

Design and Implementation of power efficient TCAM

Dr. Kiran V

Dept. of Electronics & Communication
RV college of engineering,
Bengaluru, India- 560059

Bhavar Santosh Appasaheb

Dept. of Electronics & Communication
RV college of engineering,
Bengaluru, India- 560059

Vinayak V A

Dept. of Electronics & Communication
RV college of engineering,
Bengaluru, India- 560059

Abstract: Routers are one of the integral components of any network and they play a vital role in exchange of information between the endpoints of network. The data packet moves through many routers from source to destination in a typical modern network. With advances in communication technology, networking applications require routers with higher capabilities to route data packets over a network. Increase in endpoints of a network results in additional memory needs for connected routers. Power consumption is major area of concern. Ternary Content Addressable Memory (TCAM) is convenient to implement router's memory because of its fast and parallel data searches. The designs were implemented and simulated in Cadence Virtuoso. Transient analysis was performed to verify the functionality for different cases of operation. The static power consumption was measured by ensuring there are no switching of any of the inputs and transistors are in off state. The comparison shows there is a reduction of 53.02% of static power consumption for the proposed design.

Keywords — Static memory, single-ended SRAM, bit-line, dynamic power, Low power (LP), Single-ended (SE)

I. INTRODUCTION

The electronics industry has seen a remarkable improvement in terms of efficiency and highspeed applications with the progress of technological achievement. Numerous applications are being developed as a result of advancements in telecommunications technology. Many of these applications require low connection latency, high throughput, and high data rates. With increase in telecommunication trends the requirement for large amount of data storage also increases. Because of the rising demand for vast amounts of data storage, new technologies must be developed more quickly than they currently are.

Network routers are essential tools for supporting massive and quick data transmission across the Internet [14]. The network address and the host address are the two components of each network router's address. While host addresses are made up of the remaining bits, network addresses might vary in size depending on setup. Each incoming packet is routed by the router to its destination. The router must identify the longest routing prefix that corresponds to the destination IP address in order to route an incoming packet. Whenever the packet reaches a router, the routing table memory is consulted in order to forward the packet to next router. The two vital requirements for 4G/5G network routers are efficient packet routing and data searching.

Content Addressable Memory (CAM) is a type of storage where the data is used as the address rather than the memory address. The speed of CAM lookup is significantly faster than the speed of software lookup since CAM is a hardware table and can compare the search data with all the stored data in simultaneously. Hence CAM is widely used in network routers. There are two types of CAM – Binary CAM and Ternary CAM. Ternary CAM has a “don't care” bit in addition to bit 0 and bit 1 in a Binary CAM. Fast and concurrent data searches are a feature of ternary content addressable memory (TCAM). Rapid routing is an objective that can be easily attained by the routing table used with TCAM. However, due of the high number of transistors employed, this notable benefit would result in significant leakage power.

Typically, a TCAM cell consists of three components: A match line evaluation logic, XOR type CAM cell and a SRAM cell. The data (D) is stored in the XOR type CAM cell which will be compared with search data (SL and \overline{SL}). The mask data (M) is stored in SRAM cell as shown in figure 1.1

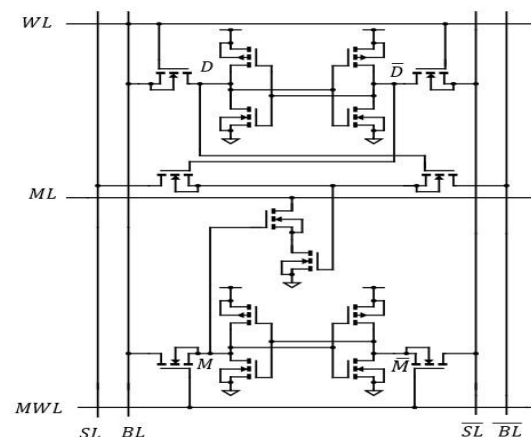


Figure 1.1. Conventional 1-bit TCAM cell

IPv4 address is displayed as IP-ADDRESS/SUBNET MASK. SUBNET MASK is a number between 0 and 32, and IP ADDRESS is of 32 bits. The MASK field indicates the number of bits to take into account from the MSB in IP ADDRESS to determine which interface the packet needs to be forwarded. The remaining (32- MASK) LSB bits are don't care bits in the IP ADDRESS field. For instance, IP address 150.140.8.14/24 can be divided into two parts [4].

