The Effects of Process Variations on Performance and Robustness of Bulk CMOS and SOI Implementations of C-Elements

A Thesis Submitted for the Degree of Doctor of Philosophy in the Faculty of Engineering

by

Ziyad AL Tarawneh

School of Electrical, Electronic and Computer Engineering
Newcastle University
Newcastle upon Tyne
United Kingdom

April-2011

Abstract

Advances in semiconductor technology have been driven by the continuous demands of market forces for IC products with higher performance and greater functionality per unit area. To date industry has addressed these demands, principally, by scaling down device dimensions. However, several unintended consequences have undermined the benefits obtained from the advances in technology, firstly, the growing impact of process variations on interconnectivity delay, aggravated by the increase in the amount of interconnectivity as circuit complexity increases. Overall, the difficulty of establishing delay parameters in a circuit is adversely impacting on the attainment of the timing closure for a design. Secondly, the increase in the susceptibility of the circuits , even at ground level, to the effects of soft errors due to the reduction in supply voltages and nodal capacitances, together with the increase in the number of nodes in a circuit as the functionality per unit area increases.

The aim of this research has been to model and analyse the reliability of logic circuits with regard to the impact of process variations and soft errors, and to finds ways to minimise these effects using different process technologies such as fully depleted silicon on insulator (FDSOI) and partially depleted silicon on insulator (PDSOI) technologies, together with the implementation of different circuit architectures.

In view of the increased susceptibility of logic elements to the effects of process variations and soft errors as device geometries are reduced, a logic element which is not only widely used but also typical to asynchronous design is the Muller C-element, which can be realised in a number of different circuit configurations. The robustness of various C-element configurations implemented in different technologies with regard to the effects of process variations and soft errors was examined using the design of the experiment (DoE) and response surface (RSM) techniques. It was found that the circuits based on SOI technology were more robust compared with bulk silicon technology. On the other hand, from the circuit architecture perspective, the differential logic implementations of C-element were found to be more resilient to the effects of process variation and soft errors in comparison with the other C-element implementations investigated.

Acknowledgment

First and foremost, I would like to express my deep and sincere gratitude to my supervisor, Dr Gordon Russell, for his continuous support and encouragement throughout this research. Without his guidance as a great mentor, this work would not have been possible. I am also deeply grateful to my supervisor, Professor Alex Yakovlev. He provided me with lots of good advice and clear guidance throughout my PhD. Their wide knowledge, understanding and encouragement provided me with a good basis for the thesis. I am really indebted to them for the valuable knowledge that they gave me throughout this research.

Special thanks to all of my friends in Newcastle who have given me a wonderful time. Without you, life would have been boring.

Finally, I thank my loving parents and each member of my family for their love and support. Without their encouragement and understanding it would have been impossible for me to finish this work.

The financial support of Mutah University is gratefully acknowledged.

List of Publications

Journal, Conferences and Workshops

[1]   Z. Al Tarawneh , G. Russell, and A. Yakovlev, "An analysis of SEU robustness of C-element structures implemented in bulk CMOS and SOI technologies," in International Conference on Microelectronics (ICM), 2010, , pp. 280-283.

[2]   Z. Al Tarawneh , G. Russell, and A. Yakovlev, "An Analysis of the Effect of Process Variations on Performance of C-Element Structures Implemented in Bulk CMOS and SOI Technologies," in 2nd European Workshop on CMOS variability, Grenoble, France, 2011.

[3]   Z. Al Tarawneh , G. Russell, and A. Yakovlev, "An Analysis and Optimisation of the Robustness of C-Element Structures to the Effects of Process Variations," in 2nd European Workshop on CMOS variability, Grenoble, France, 2011.

Papers and Talks at Internal Conferences

[1]   Z. Al Tarawneh , G. Russell, and A. Yakovlev, "Statistical Variability Analysis of C-Element Circuit for Strained Fully Depleted Silicon on Insulator Technology" in Post Graduate conference, Newcastle University, 2010.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Since the invention and fabrication of the first Metal Oxide Semiconductor Field Effect Transistors (MOSFET), it has become the basic building block in digital circuit applications such as microprocessors and memories, resulting in low-cost, high performance circuits having a high functional density.

In 1965, Gordon Moore predicted that the number of transistors in an integrated circuit (IC) would approximately double every 2 years [1]. This prediction, the so-called Moore's law, then became the guiding principle for the microelectronics industry over subsequent decades, as shown in Figure.1.1.
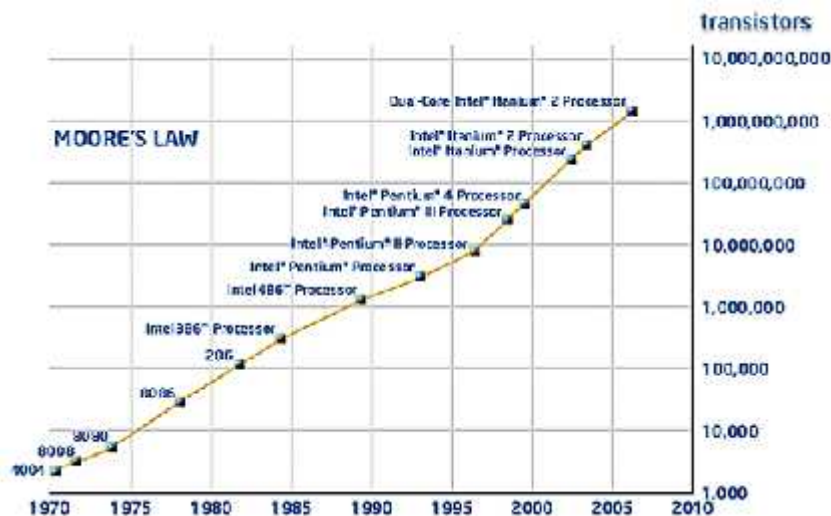
Figure 1.1: Illustration of Moore's law: the number of transistors in the different generations of Intel's microprocessors against production years [2].

Since then, semiconductor technology has advanced through the use of dimensions and voltage scaling, to create faster, lower power, and more densely packed devices, entering

the ultra large-scale integration era (ULSI). The main motive behind these trends is economic in nature, as demonstrated by Moore in 1965, According to Moore. integrated circuits and device scaling are "the cheap way to do electronics"[3] even with the large increase in the cost of manufacturing tools to fabricate nanometre scale transistors. As an example, the cost of lithographic steppers increased from $10,000 to $35 million which resulted in an increase in manufacturing costs to $2-3 billion. Nevertheless, the cost of a transistor has decreased by seven orders of magnitude during the last 40 years and it is highly likely to continue to decrease for another decade [4].

Sections 1.2 of this Chapter present a review of device scaling methods and the differences between them. Subsequent Sections 1.3.1, 1.3.2 and 1.3.3 of the Chapter present a detailed review on the challenges faced as a result of scaling. The need for new technologies and the potential of Silicon on Insulator (SOI) technology as a possible candidate to overcome the scaling issues is discussed in Section 1.4.

An outline of the contributions made in the thesis is given in Section 1.5; the Chapter finishes with an organisational roadmap of the thesis in Section 1.6.

1.2 Device Scaling

Scaling can be obtained in two basic ways: by constant-field scaling and by constant-voltage scaling. In constant-field scaling, the dimensions of the MOSFET are scaled by a factor 'S', with the aim of preserving the magnitude of the internal electric fields, especially in device channels. The power supply voltage is also scaled proportionately with the device feature sizes. On the other hand, in terms of constant-voltage scaling, the device dimensions are reduced by a factor of 'S' while the power supply is kept constant. Table 1.1 compares the effects of constant-field scaling and constant-voltage scaling on key MOSFET device parameters [5].

Constant-field scaling provides a good framework for device scaling without degrading reliability. However there are several parameters such as the thermal voltage and the energy gap of silicon material that can not be scaled with the reduced voltage and dimensions, and poses challenges in terms of device design. Another important device parameter which does not scale well is the threshold voltage. This constrains the lower limit of power supply voltage since a guard margin between the two parameters is needed

for reliable device operation. Other parameters that present a challenge to this method is the scaling of leakage current and the sub-threshold slope.

Constant-field scaling also results in the largest decrease in the power-delay product of an individual transistor. However, this requires a reduction in the power supply voltage when the minimum feature size is reduced, making it a very difficult task to scale due to the external limitations of the power supply.

One of the problems associated with constant voltage scaling is that the electric field in the channel increases as the gate length is reduced, leading to velocity saturation, mobility degradation, increased leakage currents and lower breakdown voltages which may eventually cause serious reliability problems such as hot-carrier degradation, electro-migration, and oxide breakdown.

Modifications of the constant-field and constant-voltage scaling have also been tried to overcome the high field problems. For example, in both quasi-constant scaling [5, 6] and generalised scaling [5], the device dimensions are scaled by a factor 'S' and the voltages are scaled less aggressively by a different factor. This means that the electric field is reduced and smaller compared to constant voltage scaling[5, 6].

Parameter	Symbol	Constant-field Scaling	Constant-voltage Scaling
Gate length	L	$1/S$	$1/S$
Gate width	W	$1/S$	$1/S$
Electric Field	ε	1	S
Oxide thickness	T_{ox}	$1/S$	$1/S$
Substrate doping	Na	S^2	S^2
Gate capacitance	Cg	$1/S$	$1/S$
Oxide capacitance	C_{ox}	S	S
Voltage	V	$1/S$	1
Current	I	$1/S$	S
Power	P	$1/S^2$	S
Power delay product	$P.\Delta t$	$1/S^3$	$1/S$

Table 1.1:The effects of constant field-scaling and constant-voltage scaling on key MOSFET device characteristics [5].

1.3 Scaling Challenges

Although the scaling of MOSFET devices has brought many benefits in terms of the implementation of a single chip with regard to area, performance, cost and functionality, there are still undesirable side-effects such as static power, process variability, and reliability in terms of soft errors. The following sections discuss the challenges associated with scaling in more detail.

1.3.1 Power Consumption

Since the introduction of CMOS technology in 1973, the problem of static power dissipation was noticeable, but largely minimized by adopting CMOS technology in implementing a logic gate, since it, ideally, only dissipates power as the logic circuit is in the switching mode [7, 8]. However, continuous device scaling has increased the leakage current of the device, leading to a large increase in static power dissipation which became a significant portion of the power dissipation in CMOS circuits; this is evidenced by Figure 1.2 which shows the active and standby (static) power consumption trends for Intel processors realised in various technology generations [9].
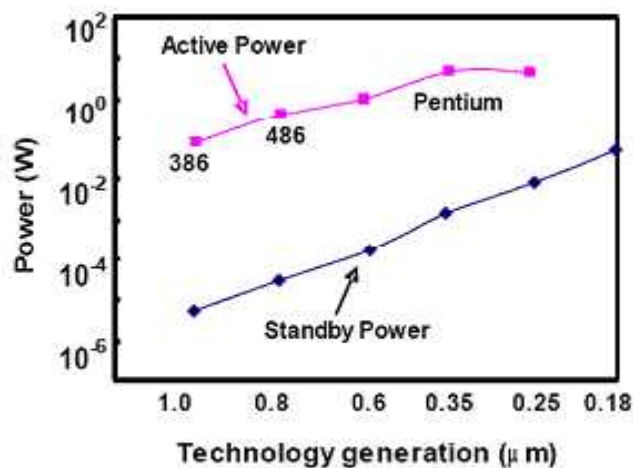
Figure 1.2 :Active and standby power trends for Intel process technologies [9].

Different leakage current mechanisms are involved in MOSFET transistors [10], but the two most dominant mechanisms are the sub-threshold leakage current and the gate

leakage current. The sub-threshold leakage current is mainly caused by the continuous reduction in the device threshold voltage to compensate for device speed loss when scaling the power supply voltage. This in turn results in an exponential increase of sub-threshold leakage current [10, 11].

The gate leakage current is relatively smaller in value compared to the sub-threshold leakage current for older technologies. However, it is expected to become comparable to the sub-threshold leakage current in deep nanometre nodes. Scaling the gate oxide into sub-nanometre thickness in the range of 20 angstroms to less than 10 angstroms is limited by quantum tunnelling effects which induce severe gate leakage currents. Alternative gate dielectrics with higher permittivity are suggested as a means of reducing leakage current and enabling continuous device scaling. However, many other challenges, including integration processes, degraded silicon channel mobility and gate oxide reliability, still need to be solved [12, 13].

Power dissipation is a major design challenge with regard to digital circuits that are based on conventional silicon CMOS technology. This needs to be addressed in order to meet the ever-growing demands for portable computing with high data rates, together with the lack of significant improvements in battery technology [7].

Supply voltage scaling has emerged as the preferable technique for reducing power consumption due to the quadratic relationship between supply voltage and dynamic power consumption [5]. However, this is associated with severe compromises such as increased circuit delay [5, 9, 14].

However, for applications such as wireless sensors, the circuit speed is generally overridden by the requirement for low power consumption. The current research into circuit designs with Si-based CMOS is struggling very hard to meet the criteria of low power consumption while achieving a high-speed of operation [15, 16].

1.3.2 Process Variation

Another important issue associated with MOSFET scaling is the large increase in process variations. With the continuous scaling of devices and interconnects, variations in key device and interconnect parameters such as device threshold voltage (V_{th}), oxide thickness (t_{ox}), wire width (W_M), and wire height (H) are growing at an alarming rate [17-

19].Subsequently, the performance of different die on the same wafer can vary widely, resulting in a significant parametric yield loss, which directly translates into higher manufacturing costs.

To demonstrate the impact of process variation on actual semiconductor products, Figure 1.3 shows the normalized distribution of the clock frequency and the static leakage current of Intel microprocessors on a single wafer [20]. It can be seen that the variations in device parameters have resulted in more than a 30% frequency spread and 20X variation in the total leakage current of the chip. The highest operating frequency chips have a wide leakage distribution while, for a given leakage current, there is a wide spread in the frequency of the chips. The highest frequency chips with a large leakage current and those low frequency chips with a reasonably high leakage current will have to be discarded, affecting the overall yield and cost.

 The large distribution of the chip operating frequency also requires frequency binning in which each chip has to be tested to find out its maximum speed and power before it can be sold in the marketplace. This is a very costly, time-consuming process. Furthermore, as static power increases as a fraction of the total power consumption, the 20X variation in leakage current results in a large increase in the variation of total power consumption. In fact, it was reported that the variation in static power can result in a variation of total power by as much as 50% [21]. As a result, a great yield loss is notably impacted on by unmitigated parameter variations (i.e. parametric yield loss).
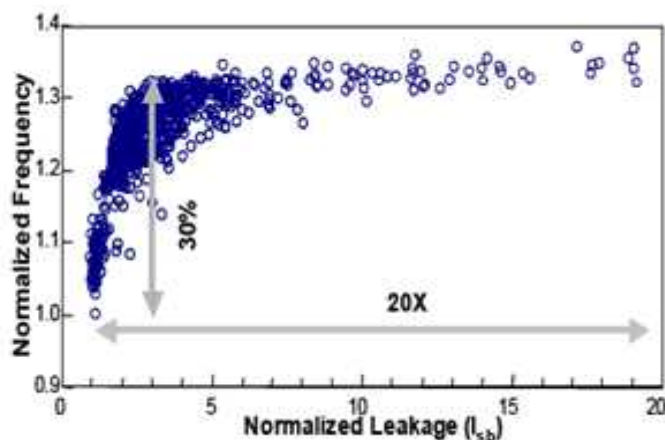
Figure 1.3: Frequency and leakage variations [20].

In general, variations in device parameters are mainly caused by (1) the limitations of the control of the manufacturing process (extrinsic causes of variations) and (b) fundamental atomic-scale fluctuations in the scaled MOSFET (intrinsic causes of variations) [17].

The variations due to the limited controllability of the manufacturing process have become extremely difficult to manage. This is because of the inability of the semiconductor industry to improve manufacturing process tolerances as device dimensions are scaled down [17]. For example, the patterning wave length (λ=193 nm), which was adopted in 130nm technology node, is still used in the 65nm and even below this technology node as can be seen in Figure 1.4. As a result, it is becoming progressively more difficult to control the channel length of transistors using technology scaling [17].
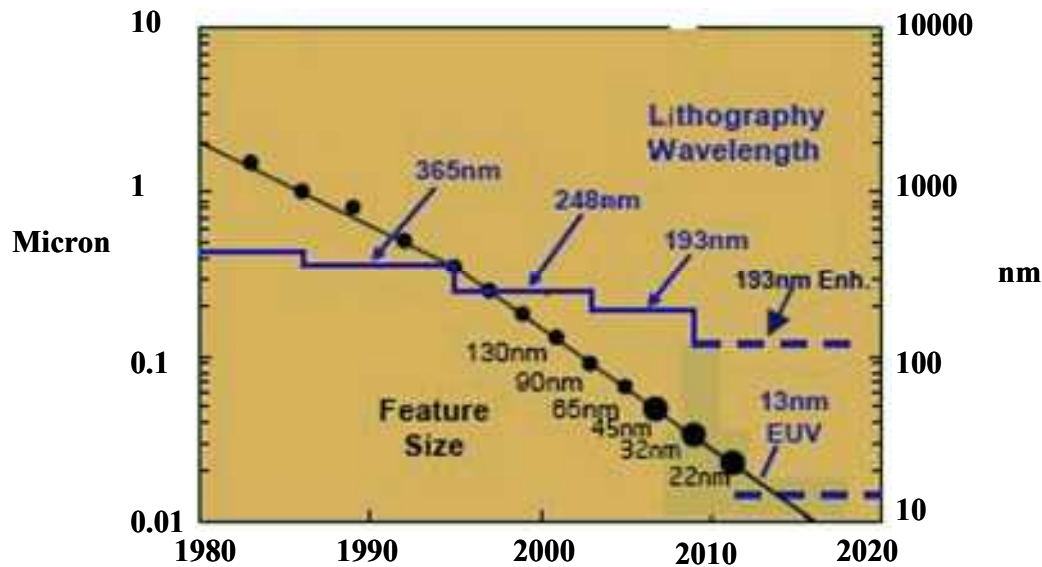
Figure 1.4: Exposure wavelength (λ) versus technology node generations [22].

The intrinsic causes of variations are also becoming a major problematic source of variation, especially in future technologies since device dimensions are reaching a scale involving silicon atomic distances. Therefore, a microscopic variation in the silicon structure has a large impact on the performance of the device [23]. For example, the threshold voltage of a MOSFET transistor relies profoundly on the dopant distribution and the density in the channel region. As a result, the threshold voltage of transistors also becomes a random variable [24].

1.3.3 Single Event Upset

The continuous scaling of device features has also led to a huge reduction in the node capacitances of integrated circuits and subsequently the charge stored. As a result, the reliability of circuits is reduced in terms of the energetic particles strikes that induce soft errors. In past technologies, this problem was limited to hostile environment applications such as space; however the effects are now being observed at ground level. Moreover the increased clock frequency and reduced supply of voltage requirements also aggravate the tolerance of the circuits to radiation induced soft errors [25, 26]. It is also worth noting that the occurrence of low energy particles is much higher than that of high energy particles. Soft errors are a major threat in critical applications where reliability is the central concern over performance and cost such as in biomedical or avionics applications. Figure 1.5 compares the critical charge needed to cause a soft error in the SRAM and the different combinational logic circuits. From Figure 1.5 it can be seen that the critical charge of SRAM cells and logic circuits has dramatically decreased as technology scales. However, the critical charge of logic circuits has fallen at a faster rate, making their robustness to soft error a significant concern. Logic circuits are also more difficult to harden than SRAMs. This is because most of the mitigation techniques have come with large area overhead and latency penalties.
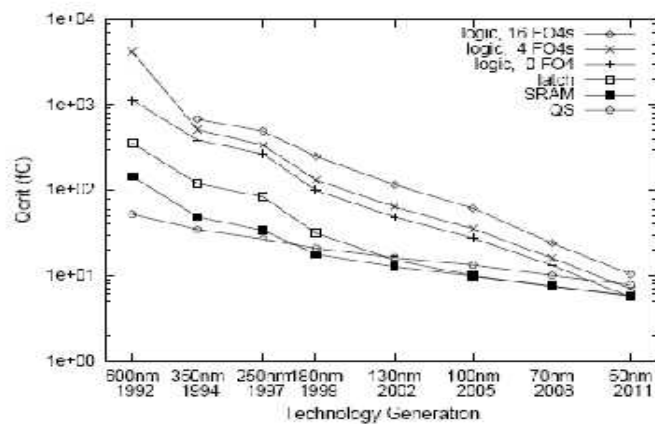
Figure 1.5 : Critical charge for SRAM, latch and logic circuits [27].

1.4 Need for New Technology

All of the above-mentioned issues regarding device scaling has resulted in the unpredictable behaviour of circuits and has severely degraded the reliability of digital systems. The widespread use of modern VLSI systems has necessitated addressing these issues during the system design phase, in order to improve system reliability and resilience to radiation and process variations.

It is widely accepted on the part of the industrial community that device scaling cannot be sustained for ever, and hence other approaches are needed to overcome the foreseeable barriers associated with the reduction in device dimensions [15, 28]. Therefore new device materials and circuit architectures are necessary to provide more solutions for these concerns, and also to enable scaling for future electronic systems.

In this regard, silicon-on-insulator (SOI) MOSFET is a competitive candidate in order to replace bulk silicon technology, especially for nanoscale circuit applications where the requirements of low power and high-speed digital applications can be met [29]. This is mainly because it provides a way to mitigate the devastating short channel effects in bulk CMOS scaling. Further benefits of the SOI technology include higher drive currents and hence produce smaller delays; the presence of the buried oxide insulation layer minimises the leakage of current from the drain-source junctions to the substrate, and subsequently reducing the static power consumption in the circuits.

In addition, the structure of an SOI device is quit similar to that of bulk CMOS, hence the fabrication process steps needed are almost identical to the bulk CMOS process [29].Consequently, it can take advantage of the recent advances in manufacturing process technology which have been introduced into bulk CMOS, such as strain silicon techniques and high-k dielectrics. All of these render SOI technology as be a suitable alternative for replacing the current silicon technology.

1.5 Motivation and Research Goals

In this chapter, we have provided an overview of scaling which is the predominant mechanism that has been employed in the semiconductor industry in the last few decades to improve device performance, increase the functionality per unit area. The challenges associated with scaling such as static power, process variability, and reliability in terms of soft errors are discussed. We have emphasised that the process variations in semiconductor manufacturing has become more of a concern and is seen to be a generic challenge to all deep submicron devices which reduces the yield of circuits. Moreover, scaling has undermined the immunity of circuits to energetic radiation particle strikes.

The need for a new technology to overcome these barriers is highlighted and, in this regard, SOI technology is proposed as a potential candidate to overcome the variability and the reliability challenges.

The main goals and contributions of this thesis are outlined as follows:

- In order to bridge the gap between the manufacturer and the designer, models for device and circuit performance parameters as a function of process variation parameters have been build using statistical techniques such as Design of Experiments (DoE) and Response Surface Modelling (RSM).

- The exploration of SOI technology as a possible solution to overcoming the performance barrier in terms of process variations at device and circuit levels.

- The investigation of the impact of circuit architecture as a way of reducing the impact of process variations on circuit-level design metrics. In this work, various C-element circuits were considered and were also used as test vehicles due to their importance in asynchronous circuit design.

- The proposal of a new method to identify the most dominant devices in a circuit that have the greatest influence on the circuit performance under process and operating condition variations.

- A demonstration of the ability of the proposed technique to reduce the impact of process variation on the variability of circuit performance metrics.

- The analysis the robustness of various C-element implementations towards soft errors in terms of key circuit design parameters.

- The exploration of the efficiency of SOI technology over bulk silicon in terms of improving circuit reliability in terms of single event upsets (SEUs).

1.6 Thesis Organisation

The subsequent chapters in this thesis describe in detail the goals outlined in the previous section.

Chapter 2 offers a review of the sources and different types of variation including inter-die, intra-die, systematic and random variations. The impact of these variations is explored and how they propagate and affect the performance at both device and circuit levels. Methodologies and approaches used for modelling variability such as the Worst Case-corner Analysis method, Monte Carlo technique, Sensitivity Analysis approach and Statistical Static Timing analysis are discussed.

To overcome the effects of process variations and their impact on overall circuit performance and yield, the need for new device materials and circuit architecture is highlighted. Silicon on Insulator (SOI) technology, in this regard, is discussed as a possible candidate. The advantages of SOI over current bulk silicon technology in terms of performance and scaling capability are discussed. Thereafter, several different SOI devices namely fully depleted FDSOI and partially depleted PDSOI, and their structural and behavioural differences are reviewed.

In Chapter 3 there is an introduction to the Design of Experiments (DoE) and Response Surface Modelling (RSM) statistical techniques. Different experimental designs such as full-factorial design, fractional factorial design, screening design and response surface designs are described. These statistical techniques were utilized to explore the potential of FDSOI and PDSOI along with bulk silicon in terms of process variations. The TCAD process and device simulations in this work were carried out using SPROCESS and SDEVICE [30] for the calibration of the device characteristics to the experimental results, and AURORA [31] (a parameter extractor) was used to extract compact models for the technologies investigated. First, eighteen process parameters arising from different process steps were identified as the important and uncontrolled sources of variation. The DoE-based screening experiment was then used to obtain the most significant process

parameters influencing the key set of the device compact model parameters such as threshold voltage for both NMOS and PMOS devices. RSM was then performed on these significant parameters and second order models for threshold voltage for each technology were developed as a function of the process parameters. Model validation was performed using a residual-analysis test and R-squared statistics were generated for all the models generated.

In Chapter 4, the response surface methodology is extended to the circuit-level, studying the impact of process variations and environmental operating conditions on various C-element circuits. The extracted variations in threshold voltage of both NMOS and PMOS devices realised on the technologies investigated, were considered along with power supply voltage and gate length variations. Subsequently, RS models were developed and used to perform response surface analysis for circuit performance metrics such as power and delay, to give useful information about the sensitivity of each metric with respect to the process parameters under consideration. The influence of circuit architectures on performance variation in terms of dynamic and static power consumptions and circuit delay was also investigated. A relative comparison between circuits based on FDSOI, PDSOI and bulk silicon in terms of delay and power was subsequently undertaken.

An in-depth analysis of the behaviour of the circuits under process variations was demonstrated by studying the impact of process variations on transistor channel currents. The investigation was carried out by measuring the maximum current in each transistor under process variation during the circuit operation. Using this method, it was possible to first, explain why the circuit topology plays a significant role in performance variation; second to identify which transistors make the most significant contribution to the final performance variation; and third to target those transistors to minimise their impact on performance variability by resizing their gate widths.

In Chapter 5, several aspects related to the single event upsets (SEUs) and their impact on VLSI circuits are reviewed, namely the charge deposition and the collection mechanisms involved, and the impact of technology scaling on the radiation tolerance of VLSI circuits. Thereafter, the impact of key design parameters such as threshold voltage, power supply and width ratio on the resilience of various C-element circuits is investigated. The RS technique is utilised to give useful measures with regard to the sensitivity of the resilience

of the circuit to radiation particle strikes with respect to the key parameters under consideration.

The use of different materials and device structures, such as FDSOI and PDSOI technologies, in order to reduce the impact of radiation particle strikes at device level is explored. A comparative study of soft error analysis of different C-element circuits is included, while the effectiveness of the differential logic with an inverter latch implementation of the C-element in terms of the prevention of soft errors is highlighted.

Finally, Chapter 6 concludes the thesis with a summary of the results and possible directions for future work in this area.

1.7 References

[1]   G. E. Moore, "Progress in digital integrated electronics," IEDM Tech. Dig, pp. 11-13, 1975.

[2]   Moore's Law, The Future-Technology & Research at Intel [Online]. Available: http://www.intel.com/technology/mooreslaw/index.htm

[3]   L. Patler, "Trend Smart: The 21 Trends That Will Change the Way You Do Business ": Sourcebooks Sep 2004

[4]   S. E. Thompson and S. Parthasarathy, "Moore's law: the future of Si microelectronics," Materials Today, vol. 9, pp. 20-25, 2006.

[5]   G. Baccarani, M. R. Wordeman, and R. H. Dennard, "Generalized scaling theory and its application to a 1/4 micrometer MOSFET design," IEEE Transactions on Electron Devices, vol. 31, pp. 452-462, 1984.

[6]   P. K. Chatterjee, W. R. Hunter, T. C. Holloway, and Y. T. Lin, "The impact of scaling laws on the choice of n-channel or p-channel for MOS VLSI," Electron Device Letters, IEEE, vol. 1, pp. 220-223, 1980.

[7]   G. Timp, A. Agarwal, F. H. Baumann, et.al , "Low leakage, ultra-thin gate oxides for extremely high performance sub-100 nm nMOSFETs," International Electron Devices Meeting, pp. 930-932, 1997.

[8]   A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low-power CMOS digital design," IEEE Journal of Solid-State Circuits, vol. 27, pp. 473-484, 1992.

[9]   S. Thompson, P. Packan, and M. Bohr, "MOS Scaling: Transistor Challenges for the 21st Century," Intel Tech. Journal Q3'98. Portland Technology Development, Intel Corp.

[10]   M. Anis, "Subthreshold leakage current: challenges and solutions," Proceedings of the 15th International Conference on Microelectronics, ICM 2003, pp. 77-80, 2003.

[11]   R. M. Rao, J. L. Burns, and R. B. Brown, "Circuit techniques for gate and sub-threshold leakage minimization in future CMOS technologies," Proceedings of the 29th European Solid-State Circuits Conference, pp. 313-316, 2003.

[12]   T. Yuan, D. A. Buchanan, C. Wei, D. J. Frank, K. E. Ismail, L. Shih-Hsien, G. A. Sai-Halasz, R. G. Viswanathan, H. J. C. Wann, S. J. Wind, and W. Hon-Sum, "CMOS scaling into the nanometer regime," Proceedings in the IEEE, vol. 85, pp. 486-504, 1997.

[13]   K. Dim-Lee, "CMOS integration issues with high-k gate stack," in Proceedings of the 11th International Symposium on the Physical and Failure Analysis of Integrated Circuits, IPFA 2004, pp. 17-20., 2004.

[14]   G. E. Moore, "Cramming More Components Onto Integrated Circuits," Proceedings of the IEEE, pp. 82-85, 1998.

[15]   T. Skotnicki, J. A. Hutchby, K. Tsu-Jae, H. S. P. Wong, and F. Boeuf, "The end of CMOS scaling: toward the introduction of new materials and structural changes to improve MOSFET performance," Circuits and Devices Magazine, IEEE, vol. 21, pp. 16-26, 2005.

[16]   International Technology Roadmap for Semiconductors, http://public.itrs.net, 2007 Ed.

[17] M. Orshansky, S. R. Nassif, and D. Boning, "Design for manufacturability and statistical design:A constructive approach," US Springer, 2008.

[18] K. Agarwal and S. Nassif, "Characterizing Process Variation in Nanometer CMOS," in Design Automation Conference, 2007. DAC '07. 44th ACM/IEEE, pp. 396-399, 2007.

[19] K. Bernstein, D. J. Frank, A. E. Gattiker, W. Haensch, B. L. Ji, S. R. Nassif, E. J. Nowak, D. J. Pearson, and N. J. Rohrer, "High-performance CMOS variability in the 65-nm regime and beyond," IBM Journal of Research and Development, vol. 50, pp. 433-449, 2006.

[20] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," Proceedings of Design Automation Conference, pp. 338-342, 2003.

[21] S. Borkar, "Designing reliable systems from unreliable components: the challenges of transistor variability and degradation," Micro, IEEE, vol. 25, pp. 10-16, 2005.

[22] Computational Lithography Technology that makes Technology Affordable Lithography.Available:http://download.intel.com/pressroom/kits/research/comput ational_litho_poster.pdf.

[23] A. Asenov, "Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 um MOSFET's: A 3-D "atomistic" simulation study," IEEE Transactions on Electron Devices, vol. 45, pp. 2505-2513, 1998.

[24] H. Mahmoodi, S. Mukhopadhyay, and K. Roy, "Estimation of delay variations due to random-dopant fluctuations in nanoscale CMOS circuits," IEEE Journal of Solid State Circuits, vol. 40, pp. 1787-1796, Sep 2005.

[25] R. Baumann, "Soft errors in advanced computer systems," Design & Test of Computers, IEEE, vol. 22, pp. 258-266, 2005.

[26] P. Hazucha, T. Karnik, J. Maiz, S. Walstra, B. Bloechel, J. Tschanz, G. Dermer, S. Hareland, P. Armstrong, and S. Borkar, "Neutron soft error rate measurements in a 90-nm CMOS process and scaling trends in SRAM from 0.25um to 90-nm generation," in IEEE International Electron Devices Meeting, 2003. IEDM '03 Technical Digest, pp. 21.5.1-21.5.4, 2003.

[27] P. Shivakumar, M. Kistler, S. W. Keckler, D. Burger, and L. Alvisi, "Modeling the effect of technology trends on the soft error rate of combinational logic," Proceedings. International Conference on Dependable Systems and Networks. DSN 2002, pp. 389-398, 2002.

[28] G. E. Moore, "No exponential is forever: but "Forever" can be delayed! ," in IEEE International Solid-State Circuits Conference, Digest of Technical Papers. ISSCC2003, vol.1, pp. 20-23, 2003.

[29] K. Bernstein and N. J. Rohrer, "SOI circuit design concepts," Kluwer Academic Publishers, 2000.

[30] Sentaurus TCAD, Release z-2007.12 ed., Synopsys, Mountain View, CA, USA, 2007.

[31] AURORA user guide: AURORA simulation program. Sunnyvale: Synopsys Inc., 2006.

Chapter 2

Process Variations and Variability Modelling

2.1 Background

In this chapter a detailed review is given of the sources and different types of process variations, including inter–die and intra-die. Systematic and random variations are introduced in Section 2.2 and the impact of these variations on the device and circuit levels is explored in Section 2.3. In the latter section the different general approaches and methodologies used to understand and address process variations are also examined, pointing out their strengths and weaknesses. The traditional techniques, such as Worst Case Analysis, Monte Carlo technique, as well as a brief introduction to DoE and RSM, Sensitivity Analysis and the Statistical Static Timing Analysis technique are given in Section 2.4.

Finally, the potential of SOI technology over bulk silicon is reviewed in Section 2.5 as a possible candidate for future MOSFET devices in order to overcome the increased effect of process variations due to scaling and their impact on circuit performance and yield. Thereafter, several different SOI devices, namely PDSOI and FDSOI and their structural and behavioural differences are reviewed, followed by the challenges associated with SOI technology.

2.2 Process Variations: An Introduction

Although the topic of process variation is sometimes considered as a new challenge associated with technology scaling, the problem of variation has been addressed for more than 40 years. In 1961, Shockley studied and analysed the random fluctuation in junction breakdown [1]. In 1974, Schemmert and Zimmer presented a number of process

parameters which can be tuned and optimised in order to minimise the threshold voltage sensitivity of a transistor, pointing to the systematic variations in transistor threshold voltage [2].

With technology moving towards the nanometre regime, process variability due to the fabrication steps or random variations from dopant fluctuations becomes an increasingly critical concern [3-5]. Finding ways to minimise the impact of the problem of these variations will remain a major challenge for future technology nodes. The contribution of fabrication process steps dominates the electrical parameter variations of a device with aggressive device scaling such as oxidation, ion implantation, lithography and chemical-mechanical planarization (CMP). Moreover, the effects of random variations in circuit operating conditions such as the temperature and the power supply voltage (V_{dd}) increases dramatically as the circuit clock frequency increases [5]. This has led to significant variations in the circuit performance and increased yield degradation which affects production costs.

The performance of a circuit is basically governed by the linear and non-linear electrical behaviour of its individual devices. Variations in electrical characteristics of these devices make the performance of the circuit deviate from its intended values and can cause performance degradation and even erroneous functionality.

Figure 2.1 shows the level of abstraction in a system design and the corresponding variability factors which will subsequently impact on the overall system performance metrics ultimately affecting yield and cost of the product.

The physical deviation of manufacturing processes such as implantation dose and energy cause a variation in device structure and doping profile. These variations together with the environmental variation sources affect the electrical behaviour of device and eventually result in performance metric variations of the circuit. All of these variation sources ultimately propagate their negative influence on the overall performance of a system manufactured on a chip (SoC).
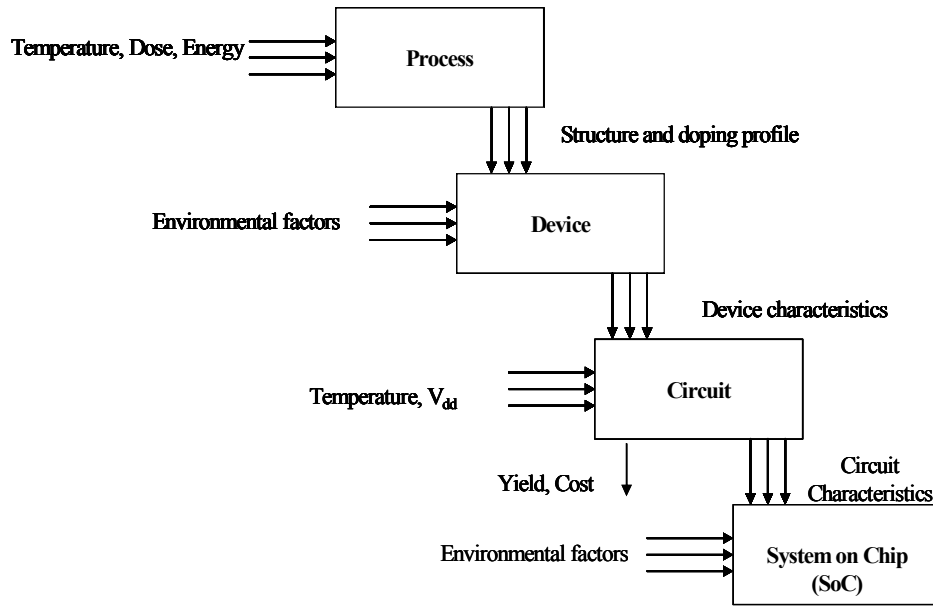
Figure 2.1: An illustration of the effect of the process and environmental factors on the device, circuit and system on chip (SoC) [6].

According to ITRS [7], as the device dimensions are scaled down together with novel process steps added to improve performance in the deep sub-micron devices, the number of process variants affecting the device and circuit performance increases. The variation is also expected to worsen for the sub-65nm technology nodes.

Table 2.1 demonstrates the relative variation of thresholds of NMOS devices for different technology nodes.

L(nm)	Nominal V_{th}(mV)	Relative change from the nominal (%)
250	450	4.7
189	400	5.8
130	330	8.2
90	300	9.3
65	280	10.7
45	200	16

Table 2.1 : Impact of process variation on threshold voltage for different technology nodes [8].

The variations will strongly impact the performance metrics of a circuit, such as delay, dynamic power and static power consumptions, which may exhibit greater variability leading to the degradation of yield, increased cost and time to market in modern technologies and their applications.

From a circuit designer perspective, the process variations can be classified into two broad types [8, 9]: inter-die variations and intra-die variations. In the following sections a detailed distinction between these sources of process variation will be discussed.

2.2.1 Inter-die Variations

Inter-die (i.e., die-to-die referred to as D2D) variations refer to the deviation of a device or circuit parameter that has the same value across a single die, and hence can represent the variations that occur from die-to-die, wafer-to-wafer and lot-to-lot, as shown in Figure 2.2. The inter-die variations of a process parameter can be expressed as a random variable as given in Equation 2.1 [10].

$$P = P_{nom} + \Delta P_{inter} \qquad (2.1)$$

where P_{nom} is the nominal value of the process parameter under consideration and P_{inter} is a random variable with a zero mean value and usually represented by a Gaussian distribution with a given standard deviation; P_{inter} has a single value for all components on the die.
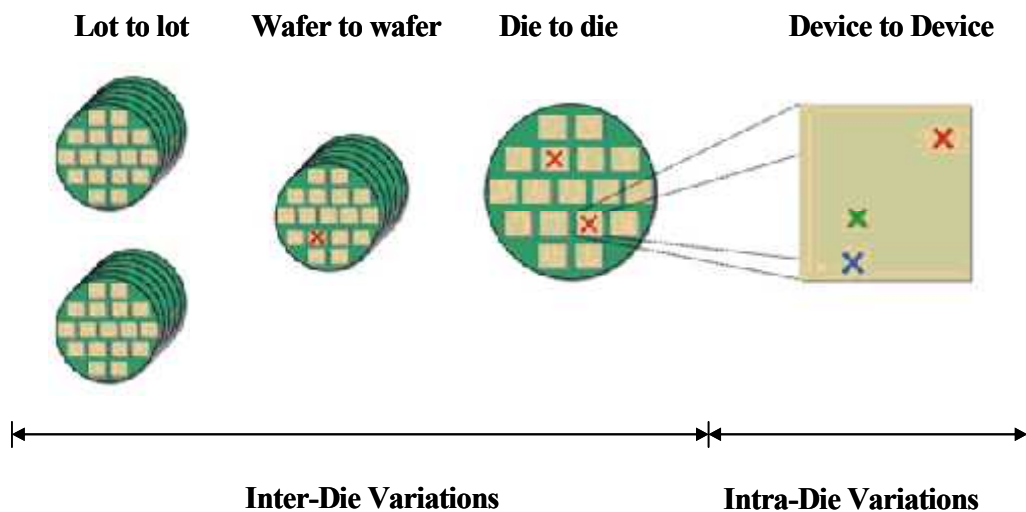
Figure 2.2: An illustration of inter-die and intra-die variations [6].

In circuit design, the inter-die variation is regarded as a shift in the mean or expected value of a parameter equally across all devices on any one die. For example, the inter-die variation in a parameter such as threshold voltage (V_{th}) changes its value in the same direction (increasing or decreasing as shown in Figure 2.3) for all transistors in a die leading to a variation in the performance metrics of a circuit such as delay and leakage current. Thus, the inter-die variation does not cause a mismatch between different transistors in a die.
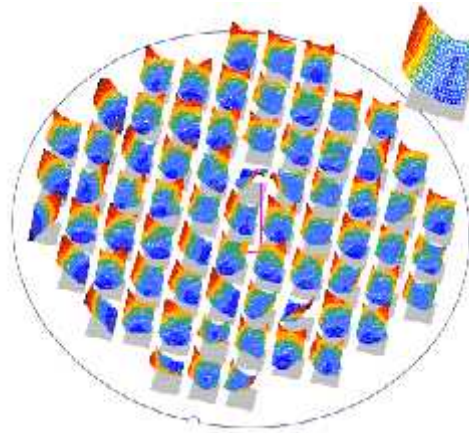
Figure 2.3: An example of the die-to-die distribution of NMOS Ids variability (systematic component) [11].

At design time it is usually assumed that each inter-die variation factor is due to different physical and independent sources [8, 10], since the circuit designer typically has no knowledge about where the chip will be placed on the wafer.

Examples of the major sources of the inter-die variation is due to materials and gas flow variation (linear variation) or due the wafer spin process and exposure time (radial variation) variations [8].

2.2.2 Intra-die Variations

Intra-die variations are those occur within a single die and cause device parameters to deviate from their intended or designed values across different locations in the die (i.e. spatially correlated). Therefore the intra-die variations of a process parameter can be expressed as a random variable as given in Equations 2.2 and 2.3 [10].

$$P = P_{nom} + \Delta P_{inter} + \Delta P_{intra}(x_i, y_i) \qquad\qquad (2.2)$$

$$P = P_{nom} + \Delta P_{inter} + \Delta P_{spatial}(x_i, y_i) + \Delta P_{random,i} \qquad (2.3)$$

where $\Delta P_{spatial}$ represents intra-die variation that consists of a spatially correlated component which is a function of the location on the die and the ΔP_{random} represents a random component which has no correlation with the other devices and normally can be considered as a single random variable for each device.

Historically, the effect of intra-die variations was negligible in the older technology generations. However, as technology moves towards a nanometre regime, it has become noticeably comparable to, and in some cases larger than, inter-die variations. As an example for the 130 nm technology node, these variations add up to around 30% of the overall performance variations [8].

There are two major sources of intra-die variations in terms of their origin, namely wafer level variations and layout-dependent variations [9]. The wafer level variations originate due to effects such as lens aberrations and result in bowl-shaped or other known distributions over the entire reticle [10]. Therefore, it can result in small trends which represent the spatial range across the die.

While the layout-dependent or die-pattern variations are due to lithographic and etching techniques used during process fabrication including process steps such as Chemical Mechanical Polishing (CMP) and Optical Proximity Correction (OPC), these dependencies create additional variations which have become a major threat in circuit fabrication. For example, two interconnected lines designed identically in different parts of the same die may result in lines with different widths, due to photolithographic interactions and plasma etch micro-loading [9, 10].

Intra-die variations have two main contributors: systematic and random variations [12]. This distinction is extremely useful for circuit designers because the impact of systematic variations can be minimised and even removed by adapting a suitable circuit design while random variations will inevitably impact the design margins in a negative manner and need to be thoroughly investigated.

2.2.3 Systematic and Random Variations

The systematic variations are mainly introduced during different process fabrication steps and they are caused by some known predictable phenomena during the manufacturing of the device. Gate length variability is an example of a systematic variation, which results in a systematic shift of their values across a reticle due to effects such as changes in the stepper-induced illumination and imaging non-uniformity due to lens irregularities [13]. Systematic variations are also considered as correlated and deterministic variations. For two transistors in close spatial proximity, these variations do not result in large differences in the transistors electrical characteristics [13-16].

On the other hand, random variations are due to the random and unpredictable phenomena in the semiconductor fabrication process such as channel doping fluctuation. The random variations are difficult to characterise and can cause a significant mismatch between the adjacent transistors [9, 14].

