The Scientific Temper (2023) Vol. 14 (2): 430-436
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2023.14.2.30
https://scientifictemper.com/

REVIEW ARTICLE

Content addressable memory for energy efficient computing applications

S. Kumar, M. Santhanalakshmi and R. Navaneethakrishnan

Abstract

Content Addressable Memory (CAM) also known as associate memory isa special kind of semiconductor memory device that works differently from conventional Random Access Memory (RAM). A Content Addressable Memory is a memory unit that matches content over a single clock rather than using addresses. Its inherent parallel search mechanism makes it more advantageous than RAM in terms of speed of search operation. Designers aim to reduce two design characteristics: increasing silicon size and power consumption. As the need for CAM increases, the problem of power consumption also increases. Recent research on CAM is concentrated around diminishing power utilization without forfeiting speed or area. The main reason for the high-power consumption in conventional CAM architecture is devoid of control over the voltage on the Match Line recharge and Search Line precharge. A novel CAM architecture is proposed by removing the necessity of the search line recharge and also by introducing a transistor with gate connected to ML_Eval input that act as a control over the search operation. An Extra transistor with gate connected to Mask_Bar decides whether the circuit can be operated as Ternary Content Addressable Memory (TCAM) or Binary Content Addressable Memory (Bi-CAM). This CAM Architecture is found to be power efficiency up to 50% due to the control over recharged voltage on ML. It is also inferred that the delay associated with the search operation can be reduced to a certain extent. The proposed CAM architecture is simulated using Cadence Virtuoso IC 6.1.6 in General Process Design Kit (GPDK) with 90 nm technology.

Keywords: CAM, Associative Memory, Computing, TCAM, Bi-CAM, Low power Memory, CAM Design, Parallel search, RAM.

Introduction

Content addressable memory is a kind of specialized hardware that allows large lookup tables to be searched in parallel for data-intensive applications (Pagiamtzis and Sheikholeslami, 2006). As the CAM can have a single clock throughput, CAMs are faster. In networking, packet forwarding and classification are accomplished with the help of a Ternary Content Addressable Memory (TCAM), which maintains wild cards (don't care bits) and finds a resemblance for each search operation.ABi-CAM can only deal with "1" and "0" but not don't cares. Figure 1 shows the theoretical representation of a CAM with w*n bits where n is the number of bits in a word and w is the number of words of n bit. The search word of n bits is fed as input into the table of stored data. A match line appears next to each saved word, indicating whether the stored data and search bit are similar (match) or not (mismatch). The match lines are passed into an encoder, creating a binary match location for each match line in the state. In systems where just one match is expected, an encoder is utilized. A priority encoder is employed in the place of a simple encoder when more than one word matches. A hit signal is used to indicate when there is no match location in the CAM. A CAM's main role is to take a search word and fetch its location that matches in the memory.

A normal CAM cell is made up of two SRAM cells, as the largest available CAM chip is around 50% of the largest memory chip presently available (Pagiamtzis and Sheikholeslami, 2006). The search data is loaded first, next by charging all match lines to V_{dd}, for a little moment, they were all in a match condition. Following that, the SL drivers send the word into the respective SL, and Every CAM checks the bit on its own search lines to the bit it has stored. When

Department of Electronics and Communication Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India

***Corresponding Author:** R. Navaneethakrishnan, Department of Electronics and Communication Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India, E-Mail: navaneethakrishnan.ece@kct.ac.in

How to cite this article: Kumar, S., Santhanalakshmi, M., Navaneethakrishnan, R. (2023). Content addressable memory for energy efficient computing applications. The Scientific Temper, **14**(2):430-436.

Doi: 10.58414/SCIENTIFICTEMPER.2023.14.2.30

Source of support: Nil

Conflict of interest: None.