150.140.8.14 is stored as IP address and 24 is the mask length which is stored as 255.255.255.0 in the SRAM cell. 24 indicates that there are 24 care bits and 8 don't-care bits in this IP.

II. LITERATURE REVIEW

[1] In the proposed design, 32-bit TCAM is partitioned into 4 segments of 8 bits. The feature of the mask is exploited to reduce power consumption. Mask LSB 1 for all 1's segment and Mask MSB 0 for all 0's segment. In both the cases, VDD and GND of SRAM cell to store mask data are maintained at same voltage level with help of couple of inverters, inputs dependent on MSB and LSB of mask data. Mask LSB 0 and MSB 1, known as boundary segment and VDD and GND of SRAM cell to store mask are maintained at different voltage levels. The power consumption in boundary segment is not reduced whereas power consumption in all other segments is reduced. The DR-TCAM can reduce TCAM leakage power by 41% for a set of real routing tables when compared to the conventional TCAM architecture. Additionally, there is a 12% drop in overall power.

[2] The paper presents design and implementation of precharge-free ternary content addressable memory (PF-TCAM). By eliminating the precharge phase before each search and completing the search in HALF clock cycle, the suggested searching strategy boosts the speed of search by minimizing the ML evaluation time in half. Despite a 50% increase in evaluation speed, a 32 16-bit suggested macro demonstrates energy efficiency benefits of 56% and 63% over 25 different search keys compared to standard TCAM and compact TCAM, respectively.

[3] Both XOR-CAM cells with NOR-based TCAM and XNOR-CAM cells with NAND-based TCAM are discussed in this paper. To reduce switching on SL and save power, it proposes the refine search enable (RSE) strategy. Additionally, it proposes using a Don't Care Gating (DCG) mechanism to stop redundant comparisons. The optimal setup for a 128-32 TCAM demonstrates that, with a 1.3% search speed boost, the DCG scheme combined with the RSE approach reduce SL energy by 72%-79% when the gating granularity is 16.

[4] A novel 8T-SRAM is implemented using transistor stacking. It is a technique where a transistor is replaced by two transistors, with width halved for both compared to original. It explains operation of working of proposed design and analyzes different power dissipations. Power and delay are calculated for write and read operations of logic 1 as well as logic 0 from simulation results. The proposed design utilizes SE scheme as well.

[5] This paper presents the working operation and performance of conventional 6T and 8T SRAM. It was observed, conventionally used SRAM cells are responsible for higher power consumption, as memory occupies around 90% of the chip area in modern semiconductor chips. Most of the power is consumed for BL and BL_{Bar} charging and discharging in SRAM. Hence proposed design, SE SRAM makes use of only single BL.

[6] To lower the power consumption of a typical 6T-SRAM cell, it uses Gated VDD and Multi Threshold CMOS (MTCMOS) design approach is presented in this paper. In Gated VDD scheme, supply voltage is not fed to SRAM when it is not used. An additional high threshold voltage transistor is

controlled through a signal, to provide or shut off ground connection to SRAM cell. Similarly, in MTCMOS technique two high threshold voltage transistors are connected between SRAM and power supply terminals. They connect to virtual ground and virtual VDD when SRAM is in operation. Low threshold transistors are used in SRAM cell.

[7] It presents a study of different parameters of a 6T-SRAM using Source Biasing in addition to stacking, Gated VDD and Multi-Threshold CMOS. In source biasing scheme, additional NMOS transistor is placed between SRAM and ground which is controlled by WL. Hence in stand-by mode, GND voltage of SRAM cell is raised contributing to less power consumption. SRAM based on different design techniques are simulated to compare leakage current and delays.

