JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Study and Analysis Resilient CMOS 6T SRAM using AI Application

Sajal Agrawal
Department of electronics and communication
Shriram group of college
Bamore (MP)

Ashish Dubey
Assistant professor
Branch - vlsi
Shriram College of Engineering & Management
Bamore (MP)

Abstract Today's automated processes and autonomous systems are only possible because to the advancements in artificial intelligence (AI). The 6T-SRAM cell, which has a performance advantage over its competitors, is a promising option for use in AI hardware implementations. However, state-of-the-art AI hardware like NNs often accesses off-chip data, which slows down the whole system. These days, speed is of the essence for any electronic system, and eighty per cent of all chips have some kind of memory. The size of memory chips, their speed, their leakage current, and their power efficiency have all decreased because to the widespread adoption of SRAM in numerous VLSI Chips. It has gained popularity as a data storage device because to its high storage density & fast access time. Recent rapid progress in low power & low voltage memory technology has prompted the research community to prioritise SRAM. The functionality and efficiency of SRAM & DRAM cells were examined. In this study, we present our work on enhancing the power efficiency of the 6T SRAM cell and its subsequent implementation.

Key Words: CMOS, D-Flip Flop, Leakage Power, Delay, Cadence.

I. INTRODUCTION

Deep neural networks (DNNs) are used in contemporary artificial intelligence (AI) to facilitate efficient and independent tasks. Large volumes of data are needed for the training and inference phases of data-centric DNNs. As a result, the flow of information between the system's memory and its processing units has slowed dramatically. To guarantee the efficiency of AI platforms, investigators have suggested the compute-in-memory idea as a means of avoiding data transportation

bottlenecks. Today's edge computing devices may produce terabytes of data per second. Large-scale data inference often consumes a great deal of time and computing power, both on- and off-chip, in conventional processing units. In cloud computing, CIM reduces the requirement for off-chip traffic and data mobility.

SRAM stands for "static random-access memory." Electronic switches are used to store information in this kind of semiconductor memory. When compared to Dynamic RAM (DRAM), which must be refreshed constantly to maintain the data, Static RAM (SRAM) is quicker and uses less power since it does not need refreshing [1]. SRAM is often utilised as cache memory due to its quick read/write access times. And it's used in all sorts of [2-4] things, including microcontrollers and networking gear. An SRAM chip's memory cells hold a binary digit as long as the power is on. After storing a bit, a flip-flop will keep that value until the opposite value is held. SRAM is small, yet it has fast data access times. The on-chip memory options additionally include SRAM [5][11]. Both SRAM and DRAM have their advantages. However, with access times as low as 10 ns, SRAM often outperforms DRAM. SRAM also doesn't need to be refreshed as often as DRAM. SRAM operates on a low, steady current and consumes less power than DRAM [6][15]. SRAM's drawbacks include a higher price tag and a larger footprint on the chip. Each chip has less storage space, and manufacturing them is more labour intensive. Since SRAM's power consumption is proportional to the frequency with which it is accessed, it consumes almost little power while it is not being actively used. Although SRAM and DRAM both consume the same amount of power at greater rates, SRAM does it at a lower cost.[7][13]. DRAM, is a kind of semiconductor memory. The computer's CPU commonly makes use of DRAM to store the software it runs. However, DRAM is not just found in desktop PCs and servers, but also in workstations and servers. RAM allows the central processing unit of a computer to bypass the need to go through memory sequentially and access data immediately. Data stored in RAM may be accessed much more quickly than data stored on a hard drive. RAM, being physically closer to the heart of the machine, is often faster.[8][16].

As transistor technology has progressed, so too has the efficiency of 6T-SRAM cells. The 6T-SRAM cell has gone through a significant transformation over the last two decades, shifting from a planar transistor to a 3D-FinFET

structure in order to boost design density, power, & performance. The current and future SRAM density trends are shown in Figure 1. The 6T-SRAM cell has several difficulties at current technological nodes, including low-voltage operation dependability, leakage current, soft mistakes, the need for security-aware design, and half-select issues. These problems are becoming more severe in FinFET 6T-SRAM cells. The risk of a soft mistake has increased as a result of the decreased geometrical dimensions, and the leakage current is now equal to the on-current. This is due to the fact that at today's technological nodes, reducing the supply voltage to conserve power actually endangers the integrity of the cell. Modifications to the cell-level design of FinFET 6T-SRAM are required to prevent data theft through side-channel assaults. Additionally, with a higher SRAM density architecture, only cells in a particular row or column may have their values altered. There has to be extensive research into the design of FinFET 6T-SRAM cells to provide consistent performance in CIM solutions for AI applications.