Received: 09/03/2023 **Accepted:** 18/04/2023 **Published**: 30/06/2023

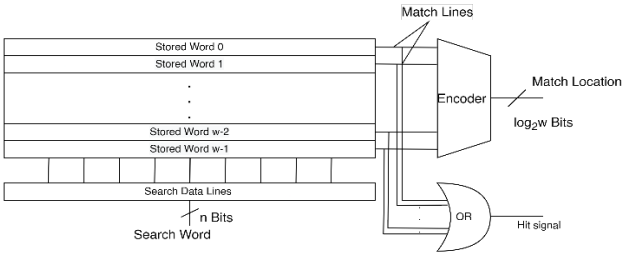


Figure 1: A theoretical representation of a word CAM

any CAM cell line has at least one bit difference, the Match line is connected to ground.

The related work on CAM architecture is discussed in section II. Section III describes the conventional CAM architectures and the proposed CAM architecture is detailed in section IV. The implementation of the proposed CAM and its results are analyzed in section V. Section VI presents the conclusion.

Literature Survey

In hardware, associative memories are implemented using CAM blocks. CMOS based TCAMs consist of two SRAM cells. Paper by Pagiamtzis and Sheikholeslami (2006) investigates the different techniques to reduce power consumption. At the circuit level, two CMOS cells are NOR and NAND cells, and the ML and SL pre-charging technique are discussed. At the architectural level,three architectures are discussed: bank selection, pre-computation, and dense encoding. In paper by Ghofrani et al. (2016), Ternary Content Addressable Memory is designed using memristive technology. The energy consumption is reduced by using an access transistor-free crossbar that realizes memristive TCAM cell structure. The Transistorlevel SPICE level simulations are done in Cadence Virtuosoto calculate the delay and power of the TCAM module. In studies conducted by Irfan, Ullah and Cheung (2019, 2020), a power saving bank selection strategy for binary CAM architecture on FPGA employing flipflops has been developed, but the design complexity will increase for larger CAMs. In Pagiamtzis and Sheikholeslami (2003) paper, a CAM architecture with pipelined MLalong with hierarchical SL mechanism is employed to decrease energy consumption. Ternary CAM of size 1024x144 bit is designed using 180 nm technology. The pipelined matchline scheme reduces the matchline power dissipation by 56%. Pipelined matchlineswithhierarchical searchlines, reduce searchline power dissipation by 60%. In the paper by Imani, Patil and ŠimuniĆ Rosing (2018), a new MultipleAccess SingleCharge (MASC) TCAM architecture is proposed and it has capacityto search TCAM contents up to several times withsingle precharge cycle. This technique resembles to XOR logic, discharges only the matched row and mismatched rows stay charged. This enables several search operations to be performed using the charge of the missed lines.This technique helps in reducing the power consumption despite accuracy ofmatching pattern.As all the previous works concentrate mainly on architectural-level power optimization, a novel technique to reduce the power consumption at circuit level is discussed in this work.It is better to concentrate on the circuit level optimization as it has a high degree of chance to optimize power performance and area.

Conventional cam structure

Conventional CAM Architectures for Binary Content Addressable Memories (Bi-CAM) and Ternary Content Addressable Memories (TCAM) are described. There are two basic CAM cell operations, namely bit storage and bit comparison. A 6T SRAM cell is used for Bit storage and for bit comparison, XNOR based logic is used to find the match/ mismatch of stored bit and search bit.

Static Random Access Memory

The circuit diagram of 6TSRAM Cell which is used to design the CAM cell is shown in Figure 2. The inverter pairs are M1, M2, M3, and M4.

The bitline and its complement are denoted by BL and BL_bar. The Write Line is abbreviated as WL. The read and write operations of 6T SRAM cell are controlled externally by the sense amplifier circuit. The values from BL and BL_bar is written into the cross-coupled inverter pair if WL is set to 1. The cross-coupled inverter pair is written into and read from using access transistors M5 and M6. If WL is set to 0, the data in the memory will remain the same and when set to 1, data is read from the memory, as BL and BL bar acts as the output terminals.