[8] Highly stable and reduced power dissipation SRAM is proposed. The primary idea behind voltage lowering scheme is to reduce supply voltage below WL voltage. Additional circuitry, consisting of three transistors is proposed to achieve voltage lowering. The implemented design is analyzed for stability, static and dynamic power dissipation, delay and area.

[9] In this paper, various SRAM cell topologies have been implemented on 90nm technology node with Cadence virtuoso tool. It has been noticed that 7T SRAM cell has minimum read power among all considered topologies. write power in 8T SRAM cell reduced by 44.15% as correlated to conventional 6T SRAM cell. The write delay in 9T SRAM cell found minimum among all considered cells. highest value of RSNM (read static noise margin) has been observed in conventional 6T SRAM cell as correlated to all simulated topologies. The WSNM (write static noise margin) of 8T SRAM cell found to 2 \times as of 6T SRAM cell.

[10] The paper proposes power consumption approaches of an SRAM based on different self-controllable voltage logic (SVL) switches. In upper SVL (USVL) switch, a circuit combining of PMOS in parallel with series of n NMOS is added, between SRAM and VDD and vice-versa in case of Lower SVL (LSVL). Both the techniques are combined to reduce power at the cost of increased area. It explains Variable Threshold CMOS (VTCMOS) and combination of VTCMOS with SVL. Different leakage currents and static power are calculated for proposed designs. VTCMOS applied to conventional SRAM cell results in 20.56% reduction in sub-threshold leakage current and 15.23% reduction in static power.

[11] An adaptive match-line (ML) discharge technique for ternary content addressable memory with low power consumption and fast speed is presented in this work. Comparing the suggested adaptive ML discharging technique to the usual method, sensing delay is improved by up to 19%, and ML power is reduced by 81%.,

[12] The CAM matrix has been proposed in this work. BCAM is used in the suggested CAM matrix architecture's design. The design method is used to filter out parallel comparisons when

searching. When compared to the standard design, the proposed design uses less power and operates faster.

III. Methodology and Related work

A. Methodology

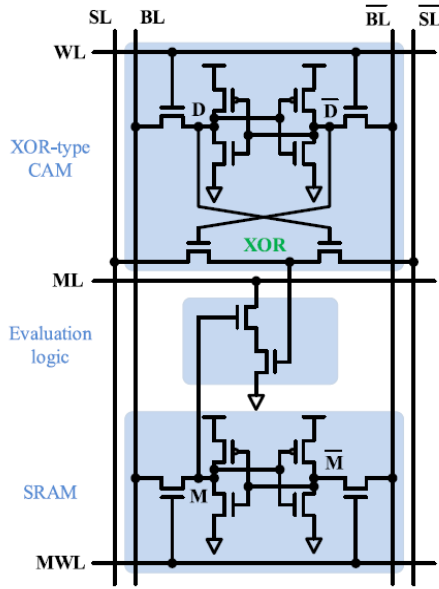


Fig. 3.1 One TCAM cell is composed of XOR-type CAM cell, evaluation logic, and SRAM cell.

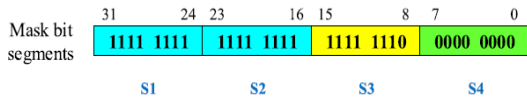


Fig. 3.2 An example of 4-segment mask data 255.255.254.0.

THE THREE KINDS OF TCAM SEGMENTS FOR IPV4

Mask Data segment	Meaning	Example in Fig. 2
1...1...1	All 1s	S1 and S2
1...10...0	Boundary segment	S3
0...0...0	All 0s	S4

The main idea of the DR-TCAM is using the feature of continuous 1s and 0s of mask data to minimize the static power consumption of TCAM. First, an N-bit TCAM entry is partitioned into S segments, and each segment has L bits data. TABLE II shows the three kinds of mask segments, i.e., all 1s, all 0s, and boundary segments. The corresponding example is shown in Fig. 2. The 32-bit mask data 255.255.254.0 is partitioned into 4 segments (S1 to S4), and each segment has 8 bits. The first two segments (S1 and S2) are the case of all 1s. The segment S3 is the boundary segment and the segment S4 is the case of all 0s.