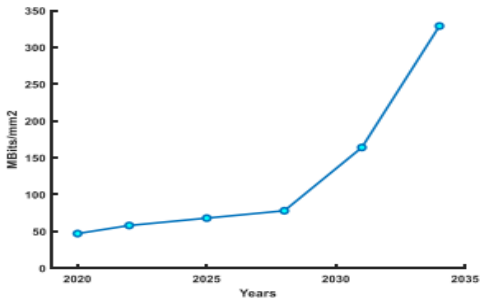


Figure 1. Guidelines for increasing SRAM density by IRDS 2022 [22].

II. Related Work

A. *Standard Architecture SRAM Cell*

Most modern microprocessors include on-chip caches to help systems run faster by reducing the time it takes to transfer data between the CPU and main memory. In order to achieve the highest possible throughput, these on-chip caches are often implemented as arrays of tightly spaced SRAM cells. A common kind of memory cell is a 6T SRAM cell. However, due to a cell size that is 1,000 times larger than a DRAM cell, the 6T cell produces memory with a very low density. In scaled CMOS technology, the leakage from SRAM has grown as a percentage of the total leakage due to the fact that the total current flowing through a chip is proportional to the number of transistors on the chip and there are a lot of transistors in a 6T SRAM. Cache memory based on six transistor SRAM cell is thus being challenged by the growing need for additional memory in nano-scaled CMOS technology. Our goal is to design a 4T SRAM cell to meet this need by decreasing cell area size & leakage current.

An SRAM is a circuit that can store and retrieve information from a location specified by an address code and accessed at random. There are three primary components of an SRAM that work together to do this job. First, the real data is stored in a cell matrix as words of a fixed length. The cells use a positive feedback loop made up of two inverters to store the data. With this iteration, data is permanently stored, eliminating the need for frequent data updates. These cells will be accessible by the sense amplifiers through bit lines. The second section consists of the sensory enhancers. They are used to convert the cells' low swing bit line signals to higher-level digital ones. The decoders are the third major component, and they translate the address into a location in the cell matrix. It goes without saying that other circuits, including a timing and control section, a pre-charge and write system, and so on, are required. When an SRAM is activated, each component must be turned on at precisely the right moment. It has a significant impact on the SRAM's performance, latency, and power consumption. Turning on a word line too soon will result in accessing the erroneous URL. Turning it on with too much of a buffer is bad for the overall SRAM's latency. Particularly important is the time of sense-amplifier activation. When triggered prematurely, sense-amplifiers may provide spurious readings. An activation that occurs too late will incur costs in terms of both power consumption and delay, since the cells need enough time to generate a voltage differential on the bit lines large enough to compensate for the mismatch in the input stage of the sense-amplifiers. As a result, you'll need to put more charge into the bit line than usual in order to get it back up to the proper pre-charge voltage and get a valid reading.

SRAM circuits must be designed with the tiniest possible transistors, making them especially vulnerable to manufacturing flaws. An crucial component of any SRAM design is striking a balance between the trade-offs of compact space, low power, and rapid reads/writes. More specifically, SRAM design calls for a delicate dance between a wide variety of design criteria, such as minimising cell area by using a smaller transistor, maintaining read/write stability, minimising power consumption by reducing supply of power, minimising power consumption, minimising read/write access time, minimising leakage current, minimising bit-line swing to reduce power consumption, increasing soft error immunity, etc. There is tension between several aspects of the design brief. For instance, a greater cell ratio, as stated in, reduces the likelihood of read failure but increases both area and leakage.

