Full-Swing Local Bitline SRAM FinFET Technology: A Review

Shipra Garg[1], Dr. Pragati Sharma[2]
[1]*M. tech Scholar, S.D College of Engineering and Technology, India*
[2]*Professor, S.D College of Engineering and Technology, India*

***Abstract*—The earlier average-8T static RAM (SRAM) design offers a competitive cell area and eliminates the need for a write-back mechanism. In this architecture, a full-swing local bitline (BL) connected to the read buffer's gate can normally be produced by applying a boosted wordline (WL) voltage. However, when implemented in more advanced technologies such as a 22-nm FinFET process, where threshold-voltage variation is significant a boosted WL voltage is not feasible because it negatively affects read stability. As a result, the local BL cannot reach a full swing, and the read buffer's gate is not driven by the full supply voltage (VDD), which leads to a substantial increase in read delay.To address this limitation, this paper introduces a differential SRAM architecture capable of generating a full-swing local BL. This is achieved by incorporating cross-coupled pMOS devices, allowing the read buffer's gate to be driven by the full VDD without relying on WL boosting. Several multi-bit configurations of the proposed architecture are evaluated with respect to minimum operating voltage and area efficiency. The version storing four bits per block demonstrates a minimum operating voltage of 0.42 V and achieves a read delay reduction by a factor of 62.6 compared to the average-8T SRAM fabricated in 22-nm FinFET technology.**

***Index Terms*—Bit-interleaving, FinFET, low-voltage operation, static random-access memory (SRAM).**

I. INTRODUCTION

Recent years, the growing use of battery-powered technologies including portable smart devices and implantable medical equipment has made low-power design a major concern in system-on-chip (SoC) development. Since static random-access memory (SRAM) occupies a substantial portion of an SoC's area, reducing its power consumption is essential for building energy-efficient SoCs. One effective way to reduce power is to lower the supply voltage, as power consumption decreases quadratically with voltage [1]. However, operating at reduced voltages increases the impact of threshold-voltage (Vth) variations, which become a dominant reliability issue. SRAM cells, built using very small transistors to achieve high density, are especially vulnerable to these Vth fluctuations. Additionally, traditional 6T SRAM architectures face an inherent trade-off between read stability and write capability, making it difficult to maintain both reliably when operating in low-voltage conditions. Several SRAM cell designs featuring an independent read port have been introduced to enable reliable operation at reduced supply voltages [2]–[6]. Incorporating a separate read path removes the inherent compromise between read stability and write capability in non–bit-interleaved SRAM arrays. As a result, both characteristics can be tuned independently, which supports dependable functionality under low-voltage conditions. SRAM cells are also vulnerable to soft-error events caused by α-particle strikes. To mitigate these faults, bit-interleaving within the memory array is required [7]. Figure 1 illustrates the structure of a bit-interleaved SRAM array. In such an architecture, "selected cells'' refer to those targeted for read or write operations. "Row half-selected cells'' are the cells on an activated wordline but on columns that are not selected, while "column half-selected cells'' lie on unselected rows but within the chosen column. During a write operation, the row half-selected cells may experience disturbances because their wordline is asserted. Consequently, their stability can be affected. In SRAM design, the stability of row half-selected cells must also be taken into account, a concern commonly known as the half-select problem. Existing solutions mentioned earlier fail to resolve this problem unless a write-back scheme is used. In such a scheme, the data stored in a cell is first read and then rewritten to reinforce its state, but this process adds extra power consumption, delay, and area overhead. To eliminate the half-select problem without relying on write-back, a 10T SRAM cell using a cross-point configuration was introduced [8]. This structure employs both vertical and horizontal wordlines, which must be activated simultaneously to access the stored nodes. Because only the targeted cell has both wordlines enabled during a write operation, half-selected cells remain undisturbed. However, the major drawback of this 10T scheme is its substantial area penalty due to the added transistors. To reduce this overhead, an "average-8T" SRAM designed in 130-nm technology was later proposed; it avoids the write-back requirement, mitigates the half-select problem, and provides a more favorable area profile [9]. Still, when this 8T structure is implemented in advanced nodes such as 22-nm FinFET technology, its read delay grows significantly due to increased Vth variability and the inherent trade-off between read stability and read speed. This paper examines that limitation of the average-8T architecture in scaled technologies and introduces an improved differential SRAM design that resolves the half-select issue without write-back, maintains competitive area efficiency, and provides full-swing local bitlines, resulting in a markedly shorter read delay compared to the average-8T design.

