



## Design of TCAM Architecture for Low Power and High Performance Applications

Sattu VEERA VENKATA SATYANARAYANA<sup>1\*</sup> , Sridevi SRIADIBHATLA<sup>1</sup> 

<sup>1</sup>*School of Electronics Engineering, Vellore Institute of Technology, 632014, Vellore, India*

### Article Info

Received: 31/01/2018

Accepted: 11/06/2018

### Keywords

Delay

Early-predict

Energy metric

Match-line

Search-line

Power

Pre-charge

### Abstract

Content Addressable Memory (CAM) is a special memory used in search engines for numerous applications, especially in network routers for packet forwarding. The CAM operation begins with pre-charging followed by evaluating the match-lines (MLs) for searching the data in the stored memory. CAM stores unique words in their array of cells such that only one word is matched for a given search word. ML associated with matched word retains its state and the remaining MLs drain their charge. Ternary content addressable memory (TCAM) is a fast lookup hardware device used for high-speed packet forwarding. However, significant power consumption and high cost limits its versatility and popularity. In this paper, a design has been made for TCAM architecture with pre-charge controller. The pre-charge controller helps in predicting the mismatched MLs during pre-charge phase. This prediction happens at an early stage and helps in terminating the pre-charging of the line. This assures the design of TCAM which consumes low power and also improves the performance. The proposed early predict  $8 \times 8$  TCAM architecture simulations were performed in 45nm technology node using Cadence Virtuoso. The proposed TCAM design exhibits 16.6% reduction in power, 24.7% decrement in delay and 37.1% minimization in energy metric than basic TCAM NOR.

## 1. INTRODUCTION

The primary function of any memory system is writing and reading the lookup data. Random access memory (RAM) is a volatile memory. It searches lookup table data serially in the memory array, and it requires more clock cycles to search the data [1]. Thus, for large capacity memory, RAM is not suitable for high-speed search. Content addressable memory (CAM) is a particular type of computer memory, also known as associative memory. From the functionality point of view, RAM is an inverse of CAM. RAM reads the content of stored memory based on the address match whereas CAM reads the content of stored memory based on the content match as shown in Figure 1. CAM cell can perform all the functions of RAM. It requires comparison circuitry in addition, to get miss/match case at ML, when search input compares with a stored data. CAM surmounts all memory search algorithms for high-speed search. CAM with a large capacity words well in various types of applications that involve searching of lookup table data with high speed. CAM is utilized in wide variety of high-speed applications like network routers [2], cache controllers [3], image-processing [4], gray coding [5] and so on. However, the parallel search lookup table within a single clock cycle offers high search speed and switching activity. As a result, CAM operates with high performance but suffers from high power consumption. Thus, researches concentrate in CAM designing to minimize power without degrading the performance. In CAM architecture, cells in a row are connected either in parallel or series to form ML architecture. The ML architecture with NOR CAM cell is widely used and it offers high performance when compared to NAND CAM cell type ML architecture.

The comprehensive literature on CAM at the circuit and architectural level has been presented. Reviews on CAM at the circuit and architectural level have been explained in detail [6,7]. In [8], a precomputation-based method is implemented to reduce the number of mismatched rows to be compared in a CAM memory array. In this method parameter of each word in a CAM memory and parameter of the search word is

\*Corresponding author, email: svvsatyanarayana8589@gmail.com

computed. Here, it searches the words of CAM memory if the parameter of a search word and the parameter of stored words match. To utilize the high-speed advantage of NOR CAM cell, and low power consumption of NAND CAM cell, hybrid CAM design has been implemented in [9] by segmenting the ML into two parts. In [10], ML power is reduced by dividing the ML into master and slaves MLs. It shares the path, if any mismatch occurs between the storage node and search input. In [11], CAM is implemented by predicting the mismatched match-line early during the precharge phase. This design terminates the mismatched ML dynamically early in precharge phase to improve the power. In [12], power efficient binary CAM design has been implemented with adiabatic logic. In [13], reduction of leakage power in a CAM design has been implemented with two-sided self-gating technique. In [1], novel CAM cell design has been implemented, which reduces power consumption by eliminating short-circuit current path due to one gate path delay. Also, it minimizes the possible routing. Multi-voltage match-line segment [14], selective match-line [15], double-bit CAM [16], Subthreshold logic designs [17] and so on are implemented to improve the energy metric of CAM design. Recently precharge free CAM, self-controlled precharge free CAM ML architectures have been introduced. It is identified that due to the existence of short circuit current path in precharge based CAM circuits, they consume high power. To avoid this and improve energy metric, precharge free CAM in [18] and self-controlled precharge free CAM in [19] have been proposed.

The rest of the paper is organized as follows. Section 2 gives information about ternary content-addressable memory (TCAM), section 3 explained about NOR TCAM architecture and the proposed early predict TCAM architecture. Section 4 presents simulation results. Section 5 concludes the brief.