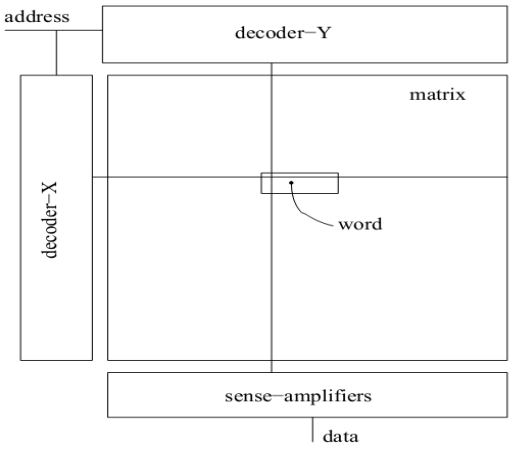


Figure 2 High level overview of the parts making up an SRAM

Typical Building Plan [drawing 3] Two inverters are cross-coupled with two access transistors to form an SRAM cell's six transistors [1]. In a positive feedback loop, two cross-coupled inverters provide the memory. Reading and writing are controlled by a pair of access transistors. [17] [18] There are four transistors total between the two cross-coupled inverters, two PMOS and two NMOS. In this circuit, NMOS transistor drains are connected to PMOS transistor gates, while PMOS transistor gates are connected to NMOS

transistor drains. This creates a recursive loop which maintains a steady cellular potential. If one inverter's output is high, it will keep the other inverter's input low, which will keep that inverter's output high, and so on.[2] [9-10].