Binary Content Addressable Memory

A Bi-CAM is shown in Figure 3. Figures 3(a) and (b) illustrate NORbased Bi-CAM cell and NANDbased Bi-CAM cell, respectively. Four nMOS transistors with one SRAM cell and one pMOS transistor is used to implement the conventional NOR based CAMcell. The pMOS transistor is used in precharge circuitry. In the event of a mismatch, four nMOS transistors are used to discharge the ML to ground. In CAM, the bit comparison is identical to an XNOR of the

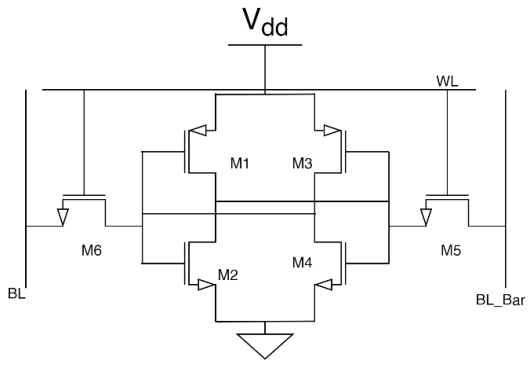


Figure 2: Circuit Diagram of 6T SRAM Cell.

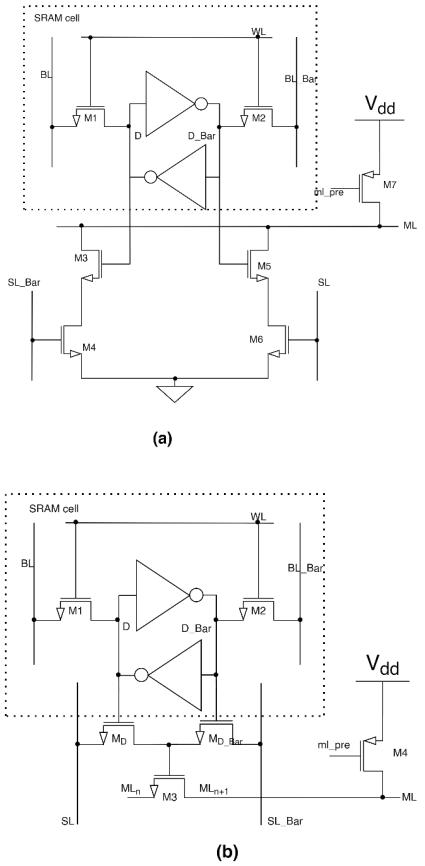


Figure 3: Conventional Bi-CAM Structure (a) NOR based cell (b) NAND based cell

search bit and the storage bit, but the NOR and NAND cells implement it differently.

The NOR based CAM cell (Figure 3(a)) uses four transistors from M_3 to M_6 to compare the search bit on search line, SLand the stored bit, D. The discharge channel of a dynamic XNOR circuit having inputs SL and D is achieved by these transistors.M3/M4 and M5/M6 transistor pairs establish a discharge path for match line to ground. At least one of the paths is activated, connecting ML to the ground, when SL and D are mismatched.

Both pulldown pathways are inhibited when SL and D are matched, and ML is detached from the ground. The NOR character of CAMs becomes obvious when numerous cells are joined in parallel to form a data word by shortening the match line of a cell to the match line of adjacent cells. The discharge paths are connected parallel, much like the pulldown network of CMOS NOR gate, such that only if every bit in the word matches the stored word, then itisa match condition on that particular match line and is disconnected from the ground.

Similar to a NOR based CAM cell, a NAND Based CAM cell (Figure 3(b)) is constructed using one SRAM cell, one pMOS Transistor and three nMOS transistors. Using the three comparison transistors M_D, M_{D_Bar}, and M_3, a NAND type CAM cell accomplishes the evaluation of stored data, D, with relevant search bits on search lines, SL. When numerous NAND cells are serially coupled, the NAND idea of this cell becomes noticeable. In this scenario, the MLn and MLn+1 nodes are connected to generate a word.A serial nMOS chain of all Mi transistors is analogous to the pulldown network of a CMOS NAND gate (Pagiamtzis and Sheikholeslami, 2006).

Ternary Content Addressable Memory

The TCAM cell is designed using the circuit schematic of 6T – SRAM cell (Figure 2).Usually, the binaryCAM cells are used to storea logic "0" orlogic "1".Notwithstanding to logic "0" or a logic "1", to store an "X" value, Ternary Content Addressable Memory (TCAM) is used.