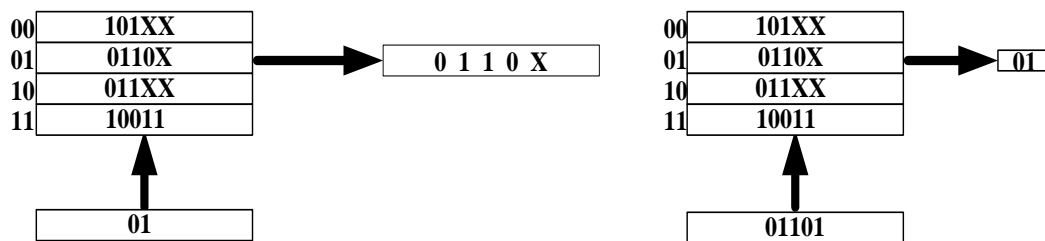


Figure 1. RAM Versus CAM

## 2. TERNARY CONTENT ADDRESSABLE MEMORY

There are two types of CAMs. One is binary content addressable memory (BCAM), and the other, ternary content addressable memory (TCAM). This paper is mainly focuses on TCAM architecture. TCAM memory consists of an array of TCAM cells. Here, the cells may be of type NOR TCAM cell or NAND TCAM cell. Memory architecture is divided into an array of rows and columns with TCAM cells and each row in the array is one word with separate match-line and pre-charge circuit as shown in Figure 2. In TCAM array, unique lookup data is stored in the form of word. TCAM cell adds a third variable, don't care 'X' other than 0 or 1. Because of allowing don't care 'X' in TCAM, there is a possibility that more than one word is matched in an array of words from the given search input. In this case, word with the longest prefix is selected and the address of that word is passed to the priority encoder. Finally, only one row is matched and remaining rows mismatch. The block diagram of TCAM cell is shown in Figure 3. TCAM stores the data with the help of two static random access memory cells (SRAM). Other than SRAM, it consists of comparison transistor, search lines, match-lines and precharge circuitry. It plays a vital role in wireless applications for speeding up search operations, and it is suitable for applications related to routing tables and access lists because of variable length prefixes. Majorly power consumption on TCAM occurs due to match-lines and search-lines. Charging and discharging of match-line is responsible for match-line power, the current path between storage nodes and search line, is responsible for search-line power. Hence number of mismatched MLs should be avoided in order to reduce power and to improve performance.