Random variations in some processes or environmental parameters can have spatial correlations, whereby parameter fluctuations of an individual transistor in a die are remarkably close in nature to those in spatially adjacent transistors but may vary significantly from the transistors that are far away [13]. Since the compensation of random variations is difficult, these variations pose a great challenge to achieving ultra Deep Sub-Micron (DSM) CMOS circuit design with an acceptable yields [8].

As the technology approaches the nanometre scale, the number of dopant atoms becomes less and small variations in their number and actual location can result in a large variation in device performance. These variations are true random variations with no correlation across devices and represent one source of intra-die random variations. Figure 2.4 shows the impact of location variations of 170 dopant atoms in a device channel on the threshold voltage of the MOSFET.

Such random variations can also result from a group of other sources, such as lithography, etching, Chemical Mechanical Polishing (CMP) etc. Despite the fact that their impact is small, it is likely to grow as process parameters are scaled down.
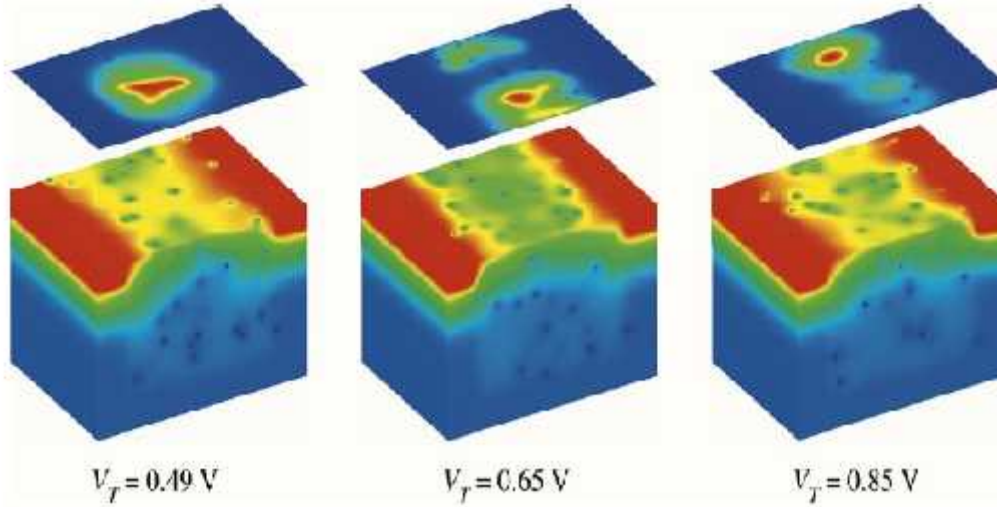
Figure 2.4: Random dopant location-induced threshold voltage fluctuation (the devices have 170 dopant atoms in the channel of the device) [17].

2.3 The Impact of Process Variations

2.3.1 Impact of Process Variations on Transistor Parameters

In a digital integrated circuit, a transistor is normally used either to charge or discharge a capacitive load (C) and the required time determines the performance of the transistor as given in Equation 2.4. For simplicity, we use the saturation current equation for a MOSFET as shown in Equation 2.5, where μ is the mobility of a charge carrier through the channel of the device, C_{ox} is the gate oxide capacitance, W and L are respectively the width and length of the transistor, V_{th} is the device threshold voltage, V_{GS} is the voltage bias between gate and source and α is the velocity saturation index where alpha is typically around 1.3 for short channel devices. Despite the fact that this equation is idealised and neglects important details in modern transistors, it is sufficient to illustrate the impacts of the main variation sources on key transistor parameters.

$$ t_{pd} = \frac{CV_{dd}}{I_D} \propto \frac{V_{dd}}{(V_{dd} - V_{th})^{\alpha}} \qquad (2.4) $$

$$ I_D \propto \mu \frac{WC_{ox}}{L} (V_{GS} - V_T)^{\alpha} \qquad (2.5) $$

Table 2.2 shows the MOSFET parameters and relevant process steps that directly influence each of those parameters. It is clear that a single process step can affect multiple transistor parameters, and thus decoupling the effects of one variation source from another is extremely difficult.

Device parameter	Relevant process step
μ	Ion implantation, diffusion , annealing, stress
C_{ox}	Gate oxide formation
W,L	Etching, lithography
V_{th}	Ion implantation, gate oxidisation, annealing, etching, lithography

Table 2.2: MOSFET parameters and their relevant process steps.

The effects of variations of the parameters shown in Table 2.2 on key electrical parameters of a device are discussed in the following sections.

2.3.1.1 Carrier Mobility (μ)

Mobility refers to the ability of the carriers (electrons or holes) to travel through the channel of a MOSFET in response to an applied electric field. It can be mathematically expressed as in Equation 2.6.

$$\mu = \frac{q\tau}{m^*} \quad\quad\quad (2.6)$$

where q is the electron charge, τ is the mean free time between carrier collisions and m is the effective mass of carriers (electron and hole). The mobility of carriers in the channel of a MOSFET device is also given as a function of the doping concentration as shown in Figure 2.5 since the doping concentration determines the mean free time between collisions (τ), and to a less significant degree, the effective mass (m^*).
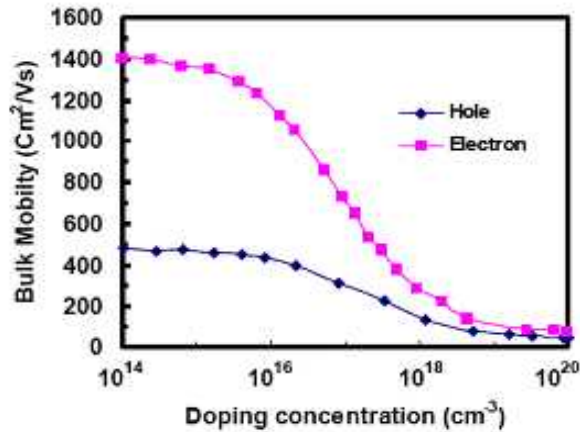
Figure 2.5 : Electron and hole mobility versus doping density for bulk Si [18].

However, in modern technology processes, strain engineering of a device channel, either by using local techniques such as nitride liners (contact-etch-stop-liner) and silicon germanium in source/drains, or global techniques such as SiGe virtual substrates, also affect the device mobility by either stretching or compressing the silicon lattice which in turn reduces the effective mass of a particular charge carrier [32].

Therefore, the mobility of carriers in the device channel is mainly determined by any manufacturing process step that affects doping concentration and the magnitude of stress in the channel. In this regard, the dose and energy of ion implantation and annealing temperature directly influence mobility since these process steps primarily determine doping concentrations.

Intentional and unintentional stresses in the channel, whether by introduced stress engineering techniques or due to proximity to Shallow Trench Isolation (STI), can have significant impacts on MOSFET mobility. Mobility enhancement greater than 10% over unstrained silicon has been achieved [19]. Even unintentional stresses such that are induced due to STI can cause intra-die mobility variations of a few percent depending on the transistor distance to the STI edge [20].

Recent reports based on the characterisation of mobility in advanced processes indicates a relatively large variation (σ_μ/μ_μ =21%), where σ_μ/μ_μ represents the coefficient of mobility variation, and may be due to fluctuations in the intentional stresses introduced in

these processes and layout variations [21]. It is also worth mentioning that the interface mobility of carriers is affected by the nature of the adjacent layers and surface roughness such as gate oxide. This results in the reduction of mobility and adds another source of variations that is directly reflected to the mobility of the carriers [2].

2.3.1.2 Gate Oxide Capacitance

Gate Oxide Capacitance (C_{ox}) is the capacitance that is formed by the silicon oxide between the polysilicon in the gate stack and the channel of the MOSFET. Equation 2.7 shows that the oxide capacitance is determined only of the oxide thickness (t_{ox}) and the dielectric constant of silicon dioxide or other gate insulator.

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \qquad (2.7)$$

The formation of gate oxide using thermal growth of silicon dioxide or silicon nitride is a relatively well-controlled process step during device fabrication.

However, with gate oxide, thicknesses are scaled to atomic level on the order of five atomic layers (10˚A). A small change in one atomic layer can greatly impact on not only the oxide capacitance, but also the threshold voltage and mobility of the MOSFET device [22]. A small variation in the thickness of just one atomic layer would result in a 20% variation in the gate oxide thickness. Controlling this variation becomes extremely difficult due to physical limitations at this atomic scale.

The effective gate capacitance is inversely proportional to the oxide thickness; therefore any change in this is directly translated to a variation in the drive current of the device. In addition, the gate leakage current is exponentially dependent on the oxide thickness by the Fowler-Nordheim tunnelling mechanism [23-25]. As a result, the variation in oxide thickness can have a catastrophic impact on the static power in modern process technologies.

At the 65nm node and beyond, the gate leakage current can become comparable to or even greater than the channel leakage current. Moreover, NMOS transistors have more gate leakage than PMOS transistors because the effective mass of electrons is much less than that of holes, thus making the probability of tunnelling in NMOS devices much higher. One solution to improve this situation and minimise gate leakage has been

implemented by Intel. They have recently started to use a "high-k" gate dielectric, hafnium dioxide (HfO_2), to allow using thicker gate oxides while maintaining oxide capacitance and gate control over the channel, but reducing gate leakage by three orders of magnitude [26, 27]. Moving to a new gate oxide material not only reduces gate leakage currents but also reduces the impact of variability on C_{ox} due to the much larger physical oxide thickness. However, variations in the oxide of "high-k" stacks interfaces are still problematic and can also affect performance [28].

2.3.1.3 Threshold Voltage

The threshold voltage of a MOSFET is defined as the gate-to-source voltage (V_{GS}) that is required to form an inversion layer in the channel at the interface between the gate oxide and the silicon surface under the gate, thus allowing a current to flow from the source to drain terminals of the transistor. The threshold voltage is one of the key device parameters in CMOS technology, since it allows transistors to act like switches and hence is a suitable device to perform logic operations.

As shown in Equations 2.4 and 2.5, threshold voltage plays a major role in device performance metrics. Not only does it determine the 'on' state current of a transistor, but also it has an exponential impact on the leakage current. Threshold voltage variation has therefore always received a great deal of attention in the circuit design community.

Threshold voltage is also one of the most difficult transistor parameters to control for a number of reasons. Firstly, as shown from Equations 2.8 and 2.9, the threshold voltage is determined only by the doping concentration (N_a) and the oxide capacitance (C_{ox}), therefore it is clear that the interaction of many processing steps can largely contribute to the variability in threshold voltage. Secondly, due to the aggressive scaling of device features, the variation behaviour of threshold voltage has become mostly random due to Random Dopant Fluctuation (RDF) in the ion implantation and thermal annealing steps. As a result, it is incredibly difficult to develop mitigation techniques to manage or reduce the variation in threshold voltage. These along with the process step variations means the threshold voltage becomes the least-controlled key parameter and accounts for 30% of the sources of variation in circuit performance [22].

$$\phi_f = 2 \frac{kT}{q} \ln \frac{N_a}{n_i} \qquad (2.8)$$

where ϕ_f is the Fermi potential, N_a is the doping concentration, n_i is the intrinsic carrier concentration in Si, T is the absolute temperature, k is Boltzman's constant and q is the electron charge.

$$V_{th} = V_{fb} + 2\phi_f + \frac{\sqrt{2 qN_a \varepsilon_s}}{C_{ox}} (\sqrt{2 \phi_f + V_{SB}} - \sqrt{2 \phi_f}) \qquad (2.9)$$

where V_{fb} is the flat-band voltage; V_{SB} is the source-body voltage; ϕ_f is the Fermi potential, C_{ox} is the oxide capacitance and ε_s is the permittivity constant of silicon.

Pelgrom et al. reported that the threshold variations are also dependent on the device area [29] (i.e. gate length and width) and also proportional to the separation distance between two adjacent devices (D) as shown in Equation 2.10, which is called Pelgrom's model. Therefore, as the device dimensions are reduced, the control of threshold voltage becomes more and more difficult.

$$\sigma_{Vth}^2 = \frac{A_{Vth}^2}{W.L} + S_{Vth}^2 D \qquad (2.10)$$

where A_{Vth} is the area proportionality constant and S_{Vth} is the distance coefficient that has to be determined through measurements.

Most recently, Asenov et al. refined Pelgrom's model based on numerical simulations and formulated an empirical model of threshold voltage variations as a function of the fundamental transistor parameters such as doping concentration (N_a), oxide thickness (t_{ox}) and transistor dimensions [42]. The model is described by Equation 2.11 [17, 30]:

$$\sigma_{Vth}^2 = 3.19 \times 10^{-8} \frac{t_{OX} N_a^{0.4}}{\sqrt{W.L}} \qquad (2.11)$$

Threshold voltage variation significantly limits the ability to design accurate analogue circuits such as amplifiers or comparators, which generally necessitate good device matching. However, under process variation they show considerable amounts of offset voltage.

In digital circuits, threshold voltage variation is important in two respects. Firstly is the variation can reduce design margins of the digital circuits and may result in faulty circuit operation. Secondly the exponential impact of threshold voltage on leakage current variations. In recent years, a number of mitigation techniques have been used to reduce inter-die variations due to threshold voltage. As an example, Tschanz et al. [31] used body bias to adaptively adjust the threshold voltage of the devices in the circuit. Techniques such as this can greatly reduce the inter-die and intra-die variations due to threshold voltage and therefore improve the yield of the circuit.

However, for low-power applications, where the voltage of the power supply is very close to the threshold voltage, the impact of variation is large and thus threshold variations jeopardise circuit operation and cannot be dealt with at circuit level.

2.3.1.4 Transistor Dimensions (W, L)

Form the saturation current of the MOSFET in Equation 2.5, it is clear that the width (W) and length (L) of a transistor are critical in influencing the current and thus performance. To increase the drive current, the width of the device must be increased or its gate length (L) decreased. Increasing the gate width, in general, is not as effective as decreasing the length, since it will also increase the load capacitance. While shrinking gate length not only increases the saturation current, it also reduces gate capacitances and increases transistor density as well. Therefore (L) is continually reduced in order to pursue improvement in performance, making it the most vital dimension in a transistor today.

A large number of processing steps contribute to the overall variation in gate length and width. These factors include the wafer mask, exposure time, etching process, spacer definition, source/drain implantation and even the environment during the manufacturing process. Of these factors, the primary sources of variation are the steps involved in the photolithographic and plasma etch processes which are considered as systematic variations and hence can be compensated, and random variations such as line-edge roughness.

Due to the fact that (W) is often larger than (L), the variation in channel length is generally a major concern in device manufacturing (with the exemption of the shortest width transistors). Equations 2.4 and 2.5 show the impact of the transistor gate length on

the delay. Any variation in channel length is directly reflected in device delay which is directly proportional to the channel length. In addition, shrinking the device channel length is physically limited by the patterning wave length (λ=193 nm for 45 technology node), therefore patterning a very short channel length below this wavelength becomes extremely difficult to control leading to an increase in gate length variation [32].

2.3.2 The Impact of Process Variation on Performance of Circuit

At circuit level, the variation in transistor parameters affects performance metrics differently and depends on a variety of factors, together with circuit implementation, logic style and the region of transistor operation. Additionally, the impact of variation in a particular parameter is different for different performance metrics such as power consumption and delay.

Tables 2.3 and 2.4 demonstrate the impact of the variation in the transistor parameters on common digital blocks, such as adders based on a 90nm process. The data was obtained using the Monte-Carlo simulation method [22]. Table 2.3 shows the variability in delay and power as a function of different circuit implementations, whilst Table 2.4 decomposes the overall variation into contributions from individual parameters. It is also noticeable that the variations in threshold voltage and channel length contribute most seriously to the overall variation in delay and power consumption of the circuit.

Circuit style	Delay Variability (σ/μ) (%)	Power Variability (σ/μ) (%)
Static CMOS	6.1	4.1
Pulsed-Static CMOS	6.5	5
Domino	6.6	4.3

Table 2.3: Circuit-style impact on delay and power variability [22].

However, it should also be noted that temporal sources of variation, especially V_{dd} variations, are not included which also considerably contributes to the performance metrics variation.

Parameter	Delay (σ/μ)	Power(σ/μ)
t_{ox}	1-2 %	1-2 %
W	<<1 %	0.5-1 %
L	<3 %	< 2 %
V_{th}	2.5-6 %	1.75-4.75 %

Table 2.4: Contributions of key transistor parameters of variation on delay and power variation [22].

2.4 Analysis Methods for Studying Variability

Digital circuits are traditionally designed in such a way that the manufactured circuits should meet the performance specifications such as speed and power consumption under all operating conditions. However, the statistical fluctuations in the semiconductor fabrication processes have resulted in undesirable variations in circuit performance. The excessive spread of circuit performance can lead to a significant yield loss and hence can increase the unit cost of the product. Therefore it is necessary to understand and model manufacturing process variations for the prediction of device and circuit performance and also to provide enough information for circuit designers in order to minimise the impact of parameter variation on the circuit performance and maximise the yield.

In the following Sections, the different general techniques and methodologies used to handle the impact of process variations in circuit design are briefly reviewed, pointing out their strengths and weakness.

2.4.1 Worst Case-corner Analysis

Worst Case-corner Analysis (WCA) was one of the most common and widely-practiced techniques used to verify circuit performance under process variation. This was mainly because of its efficiency in terms of computational time and design efforts.

In this approach, numbers of NMOS and PMOS parameters are used to assess the circuit performance at the nominal and the worst and best extremes of the process corners [33].

However, the correlation between different device parameters is neglected [34], thus possibly leading to the use of worst case models that are extremely unlikely or even physically impossible to occur.

This is illustrated in Figure 2.6, which shows a scatter plot of the NMOS and PMOS I_{Dsat} measurements over numerous wafer lots. It is clear that the corners, which here are fast–slow and slow–fast (FS and SF), rarely occur.
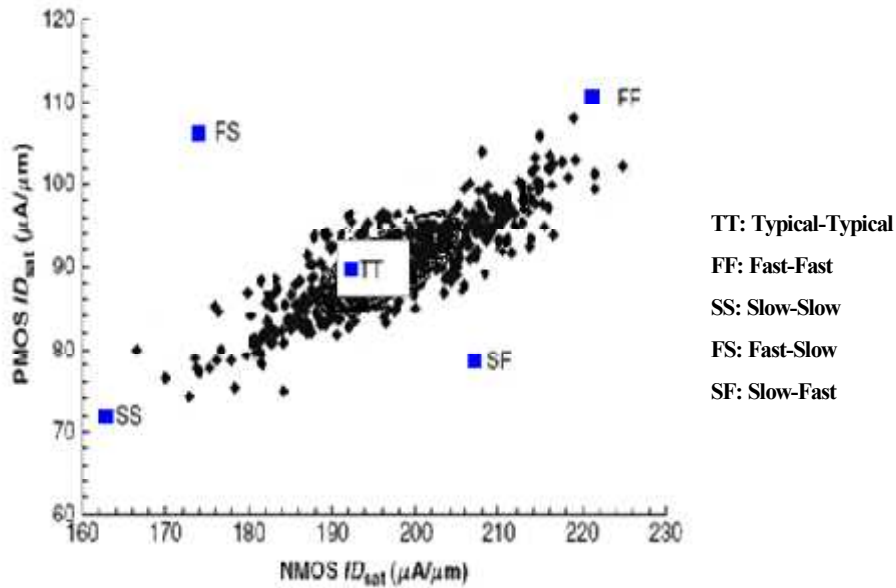
Figure 2.6 : Process variation map for N and P-MOSFET devices [8].

Moreover, with the increase of the impact of local process variation (intra-die variations) where process parameters tend to randomly fluctuate in any direction within the die [8], the WCA faces serious limitations for modelling due to its inability to partition the effect of localised variation between devices based on corner models [8, 33].

The shortcoming of considering unrealistic process combinations leads to overdesign and makes it harder for circuits to meet their design specification, especially as design margins become smaller with the continued trends in scaling. It is also worth mentioning that the use of WCA does not give enough information about the robustness of the design to the effects of process variations [14, 33].

2.4.2 Monte Carlo Analysis Technique

The Monte Carlo (MC) technique was first introduced and applied by Metropolis and Ulam in 1949 [35]. The technique is based on iteratively evaluating the response of the deterministic model using sets of random numbers as inputs within certain specified ranges [36]. The simulation flow for the technique is shown in Figure 2.7 [37] .

Despite the results being valuable and the accuracy being good in this technique, it is computationally expensive, especially when a large number of variables are required to be involved such as in studying the impact of parameter variation in semiconductor processes [37]. However, the accuracy of the estimated results could be traded-off against computation times.
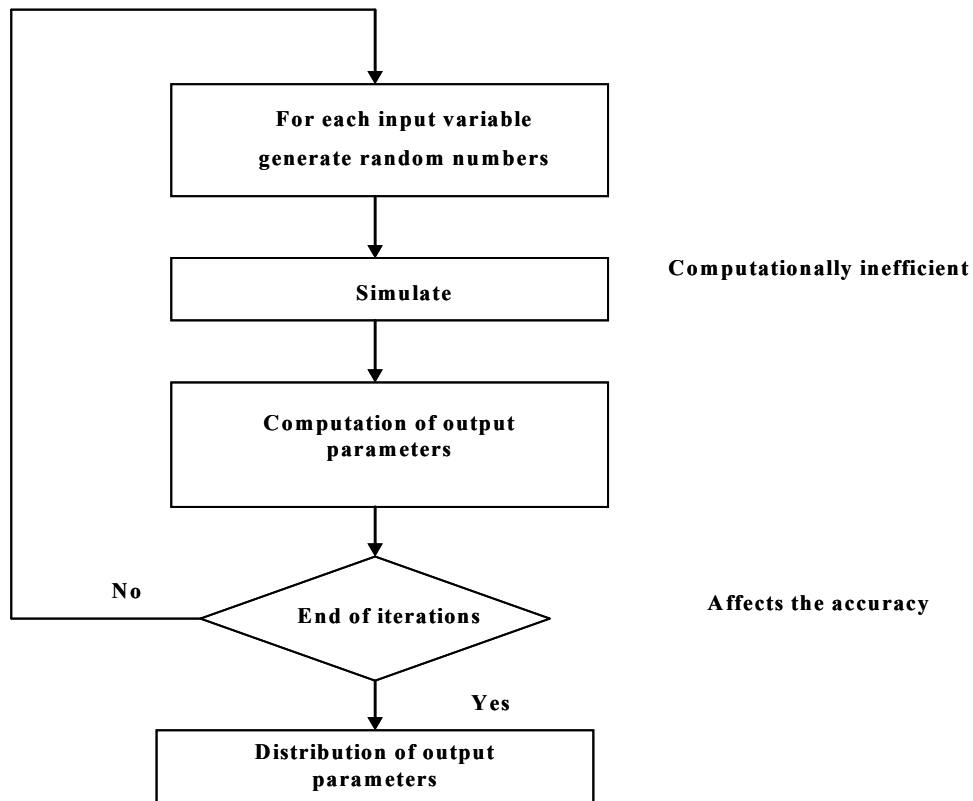
Figure 2.7: Monte Carlo analysis (flow diagram) [37].

 A further disadvantage of this approach, because of the random nature of the sampling technique involved in MC, is that a set of independent input samples sometimes can leave large regions of the design space unexplored leading to underestimations of the design space [36].It is also worth noting that the MC method is suitable for simulating both local

and global variations, in contrary to the worst case corner method where the local variations cannot be considered.

Cheng et al., [38] has demonstrated a MC-based circuit simulation methodology to study the impact of random dopant fluctuations on an SRAM cell. The MC-based approach is used to collect the intrinsic parameter fluctuation information into compact model sets and thereafter the impact of random device doping on SRAM static noise margins, read and write characteristics was studied.

2.4.3 Design of Experiments and Response Surface Modelling (DoE/RSM)

Design of Experiments (DoE) and Response Surface Modelling (RSM) are well-established branches of statistics which have been successfully adopted since the 1920s in many manufacturing fields such as the chemical and aeroplane industries [39, 40].

In these techniques, a systematic method for experiment planning is used in order to conduct the experiments in an efficient way and enable designers to construct empirical models from which the output responses can be determined as a function of the input factors or parameters.

Response Surface Modelling methodology is a combination of mathematical and statistical methods which are useful in developing analytical models and the analysis of problems in which a response of interest is affected by several input variable factors and the aim is to optimise this output response [39]. The response is usually represented graphically by a surface.

For a given system with an output response, y, and a vector of independent input variables, x, considered in planned experiments, the relationship between x and y is given by Equation 2.12.

$$y(x) = f(x) + \varepsilon \qquad (2.12)$$

where ε represents a normal distribution random error with a zero mean value and constant standard deviation σ.

Applying RSM techniques, an approximate or predictive value of the unknown $f(x)$ function can be built and expressed as $g(x)$ given in Equation 2.13.

$$\bar{y}(x) = g(x) \qquad\qquad (2.13)$$

The resulting RSM model of ($\bar{y}$) can be either in a form of a linear or a polynomial function. The linear models are typically obtained by designing a two-level fractional factorial experiment [39].

$$\bar{y}(x) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i \qquad\qquad (2.14)$$

A second order polynomial is most commonly used to model larger variations to account for all two-parameter interactions

$$\bar{y}(x) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_{ii}^2 + \sum_{i=1}^{k} \sum_{j=1}^{k} \beta_{ij} x_i x_j \qquad (2.15)$$

where, in Equations 2.14 and 2.15, k is the number of input variables, x_i is the i_{th} input variable and β is the RSM coefficient which is calculated using least squares regression analysis to fit the response approximation $\bar{y}$.

In general, the methodology adopted to model a circuit for the analysis of the effects of variability using the DoE/RSM techniques consist of a three step process including screening, model building and model analysis.

The screening step is usually adopted to identify the most significant parameters influencing the output response and thus reduce the input design space if there are a large number of parameters to be modelled. In screening, design techniques such as fractional factorial and Plackett-Burman (PB) design [39] along with other techniques are used to screen out the input parameters that influence the output response significantly.

In the second step, an approximation model is constructed based on the most significant (screened) parameters using a suitable RS design method such as central composite design (CCD), Box-Behnken [39] etc.

The decision on type of DoE plan is based on the polynomial order of the model, the number of parameters, and the model accuracy requirements. The number of simulation runs that are needed to build a quadratic model is typically less than one hundred runs when considering 10 to 12 input parameters for the MOSFET device implemented on the current technology. Subsequently, it reduces the runtime complexity compared to the

Monte-Carlo technique. In other words, the power of DoE/RSM techniques over the MC approach is its computational efficiency and reduced run time with acceptable accuracy.

The RSM technique, along with other variability modelling techniques, can also be used in the optimization of semiconductor manufacturing processes and hence the yield as reported in various investigations [14, 41, 42].

Aoki et al., [43] has illustrated the use of the DoE/RSM technique for device optimization and accurate prediction of process sensitivity in device performance. They have applied this methodology for the optimization and calculation of statistical variations for a 0.5μm MOSFET using a two-dimensional device simulator. Device performance parameters such as threshold voltage, output conductance, and drain current have been studied with respect to some device parameters such as gate length, oxide thickness, source/drain doping concentration, and substrate doping concentration.

2.4.4 Sensitivity Analysis

Sensitivity analysis is a method that is used to determine how "sensitive" the output of a model is to changes in the value of the input parameters and to changes in the structure of the model. This technique is suitable for studying the impact of one set of input parameters, for example, geometrical parameters.

The sensitivity analysis for a linear function of (X) is performed by propagating the variance from X to Y as expressed in Equations 2.18 and 2.19.

$$Y + \Delta Y = f(X + \Delta X) \qquad (2.18)$$

$$\Delta Y \approx \left| \frac{\partial f}{\partial X} \right| \Delta X \qquad (2.19)$$

where ΔX and ΔY are the standard deviations of the parameters X and Y respectively. If ΔY has small deviation value then it can be approximated by a normal distribution variance as follows [9]:

$$\sigma_Y^2 \approx \left[\frac{\partial f}{\partial X} \right]^2 \sigma_X^2 \qquad (2.20)$$

The sensitivity approach can be used for analytical-based modelling approaches or even along with other modelling approaches. Blaauw et al.,[44] presented the concept of statistical sensitivities which are used to perform sensitivity analysis based on statistical optimization of delay, power and leakage power. The optimization is performed by utilizing a combination of the existing statistical static timing analysis and sensitivity calculation approach. The disadvantage of a sensitivity-based approach is that it is unable to model the non-linear or interactive effects between the parameters.

2.5 Need for Silicon on Insulator Technology

The semiconductor industry is driven by the continuous scaling of device feature sizes. However, at the nanometre scale, the performance and yield of bulk silicon technology is greatly degraded due to short channel effects, the increase of process variation effects and the reduced immunity to single event upsets induced by radiation particles. Therefore the silicon technology is almost reaching its physical limits and the era of bulk MOSFET is nearing its end.

Silicon-On-Insulator (SOI) technology, in the view of scaling barriers such as the increased effects of process variation and the reduced immunity to single event upsets, is gaining large attention to be the next driver of technology scaling due to its capability of providing more speed, less power consumption and enhanced scalability as demanded by the future CMOS generations.

Comparing both bulk CMOS and SOI technologies, it was reported that CMOS circuits realised by SOI can work at a 20-35% higher speed than their bulk counterparts, and 2-4 times less power consumption when running under the same operating conditions [45].

Research on SOI technology dates back to 1960s [46],when this technology was adopted in military and space applications due mainly to its robustness to radiation-induced soft errors. However, the high cost of manufacturing SOI wafers has been the main barrier to their widespread growth and prevented them from entering volume production.

Today, with the continuous market demands for higher speed and lower power consumption devices, the potential advantages of SOI look more attractive and help it find its way into the semiconductor roadmap [47].

In recent years, most of the semiconductor companies have either utilized or considered SOI technology as a potential solution to device performance improvements. IBM, AMD, Sony Group and Toshiba have already adopted SOI for cell processors and ST Microelectronics have also switched to using SOI wafers.

Several SOI device structures varying from single gate to multiple gate structures have evolved such as planer SOI, double-gate SOI, FINFET and so on, and the topic of current research.

It is fair to say that Moore's law which has driven the past device scaling will continue to be alive when SOI technology is adopted into mainstream manufacturing.

2.5.1 Silicon on Insulator Advantages over Bulk Silicon

The SOI substrate consists of a silicon film, usually in the range of tens to hundreds of nanometres thick, isolated from the substrate by a relatively thick layer of silicon dioxide or a suitable insulating material, usually, called buried oxide. Figure 2.9 shows a cross Section of a SOI device.
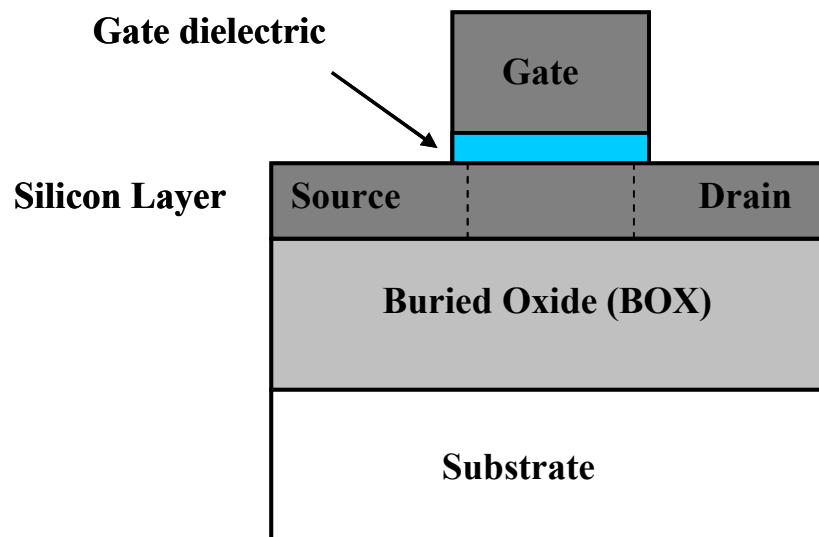
Figure 2.8: Cross Section of a MOSFET device realised on SOI technology.

The SOI device has the advantage of its small source/drain junction capacitance which provides high-operating speeds, less power consumption and higher device reliability. One of the major advantages of the SOI technology is the suppressed short channel

effects which means that the SOI device has a steeper sub-threshold slope which in turn can be translated into higher drive current(I_{on}) and less source/drain leakage current (I_{off}).

The vertical isolation from the silicon substrate, by the Buried Oxide (BOX) and the lateral isolation by the shallow trench oxide, provides this technology with three significant advantages over the conventional bulk silicon technology. Firstly, the latch-up [48] and cross-talk between devices which are prominent in bulk silicon devices are eliminated [49]. Secondly, the immunity to soft errors due to radiation particles, especially in SRAM circuits [48], is increased. Thirdly, there is flexibility in using different voltages on different devices without any need to add extra processing steps for triple wells as in bulk silicon. In addition, the same circuit design methodologies employed in bulk silicon can also be applied to SOI devices [48].

With semiconductor process technologies moving down into the nanometre regime, the inherent benefits of SOI technology in reducing junction capacitance, suppressing the short-channel effects, reducing leakage and increasing soft error immunity become more and more attractive as an alternative technology to current bulk silicon technology.

2.5.2 Silicon on Insulator Devices

Depending on the active body thickness of the silicon layer, the SOI MOSFET operation can be classified into two types, namely fully-depleted SOI (FDSOI) and partially-depleted SOI (PDSOI) [47].

In partially-depleted SOI devices, the silicon layer is thick enough so that the depletion region under the channel does not reach the silicon body-Buried Oxide (BOX) interface. This results in the creation of a floating region (without charges) and may lead to floating body effects such as the presence of kink in the I_d-V_d characteristics, and the activation of the lateral parasitic bipolar transistor. Therefore, a proper design is essential to minimise the floating body effects.

In a fully-depleted device, the depletion region under the channel extends up and covers the whole silicon layer providing a better coupling between the gate and the inversion layer in the channel, which in turn improves the device drive current.

The FDSOI transistors have several advantages over the PDSOI counterparts. The FDSOI has the highest gains and faster circuits as well as the highest level of radiation immunity

and better sub-threshold swing [48]. It also is free from kink effects in the current voltage characteristics.

In order to distinguish between the fully and partially depleted SOI MOSFETs, a coefficient (α) is introduced to represent the total capacitance that is formed from the gate capacitance and the capacitance of the silicon and the buried oxide layers as shown in Figure 2.10 [50]. The value of "α" is governed by the Equations 2.21 and 2.22 for FDSOI and PDSOI devices respectively [48, 50].
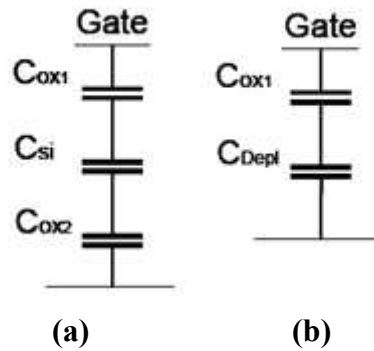
Figure 2.9 : A simplified capacitive network of (a) a fully depleted and (b) a partially depleted SOI MOSFET.

$$\alpha_{FD} = \frac{C_{ox\,2}C_{si}}{C_{ox\,1}(C_{ox\,2}+C_{si})} \qquad (2.21)$$

$$\alpha_{PD} \approx \frac{C_{dep\,1}}{C_{ox\,1}} \qquad (2.22)$$

where C_{ox1} and C_{ox2} represent the gate oxide and the buried (BOX) capacitances respectively. C_{depl} and C_{si} are the capacitances that relate to the depletion region of the PDSOI and that of the thin film silicon respectively. The coefficient α also describes the efficiency of the coupling between the gate and the channel of the device.

Typically, this coefficient is very small and close to zero for FDSOI , while it is around 0.3-0.5 for PDSOI and bulk silicon [50].

Moreover, the value of α increases when reducing device sizes, this means that the driving controllability of the gate is also reduced with α alpha reduction.

The saturation current and transconductance can be also expressed in terms of α for a long channel SOI device as in Equations 2.23 and 2.24 [50]

$$I_D = \frac{W}{L} \cdot \frac{\mu C_{ox\,1}}{2\,(1+\alpha)} \cdot (V_{gs} - V_{th})^2 \qquad (2.23)$$

$$g_m = \frac{\partial I_D}{\partial V_{gs}} = \frac{W}{L} \cdot \frac{\mu C_{ox\,1}}{(1+\alpha)} \cdot (V_{gs} - V_{th}) = \sqrt{2\,I_D \cdot \frac{W}{L} \cdot \frac{\mu C_{ox\,1}}{(1+\alpha)}} \quad (2.24)$$

where W and L are the width and the length of the gate of the transistor, μ is the carrier mobility, V_{gs} is the gate to source voltage, and V_{th} is the threshold voltage.

From Equations 2.23 and 2.24 it can be concluded that for a given technology node, the FDSOI devices have a better transconductance and current drive compared with PDSOI or bulk devices. This is mainly because α is smaller in the case of FDSOI technology.

Despite the potential benefits of SOI technology such as its high performance and low power consumption, the wide adoption of SOI technology still faces both real and perceived challenges. In the past, the main barrier which prevented SOI technology being used to be in mainstream CMOS fabrication industry is controlling the silicon layer or silicon dioxide layer quality, which in turn means a higher cost of SOI wafers. The key materials quality issues are firstly the continuity uniformity of the BOX thickness, secondly, thickness uniformity of the active silicon layer and thirdly the defects in silicon layer. Moreover, the presence of the defects in the BOX layer may increase the number of defects and threading dislocations in top silicon layer.

2.6 Summary

This chapter reviews the sources and different types of process variations, including inter–die and intra-die. The impact of these variations on the device and circuit levels is explored. The different approaches and methodologies used to handle the impact of process variations in circuit design are also examined, pointing out their strengths and weaknesses. The traditional techniques, such as Worst Case Corner Analysis, Monte Carlo technique (MC), as well as a brief introduction to DoE and RSM, Sensitivity

Analysis and the Statistical Static Timing Analysis technique are reviewed. The DoE and RSM techniques can provide a reasonable balance between accuracy and the computational efficiency as compared to the MC simulations by limiting the number and hence the complexity of simulations. Finally, the need for new process technology to overcome the scaling associated barriers such as SOI technology is highlighted. The SOI technology is discussed as a possible candidate for future MOSFET devices in order to overcome the increased effect of process variations due to scaling and their impact on circuit performance and yield.

2.7 References

[1]   W. Shockley, "Problems related to p-n junctions in silicon," Solid-State Electronics, vol. 2, pp. 35-67, 1961.

[2]   W. Schemmert and G. Zimmer, "Threshold-voltage sensitivity of ion-implanted MOS transistors due to process variations," Electronics Letters, vol. 10, pp. 151-152, 1974.

[3]   D. Sylvester, K. Agarwal, and S. Shah, "Variability in nanometer CMOS: Impact, analysis, and minimization," Integration-the VLSI Journal, vol. 41, pp. 319-339, May 2008.

[4]   B. H. Calhoun, C. Yu, L. Xin, M. Ken, L. T. Pileggi, R. A. Rutenbar, and K. L. Shepard, "Digital Circuit Design Challenges and Opportunities in the Era of Nanoscale CMOS," Proceedings of the IEEE, vol. 96, pp. 343-365, 2008.

[5]   M. Alioto and G. Palumbo, "Impact of supply voltage variations on full adder delay: Analysis and comparison," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 14, pp. 1322-1335, Dec 2006.

[6]   S. Kang and Y. Leblebici, "CMOS Digital Integrated Circuits," 2nd ed: Mc-Graw Hill, 1999.

[7]   International Technology Road Map for Semiconductors (ITRS), http://public.itrs.net.

[8]   B. P. Wong, Nano-CMOS Circuit and Physical Design. New Jersey John Wiley 2005.

[9]   D. Boning and S. Nassif, "Models of Process Variations in Device and Interconnect," Design of High-Performance Microprocessors Circuits, pp. 98-116, 2000.

[10]   A. Srivastava, D. Sylvester, and D. Blaauw, "Statistical Analysis and Optimization for VLSI: Timing and Power " New York, NY: Springer, 2005.

[11]   H. Masuda, S. Ohkawa, A. Kurokawa, and M. Aoki, "Challenge: variability characterization and modeling for 65- to 90-nm processes," in *Proceedings of the IEEE 2005 Custom Integrated Circuits Conference.* , pp. 593-599,2005.

[12]   M. Alioto, G. Palumbo, and M. Pennisi, "Understanding the Effect of Process Variations on the Delay of Static and Domino Logic," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 18, pp. 697-710, 2010.

[13]   S. S. Sapatnekar, "Variability and Statistical Design," IPSJ Transactions on System LSI Design Methodology, vol. 1 pp. 18-32, August 2008.

[14]   A. A. Mutlu and M. Rahman, "Statistical methods for the estimation of process variation effects on circuit operation," IEEE Transactions on Electronics Packaging Manufacturing, vol. 28, pp. 364-375, 2005.

[15]   S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in Design Automation Conference, pp. 338-342, 2003.

[16]   M. Orshansky, L. Milor, C. Pinhong, K. Keutzer, and H. Chenming, "Impact of spatial intrachip gate length variability on the performance of high-speed digital circuits," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 21, pp. 544-553, 2002.

[17]   A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs," IEEE Transactions on Electron Devices, vol. 50, pp. 1837-1852, 2003.

[18]   B. G. Streetman and S. K. Banerjee, "Solid State Electronic Devices," United States of America,Prentice Hall.Inc, pp. 295-297, 2000.

[19]   P. Bai, C. Auth, S. Balakrishnan, M. Bost, and e. al., "A 65nm logic technology featuring 35nm gate lengths, enhanced channel strain, 8 Cu interconnect layers, low-k ILD and 0.57 µm2 SRAM cell," IEEE International Electron Devices Meeting, pp. 657-660, 2004.

[20]   C. Gallon, G. Reimbold, G. Ghibaudo, R. A. Bianchi, R. Gwoziecki, S. Orain, E. Robilliart, C. Raynaud, and H. Dansas, "Electrical analysis of mechanical stress induced by STI in short MOSFETs using externally applied stress," IEEE Transactions on Electron Devices , vol. 51, pp. 1254-1261, 2004.

[21]   Z. Wei, F. Liu, K. Agarwal, D. Acharyya, S. R. Nassif, K. J. Nowka, and C. Yu, "Rigorous Extraction of Process Variations for 65-nm CMOS Design," IEEE Transactions on Semiconductor Manufacturing, vol. 22, pp. 196-203, 2009.

[22]   K. Bernstein, D. J. Frank, A. E. Gattiker, W. Haensch, B. L. Ji, S. R. Nassif, E. J. Nowak, D. J. Pearson, and N. J. Rohrer, "High-performance CMOS variability in the 65-nm regime and beyond," IBM Journal of Research and Development, vol. 50, pp. 433-449, 2006.

[23]   H. Sasaki, M. Ono, T. Yoshitomi, T. Ohguro, S. Nakamura, M. Saito, and H. Iwai, "1.5 nm direct-tunnelling gate oxide Si MOSFET's," IEEE Transactions on Electron Devices, vol. 43, pp. 1233-1242, 1996.

[24]   A. Marras, I. De Munari, D. Vescovi, and P. Ciampolini, "Impact of gate-leakage currents on CMOS circuit performance," Proceeding of the 24th International Conference on Microelectronics, vol. 2, pp. 653-656,2004.

[25]   J. Chen, T. Y. Chan, P. K. Ko, and C. Hu, "Gate current in OFF-state MOSFET," Electron Device Letters, IEEE, vol. 10, pp. 203-205, 1989.

[26]   S. Datta, G. Dewey, M. Doczy, B. S. Doyle, B. Jin, J. Kavalieros, R. Kotlyar, M. Metz, N. Zelick, and R. Chau, "High mobility Si/SiGe strained channel MOS transistors with HfO/TiN gate stack," in IEEE International Electron Devices Meeting, 2003. IEDM '03 Technical Digest. pp. 28.1.1-28.1.4, 2003.

[27]   K. Mistry, C. Allen, et.al, "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," in IEEE International Electron Devices Meeting, 2007, pp. 247-250.

[28]   M. Cho, K. Maitra, and S. Mukhopadhyay, "Analysis of the impact of interfacial oxide thickness variation on metal-gate high-K circuits," IEEE Custom Integrated Circuits Conference, pp. 285-288, 2008.

[29]   M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," IEEE Journal of Solid-State Circuits, vol. 24, pp. 1433-1439, 1989.

[30]   A. Asenov, S. Kaya, and J. H. Davies, "Intrinsic threshold voltage fluctuations in decanano MOSFETs due to local oxide thickness variations," IEEE Transactions on Electron Devices, vol. 49, pp. 112-119, 2002.

[31]   J. W. Tschanz, J. T. Kao, S. G. Narendra, R. Nair, D. A. Antoniadis, A. P. Chandrakasan, and V. De, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," IEEE Journal of Solid-State Circuits, vol. 37, pp. 1396-1402, 2002.

[32]   L. Xi-Wei and V. Moroz, "Layout Proximity Effects and Modelling Alternatives for IC Designs," IEEE Design & Test of Computers, vol. 27, pp. 18-25, 2010.

[33]   K. Singhal and V. Visvanathan, "Statistical device models from worst case files and electrical test data," IEEE Transactions on Semiconductor Manufacturing, vol. 12, pp. 470-484, 1999.

[34]   A. Nardi, A. Neviani, E. Zanoni, M. Quarantelli, and C. Guardiani, "Impact of unrealistic worst case modelling on the performance of VLSI circuits in deep submicron CMOS technologies," IEEE Transactions on Semiconductor Manufacturing, vol. 12, pp. 396-402, 1999.

[35]   N. Metropolis and S. Ulam, "The Monte Carlo Method," Journal of the American Statistical Association, vol. 44, pp. 335-341, 1949.