According to the continuous feature, the segment type can be determined by checking the MSB and LSB of a segment. As shown in TABLE II, when the LSB is 1, this segment must be all 1s segment; when the MSB is 0, this segment must be all 0s segment; otherwise, the segment is called the boundary segment.

B. Design

Read operation:

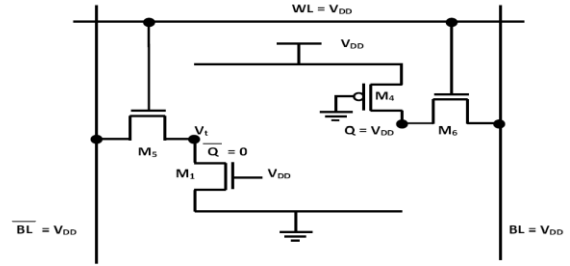


Fig 3.3 conducting transistors in read operation

M 1: Linear, M 4: Linear, M 5: Saturation, M 6: Cut-off.

Power and Delay Calculations:

Here in read operation, we take transistor M1 > M5.

For M5 => Triode Region:

$$ID = Kn [(VGS - Vth) VDS (1/2) VDS2]$$

For M1 => Saturation Region:

$$ID = (1/2) Kn (VGS - Vth) ^2$$

Comparing both equations

$$Kn [(VGS - Vth) VDS - (1/2) VDS^2] > (1/2) Kn (VGS - Vth) ^2$$

$$W1 > W5$$

So, we take width of NMOS transistor approximately double than access transistor.

Write operation:

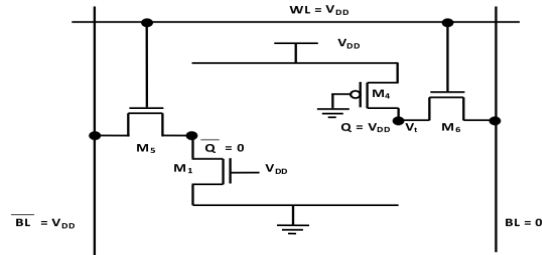


Fig 3.4 conducting transistors in read operation

M 1: Linear, M 4: Linear, M 5: Saturation, M 6: Linear.

Power and Delay Calculations:

Here in write operation, we take transistor M6 > M4.

For M6 => Triode Region:

$$ID = Kn [(VGS - Vth) VDS - (1/2) VDS2]$$

For M4 => Saturation Region:

$$ID = (1/2) Kn (VGS - Vth)2$$

$$Kn [(VGS - Vth) VDS - (1/2) VDS^2] > (1/2) Kn (VGS - Vth) ^2$$

$$W6 > W4$$

So, we take width of access transistor approximately double than PMOS transistor.

IV. Implementation:

Implementing a 6T and 8T gated VDD (Voltage Drain Driven) and stacked involves integrating these different types of memory cells into a larger memory architecture. The implementation of these cells is done to compare their power and delay [6].

1) Conventional 6T SRAM:

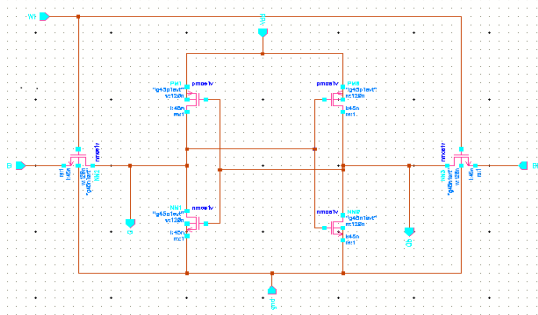


Figure 3.5. Schematic of Conventional 6T SRAM Fig 3.6 shows transient response for conventional 6T SRAM. When the WL is high the SRAM cell is in write mode and the value of BL is stored at D. When the WL is low the SRAM cell is in hold mode. The write and hold modes for different values of input are verified by providing suitable inputs to BL, BL_Bar and WL. The static power dissipation of conventional 6T SRAM with NMOS access transistor is 12.474pW.