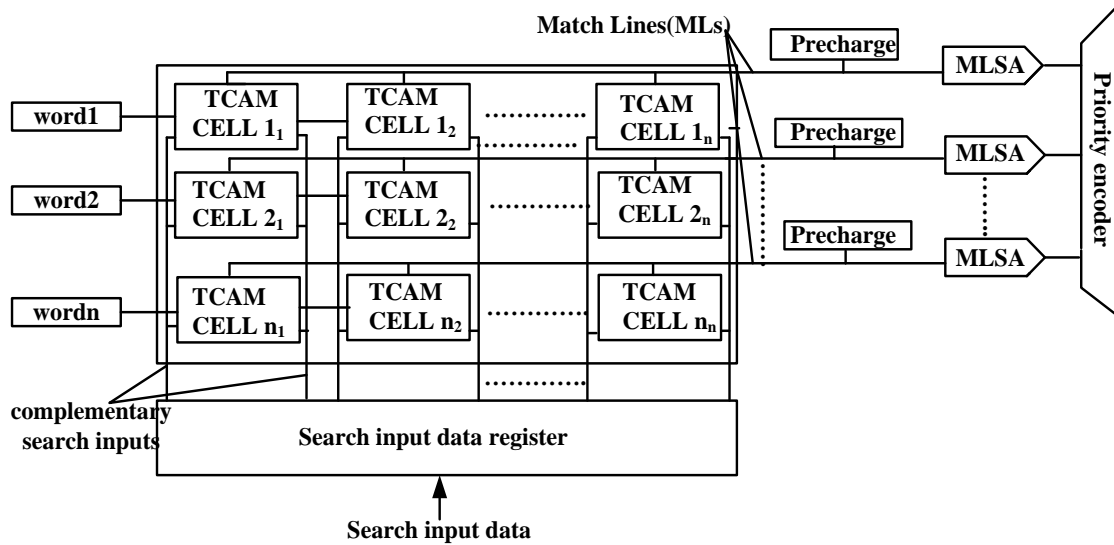


Figure 2. Organisation of TCAM array

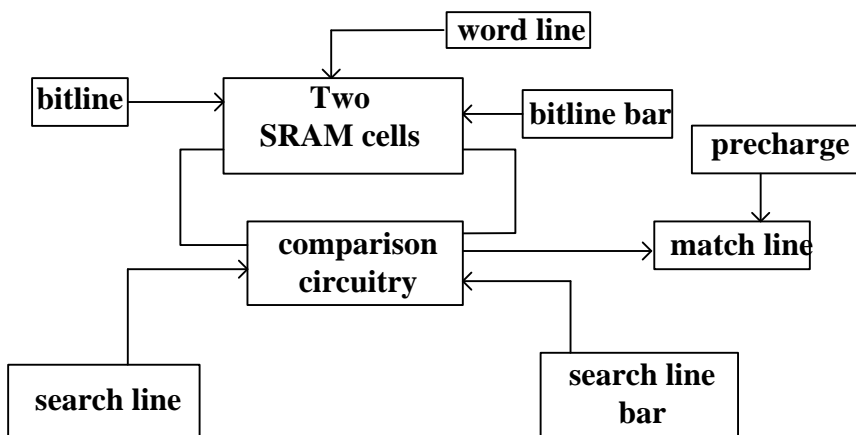


Figure 3. Block diagram of TCAM cell

### 3. TCAM MATCH-LINE STRUCTURE

#### 3.1. NOR TCAM Match-line Structure

TCAM NOR cells are connected in parallel to form NOR ML structure as shown in Figure 4. TCAM NOR cell consists of a pair of SRAM cells for storing ternary logic states 0, 1 and X. Two pairs of pull-down transistors  $T_1, T_3$ , and  $T_2, T_4$  are used for comparing stored bit and search bit. Precharge transistor  $P_1$  is connected to ML. When precharge signal is low the transistor  $P_1$  is ON, then ML is high irrespective of search bit and store bit. If precharge signal is high precharge transistor  $P_1$  is OFF, then evaluation phase starts to compare search input with the stored data. Here, the ML output may be a match, miss or wild-match. In case of a match, ML is isolated from the ground since there is no pull-down path connected to the ground. In case of a miss, at least one pull-down path is short circuited to the ground from the ML. The wild match is the important ternary logic state X used in TCAM design. In case an X value occurs, the cell always produces a match regardless the input search value. Table 1 shows the truth table of TCAM cell design for three different cases. CAM array stores unique lookup data in each word, thus only one word is matched at a time for a given search input. However, unnecessary charging and discharging of mismatched MLs adds considerable dynamic power consumption to the CAM architecture. NOR nature of CAM cell can be viewed by observing the timing waveforms. In TCAM NOR cell, a match indicates high and miss indicates low. The timing diagram of NOR ML for single bit TCAM cell is as shown in Figure 5. NAND

TCAM cell is power efficient, but it is slow because of the long pull-down paths and it also has a charge sharing problem. Thus, NOR TCAM cells are preferred over NAND TCAM cells. BCAM architecture proposed in [11] reduces power by minimizing the ML voltage dynamically with a precharge controller. The idea here is to design TCAM architecture by using already available precharge controller for high performance and low power.