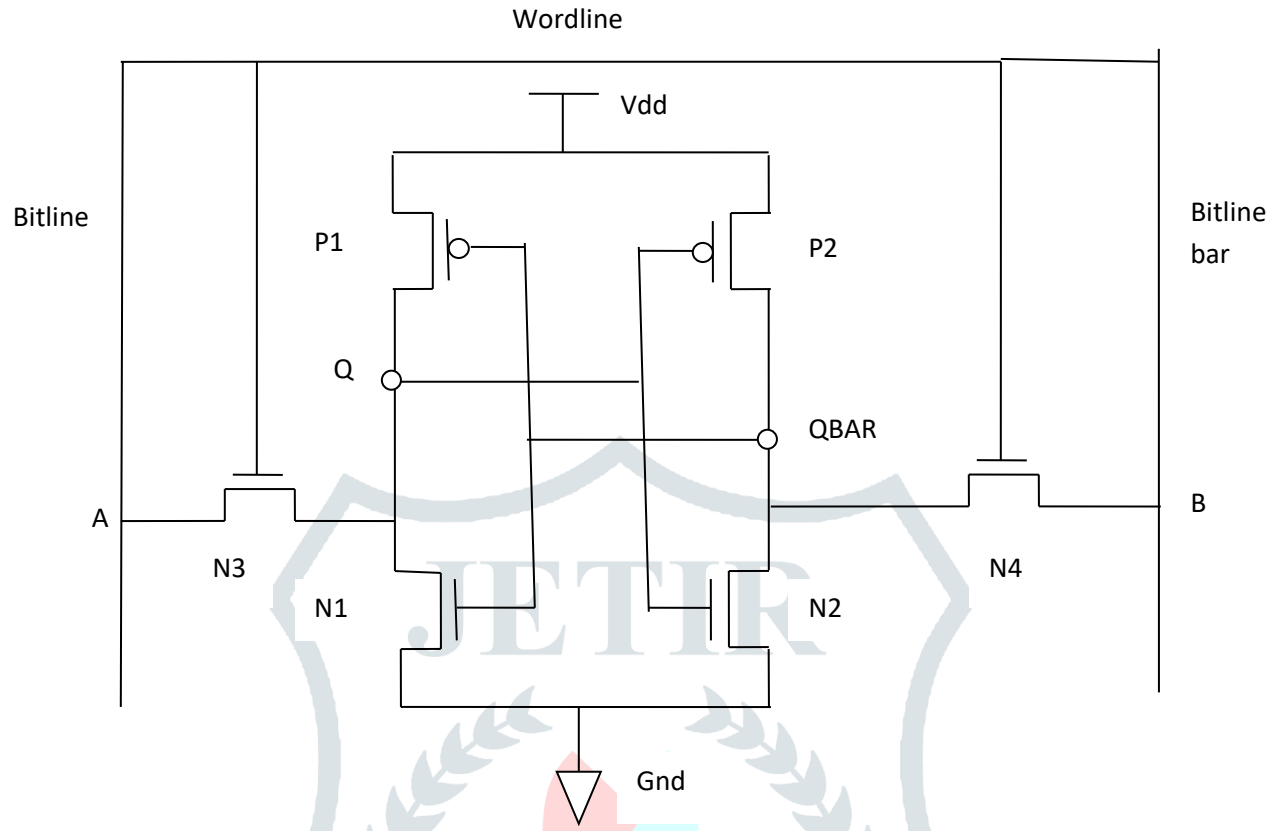


Figure 3. Standard Architecture of 6T SRAM Cell

There are a total of six transistors in a single Standard Architecture SRAM cell; two inverters, one cross-coupled inverter, and 2 access transistors. The memory function is supplied by the positive feedback loop produced by the 2 cross-coupled inverters, and the read and write operations are executed by the two access transistors.[4][12]. This fundamental design has been in use for a long time because of its simplicity and dependability, although newer versions have been created that employ fewer transistors to decrease energy consumption and boost density.[14][19].

B. Standard Architecture DRAM CELL

DRAM, is another form of semiconductor memory utilised extensively in today's digital systems. In contrast to SRAM, DRAM stores a single bit using only a transistor and a capacitor. The typical layout of a DRAM cell will be discussed below.

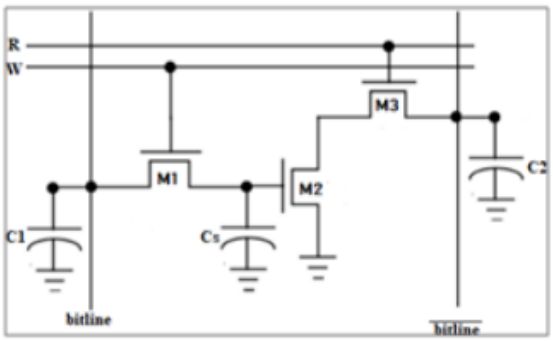


Figure.4 Standard Architecture 3T DRAM

One transistor and one capacitor make up the Standard Architecture DRAM cell structure [7]. In most cases, the transistor will be a normal metal oxide semiconductor (NMOS) type, and the capacitor will consist of two metal plates separated by a dielectric. Any time a transistor is used to charge or discharge a capacitor, the data contained in it is at risk of corruption if the refresh cycle is not carried out correctly.[20][21] Finally, a single bit of information may be stored in a Standard Architecture DRAM cell, which consists of a single transistor & a capacitor. The cell structure of this SRAM is inferior to the industry standard in terms of speed, refresh rate, and power consumption. DRAM's great density and cheap cost per bit make it popular for uses like primary memory in computers & mobile devices, despite the technology's flaws.

Both 3T DRAM and 6T SRAM Cells have serious leakage problems. Power dissipation in semiconductor memories is aided by off-state current in MOSFET devices.

C. Sub-threshold Leakage Current

Vgs Vth of mos transistor is the primary reason for sub-threshold leakage current. In this condition, the mos transistor is in its off state, operating in a weak inversion zone. Leakage current below the threshold, defined mathematically as

$$I_{subthreshold} = I_0 e^{\frac{(Vgs-Vth)}{nV_T}} [1-e^{-(\frac{V_{ds}}{V_T})}] \quad \ldots\ldots(1)$$

Where,

$$I_0 = \frac{W \mu_0 C_{OX} V_T^2 e^{1.8}}{L},$$

$$V_T = \frac{KT}{q} = \text{Thermal voltage},$$

Vth = Threshold voltage (Volt),

Vds = Drain to source voltage (Volt),

Vgs = Gate to source voltage (Volt),

Cox = Gate oxide capacitance (F/m2),

μ0 = Carrier mobility,

n = Sub-threshold swing coefficient.

L = Transistor Length

W = Transistor Width.

D. Leakage Power

Scaling methods reduce the overall footprint of the mos transistor. Leakage power increases when the size of devices decreases due to power dissipation in the devices. Leakage power in a mos device is defined as:

$$P_{Leak} = I_{Leak} * V_{dd} \quad \text{.....(2)}$$

Where, P_{Leak}= Leakage power,

I_{Leak} = Leakage current,

V_{dd} = DC Power supply

III. SRAM Cell Design Metrics, Failure Mechanisms & Operations

A. SRAM Design Metrics

SRAM circuits must be designed with the tiniest possible transistors, making them especially vulnerable to manufacturing flaws. A crucial component of any SRAM design is striking a balance between the trade-offs of compact space, low power, and rapid reads/writes. To rephrase, SRAM design necessitates a careful juggling act between competing design criteria like minimising cell area by using smaller transistors, preserving read/write stability, cutting down on energy consumption by decreasing the energy supply, shortening the time it takes to access data during reading and writing, minimising leakage current, decreasing bit line swing to lessen energy consumption, enhancing soft error immunity, and so on. There is tension between several aspects of the design brief. For instance, a greater cell ratio, as stated in, reduces the likelihood of read failure but increases both area & leakage.