[36]   G. AA, W. SF, and E. MS, "Overview of modern design of experiments methods for computational simulations," 41st AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV, 2003.

[37]   C. Jacoboni and P. Lugli, "The Monte Carlo method for semiconductor device simulation," Springer, Ed. New York 1989.

[38]   B. Cheng, S. Roy, and A. Asenov, "The impact of random doping effects on CMOS SRAM cell," in Proceeding of the 30th European Solid-State Circuits Conference, 2004, pp. 219-222, 2004.

[39]   Montgomery, "Design and Analysis of Experiments," 5th ed: John Wiley and Sons, 2001.

[40]   G. E. P. Box and N. R. Draper, "Empirical Model-Building and Response Surfaces," John Wiley and Sons, 1987.

[41]   D. S. Boning and P. K. Mozumder, "DOE/Opt: a system for design of experiments, response surface modelling, and optimization using process and device simulation," IEEE Transactions on Semiconductor Manufacturing, vol. 7, pp. 233-244, 1994.

[42]   A. R. Alvarez, B. L. Abdi, D. L. Young, H. D. Weed, J. Teplik, and E. R. Herald, "Application of statistical design and response surface methods to computer-aided VLSI device design," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 7, pp. 272-288, 1988.

[43]   Y. Aoki, H. Masuda, S. Shimada, and S. Sato, "A New Design-Centering Methodology for VLSI Device Development," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 6, pp. 452-461, 1987.

[44]   A. Srivastava, D. Sylvester, and D. Blaauw, "Statistical optimization of leakage power considering process variations using dual-Vth and sizing," in Proceedings.of the 41st Design Automation Conference, pp. 773-778, 2004.

[45]   D. K. Sadana and M. Current, "Fabrication of Silicon On Insulator (SOI) Wafers Using Ion Implantation," in Ion Implantation Science and Technology, J. F. Ziegler, Ed.: Ion Implantation Technology Co., 2000.

[46]   M. Alexe and U. Gösele, "Wafer bonding: applications and technology "Berlin: Springer, 2004.

[47]   K. Bernstein and N. J. Rohrer, "SOI circuit design concepts," Kluwer Academic Publishers, 2000.

[48]   J. P. Colinge, "Silicon-on-Insulator Technology: Materials to VLSI ", 2 ed,Boston: Kluwer Academic Publishers, 1997.

[49]   M. Rahim, H. Bor-Yuan, and J. Foerstner, "Comparison of SOI Versus Bulk Silicon Substrate Crosstalk Properties for Mixed-Mode IC's," in IEEE International SOI Conference, pp. 170-171, 1992.

[50]   J-P.Colinge, "Silicon-on-Insulator Technology: Materials to VLSI," Kluwer Academic Publisher, 1991.

Chapter 3

Variability Analysis of Compact Model Parameters of Bulk Silicon and SOI Technologies

3.1 Background

The different statistical techniques used to model and analyse variability have been reviewed in Chapter 2. Two techniques such as DoE and RSM have been commonly used to model variability and to optimise the manufacturing process in many industrial fields such as chemicals, aerospace, etc. These techniques have also been used in this research to model process variability in the semiconductor field because of their computational efficiency compared to the Monte Carlo (MC) method.

This chapter continues with an introduction to the DoE approach in Section 3.2 which includes a review of a range of two-level DoE categories such as full factorial design, fractional factorial design, and screening design. The Response Surface (RS) design has been used since it has this capability to fit a second order or curved surface. This design is discussed in Section 3.2. Together with several variants such as Central Composite Design (CCD), Face Centred Central Composite Design (FCCC) and Box-Behnken are illustrated. In Section 3.4, these statistical techniques are utilised to model the impact of process variations with regard to 60 nm devices realised by FDSOI, PDSOI and bulk silicon technology processes on key electrical device parameters, and are used to analyse the effect process variability on the device characteristics.

Previous research works had focused on the investigation of the effect of process variations at device level using different process technologies such as FDSOI as in [1, 2] and FinFET as in [3], however most of those works have not studied the impact of those effects on the compact model parameters of the devices and the circuit levels. This chapter aims to model and analyse the impact of the effects of process variations on key compact model parameters of SOI and bulk silicon devices such as threshold voltages (V_{th0}) for both NMOS and PMOS devices.

Finally, the chapter is concluded in Section 3.5 with a summary of the results.

3.2 Design of Experiments

Design of Experiments (DoE) is an established branch of statistics which has been widely adopted since the 1920s in many manufacturing fields such as the chemical industry in order to obtain quality and yield improvements [4-6].

In this technique, a systematic method for experiment planning is used in order to run the experiments efficiently, and to help the designer to construct empirical models with regard to which responses of interest can be determined as a function of the input factors or parameters. As an example, Xu et al. used statistical based experimental designs to optimise the medium of an important medical microorganism [7].

The advantage of DoE over other statistical techniques such as the Monte-Carlo technique (MC) is that they are much more computationally efficient, easier to develop and it is possible to build a model that can be used for predicting variability instead of performing exhaustive simulation or experimental runs.

In general, the DoE method is divided into three main categories: full factorial, fractional factorial, screening and response surface (RS) designs [4, 5]. In the following sections, the main DoE methods will be discussed in more detail.

3.2.1 Full Factorial Design

Full factorial design is one of the most basic techniques in experimental design. In this method, all of the possible combinations of the main parameters or factors and their interactions are considered. Two-level factorial designs are the most widely used method for modelling main effects and interactions as they need a smaller number of experimental runs compared to higher order factorial designs. In two-level designs, each parameter has two levels, 'high value (+1)' and 'low value (-1)', and 2^k number of experiments is required to incorporate all the possible combinations of the k factors. Figure 3.1 shows a geometrical representation of a three parameter (x, y, z) with two-level full factorial design. Here, eight (2^3) experimental runs are required to consider the main effects (x, y, z), the two-factor interactions (xy, yz, xz), while the three-factor

interaction is (xyz) and (- xyz). Table 3.1 shows the factorial design matrix for two-level of 3 factors.
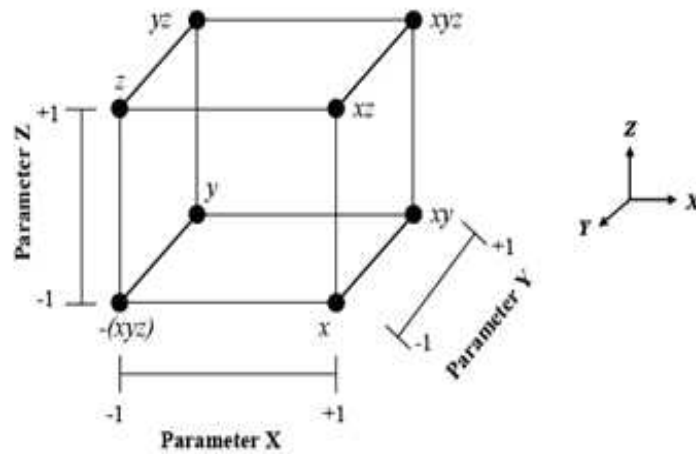
Figure 3.1: Two level full factorial design for three factors (k=3) [8].

Run	Parameter			Labels
	x	y	z	
1	-1	-1	-1	-(xyz)
2	1	-1	-1	x
3	-1	1	-1	y
4	1	1	-1	xy
5	-1	-1	1	z
6	1	-1	1	xz
7	-1	1	1	yz
8	1	1	1	xyz

Table 3.1: Design matrix for 2^3 factorial designs.

One of the most popular factorial designs is the three-level full design, such as a 3^k factorial design, which is widely used to model the quadratic effects of curved surface responses. The three levels are referred to as low, intermediate and high, and the number of runs for k factors is 3^k factorial runs which are required in order to capture the impact of all of the main effects, and their interactions in terms of an output response.

Despite this design giving more accurate results, it is computationally inefficient when using a large number of factors, since the experimental size increases exponentially with the number of factors, and hence is too expensive to run.

3.2.2 Fractional Factorial Design

As the size of the experiment increases exponentially with the number of factors in a full factorial design, e.g. for k=10 factors, this will require $2^{10} =1024$ experiments or runs. The use of fractional factorials designs can be considered as a solution by taking a fraction of the full factorial design, and hence achieving accuracy with fewer runs compared to the full factorial designs. In these designs, for k parameters, the number of experimental runs are 2^{k-q}, where the term 2^{-q} is the considered fraction and $q < k$. Figure 3.2 shows a half-fraction of the two-level fractional factorial design for three factors (k=3) and a design matrix for a half-fraction of the 2^3 factorial design is shown in Table 3.2.
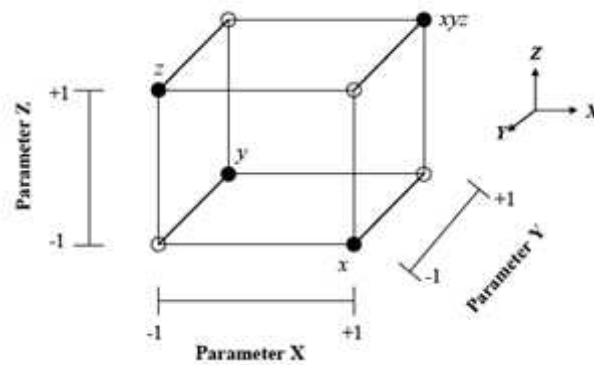
Figure 3.2: Two level fractional factorial design for three factors (k=3) [8].

Run	Parameter			Labels
	x	y	z	
1	1	-1	-1	x
2	-1	1	-1	y
3	-1	-1	1	z
4	1	1	1	xyz

Table 3.2: Design matrix for half fractional of a 2^3 factorial design.

Despite the reduction in the number of runs (design points), the aliasing and design resolution are considered to be the key issues in these designs.

The aliasing of an effect is when the impact of factors in a design cannot be distinguished separately from one another. For example, in the 2^3 full factorial design as shown in Figure 3.1, all the main effects (x, y, z), all two factor interactions (xy, yz, xz) and the three factor interaction (xyz) are considered. However, in the 2^{3-1} fractional factorial design shown by the solid dots in Figure 3.2, the main effects (x, y, z) are aliased or confounded with the two factor interactions (xy, yz, xz). These aliased effects cannot be estimated independently unless all but one of the aliased effects are known or assumed to be negligible.

The design resolution describes the extent to which the effects are aliased or confounded with other effects. The higher the resolution of a fractional factorial design, the less the impact of aliasing, but more simulation runs are required. In general, Designs with Resolution III, IV and V are most commonly used [4, 5].

3.2.3 Screening Design

In the screening designs, the fractional factorial designs (2^{k-q}) are commonly used to identify or screen the most significant factors when the number of factors is too large to assess the higher order effects, and hence reduces the number of the input factors to the system to be investigated.

Screening designs of an experiment are usually carried out in the early stages of the experimentation phase, when it is likely that the output response of a system is dominated or driven primarily by the main effects and their low-order interactions, and it is assumed that the higher order of multiple factor interactions have negligible or no effect on the responses of interest. The significant factors in a system are then investigated more thoroughly, either using full factorial or RSM designs as the next phase in the experimentation.

One of the most commonly used techniques for screening is the **Plackett-Burman (PB) design** [9] which is considered as a 2-level fractional factorial design with a class of

Resolution III. Using the Plackett-Burman (PB) design, it is efficient to run N experiments for studying k=N-1 factors.

The Plackett–Burman designs have also the following orthogonal property where for any given two columns in the N experiment design matrix, there are N/4 plus signs (+1) and N/4 minus signs (-1) in the first column corresponding to N/2 plus signs in the second column. Similarly, for N/2 minus signs in the second column, there will be N/4 plus and N/4 minus signs in the first column [10].

Table 3.3 shows that for a PB design for 11 parameters, only 12 runs are required. The design matrix is built first by constructing the first column according to the PB design [9]. A second column is then generated from this first one by shifting the elements of the column down by one position and moving the last element (+1 or -1) into the first position. In a similar manner, the third to the eleventh columns are constructed from the following column. Finally, a row of lower values (-1) is added as part of the design, hence completing the design.

Runs	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11
1	1	-1	1	-1	-1	-1	1	1	1	-1	1
2	1	1	-1	1	-1	-1	-1	1	1	1	-1
3	-1	1	1	-1	1	-1	-1	-1	1	1	1
4	1	-1	1	1	-1	1	-1	-1	-1	1	1
5	1	1	-1	1	1	-1	1	-1	-1	-1	1
6	1	1	1	-1	1	1	-1	1	-1	-1	-1
7	-1	1	1	1	-1	1	1	-1	1	-1	-1
8	-1	-1	1	1	1	-1	1	1	-1	1	-1
9	-1	-1	-1	1	1	1	-1	1	1	-1	1
10	1	-1	-1	-1	1	1	1	-1	1	1	-1
11	-1	1	-1	-1	-1	1	1	1	-1	1	1
12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Table 3.3: PB design for 11 factors.

3.3 Response Surface Designs

The Response Surface (RS) design is normally used to overcome the disadvantages of the full factorial design by mathematically modelling or fitting a second order function of curved response surfaces.

With a full factorial design, two level factorial designs (2^k) are not able to fit curved surfaces, as three distinct values for each factor are required to fit a quadratic function. However, while using higher levels, such as a 3^k design, it is not computationally efficient to model a curved or quadratic relationship.

RS designs are actually a two-level design (2^k) with added centre and axial points, where the centre and axial points are used to obtain an indication of the curvature and fit of a quadratic function.

The Central Composite Design (CCD) [9-11] and the Box-Behnken designs [10, 12] are the most widely used RS designs [4, 5].

In a CCD design, the curvature of a response is estimated by combining a two-level factorial or fractional-factorial design with centre points (all the values of a parameter are at their midrange) and axial points. In other words, a CCD is a modified first order design augmented with a number of centre and axial points, making the design very efficient for estimating a second order curvature.

The number of centre runs (n_c) and the distance (α) of the axial point are considered as design specifications of the CCD [4]. Figure 3.3 shows a geometrical representation of the CCD for k=2 factors, with 2^k, i.e. 4 factorial points and 2k, i.e. 4 axial points, and one centre point.
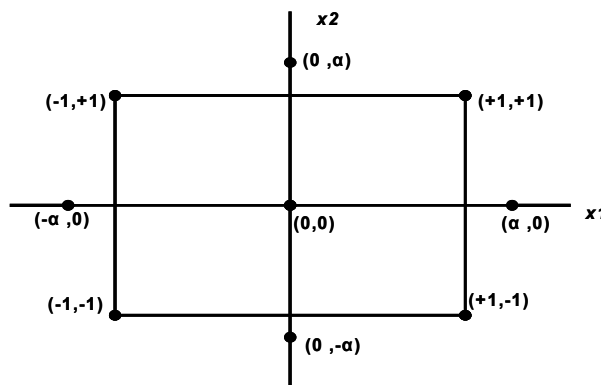
Figure 3.3: A geometrical representation of CCD for k=2 factors.

A Face Centred Central Composite design (FCCC) can be constructed by setting α =1, where the locations of the axial points will be on the centres of the faces of the square. Figure 3.4 shows a FCCC design with k=3 factors. Here 2^k, i.e. 8 factorial points, 2k, i.e. 6 axial points, and one centre point are required.
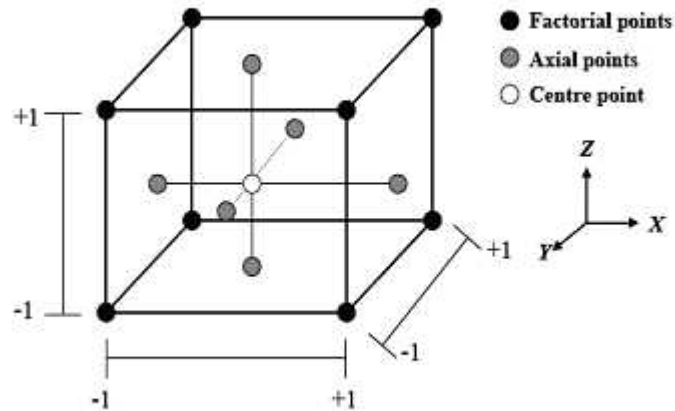
Figure 3.4: A geometrical representation of FCCC design for three factors (k=3) [8].

The Box-Behnken design is another type of computationally cost-effective response surface technique. The Box-Behnken design contains three levels of the input parameters design but does not contain an embedded factorial or fractional factorial design. As shown in Figure 3.5, these designs do not have any extreme points at the vertices of the cube as seen in both the CCD or FCCC designs.

The advantage of this is that running extreme points, which might be difficult or expensive, is not needed to model a curve response. However, compared to the CCD, some accuracy will be lost due to high uncertainty of prediction near the vertices.
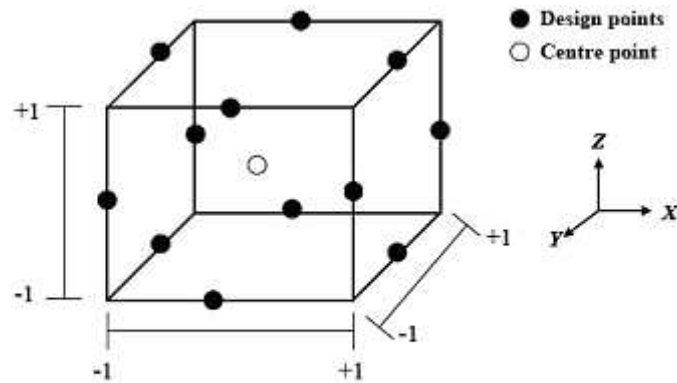
Figure 3.5: A geometrical representation of Box-Behnken design for k=3 factors [8].

After choosing an appropriate design, the next step in the RS modelling approach involves estimating or modelling the response parameter as a function of the input parameters.

3.3.1 Response Surface Modelling

Response Surface Modelling (RSM) is a combination of mathematical and statistical methods which are valuable for developing analytical models and for the analysis of problems in which the response of interest is governed by several input variables, and the aim is to optimise this output response. The response is usually represented graphically by a surface plot.

For a given system with an output response, y, and a set of independent input variables, x, in a planned experiment, the relationship between x and y is given as in Equation (3.1):

$$y(x) = f(x) + \varepsilon \qquad (3.1)$$

where ε represents a normal distribution random error with zero mean value and constant standard deviation σ.

Applying RSM techniques results in an approximation or prediction of the unknown $f(x)$ function, referred to as $\bar{y}$, which can be built and expressed as $g(x)$ in Equation (3.2):

$$\bar{y}(x) = g(x) \qquad (3.2)$$

The subsequent RSM model of $\bar{y}$ can be in the form of a linear or polynomial function.

The first-order models, sometimes called main effects models, are typically obtained by designing a two-level fractional factorial experiment. The first order models approximate the response surface over a small range of input factors, where the curvature of y is assumed to be negligible. The form of the first-order model is given in the following equation:

$$\bar{y}(x) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i \quad (3.3)$$

However, if the curvature in the response surface is sufficiently pronounced, the first order model is inadequate, even when considering the effect of the interaction terms. In this situation it is more likely to require a second- order model in order to account for the large response curvature due to the increased effect of input interactions.

In addition, a second order polynomial is most commonly used to model larger variations to account for all two-parameter interactions:

$$\bar{y}(x) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_{ii}^2 + \sum_{i=1}^{k} \sum_{j=1}^{k} \beta_{ij} x_i x_j \quad (3.4)$$

where, in equations (3.3) and (3.4), k is the number of independent input variables, x_i is the i_{th} input variable, and β is the RSM coefficient which is calculated using least squares regression analysis to fit the response approximation $\bar{y}$.

In general, the methodology involved in using the DoE/RSM techniques consists of three steps involving screening, model building and model analysis.

The screening step is usually utilised in order to identify the most significant parameters that influence the output response, and thus reduces the number of variables in the input design space if there are a large number of parameters to be modelled. In screening, design techniques such as fractional factorial, PB design, along with other techniques, are used to screen out the input parameters that influence the output response significantly.

In the second step, an approximation model is constructed based on the most significant (screened) parameters, using a suitable RS design method such as CCD or Box-Behnken. Finally, the RS models obtained are often analysed using graphical plots such as contour and surface plots.

3.4 Variability Modelling

A statistical approach based on TCAD and statistical techniques to model the impact of process variation effects on the device performance metrics for 60 nm NMOS and PMOS devices realised by strained FDSOI, PDSOI and the bulk Si processes, is presented in this section. The general methodology for studying variability is shown in Figure 3.6 and involves three main steps: parameter screening, model building and model analysis and validation. The methodology begins with the identification of uncontrollable process variation parameters which influence the output response investigated, in terms of a set of compact model parameters of the device. This is followed by the TCAD process and device simulations, the calibration of the device electrical characteristics with the experimental data, and the extraction of the compact model parameters of nominal devices.

The compact model parameters chosen in this work were the V_{th0} for both NMOS and PMOS transistors, where V_{th0} represents the threshold voltage for long channel devices at zero bias voltage [13, 14].

The reason for choosing V_{th0} for both NMOS and PMOS transistors as the design parameters among numerous other compact model parameters is that they show a strong statistical relationship with circuit performance metrics. In addition, they abstract complex physical process variation phenomena into a single number, and are commonly used as a design variable, both in digital and analogue designs, and hence bridge the gap between design and manufacturing processes.

Subsequently, a statistical screening process is performed if the number of process parameters to be investigated is large. Screening provides information about the statistical significance of process variation parameters from an initial large set of parameters, hence it reduces the number of process parameters to be modelled.

However, if the number of parameters to be modelled is not large, RSM is performed directly without any screening.

In the RSM step, the simulation experiments are designed to thoroughly investigate and model the output responses in terms of the initially identified process parameters, or the most significant process parameters obtained from screening.

The target of the RSM in this work is the compact model parameters (V_{th0}) for both NMOS and PMOS devices that give an insight into electrical device performance.

Finally, the RSM model validity is assessed in terms of statistical residual analysis [4]. In addition, the response surface plots are generated to visualise and study the behaviour of device responses under various process variations.
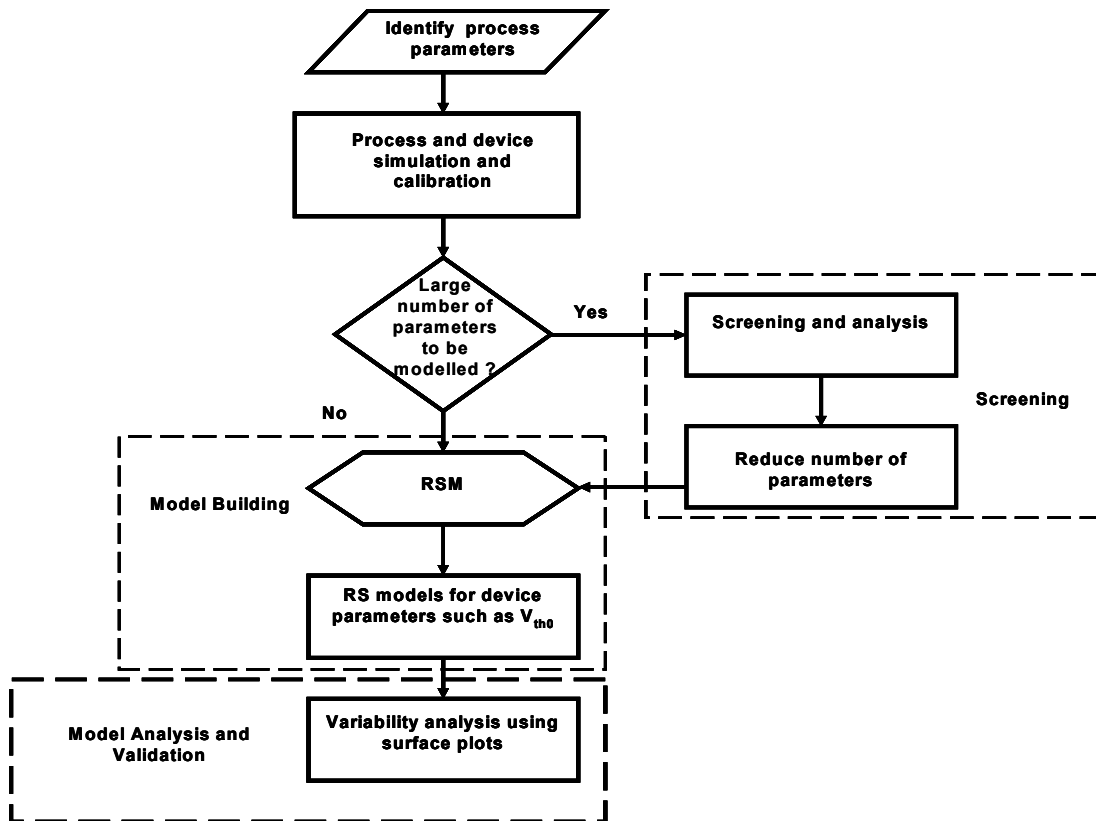
Figure 3.6 : Flow chart of variability analysis utilized by DoE and RSM statistical techniques.

3.4.1 TCAD Simulations

Three types of state of the art technologies are investigated in this work. To begin with, strained fully depleted silicon on insulator (FDSOI) devices designed for high performance applications were first calibrated with experimental data using the TCAD Sentaurus package [15]. Here, the gate length was 60nm, the thickness of the active silicon body was 16 nm, and the buried oxide thickness (BOX) was 145 nm. The gate

dielectric was composed of a thin layer of oxide (0.5 nm) with 2.5 nm of high-k dielectric gate material (HfO_2) to reduce the gate leakage. Figure 3.7 shows a cross section of the 60 nm FDSOI NMOS transistor (nominal case).

The electrical measurements of were performed using an Aglient 4155C semiconductor parameter analyzer connected to a cascade microtech probe station (semiautomatic probe station) through an Aglient E52550A low leakage switch mainframe. The measurements were programmed and controlled using a computer. The drain current–gate voltage characteristics of FDSOI (NMOS) devices were obtained by measuring the drain current as a function of gate voltage, where the gate voltage is swept from 0 volt to 1 volt while the drain voltage is kept at 0.1 volt and 1 volt respectively. While the output characteristics (I_d–V_d) of device were characterised by connecting the source and substrate body of the transistor to the ground .The drain voltage is then swept from 0 to 1 volt, while the gate overdrive voltage is stepped from 0 to 1 volt.

The calibration of the process and device simulator was performed by using the data available such as the device feature sizes. The doping profile in the device was adopted form a technical example of 60nm SOI in Sentaurus package [15]. The electrical characteristics obtained was fitted to that of the measured ones by iteratively tuning the doping values such as doping dose and energy of different process steps in order to match the measured data. For example, the threshold voltage adjustment doping dose and energy were used in order to fit the simulated threshold voltage with that extracted form measurements.
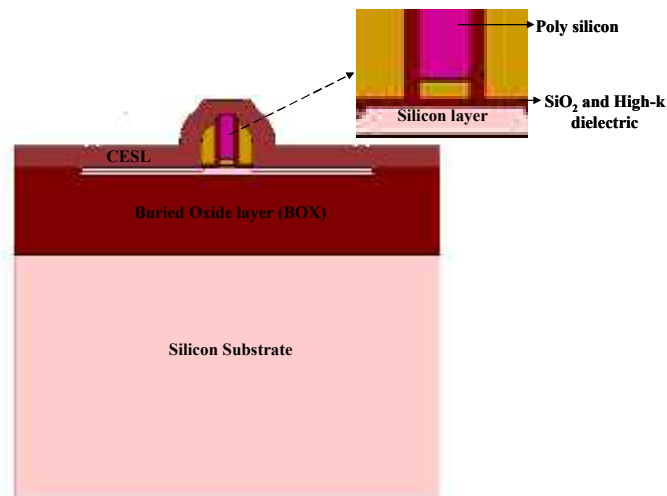
Figure 3.7 : A cross section of the 60 nm FDSOI NMOS transistor (nominal case).

The second type of CMOS devices investigated was based on strained partially depleted silicon on insulator technology (PDSOI), where the silicon and the buried oxide thickness were 100 nm and 300 nm respectively.

Additionally, strained bulk NMOS and PMOS transistors with the same gate length and gate material were simulated for comparison, in order to investigate the possible advantages of FDSOI and PDSOI MOSFETs over bulk silicon with regard to process variation.

Silicon nitride layer (or contact etch stop liner (CESL)) was used as a dual stressor to increase the stresses in the channel for both the N and P type MOSFETs. A compression stress of 1.6 GPa to improve the hole mobility in the PMOS device was used, while a tension stress of 1.6 GPa was incorporated in the case of the NMOS device. The goal of the high-k dielectric was to continue the reduction of the effective oxide thickness (EOT), while keeping the gate leakage current under control.

In the process simulation, all necessary physical models were incorporated with the intention of defining and using the necessary models for a realistic simulation of the process and device conditions. The basic model for the complete process simulation consisted of a diffusion model with charged point defects, a transient dopant clustering model, a three-phase segregation model and a mechanical stress model, including the thermal and lattice mismatch as well as intrinsic stresses [16].

For an adequate simulation of the electrical performance of short channel MOSFETs, the correct choice of the device simulation models is important. The simulations included a hydrodynamic carrier transport and quantum corrections based on the density gradient model [16]. The mobility models included high-k dielectric enhancements to the Lombardi mobility model [16] to simulate the mobility degradation at the high-k dielectric interface with oxide in the gate stack. An important part of this work was the incorporation of quantum mechanical effects within the simulation modules developed for the NMOS and PMOS devices. This was necessary to determine the device characteristics more accurately.

To account for the mechanical stress effect on the electrical performance of the devices investigated based on SOI and bulk technologies, such as carrier mobility, threshold voltage and leakage current, the deformation potential model [16] was used. The

enhancement in terms of carrier mobility due to the carrier redistribution in the bands, inter-valley scattering, and stress-induced electron mobility, as well as the Intel stress-induced hole-mobility models [16], was also included.

Figure 3.8 shows the drain current-gate voltage (I_{ds}-V_{gs}) characteristics for simulated and experimental data under nominal process and operating conditions.
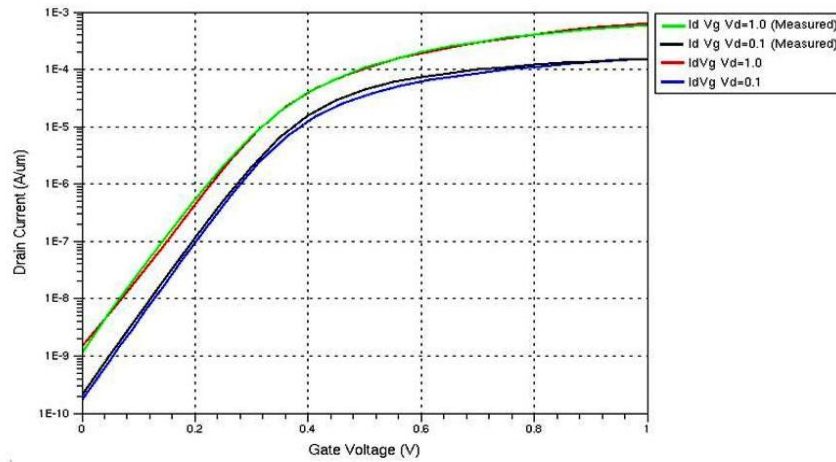
Figure 3.8: I_d-V_{gs} simulated and calibrated characteristics for NMOS (FDSOI) process at V_{ds}=0.1 and 1 Volts.

The compact models for the investigated technologies, SOI and bulk silicon (BSIMSOI [13] and BSIM3 [14]), were then extracted for the nominal NMOS and PMOS devices using I-V data under different bias conditions and for different device geometries using AURORA which is a parameter extraction and optimisation program that can fit analytical models to electrical data (either measured directly from a real device or simulated using TCAD) by varying one or more of the model parameters. A good match of gate and drain characteristics between the analytical models obtained with the compact model and the measured characteristics can be seen in Figure 3.9.

(a) **(b)**
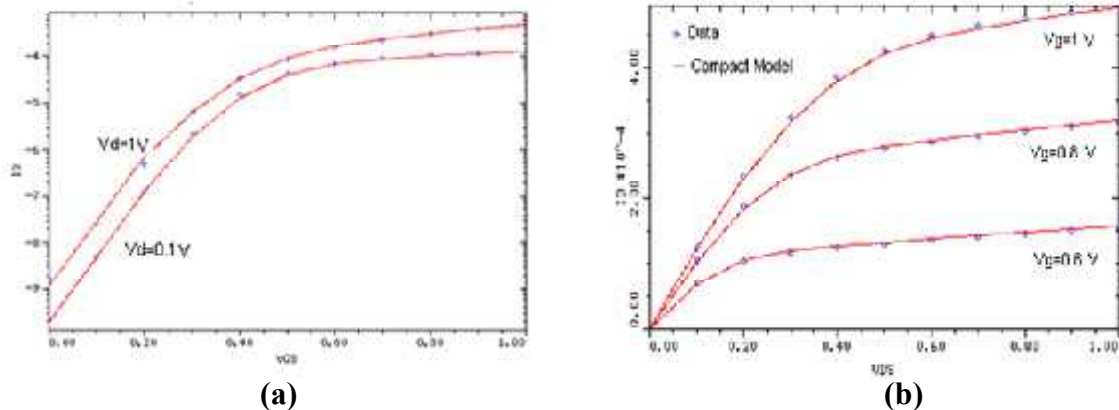
Figure 3.9: Compact model calibration of (a) gate characteristics and (b) drain characteristics for the nominal NMOSFET (FDSOI).

To study and analyse the impact of process variability on the compact model parameter (V_{th0}) of the investigated technologies, eighteen process parameters, shown in Table 3.4, were identified as potential sources of uncontrollable variation at different process steps for the devices based on FDSOI, PDSOI and bulk silicon technologies. The process parameters chosen were the same for both the NMOS and the PMOS devices.

All the process parameters were varied by ±10% of their mean values (i.e. the same standard deviations). However the process temperatures during the different manufacturing steps were set at ±10 ^{o} C from the nominal. This is mainly because the temperature values are very high and, in practice, would not drift in the range of ±10%. As shown in Table 3.4, the '-1' and '1' signs correspond to the -10% and +10% deviations respectively, assuming that the range of variation (± 10% and ±10° C) correspond to ±3σ variation. The assumption of a ±3σ variation is consistent with those adopted in previous research work [10, 17, 18].

Symbol	Description	Unit	Nominal level value	Low level value	High level value
			0	**-1**	**1**
x_1	Gate oxide thickness	nm	0.5	0.45	0.55
x_2	High-k dielectric thickness	nm	2.5	2.25	2.75
x_3	Buried oxide (Box) thickness	nm	145	130.5	159.5
x_4	Silicon layer thickness	nm	16	14.4	17.6
x_5	CESL layer thickness	nm	40	36	44
x_6	Intrinsic-Stress	GPa	1.6	1.44	1.76
x_7	Threshold voltage adjustment implantation dose	atoms/cm^2	$2e^{13}$	$1.8\,e^{13}$	$2.2\,e^{13}$
x_8	Threshold voltage adjustment implantation energy	KeV	25	22.5	27.5
x_9	Halo implantation dose	atoms/cm^2	$7.7e^{13}$	$6.93e^{13}$	$8.47e^{13}$
x_{10}	Halo implantation energy	KeV	25	22.5	27.5
x_{11}	Substrate doping (N_{sub})	atoms/cm^2	$1e^{17}$	$0.9\,e^{17}$	$1.1e^{17}$
x_{12}	Source/drain implantation dose	atoms/cm^2	$1.5e^{15}$	$1.35e^{15}$	$1.65e^{15}$
x_{13}	Source/drain implantation energy	KeV	0.6	0.54	0.66
x_{14}	Heavily doped drain implantation dose	atoms/cm^2	$7.5\,e^{15}$	$6.75e^{15}$	$8.25e^{15}$
x_{15}	Heavily doped drain implantation energy	KeV	7.5	6.75	8.25
x_{16}	Threshold voltage adjustment temperature	˚C	1050	1040	1060
x_{17}	Spike annealing temperature	˚C	950	940	960
x_{18}	Laser annealing temperature	˚C	600	590	610

Table 3.4 : Process parameters for variability study of 60 nm NMOS realised by strained FDSOI process.

3.4.2 Plackett-Burman Parameter Screening

Parameter screening is a crucial step with regards to identifying the most significant process parameters which will produce the greatest fluctuation in device electrical performance. Hence, it is extremely useful in understanding which manufacturing steps require greater control and focus.

Due to the fact that performing RSM analysis for a large input space requires a very large number of experimental runs (in the order of (2^n+2n+1), where n is the number of parameters [4] , so it becomes computationally inefficient, as is the case with the Monte-Carlo technique. In other words, screening analysis is adopted to overcome the deficiency of the RSM techniques by reducing the dimensionality of the input space. Therefore, RSM is followed by the screening step, wherein the relatively insignificant input parameters are eliminated, since not all the input variables are influential with regard to the output response to the same degree.

Plackett-Burman (PB) screening [4], due to its computational efficiency, was used to screen and identify the most dominant input parameters from an initial set of eighteen (18) process parameters. For the PB design in which k=18, the number of runs required, N, was 20. Table 3.5 shows the subsequent PB design matrix built using Minitab [19]. Twenty process and device simulations and compact model extractions were undertaken as part of the PB screening process, and the key device model parameters (V_{th0}) for both NMOS and PMOS devices were used to evaluate the screening experiment as output responses.

Runs	Process Parameters																	
	x1	x2	X3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15	x16	x17	x18
1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1
2	-1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1
3	-1	-1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1
4	1	-1	-1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1
5	1	1	-1	-1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1
6	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1
7	1	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1
8	-1	1	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1	-1	1	-1	1
9	1	-1	1	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1	-1	1	-1
10	-1	1	-1	1	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1	-1	1
11	1	-1	1	-1	1	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1	-1
12	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	-1	1	1	-1	-1	-1
13	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	-1	1	1	-1	-1
14	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	-1	1	1	-1
15	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	-1	1	1
16	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	-1	1
17	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1	-1
18	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1	1
19	1	-1	1	1	-1	-1	-1	-1	1	-1	1	-1	1	1	1	1	-1	-1
20	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Table 3.5: PB design matrix for k=18, N=20

The statistical analysis of the responses was subsequently performed using Pareto analysis. This analysis compares the relative magnitude of the influence of all the main input parameters on the output responses, and arranges them in order of the decreasing absolute value of the effect. The statistical significance of each input then can be identified.

A Pareto chart is a bar chart which lists all the factors according to the potential impact of their effects on the output response, ranging from the strongest to the weakest. All factors with demonstrate an effect above the dotted line have a significant influence on the response. This line is calculated by Minitab using the significant level alpha (α) which shows the maximum probability at which one could be wrong in making a statement about the significance of a factor. In other words, the confidence level, $100(1-\alpha)$ % actually refers to the probability that the statement is correct [19].

The Pareto analysis of effects for the compact model with parameters V_{th0} (NMOS) based on FDSOI, PDSOI and bulk Silicon, are shown in Figures 3.10, 3.11 and 3.12 respectively.
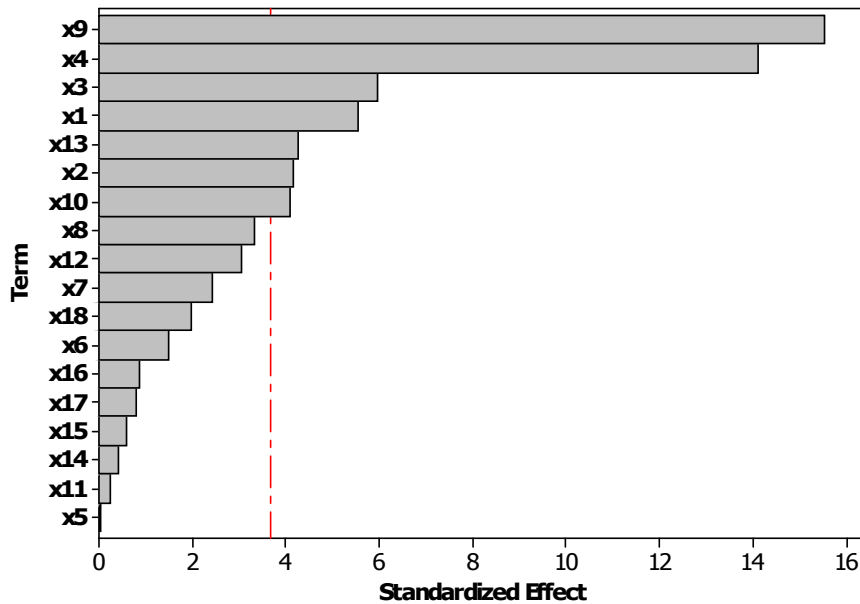
Figure 3.10: Pareto plot of the most significant process parameters on the V_{th0} of FDSOI-NMOS device.
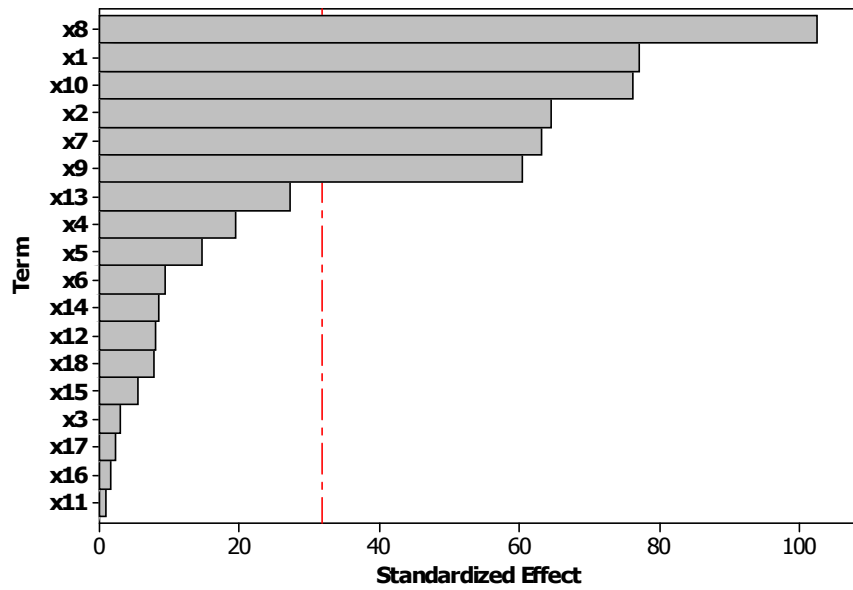
Figure 3.11: Pareto plot of the most significant process parameters on the V_{th0} of PDSOI-NMOS device.
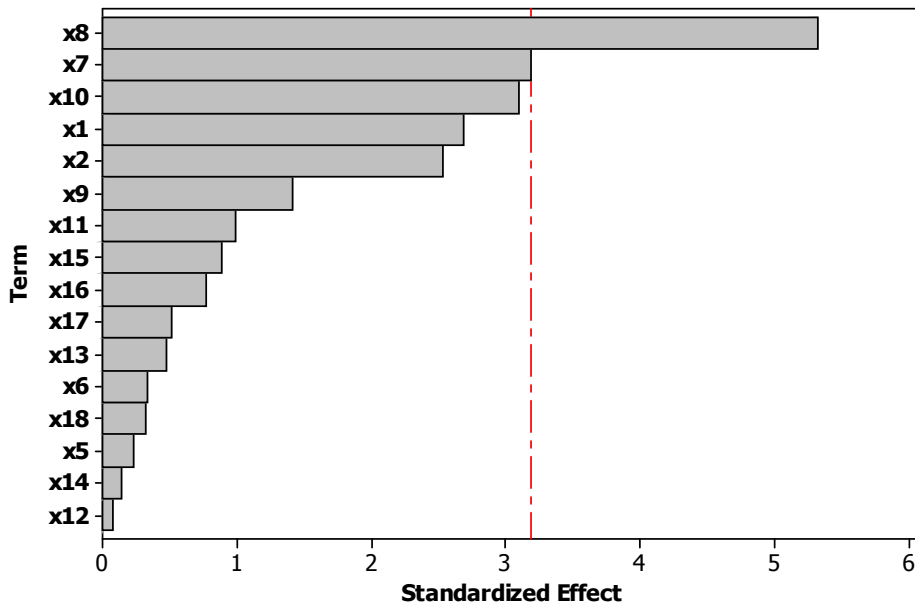
Figure 3.12: Pareto plot of the most significant process parameters on the V_{th0} of Silicon-NMOS device.