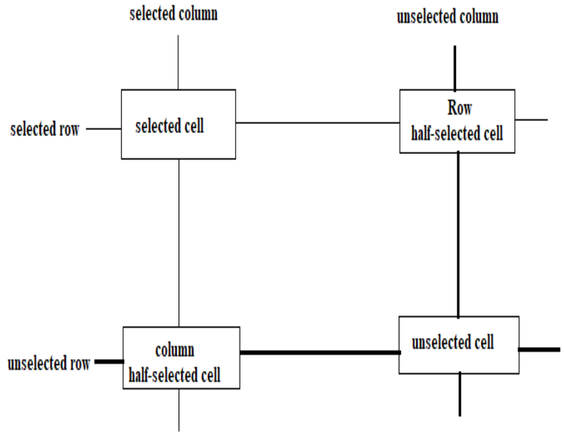

Fig.1. Selected, half-selected, and unselected cells in a bit-interleaved SRAM array

In SRAM design, the stability of row half-selected cells must also be taken into account. This concern is known as the half-select problem. Existing solutions generally fail to eliminate this issue unless a write-back mechanism is used. In a write-back scheme, the data in a cell is first read and then rewritten to reinforce its stored value, but this approach increases power consumption, delay, and area. To avoid relying on write-back, a 10T cross-point SRAM cell was introduced in [8]. This design uses both vertical and horizontal wordlines, and access to the stored nodes occurs only when both are enabled. Since only the targeted cell activates both wordlines during a write operation, the half-select problem is effectively removed. However, the 10T cell incurs a substantial area penalty because of the additional transistors. To reduce this overhead, a compact average-8T SRAM implemented in 130-nm technology was later proposed as an alternative. It suppresses the half-select issue without write-back while keeping the area competitive [9]. Its main limitation appears in advanced nodes such as 22-nm FinFET technology: the significant Vth variations create a strong tradeoff between read stability and read delay, leading to much slower reads. This paper examines that limitation in the average-8T design and introduces an improved SRAM cell architecture that resolves it. The proposed differential SRAM design eliminates the half-select issue without write-back, maintains competitive area efficiency, and uses a full-swing local bitline to achieve a much shorter read delay than the average-8T cell. The rest of the paper is structured as follows: Section II reviews and analyzes the average-8T SRAM. Section III presents the structure and operation of the proposed cell. Section IV compares simulation results of both designs in 22-nm FinFET technology. Section V concludes the work.

II. AVERAGE-8T SRAM ARCHITECTURE

Fig. 2 shows the average-8T SRAM architecture and its operational waveform. A block that stores four bits consists of four pairs of cross-coupled inverters, pass gate transistors (PGL1~4 and PGR1~4), block mask transistors (MASK1 and MASK2), write access transistors (WR1 and WR2), and read buffers (RD1~4). A stacked nMOS structure is used as a read buffer to reduce the read BL (RBL and RBLB) leakage. It is important to note that the block select signal (BLK) and WLs (WL1~4) are row-based signals, whereas the RBLs and write BLs (WBL and WBLB) are column-based signals. During the hold state, the WLs are held at 0 V to isolate the storage node from the local BLs (LBL and LBLB), and BLK is held at supply voltage (VDD) to discharge the LBLs and to turn OFF the read buffers. The WBLs are held at 0 V, and the RBLs are precharged to VDD. For the read operation, BLK of the selected block is forced to remain at 0 V to turn OFF the block mask transistors, and the selected WL is enabled to turn ON the pass gate transistors. Thus, the stored data in the selected cell are transferred to the LBLs through the pass gate transistors, and one of RBLs is discharged on the basis of the stored data. In the column halfselected block, which is located on the selected column and the unselected row, during the read operation, the WLs are held at 0 V to turn OFF the pass gate transistors, and the BLK is forced to VDD to turn ON the block mask transistors. In the column half-selected block, which is located on the selected column and the unselected row, during the read operation, the WLs are held at 0 V to turn OFF the pass gate transistors, and the BLK is forced to VDD to turn ON the block mask transistors. In the column half-selected block, the two stacked nMOSs in the read buffer are turned OFF, irrespective of the stored data, because the LBLs are discharged to 0 V by the block mask transistors. Thus, the RBL leakage is said to be data-independent. This property of the RBL leakage facilitates the existence of a large number of cells per RBL.