B. SRAM Failure Mechanisms

To well as the aforementioned design requirements, the designer of SRAM must also account for the numerous SRAM cell parametric failures that may occur, most often as a result of read, write access, & hold failures.

Read Failure – This error manifests itself during the process of reading the contents of an SRAM cell. In Fig.5 (a), we assume that BLB is discharging via nodes MN3 and MN1 and that node QB is holding a "0." A voltage ripple VQB is produced by the resistive divider formed by the pull-down transistor MN1 and the pass transistor MN3 if MN1's resistance is greater than MN3's. The cell state will transition if the read value of VQB exceeds the switching threshold of the inverter produced by MP2 and MN2. If the voltage increase at the "0" storage node is greater than the voltage rise at the "1" storage node, the read failure will be mitigated.

Write Failure – The attempt to write to the SRAM cell failed. When the wordline is turned on, the node containing the "1" is drained using access transistors, and the write operation succeeds if this happens at the same time. While decreasing write failure, increasing the wordline turn on time also slows down SRAM.

Access Failure – If the sense amplifier fires and the voltage gap between the bitlines is less than the offset voltage, the access will fail. As the discharge current of the bit line is reduced by the pass transistor and the pull-down transistor, access is denied. Clearly, increasing the strength of the pull-down transistor will result in a quicker bitline discharge by lowering the resistance in the discharge route. However, higher density SRAMs should avoid using cells with a bigger size because of the performance hit they would take from doing so.

Hold Failure - The hold is lost because the pull-down NMOS transistors connected to the node holding "1" are leaking significantly. Because of high leakage of the pull-down transistor, the node storing a "1" drops from VDD in scaled technology at lower supply voltage (VDD). As soon as the cell's voltage falls below the inverter's holding threshold of zero, the device enters hold mode. Leakage in standby mode may be reduced by high-Vt pull-down transistors, preventing this failure. The read latency is a tradeoff for the benefit.

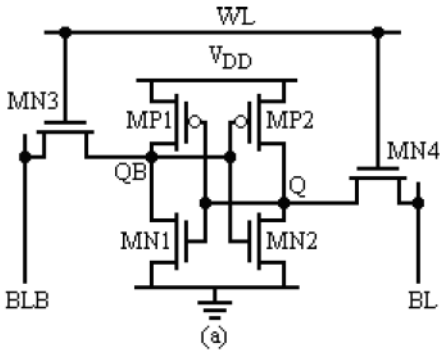
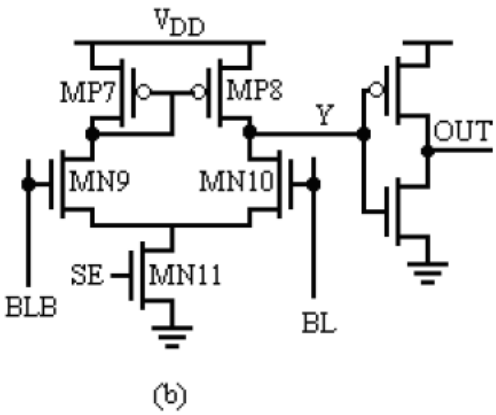


Figure 5 (a): Standard 6T SRAM cell



(b): Sense amplifier

IV. Implementation System

In terms of leakage current, the proposed 6T SRAM cell significantly outperforms both the normal 6T SRAM and the 8T SRAM cell. The suggested 6T SRAM cell consumes less power without sacrificing performance compared to a conventional SRAM cell.