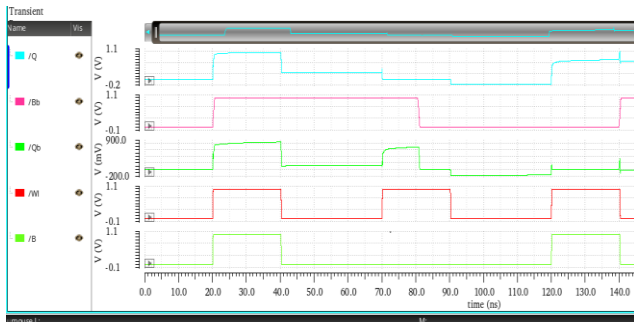


Fig 3.6 Transient response of conventional 6T SRAM

2) Gated VDD:

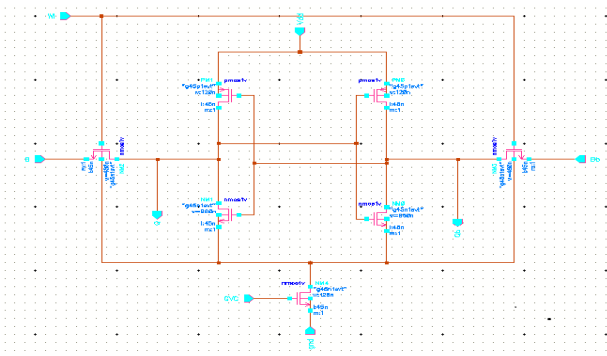


Figure 3.7. Schematic of Gated VDD SRAM

Fig 3.6 shows transient response for Gated VDD based SRAM. The supply is gated through control signal GVC. GVC control is given such that only during write mode, non-

degraded ground potential is supplied. The static power dissipation is 5.0712pW.

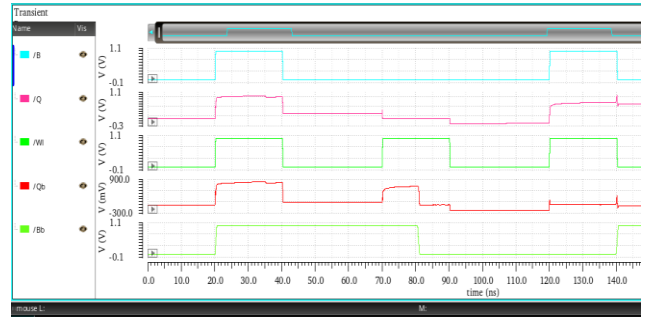


Fig 4.6 Transient response of Gated VDD SRAM

3) Transistor Stacking:

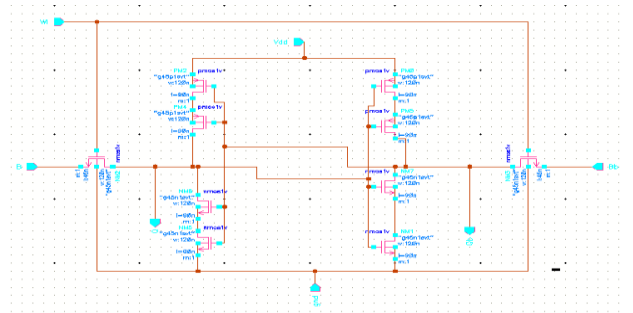


Figure 3.8. Schematic of Transistor Stacking SRAM cell

Fig 4.7 shows transient response for Transistor Stacked SRAM. The write and hold operations are similar to conventional 6T SRAM. The static power dissipation is 6.1337pW.