During the read operation, the 1 storage node is disturbed because it is connected to the predischarged LBL via the pass gate transistor. However, a sufficiently high read stability for a low-voltage operation can be achieved because the LBL exhibits a small capacitance. It is important to accurately control the signal timing to achieve robust read stability. If both the WL and the BLK are high, simultaneously, the 1 storage node and the source of the block mask transistor (VSS) will be connected, and the stored data can be flipped. Thus, the WL should rise after the fall of the BLK is completed. For the write operation, the BLK of the selected block is forced to remain at 0 V, the selected WL is enabled, and one of the WBLs is forced to switch to VDD to turn ON the write access transistor on the basis of the write data.

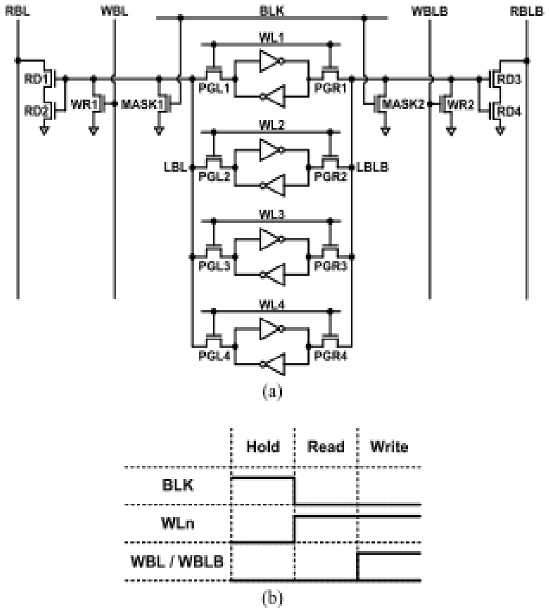

Fig. 2. (a) Average-8T SRAM architecture and (b) its operational waveform.

While the read operation is a differential operation, the write operation is a single-ended operation that discharges only the 1 storage node through the pass gate and write access transistors. It is important to note that during the write operation, the row half-selected block is in the same condition as it was in the read operation. The stability of the row half-selected block is ensured owing to the fact that the LBL exhibits a small capacitance, eliminating the need for a write-back scheme. Furthermore, the area overhead is reduced because the additional transistors such as the block mask and the write access transistors, as well as the read buffers, are shared among the four cells. However, the average-8T SRAM suffers from a drawback associated with the read stability and read delay. During the read operation, the stored 1 cannot be completely delivered to the LBL owing to the Vth drop in the nMOS pass gate transistor. Thus, the gate of the read buffer is driven by VDD–Vth, which results in a considerably large read delay in a low-voltage region. The average-8T SRAM architecture, which was proposed based on the 130-nm technology, alleviates this drawback by boosting the WL voltage, causing the read stability to degrade. Despite the fact that the read stability is degraded by the boosted WL voltage, a sufficiently high read stability is ensured owing to the small capacitance at the LBL. Thus, both a high read stability and a small read delay are ensured in the case of the 130-nm technology. However, in the case of an advanced technology, where the variation in Vth is large, it is difficult to achieve a sufficiently high read stability with the boosted WL voltage in spite of the small capacitance at the LBL. Instead, the suppressed WL voltage is required to achieve a sufficiently high read stability in the case of the advanced technology. However, the suppressed WL voltage intensifies the effect of the Vth drop and considerably worsens the read delay. In light of this discussion, it can be concluded that the average-8T SRAM architecture exhibits a tradeoff between the read stability and the read delay, and it cannot simultaneously ensure a high read stability and a small read delay in the low-voltage region when it is fabricated using an advanced technology.