The don't care values are represented by the letter "X", which can be used to represent both "0" and "1," permitting for a wildcard operation. A "X " value placed in a TCAM produces a similarity independent of the search data, which is referred to as wild card operation. The second SRAM cell is added to store a don't care value in a NOR cell, as shown in Figure 4(a). As shown in Figure 4(b), a NAND-based Bi-CAM cell can be converted to TCAM by adding an inverter pair as a mask bit store at node M.

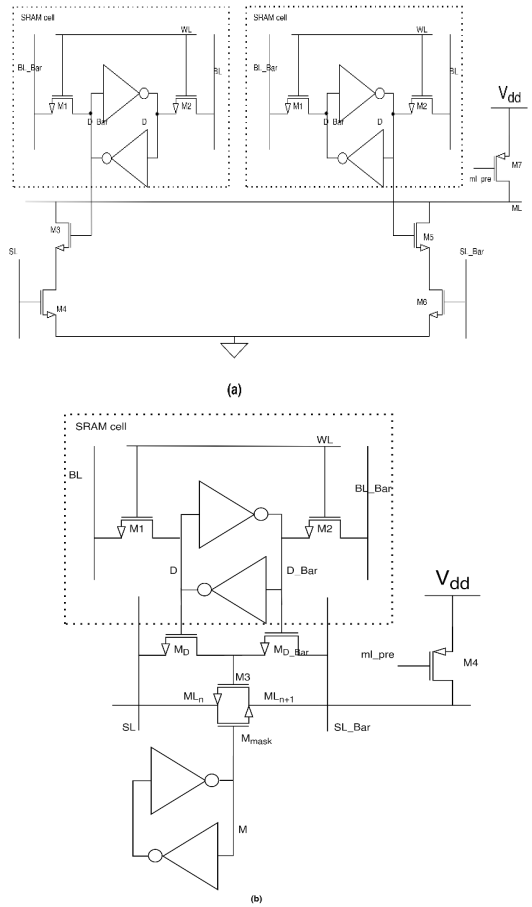


Figure 4: Conventional TCAM Structure (a) NOR type TCAM (b) NAND type TCAM

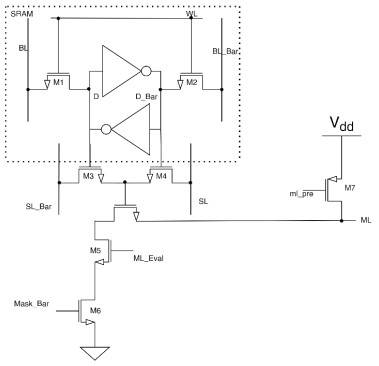


Figure 5: Proposed CAM Structure

In all conventional CAM architectures, CAM search operation consists of three phases: SL precharge, ML precharge, and ML evaluation. So, a novel CAM structure is proposed by removing the necessity of SL precharge by introducing the two new transistors with inputs, namely ML_Eval and Mask_Bar. Thus, the proposed CAM architecture has only two phases: ML precharge and ML evaluation. The XNOR based bit comparison logic is retained from the conventional CAM architecture.

Proposed cam structure

The proposed CAM structure as shown in Figure 5 is implemented using

- 6T SRAM cell (Figure 2) as in conventional CAM structure.
- Five nMOS transistors, three nMOS transistors and
- one pMOS transistor

As in conventional CAM, the pMOS transistor is used in the precharge circuitry, and the nMOS transistors are used in the ML's discharge path in the event of a mismatch. The two additional transistors introduced are M5 and M6 with ML_Eval and Mask_bar inputs, respectively. In Figure 5, the data bit and its complement are stored in the SRAM cell as D and D_Bar. The search line and its complement are SL and SL_Bar. The match line ML is precharged to V_{dd} with ml pre. The signal ML Eval is used to control the search operation.

As ml_pre was set to "0", the pMOS transistor will turn on and act as a short between Vdd and ML during the precharge phase. During this period, ML_Eval and Mask_barmust be set to "0" to avoid short between V_{dd} and Ground, thus significantly reducing power.