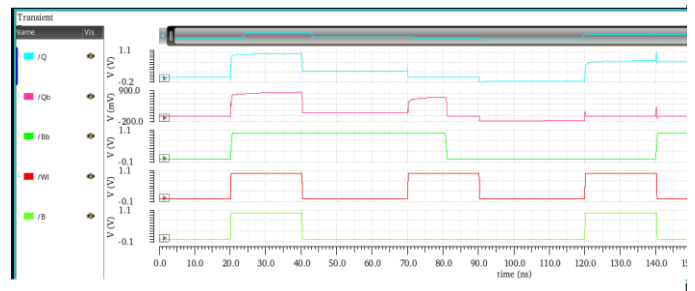


Fig 3.9 Transient response of Transistor Stacked SRAM

4) 8T SRAM:

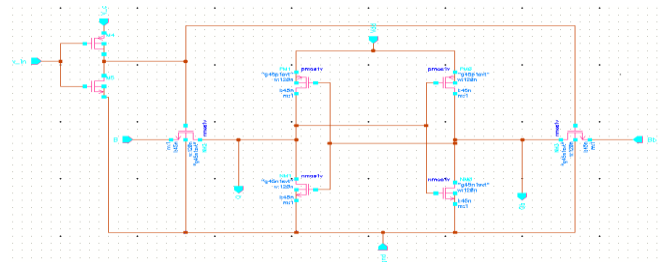


Figure 3.10. Schematic of 8T SRAM cell

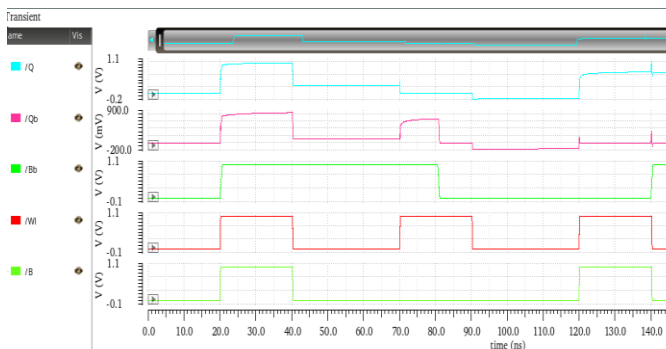


Fig 3.11 Transient response of 8T SRAM

TCAM Implementation:

Creating a TCAM (Ternary Content Addressable Memory) by amalgamating stacked and 6T SRAM cells entails devising a specialized architecture optimized for efficient ternary content search operations. This integration enables the construction of a high-performance memory subsystem capable of swiftly executing ternary content searches, suitable for various applications such as networking. Below the implementation of 1Bit TCAM using 6T SRAM and Stacked SRAM are given.

5) 1 Bit TCAM:

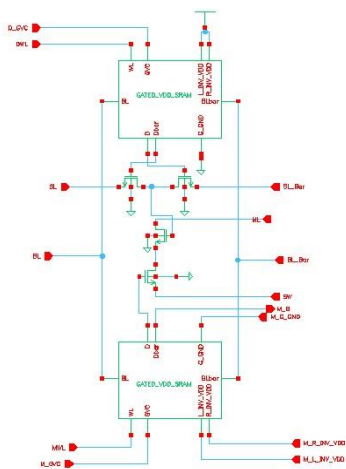


Figure 3.12. Schematic of 1-bit TCAM using Stacked SRAM

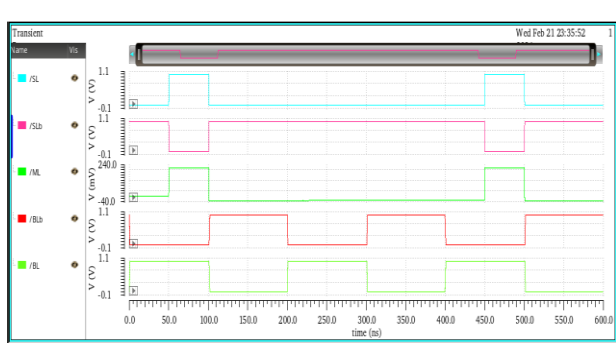


Fig 3.13 Transient response of 1X8 TCAM

6) 1X8 TCAM:

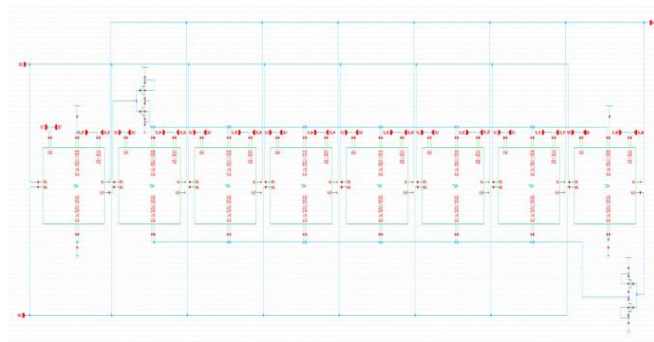


Figure 3.14 Schematic of 1X8 TCAM

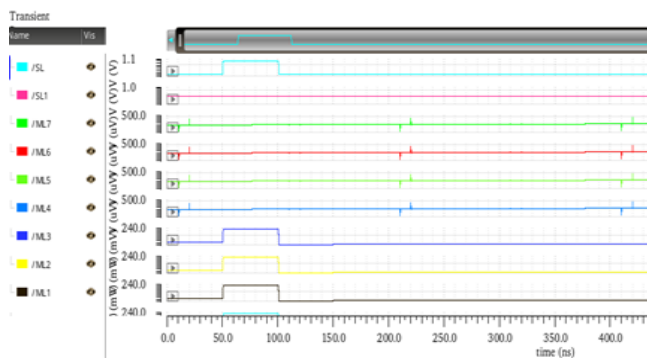


Fig 3.15 Transient response of 1X8 TCAM

V. RESULTS AND COMPARISON

Table 5.1 comparison of results of all implementations

Structure implemented	Average Power	Delay
6T SRAM	33.4nW	114.71ps
8T SRAM	39.5nW	42.4ps
Gated VDD SRAM	12.18nW	20.24ps
Stacked SRAM	5.54nW	29.30ps
TCAM with 6T SRAM cell	120nW	64.5ps
TCAM with Stacked SRAM	46.25nW	55.2ps

Graph below shows the power consumption and time taken by different SRAM structures and also by the TCAM and it implies that by using stacked SRAM structure we can reduce power consumption and improve speed.

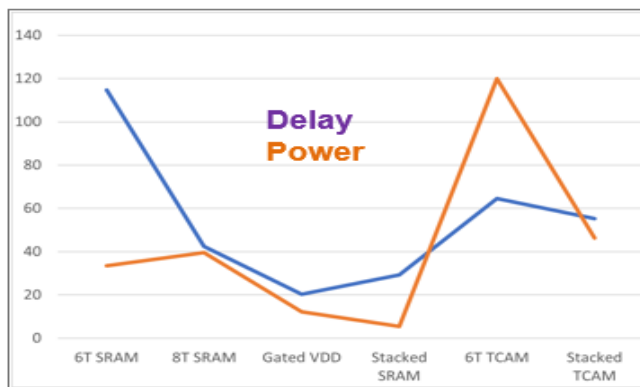


Fig: graph for comparison of delay and power

VI. CONCLUSION

conventional 6T SRAM cell has been correlated the Gated VDD, 8T and Stacked SRAM cells in terms of read power and write power dissipation, read delay and write delay of operation. All the results have been generated on Cadence SPECTRE simulator. Stacked SRAM cell got minimum read power loss as correlated to 6T and 8T SRAM cells. In terms of write power 8T SRAM cell achieves less power consumption. The comparison between existing conventional 1X8 TCAM and optimized 1X8 TCAM shows that static power dissipation is reduced by 61.45% in proposed design.