III. PROPOSED DIFFERENTIAL SRAM ARCHITECTURE

The proposed differential SRAM stores multiple bits in one block, as in the case of an average-8T SRAM. Fig. 3 shows the architecture of the proposed SRAM that stores i bits in one block. The minimum operating voltage and area per bit of the proposed SRAM depend on the number of bits in one block. A configuration that stores four bits in one block is selected as the basic configuration by considering the balance between the minimum operating voltage and the area per bit, which will be described in Section IV. The basic configuration of the proposed SRAM includes four cross-coupled inverter pairs, pass gate transistors (PGL1~4 and PGR1~4), block mask transistors (MASK1 and MASK2), write access transistors (WR1 and WR2), read buffers (RD1 and RD2), a head switch (P1), and cross-coupled pMOSs (P2 and P3). The head switch and cross-coupled pMOSs of the proposed SRAM are notable differences from the average-8T SRAM. WLs (WL1~4), the block select signal (BLK), and the read WL (RWLB) are row-based signals, whereas the write WL (WWL), write BLs (WBL and WBLB), and read BLs (RBL and RBLB) are column-based signals. During the hold state, WLs, WWL, and WBLs are held at 0 V. BLK is held at VDD to connect the WBLs and the LBLs, so that the LBLs are discharged to 0 V and the read buffers are turned OFF. Further, the RWLB is also held at VDD to turn OFF the head switch and to eliminate the RBL leakage current. A. Read Operation The read operation of the proposed SRAM architecture is described in Fig. 4(a). This operation is performed in two phases. During the first phase, BLK of the selected block is forced to remain at 0 V, and the selected WL is enabled. On the basis of the stored data, although the voltage of the LBL that is connected to the 1 storage node becomes high, its value cannot be as high as that of the full VDD because of the Vth drop through the pass gate transistor, and the voltage of the other LBL remains low. The read operation in the first phase is similar to that of the average-8T SRAM, except that the RBL is not discharged because the RWLB is high in the first phase. With the assertion of WL, although the 1 storage node is disturbed, the read disturbance is small because of the small capacitance at the LBL. This smaller read disturbance makes the proposed SRAM be able to operate in significantly lower operating voltage compared with 6T SRAM cell. The second phase starts with the falling of the RWLB. The assertion of the RWLB enables not only the discharge of the RBL but also the feedback of cross-coupled pMOSs. Positive feedback of the cross-coupled pMOSs increases the LBL to the value of the full VDD, owing to which the LBL can achieve a full swing, and the gate of the read buffer.

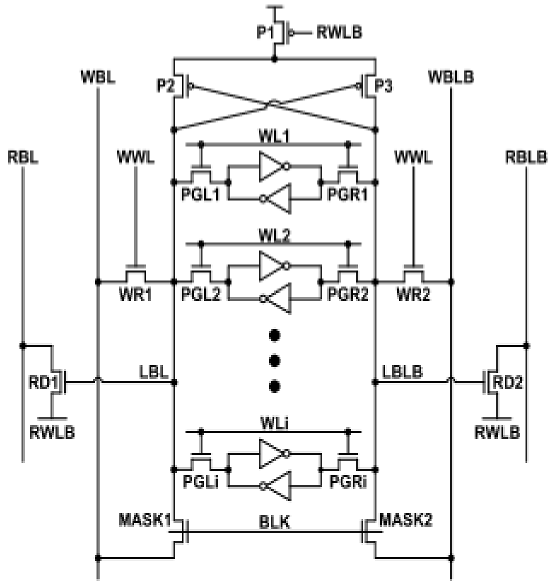

Fig. 3. Proposed SRAM architecture that stores i bits in one block.

IV. REVIEW OF PREVIOUS WORK

The intention of this paper is to reduce power and area of 2:1 multiplexer (MUX) while maintaining the competitive performance. The various configurations are designed using different topology of 2:1 MUX such as CMOS-based MUX, transmission gate and pass transistor using fin-shaped field effect transistor (FINFET). The mobility was enhanced in devices with taller fins due to increase tensile stress. In DG, FINFET can be efficiently used todevelop performance and reduce power consumption. In noncritical paths self-determining gate control can be used to join together parallel transistors. We have estimated the optimum power, optimum current, leakage power, leakage current, operating power, operating current and delay in voltage supply 0.7 V at different temperature such as 10C, 27C and 50C, respectively. A 20 ns access time and frequency 0.5 GHz provide 45 nm CMOS process technology with 0.7 V power supply is employed to carry out different topology of 2:1 MUX using FINFET.[2]

Double-gate FinFET is a novel device structure that is used in the nanometre regime, whereas the regular CMOS technology's performance decreases because of excess short channel effects (SCEs). Double-gate (DG) FinFETs has much better SCEs performance in comparison to the regular CMOS and stimulates technology scaling. In this paper, we are designing 32nanometer DGFinFETs and observing their characteristics by using Sentaurus TCAD, Simulated results of the device shows that it can be governed at the nanometer - scale regime. DGFinFET has independent gates; threshold voltage of one gate can be controlled by varying the voltage at the other gate. By using this phenomenon, logic circuit can be configured in one of the following modes such as SG mode, LP mode, IG mode and IG/LP mode. INVERTER and NAND gate are constructed in the above mentioned node and comparison has been drawn between them. Based on the simulated results, SG-mode is more than sufficient for high performance design.[3]