The Pareto analysis of the effects for the compact model parameters V_{th0} (PMOS) based on FDSOI, PDSOI, and bulk Silicon are shown in Figures 3.13 through Figure 3.15 respectively.
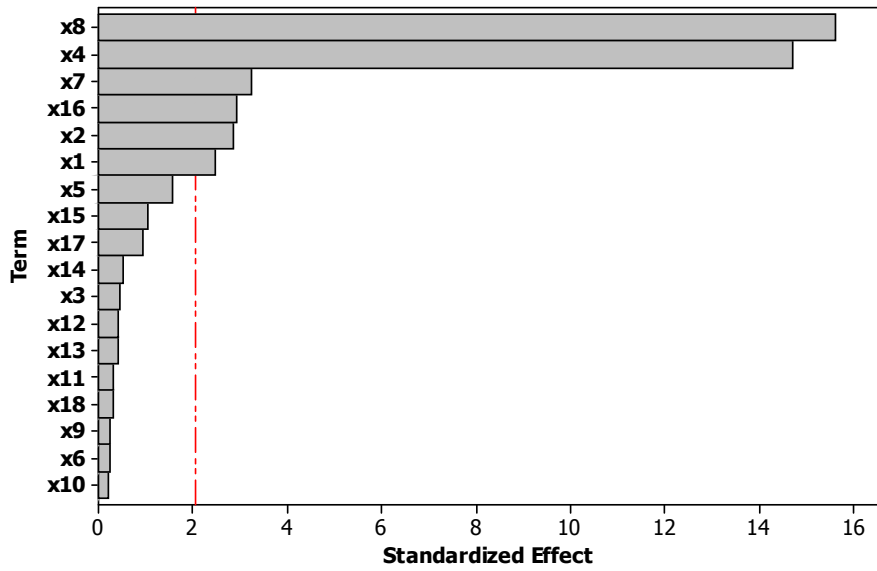
Figure 3.13: Pareto plot of the most significant process parameters on the V_{th0} of FDSOI-PMOS device.
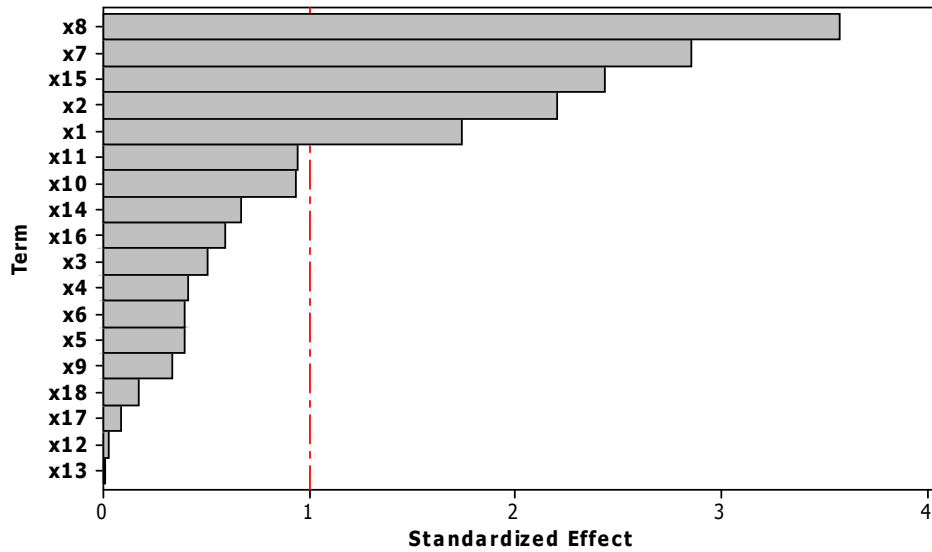
Figure 3.14 : Pareto plot of the most significant process parameters on the V_{th0} of PDSOI-PMOS device.
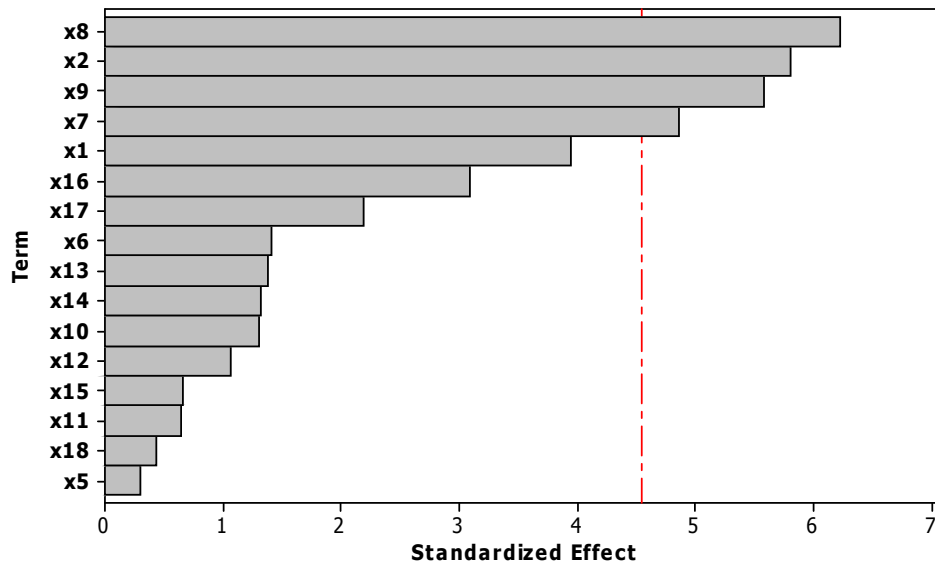
Figure 3.15 : Pareto plot of the most significant process parameters on the V_{th0} of Silicon-PMOS device.

From the Pareto plots of V_{th0} of the NMOS and PMOS devices based on FDSOI technology, it can be seen that the silicon layer thickness (x_4) and the buried oxide thickness (x_3) have an influential role on the threshold voltage of the devices. It can also be seen that the effect of the process parameters is not the same for both the NMOS and PMOS devices. This might be explained by the use of different dopants and stressors in the case of PMOS with regard to NMOS devices.

Focussing on the PDSOI devices, it can be seen that the V_{th0} sensitivity for both NMOS and PMOS devices to the SOI structural features (i.e. silicon layer thickness and buried oxide thickness) is minimal and the doping related processes such as the threshold voltage adjustment implantation dose and the halo dose have become the dominant parameters. This overcomes the problem associated when using chemical mechanical polishing (CMP) and lithography in terms of controlling the thin silicon layer in FDSOI technology.

In the case of silicon technology, it can be seen from the graphs in Figures 3.12 and 3.15, that the dominant process parameters that need to be controlled to reduce the impact, are those related to gate oxide and high k dielectrics, and the doping related process steps, in particular those used for adjusting the threshold voltage of the devices.

From Pareto plots shown in Figures 3.10-3.15, it can also be seen that the halo implantation dose (x_9) plays an important role in controlling V_{th0} of NMOS device for both FDSOI and PDSOI technologies, while it has a minimum effect on V_{th0} of PMOS devices. This is in contrary to silicon devices, where the halo implantation step becomes an influential step that affects V_{th0} for both NMOS and PMOS devices as can be seen from Figures 3.12 and 3.15. This might be explained by the presence of buried oxide (BOX) and silicon layers which might affect on the doping profile in SOI devices and hence their V_{th0} Values.

Table 3.6 summarises the most significant process steps that influence V_{th0} for the FDSOI, PDSOI, and Silicon technologies for both NMOS and PMOS devices.

Technology/Device type	NMOS	PMOS
FDSOI	Halo implantation dose (x_9) Silicon layer thickness (x_4). Buried oxide thickness (x_3). Gate oxide thickness (x_1). Source/drain implantation energy (x_{13}). High-k dielectric thickness (x_2). Halo implantation energy (x_{10}).	Threshold voltage adjustment implantation energy (x_8). Silicon layer thickness (x_4). Threshold voltage adjustment implantation dose (x_7). Threshold diffusion temperature (x_{16}). High-k dielectric thickness (x_2). Gate oxide thickness (x_1).
PDSOI	Threshold voltage adjustment implantation energy (x_8). Gate oxide thickness (x_1). Halo implantation energy (x_{10}) High-k dielectric thickness (x_2). Threshold voltage adjustment implantation dose (x_7). Halo implantation dose (x_9).	Threshold voltage adjustment implantation energy (x_8). Threshold voltage adjustment implantation dose (x_7). Heavily doped drain implantation energy (x_{15}). High-k dielectric thickness (x_2). Gate oxide thickness (x_1).
Silicon	Threshold voltage adjustment implantation energy (x_8). Threshold voltage adjustment implantation dose (x_7). Halo implantation energy (x_{10}). Gate oxide thickness (x_1). High-k dielectric thickness (x_2).	Threshold voltage adjustment implantation energy (x_8). High-k dielectric thickness (x_2). Halo implantation dose (x_9). Threshold voltage adjustment implantation dose (x_7).

Table 3.6: The most significant process parameters that impact V_{th0} for both NMOS and PMOS devices based on FDSOI, PDSOI, and bulk silicon.

Having identified the most significant process parameters for the device responses in the screening steps for all technologies investigated in terms of the compact model parameter, namely (V_{th0}), the RS models for these compact model parameters, in terms of the significant process parameters, were subsequently built. The FCCC design was used to investigate process variability, and to build second order RS models of the responses.

3.4.3 Response Surface Modelling for the Device Responses

After the process parameters were identified from the PB screening analysis as significantly influencing the compact model parameters in terms of responses, a second order RSM procedure was performed on these significant parameters. An FCCC design for k significant parameters with 2^k factorial runs, 2k axial runs and one centre run requires, in total, (2^k +2k+1) runs. As an example, 143 process and device simulations have been performed to generate a second order RSM for V_{th0} of an FDSOI-NMOS device with k=7. The second order model was obtained by a regression technique using the simulation results. It is in the form of equation (3.4) and given by equation (3.5).

$$
\begin{aligned}
V_{th0}(V) =\ & 0.25597 + 0.00861x_1 + 0.008527x_2 - 0.00907x_3 + 0.022255x_4 + 0.026563x_9 \\
& + 0.007636x_{10} - 0.00666x_{13} + 0.00323x_1^2 + 0.005011x_2^2 - 0.00801x_3^2 - 0.02236x_4^2 \\
& + 0.002965x_{10}^2 + 0.008698x_{13}^2 + 0.001963x_3x_{10} - 0.00566x_3x_{13} + 0.001339x_4x_9 \\
& - 0.00146x_9x_{10} - 0.00177x_9x_{13}
\end{aligned}
\tag{3.5}
$$

In a similar manner, the RSM models for V_{th0} of an NMOS device realised by PDSOI and bulk silicon technologies are given by Equations (3.6) and (3.7) respectively.

$$
\begin{aligned}
V_{th0}(V) =\ & 0.370225 + 0.010762x_1 + 0.010152x_2 + 0.009186x_7 - 0.0187x_8 + 0.010069x_9 \\
& + 0.012324x_{10} - 0.00352x_1^2 - 0.00143x_2^2 - 0.00127x_7^2 + 0.001932x_8^2 + 0.005311x_9^2 \\
& - 0.00183x_{10}^2
\end{aligned}
\tag{3.6}
$$

$$
\begin{aligned}
V_{th0}(V) =\ & 0.393716 + 0.015071x_1 + 0.013385x_2 + 0.015837x_7 - 0.02599x_8 + 0.01428x_{10} \\
& + 0.001913x_1^2 + 0.001699x_8^2 + 0.001875x_1x_2 + 0.002367x_1x_7 - 0.00346x_1x_8 + 0.002971x_2x_{10} \\
& - 0.00239x_7x_8 + 0.002015x_7x_{10} - 0.00213x_8x_{10}
\end{aligned}
\tag{3.7}
$$

The RSM models were also generated for V_{th0} of PMOS devices realised by FDSOI, PDSOI and bulk silicon technologies, the models are given by Equations (3.8), (3.9), and (3.10) respectively.

$$V_{th0}(V) = -0.30856 - 0.05662x_4 - 0.01573x_1 - 0.01757x_2 - 0.01673x_7 + 0.066164x_8$$
$$+ 0.014206x_{16} + 0.039041x_4^2 + 0.003412x_1^2 + 0.002989x_2^2 - 0.02047x_7^2$$
$$- 0.02895x_8^2 + 0.005407x_{16}^2 - 0.00111x_4x_1 - 0.001x_4x_2 + 0.002876x_4x_7$$
$$- 0.01573x_5x_8 - 0.00691x_4x_{16} + 0.00284x_1x_8 + 0.002739x_2x_8 0.005286x_7x_8$$
$$+ 0.00394x_8x_{16}$$

(3.8)

$$V_{th0}(V) = -0.3234 - 0.02133x_1 - 0.02713x_2 + 0.023471x_{15} - 0.01096x_7$$
$$+ 0.04828x_8 - 0.00915x_1^2 - 0.00968x_2^2 + 0.003442x_{15}^2 - 0.02662x_7^2$$
$$- 0.00061x_8^2 + 0.003405x_1x_2 + -0.00119x_1x_7 + 0.001843x_1x_8$$
$$- 0.00134x_2x_7 + 0.002312x_2x_8 + 0.005018x_{15}x_8$$

(3.9)

$$V_{th0}(V) = -0.38124 - 0.02491x_2 - 0.01467x_7 + 0.037639x_8 - 0.02754x_9$$
$$+ 0.002278x_2^2 - 0.0219x_7^2 - 0.00671x_8^2 + 0.006229x_9^2 + 0.001518x_2x_8$$
$$- 0.00167x_2x_9 + 0.005764x_7x_8$$

(3.10)

Having built the second-order variability models for the compact model parameters of the devices, a variability analysis was performed on each of the responses with respect to the process parameters of each device, and are described below.

3.4.4 Variability Analysis of the Compact Model Parameters

The variability analysis was carried out for all the devices responses based on the model equations. The analysis was performed using the RS plots such as contour and three dimensional (3D) surface plots. Where, in a surface plot, the response surface is viewed as a three-dimensional (3D) plane, and the response is represented as the functional relationship with respect to two input parameters of interest, while all the other parameters are held at their nominal values.

Figure 3.16(a) shows the response surface for V_{th0} for an FDSOI-NMOS device as a function of gate-oxide and high-k dielectric thickness, while the other significant parameters are kept at their nominal values. The impact of gate-oxide thickness and high-k dielectric on the V_{th0} is obvious from the response surface plot. At low levels of gate-oxide and high k thicknesses corresponding to thinner gate-oxide and high-k dielectrics

thicknesses, the threshold voltage has decreased. However at higher levels, the V_{th0} has increased by almost 16%, it can also be seen that the curvature is reduced when using thicker high-k dielectric thickness.
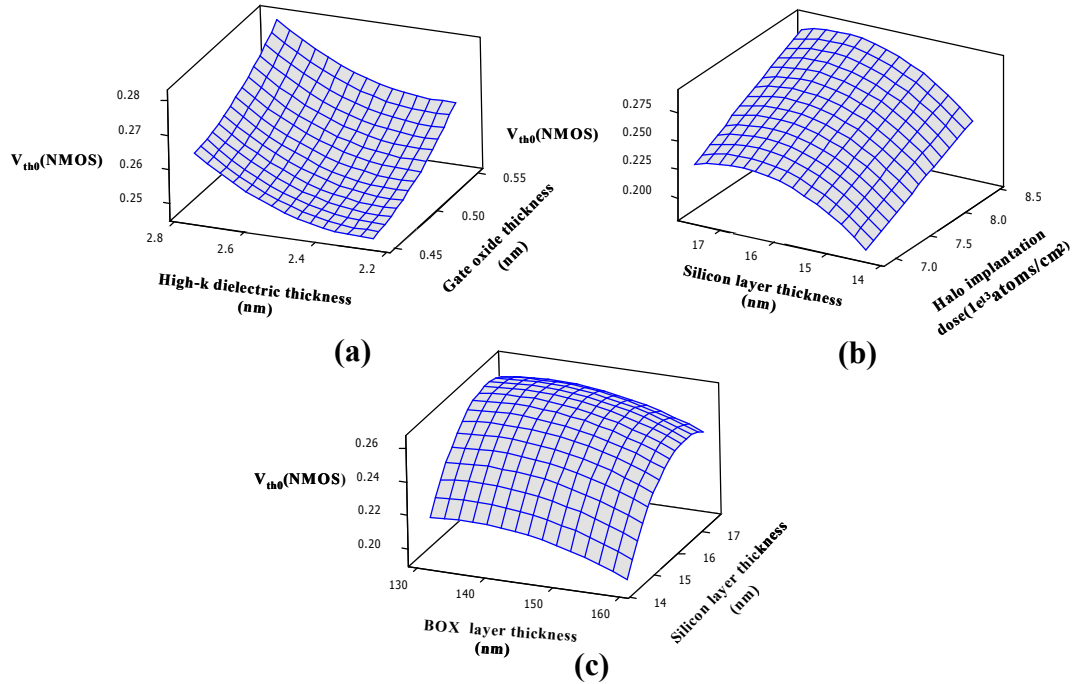
Figure 3.16: Response surface for V_{th0} (FDSOI-NMOS) with respect to(a) gate-oxide thickness and high-k dielectric thickness (b) silicon layer thickness and halo implantation dose.(c) silicon layer thickness and buried oxide (BOX) thickness.

In Figure 3.16(b), the response surface for V_{th0} of the FDSOI–NMOS device as a function of the halo implantation dose and the silicon layer thickness, while the other significant parameters are kept at their nominal values. The impact of silicon layer thickness on the V_{th0} is large, as can be seen from the response surface plot. For thinner values of silicon layer thickness, the threshold voltage is low. However, the curvature between the low and the nominal values is large, while for higher value of the silicon layer thickness, the V_{th0} curvature is reduced. This means that adopting a thicker silicon layer can reduce the impact of the silicon layer thickness variation on the threshold voltage.

In a similar manner, the response surface for V_{th0} of the FDSOI–NMOS device as a function of halo implantation dose and silicon layer thickness is shown in Figure 3.16(b). The impact of silicon layer thickness and buried oxide thickness (BOX) on the V_{th0} is

large, and gives a variation of almost 25% as shown in Figure 3.16(c). However, this impact can be minimised when using higher values of silicon layer thickness.

Figure 3.17(a) shows the response surface for V_{th0} of the PDSOI–NMOS device as a function of gate-oxide thickness and high-k dielectric thickness, while the other significant parameters are kept at their nominal values. The impact of gate-oxide thickness and high-k dielectric on V_{th0} is obvious from the response surface plot. At low levels of gate-oxide and high k thicknesses corresponding to thinner gate-oxide and high-k dielectric, the threshold voltage has decreased. However, at their high levels, V_{th0} has increased by almost 9%.
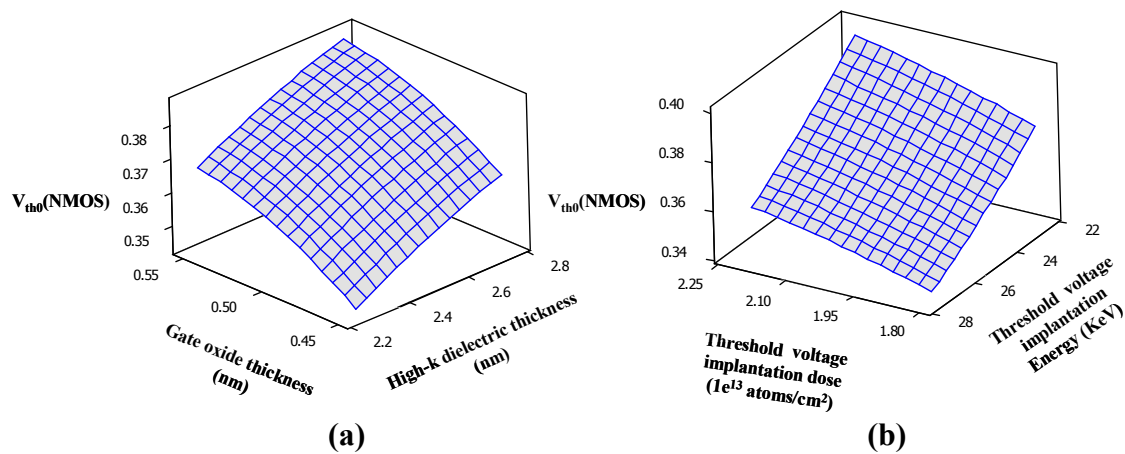
Figure 3.17: Response surface for V_{th0} (PDSOI-NMOS) with respect to gate-oxide thickness and high-k dielectric thickness (b) threshold voltage adjustment implantation energy and threshold voltage adjustment implantation dose.

Figure 3.17(b) shows the response surface for V_{th0} of the PDSOI–NMOS device as a function of threshold voltage adjustment implantation energy and threshold voltage adjustment implantation dose. The impact of these parameters on V_{th0} is large, and the variation of V_{th0} under these conditions is almost 14%.

Figure 3.18(a) shows the response surface for V_{th0} of the bulk silicon–NMOS device as a function of gate-oxide thickness and high k thickness. The impact of gate-oxide thickness and high-k dielectric thickness on V_{th0} is apparent from the response surface plot. The variation in V_{th0} ranges by almost 9 % in response to gate-oxide and high k dielectric thicknesses variations.
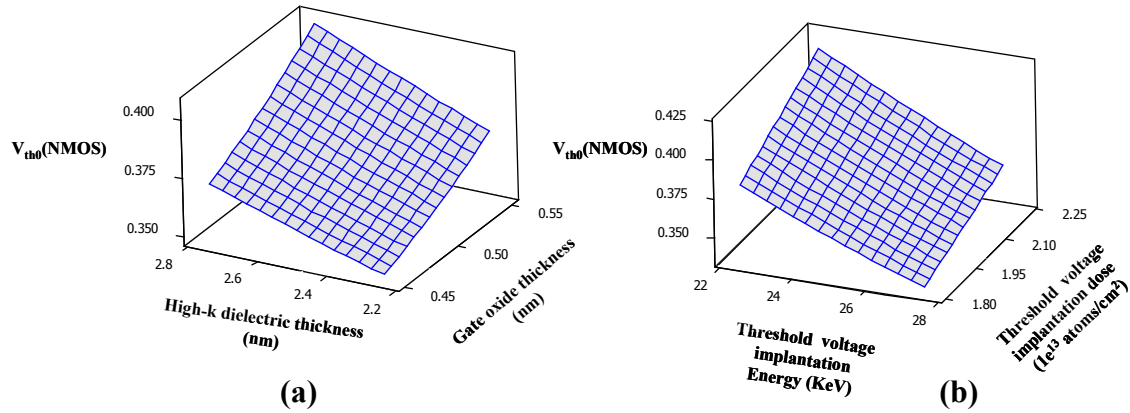
Figure 3.18: Response surface for V_{th0} (bulk silicon-NMOS) with respect to (a) gate-oxide thickness and high-k dielectric thickness (b) threshold voltage adjustment implantation energy and dose.

In addition, as can be seen from Figure 3.18(b), the variation of V_{th0} relative to variations in the threshold voltage adjustment implantation energy and threshold adjustment dose is almost 17%.

RSM analysis was also performed for V_{th0} of the PMOS devices realised by different technologies. Figures 3.19(a) and (b) show the RS plots for the PMOS device based on FDSOI technology. It can be seen that the silicon layer thickness, as with FDSOI-NMOS, is dominant and causes large variations in V_{th0} (PMOS). However, adopting higher silicon layer values can reduce the variation due the small curvature of V_{th0} as shown in Figure 3.19(a).

In Figure 3.19(b), lower threshold values can be achieved when using thinner gate oxide and high-k dielectric. However, they have a great influence on the total variation in V_{th0} (PMOS).
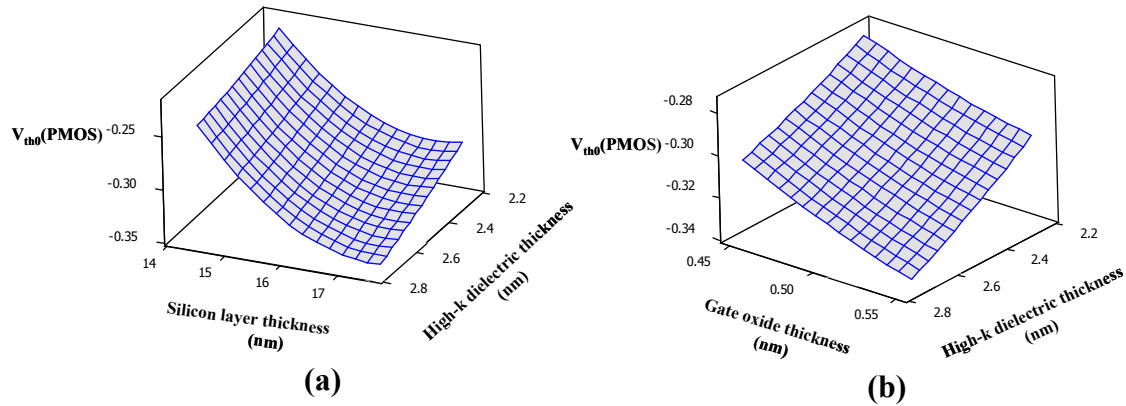
Figure 3.19: Response surface for Vth$_0$ (FDSOI-PMOS) with respect to (a) silicon layer thickness and high-k dielectric thickness (b) gate-oxide thickness and high-k thickness.

Figure 3.20(a) shows the response surface for V_{th0} of the PDSOI–PMOS device as a function of threshold voltage adjustment implantation energy and dose. It can be seen from the RS plot that the impact of these process parameters on V_{th0} is large and is mainly due to the variations in the threshold voltage adjustment implantation energy. Moreover, the impact of both gate oxide thickness and high-k dielectric thickness on V_{th0} of the PDSOI–PMOS is large as shown in Figure 3.20(b).
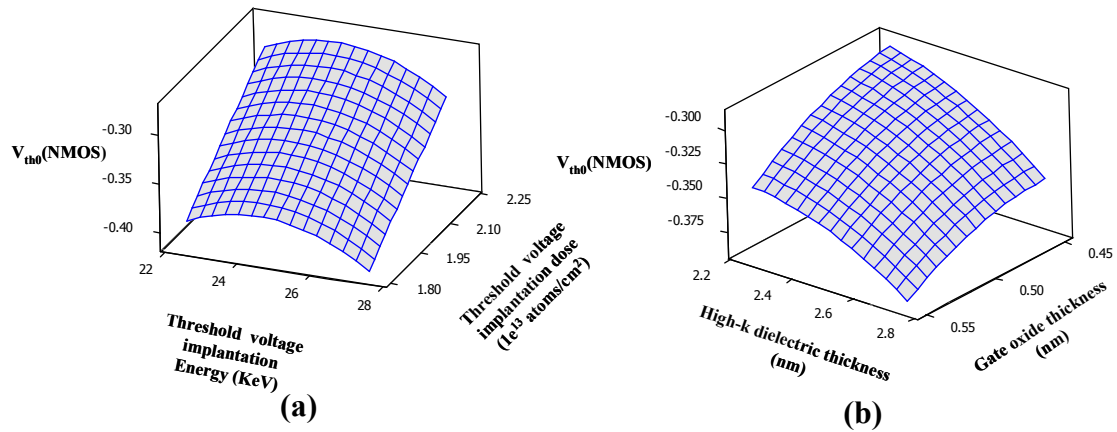
Figure 3.20: Response surface for Vth$_0$ (PDSOI-PMOS) with respect to threshold voltage adjustment implantation dose and threshold voltage adjustment implantation energy (b) gate-oxide thickness and high-k dielectric thickness.

In the case of the silicon device, Figure 3.21(a) shows the response surface for V_{th0} of the PMOS device with respect to threshold voltage adjustment implantation energy and threshold voltage adjustment implantation dose. The impact of these parameters on V_{th0} is large, especially due to the energy of the implantation.

The impact of the high-k dielectric thickness variation on V_{th0} (PMOS) is large as can be seen in Figure 3.21(b), while the variation in the threshold implantation dose shows less impact on the threshold voltage of the PMOS-Silicon device.

<div align="center">(a)   (b)</div>
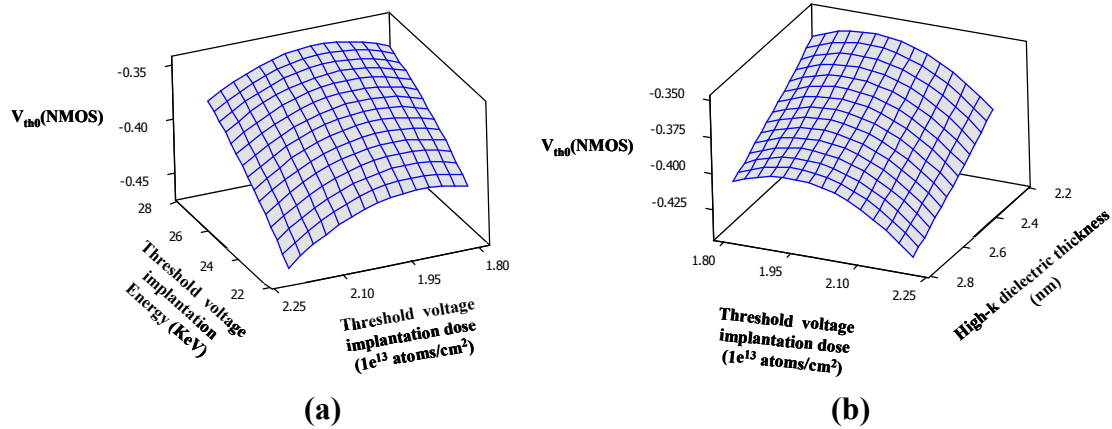
Figure 3.21: Response surface for V_{th0} (Silicon-PMOS) with respect to (a) threshold voltage adjustment implantation dose and threshold voltage adjustment implantation energy (b) threshold adjustment implantation dose and high-k dielectric thickness.

In general, from the results obtained using DoE/RSM, it is possible to improve device yield by optimizing the process parameters such as silicon layer thickness and buried oxide thickness in order to reduce their impact electrical characteristics of the device.

Moreover, RSM models can bridge the wide gap between the circuit designer and the device manufacturer and hence can improve both circuit performance and yield.

Table.3.8 compares the mean (μ) and (σ) values of V_{th0} for both NMOS and PMOS devices realised by FDSOI, PDSOI, and bulk silicon technologies.

It can be seen from Table 3.8 that the NMOS and PMOS devices realised in FDSOI technology have the lowest threshold voltage mean values compared to PDSOI and bulk silicon devices, however in terms of the effect of process variations, the FDSOI devices show the highest standard deviations (σ) of V_{th0} for both NMOS and PMOS devices compared to PDSOI and bulk silicon counterparts. It can also be seen that the variation in PMOS devices is much higher compared to NMOS devices for all technology investigated.

The variation in the threshold voltage of PMOS bulk silicon is the lowest compare to FDSOI and PDSOI devices. However, the PMOS–bulk silicon device has higher mean value of the threshold voltage in comparison to FDSOI and PDSOI devices.

Device Type	Technology	V_{th0}			
		μ	σ	Min	Max
NMOS	FDSOI	0.2463	0.039	0.147	0.333
	PDSOI	0.3695	0.029	0.29	0.453
	Silicon	0.3975	0.0362	0.307	0.4733
PMOS	FDSOI	-0.3192	0.0707	-0.408	-0.236
	PDSOI	-0.3562	0.061	-0.506	-0.239
	Silicon	-0.3957	0.049	-0.531	-0.29

Table 3.7: .The μ , σ, minimum and maximum values for V_{th0} for both NMOS and PMOS devices realised by FDSOI, PDSOI, and bulk silicon technologies.

3.4.5 Models Validity of the Device Responses

In this work, the accuracy of the response surface models were checked and verified using the 'goodness' of the second-order fit, such as R^2 (R-square) and R^2_{adj} (adjusted R-square) statistical measures [4, 5], where R^2 is a statistical measure which indicates how close the regression line is to the actual data points. The R^2 value can be increased by incorporating additional parameters into the model which may not be statistically significant. Hence, models with large R^2 values can give an undesired fit, and hence a poor estimation of the output response. R^2_{adj} is a modified R^2 for the number of terms in the model. In contrast to R^2, R^2_{adj} becomes smaller when unnecessary terms are added to the model. Tables 3.8 presents the resulting model fits obtained for the device compact model parameters (V_{th0}) for both NMOS and PMOS devices. It can be seen that the R^2 and R^2_{adj} values are very close to 100% for all responses. This is desirable and therefore ensures that the models accurately highlight the variability due to process fluctuations.

Device Type	Technology	V_{th0}	
		R^2	R^2_{adj}
NMOS	FDSOI	97.58%	96.79%
	PDSOI	94.09%	90.84%
	Silicon	97.93%	96.05%
PMOS	FDSOI	98.88%	97.11%
	PDSOI	98.53%	96.99%
	Silicon	98.38%	96.10%

Table 3.8: The R^2 and R^2_{adj} values for V_{th0} both NMOS and PMOS devices realised by FDSOI, PDSOI, and bulk silicon technologies.

3.5 Summary and Conclusions

An overview of DoE and RSM techniques has been presented in some detail in this chapter. The different experimental design methods such as factorial, screening and RS designs have also been discussed.

A statistical methodology based on the DoE and RSM techniques with the aid of TCAD, has been utilised to study and analyse process variability in state of the art technologies, namely strained fully depleted silicon on insulator, strained partially depleted silicon on insulator and strained bulk silicon, all with high-k dielectrics in the gate stack.

The process and device simulations were first performed, followed by the calibration of the device characteristics to the experimental results, and suitable compact models have then been extracted.

The PB screening technique has been used to characterise and screen the most dominant process parameters from an initial set of 18 which have greatest influence on the device compact parameters. The screening step improves the computational efficiency for the second order RSM by reducing size of the input space, and also helps in identifying input parameters that need to be kept under control.

Second-order RSM models, using FCCC design, were subsequently developed to model the device response parameter (V_{th0}) and evaluate the variability. A thorough RSM

investigation has been undertaken through response surface plots, using the models obtained for the device compact model parameters as functions of the process variants.

It was found that the variation in silicon layer thickness greatly impacts on the threshold voltage (V_{th0}) for both NMOS and PMOS devices based on FDSOI technology. However, this sensitivity is largely mitigated when using PDSOI technology.

The threshold voltage adjustment implantation dose and energy, along with gate-oxide thickness and high-k dielectric thickness, are found to be the most influential and common process parameters for most of the threshold voltage responses based on the technology investigated.

The MOSFET devices realised in FDSOI technology were found to have the lowest threshold voltage mean values compared to PDSOI and bulk silicon devices, however in terms of the effect of process variations, the FDSOI devices show the highest standard deviations (σ) of V_{th0} for both NMOS and PMOS devices compared to PDSOI and bulk silicon counterparts.

It was also found that the variation in PMOS devices is much higher compared NMOS devices for all technology investigated.

The variation in the threshold voltage of PMOS bulk silicon is the lowest compare to FDSOI and PDSOI devices. However, the PMOS–bulk silicon device has higher mean value of the threshold voltage in comparison to FDSOI and PDSOI devices.

The models obtained can be used for optimisation and process control purposes in the semiconductor manufacturing environment.

Finally, the accuracy of the models has been successfully examined and the goodness of fit, R^2 and R^2_{adj} are close to 100% for most of the models in terms of the device responses with regard to all the technologies investigated.

3.6 References

[1] T. J. Sanders and M. J. Phelps, "Simulation of the effects of manufacturing process variations on the characteristics of SOI MOSFETs," in SOI Conference, 1995. Proceedings., 1995 IEEE International, pp. 24-25,1995.

[2] M. J. Phelps, G. T. Hess, and T. J. Sanders, "Statistical design for manufacturing software for low power radiation hardened SOI MOSFETs," in University/Government/Industry Microelectronics Symposium, 1997., Proceedings of the Twelfth Biennial, pp. 98-101, 1997.

[3] A. N. Chandorkar, S. Mande, and H. Iwai, "Estimation of process variation impact on DG-FinFET device performance using Plackett-Burman design of experiment method," in Solid-State and Integrated-Circuit Technology, 2008. ICSICT 2008. 9th International Conference on, pp. 215-218, 2008

[4] Montgomery, "Design and Analysis of Experiments," 5th ed: John Wiley and Sons, 2001.

[5] G. E. P. B. a. N. R. Draper, "Empirical Model-Building and Response Surfaces," John Wiley and Sons, 1987.

[6] S. K. a. Y. Leblebici, "CMOS Digital Integrated Circuits," 2nd ed: Mc-Graw Hill, 1999.

[7] S. W. K. C.P. Xu, H.J. Hwang, J.W. Yun "Application of statistically based experimental designs for the optimization of exopolysaccharide production by Cordyceps milltaris NG3," Appl. Biochem, vol. 36, pp. 127–131, 2002.

[8] S. SHEDABALE, "Statistical Modelling of Process Variations in CMOS Devices and Circuits," PhD thesis,Newcastle University, April 2009.

[9] R. L. Plackett and J. P. Burman, "The design of optimum multifactorial experiments," Biometrika, vol. 33, pp. 305-325, 1946.

[10] A. A. Mutlu and M. Rahman, "Statistical methods for the estimation of process variation effects on circuit operation," IEEE Transactions on Electronics Packaging Manufacturing, vol. 28, pp. 364-375, 2005.

[11] D. S. Boning and P. K. Mozumder, "DOE/Opt: a system for design of experiments, response surface modeling, and optimization using process and device simulation," IEEE Transactions on Semiconductor Manufacturing, vol. 7, pp. 233-244, 1994.

[12] A. R. Alvarez, B. L. Abdi, D. L. Young, H. D. Weed, J. Teplik, and E. R. Herald, "Application of statistical design and response surface methods to computer-aided VLSI device design," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 7, pp. 272-288, 1988.

[13] BSIMSOI [Online]. Available:http://www-device.eecs.berkeley.edu/~bsimsoi.

[14] BSIM3 [Online]. Available:http://www-device.eecs.berkeley.edu/~bsim3.

[15] Sentaurus TCAD, Release z-2007.12 ed., Synopsys, Mountain View, CA,USA, 2007.

[16] Sentaurus Device User Guide. Version Z-2007.12.2007. Synopsys.

[17] S. Williams and K. Varahramyan, "A new TCAD-based statistical methodology for the optimization and sensitivity analysis of semiconductor technologies," IEEE Transactions on Semiconductor Manufacturing, vol. 13, pp. 208-218, 2000.

[18]   H. Ramakrishnan, S. Shedabale, G. Russell, and A. Yakovlev, "Analysing the effect of process variation to reduce parametric yield loss," IEEE International Conference on Integrated Circuit Design and Technology and Tutorial, pp. 171-176, 2-4 June 2008.

[19]   Minitab, Minitab Release 15, Statistical Software. Minitab Inc., 2006.

Chapter 4

An Analysis of the Effect of Process Variations on Performance of C-Element Structures

4.1 Background

In this chapter, the response surface methodology is extended to the circuit-level, by studying the impact of process variations and environmental operating conditions on various C-element circuits. The extracted variations in terms of threshold voltages of both NMOS and PMOS devices realised by the technologies investigated in Chapter 3, are considered, together with the gate length of the devices and power supply voltage variations. Subsequently, RS models are developed and used to perform response surface analysis for circuit performance metrics such as dynamic and static power consumptions and propagation delay; subsequently giving useful information about the sensitivity of each metric with respect to the process parameters under consideration. The influence of circuit architectures, together with different process technologies on performance variations in terms of dynamic and static power consumption and circuit delay, is also investigated. A relative comparison between circuits based on FDSOI, PDSOI and bulk silicon in terms of delay and power consumption is subsequently undertaken.

An in-depth analysis of the behaviour of the circuits under process variations was undertaken by studying the impact of process variations on transistor channel currents. The investigation was carried out by measuring the maximum current in each transistor under process variations during circuit operation. Using this method, it was possible to first, explain why the circuit topology plays a significant role in performance variation; second to identify which transistors make the most significant contribution to the final performance variation; and third to target those transistors in such a way as to minimise their impact on performance variability by resizing their widths.

The DoE and RSM is utilised to investigate the direct impact of manufacturing process steps on the performance of circuits based on the technologies investigated. This is done

with the aim of determining and identifying the critical process parameters that need to be under control and hence finding ways to reduce their effect on performance and on the yield of a circuit.

4.2 Process Variation

As technology scales down to the nanometre regime, the effects of semiconductor process variations are an increasingly significant and critical factor in the design of high performance CMOS circuits. Furthermore, as the process steps involved in the manufacture an IC becomes more complex as the overall circuit performance is now more sensitive to the underlying statistical process variations. Hence a precise knowledge of the effects of process variations is emerging as an integral part of the design methodology.

This variation negatively impacts on circuit performance, and generates design specifications which are harder to meet, resulting in functional and parametric yield loss. Moreover, with technology scaling, these problems are becoming increasingly critical due to the tighter design requirements. Novel process steps are used in the nanometre regime to enhance the performance of the device and of circuits, such as the introduction of high –k dielectric and strained silicon technologies. However, this has also increased the number of process variants affecting the device and circuit performance. Hence there is a need to understand and model manufacturing process variations for the prediction of device and circuit performance as a result of such variations.

On the other hand, accommodating effects of a random nature in process parameters on circuit functionality has become a major design challenge in the ultra DSM technology regime. Therefore new material and device structures are required to minimise the variation effects. In this regard, SOI technology can be seen as a possible candidate as a means of improving the resilience of circuits to performance variation, and hence to increase its yield. Another way of combating the effect of variations, is the use of different circuit architectures. As an example consider a logic element which is not only widely used, but also peculiar to asynchronous design, namely the Muller C-element,

which can be realised in a number of different configurations. In view of the increased uncertainty of logic elements with regard to the effects of manufacturing processes and environment variations as device geometries are reduced, this chapter reports on the analysis of the robustness of various C-element configurations implemented in different technologies, with regard to the effects of process variations.

4.3 The C-Element

In recent years a great deal of research and effort has demonstrated the potential of asynchronous circuits through many low-power and high-speed applications. For example, The Philips research group had designed a fully asynchronous error corrector for the digital compact-cassette (DCC) player which consumed only a fifth of the power in a similar synchronous version [1]. Some of the advantages of asynchronous circuits are high speed, low power consumption, design modularity, and freedom from clock skew.

One of the most basic circuits that is frequently used in asynchronous design as a control circuit is the C-element [2]. For example, the C-element played an important role in the micro pipeline of the ARM microprocessor [2] .

In 1959, David E. Muller introduced the C-element circuit. Consequently, the C-element circuit is often referred as the "Muller C-element" [3]. A C-element has two inputs and one output and its function can be described as follows: If both inputs have the same logic level 0/1, then the output becomes 0/1; otherwise the output is maintained at the pervious logic level. Figure 4.1 shows the commonly used schematic of the C-element. The function of the C-element can be described in terms of a Boolean function of its inputs and output as in Equation (4.1).
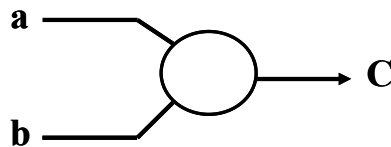
Figure 4.1: The schematic diagram of the C-element.

$$C = \overline{C} \bullet (a + b) + a \bullet b \qquad (4.1)$$

In asynchronous circuits, the C-element is often referred as a Join or Rendezvous element [2] and it could be expressed as an AND operator of events, where the event could be a rising or falling transition. Thus, it only produces an output transition (event) when both inputs have arrived.

In the following section, a brief view of the implementations of the C-elements studied is presented. This includes the dynamic implementation, the static implementation (standard, weak feedback, resistive weak feedback and symmetric configurations), and the differential implementations of the C-element.

4.3.1 C-Element Configurations

Figure 4.2(a) shows the dynamic implementation of the C-element. This consists of the basic parts of all the C-element implementations discussed below. The other static and differential implementations of the C-element differ only in terms of preserving the output logic level when the inputs are different.

The implementation of the C-element with weak feedback inverter, as shown in Figure 4.2(b), has been introduced by Martin [6]. The circuit contains a weak inverter in order to create a latch to maintain the output logic level when the inputs do not match. The reason for choosing a weak inverter is to allow changes in the output of the latch, and to reduce the inherent resistance of the latch to switching its logic state. For a proper circuit operation, certain width sizes must be imposed on the transistors.
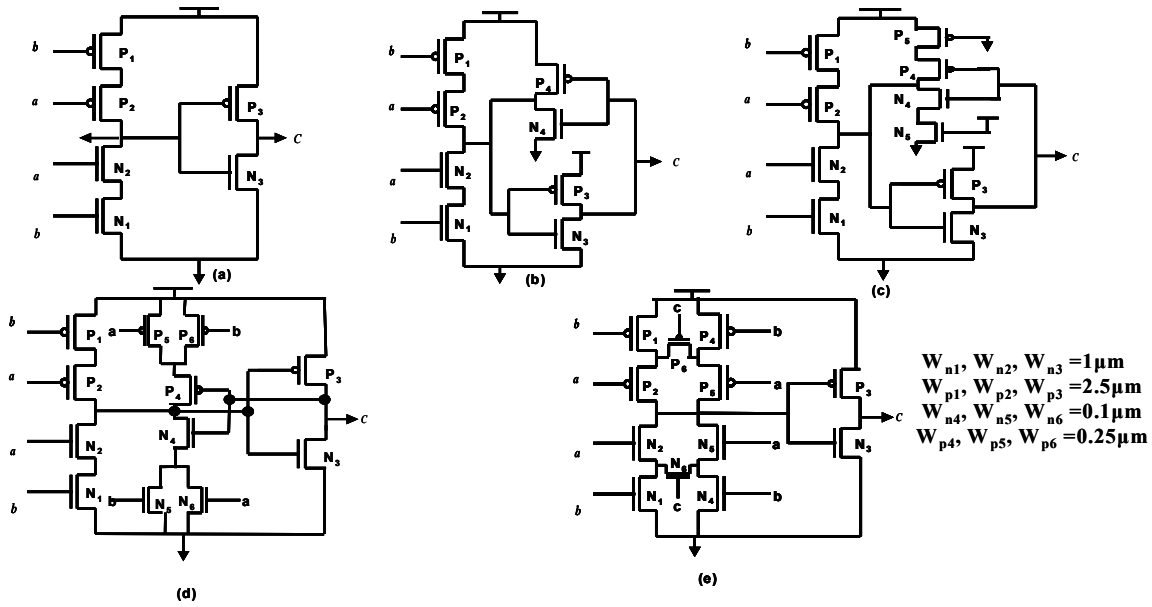
Figure 4.2: Implementation of the C-element: (a) dynamic, (b) weak feedback, (c) resistive weak feedback (d) standard and (e) symmetric [4, 5].

Figure 4.2(c) shows the resistive implementation of the C-element which can be considered as a modified version of the weak feedback implementation. The weak inverter has two more transistors. The additional transistors, namely N5 and P5 which act as resistors to reduce the current flow from the weak inverter during the switching, and hence improving the charging and discharging mechanism of the $\overline{C}$ node and improving circuit performance.