A cell with the value "0" triggers the activation of M3 and the deactivation of M4 by connecting the Qbar node to VDD. In response, M2 is disabled, and the logic '1' is generated at the Qbar node. In a typical SRAM cell, the current Ids-M2 will increase the voltage at node Q, which may lead to an error. The subthreshold voltage VIDLE and the subthreshold current M1 keep the node value at Q constant. The value of the Q node must be discharged to GND along the bit-line bar when the system is not in use. Unlike conventional cells, which need periodic data refreshes, this one can keep its "0" value indefinitely because to a new technique that takes advantage of leakage current. The analysis shows that the layout is very reliable when a value of zero is stored. Q node and BLB-line get power from VDD; 1 data. In this circuit, M1 is the NMOS access transistor, which turns on M4 and triggers M2 by raising the Qbar node to VDD-Vth. When M2 and M4 interact, a positive feedback loop will result. Keeping the BLB-line grounded, having M1 pull the Q node to VDD while M3 is on, and then having M1 pull the Q node down to GND results in a data value of 0. After the writing procedure is complete, the cell goes into a dormant condition. When M6 is active, the word-line is disconnected from the cell and sent to GND during read operations. A charge of VDD is applied to the bit-line before reading. Even when the voltage at the Q node is high, the performance of the bitline will not degrade. Q node releases the bitline to GND when it saves the value 0. Afterward, information may be retrieved by connecting a sensing amplifier to a bit-line.

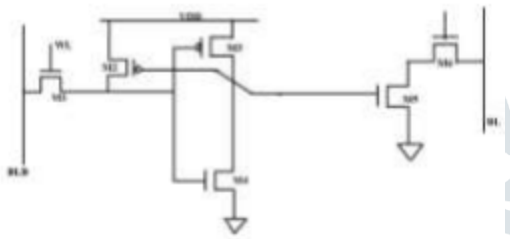


Figure 6. CMOS 6T SRAM Cell

V. Result Analysis and Discussion

This means that the static noise margin of the CMOS 6T SRAM cell is less than that of the Standard Architecture 6T SRAM cell. Power consumption and leakage current are lower than those of the other two SRAM cells.

Comparisons Result Summary is shown below table 1

Table 1

Parameters	Standard Architecture 6T SRAM cell	CMOS 6T SRAM cell
Technology	90nm	90nm
Supply voltage	1.2V	1.2V
Power Consumption	0.262mW	0.155mW
Leakage Current	0.918mA	0.237mA

VI. Conclusion

In this way, the proposed CMOS 6T SRAM uses less power than the Standard Architecture 6T while maintaining the same level of efficiency as both SRAM and DRAM designed using the latter. This study unveils a brand-new, very reliable 6T-SRAM fabricated using a 90 nm process. This novel architecture employs distinct modes of operation for writing and reading, limiting its usage to a single bitline each time. Cache relies heavily on low-power states, therefore the new cell is a significant improvement. When the cell is not actively being used, the voltage on the word line stays at Vidle. Leakage current and positive feedback allow the information to be kept in the cell indefinitely, eliminating the requirement for a refresh cycle. The innovative structure enhances the parameters by separating the reading and writing processes.