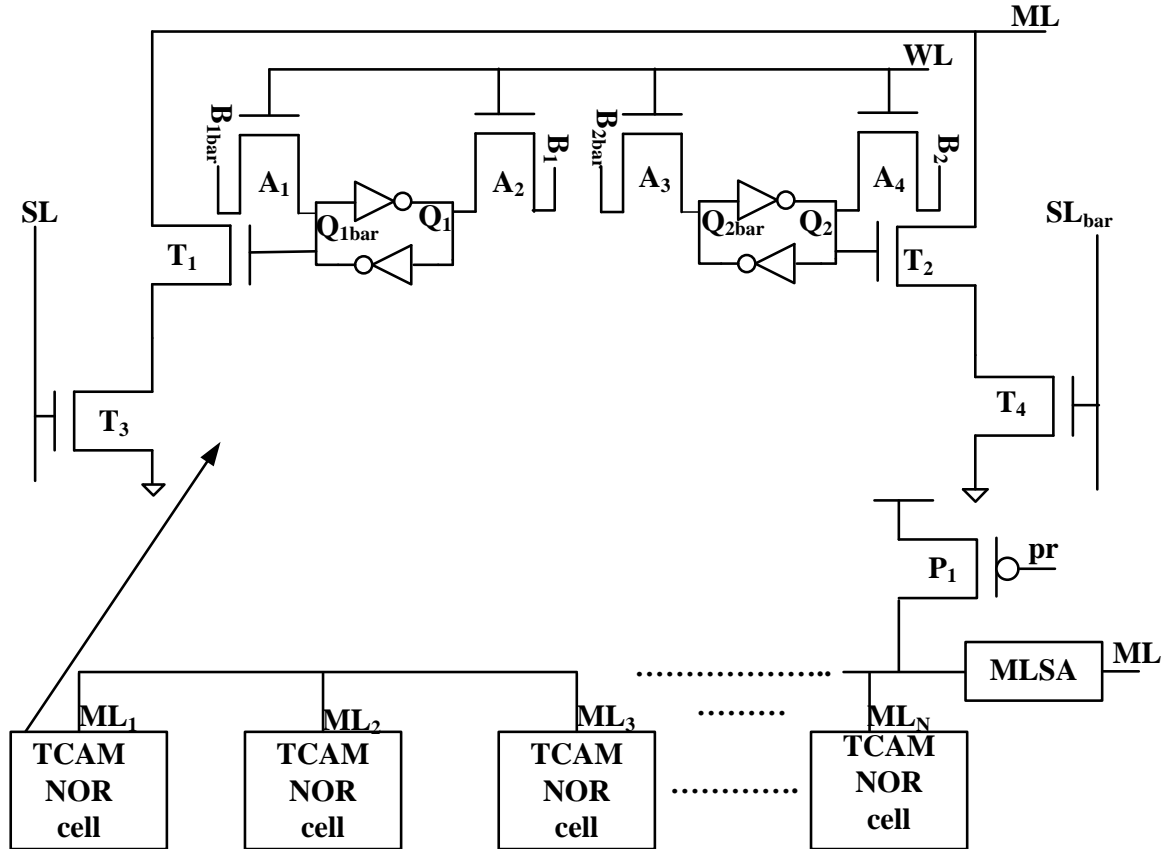


Figure 4. TCAM NOR match line structure

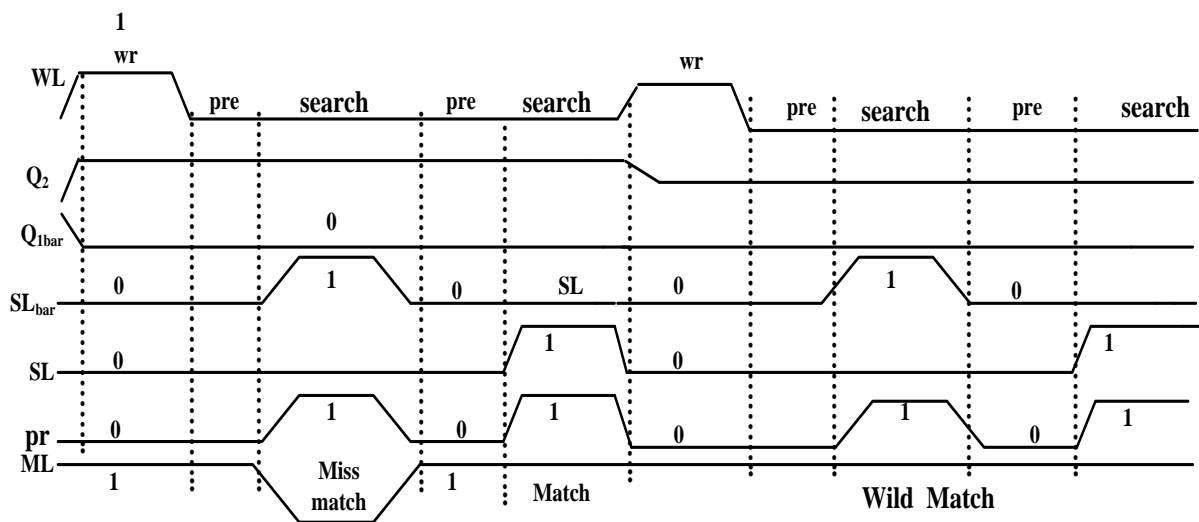


Figure 5. Timing wave form of TCAM NOR cell for different match-line cases

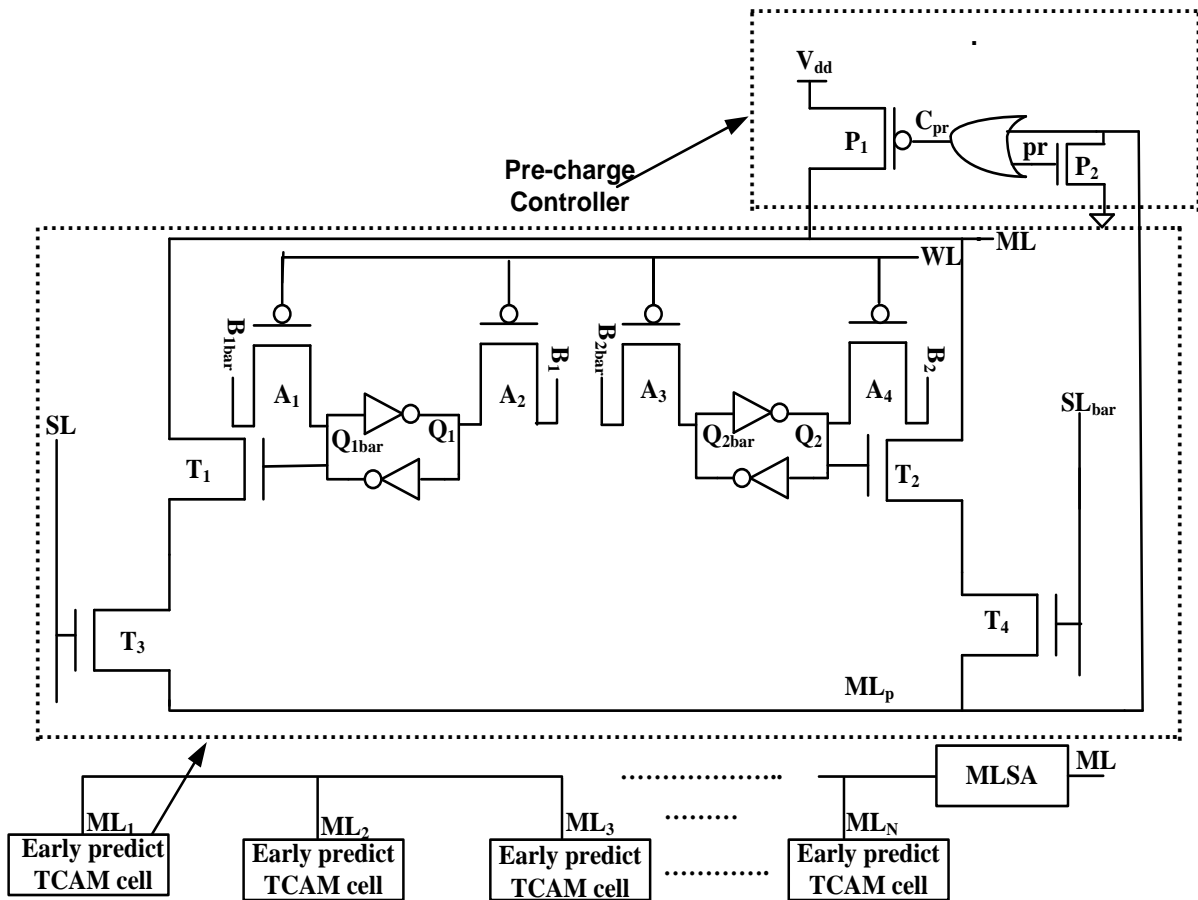


Figure 6. TCAM Early predict match line structure

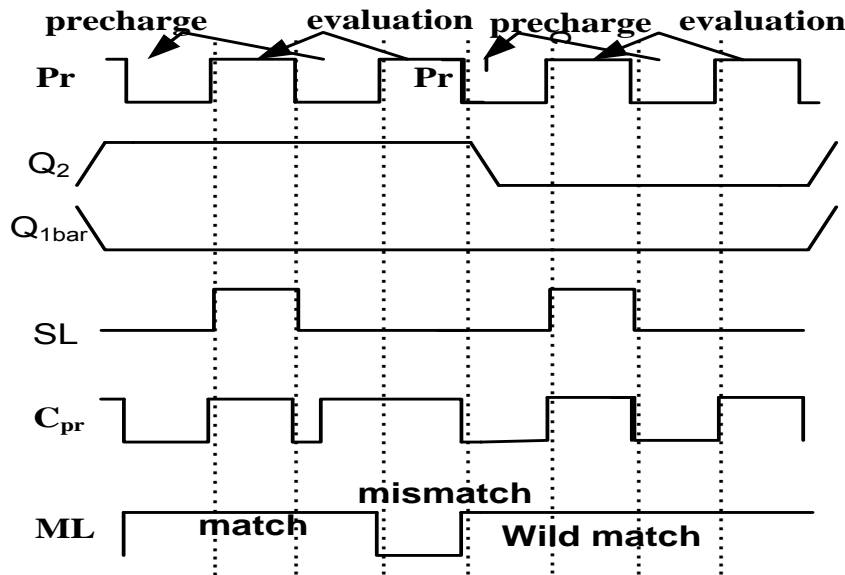
### 3.2. Proposed TCAM Match-line Structure

Early predict TCAM cells are connected in parallel along with a precharge controller to form an early predict ML structure as shown in Figure 6. Each TCAM cell consists of two SRAM cells, four comparison transistors  $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$ . The two SRAM cells are controlled by PMOS access transistors  $A_1$ ,  $A_2$ ,  $A_3$ , and  $A_4$  which improve the stability of TCAM memory design [20]. The SRAM cell with the output nodes  $Q_2$  and  $Q_{1bar}$  is used to store a new logic state  $X$ , which allows the ML output for global matching irrespective of search input. TCAM write operation starts with passing complementary bit information through SRAM cells, by enabling the word line (WL) to low. Precharge and evaluation phase is performed in hold mode by disabling the WL to high. The aim of this architecture is to predict the mismatched ML earlier and to terminate that ML in the precharge phase by using a precharge controller. Precharge controller varies dynamically and it helps to reduce the number of comparison operations.  $pr$  is a precharge signal with fixed width and  $C_{pr}$  is a precharge signal whose width varies dynamically. If input search lookup data is mismatched with any of the words in the ML structure, a path is created between the nodes  $ML_p$  and  $ML$ . The moment match line starts pre-charging, the node  $ML_p$  charges through the comparison circuitry of the TCAM cell