During the evaluation phase, ml pre is set to 1, making the pMOS transistor inactive. ML_Eval is set to 1 to allow discharge to ground in case of difference in stored and search bits. The Mask_bar signal in this proposed CAM structure determines whether the circuit can be operated as TCAM or Bi-CAM. Here if Mask_bar = 1, the circuit operates as Bi-CAM, else if Mask_bar = 0, the circuit will work as a TCAM.

In the proposed CAM structure, when Mask_Bar = 1, M_6 is turned on, providing a path for ground during mismatch. In case match occurs, ML holds the precharged value though M_6 is ON. Thus, the CAM operation is similar to a Bi-CAM. In case, Mask_bar is "0" $M_6$is turned OFF so that the Match line is held at V_{dd}, though match or a mismatch occurs. This avoids the necessity for the evaluation phase, which was mandatory in Bi-CAM and significantly reduces the power consumption. So, in a TCAM, the match line is held high always irrespective of the Search line (SL) and Data line (D). The operation of the proposed CAM Cell is explained in a truth table given in Table 1.

Whenever the Mask_Bar is "0", the output ML is always high irrespective of the inputs on D and SL, since it is working as a TCAM. But if mask_bar is "1", the circuit operates as Bi-CAM and ML will be high only when matchoccur (D and SL are same) and low for mismatch (D and SL are not same).

Implementation and results

This section discusses the implementation of conventional CAM and Proposed CAM cellsusingGPDK90nm technology.

SRAM cell

The implemented schematic of the 6TSRAM cell (Figure 2) and its transient analysis are as shown in Figures 6 and 7, respectively.

From Figure 7, it is clear that when the bit value of WL is 1. The SRAM Cell's D output is equal to the BL input bit.

Conventional CAM

The schematic of the Bi-CAM cell (Figure 3a) is implemented and shown in Figure 8. The transient analysis of the implemented schematic of the Bi-CAM cell is shown in Figure 9.

From Figure 9, ml_pre is made "0" for some duration to precharge the ML to V_{dd}. Once ML reaches V_{dd}, ml_pre is deasserted to turn off the pMOS transistor. SL is the search line bit which has "0" for the first 5 ns and "1" for next 5 ns. The D output of the SRAM cell is at "1" for the whole 10ns. The ML signal shows a Mismatch for the first 5ns (ML=0) and Match for next 5ns(ML=1).

Proposed CAM

The schematic of the Proposed CAM cell (Figure 5) is implemented and shown in Figure 10. The transient analysis of the implemented proposed CAM cell is in Figure 11.

Table 1: Truth table of proposed CAM Structure

Input			*Output*
D	SL	Mask_bar	ML
0	0	0	1
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	1