The standard implementation of the C-element, shown in Figure 4.2(d), was first presented by Sutherland [7] and is often used in high-speed micro pipelines [2]. In this implementation of the C-element, transistors N1, N2 and N3 are the main pull down devices, while P1, P2 and P3 are the main pull up devices. Meanwhile, the transistors N4, and P4 form the necessary feedback inverter to maintain the output logic value when both inputs do not match. Therefore these transistors have to be reduced in size as much as possible, to reduce their loading effect on the performance of the circuit. In addition, as can be seen from Figure 4.2(d), this implementation can be considered as a modified version of the weak feedback implementation, where the weak feedback inverter is controlled by the transistors N5, N6, P5 and P6, and only connect the weak inverter to the power supply and the ground when the inputs of the circuit do not match.

88

The symmetric implementation of the C-element, which is shown in Figure 4.2(e) is presented by Van Berkel [8]. In this implementation, the output is preserved through a feed-back conducting path of three transistors in the pull-up or pull-down networks. If the inputs are different and the output is low, the conducting path to retain the output node value consists of either P1, P5 and P6 or P4, P2 and P6. In a symmetrical way, when the output is high, the output is held to the previous value by the path which consists of either N1, N5 and N6 or N2, N4 and N6. If both inputs are similar in terms of logic value, the parallel PMOS transistors contribute to the rise time of the output, while the parallel NMOS transistors contribute to the fall time of the output.

Figure 4.3(a) shows the basic differential logic, and an inverter latch implementation of the C-element (DIL) [5]. It is composed of two pull-down networks consisting of NMOS devices, and an inverter latch formed by the transistors P_{FL}, N_{FL}, P_{FR} and N_{FR}.

When both inputs (a, b) are high, then $\overline{C}$ becomes low, and C is pulled up by the PMOS device of the right hand side inverter of the latch, P_{FR}. In a similar manner, when the inputs are low, then C becomes low and $\overline{C}$ is pulled up by the PMOS device of the left inverter of the latch, P_{FL}. However, when the inputs do not match ($a \neq b$), the inverter latch maintains the previous state of the output. One of the main drawbacks of this particular implementation is that it suffers from a large rising delay due to the weak PMOS devices of the inverter latch which are responsible for pulling up the C and the $\overline{C}$ nodes.

Another possible modification of this implementation is DILP as shown in Figure 4.3(b). The DILP implementation is intended to reduce the rising delay of the basic DIL implementation by adding two pull-up networks composed of PMOS devices.

$W_{NL1}, W_{NL2}, W_{NR1}, W_{NR2} = 1\mu m$
$W_{PL1}, W_{PL2}, W_{PR1}, W_{PR2} = 2.5\mu m$
$W_{NL1}, W_{NL2}, W_{NR1}, W_{NR2} = 1\mu m$
$W_{NFL}, W_{NFR} = 0.1\mu m$
$W_{PFL}, W_{PFR} = 0.25\mu m$
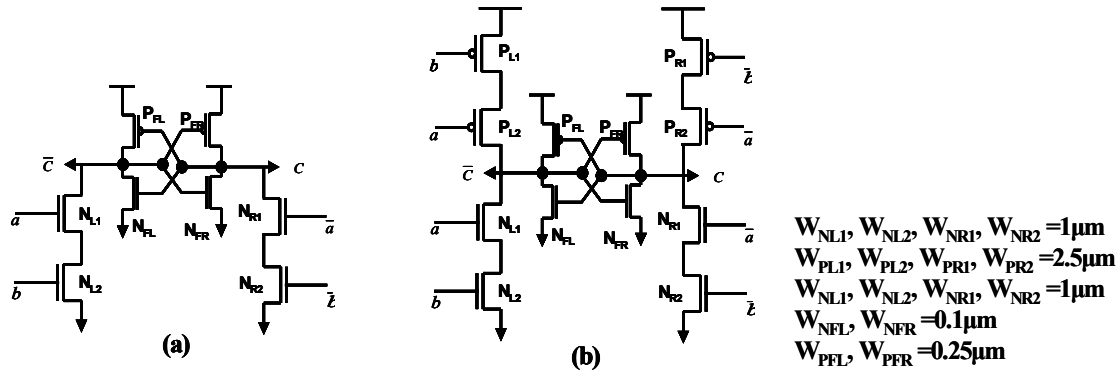
Figure 4.3: Differential logic with an inverter latch implementations of the C-element (a) DIL and (b) DILP.

All C-element circuits can be considered as modified versions of the basic implementation, which is the dynamic implementation as shown in Figure 4.2(a). The difference between them is only by way of preserving the output when the inputs have different logic values. The large variety of C–element implementations makes it a suitable test vehicle to study the effects of circuit topologies on achieving a high level of robustness against process variations, together with novel process technologies such as SOI technology.

In the next sections, the analysis of the robustness to the effects of process variation of the various C-element configurations (dynamic, static, and differential implementations) realised in different process technologies, is presented.

4.4 C-Element Variability Analysis

The analysis associated with studying the impact of process variability on circuit performance metrics, as demonstrated in Figure 4.4, is broadly divided into two parts with the aim of, firstly, studying the impact of variation of a number of process parameters of 60nm strained FDSOI, PDSOI and bulk silicon technologies on key compact model parameter such as (V_{th0}) of both NMOS and PMOS devices, where V_{th0} represents the threshold voltage for long channel devices at zero bias voltage. The reason for choosing V_{th0} for both NMOS and PMOS transistors as the design parameters among numerous other compact model parameters is that they show a strong statistical relationship with circuit performance metrics. In addition, they abstract complex physical

process variation phenomena into a single number, and are commonly used as a design variable, both in digital and analogue designs, and hence bridge the gap between design and manufacturing processes.

As previously motioned in Chapter 3, the variability analysis of the V_{th0} was undertaken using the Plackett-Burman (PB) screening technique in order to identify the most influential process parameters on the V_{th0} of both NMOS and PMOS devices implemented in FDSOI, PDSOI, and bulk silicon technologies, Thereafter the RSM technique is used to extract the spread of V_{th0} for the devices investigated. Secondly, studying the impact of process and operating condition variations on key performance metrics of a circuit such as propagation delay, dynamic power, and static power. The extracted variations in V_{th0} of the NMOS and PMOS devices for each technology, as tabulated in Table 3.7, together with the variations of the gate length (10% of the mean value) were considered.

The simulations were performed at supply voltages ranging from 0.81V to 0.99V for the transistors under different process conditions.
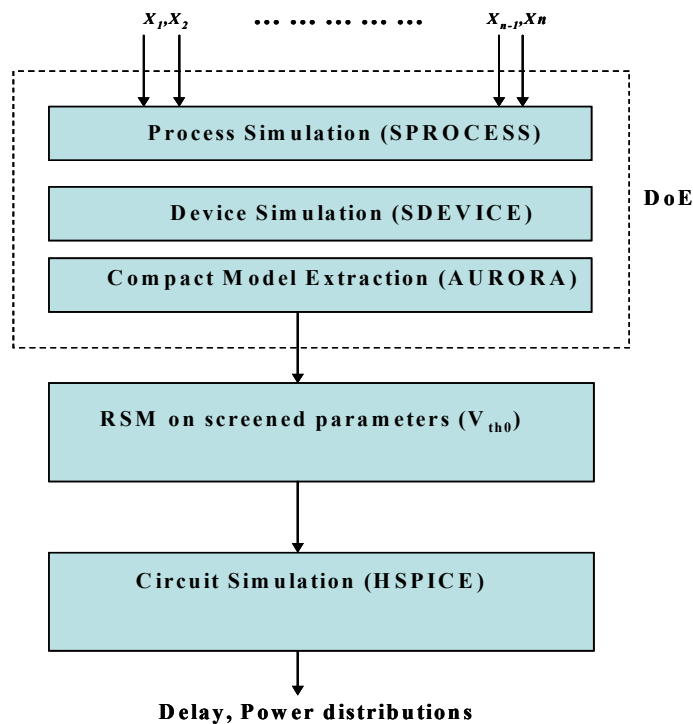
Figure 4.4: Flow chart of variability analysis.

Seven different C-element implementations were considered in this work, namely, the dynamic, standard, weak feedback, resistive weak feedback, symmetric, DIL, and DILP implementations. The width ratio of the PMOS to NMOS devices was chosen to be 2.5. Figures 4.5 to 4.7 show the RS plots obtained for the key performance metrics of the dynamic implementation of the C-element based on an FDSOI technology as it shows a very high robustness to the effect of process variations.

From the surface plots of the propagation delay (T_{pd}) for the dynamic C–element implementation under different process variation and operating conditions, it was found that the spread in T_{pd} is large due to the variation in the device threshold voltages compared to the gate length variation of the PMOS and NMOS devices.

In addition, it can be observed from Figure 4.5 that assigning higher power supply voltages (V_{dd}) reduces the effects caused by variations in the threshold voltage of the devices. It also can be seen that the use of low threshold voltage devices improves circuit performance and yield (i.e. reduces variation). This can be explained analytically by Equations 4.2 and 4.3 [9].

$$t_{pd} \propto \frac{V_{dd}}{(V_{dd} - V_{th})^{\alpha}} \qquad (4.2)$$

$$\frac{\partial t_{pd}}{\partial V_{th}} \propto \frac{\alpha V_{dd}}{(V_{dd} - V_{th})^{\alpha+1}} \qquad (4.3)$$

where V_{dd} is the supply voltage, V_{th} is the threshold voltage of the device α and is a parameter with a value between one and two, modelling the effects of velocity saturation. At higher V_{dd} values, the delay sensitivity to threshold voltage variation decreases, consequently, the delay distribution spread reduces.
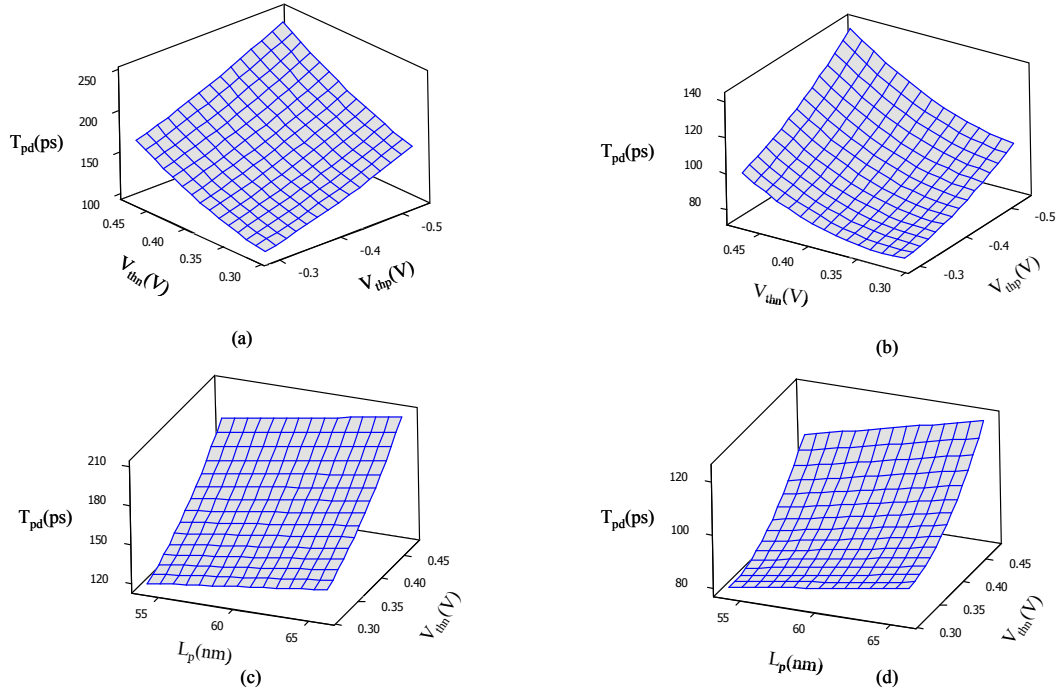
Figure 4.5: Surface plot of T_{pd} of dynamic C-element with respect to (a) the variation of threshold voltage of P and NMOS devices at Vdd=0.81V (b) the variation of threshold voltage of P and NMOS devices at Vdd=0.99V (c) the variation of threshold voltage of the NMOS device and PMOS gate length (L_p) at V_{dd}=0.81V (d) the variation of threshold voltage of the NMOS device and PMOS gate length (L_p) at V_{dd}=0.99V.

Figure 4.6 shows the response surface plots of the dynamic power consumption for the dynamic implementation of the C-element under different process variations (i.e. threshold voltage and gate length variations of both NMOS and PMOS devices) and operating conditions (V_{dd}).
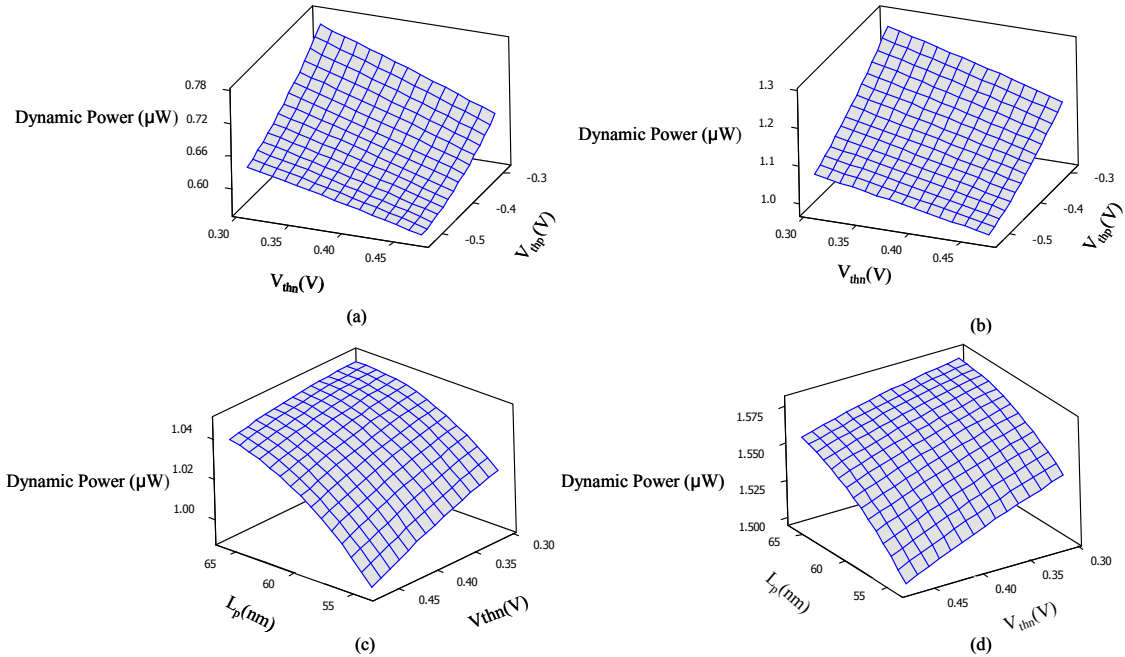
Figure 4.6: Surface plots of dynamic power of the dynamic C-element with respect to (a) the variation of threshold voltage of P and NMOS devices at Vdd=0.81V (b) the variation of threshold voltage of P and NMOS devices at Vdd=0.99V (c) the variation of threshold voltage of the NMOS device and PMOS gate length (L_p) at V_{dd}=0.81V (d) the variation of threshold voltage of the NMOS device and PMOS gate length (L_p) at V_{dd}=0.99V.

From the surface plot of the interaction between the threshold voltage of the PMOS and NMOS devices at different power supply voltages, as shown in Figure 4.6(a) and (b), it can be seen that the variation in the dynamic power consumption of the dynamic implementation of the C-element cab be largely minimised by applying a lower power supply voltage. However, this compromises the circuit performance in terms of propagation delay and its spread.

The surface plots of the static power, in Figures 4.7 show that the threshold voltage variation plays a significant role in static power variations. The use of higher threshold voltage reduces the variations, however, this negatively affects the circuit performance in terms of propagation delay.

Consequently, trade-offs between variations in delay and static and dynamic power consumptions can be made. Thus, the optimization of the threshold voltage and the power supply is essential depending upon the application of the circuit, to ensure a high performance and yield.

94
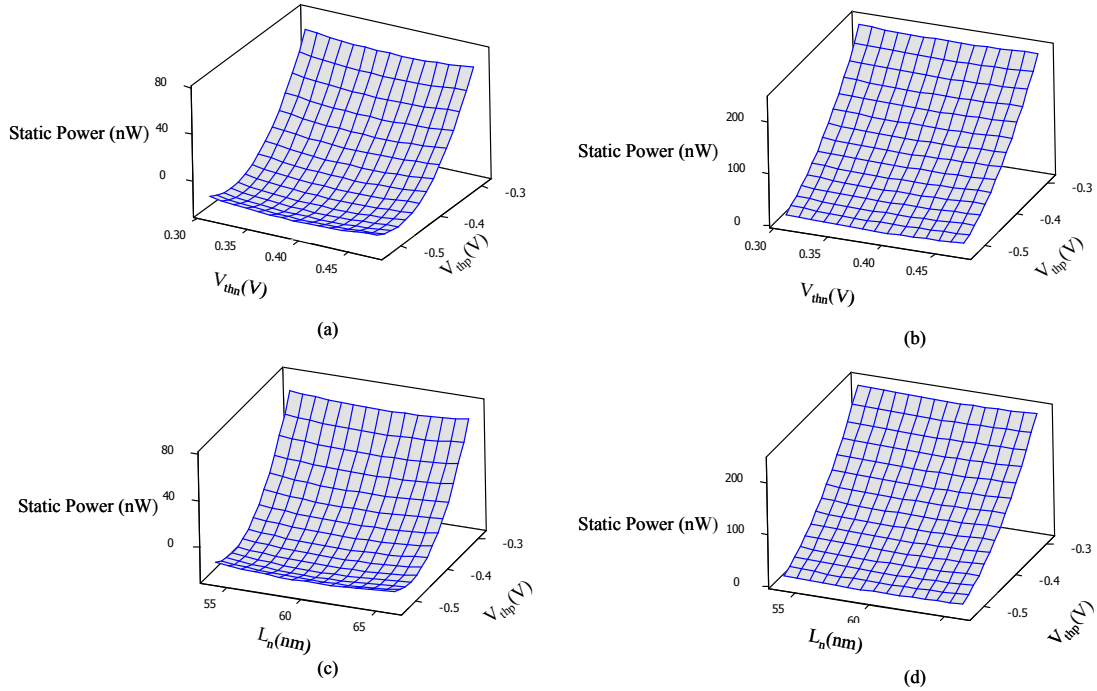
Figure 4.7: Surface plots of static power of the dynamic C-element with respect to (a) the variation of threshold voltage of P and NMOS devices at Vdd=0.81V (b) the variation of threshold voltage of P and NMOS devices at Vdd=0.99V(c) the variation of threshold voltage of the PMOS device and NMOS gate length at V_{dd}=0.81V (d) the variation of threshold voltage of the PMOS device and NMOS gate length at V_{dd}=0.99V.

All of other C-element implementations investigated were studied using the same RSM methodology.

Figures 4.8 to 4.10 show the histogram plots for T_{pd} and the dynamic power and static power of dynamic and static configurations respectively of the C-element based on FDSOI technology.
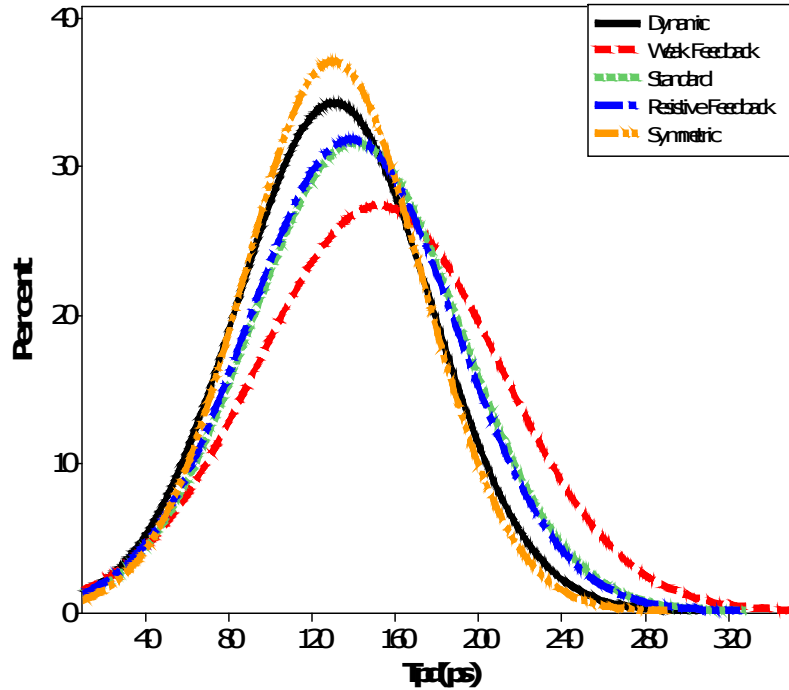
Figure 4.8: Histogram plots of propagation delay (T_{pd}) for dynamic and static configurations of the C-element based on bulk silicon technology under process and operating condition variations.

From Figure 4.8, it is clear that the symmetric and the dynamic implementations of the C-element are more robust in terms of delay variations.

Regarding the weak feedback implementation of C-element, the presence of a weak feedback inverter (the keeper) in the circuit degrades circuit tolerance to variation by adding an extra source of variation through the feedback path into the final output.

However, as in the standard implementation, the feedback is controlled by the transistors P5, P6, N5 and N6 as shown in Figure 4.2(d). Therefore, the weak feedback inverter is only connected to the V_{dd} and the ground when the inputs are different, in order to help maintain the previous logic state.

Another way to boost the weak feedback C-element is by adding transistors acting like resistors, N5 and P5, to the weak feedback inverter as shown in Figure 4.2(c), to reduce the impact of V_{dd} as a source of variation on the weak feedback path.

In terms of circuit resilience to dynamic power variations, it can be seen from Figure 4.9 that the dynamic implementation has the lowest mean and standard deviation values

among the other static C-element implementations. In a similar manner as with delay variation analysis, the results show that the weak feedback inverter is adding more variation to the final power variation, and the modification of this path can reduce this effect, either by controlling the weak inverter connection to V_{dd} and ground by the input signals as in the standard implementation, or by reducing the impact of V_{dd} as a source of variation on the feedback loop as in the resistive weak feedback implementation.

However, among the other circuits, the symmetrical implementation has the largest mean and standard deviation values as shown in Figure 4.9, this can be explained by the fact that the symmetrical implementation has larger number of transistors in comparison to other C-element implementations (i.e. higher input capacitance), therefore the dynamic power consumption is expected to be larger than that of their C-element counterparts.
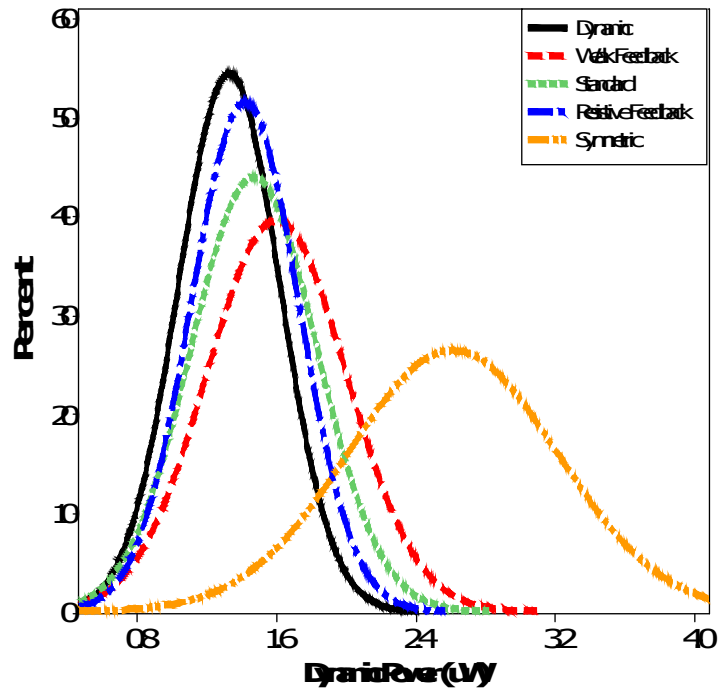
Figure 4.9: Histogram plots of dynamic power for the dynamic and static configurations of the C-element based on bulk silicon technology under process and operating conditions variations.
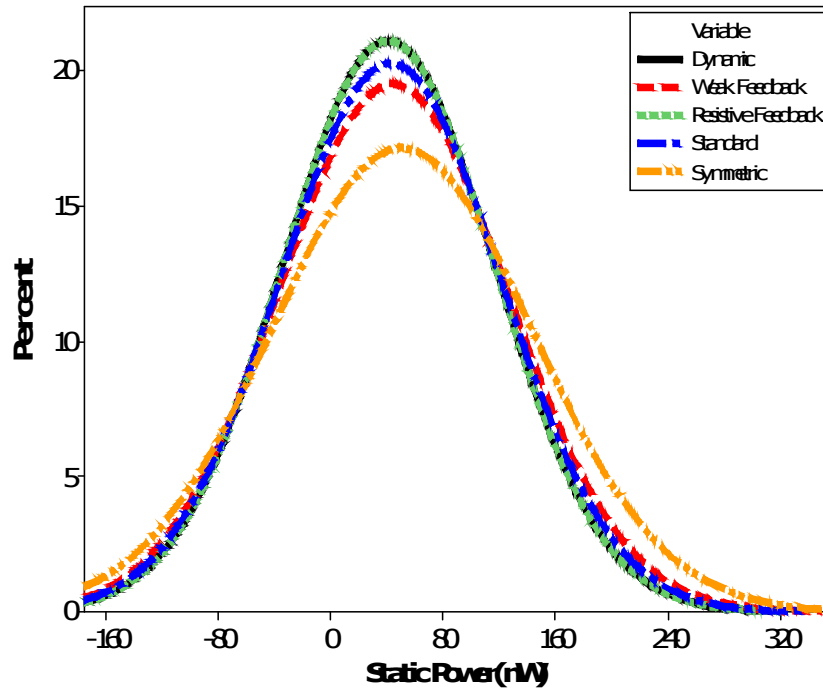
Figure 4.10: Histogram plots of static power consumption for the dynamic and static configurations of the C-element based on bulk silicon technology under process and operating conditions variations.

In terms of static power consumption under process variation, all of the dynamic and static circuits have almost the same standard deviation value as shown in Figure 4.10. This could be explained by the exponential dependence of the static power consumption on the threshold voltage of the devices.

Considering the differential implementation of the C-element realised in bulk silicon technology, it can be seen from the appropriate columns in Tables 4.1-4.3 that the mean and standard deviation values of the performance metrics are the lowest when compared to the dynamic and static implementations of the C-element.

In similar manner, the variability analysis is applied to the C-element circuits based on FDSOI and PDSOI technologies. It is observed that all C-element circuits realised on SOI technology are more resilient to process variation, with higher performance metrics. At the same time, the bulk silicon circuits are the worst in terms of performance metrics, and are affected more by the effect of variations in threshold voltage, gate length and power supply voltage.

In addition, as can be seen from Tables 4.1 and 4.3, FDSOI circuits are more robust in terms of propagation delay variations, and they show the lowest propagation delay compared to PDSOI and bulk silicon circuits.

Moreover, the circuits based on PDSOI are more resilient to dynamic and static power variations in comparison to FDSOI and silicon counterparts.

The results suggest that silicon on insulator technology (SOI) is a potential candidate to improve a performance and yield of a circuit.

C-Element Circuit	FDSOI		PDSOI		Silicon	
	μ	σ	μ	σ	μ	σ
Dynamic	43.36	8.66	52.81	12.39	130.43	46.59
Weak Feedback	49.04	8.83	59.62	13.9	151.85	58.39
Resistive weak Feedback	45.3	8.88	55.63	13.03	140.62	50.53
Standard	45.19	8.55	55.63	12.57	138.64	50.21
Symmetric	41.03	8.44	51.3	12.8	129.89	43.1
DILP	23.473	5.108	29.64	7.35	67.1	16.73
DIL	29.029	5.398	38.33	11.29	110.24	45.34

Table 4.1: Propagation delay (T_{pd}) (ps) for different C-element implementations under process and operating conditions variation.

C-Element Circuit	FDSOI		PDSOI		Silicon	
	μ	σ	μ	σ	μ	σ
Dynamic	0.506	0.133	0.356	0.091	1.285	0.257
Weak Feedback	0.603	0.161	0.411	0.109	1.554	0.349
Resistive	0.545	0.145	0.386	0.102	1.376	0.272
Standard	0.546	0.149	0.376	0.099	1.425	0.315
Symmetric	1.078	0.274	0.799	0.194	2.56	0.549
DILP	0.479	0.122	0.334	0.093	1.226	0.264
DIL	0.269	0.081	0.209	0.0532	0.873	0.181

Table 4.2: Dynamic power consumption (μW) for different C-element implementations under process and operating conditions variation.

C-Element Circuit	FDSOI		PDSOI		Silicon	
	μ	σ	μ	σ	μ	σ
Dynamic	49.34	55.52	7.27	7.55	41	75.6
Weak Feedback	45.34	55.72	8.09	8.63	44.3	81.8
Resistive	42.99	52.83	7.52	8.04	41	75.6
Standard	43.63	53.52	7.74	8.23	42.6	78.7
Symmetric	50.25	58.22	11.21	11.64	50.6	93.1
DILP	31.78	38.3	5.589	5.738	27.3	49.33
DIL	29.3	37.7	2.803	3.672	6.86	12.5

Table 4.3 : Static power consumption (nW) for different C-element implementations under process and operating conditions variation.

4.5 Transistor Current Variation Analysis and Optimization

For further analysis on circuit behaviour under process variation, the impact of process variations on channel current that pass through each transistor in a circuit was studied. The investigation was carried out by measuring the maximum current in each transistor under process variation during a circuit's operation. Using this method, it was possible to first, explain why the circuit topology plays a significant role in performance variation, second, identify which transistors make the most significant contribution to the final performance variation, and finally target these transistors in order to minimise their impact on performance variability, by resizing their widths.

In order to compare the performance of the circuits investigated under process and operating condition variations, the N1, N2, N3 and P1, P2, P3 transistors that form the basic dynamic implementation of the C-element were considered in this study.

Figure 4.11 shows a histogram plot of the maximum current in the transistors that make up a dynamic implementation of the C-element.
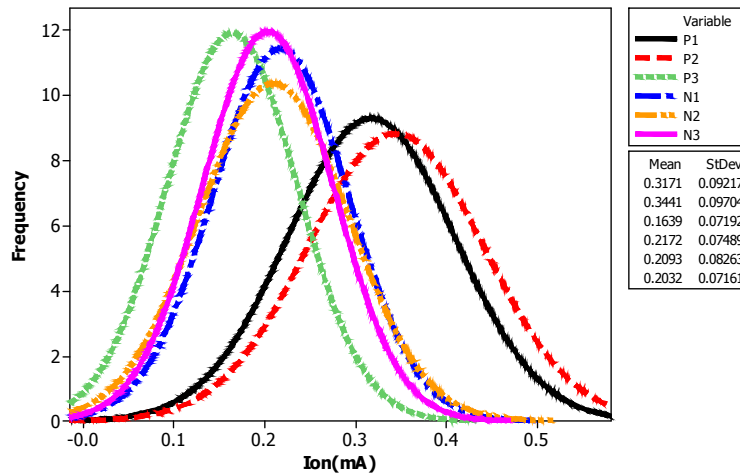
Figure 4.11: Histogram plots of maximum current of transistors composing the dynamic C-element implementation under process and operating condition variations.

By comparing the standard deviation over the mean values (σ/μ) of the current in the input transistors of a C-element (i.e.N1, N2, P1, P2), tabulated in Table 4.4, it can be observed that the transistors P1,P2 in the weak feedback C-element implementation have a larger (σ/μ) of the channel current values compared to their counterparts in the dynamic implementation of C-element. This explains the higher variation in both delay and power consumption compared to the dynamic implementation of the C-element. In other words, the presence of the weak feedback inverter in these implementations has degraded the circuit performance and its resilience towards variation, compared to the dynamic C-element implementation.

It was also observed that controlling the weak feedback inverter, as in the standard and resistive weak feedback implementations, reduces the standard deviation of the mean values (σ/μ) of the current in the input transistors of a C-element (i.e.N1, N2, P1, P2) as shown in Table 4.4, in comparison to the weak feedback implementation of the C-element. Therefore, this improves its robustness in terms of process variation.

From Table 4.4, it can be seen that the transistors N2, N1 have larger (σ/μ) values compared to transistors P1, P2 which means that they contribute more to the final delay variations.

Circuit /Transistor	P1	P2	P3	N1	N2	N3
Dynamic	0.291	0.282	0.439	0.345	0.395	0.352
Standard	0.333	0.32	0.378	0.333	0.377	0.363
Resistive Weak Feedback	0.306	0.293	0.443	0.331	0.378	0.361
Weak Feedback	0.339	0.319	0.382	0.342	0.387	0.37
Symmetric	0.263	0.224	0.387	0.355	0.388	0.349

Table 4.4: The standard deviation over the mean values (σ/μ) of the current of common transistors composing the dynamic and static implementation of the C-element under process and operating condition variations.

Table 4.5 shows the mean (μ), standard deviation (σ), and standard deviation over mean (σ/μ) values of the maximum current of the transistors in the weak feedback implementation of the C-element. It can be seen that the transistors N1 and N2 have the lowest current mean values compared to the input transistors (P1, P2) of the C-element, i.e. the highest σ/μ values. This means that the N1 and N2 transistors significantly contribute to the large variation in the propagation delay of the weak feedback C-element.

Transistor	μ(mA)	σ(mA)	σ/μ
P1	0.33	0.112	0.34
P2	0.35	0.1124	0.32
P3	0.155	0.059	0.38
P4	0.063	0.017	0.27
N1	0.21	0.071	0.34
N2	0.19	0.077	0.39
N3	0.197	0.073	0.37
N4	0.042	0.011	0.25

Table 4.5: The mean, the standard deviation, and standard deviation over mean values of the current of transistors composing the weak feedback implementation of C-element under process and operating condition variations.

Having identified the most dominant devices in terms of the circuit performance metrics, it is now possible to reduce the variations in performance by varying their widths.

The weak feedback and the standard and weak feed back implementations of the C-element are used as test vehicles to prove the concept of variability analysis based on the

channel current of transistors and on the ability to improve the robustness of the circuit with regard to process variations.

As observed from the above results in Tables 4.4 and 4.5, the transistors N1 and N2 in the weak feedback implementation are the dominant devices that largely contribute to performance variation, in particular, the propagation delay (T_{pd}) of the circuit, since they have lower current values. One way of increasing the mean value of the current in N1 and N2 is by doubling their original widths.

It can be seen from Figures 4.12-4.14, that when doubling the width of the N1, N2 transistors, the propagation delay of the weak feedback implementation of C-element is reduced in terms of the mean and standard deviation values by 17.5% and 23.6%, respectively, however the dynamic power consumption is increased in terms of the mean and standard deviation by 7.3% and 6.5%, respectively. The static power consumption remained almost the same.

Another way to improve the circuit performance and its robustness to the effects of process variation is by reducing the impact of loading on the input transistors of a C-element. This can be done by reducing the width of the transistors that compose the weak feedback inverter and the output inverter (P3, N3).

Figures 4.12-4.14 show that the improvements in circuit performance and its robustness to variability that can be obtained by resizing the width of the output inverter (P3, N3) to half of their original widths. The mean and standard deviation values of the propagation delay has reduced by 13% and 18%, while mean and standard deviation values of the dynamic power consumption has reduced by 25.2 % and 27.5%, and the values of the static power has also been reduced by 34.6% and 34.9%, respectively.

Similarly, reducing the width of the transistors which compose the weak feedback inverter, can slightly improve the mean and standard deviation values of the propagation delay and the dynamic power consumption by 8% and 12% and 9%, 9.3%, respectively, while the static power consumption remains almost unchanged.

Finally, by increasing the width of transistors (N1, N2), together with reducing the size of the transistors that compose both the output inverter and weak feedback inverter (i.e. reduce the loading effect), it is possible to improve the circuit performance and its robustness to the effects of process variations in terms of propagation delay, dynamic

power and static power consumptions. From Figures 4.12-4.14, it can be seen that the mean and standard deviation values of the propagation delay has reduced by 30.2% and 40% respectively, while mean and standard deviation values of the dynamic power consumption has reduced by 23.6% and 30% respectively and the values for the static power consumption has been also reduced by 38.6% and 39%.
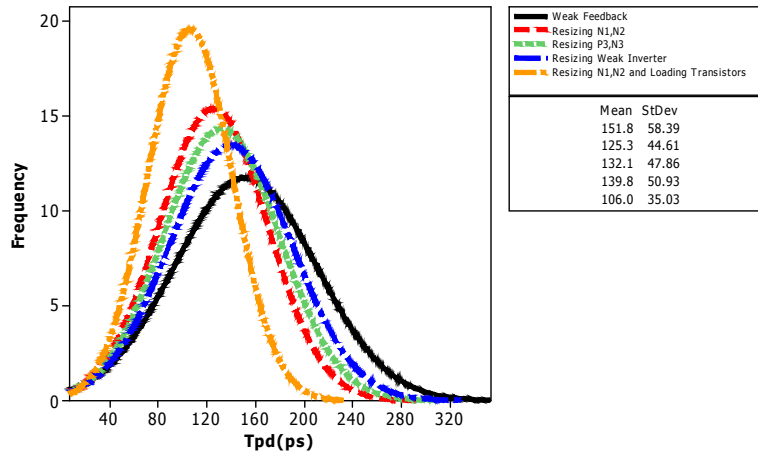
Figure 4.12: Optimisation of the propagation delay (T_{pd}) the weak feedback implementation of the C-element under process and operating conditions.
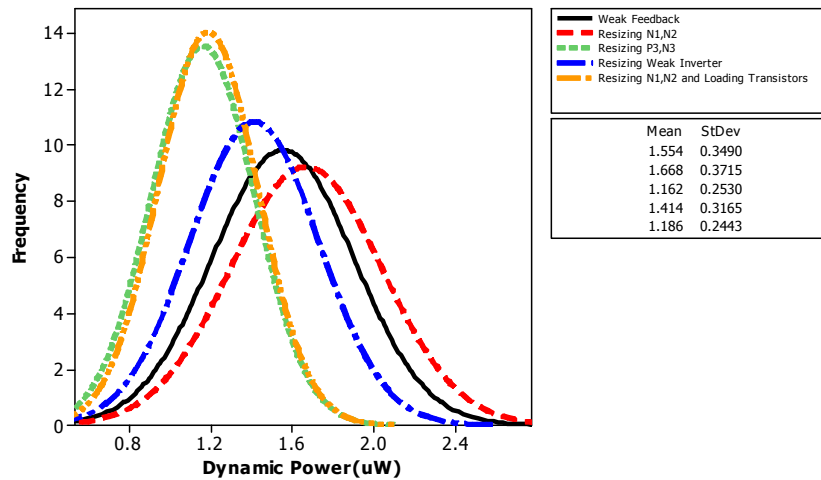
Figure 4.13: Optimisation of the dynamic power consumption of the weak feedback implementation of the C-element implementation under process and operating condition variations.
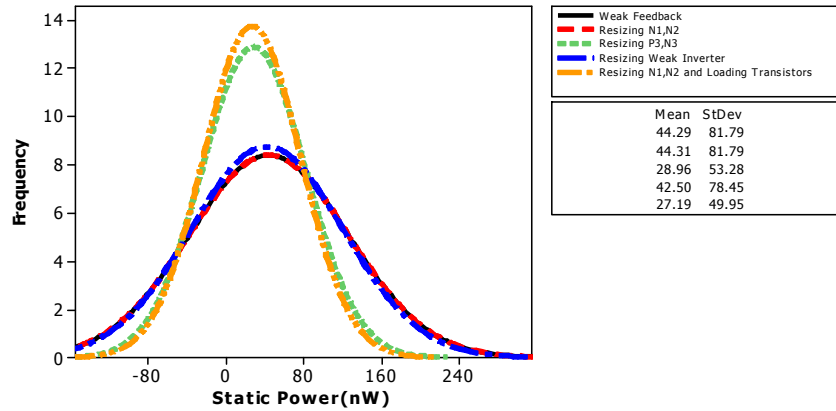
Figure 4.14: Optimisation of the static power consumption of the weak feedback implementation of the C-element implementation under process and operating condition variations.

The optimisation of the propagation delay (T_{pd}) of the standard implementation of C–element was also undertaken using the same approach. Table 5 shows the improvements that can be achieved in terms of the performance of the circuit and its robustness to the effect of process variations. It can be seen that by doubling the width of transistors N1, N2, together with the reduction of the size of the loading transistors, it was possible to reduce the mean values of the propagation delay, dynamic power and static power by almost 25%, 18.6 % and 38%, respectively. While the standard deviation values are also reduced by 30.7%, 31% and 38.3%, respectively.

Circuit Modification	Propagation Delay(ps)		Dynamic Power(µW)		Static Power(nW)	
	µ	σ	µ	σ	µ	σ
Standard implementation	138.64	50.21	1.4252	0.3147	42.6	78.7
Resizing N1,N2	120.52	41.93	1.5876	0.3378	42.7	78.7
Resizing P3,N3	123.37	42.71	1.0608	0.2201	27.3	50.17
Resizing weak inverter transistors	136.50	48.81	1.3685	0.2973	41.7	77
Resizing N1,N2 and loading transistors	104.55	34.79	1.1595	0.2266	26.42	48.5

Table 4.6: The mean (µ) and standard deviation (σ) values of the propagation delay, dynamic power and static power consumptions for standard implementation of C-element.

4.6 Process - Circuit Variation Analysis

In this section, the design of experiment and response surface methodologies are extended to investigate the direct impact of manufacturing process steps on the performance of circuits based on the technologies investigated. The aim is to determine and identify the critical process parameters that need to be controlled, and to find ways to reduce variability in terms of the sources, hence improving the performance and the yield of the circuit.

To study and analyse the impact of process variability on the key performance metrics of a circuit, such as propagation delay, dynamic power, and static power, eighteen process parameters, - shown in Table 3.4 - and the gate length of the device, referred as (x_{19}), were identified as potential sources of uncontrollable variations during different manufacturing process steps for the devices based on FDSOI, PDSOI and bulk silicon technologies. The process parameters chosen were the same for both the NMOS and the PMOS devices.

All the process parameters were varied by ±10% of their mean values (i.e. the same standard deviations). However the process temperatures during the different manufacturing steps were set at ±10° C from the nominal. This is mainly because the temperature values are very high and, in practice, would not drift in the range of ±10%.

 Plackett-Burman (PB) screening [10], due to its computational efficiency, was used to screen and identify the most dominant input parameters from an initial set of nineteen process parameters. For the PB design of k=19 parameters, the number of simulation runs required, N, was 20.

Twenty process and device simulations, compact model extractions and circuit simulations were undertaken as part of the PB screening process, and key performance metrics of the circuit such as the propagation delay, dynamic power, and static power consumptions were considered as output responses in order to evaluate the screening experiment.

The statistical analysis of the responses was subsequently performed using Pareto analysis. This analysis ranks the relative magnitude of the influence of all the main input

parameters on the output responses, and arranges them in order of the decreasing absolute value of the effect. Hence the statistical significance of each input can then be identified. The Pareto analysis of the performance metrics for the standard implementation of C-element based on FDSOI are shown in Figures 4.15 to 4.17 respectively.
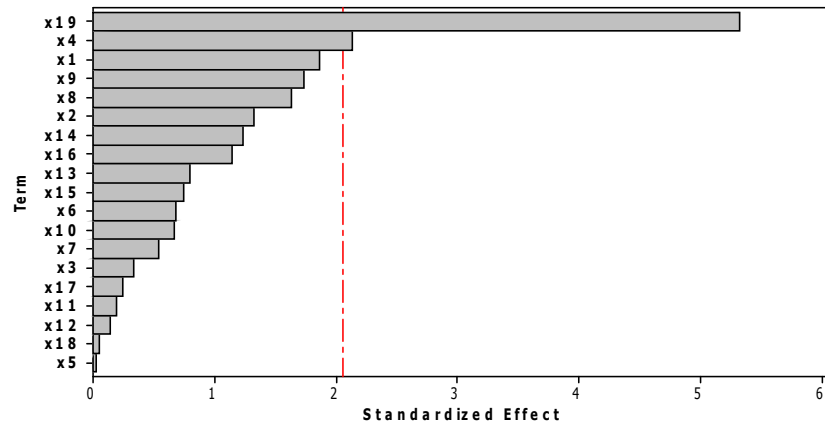
Figure 4.15: Pareto plot of the most significant process parameters that impact on the propagation delay (T_{pd}) of the standard C-element based on FDSOI technology.

Figure 4.16 : Pareto plot of the most significant process parameters that impact on the dynamic power consumption of the standard C-element based on FDSOI technology.

Figure 4.17 : Pareto plot of the most significant process parameters that impact on the static power consumption of the standard C-element based on FDSOI technology.

From Figures 4.15 to 4.17, it can be seen that the variations in silicon layer thickness (x_4) largely contributed to the variations in the performance metrics of the standard implementation of the C-element.

In addition, the gate length variation (x_{19}) mainly impacts on both the propagation delay and the dynamic power consumption of the standard implementation of the C-element based on FDSOI technology.

Moreover, the static and dynamic power consumptions of the standard implementation of the C-element is affected by other common sources such as the threshold voltage adjustment implantation energy (x_8), the halo implantation dose (x_9) and source and drain implantation dose (x_{12}).