from the node  $ML$ . As soon as  $ML_p$  reaches the threshold value,  $C_{pr}$  begins to rise and drains  $ML$ . This overcomes the problem of unnecessary charging of  $ML$  by approximately 45% to 50% of normal voltage swing of mismatch. Table 1 shows the truth table of TCAM cell for three different cases. In case 1  $Q_{1bar}=0$  and  $Q_2=1$ , the comparison transistors  $T_1$  is OFF and  $T_2$  is ON, and then the ML output for miss/match depends on  $SL_{bar}$  and  $Q_2$ . In this case, let us consider  $SL_{bar}=1$  such that  $T_4$  and  $T_2$  transistors are ON, then the path is formed from match line to the ground resulting in a mismatch condition. If  $SL_{bar}=0$  such that  $T_4$  transistor is ON and  $T_2$  transistor is OFF, then the  $ML$  is isolated from the ground which in turn follows the precharge phase output, results in a match condition. In case 2,  $Q_{1bar}=1$  and  $Q_2=0$ , the comparison transistor  $T_1$  is ON and  $T_2$  is OFF, then the  $ML$  output for miss/match depends on  $SL$  and

$Q_{2bar}$ . In case 3,  $Q_{1bar}=0$  and  $Q_2=1$ , both comparison transistors  $T_1$  and  $T_2$  are OFF and ML is isolated from the comparison circuitry. The output, in this case, is a wild match irrespective of search input. The main advantage of proposed TCAM memory design over [11] is an improvement in the stability of TCAM due to use of PMOS access transistors. In addition, it allows storing of ternary logic state value X for wildcard functionality, which is an import logic state useful in IP routing for the network router. The timing diagram shows the ML output of early predict TCAM cell design and is shown in Figure 7.

**Table 1.** Truth table of TCAM

SL	SL <sub>bar</sub>	$Q_{1bar}Q_2$	ML
0	1	01	Miss match
1	0		Match
0	1	10	Miss match
1	0		Match
0	1	11	Wild match
1	0		