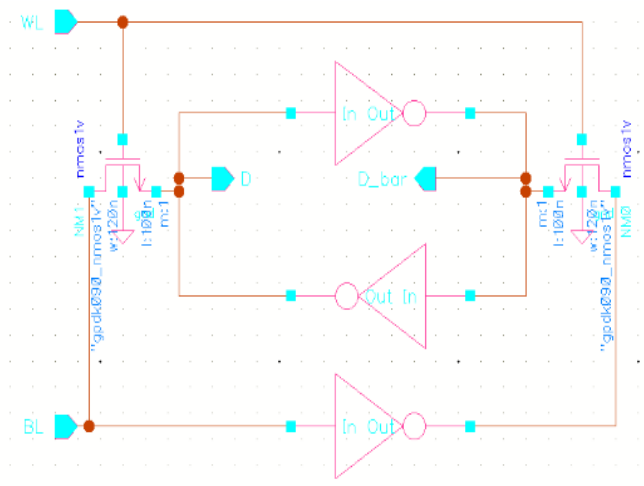


Figure 6: Implemented schematic of 6TSRAM cell

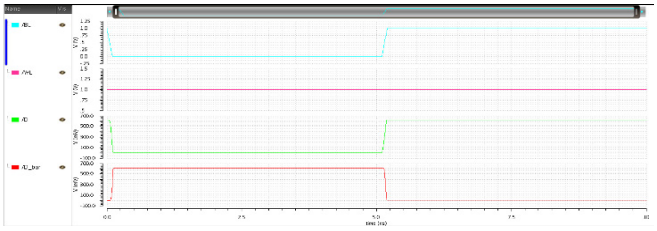


Figure 7: Transient analysis of 6TSRAM cell

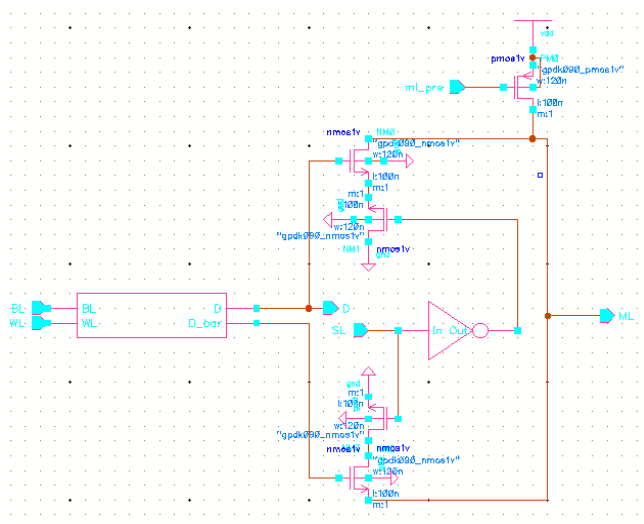


Figure 8: Implemented schematic of conventional cam cell

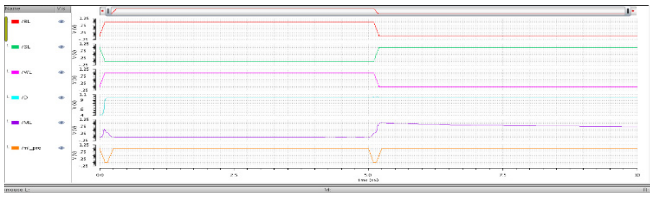


Figure 9: Transient analysis of conventional cam cell

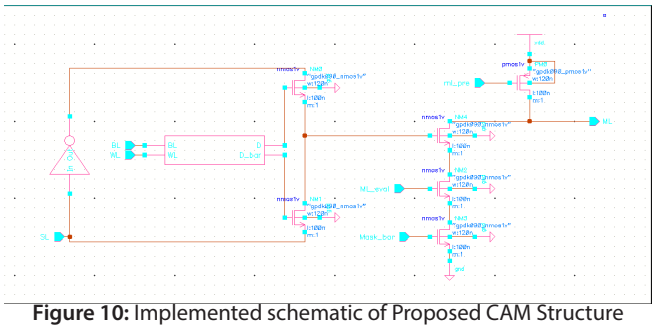


Figure 10: Implemented schematic of Proposed CAM Structure

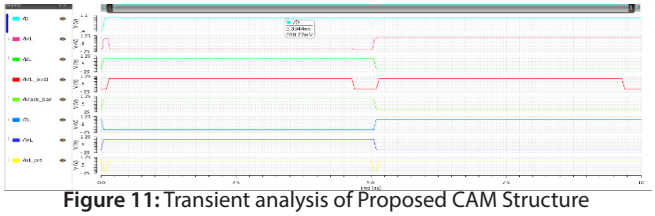


Figure 11: Transient analysis of Proposed CAM Structure

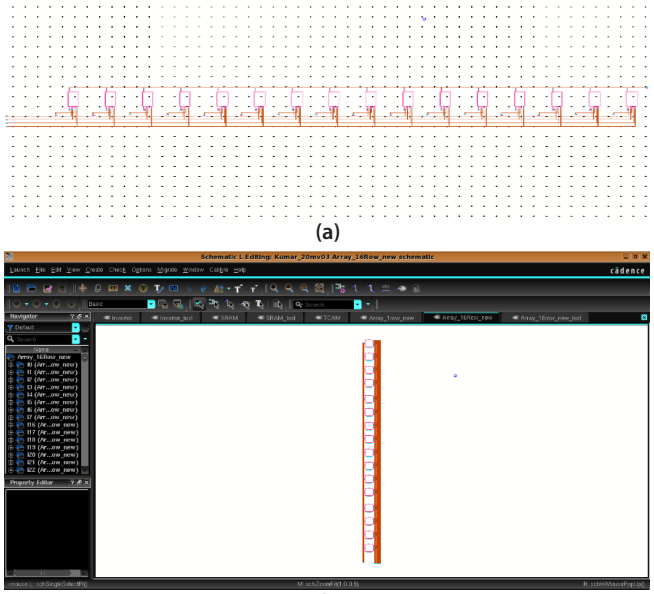


Figure 12: Array of CAM Cells (a) 1*16 CAM Cells (b) 16*16CAM Cells

In Figure 11, when ml_pre is made "0" for some duration to precharge the ML to V_{dd}, ML_Eval is also set to "0" to disconnect its path from ground. Once ML reaches V_{dd}, ml_pre is de-asserted to deactivate the pMOS transistor and ML_Eval is made to "1" to activate the nMOScircuit. SL is the search bit with "0" for the first 5 ns and "1" for the next 5 ns. The D output of the SRAM cell is "1" for the whole 10ns. The ML signal shows mismatch for the first 5ns(ML=0) and match for next 5ns(ML=1).

A 1x16 CAM array and 16x16 CAM structure constructed using 1x16 CAM are shown in Figures 12(a) and 12(b), respectively. Both the CAM arrays are designed using the proposed CAM structure shown in Figure 5.

The comparison of the Proposed CAM and conventional CAM structures in terms of average power and delay associated with the search operation is tabulated in Table 2.

Table 2: The comparison table of the proposed CAM Structure with existing CAM structure in terms of power and delay

Configuration Parameter	*Existing Bi-CAM [8]*	*Existing TCAM [8]*	*Conventional CAM [1]*	*Single Proposed CAM cell*	*An array of 1*16 using proposed CAM*	*An array of 16*16 using Proposed CAM*
Average Power	48.84uW	97.27uW	1.074uW	492.9nW	7.558uW	152.8uW
Delay (ml_pre - ML)	NA	NA	103.3ps	12.39ps	19.78ps	14.40ps
Delay (ML_Eval-ML)	NA	NA	NA	4.638ns	5.204ns	14.40ps
Delay (SL-ML)	46.55ps	8.032ns	5.059ns	5.037ns	117.0ps	10.51ps
No. of Transistors	14	20	15	16	256	4096