Having identified the most significant process parameters for the circuit responses in the screening steps, the response surface analysis method for circuit performance metrics was performed to obtain useful information about the sensitivity of each metric with respect to the process parameters under consideration.

The RS plots obtained for the performance metrics for the standard implementation of the C-element based on FDSOI technology are shown in Figures 4.18 to 4.20.
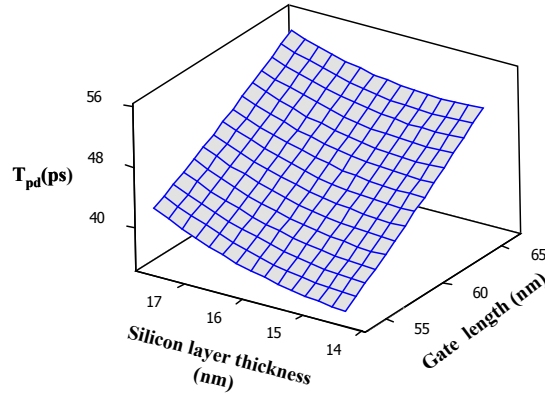
Figure 4.18: Response surface for the propagation delay of the standard C-element based on FDSOI technology with respect to gate-length and silicon layer thickness variations.

(a)

(b)
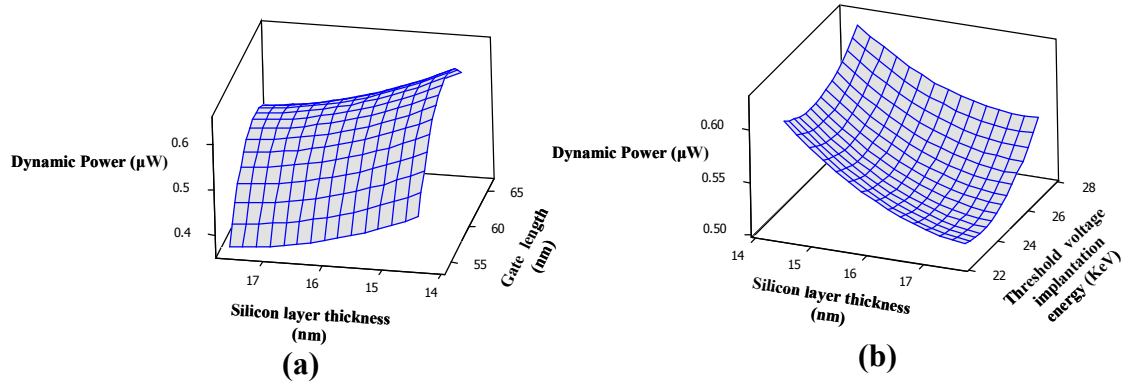
Figure 4.19: Response surface for the dynamic power consumption of the standard C-element based on FDSOI technology with respect to (a) gate length and silicon layer thickness (b) silicon layer thickness and threshold voltage adjustment implantation energy.

(a)

(b)

Figure 4.20: Response surface for the static power consumption of the standard C-element based on FDSOI technology with respect to (a) threshold voltage adjustment implantation energy and halo implantation dose (b) silicon layer thickness and source and drain implantation dose.

109

From Figures 4.18 and 4.19(a), it can be seen that the gate length variation largely impacts on the propagation delay and on the dynamic power consumption of the circuit in comparison with the variation in silicon layer thickness.

The variations in dynamic power consumption can be reduced when using longer channel devices. However, this adversely impacts on both the performance of the circuit in terms of propagation delay, and dynamic power consumption.

The static power consumption can be significantly minimised by adopting lower values of threshold voltage adjustment implantation energy, as shown in Figure 4.20 (a). This helps to reduce the sensitivity of the static power of the circuit to the halo implantation dose variations. However this might impact on the propagation delay of the circuit as can be seen in Figure 4.21, where the propagation delay becomes more sensitive to the halo implantation dose at lower values of threshold voltage adjustment implantation energy.
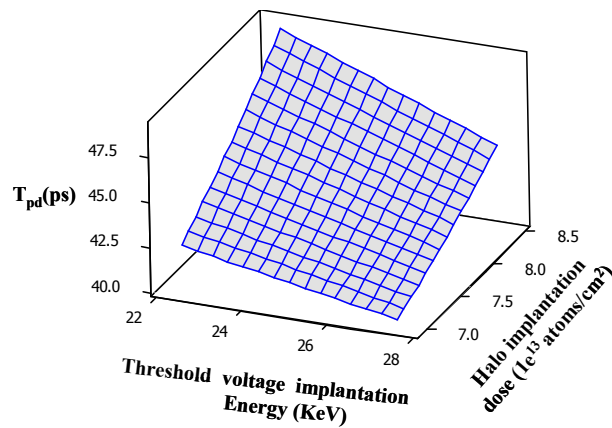
Figure 4.21: Response surface for propagation delay of the standard C-element based on FDSOI technology with respect to threshold voltage adjustment implantation and halo implantation dose.

Figure 4.20 (a) shows the impact of the silicon layer thickness and the source and drain implantation dose on the static power of the circuit investigated. From this figure it can be seen that the thicker the silicon layer, the lower the static power consumption. This can be explained by the large impact of silicon layer thickness on the threshold voltage of the FDSOI devices as demonstrated in Chapter 3 which, in turn, influences the static power of the circuit.

The Pareto analysis of the performance metrics for the standard implementation of the C-element based on PDSOI technology are shown in Figures 4.22 to 4.24.

Figure 4.22: Pareto plot of the most significant process parameters that impact on the propagation delay (T_{pd}) of the standard C-element based on PDSOI technology.
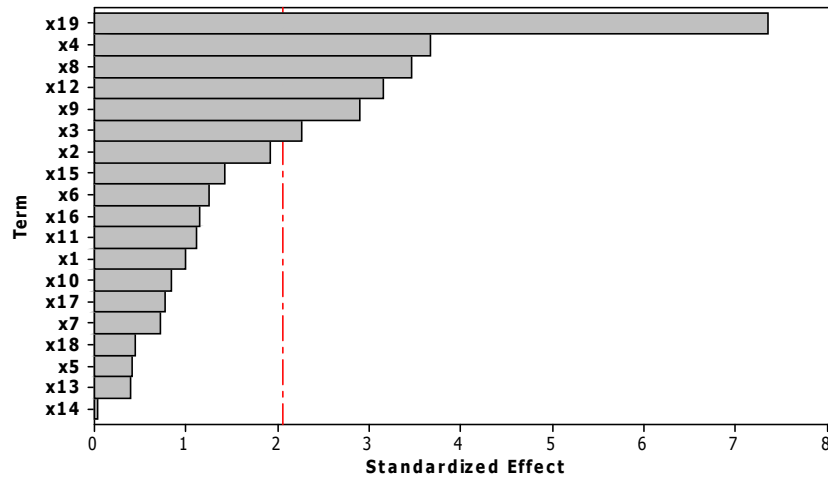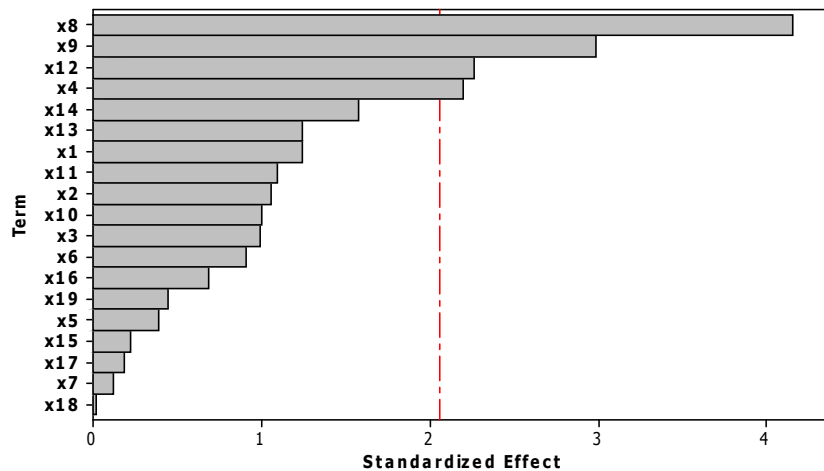

Figure 4.23: Pareto plot of the most significant process parameters that impact on the dynamic power consumption of the standard C-element based on PDSOI technology.

Figure 4.24: Pareto plot of the most significant process parameters that impact on the static power consumption of the standard C-element based on PDSOI technology.

From the Pareto plots for the key performance metrics of the standard C-element based on PDSOI technology, it can be seen that the gate length (x_{19}) and the threshold voltage adjustment implantation energy (x_8) largely contribute to the variations in propagation delay and dynamic power consumption of the standard implementation. At the same time, the static dynamic power consumption of the circuit investigated is largely affected by the threshold voltage adjustment implantation energy (x_8) and the gate oxide thickness (x_1).

Figures 4.25 to 4.27 show the RS plots obtained with regard to the propagation delay and the dynamic power and static power consumptions of the standard C-element realised by PDSOI technology, respectively.

Figure 4.25 : Response surface for propagation delay of the standard C-element based on PDSOI technology with respect to gate length and threshold voltage adjustment implantation energy variations.

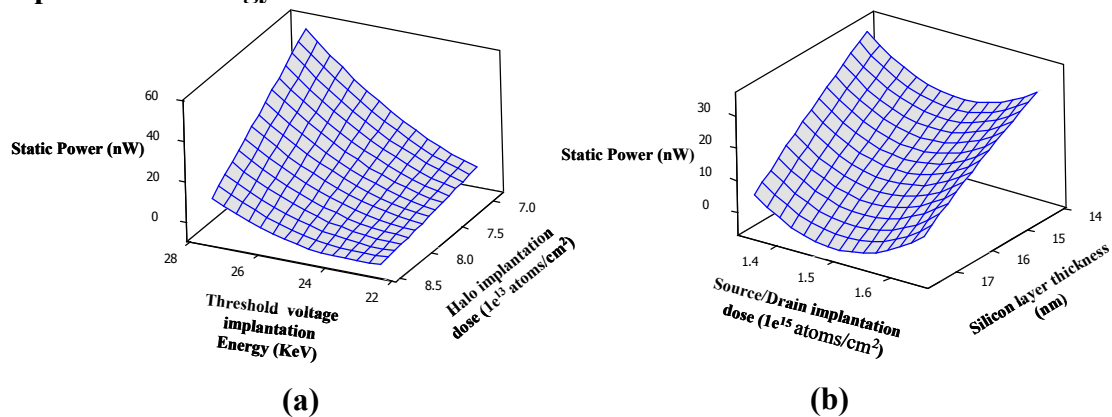**Figure 4.26: Response surface for dynamic power consumption of the standard C-element based on PDSOI technology with respect to gate length and threshold voltage adjustment implantation energy variations.**

Figure 4.27: Response surface for static power consumption of the standard C-element based on PDSOI technology with respect to gate oxide thickness and threshold voltage adjustment implantation energy variations.

From Figures 4.25 and 4.26, it can be seen that the gate length variation (x_{19}) largely impacts on the propagation delay and the dynamic power consumption of the standard C-element based on PDSOI technology.

While the static power consumption is significantly influenced by threshold voltage adjustment implantation energy (x_8) and gate oxide thickness (x_1), as shown in Figure 4.27, it can be seen that the static power consumption of the circuit investigated is reduced when using low threshold voltage adjustment implantation energy.

The sensitivity of the static power of the circuit to the threshold voltage adjustment implantation energy can be reduced by increasing the gate oxide thickness. However this will compromise the circuit performance in terms of delay and dynamic power consumption.

In a similar manner, the analysis was performed for the standard implementation of the C-element. Figures 4.28 to 4.30 show the Pareto analysis of the performance metrics for the standard implementation of the C-element based on bulk silicon technology.
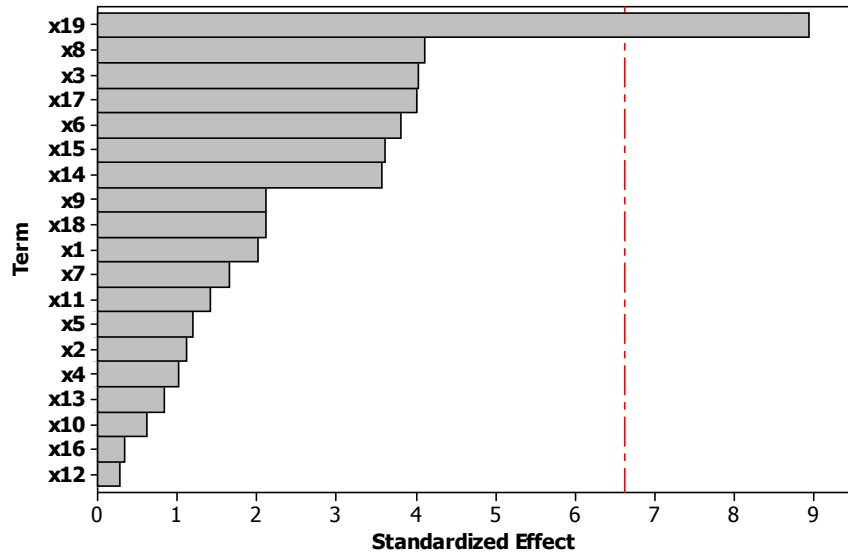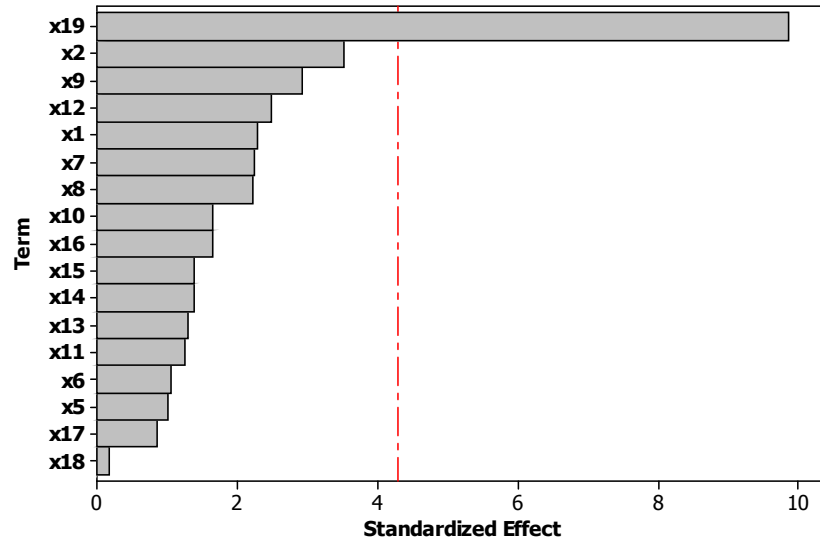

Figure 4.28: Pareto plot of the most significant process parameters that impact on the propagation delay (T_{pd}) of the standard C-element based on bulk silicon technology.
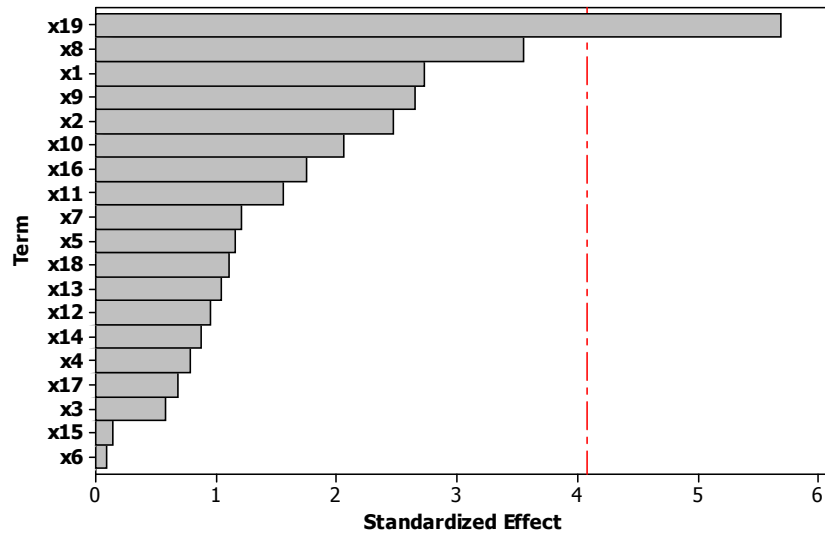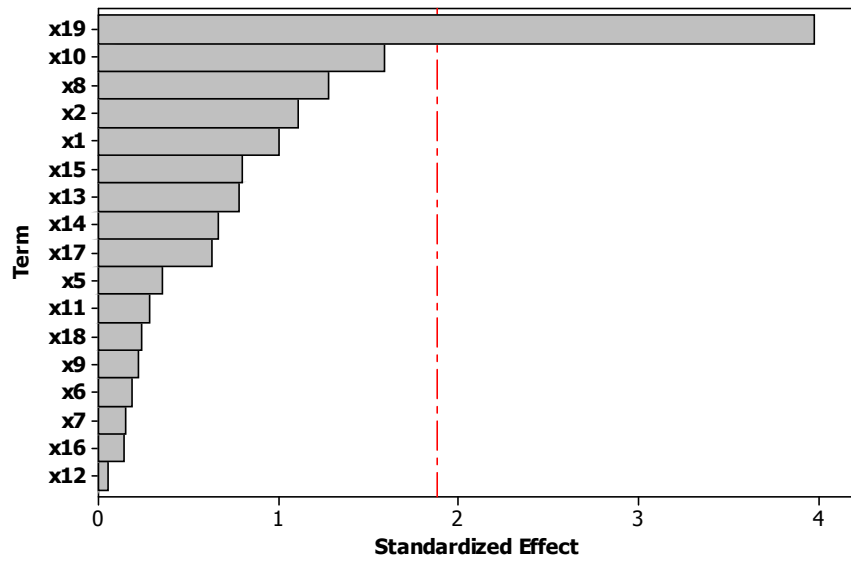
Figure 4.29: Pareto plot of the most significant process parameters that impact on the dynamic power consumption of the standard C-element based on bulk silicon technology.
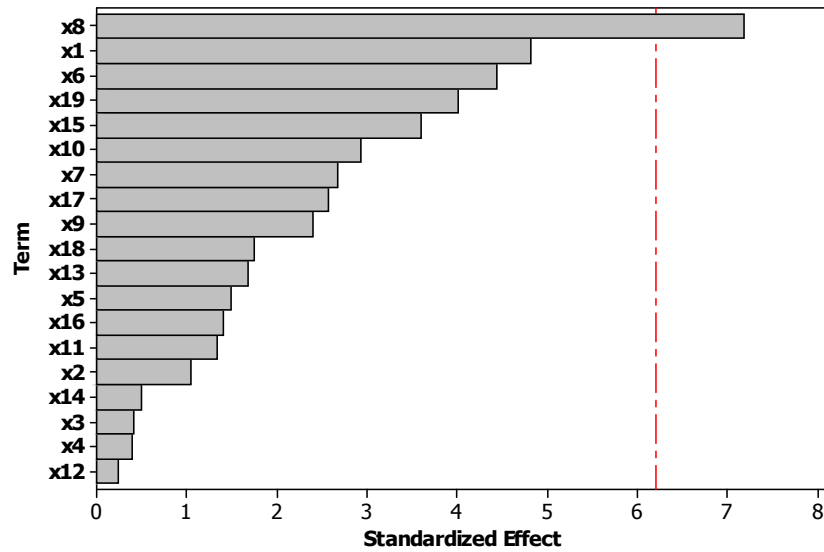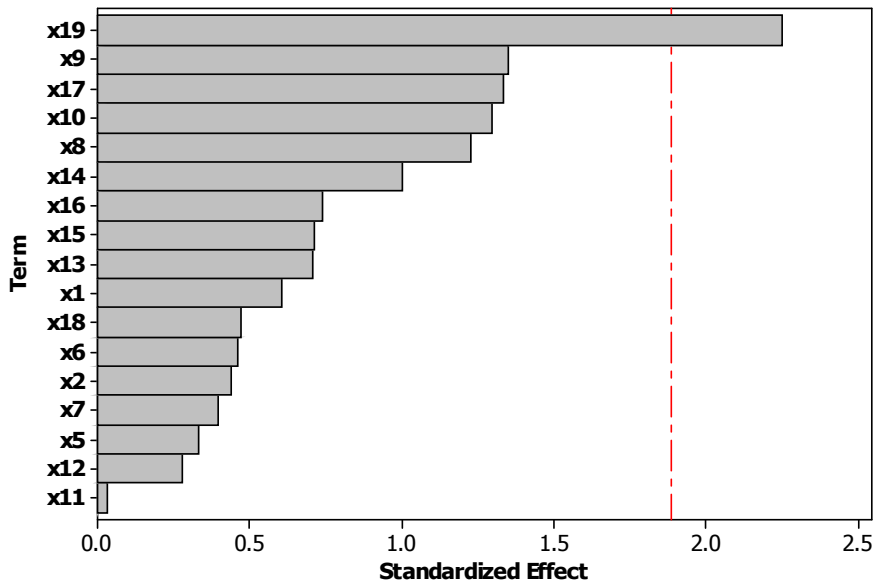

Figure 4.30: Pareto plot of the most significant process parameters that impact on the static power consumption of the standard C-element based on bulk silicon technology.
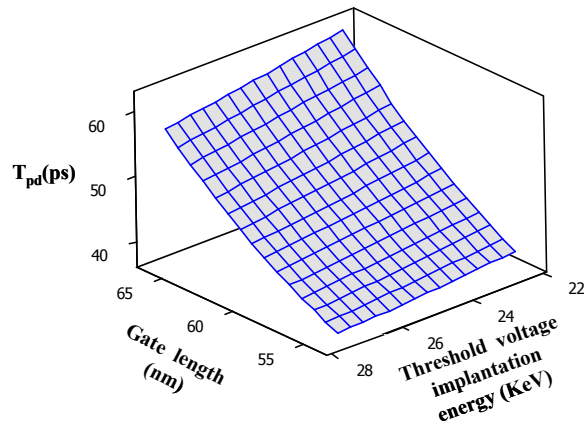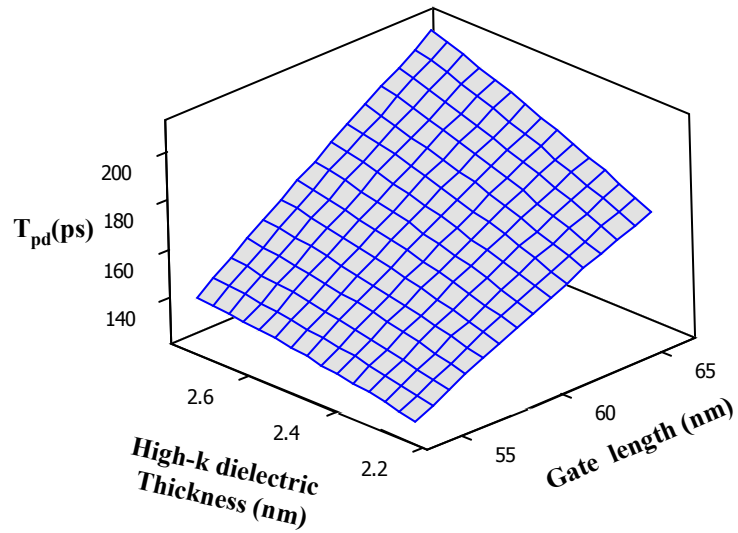
Figure 4.31: Response surface for propagation delay of the standard C-element based on bulk silicon technology with respect to gate length and high-k dielectric thickness.
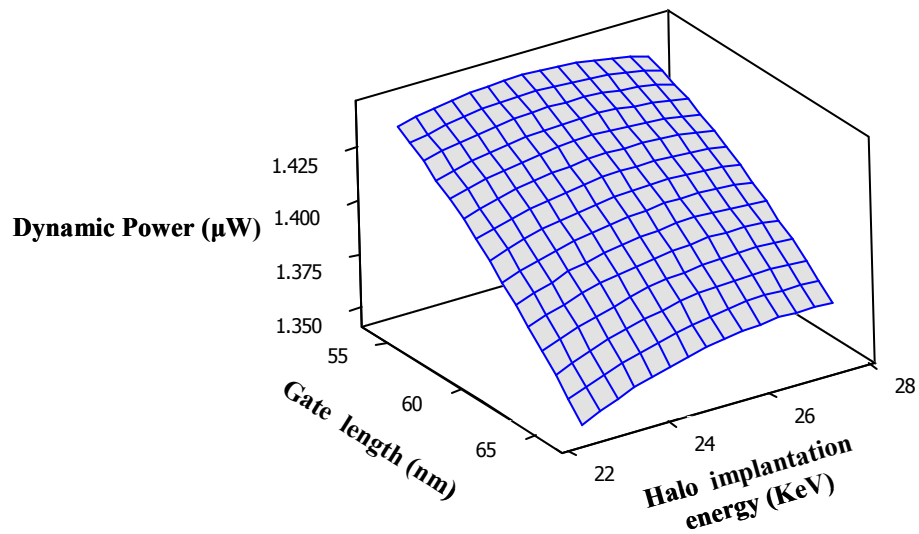
Figure 4.32: Response surface for dynamic power consumption of the standard C-element based on bulk silicon technology with respect to gate length and halo implantation energy.
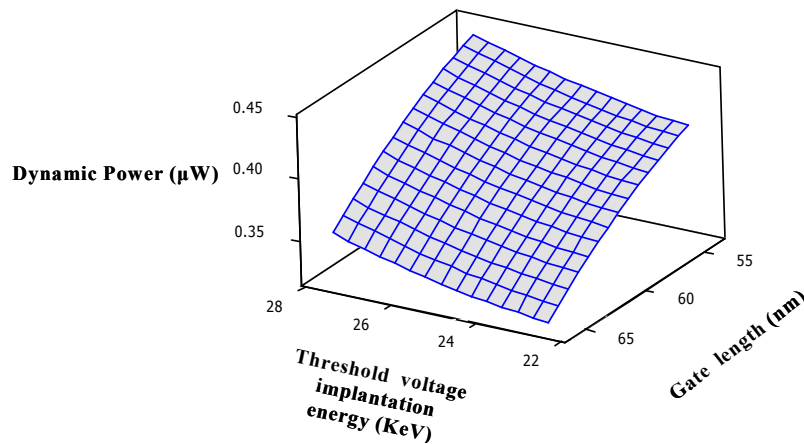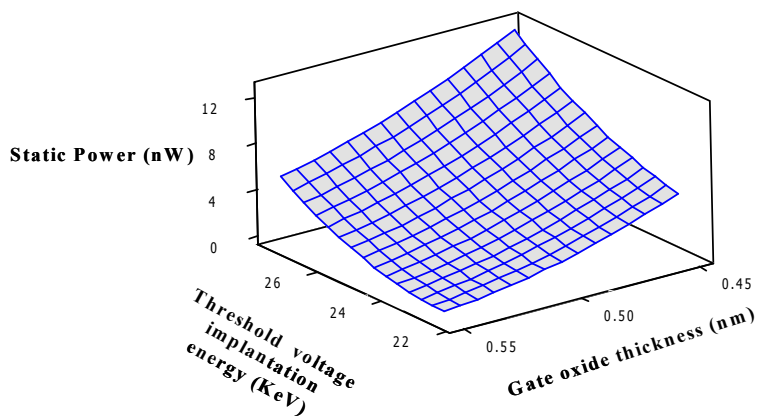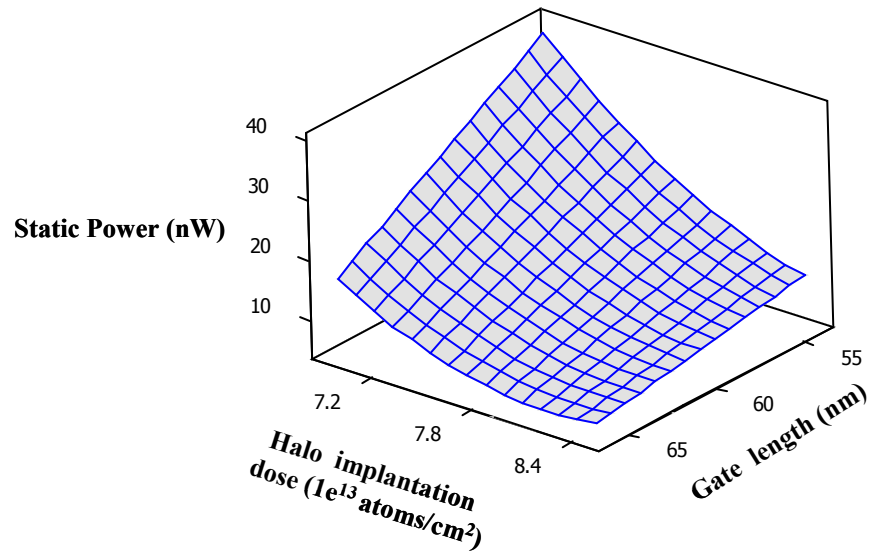
Figure 4.33: Response surface for static power consumption of the standard C-element based on bulk silicon technology with respect to gate length and halo implantation dose.

From the RS plots for the standard C-element based on bulk silicon technology, it can be seen that the gate length (x_{19}) variation largely impacts on the propagation delay and the dynamic and static power consumption of the circuit.

The impact of the variations in gate length (x_{19}) and high-k dielectric thickness (x_2) on the propagation delay of the circuit is shown in Figure 4.31. It can be seen that the use of a shorter gate length can largely reduce the sensitivity of the propagation delay in terms of the high-k dielectric thickness variations, as well as reducing the values of both propagation delay and dynamic power consumption of the circuit.

From Figure 4.33 it can be seen that the variations in gate length (x_{19}) and halo implantation dose (x_9) largely impact on the static power consumption of the circuit. However, in contrast to the propagation delay of the circuit, the use of a longer gate length reduces the sensitivity of static power consumption to the variation in the halo implantation dose.

Considering the results obtained for each of the process technologies investigated, it can be seen that the gate length and the silicon layer thickness variation dominates the performance metrics of the circuit based on FDSOI technology, in terms of the propagation delay and dynamic power consumption, while this sensitivity to silicon layer thickness variation in the case the circuit based on PDSOI is largely minimised.

In the case of the circuit based on bulk silicon, the performance metrics are significantly influenced by gate length variations and the halo implantation process steps.

Finally, form the results of standard implementation of C-element based on all process technologies investigated, it can be seen that the gate length is a common source of process variation. Therefore, it is necessary to find ways and methods of reducing gate length variation in order to reduce its impact from the early stages of process manufacturing on the key performance metrics and yield.

4.7 Summary and Conclusions

The DoE/RSM methods were utilised in order to study the impact of process variability in FDSOI, PDSOI, and bulk silicon technologies on propagation delay and the static and dynamic power consumption of different C-element configurations.

The threshold voltage variations for both N and PMOS devices have a significant role in terms of variations in delay and power, and optimisation of threshold voltage must be applied to reduce its variations. Assigning higher power supply voltages and using lower threshold voltage devices improves the circuit performance in terms of delay variations. However, trade-offs between the variation in delay and power must be considered.

It was found that the dynamic implementation of the C-element is more robust in terms of process variations compared to the other implementations in terms of power consumption and delay variations.

The presence of the weak feedback inverter in the weak feedback, standard and resistive weak feedback implementations, has degraded the circuit performance and its resilience towards variation compared to the dynamic C-element implementation. It also was observed that controlling the weak feedback inverter, as in standard and resistive weak feedback implementations, improves its robustness to the effects of process variations.

The differential implementation of the C-element shows much more robustness to process variation in comparison to the static and dynamic implementations of the C-element in terms of the performance metrics investigated.

It was found that all C-element circuits realised in an SOI technology are more resilient to process variations with higher performance metrics compared to their counterparts

based on bulk silicon. The SOI technology (FDSOI and PDSOI) has been explored in terms of process variations. It can be concluded that the SOI technology has the potential to reduce the impact of process variations over silicon based circuits.

In depth variability analysis was performed by analysing the current variation of each transistor in the circuit. Using such methods helps to explain the influence of circuit topology on performance variability and to identify the most dominant devices on circuit performance variability. Hence, by optimising the size of such devices, it is possible to reduce their impact on performance variability.

The direct impact of the manufacturing process steps on the performance of standard implementation of a C-element circuit based on the technologies investigated, was undertaken with the aim of determining and identifying the critical process parameters that need to be controlled, and to find ways to reduce the variability in terms of the sources, hence improving the performance and yield of a circuit.

It was found that the gate length and the silicon layer thickness variations dominate the performance metrics of circuits based on FDSOI technology in terms of the propagation delay and dynamic power consumption. At the same time, this sensitivity to silicon layer thickness variation in the case circuits based on PDSOI has been largely reduced.

In the case of circuits based on bulk silicon, the performance metrics are significantly influenced by gate length variations and the halo implantation process steps.

It was found that the impact of gate length variation on the propagation delay and the dynamic power consumption of the circuit studied is commonly significant in terms of process parameters for all the process technologies investigated.

Therefore, new methods are required for minimising the impact of gate length variation, during the early stages of process manufacturing and circuit design on the key performance metrics and yield.

4.8 References

[1]   K. Van Berkel, R. Burgess, J. L. W. Kessels, A. Peeters, M. Roncken, and F. Schalij, "A fully asynchronous low-power error corrector for the DCC player," IEEE Journal of Solid State Circuits, vol. 29, pp. 1429-1439, 1994.

[2]   M. Shams, "Modelling and Optimization of CMOS Logic Circuits with Application to Asynchronous Design," PhD thesis, University of Waterloo, 1999.

[3]   D. Muller and W. Bartky, "A theory of asynchronous circuits," Proceedings of the International Symposium on the Theory of Switching, pp. 204-243, 1959.

[4]   M. Shams, J. C. Ebergen, and M. I. Elmasry, "Optimizing CMOS implementations of the C-element," Proceedings of the 1997 IEEE International Conference on Computer Design: VLSI in Computers and Processors, pp. 700-705, 1997.

[5]   M. Shams, J. C. Ebergen, and M. I. Elmasry, "A comparison of CMOS implementations of an asynchronous circuits primitive: the C-element," International Symposium on Low Power Electronics and Design, pp. 93-96, 1996.

[6]   A. J. Martin, "Formal Program Transformations for VLSI Circuit Synthesis," UT Year of Programming Inst. Formal Developments of Programs and Proofs, E.W. Dijkstra, ed. Reading, Mass.: Addison-Wesley, 1989.

[7]   Sutherland I E, "Micropipelines," Communications of the ACM, , vol. 32,no 6, pp. 720-738, June 1989.

[8]   K. Vanberkel, "Beware the Isochronic Fork," Integration, the VLSI Journal, vol. 13, pp. 103-128 Jun 1992.

[9]   H. Mahmoodi, S. Mukhopadhyay, and K. Roy, "Estimation of delay variations due to random-dopant fluctuations in nanoscale CMOS circuits," IEEE Journal of Solid State Circuits, vol. 40, pp. 1787-1796, 2005.

[10]   V. Wang and D. Markovic, "Linear analysis of random process variability," IEEE/ACM International Conference on Computer-Aided Design, pp. 292-296, 2008.

[11]   Montgomery, "Design and Analysis of Experiments," 5th ed: John Wiley and Sons, 2001.

Chapter 5

Single Event Upset Analysis of C-Element Circuits

5.1 Background

In this chapter, several aspects related to the single event upsets (SEUs) and their impact on VLSI circuits are reviewed, namely the physical origin of radiation particles, the charge deposition and the collection mechanisms involved, and the impact of technology scaling on the radiation tolerance of VLSI circuits. SEU mitigation techniques are also briefly discussed.

Thereafter, the impact of key design parameters such as threshold voltage, power supply and width ratio on the resilience of various C-element configurations are investigated. The RS technique is utilised to give useful measures with regard to the sensitivity of the resilience of the circuit to radiation particle strikes with respect to the key parameters under consideration such as threshold voltage, power supply voltage, and width ratio.

The use of different materials and device structures such as FDSOI and PDSOI technologies, in terms of the reduction of the impact of radiation particle strikes at device level, is explored. A comparative study using soft error analysis of different C-element circuits is included; the effectiveness of the differential logic with an inverter latch implementations of the C-element in terms of the prevention of soft errors is highlighted. Finally, the chapter is concluded with a summary of the results.

5.2 Radiation Particle Strikes

The continuous scaling of device features has led to a huge reduction in the node capacitances of integrated circuits, and subsequently in the charge stored. As a result, the reliability of circuits is reduced in terms of the energetic particle strikes that induce soft errors. In previous technologies, this problem was limited to hostile environment applications such as space. However, the effects are now being observed at ground level.

In fact, the first recorded incident with regard to radiation particles inducing upsets in space (four upsets in 17 years of satellite operation) was reported in 1975 [1]. However, just four years later, soft errors were also observed at ground level [2]. Since then, with technology scaling, many cases of soft errors have been observed with regard to both space and terrestrial electronics [3].

In addition, the increased clock frequency and reduced supply of voltage requirements have also aggravated the intolerance of electronic circuits to radiation-induced soft errors [4, 5]. It is also worth noting that the incidence of low energy particles is much greater than that of high energy particles [6]. Soft errors are a major threat in critical applications where reliability is the central concern rather than performance and cost, such as in biomedical or avionics applications.

In general, single event upsets in microelectronic circuits are caused when radiation particles such as protons, neutrons, alpha particles, or heavy ions hit sensitive regions (typically reverse-biased p-n junctions) in digital circuits. The particle strikes can deposit a charge at the affected node, resulting in a voltage pulse or glitch, which in turn can result in a soft or transient error.

Radiation particle strikes are a very problematic issue for memories (latches, SRAMs, and DRAMs) as they can directly change the stored logic state of a memory element, resulting in a single event upset (SEU) [2, 7]. Although radiation-induced errors in sequential elements will continue to be problematic for high performance circuits, it is expected that the soft errors effects in combinational logic circuits will dominate future technology nodes [3, 8, 9]. Figure 5.1 compares the critical charge needed to cause a soft error with regard to the SRAM and the different combinational logic circuits. From Figure 5.1 it can be seen that the critical charge of SRAM cells and logic circuits has dramatically decreased as technology scales. However, the critical charge of logic circuits has fallen at a faster rate, making their robustness to soft error a significant concern. Logic circuits are also more difficult to harden than SRAMs. This is because most of the mitigation techniques have come with large area overhead and latency penalties.
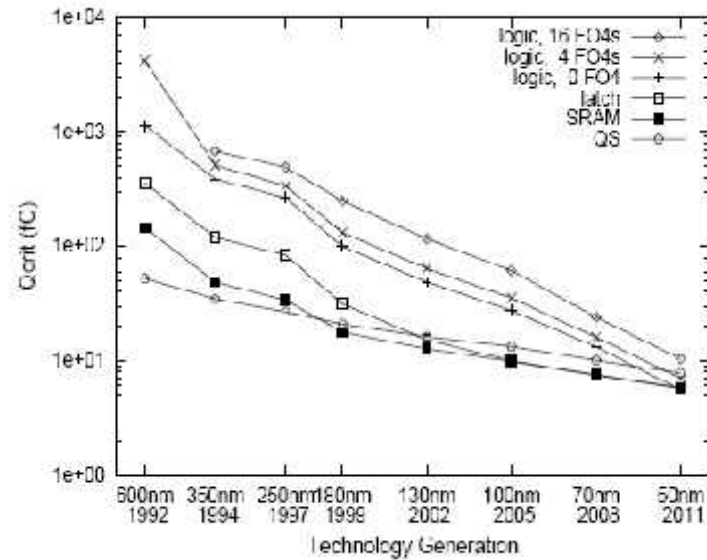
Figure 5.1: Critical charge for SRAM, latch and logic circuits [8].

The upset mechanism in logic circuits is different from that of memory, as it depends on many factors such as the drive strength of the gate, the fan out capacitance of the gate, clock speed, and logic depth.

In a logic circuit, a voltage glitch due to a radiation particle strike can propagate to the input(s) of a memory element, which might result in an incorrect value being latched, leading to single or multiple upsets.

The propagation of the voltage glitch caused by a radiation particle strike in a logic (combinational) circuit depends on three masking factors as follows [3, 9]:

- **Electrical masking:** this occurs when a voltage glitch is attenuated by subsequent gates in the logic path. Such masking can minimise the voltage magnitude of a glitch, resulting in a small voltage value which cannot cause any soft errors.

- **Logical masking:** this occurs when there is no functionally sensitive path from the node in the circuit (where a radiation particle may strike) to the input(s) of the memory element.

- **Temporal masking:** this occurs when a voltage glitch induced by a radiation particle strike reaches the inputs of a circuit at an instant far from the latching window of the sequential elements of the circuit. Temporal masking is only affected by the operational frequency of the circuit.

123

5.3 Physical Origin of Radiation Particles

In space, the main source responsible for soft errors in the electronic applications are cosmic rays, often referred to as galactic cosmic rays, which are high-energy charged particles composed of protons, electrons, and heavier nuclei [3].

Apart from cosmic rays particles, protons caused by solar events, and those trapped in the radiation belts of the earth, are other sources of protons present in the atmosphere of the earth which are also capable of inducing SEUs in a circuit [3].

SEUs are also known to be caused by alpha particles emanating from the naturally available radioactive elements present on the surface of the Earth, and might also originate from radioactive contaminations in IC packaging materials [3, 10]. Recently, flip-chip packages have been recognised as a source of radiation particles (from the Pb-Sn solder bumps) [11]. This expands the problem of developing hardening techniques against radiation, because a source of radiation particles is extremely close to the die. Moreover, on the surface of the earth, neutrons that induced upsets have been found to be very problematic. Several studies have reported that the neutrons from cosmic rays are a significant source of soft errors for SRAMs operating at the ground level [3, 8].

5.4 Charge Deposition Mechanisms

In VLSI designs, the radiation particles can deposit charges on the nodes of a circuit through two mechanisms: direct ionization and indirect ionization, depending on the type of the radiation particles [3].

Direct ionization is the process of generating electron-hole pairs in a semiconductor material due to radiation particles. When a radiation particle strikes a semiconductor material, it passes through the material. Along its path it loses its energy, leading to the generation of electron-hole pairs. The energy released by radiation particles is often described by linear energy transfer (LET) value [3]. LET (MeV-cm^2/mg) determines the energy transferred to the target material (for electron–hole pair generation) by the radiation particle per unit length, normalized by the density of the target material - such

as silicon - in VLSI designs. This also corresponds to the amount of charge per unit length deposited by the radiation particles.

Heavy ions and alpha-particles mainly deposit charge in a semiconductor by the direct ionization mechanism. However, light particles such as protons and neutrons do not deposit enough charge by direct ionization to cause a single event upset.

Indirect ionization is a process that occurs when high-energy and light particles such as protons and neutrons pass through a semiconductor material, and collide with nuclei, resulting in the production of secondary particles such as alpha-particles or heavy ions.

The resulting particles then deposit their charge near the impact area by the direct ionization mechanism. The charge deposited by such indirect ionization is strongly dependent on the location and the angle of the incident strike.

5.5 Charge Collection Mechanisms

There are three major mechanisms involved in the collection of charges which cause SEUs, namely, drift, diffusion and funnelling mechanisms. These are depicted in Figure 5.2.
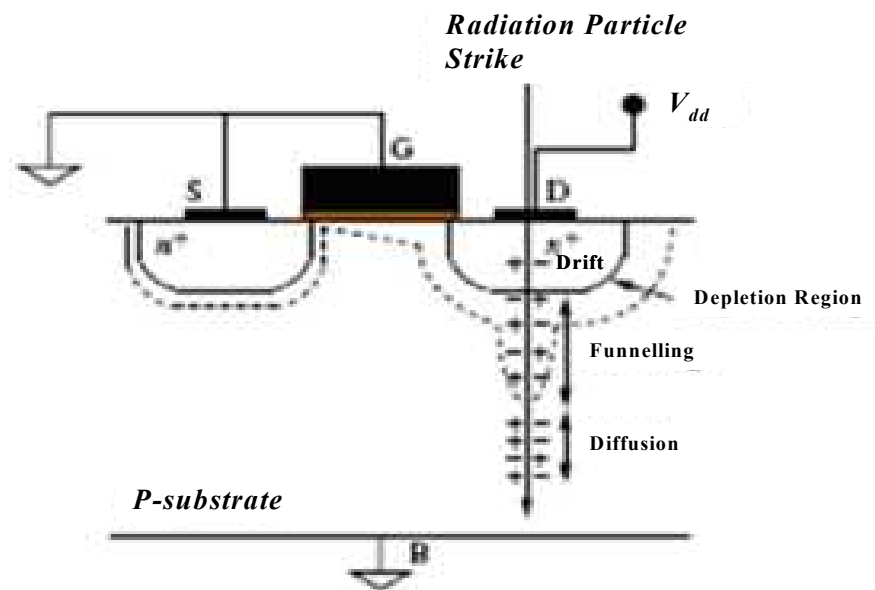
Figure 5.2: Charge deposition and collection caused by a radiation strike [12].

125

When a particle strikes a sensitive region in a circuit such as a reverse biased p/n junction, most of the hole-electron pairs induced by the strike can be collected by the drift mechanism due to the presence of the strong electric field in the reverse biased junction, leading to a high current in a short period of time at the junction contact.

The funnelling process then assists in collecting the charge by extending the depletion region, and thus increases the junction electric field. This leads to more charge collection as shown in Figure 5.3 which shows electron concentration due to funnelling [3]. Thereafter the diffusion mechanism, mainly due to the concentration gradient of the carriers, slowly collects the remaining carrier in the junction region, leading to a lengthy collection time in comparison to the drift mechanism.