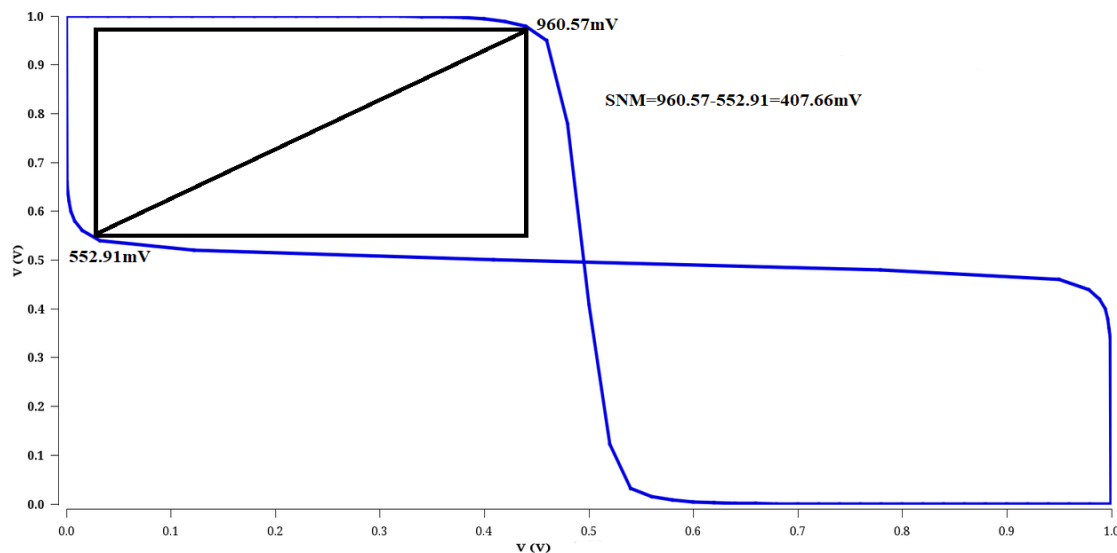
**Figure 7.** Timing wave form of early predict TCAM cell for different ML case

#### 4. SIMULATION RESULTS

To assess the TCAM design for high performance and low power, TCAM architectures are implemented and simulated in Cadence Virtuoso at a supply voltage  $V_{dd}=1V$  and frequency 200 MHz using 45nm technology node. Stability, search delay, and power are the three important parameters which determine the performance of memory design. Static noise margin (SNM) determines the stability of the system. Graphically, stability of the memory design is calculated from the butterfly curve. SNM value is obtained by inserting a largest possible square inside the lobe of the graph, thereby subtracting the diagonal value. The butterfly curve of the proposed TCAM cell design is shown in the Figure 8. The SNM values of the proposed design and conventional design are 407.66mV and 398.43mV respectively. SNM values show that the proposed design has better stability than the conventional TCAM cell design. The Search delay is calculated from the transient response graph. The time difference between the miss/match line output and rising edge of the precharge signal will give the value of search delay. The power consumption of proposed design is calculated from the parameter storage format (PSF) file generated by the Virtuoso tool. Energy metric (EM) is one of the important performance metrics which is calculated by equation 1

$$EM = \frac{power \times delay}{Total\ number\ of\ bits} (j/bit/search) \tag{1}$$

Typical comparison results of conventional and proposed TCAM architecture of 8(words)  $\times$  8(bits) for power, search delay and energy metric are shown in Table 2. Typical simulation results of proposed design show 35.15% lower energy metric, 24.61 % lower search delay and 16.63% lower power consumption. To determine the functionality of TCAM design, Monte Carlo (MC) simulations were also performed for 500 runs with a Gaussian  $3\sigma$  variation on different parameters of the device. CAM stores unique data in their words. Search input word is selected randomly for performing the MC to ascertain performance metric. Search delay and power consumption for proposed TCAM design are averaged in MC simulations for 500 runs as shown in Figure 9, Figure 10. In MC simulation the average search delay obtained is 262pS and average power consumption per bit is 586.102nW. Different process corner simulations like FF, SS, SF, & FS for proposed TCAM design are also performed. Table 3 shows process corner simulation results of power, search delay and energy metric for proposed 8 $\times$ 8 early predict TCAM architecture. It is observed that the worst case of energy metric is 0.222 fJ/bit/search and best case of energy metric is 0.0291fJ/bit/search. For functionality check, the timing simulation waveforms of conventional and proposed TCAM designs are shown in Figure 11 and Figure 12.



**Figure 8.** Static noise margin of proposed early predict TCAM design

**Table 2.** Performance comparison of TCAM

8 $\times$ 8 TCAM	Power consumption ( $\mu$ W)	Search delay (pS)	Energy Metric (fJ/bit/search)
NOR	28.322	349.36	0.15459
Proposed TCAM	23.610	263.36	0.09715

**Table 3.** Process corner simulations for proposed TCAM

Parameters	TT	FF	SS	SF	FS
Power ( $\mu$ W)	23.16	10.46	35.10	39.27	50.24
Delay (pS)	263.36	178.32	405.36	316.16	238.08
Energy Metric (fJ/bit/search)	0.09715	0.0291	0.222	0.1940	0.186

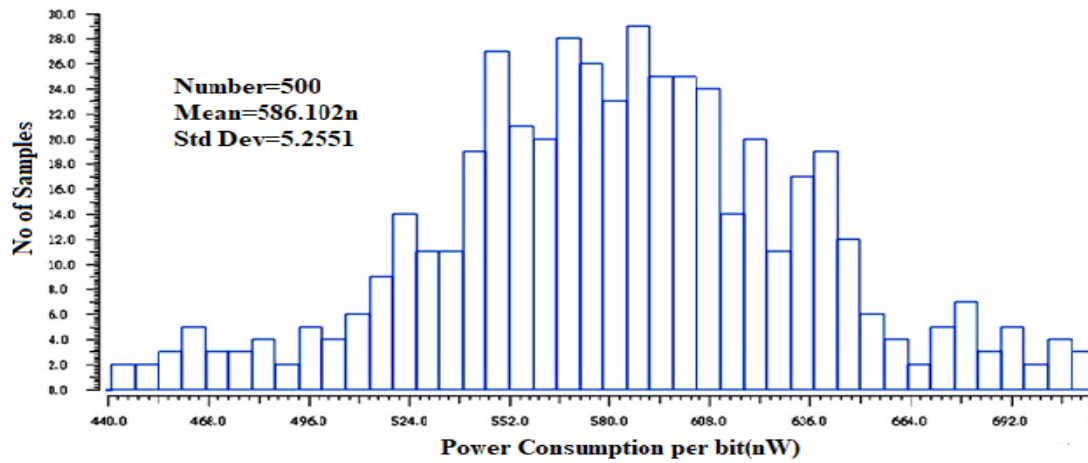


Figure 9. Performance metric of early predict TCAM on MC simulations for 500 runs- histogram of power consumption