VII. REFERENCES

- [1] Y. -J. Chang, K. -L. Tsai and Y. -C. Cheng, "Data Retention-Based Low Leakage Power TCAM for Network Packet Routing," in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 2, pp. 757-761, Feb. 2021
- [2] T. Venkata Mahendra, S. Wasim Hussain, S. Mishra and A. Dandapat, "Energy-Efficient Precharge-Free Ternary Content Addressable Memory (TCAM) for High Search Rate Applications," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 7, pp. 2345-2357, July 2020, doi: 10.1109/TCSI.2020.2978295.
- [3] Y. -J. Chang, "Don't-Care Gating (DCG) TCAM Design Used in Network Routing Table," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 11, pp. 1599-1607, Nov. 2010, doi: 10.1109/TVLSI.2009.2025951
- [4] Kiran V & H V Sampad "Data Clustering Algorithm" *International journal in advances in Engineering and Management (IAEM)* Volume 3, Issue 9 Sep @021, pp: 84-89, ISSN:2395-5252.
- [5] H. V. Ravish Aradhya, J. Fadnavis and S. G. Gojanur, "Memory Design and Verification of SRAM-based Energy Efficient Ternary Content Addressable Memory," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), 2021, pp. 1-7, doi: 10.1109/ISCON52037.2021.9702386.
- [6] Kiran V & Nakul C Kubsad "Simulation and analysis of Different CMOS Full Adder for Delay Optimization" *International journal for research in applied Science & Engineering Technology*, Volume 9 Issue IX Sep 2021. ISSN : @321-9653-C Value: 45.98; SJ Impact Factor:&.429.
- [7] A. Chauhan, D. S. Chauhan and N. Sharan, "Characterization of 6T CMOS SRAM in 90nm technology for various leakage reduction techniques," 2016 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 2016, pp. 1-5, doi: 10.1109/SCEECS.2016.7509333.
- [8] J. K. Mishra, P. K. Misra and M. Goswami, "Design of SRAM cell using Voltage Lowering and Stacking Techniques for Low Power Applications," 2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), 2020, pp. 50-53, doi: 10.1109/APCCAS50809.2020.9301672.
- [9] D. Mittal and V. K. Tomar, "Performance Evaluation of 6T, 7T, 8T, and 9T SRAM cell Topologies at 90 nm Technology Node," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-4, doi: 10.1109/ICCCNT49239.2020.9225554.
- [10] A. Pandey, P. Kumar Sahu, R. Dwivedi, A. Kumar, L. Chandra and V. N. Mishra, "Design and Analysis of Low Leakage SRAM cell at 45nm Technology," 2019 International Conference on Computing, Power and Communication Technologies (GUCON), 2019, pp. 411-415.
- [11] R. Schoop, A. W. Colombo, B. Suessmann and R. Neubert, "Industrial experiences, trends and future requirements on agent-based intelligent automation," IEEE 2002 28th Annual Conference of the Industrial Electronics Society. IECON 02, Seville, Spain, 2002, pp. 2978-2983 vol.4, doi: 10.1109/IECON.2002.1182870.
- [12] S. Vijayalakshmi, B. Elango and V. Nagarajan, "Content addressable memory using XNOR CAM matrix," 2016 International Conference on Communication and Signal Processing (ICCS), 2016, pp. 2319-2322, doi: 10.1109/ICCS.2016.7754110.
- [13] D. Mittal and V. K. Tomar, "Performance Evaluation of 6T, 7T, 8T, and 9T SRAM cell Topologies at 90 nm Technology Node," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-4, doi: 10.1109/ICCCNT49239.2020.9225554.
- [14] Kaur V & AV Skandashree "WLAN Controller for priority setting of Data packets using EDCA" *Journal of unification and Computational science*, volume 10, Issue2 Feb 2020. ISSN 1548-7741.
- [15] Karan V "A Heuristic Analysis Approach to Prioritize Voice in Wireless Networks published in scopus indexed Seybold journal, Volume 15 Issue 7-2020 ISSN: 1533-9211