Some studies have reported that in a lightly-doped substrate, the dominant collection charge mechanism is the drift process, whereas with higher doped substrates, both the drift and the diffusion processes are responsible for charge collection [13, 14].
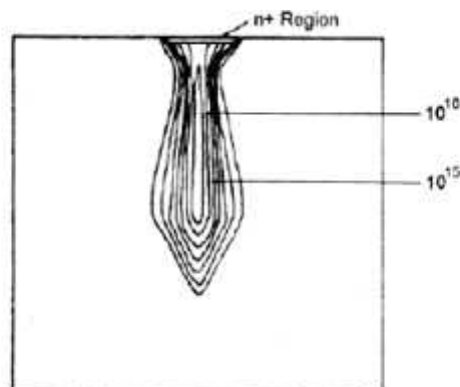
Figure 5.3: Electron concentration due to funnelling in an n^+/p silicon junction following an electron strike [3].

In deep submicron devices, another collection mechanism, ALPEN, is capable of causing an SEU in a circuit [3]. ALPEN, or Alpha Penetration, occurs when a radiation particle hits a MOSFET device in the form of a near-grazing incident, as shown in Figure 5.4. The particles penetrate through both the source and the drain regions of an (off) MOS transistor, resulting in a significant transient current that mimics the "on" state of the transistor. Dodd et al. reported that the charge collection of an ALPEN mechanism, tends to increase as the channel length decreases below 0.5 μm [13]. This increases the susceptibility of future generation technology processes to radiation effects.
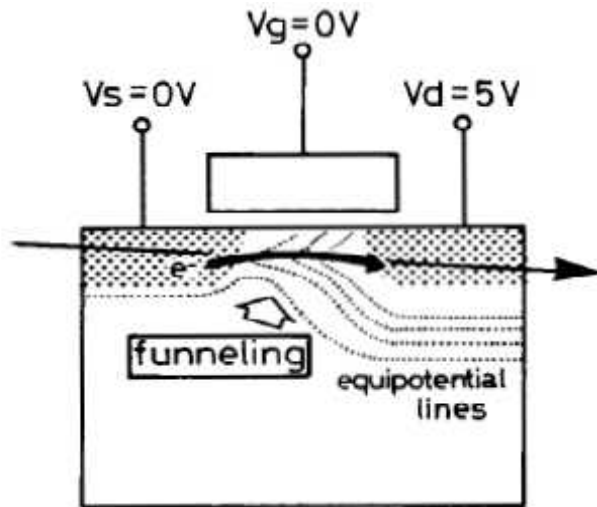
Figure 5.4 : The ALPEN effect [15].

A further known effect is the bipolar transistor effect. This occurs when a particle strikes an n-channel MOSFET transistor as an example, the holes generated by the particles are attracted to the substrate leading to an increase in the substrate potential, which results in effectively lowering the source/substrate potential barrier. The lowered potential barrier makes the source inject electrons into the device channel, and hence these can be collected at the drain terminal; this increases the particle-induced current. The bipolar effect mimics the "on" state of the parasitic bipolar transistor where the source acts as the emitter, the channel as the base region, and the drain as the collector. Scaling the channel length largely decreases the base width of the parasitic bipolar transistor. Therefore this effect becomes more prominent in scaled technologies [3, 13, 16].

5.6 Impact of Technology Scaling on the Radiation Tolerance of VLSI Design

The continuous scaling of device features has led to a huge reduction in the node capacitances of VLSI circuits. As a result, a small amount of charge deposited by a radiation strike is sufficient to cause an SEU at the node. Moreover the increased clock frequency and the reduced supply voltage requirement also aggravate the tolerance of the circuits to SEUs [5, 17].

The increased demand for reducing the gate length of the devices also increases the impact of ALPEN and bipolar transistor effects which, in turn, increases the susceptibility of the circuits to soft errors. Dodd et al. reported that the ALPEN effect can occur in 300nm gate length devices, even for normal incidences, and can lead to charge multiplications [3]. It is also reported that the bipolar effect becomes more effective even with light particle strikes, leading to direct ionization, and hence increases the efficiency of the charge collection mechanism in advanced MOSFET structures [3].

Although technology scaling has a severe impact on reducing tolerance of circuits to radiation, there are advantageous factors associated with scaling that improve the radiation tolerance of circuits. The drain area reduction due the scaling helps to reduce the probability of striking a sensitive transistor, as well as reduces the collection volume of the drain depletion area. The reduction of the power supply voltage also reduces the charge collection efficiency which helps to reduce the impact of soft error rates as a result of scaling [3, 16, 18].

5.7 SEU Mitigation Techniques

The most fundamental method for hardening against SEU is to reduce the charge collection volume in a device [19, 20]. This can be done be using an epitaxial substrate instead of a bulk substrate and/or using extra doping layers [21]. Figure 5.5 illustrates the reduction of the collected charge volume using an epitaxial layer compared to bulk substrate.
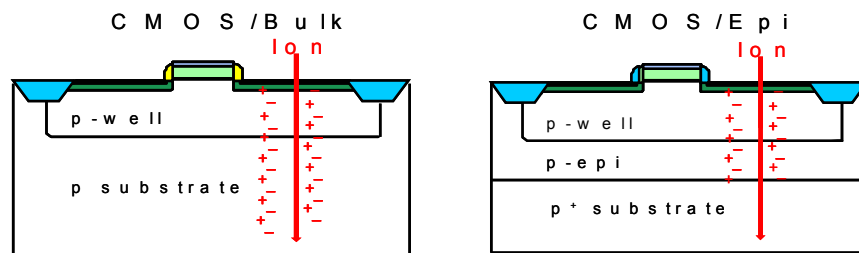
Figure 5.5: An illustration of the charge volume reduction using an epitaxial layer compared to bulk substrate.

Another effective way of reducing the collection volume of the charge in silicon devices is the use of SOI substrates [22]. In this case, as shown in Figure 5.6, the collection volume is reduced since the thin silicon layer that forms the active device is isolated from the substrate. This substantially reduces the SEU-sensitive area because the reverse-biased drain junction area is limited to the depletion region between the drain and the body of the transistor. The charge deposited underneath the buried oxide layer (BOX) cannot be collected at the drain, as it is electrically isolated from the silicon substrate. However, recent research has indicated that capacitive coupling across the BOX can lead to unexpected charge collection in SOI devices [3, 23].
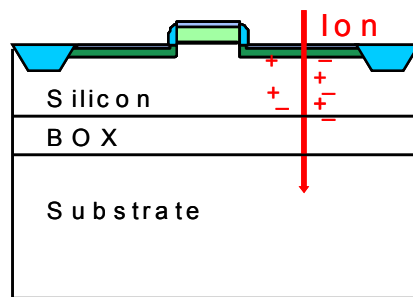
Figure 5.6: An illustration of the charge volume reduction using SOI technology.

Improvements in SEU hardness can also be obtained at circuit-level, the circuit hardening techniques aim to reduce single event upset effects by reducing the sensitivity of the gate to SEUs. Circuit-level hardening techniques mostly involve SER prediction, duplicating the sensitive gates [24], and sizing the gates to reduce SEU sensitivity [25, 26].

In the next sections, the analysis of the robustness to the effects of soft errors of the various C-element configurations (dynamic, static, and differential implementation) realised in different process technologies, is presented.

5.8 Analysis of Single Event Upset of C-Element Circuits

The C-element configurations analysed, shown in Figure 5.7, comprise dynamic, standard and weak feedback, resistive weak feedback and symmetric implementations [27, 28]. All of the C-element circuits investigated can be considered to be modified versions of the

simplest circuit, that is, the dynamic C-element shown in Figure 5.7 (a); subsequent circuits differ only in terms of the method of preserving the output state when the inputs have different logic values.

$$W_{n1}, W_{n2}, W_{n3} = 1\mu m$$
$$W_{p1}, W_{p2}, W_{p3} = 2.5\mu m$$
$$W_{n4}, W_{n5}, W_{n6} = 0.1\mu m$$
$$W_{p4}, W_{p5}, W_{p6} = 0.25\mu m$$
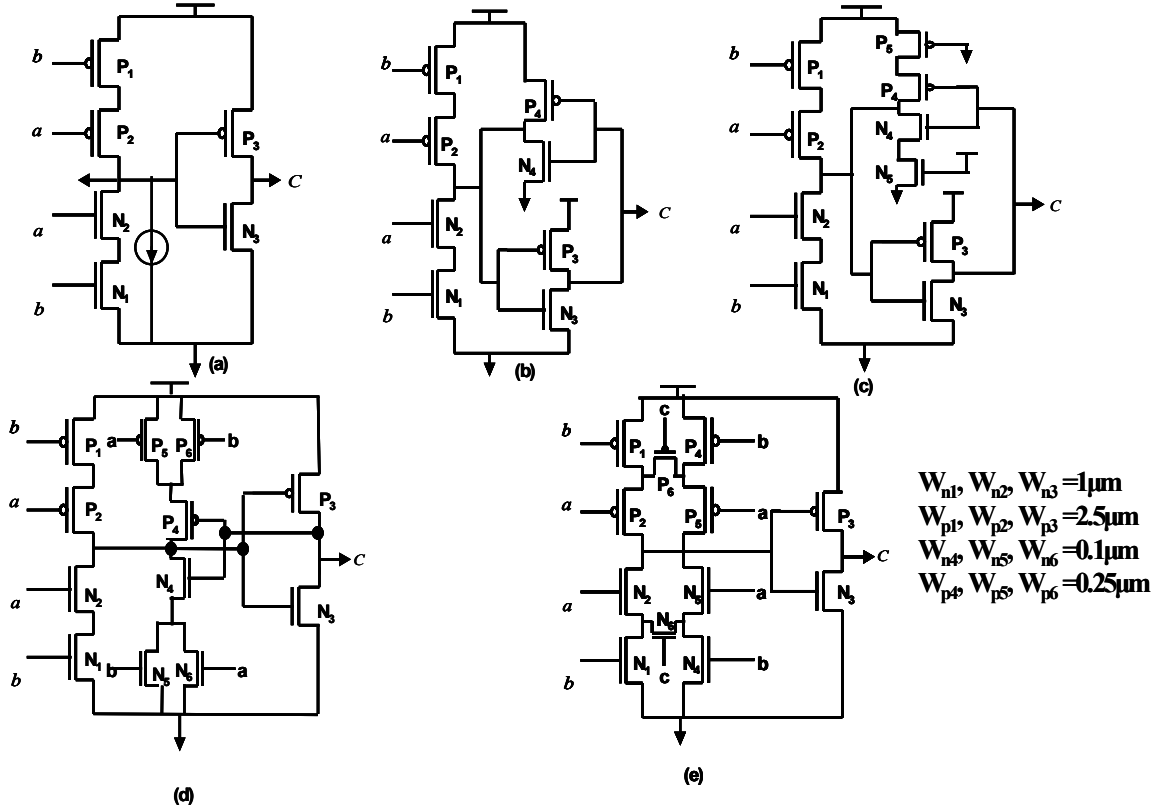
Figure 5.7: Implementation of the C-element: (a) dynamic, (b) weak feedback (c) resistive weak feedback (d) standard and (e) symmetric [27, 28].

In investigating the robustness of the various C-Element configurations to an SEU at circuit level, simulations were undertaken with a particle strike modelled as a double exponential current source [29] with a fast rise and slow fall time using equation (5.1).

$$I_{SEU}(t) = I_{peak}.(e^{\frac{-t}{\tau_a}} - e^{\frac{-t}{\tau_b}}) \quad (5.1)$$

$$I_{peak} = \frac{Q}{\tau_a - \tau_b} \quad (5.2)$$

where I_{peak} is the current pulse amplitude which is given in equation (5.2), τ_a is the collection time constant, and τ_b is the ion-track establishment time constant.

130

The current source is connected to the $\overline{C}$ node as shown in Figure 5.7(a). This node was selected because it has the strongest reversed biased pn junction (where the N1, N2 are off) among the other nodes in the C-element circuit studied. This means that this node is more susceptible to the effects of a particle strike on the circuit.

The time constant parameters τ_a and τ_b are taken as 250ps and 10ps respectively. The peak of the current source is varied iteratively to find out the minimum amount of charge, which causes the output node (C) to flip its logic value, and is referred to as critical charge $Q_{crit.}$

The response surface modelling (RSM) technique was used to analyse the impact of variations in the supply voltage (V_{dd}), the threshold voltage of both N and P MOSFET devices and the width ratio ($K = W_P/W_n$) on the value of the critical charge.

Figures 5.8 and 5.9 show how the variations in V_{dd}, V_{th} for P and NMOS devices and width ratio (K) affect the value of the critical charge in the dynamic implementation of the C-element, as shown in Figure 5.7(a), realised in bulk CMOS technology.

Figure 5.8: Surface plots of critical charge (Q_{crit}) of the dynamic C-element in terms of variations of (a) the threshold voltage of PMOS devices and width ratio (K) (b) the threshold voltage of PMOS devices and power supply voltage (V_{dd}).

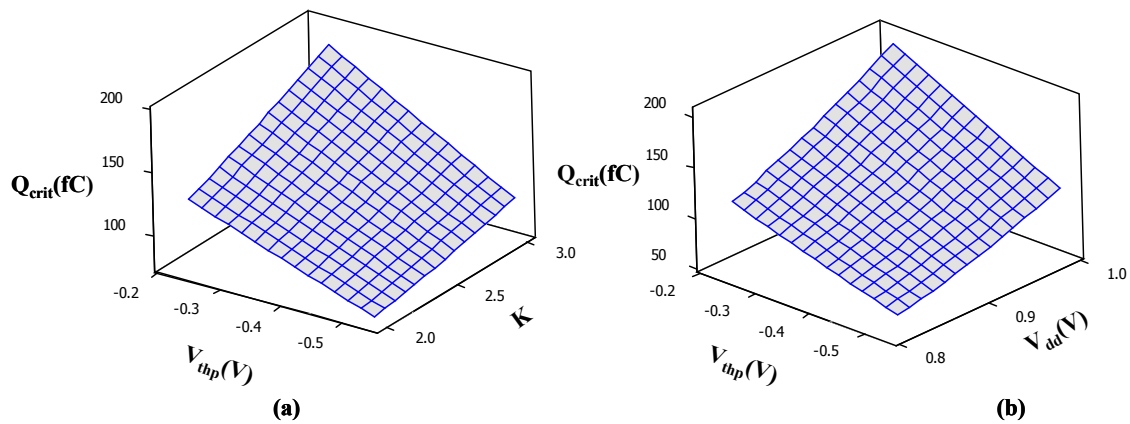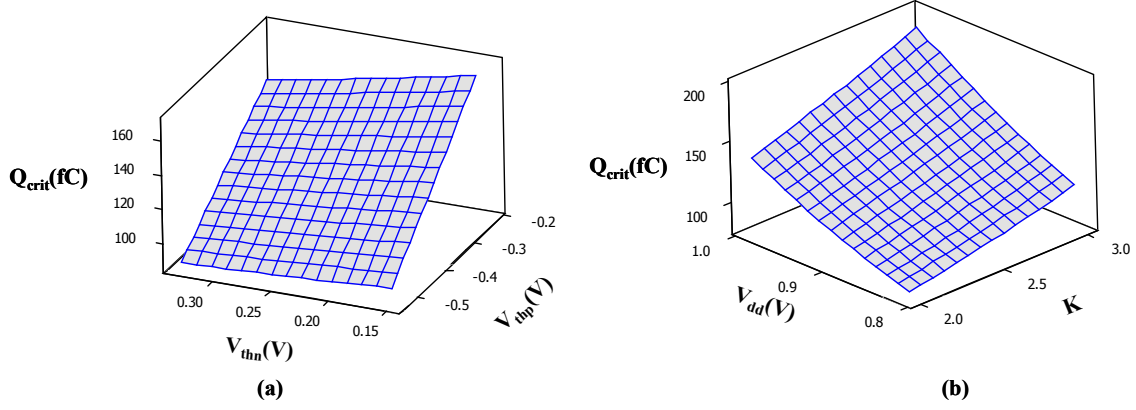**Figure 5.9: Surface plots of critical charge (Q_{crit}) of the dynamic C-element in terms of variations of (a) the threshold voltage of PMOS and NMOS devices (b) the power supply voltage (V_{dd}) and width ratio (K).**

From the above results, it can be seen that assigning a higher power supply voltage increases the value of Q_{crit} and hence increases the robustness of the circuit to soft errors. The use of lower threshold voltages for PMOS devices improves the robustness of the circuit with regard to soft errors. However, the threshold voltage of NMOS devices has a very small impact on the Q_{crit}. Similarly, increasing the size of the pull–up transistors (P1, P2) subsequently increasing the width ratio (K) improves the robustness of the circuits to soft errors. This can be explained by expressing the current at node $\overline{C}$ as follows:

$$i_{\overline{c}}(t) = i_{on}(t) - i_{SEU}(t) \quad (5.3)$$

$$i_{\overline{c}}(t) = C_{\overline{c}} \cdot \frac{dv_{\overline{c}}}{dt} \quad (5.4)$$

where $C_{\overline{C}}$ is the capacitance at node $\overline{C}$, and $V_{\overline{C}}$ is the voltage at node $\overline{C}$, while $i_{on}(t)$ represents the current required to charge up the node $\overline{C}$ and $i_{SEU}(t)$ which is given in equation (5.1).

From equations (5.3) and (5.4), by increasing the strength of the restoring devices, in this case the pull up network (P1, P2), it is possible to increase the rate of the removal of the charge collected at a node, and hence reduce the sensitivity of the node to a single event upset. In other words, it is possible to improve the robustness of circuit to soft errors by (1) increasing the node capacitance. This can be done by connecting a capacitor to the selected node. However this will degrade circuit performance; (2) increasing the driving

strength, that is $i_{on}(t)$ of the restoring devices that supply charges to the node subjected to a single event strike.

Another important factor to consider is the pulse width of the voltage glitch induced as a result of a radiation practical strike, the propagation of the voltage glitch to the primary outputs of the circuit can increase the pulse width resulting in pulse spreading [30]. This will consequently increase the possibility of the soft error being latched in a memory element. In addition, the increased pulse width of the voltage glitch reduces the efficiency of the electrical and temporal masking in the logic circuits.
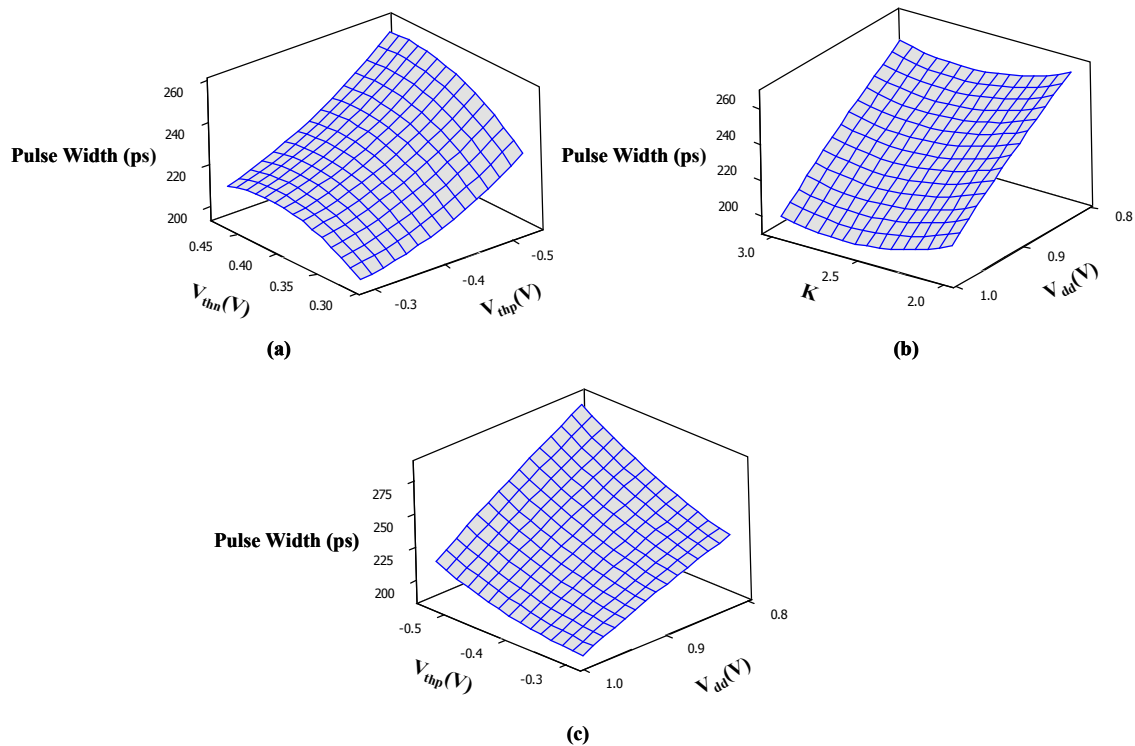
Figure 5.10: Surface plots of pulse width of the dynamic C-element in terms of variations of (a) the threshold voltage for N and P devices (b) the power supply voltage (V$_{dd}$) and width ratio (K).and (c) the power supply voltage (V$_{dd}$) and the threshold voltage of PMOS devices.

From the above results, it can be seen that the power supply voltage and the threshold voltage of the PMOS devices have a major influence on the width of the voltage pulse, while the sizing ratio and threshold voltage of NMOS devices have less impact on the pulse width.

It can also be observed that assigning a higher power supply voltage can significantly improve the robustness to soft errors by reducing the induced pulse width and hence reducing the rate of soft error from being latched in a memory element. In a similar manner, the use of lower threshold voltage devices, especially PMOS devices, (i.e. ones that improve the driving strength of the restoring devices) can reduce the width of the pulse induced by a radiation strike.

One way to improve the robustness of circuits to soft errors is by using silicon on insulator (SOI) technology. In this work, the impact of SEU on circuits based on FDSOI, PDSOI and bulk silicon have also been studied.

The same analysis is undertaken for SOI technology. It is found that all circuits have similar trends in terms of soft error robustness, and that the circuits based on FDSOI are the most robust because the charge collection volume is less compared to those based on PDSOI and on bulk silicon technologies. Therefore, it can be concluded that silicon layer thickness is an important factor in SOI technology in that it determines the value of the charge collection volume.

Figure 5.11 compares the minimum amount of charge, which causes the output node (C) to flip its logic value (i.e. the critical charge (Q_{crit})) of different C-element circuits based on FDSOI, PDSOI and bulk silicon technologies.

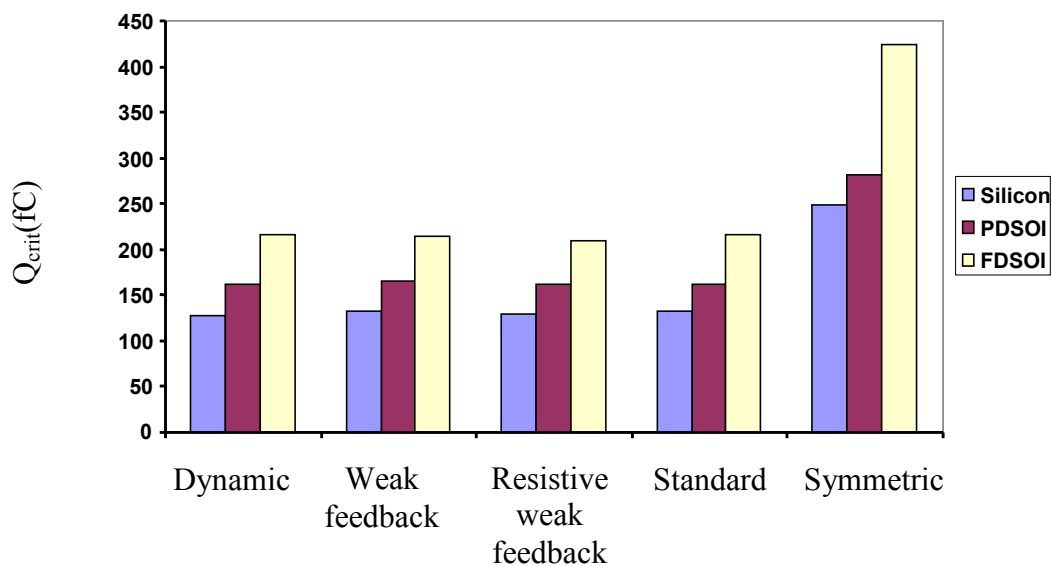**Figure 5.11: Comparison of the critical charge (Q_{crit}) of different C-element circuits based on FDSOI, PDSOI and bulk silicon devices.**

From Figure 5.11, it can be seen that the symmetric implementation of the C-element circuit is more robust in terms of soft errors since it requires a higher Q_{crit} to cause an SEU. This is because the node $\overline{C}$ is strongly pulled up by parallel transistors (Figure 5.7(e)), leading to an increase in i_{on} (t) at node $\overline{C}$ as in equation (5.3). When both inputs are low, the transistors, P1 and P2, are in parallel with P4 and P5 respectively. This means that the node $\overline{C}$ is pulled up strongly in comparison to other C-element implementations. At the same time, the other implementations have almost the same drive current that supplies node $\overline{C}$, and hence they have very close values to Q_{crit}.

Figure 5.12(a) shows the basic differential logic and an inverter latch implementations of the C-element [31]. It is composed of two pull-down networks consisting of NMOS devices, and an inverter latch formed by the transistors P_{FL}, N_{FL}, P_{FR} and N_{FR}.

When both inputs (a , b) are high, then $\overline{C}$ becomes low, and C is pulled up by the PMOS device of the right hand side inverter of the latch, P_{FR}. In a similar manner, when the inputs are low, then C becomes low and $\overline{C}$ is pulled up by the PMOS device of the left inverter of the latch, P_{FL}. However, when the inputs do not match ($a \neq b$), the inverter latch maintains the previous state of the output. As a result, the inverter latch is only used for holding the output logic states when the inputs do not match.

Another possible modification of the DIL implementation is the DILP, shown in Figure 5.12(b). The DILP implementation is intended to reduce the rising delay of the basic DIL implementation by adding two pull-up networks composed of PMOS devices.

After applying soft error analysis, it was found that in the C-element implementations that use differential logic and an inverter latch, the output logic was unaffected by particle strike on the $\overline{C}$ node. This is independent of the amount of charge deposited at the $\overline{C}$ node, as shown in Figures 5.13 and 5.14.

In this implementation, if the node $\overline{C}$ is hit by a particle radiation strike, then $\overline{C}$ goes from high to low. This makes the PMOS device (P_{FR}) on the right hand side inverter of the latch conduct and try to pull up node C (output of the C-element circuit). However, because the transistors that are connected to the inverted inputs ($\overline{a}$, $\overline{b}$) are already switched on (N_{R1}, N_{R2}), the charge supplied by the PMOS transistor due to the strike is

discharged very quickly during the very small duration of the strike. It can also be seen that the magnitude of the output voltage is very small in comparison to V_{dd}.

Moreover, as can be seen from Figures 5.13 and 5.14 that the effect of the deposited charge Q from which the output voltage reaches its maximum value is higher for DILP implementation compared to DIL implementation of C-element.

$W_{NL1}, W_{NL2}, W_{NR1}, W_{NR2} = 1\mu m$
$W_{PL1}, W_{PL2}, W_{PR1}, W_{PR2} = 2.5\mu m$
$W_{NL1}, W_{NL2}, W_{NR1}, W_{NR2} = 1\mu m$
$W_{NFL}, W_{NFR} = 0.1\mu m$
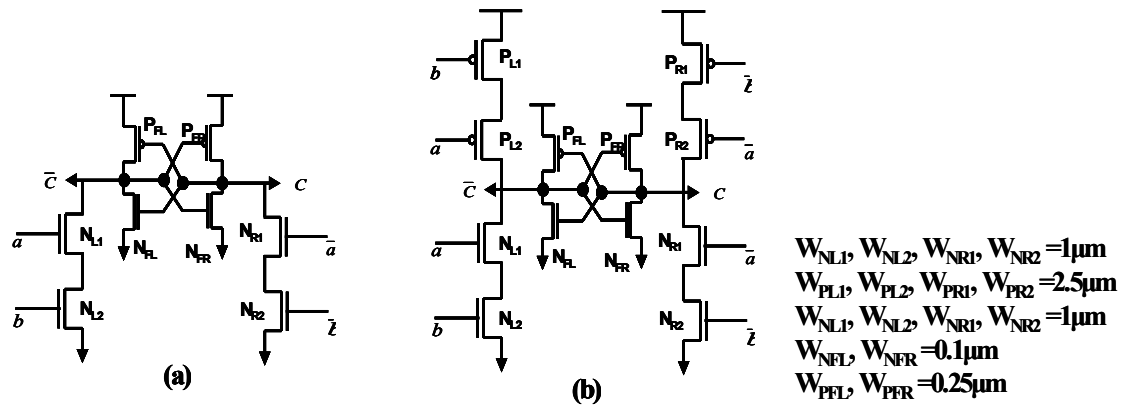$W_{PFL}, W_{PFR} = 0.25\mu m$

Figure 5.12: Differential logic with an inverter latch implementations of the C-element (a) DIL and (b) DILP.
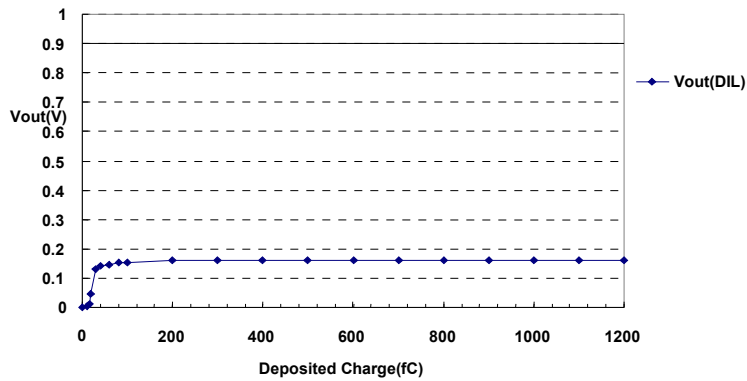

Figure 5.13: Output voltage of (DIL) implementation of the C-element versus the deposited charge (Q) by a particle strike.
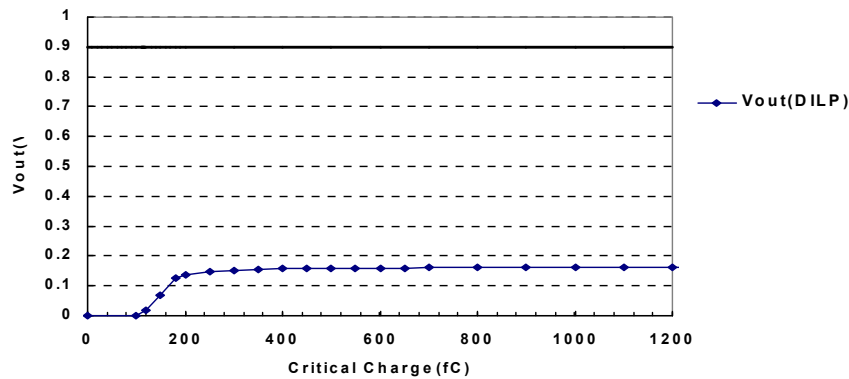
Figure 5.14: Output voltage of (DILP) implementation of the C-element versus the deposited charge (Q) by a particle strike.

5.9 Summary and Conclusions

An overview of different aspects related to the single event upsets (SEUs) and their impact on VLSI circuits has been presented in some detail in this chapter.

Soft error analysis on various implementations of the C-element has been carried out. The analysis involved the RSM method for studying the impact of each of the threshold voltages of NMOS and PMOS devices, power supply voltage (V_{dd}), the sizing ratio (K) and their interaction on the critical charge (Q_{crit}) necessary to create a single event upset (SEU) . It was found that the sensitivity of a node to a single event upset can be reduced by increasing the drive strength of the restoring devices which, in turn, increases the rate of removal of the charge collected at a node by a particle strike. This can be done by reducing the threshold voltages of the devices and increasing the power supply voltage.

In a similar manner, the impact of threshold voltage, power supply, and the sizing ratio on the width of the induced voltage pulse is analysed. It was found that the pulse width is largely influenced by the power supply and the threshold voltage of the PMOS devices. The pulse width can be reduced by adopting higher power supply voltages and lower threshold voltage devices, hence improving the robustness circuits to soft errors.

The analysis of different C-element implementations subsequently permits a relative comparison of Q_{crit} to be undertaken. The symmetric implementation of the C-element is

found to be the most robust in terms of soft errors compared to the dynamic and static implementations of the C-element.

The potential of silicon on insulator technology (SOI) in enhancing the resilience of a circuit to soft errors has been explored. It is observed that circuits based on FDSOI technology show higher tolerance compared to PDSOI and bulk silicon technologies.

Finally, The differential logic with an inverter latch implementations of the C-element were found to be the most robust with regard to soft errors as they were unaffected by radiation strikes and independent of the amount of charge deposited. This can be used as a possible way to mitigate the impact of soft errors in asynchronous designs.

5.10 References

[1]   D. Binder, E. C. Smith, and A. B. Holman, "Satellite Anomalies from Galactic Cosmic Rays," IEEE Transactions on Nuclear Science, vol. 22, pp. 2675-2680, 1975.

[2]   T. C. May and M. H. Woods, "Alpha-particle-induced soft errors in dynamic memories," IEEE Transactions on Electron Devices, vol. 26, pp. 2-9, 1979.

[3]   P. E. Dodd and L. W. Massengill, "Basic mechanisms and modelling of single-event upset in digital microelectronics," IEEE Transactions on Nuclear Science, vol. 50, pp. 583-602, 2003.

[4]   R. Baumann, "Soft errors in advanced computer systems," Design & Test of Computers, IEEE, vol. 22, pp. 258-266, 2005.

[5]   P. Hazucha, T. Karnik, J. Maiz, S. Walstra, B. Bloechel, J. Tschanz, G. Dermer, S. Hareland, P. Armstrong, and S. Borkar, "Neutron soft error rate measurements in a 90-nm CMOS process and scaling trends in SRAM from 0.25um to 90-nm generation," IEEE International Electron Devices Meeting, pp. 21.5.1-21.5.4, 2003.

[6]   A. V. Belov and E. A. Eroshenko, "Proton spectra of the four remarkable gle in the 22nd solar cycle," Radiation Measurements, vol. 26, pp. 461-466, 1996.

[7]   J. C. Pickel and J. T. Blandford, "CMOS RAM Cosmic-Ray-Induced-Error-Rate Analysis," IEEE Transactions on Nuclear Science, vol. 28, pp. 3962-3967, 1981.

[8]   P. Shivakumar, M. Kistler, S. W. Keckler, D. Burger, and L. Alvisi, "Modelling the effect of technology trends on the soft error rate of combinational logic," Proceedings of the International Conference on Dependable Systems and Networks, pp. 389-398, 2002.

[9]   Z. Quming and K. Mohanram, "Gate sizing to radiation harden combinational logic," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 25, pp. 155-166, 2006.

[10]   R. C. Baumann, "Soft errors in advanced semiconductor devices-part I: the three radiation sources," IEEE Transactions on Device and Materials Reliability, vol. 1, pp. 17-22, 2001.

[11]   B. M. Clark, M. W. Weiser, and I. J. Rasiah, "Alpha radiation sources in low alpha materials and implications for low alpha materials refinement," Thin Solid Films, vol. 462-463, pp. 384-386, 2004.

[12]   R. GARG, "Analysis and Design of Resilient VLSI Circuits," PhD thesis, Texas A&M University, May 2009.

[13]   P. E. Dodd, "Device simulation of charge collection and single-event upset," IEEE Transactions on Nuclear Science, vol. 43, pp. 561-575, 1996.

[14]   P. E. Dodd, F. W. Sexton, and P. S. Winokur, "Three-dimensional simulation of charge collection and multiple-bit upset in Si devices," IEEE Transactions onNuclear Science, vol. 41, pp. 2005-2017, 1994.

[15]   E. Takeda, K. Takeuchi, D. Hisamoto, T. Toyabe, K. Ohshima, and K. Itoh, "A cross section of alpha -particle-induced soft-error phenomena in VLSIs," IEEE Transactions on Electron Devices, vol. 36, pp. 2567-2575, 1989.

[16]   A. Mallajosyula, "Single Event Upset Hardened CMOS Combinational Logic and Clock Buffer Design," Master Thesis, University of New Mexico, December 2008.

[17]   S. Borkar, "Designing reliable systems from unreliable components: the challenges of transistor variability and degradation," Micro, IEEE, vol. 25, pp. 10-16, 2005.

[18]   T. Karnik and P. Hazucha, "Characterization of soft errors caused by single event upsets in CMOS processes," IEEE Transactions on Dependable and Secure Computing, vol. 1, pp. 128-143, 2004.

[19]   D. Burnett, C. Lage, and A. Bormann, "Soft-error-rate improvement in advanced BiCMOS SRAMs," Proceedings of the International Reliability Physics Symposium, pp. 156-160, 1993.

[20]   T. T. Kishimoto, M., Ohno, Y., Nishimura, T., and Inuishi, M., " Control of carrier collection efficiency in n+p diode with retrograde well and epitaxial layer," Jpn. J. Appl. Phys., vol. 36, pp. 3460 - 3462 1997.

[21]   F. Sai-Wai, A. M. Mohsen, and T. C. May, "Alpha-particle-induced charge collection measurements and the effectiveness of a novel p-well protection barrier on VLSI memories," IEEE Transactions on Electron Devices, vol. 32, pp. 49-54, 1985.

[22]   O. Musseau, "Single-event effects in SOI technologies and devices," IEEE Transactions on Nuclear Science, vol. 43, pp. 603-613, 1996.

[23]   P. E. Dodd, A. R. Shaneyfelt, K. M. Horn, D. S. Walsh, G. L. Hash, T. A. Hill, B. L. Draper, J. R. Schwank, F. W. Sexton, and P. S. Winokur, "SEU-sensitive volumes in bulk and SOI SRAMs from first-principles calculations and experiments," IEEE Transactions on Nuclear Science, vol. 48, pp. 1893-1903, 2001.

[24]   K. Mohanram and N. A. Touba, "Partial error masking to reduce soft error failure rate in logic circuits," Proceedings of the18th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems, pp. 433-440, 2003.

[25]   Y. S. Dhillon, A. U. Diril, A. Chatterjee, and A. D. Singh, "Sizing CMOS circuits for increased transient error tolerance," pp. 11-16, 2004.

[26]   R. R. Rao, D. Blaauw, and D. Sylvester, "Soft Error Reduction in Combinational Logic Using Gate Resizing and Flip flop Selection," IEEE/ACM International Conference on Computer-Aided Design, pp. 502-509, 2006.

[27]   M. Shams, J. C. Ebergen, and M. I. Elmasry, "Optimizing CMOS implementations of the C-element," Proceedings of the 1997 IEEE International Conference on Computer Design: VLSI in Computers and Processors, pp. 700-705, 1997.

[28]   M. Shams, J. C. Ebergen, and M. I. Elmasry, "A comparison of CMOS implementations of an asynchronous circuits primitive: the C-element," International Symposium on Low Power Electronics and Design, pp. 93-96, 1996.

[29]   Z. Quming and K. Mohanram, "Cost-effective radiation hardening technique for combinational logic," IEEE/ACM International Conference on Computer Aided Design, pp. 100-106, 2004.

[30]   P. Dahlgren and P. Liden, "A switch-level algorithm for simulation of transients in combinational logic," in Twenty-Fifth International Symposium on Fault-Tolerant Computing, pp. 207-216, 1995.

[31]   M. Shams, J. C. Ebergen, and M. I. Elmasry, "A comparison of CMOS implementations of an asynchronous circuits primitive: The C-element," in 1996 International Symposium on Low Power Electronics and Design - Digest of Technical Papers New York: IEEE, pp. 93-96, 1996.

Chapter 6

Conclusions and Future Work

This concluding chapter summarizes the main points presented in the thesis, and highlights the important conclusions. This is followed by several key points with regard to future work.

6.1 Summary

In the last few decades, VLSI technology scaling has prompted a rapid growth in the semiconductor industry. With CMOS device dimensions descending into the nanometre regime achieving higher performance and increased functionality per unit area in integrated circuits has become much easier. However, the scaling trend raises new challenges to circuit designers and process manufacturers. First, larger process parameter variations in the current technologies cause a larger spread in the delay and power distribution in the circuits, and results in parametric yield and cost loss. In addition, ensuring the reliability of deep sub-micron (DSM) technologies against soft errors is a significant challenge. Soft errors occur because of the combined effects of the particle radiation strikes and the significantly reduced node capacitance of scaled technologies.

This thesis focuses on the issues related to process variations and reliability in deeply scaled CMOS and SOI technologies. The aim of this research has been to model and analyse the reliability of logic circuits with regard to the impact of process variations and soft errors and to find ways to minimise these effects using different process technologies, together with the implementation of different circuit architectures.

The research can be divided into two parts, the first of which addresses the issues related to the effects of process variation at device and circuit levels. Herein, the potential of SOI technology, based on a statistical TCAD framework, is explored as a possible way to

reduce the variation effects on the power and performance distribution of a circuit; modifications to the circuit topology are also considered as another route to improving the robustness of a circuit to the effects of process variation. The second part deals with the reliability of circuits investigating possible ways to reduce the effect of soft errors by either adopting process technologies other than bulk silicon, such as SOI or by modifying circuit design.

In Chapter 1, the background in terms of the scaling adopted in the semiconductor industry in the last few decades to improve device performance, and to increase the functionality per unit area was provided. The challenges associated with scaling such as static power, process variability, and reliability in terms of soft errors were outlined. The need for a new technology to circumvent these barriers was highlighted. In this regard, SOI technology was proposed as a potential candidate for overcoming the variability and the reliability challenges. The need for high yield and error resilient circuit architectures was also highlighted.

A detailed review of the sources and different types of variation, including inter-die, intra-die, systematic and random variations, was provided in Chapter 2. The effect of these variations on the performance at both device and circuit levels was also explored. Methodologies and approaches used to model variability such as the Worst Case-corner Analysis method, Monte Carlo technique, the Sensitivity Analysis approach and Statistical Static Timing analysis were reviewed.

Thereafter, the need of Silicon on Insulator (SOI) technology, in the view of the effects of process variation and soft errors, was discussed as a possible candidate to overcome scaling barriers. The advantages of SOI over current bulk silicon technology in terms of performance and scaling capability were demonstrated. Subsequently, several different SOI devices, namely fully depleted FDSOI and partially depleted PDSOI, and their structural and behavioural differences, were reviewed.

In Chapter 3, a detailed introduction to the statistical techniques such as the Design of Experiments (DoE) and Response Surface Modelling (RSM) was given, followed by an analysis of the impact of process variations on key device parameters such as threshold voltage of both NMOS and PMOS devices, as realised in FDSOI, PDSOI and bulk silicon technologies. These techniques offer a reasonable balance between accuracy and

computational efficiency as compared to the conventional Monte Carlo technique by limiting the complexity of the simulation and modelling process.

The methodology implemented within the TCAD framework for characterizing process variations incorporates three main steps, namely: screening, model building, and model analysis. The methodology uses existing TCAD tools, namely, SPROCESS and SDEVICE, for process and device simulations respectively, and AURORA for compact model parameter extraction. PB screening has been used to characterize and identify the most significant process parameters from an initial set of 18 parameters for the process which would most likely influence the key device electrical characteristics such as threshold voltage for both NMOS and PMOS devices realised in FDSOI, PDSOI and bulk silicon technologies. Second order models for threshold voltage for each technology were then developed as a function of the significant process parameters. The models obtained can be used for optimisation and process control purposes in the semiconductor manufacturing environment. Finally, model validation was performed using a residual-analysis test, and R-squared statistics were generated for all the models developed.

In Chapter 4, the response surface methodology (RSM) was extended to the circuit-level, by studying the impact of process variations and environmental operating conditions on various C-element circuits which are not only widely used but also peculiar to asynchronous design. Subsequently, the response surface analysis for circuit performance metrics such as power and delay was undertaken, giving useful information about the sensitivity of each metric with respect to the process parameters under consideration. The influence of circuit architectures on performance variation in terms of dynamic and static power consumptions and delay of the circuit was also investigated. A comparative study on the behaviour of circuits based on FDSOI, PDSOI, and bulk silicon in terms of delay and power, was subsequently performed.

The behaviour of the various circuit topologies under process variations was investigated by studying their impact on transistor channel currents. The investigation was carried out by measuring the maximum current in each transistor under process variation during the operation of the circuit. Using this method, it was possible to firstly explain why the circuit topology plays a significant role in performance variation; secondly to identify which transistors make the most significant contribution to the final performance

variation; and thirdly to target those transistors in order to minimise their impact on performance variability by resizing the width of the dominant transistors.

The direct impact of the manufacturing process steps on the performance of the standard implementation of the C-element circuit based on the technologies investigated, was undertaken with the aim of determining and identifying the critical process parameters that need to be controlled, and to find ways to reduce the variability in terms of the sources, thereby improving the performance and yield of the circuit.

An overview of different aspects related to the single event upsets (SEUs) and their impact on VLSI circuits was reviewed in Chapter 5, namely the charge deposition and the collection mechanisms involved, and the impact of technology scaling on the radiation tolerance of VLSI circuits. Thereafter, the impact of key design parameters such as threshold voltage, power supply and width ratio on the resilience of various C-element circuits was investigated.

The use of different materials and device structures such as FDSOI and PDSOI technologies to reduce the impact of radiation particle strikes at device level was explored. A comparative study of soft error analysis of different C-element circuits was undertaken.

6.2 Conclusions

From the analysis of the impact of process variations on key device parameters such as threshold voltage of both NMOS and PMOS devices, realised in FDSOI, PDSOI and bulk silicon technologies, it can be concluded that:

- The variation in silicon thickness greatly influence threshold voltage (V_{th0}) for both NMOS and PMOS devices based on FDSOI technology. However, this sensitivity is mostly reduced when using PDSOI technology.