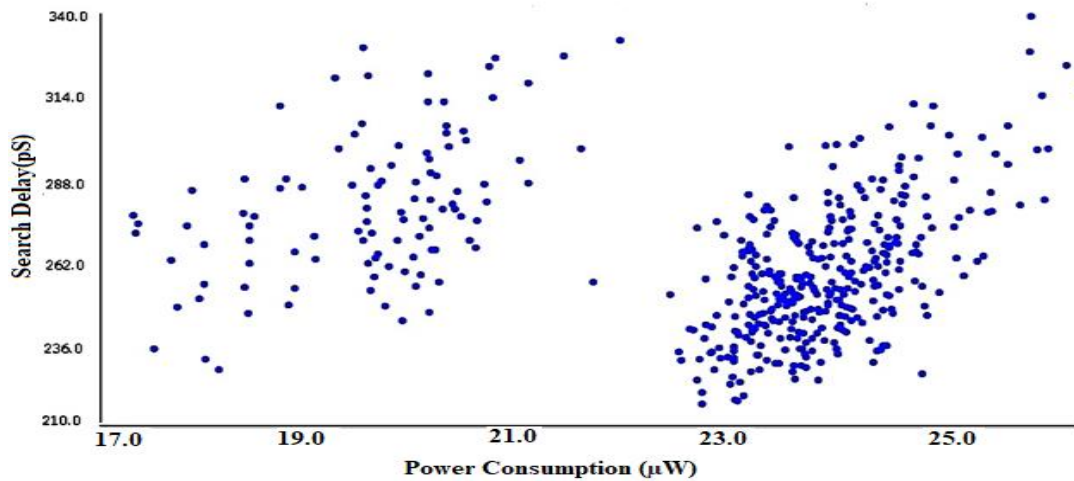


Figure 10. Performance metric of early predict TCAM on MC simulations for 500 runs- power consumption and search delay scattered plot

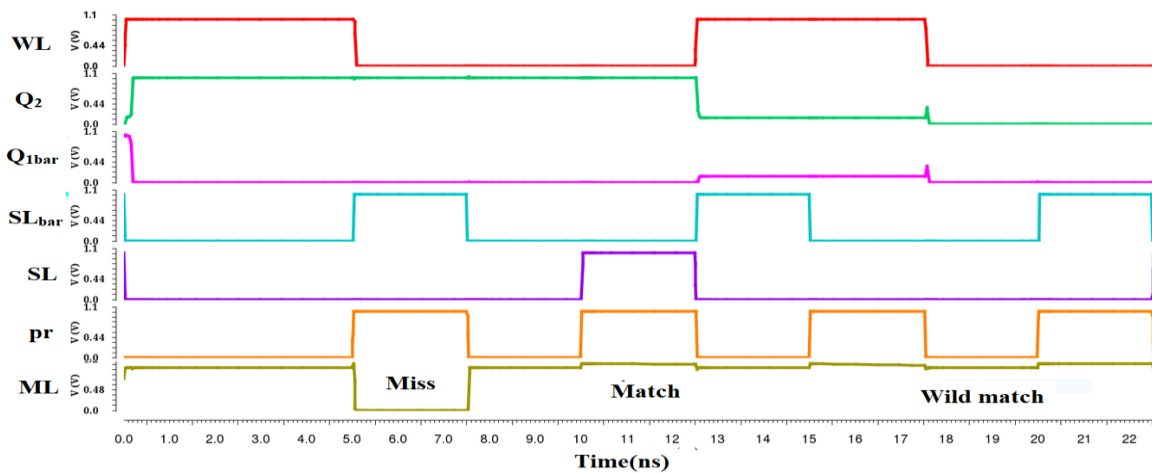
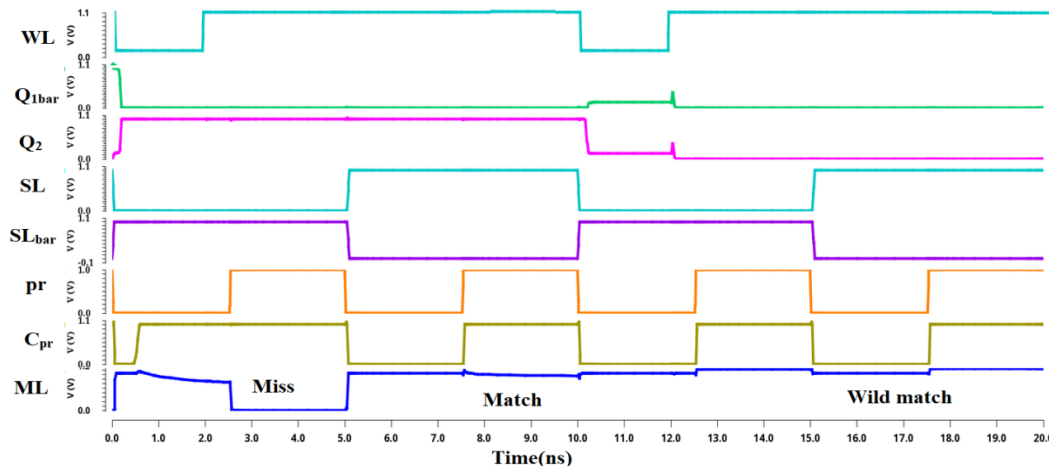