Static Random-Access Memory (SRAM) plays a pivotal role in modern integrated circuits, serving as a critical component for data storage and fast data access. This review paper provides a comprehensive examination of SRAM cells, focusing on their implementation in two prominent semiconductor technologies: Complementary Metal-Oxide-Semiconductor (CMOS) and FinFET. The performance parameter analysis of the various SRAM design configurations i.e., 6T, 8T and 10T SRAM and challenges encountered in CMOS technology, highlighting key parameters such as speed, power consumption, and stability. Radiation hardness is a critical attribute for SRAM (Static Random-Access Memory) in mission-critical and radiation-prone applications due to the potential adverse effects of ionizing radiation on electronic devices like Single Event Upset (SEU) error occurring in SRAM. This research lays emphasis on the selection of an optimal SRAM cell design, whether 6T, 8T, or 10T in CMOS or FinFET technology depends on the specific application requirements, balancing tradeoffs between area, efficiency, performance metrics, and resilience to radiation-induced faults. It also establishes the superiority of FinFET technology compared to the CMOS technology which suffers from Short Channel Effect at lower technology nodes.[4]

As this technology is as scaled down, the need of important of leakage current and the power analysis for the memory design is exceeding. We consider a mechanism for improving FinFETs efficiency, called variable-supply voltage schemes. It is well known that leakage savings using transistor stacks is not effective in double-gate technologies such as FinFETs, due to the absence of body effect. However, transistor stacking along with variable supply voltage operation of FinFETs can offer larger leakage savings compared to that of bulk devices. the FinFET based 4x4 size SRAM cell array this means of as one bit 7T SRAM. The FinFET on the basis of 7T SRAM has been as designed and the analysis have been as carried out for the leakage current, the dynamic power and the delay. After there, 2:4 decoder has been designed and results obtained through proposed model have been verified. For the validation of our design approach, output of FinFET SRAM array have been compared with standard CMOS SRAM and significant improvements are obtained in proposed model.[5]

MOS is one of the old transistor technologies that provides low power consumption, but also has a shorter channel for the flow of current and thereby has some drawbacks like unnecessary current usage and bigger size. The FinFET (Fin Field Effective Transistor) is an upcoming technology that has a longer channel gate. The size ranges from 32nm, 22nm by Intel at first and then it has been shrunk to 14nm by Samsung. This project is about the study of these two technologies and how we are able to distinguish these techniques and make CMOS as worth as FinFET.[6]

Circuit design of 2-input re-configurable dynamic logics that is based on double gate MOSFETs with the entire set of 16 functions which has been newly explained. 16 function 12T DRDLC with two states (+V, 0) of control gate voltages and 14T DRDLC with two states (0, -V) of control gate voltages have been newly introduced. From these two case which are state control gate, 12T DRDLC with three states (+V, 0, -V) of control gate voltages is successfully obtained. Newly proposed DG-MOSFET with two states (+V, 0), (0, -V) cases are often successfully realized using presently available OR type, AND type IDG-MOSFET, respectively. Newly proposed circuit design, especially for 16 functions 12T DRDLC with two states (+V, 0) case maybe a promising candidates for realizing future re-configurable LSIs.[7]

The FinFET, which we use in a double-gate field effect transistor (the DGFET), it is more versatile with comparison to traditional single-gate field effect transistors the reason is it has 2 gates which can be controlled as independently. Generally, the 2nd gate of FinFET, the transistors is used to dynamically control as the threshold voltage having first gate in the order to improve the performance and as well as reduce the leakage power. After this, we can utilize also the FinFET's 2nd gate for the implement circuits with lesser no. of transistors. There is importanted since the area efficiency is 1 of the main concerns for this circuit design. In the Presented paper, the novel scheme for implementing the majority gate and the 2-1 MUX with the help of both these gates of FinFET transistors as the inputs is shown . Simulation results shows that FinFET logic implementation has a significant advantage over static CMOS logic and pass transistor logic in terms of power consumption and cell area. [8]