- The threshold adjustment implantation dose and energy, along with gate-oxide thickness and high-k dielectric thickness, were found to be the most influential and common process parameters for most of the threshold voltage responses based on the technologies investigated.

- The MOSFET devices realised in FDSOI technology were found to have the lowest threshold voltage mean values compared to PDSOI and bulk silicon devices, however in terms of the effect of process variations, the FDSOI devices show the highest standard deviations (σ) of V_{th0} for both NMOS and PMOS devices compared to their PDSOI and bulk silicon counterparts.

- It was also found that the variation in PMOS devices is much higher compared NMOS devices for all technology investigated.

The analysis of the effect of process variations on the performance of various C-element implementations realised in FDSOI, PDSOI and bulk silicon technologies, has resulted in the following conclusions.

- The threshold variations in both N and PMOS devices were found to have a significant role in terms of variations in delay and power consumption of a circuit. Assigning higher power supply voltages and using lower threshold voltage devices improves the circuit performance in terms of delay variations. However, trade-offs between the variations in delay and power must be considered, depending upon the application of the circuit.

- The circuit topology has a key impact on both the performance metrics of a circuit and its robustness with regard to process variations. It was found that the dynamic implementation of the C-element is more robust in terms of process variations compared to the other implementations of the C-element in terms of delay and power consumption variations.

- The presence of the weak feedback inverter in the weak feedback, standard and resistive weak feedback implementations of a C-element, degrades the circuit performance and its resilience towards variation, compared to the dynamic C-element implementation. It also was observed that controlling the weak feedback inverter, as in the standard and resistive weak feedback implementations, improves its robustness in terms of the process and environmental variations.

- The differential implementation of the C-element shows much more robustness to process variation in comparison to the static and dynamic implementations in

terms of the performance metrics such as propagation delay and dynamic and static power consumptions of the circuit investigated.

- From the process technology point of view, it was found that all C-element circuits realised in the SOI technology (FDSOI and PDSOI) are more resilient to process variations with higher performance metrics compared to their counterparts based on bulk silicon. It can be concluded that the SOI technology has the potential to reduce the impact of process variations over silicon based circuits.

- The impact of process variations on the delay and dynamic power consumption of the weak feed and standard implementations of a C-element were reduced by using the variability analysis of transistor current method and varying the width of the most influential transistors on the performance variation of a circuit. As an example, In the weak feedback implementation, the mean and standard deviation values of the propagation delay were reduced by 30.2% and 40% respectively, while the dynamic power is reduced in terms of the mean and standard deviation by 23.6% and 30% respectively and the values for the static power consumption were also reduced by 38.6% and 39%.

- It was found that the gate length and the silicon layer thickness variations dominated the performance metrics of circuits based on FDSOI technology in terms of the propagation delay and dynamic power consumption. At the same time, this sensitivity to silicon layer thickness variation in the circuits implemented using PDSOI has been significantly reduced.

- In the case of circuits based on bulk silicon, the performance metrics were significantly influenced by gate length variations and by the halo implantation process steps.

- It was found that the impact of gate length variation on the propagation delay and the dynamic power consumption of the standard implementation of the C-element was commonly significant in terms of process parameters for all the technologies investigated.

Focusing on the results obtained from the analysis of soft errors robustness of various C-element circuits implemented in different process technologies, it can be concluded that:

- The sensitivity of a node in a circuit to a single event upset can be reduced by increasing the drive strength of the restoring devices which, in turn, increases the rate of removal of the charge collected at a node by a particle strike. This can be done by reducing the threshold voltages of the devices and increasing the power supply voltage.

- The pulse width of the induced glitch due a radiation strikes is largely influenced by the power supply and the threshold voltage of the PMOS devices. The pulse width can be reduced by adopting higher power supply voltages and lower threshold voltage devices.

- The analysis of different C-element implementations subsequently permits a relative comparison of Q_{crit} to be undertaken. The symmetric implementation of the C-element is found to be the most robust in terms of soft errors compared to the dynamic and static implementations of the C-element.

- The differential logic plus an inverter latch implementation of the C-element was found to be the most robust with regard to soft errors as they were unaffected by radiation strikes and independent of the amount of charge deposited. This can be used as a possible way to mitigate the impact of soft errors in asynchronous designs.

- The potential of silicon on insulator technology (SOI) in enhancing the resilience of a circuit to soft errors has also been explored. It was observed that circuits based on FDSOI technology show higher tolerance compared to PDSOI and bulk silicon technologies.

6.3 Future Work

As CMOS VLSI technology in the nanometre regime continues to scale aggressively for increased performance and integration density, designing reliable and robust devices and circuits to mitigate the effects of particle strikes and process variations will, therefore, continue to pose a challenge for future generations of technologists. The following section highlights the possible key points for future work in this area.

- The DoE/RSM statistical techniques used in this work seem to be limited and become inefficient when a large number of process parameters are considered together with their interactions. Therefore, new statistical techniques and tools need to be developed to perform the variability analysis of deep sub-micron devices which are particularly vulnerable to process variability.

- The investigation of the impact of process variation on non-planar device structures based on SOI technology such as FinFET and Nanowires as a possible way to circumvent the variability challenge for future technology generations is also a possible research area.

- Furthermore, as device dimensions are reduced, the random variations due to dopant fluctuations and line edge roughness (LER) will have a profound effect on devices and behaviours, and the subsequent yield of manufactured circuits. Therefore, it is essential to be able to analyse the effect of process variations and their impact on various aspects of overall circuit behaviour. However, present day process/device models are too simplistic to accurately predict process/device behaviour in terms of these variations. Consequently new process/device models need to be developed to allow accurate variability prediction, together with the use of 3D rather than 2D simulations of device structures.

- The investigation, is required, of the impact of temporal variability sources such as the history effects in PDSOI technology and self heating, together with the

impact of transistor aging such as negative bias temperature instability (NBTI) which affects the threshold voltage of PMOS devices.

- The investigation, is also required, of the impact of process variation and radiation effects on sub-threshold digital circuits which have emerged as a low energy solution for applications with strict energy constraints.

- The dramatic increase in leakage current combined with the large increase in variability in highly scaled CMOS technologies, pose a major challenge for future IC design. Therefore leakage reduction techniques should be investigated.

Appendix A

The response surface (RS) plots of the threshold voltage (V_{th0}) for both NMOS and PMOS devices realised by FDSOI, PDSOI and bulk silicon technologies.

A.1 RS plots of V_{th0} for NMOS- FDSOI
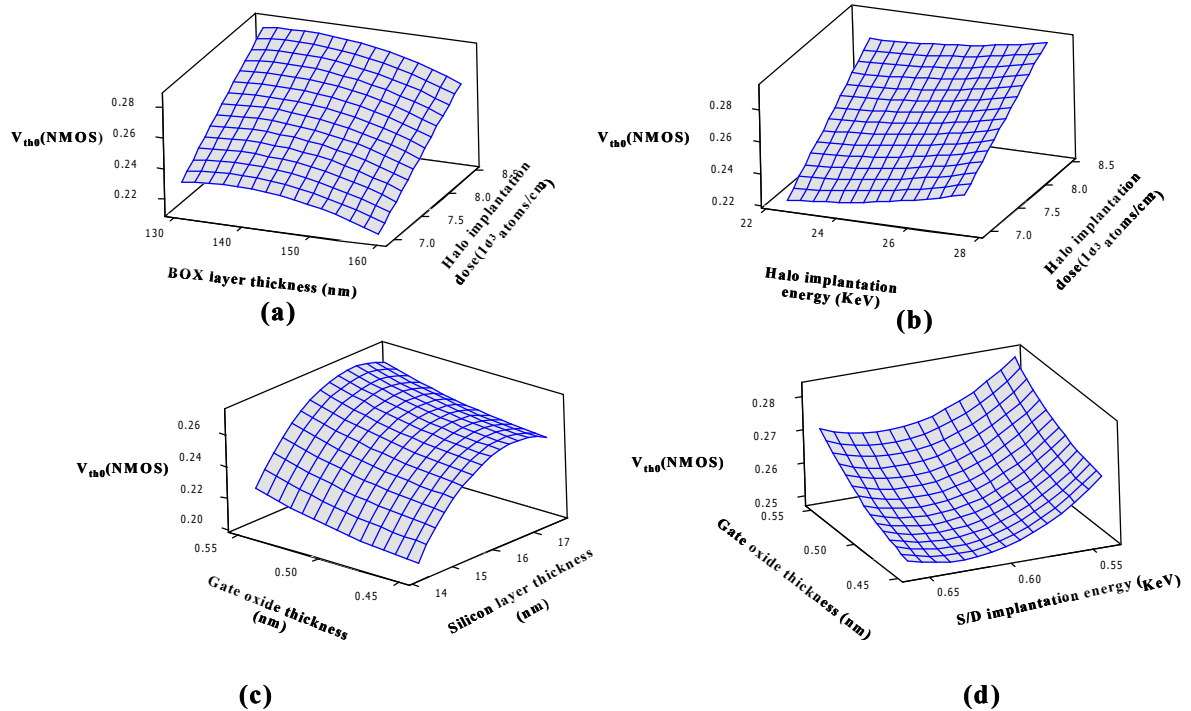
Figure A. 1: Response surface for V_{th0} (FDSOI-NMOS) with respect to (a) halo implantation dose and buried oxide (BOX) layer thickness (b) halo implantation dose and halo implantation energy (c) silicon layer thickness and gate oxide thickness (d) silicon layer thickness and source/drain implantation energy.

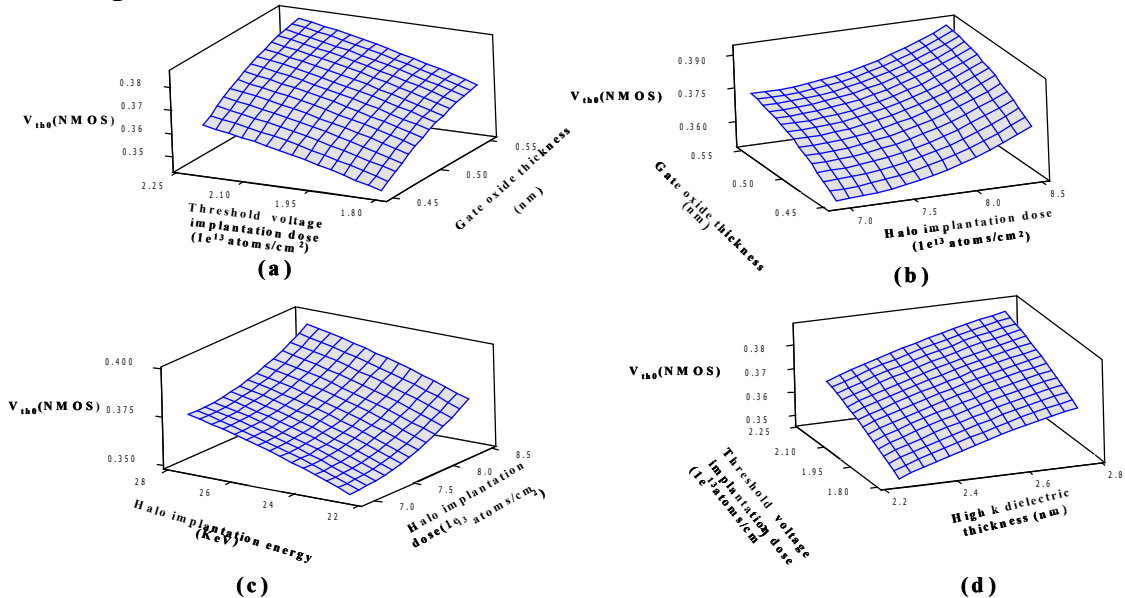## A.2 RS plots of V_{th0} for PMOS- FDSOI
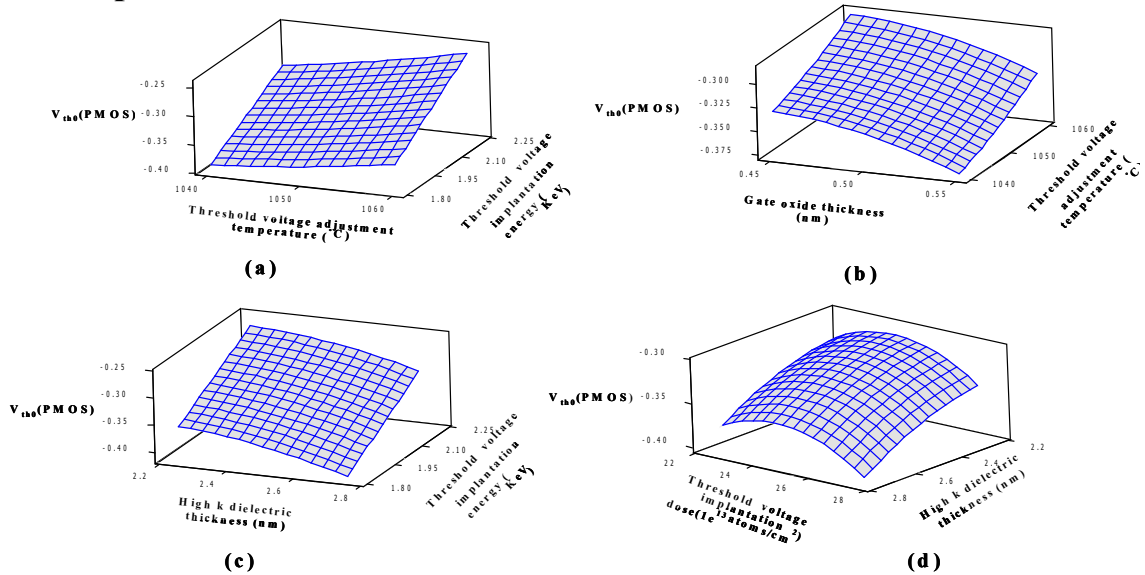

(a)

(b)

(c)

(d)
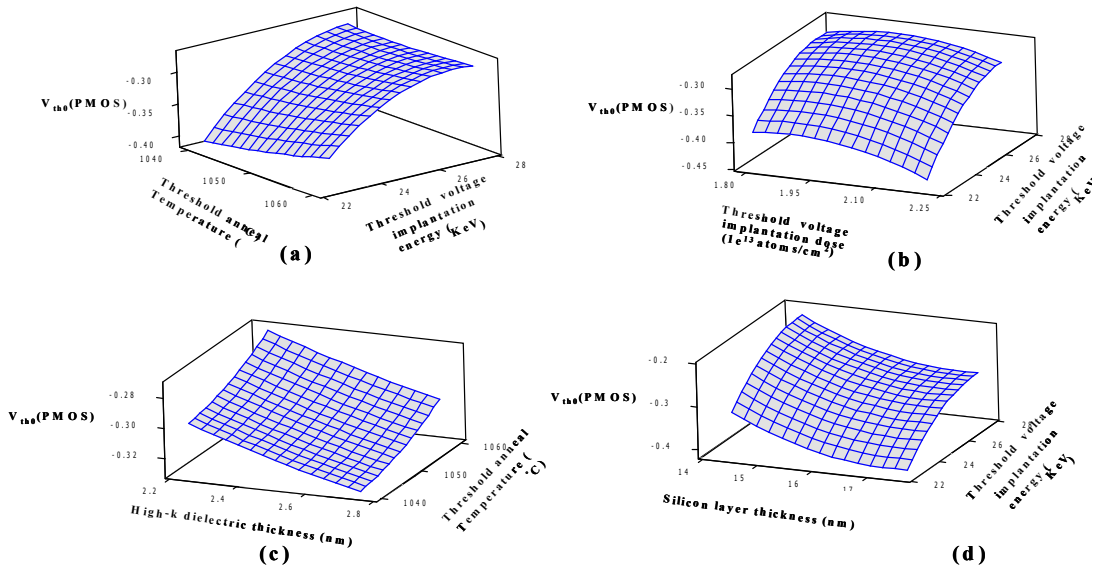
Figure A. 2 : Response surface for V_{th0} (FDSOI-PMOS) with respect to (a) threshold voltage adjustment implantation energy and threshold annealing temperature (b) threshold voltage adjustment implantation energy and threshold voltage adjustment implantation dose (c) gate oxide thickness and threshold annealing temperature (d) threshold voltage adjustment implantation energy and silicon layer thickness.

A.3 RS plots of V_{th0} for NMOS- PDSOI

(a)

(b)

(c)

(d)

Figure A. 3 : Response surface for V_{th0} (PDSOI-NMOS) with respect to (a) threshold voltage adjustment implantation dose and gate-oxide thickness (b) gate-oxide thickness and halo implantation dose (c) halo implantation dose and halo implantation energy (d) threshold voltage adjustment implantation dose and high-k dielectric thickness.

A.4 RS plots of V_{th0} for PMOS- PDSOI

(a)

(b)

(c)

(d)

Figure A. 4 : Response surface for V_{th0} (PDSOI-PMOS) with respect to (a) threshold voltage adjustment implantation energy and threshold annealing temperature (b) threshold voltage adjustment implantation energy and threshold voltage adjustment implantation dose (c) threshold annealing temperature and high-k dielectric thickness (d) silicon layer thickness and threshold voltage adjustment implantation energy.

A.5 RS plots of V_{th0} for NMOS- bulk Silicon

(a)

(b)

(c)

(d)
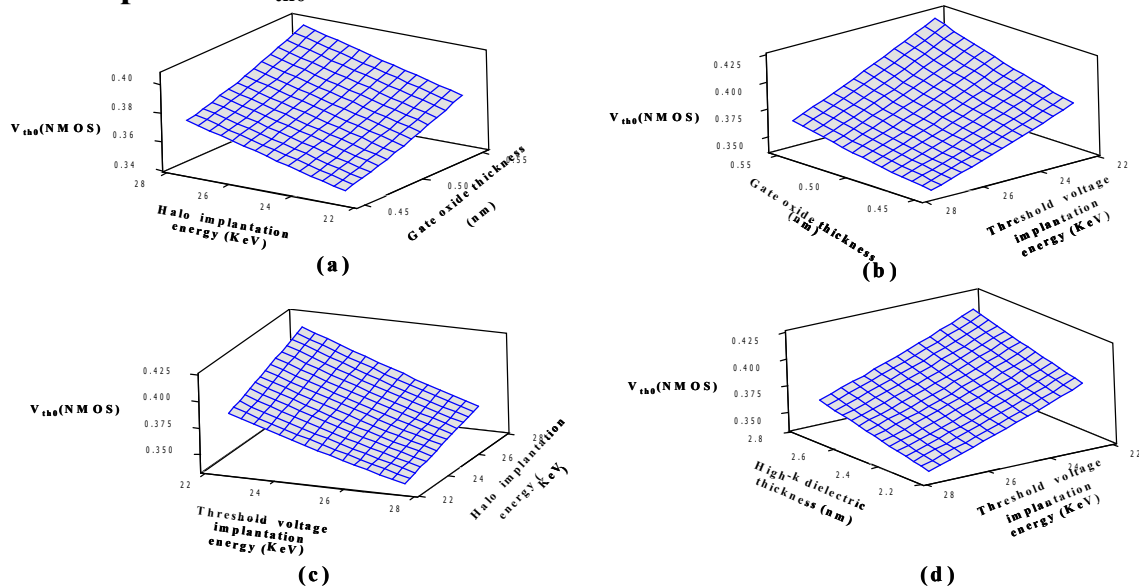
Figure A. 5 : Response surface for V_{th0} (Silicon-NMOS) with respect to (a) gate-oxide thickness and halo implantation energy (b) gate-oxide thickness and threshold voltage adjustment implantation energy (c) threshold voltage adjustment

implantation energy and halo implantation energy (d) threshold voltage adjustment implantation dose and high-k dielectric thickness.

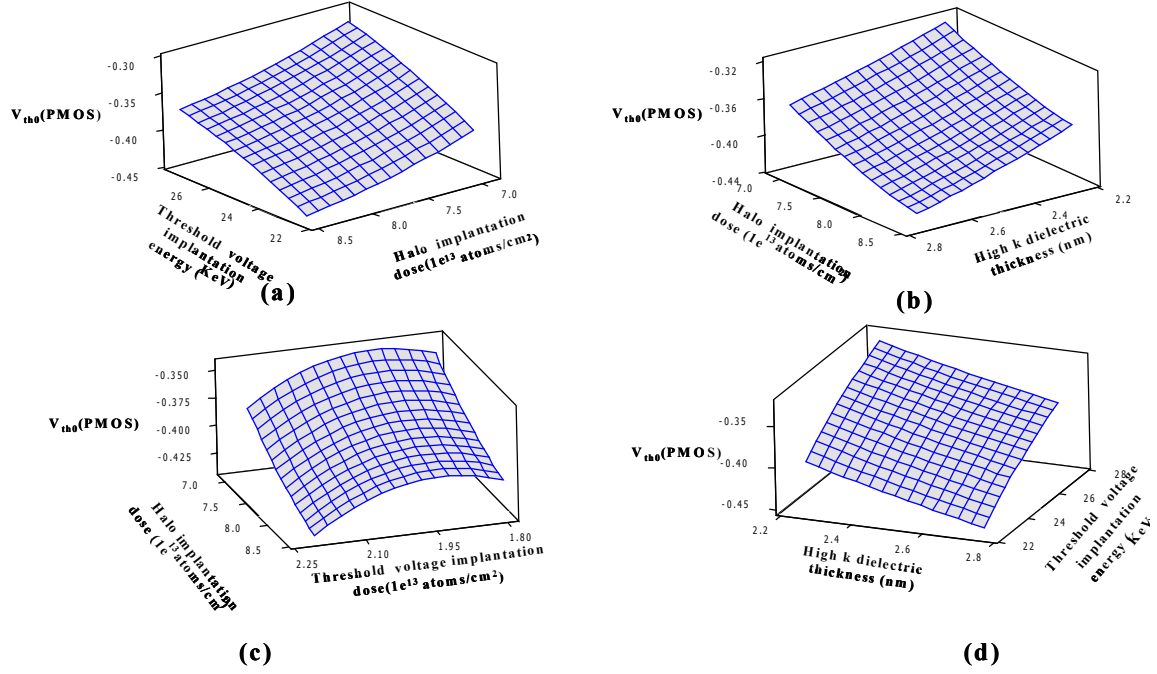## A.6 RS plots of V_{th0} for PMOS- bulk Silicon

(a)

(b)

(c)

(d)

Figure A. 6 : Response surface for V_{th0} (Silicon-PMOS) with respect to(a) threshold voltage adjustment implantation energy and halo implantation dose (b) high-k dielectric thickness and halo implantation dose (c) threshold voltage adjustment implantation dose and halo implantation dose (d) threshold voltage adjustment implantation energy and high-k dielectric thickness.

Appendix B

The response surface (RS) plots key performance metrics (propagation delay, dynamic power consumption and static power consumption) of the standard implementation of C-element realised by FDSOI, PDSOI and bulk silicon technologies.

B.1 RS plots of key performance metrics of the standard implementation of C-element realised by FDSOI technology
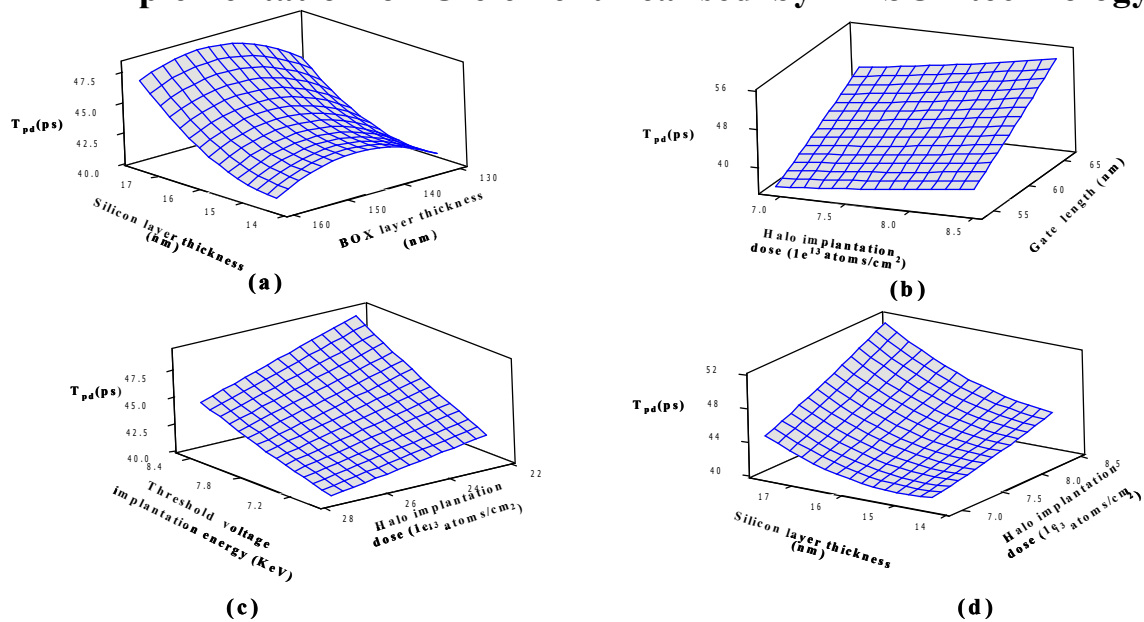

Figure B. 1 : Response surface for propagation delay of the standard C-element realised on FDSOI technology with respect to (a) silicon layer thickness and buried oxide (BOX) layer thickness (b) halo implantation dose and gate length (c) threshold voltage adjustment implantation energy and halo implantation dose (d) halo implantation dose and silicon layer thickness.
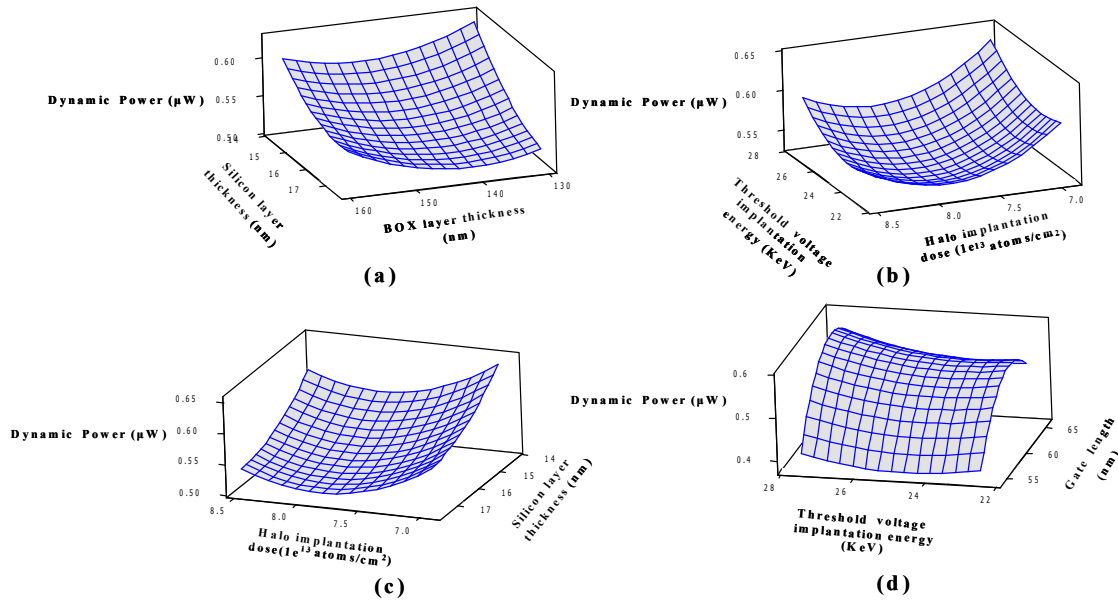
(a)

(b)

(c)

(d)

Figure B. 2 : Response surface plots for dynamic power consumption of the standard C-element realised on FDSOI technology with respect to (a) silicon layer thickness and buried oxide (BOX) layer thickness (b) halo implantation dose and threshold voltage adjustment implantation energy (c) halo implantation dose and silicon layer thickness (d) threshold voltage adjustment implantation energy and gate length.
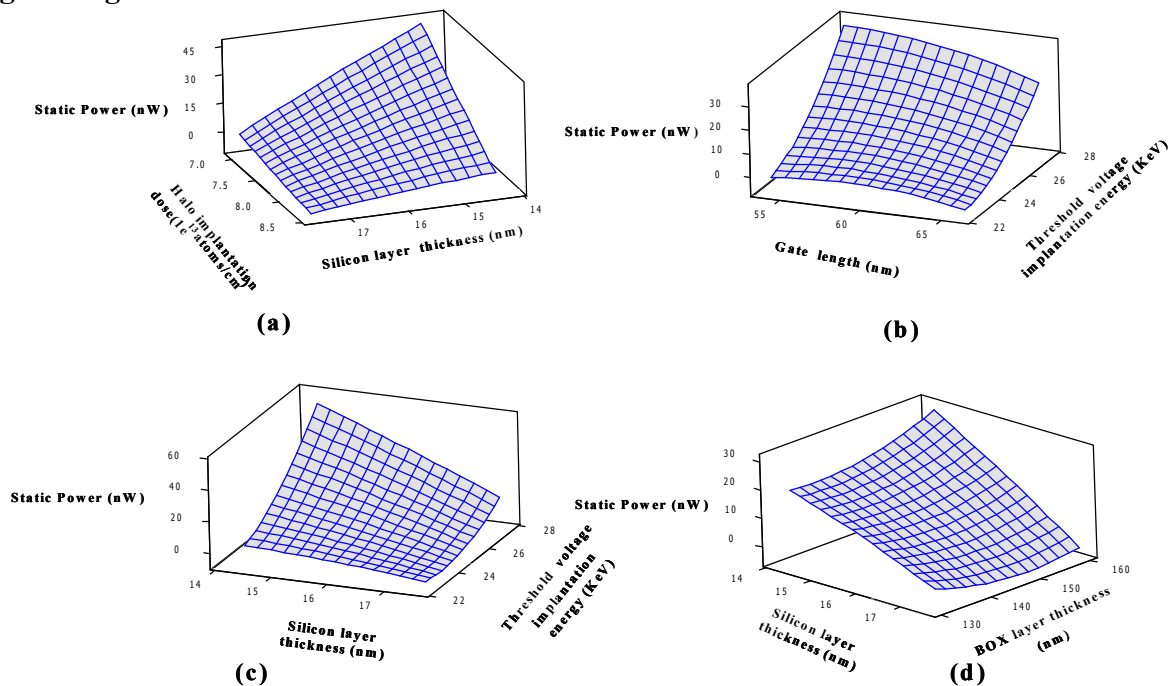

(a)

(b)

(c)

(d)

Figure B. 3 :Response surface plots for static power consumption of the standard C-element realised on FDSOI technology with respect to (a) halo implantation dose and silicon layer thickness (b) threshold voltage adjustment implantation energy and gate length (c) threshold voltage adjustment implantation energy and silicon layer thickness (d) silicon layer thickness and buried oxide (BOX) layer thickness.

B.2 RS plots of key performance metrics of the standard implementation of C-element realised by PDSOI technology
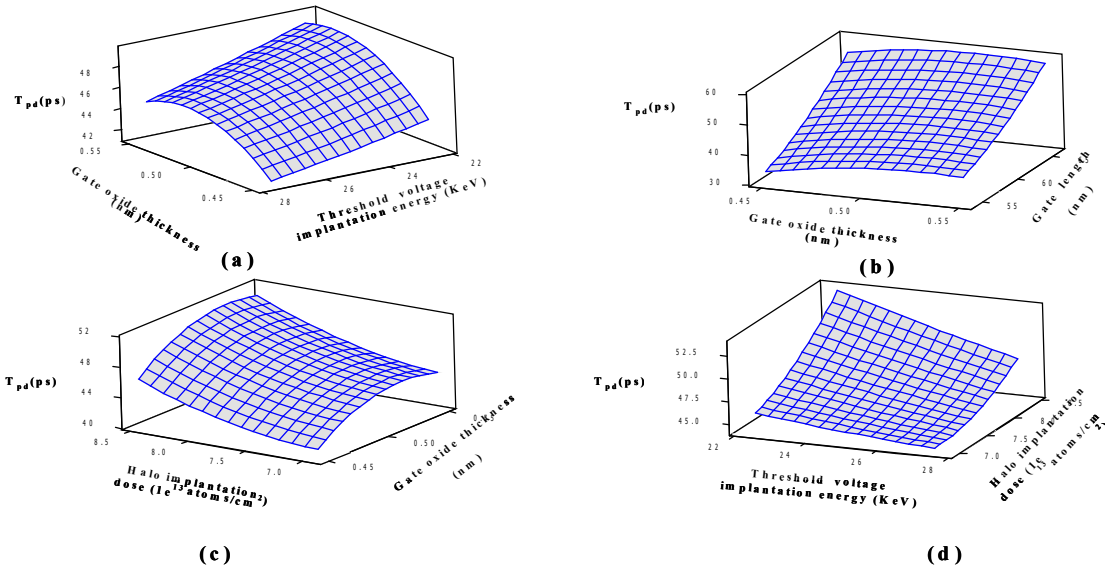

(a)

(b)

(c)

(d)

Figure B. 4 : Response surface plots for propagation delay of the standard C-element realised on PDSOI technology with respect to (a) gate oxide thickness and threshold voltage adjustment implantation energy (b) gate oxide thickness and gate length (c) halo implantation dose and gate oxide thickness (d) threshold voltage adjustment implantation energy and halo implantation dose.
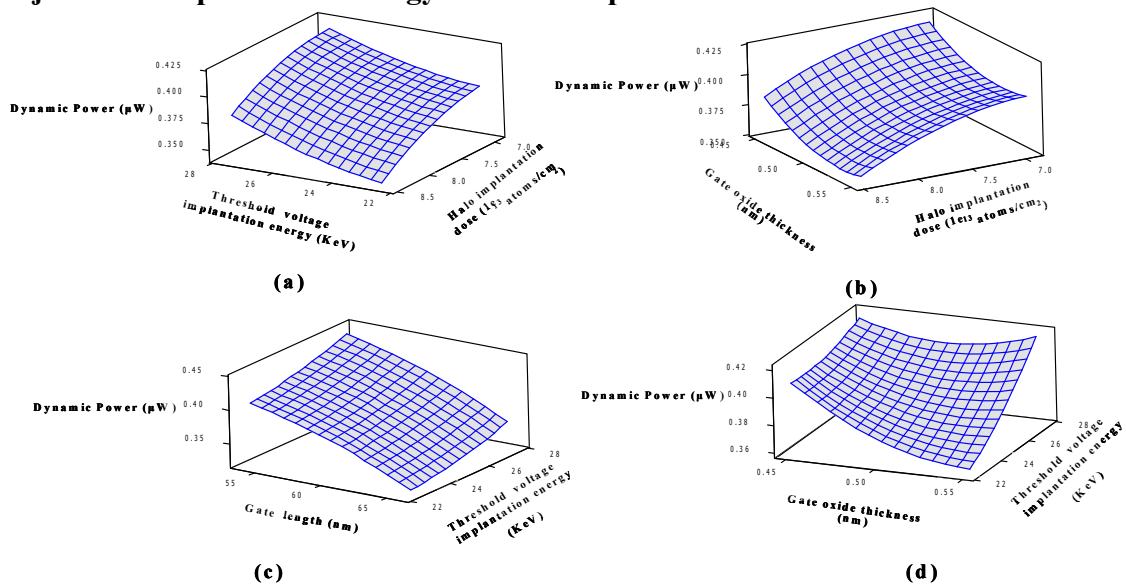

(a)

(b)

(c)

(d)

Figure B. 5 : Response surface plots for dynamic power consumption of the standard C-element realised on PDSOI technology with respect to (a) threshold voltage adjustment implantation energy and halo implantation dose (b) halo implantation dose and gate oxide thickness (c) threshold voltage adjustment implantation energy and gate length (d) gate oxide thickness and threshold voltage adjustment implantation energy.
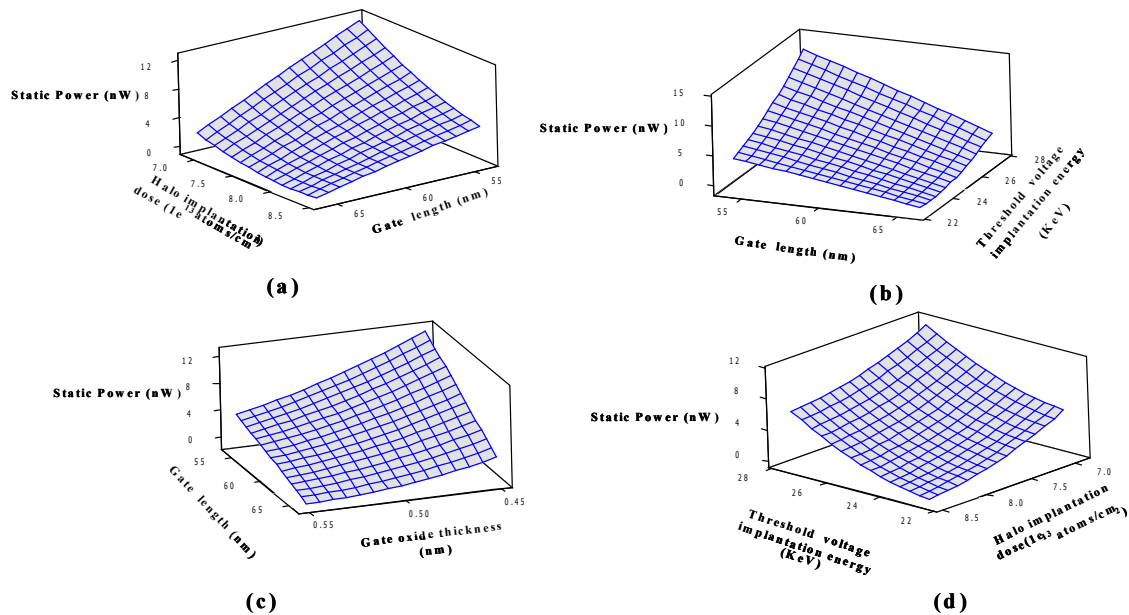
157

(a)

(b)

(c)

(d)

Figure B. 6 :Response surface plots for static power consumption of the standard C-element realised on PDSOI technology with respect to (a) halo implantation dose and gate length (b) threshold voltage adjustment implantation energy and gate length (c) gate oxide thickness and gate length (d) threshold voltage adjustment implantation energy and halo implantation dose.

B.3 RS plots of key performance metrics of the standard implementation of C-element realised by bulk silicon technology
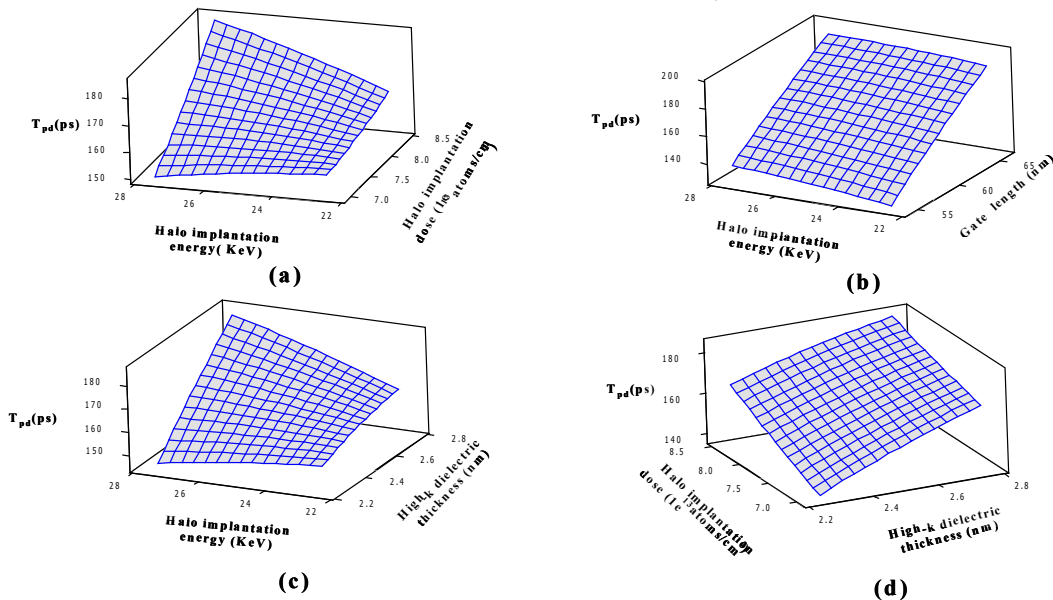

(a)

(b)

(c)

(d)

Figure B. 7 : Response surface plots for propagation delay of the standard C-element realised on bulk silicon technology with respect to (a) halo implantation energy and halo implantation dose (b) halo implantation energy and gate length (c)

halo implantation energy and high k dielectric thickness (d) halo implantation dose and high k dielectric thickness.
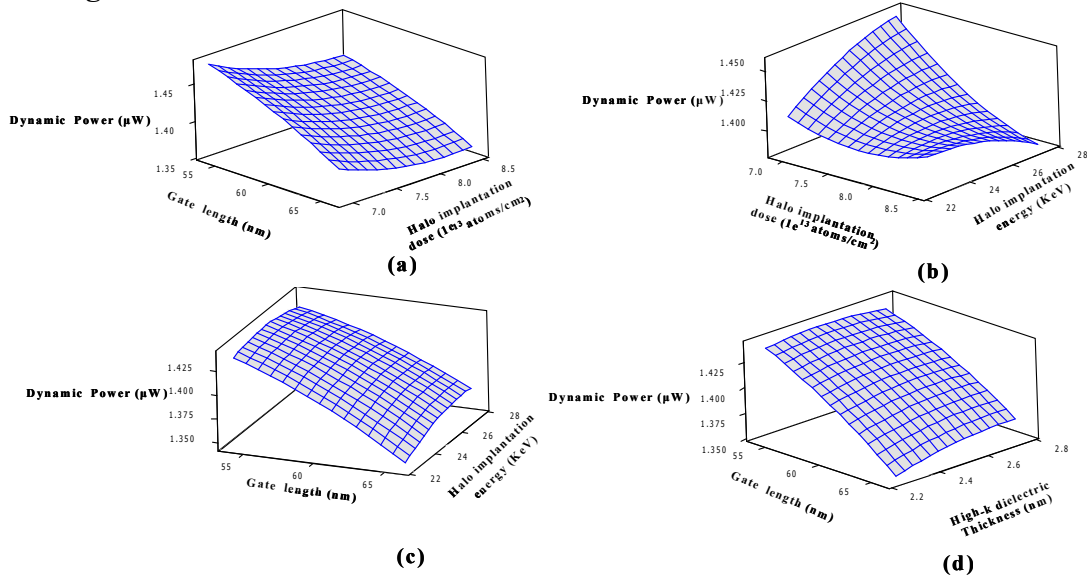

Figure B. 8: Response surface plots for dynamic power consumption of the standard C-element realised on bulk silicon technology with respect to (a) halo implantation energy and gate length (b) halo implantation energy and halo implantation dose (c) halo implantation dose and gate length (d) gate length and high k dielectric thickness.
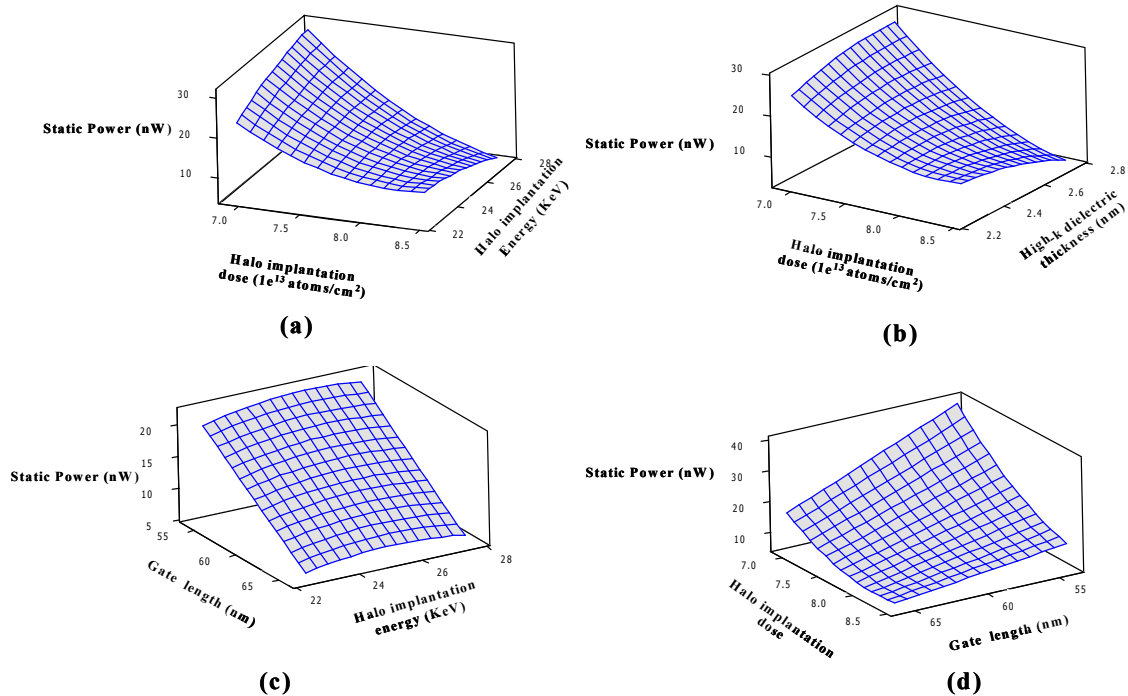

Figure B. 9 : Response surface plots for static power consumption of the standard C-element realised on bulk silicon technology with respect to (a) halo implantation energy and halo implantation dose (b) halo implantation energy and high k dielectric thickness (c) halo implantation energy and gate length (d) halo implantation dose and gate length.