References

1. Y. Murali Mohan Babu, Suman Mishra, K. Radhika., proposed the title “Design Implementation and Analysis of Different SRAM Cell Topologies” that was published in the year of 2021 by the journal of 2021 International Conference on Emerging Smart Computing and Informatics (ESCI).
2. Hansraj, Alka Chaudhary, Ajay Rana., proposed the title “Ultra Low power SRAM Cell for High Speed Applications using 90nm CMOS Technology” that was published in the year of 2020 by the journal of 8th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO).
3. Tanisha Gupta, Pankaj Naik., proposed the title “Comparative Analysis of 2T, 3T and 4T DRAM CMOS Cells” that was published in the year of 2017 by the journal of the IEEE Custom Integrated Circuits Conference.
4. V. Rukkumani, M. Saraanakumar, K. Srinivasan., proposed the title “Design and Analysis of Sram Cells for Power Reduction Using Low Power Techniques” that was published in the year of 2016 by the journal of the IEEE Custom Integrated Circuits Conference.
5. Cagla Cakir, Mudit Bhargava, Ken Mai., proposed the title “6T SRAM and 3T DRAM Data Retention and Remanence Characterization in 65nm bulk CMOS” that was published in the year of 2012 by the journal of the IEEE Custom Integrated Circuits Conference.
6. T. Gupta and P. Naik, "Comparative analysis of 2T, 3T and 4T DRAM CMOS cells," 2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC), Indore, India, 2017.
7. S. Akashe, A. Mudgal and S. B. Singh, "Analysis of power in 3T DRAM and 4T DRAM Cell design for different technology," 2012 World Congress on Information and Communication Technologies, Trivandrum, India, 2012.
8. P. Kushwah, N. Saxena, S. Akashe and Saurabh, "Reduction of Leakage Power & Noise for DRAM Design Using Sleep Transistor Technique," 2015 Fifth International Conference on Advanced Computing & Communication Technologies, Haryana, India, 2015.
9. P. Bikki, M. K. R. T, M. Annapurna and S. Vujwala, "Analysis of Low Power SRAM Design with Leakage Control Techniques," 2019 TEQIP III Sponsored International Conference on Microwave Integrated Circuits, Photonics and Wireless Networks (IMICPW), Tiruchirappalli, India, 2019.
10. S. H. Choudhari and P. Jayakrishnan, "Structural Analysis of Low Power and Leakage Power Reduction of Different Types of SRAM Cell Topologies," 2019 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, India, 2019.
11. Vijayaraja L., Gokulnath B., Sai Ganapathy S., Dhanasekar R., Kesavan R.,(2018),"Modified SEPIC converter based PV Inverter for Power Quality Improvement",7th IEEE International Conference on Computation of Power, Energy, Information and Communication, ICCPEIC 2018,Vol.2018- January,no.,pp.280-285.doi:10.1109/iccpeic.2018.8525223
12. Senthilkumar K.K., Kumarasamy K., Dhandapani V.,(2021),"Approximate Multipliers Using BioInspired Algorithm",Journal of Electrical Engineering and Technology,Vol.16,no.1,pp.559- 568.doi:10.1007/s42835-020-00564-w.

13. Slacer P.P., Priyadharshini R.I., Benila A., Rajam P.P., Jacob J.J.,(2022),"A contemporaneous input vector monitoring Bist architechture using memory",AIP Conference Proceedings,Vol.2393,no.,pp.- .doi:10.1063/5.0074185
14. Harish R., Indira G., Kalapriyadarshini G., Valarmathy A.S., Chandrakala P.,(2022),"Analysis and implementation of DC power and appliances in conjunction with solar for large independent units",AIP Conference Proceedings,Vol.2393,no.,pp.- .doi:10.1063/5.0079651
15. Srinivasan V., Varshad Venkatraman R., Senthil Kumar K.K.,(2013),"Schmitt trigger based SRAM cell for ultralow power operation-a CNFET based approach",Procedia Engineering,Vol.64,no.,pp.115-124.doi:10.1016/j.proeng.2013.09.082
16. Prof. Virendra Umale. (2020). Design and Analysis of Low Power Dual Edge Triggered Mechanism Flip-Flop Employing Power Gating Methodology. International Journal of New Practices in Management and Engineering, 6(01), 26–31.
17. Dr. Bhushan Bandre. (2013). Design and Analysis of Low Power Energy Efficient Braun Multiplier. International Journal of New Practices in Management and Engineering, 2(01), 08–16.
18. Vijay, J.., & Hema, L. K. . (2023). Analysis of Solid-State Transformer Enabled Hybrid Microgrid using Resilience Energy Amendment Control Algorithm. International Journal of Intelligent Systems and Applications in Engineering, 11(2), 572–581.
19. Karthikeyan, A., Balaji, S. ., Santhakumar, R. ., Rajalakshmi, S. ., & Jayakrishnan, P. . (2022). A Leakage Reduction Charge Pump based Domino Logic for Low Power VLSI Circuits. International Journal of Intelligent Systems and Applications in Engineering, 10(4), 211–215.
20. Eamani, R. R. ., & Kumar, N. V. . (2023). Design and Analysis of Multiplexer based Approximate Adder for Low Power Applications. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3), 228–233.
21. C. S., S., Kunju, N. ., & Shahul Hameed, T. A. . (2023). Design and Simulation of Two Stage Wideband CMOS Amplifier in 90 NM Technology. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2s), 249–258.
22. International Roadmap for Device and Systems (IRDS) 2022. Available online: https://irds.ieee.org/editions/2022 (accessed on 15 April 2023).