Fin Field-Effect Transistor (FinFET) architecture becomes popular in order to develop a Static Random-Access Memory (SRAM) cells with high performance as well as low power consumption with the semiconductor advancement. A comprehensive analysis of 6-transistor (6T), 7-transistor (7T), and 8-transistor (8T) SRAM cells, are developed using FinFET technology at the 7nm node, is presented in this research. The approach compares and assesses the performance based on various critical parameters such as read/write delay, power consumption, and leakage currents at nominal as well as varying environmental conditions. FinFET uses its intrinsic advantages of lower leakage and better electrostatic control to overcome the challenges pertaining to stability and power dissipation in nanoscale SRAM designs. It is further demonstrated that the 7T and 8T SRAM cells exhibit superior stability and less interference from read and write noise than 6T using thorough simulation with industry-standard EDA tools.[9]

This research implements a 7nm FinFET-based 7T SRAM cell, designed using Xschem and simulated with Ngspice. Compared to the conventional 7nm 6T SRAM, the 7T cell achieves reductions of 9.20% in read power, 7.72% in write power, and a substantial 32.46% decrease in leakage power. Additionally, read and write delays are reduced by 10.23% and 4.21%, respectively. When benchmarked against the older 130nm 6T SRAM, the 7nm 7T design demonstrates up to 44.00% lower leakage power and over 16% reduction in dynamic power metrics, albeit with a slight trade-off in write delay. These results underscore the energy efficiency and speed advantages of FinFET-based SRAM, establishing it as a promising candidate for next-generation memory applications. Future research can explore further optimizations to enhance performance under varying operating conditions.[10]

V. CONCLUSIONS

The average-8T SRAM architecture offers the benefits of bit-interleaving without the need for a write-back scheme and maintains a competitive area. However, when implemented in advanced technologies such as 22-nm FinFET, it suffers significant limitations. The read stability–read delay tradeoff prevents full-swing LBL operation, causing the read buffer gate to fall short of full VDD and resulting in substantial read delays, especially at low voltages. Moreover, unnecessary RBL discharge during write operations leads to excessive dynamic power consumption. In contrast, the proposed differential SRAM effectively addresses these shortcomings. By employing cross-coupled pMOS devices, it achieves full-swing LBL while using a suppressed WL read-assist technique to ensure strong read stability. The single-nMOS read buffer further reduces read delay, and the inclusion of a buffer-foot structure prevents unnecessary RBL discharge, improving write-energy efficiency. Overall, the proposed 22-nm FinFET-based SRAM demonstrates clear advantages: significantly reduced read delay, lower energy consumption, and even a slightly smaller area compared with the conventional average-8T SRAM design.

REFERENCES

[1] Kyoman Kang et al. , "Full-Swing Local Bitline SRAM Architecture Based on the 22-nm FinFET Technology for Low-Voltage Operation" 2016 IEEE

[2] Khushboo mishra et al., "Design Different Topology For Reduction Of Low Power 2:1 Multiplexer Using Finfet In Nanometer Technologies" International Journal of Nanoscience Vol. 12, No. 4 2013

[3] V Narendar et al., "Design of High-performance Digital Logic Circuits based on FinFET Technology" International Journal of Computer Applications (0975 – 8887) Volume 41– No.20, March 2012

[4] Monika Sharma et al,"Comparative Analysis 6T, 8T and 10T SRAM in CMOS and FinFET Technologies: A Review"2024 IEEE.

[5] Saurabh khandelwal et al., "Leakage Current And Dynamic Power Analysis Of Finfet Based 7t Sram At 45nm Technology" The International Arab Conference on Information Technology ACIT'2013

[6] R.Rajprabu et al., "Battery Performance Analysis of CMOS and FinFET Logic" IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 2, Issue 1 Mar. – Apr. 2013

[7] Junki Kato et al., "Circuit Design of 2-Input Reconfigurable Dynamic Logic Based on Double Gate MOSFETs with Whole Set of 16 Functions" Contemporary Engineering Sciences, Vol. 7, 2014

[8] Michael C. Wang et al., "Low Power, Area Efficient FinFET Circuit Design" Proceedings of the World Congress on Engineering and Computer Science Vol I 2009

[9] Neerajaksha Monish et al,"Analysis of 6T, 7T, 8T SRAM Cells Using FinFET at 7nm Technology" 2025 IEEE.

[10] K. meenakshi et al,"Enhanced Performance SRAM Design using 7nm FinFET Technology: Optimization of Propagation Delay and Power Consumption"2025 IEEE.