Figure 11. NOR TCAM timing wave form for single bit miss followed by match condition





**Figure 12.** Early Predict TCAM timing wave form for single bit miss followed by match condition

## 5. CONCLUSION

In TCAM search operation, during miss-match, voltage swing in ML contributes to large search delay and power consumption. To overcome these problems and to reduce the miss-matched voltage swing, early predict TCAM architecture is designed. Early predict precharge circuitry helps to reduce the voltage swing on miss-matched ML dynamically, to improve the performance metric. In addition, proposed TCAM has greater stability than conventional TCAM due to use of PMOS access transistors for controlling the static random access memories. Proposed TCAM architecture was simulated in CMOS process at 45nm technology node using Cadence Virtuoso and results are compared with TCAM NOR architecture. When compared with NOR TCAM architecture the proposed design offers 16.6% reduction in power, 24.7% minimization in search delay and it reduces the energy metric by 37.1%. Monte Carlo simulations and Process Corner simulations are also performed to validate the proposed TCAM for low power and low search delay. Simulation results show that the proposed design is useful for high speed, low power network packet forwarding applications with longer word lengths memory designs.

## CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

## REFERENCES

- [1] Mohammad, K., Qaroush, A., Washha, M. and Mohammad, B., "Low-power content addressable memory (CAM) array for mobile devices", *Microelectronics Journal*, 67: 10-18, (2017)..
- [2] Maurya, S.K. and Clark, L.T., "A dynamic longest prefix matching content addressable memory for IP routing", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 19(6): 963-972, (2011).
- [3] Lines, V., Ahmed, A., Ma, P., Ma, S., McKenzie, R., Kim, H. S. and Mar, C., "66 MHz 2.3 M ternary dynamic content addressable memory", In *Memory Technology, Design and Testing, USA*, 101-105, (2000).
- [4] Shin, Y.C., Sridhar, R., Demjanenko, V., Palumbo, P.W. and Srihari, S.N., "A special-purpose content addressable memory chip for real-time image processing", *IEEE Journal of Solid-State Circuits*, 27(5): 737-744, (1992).

- [5] Bremler-Barr, A. and Hendler, D., "Space-efficient TCAM-based classification using gray coding", *IEEE Transactions on Computers*, 61(1): 18-30, (2012).
- [6] Pagiamtzis, K. and Sheikholeslami, A., "Content-addressable memory (CAM) circuits and architectures: A tutorial and survey", *IEEE Journal of Solid-State Circuits*, 4(3): 712-727, (2006).
- [7] Schultz, K.J., "Content-addressable memory core cells A survey", *Integration, the VLSI journal*, 23(2): 171-188, (1997).
- [8] M Ruan, S.J., Wu, C.Y. and Hsieh, J.Y., "Low power design of precomputation-based content-addressable memory", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 16(3): 331-335, (2008).
- [9] Chang, Y.J. and Liao, Y.H., "Hybrid-type CAM design for both power and performance efficiency", *IEEE transactions on very large scale integration (VLSI) systems*, 16(8): 965-974, (2008).
- [10] Chang, Y.J. and Wu, T.C., "Master-Slave Match Line design for low-power content-addressable memory", *IEEE transactions on very large scale integration (VLSI) systems*, 23(9): 1740-1749, (2015).
- [11] Kittur, H.M., "Content Addressable Memory—Early Predict and Terminate Precharge of Match-Line", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 25(1): 385-387, (2017).
- [12] Jothi, D. and Sivakumar, R., "Design and Analysis of Power Efficient Binary Content Addressable Memory (PEBCAM) Core Cells", *Circuits, Systems, and Signal Processing*, 37(4): 1422-1451, (2018).
- [13] Chang, Y.J., Tsai, K.L. and Tsai, H.J., "Low leakage TCAM for IP lookup using two-side self gating", *IEEE Transactions on Circuits and Systems*, 60(6): 1478-1486, (2013).
- [14] Zackriya, M.V. and Kittur, H.M., "Low Energy Metric Content Addressable Memory (CAM) with Multi Voltage Matchline Segments", *Journal of Circuits, Systems and Computers*, 25(02): 1650002, (2016).
- [15] Zackriya, M.V. and Kittur, H. M., "Selective match-line energizer content addressable memory (SMLE-CAM)", (2014).
- [16] Kayal, D., Dandapat, A. and Sarkar, C. K., "Design of a high performance memory using a novel architecture of double bit CAM and SRAM", *International Journal of Electronics*, 99(12): 1691-1702, (2012).
- [17] Sanapala, K. and Sakthivel, R., "Two Novel Subthreshold Logic Families for Area and Ultra Low-Energy Efficient Applications: DTGDI & SBBGDI", *Gazi University Journal of Science*, 30(4): 283-294, (2017).
- [18] Kittur, H.M., "Precharge-Free, Low-Power Content-Addressable Memory", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 24(8): 2614-2621, (2016).
- [19] Mahendra, T.V., Mishra, S. and Dandapat, A., "Self-Controlled High-Performance Precharge-Free Content-Addressable Memory", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 25(8): (2017).
- [20] Arulvani, M. and Ismail, M.M., "Low power FinFET content addressable memory design for 5G communication networks", *Computers & Electrical Engineering*, (2018).