Though two additional transistors are introduced, their operation produces 54.18% of power reduction in proposed CAM cell compared to the conventional CAM cell. Monte Carlo simulation was performed on the proposed CAM structures to validate the simulated results.

Monte Carlo Simulation

A Monte Carlo simulation is a model used to predict the probability of different outcomes when the intervention of random variables is present.It can be used to show how risk and uncertainty affect prediction and forecasting models. Monte Carlo simulations are used in a range of industries, including engineering, supply chain, and science.It is based on giving many values to an uncertain variable to generate different outcomes, which are then averaged to obtain an estimate.

Statistical distributions are used in Monte Carlo analysis. Mismatching and process variation are realistically simulated. Every parameter is calculated randomly according to a statistical distribution model during each simulation run. Monte Carlo has the disadvantage of requiring a large number of simulations to achieve acceptable results. The number of runs should be at least 250 to have a significant sample. The minimum number of simulations is not a constraint, but the test is more significant for a greater number of simulations.So, the Monte Carlo simulation are carried out for proposed CAM cell and array of CAM for 1000 runs to obtain accurate values.

Figure 13 (a-c) shows the Monte Carlo simulation for single CAM cell in terms of its average power, delay between ml_pre and ML and delay between SL and ML, respectively.

The average power, delay between ML_Eval and ML and delay between ml_pre and ML of the 1x16 CAM array are as shown in Figures 14 (a-c), respectively.

The average power, delay between ml_pre and ML and delay between ML and SL of the 16x16 CAM array are shown in Figure 15 (a-c).

The Monte Carlo simulation of the proposed CAM cell and array of CAMs is performed for 1000 Runs. From the Monte Carlo simulations, it is evident that the mean power consumption and mean delay of the circuits are found to be identical with tabulated values.

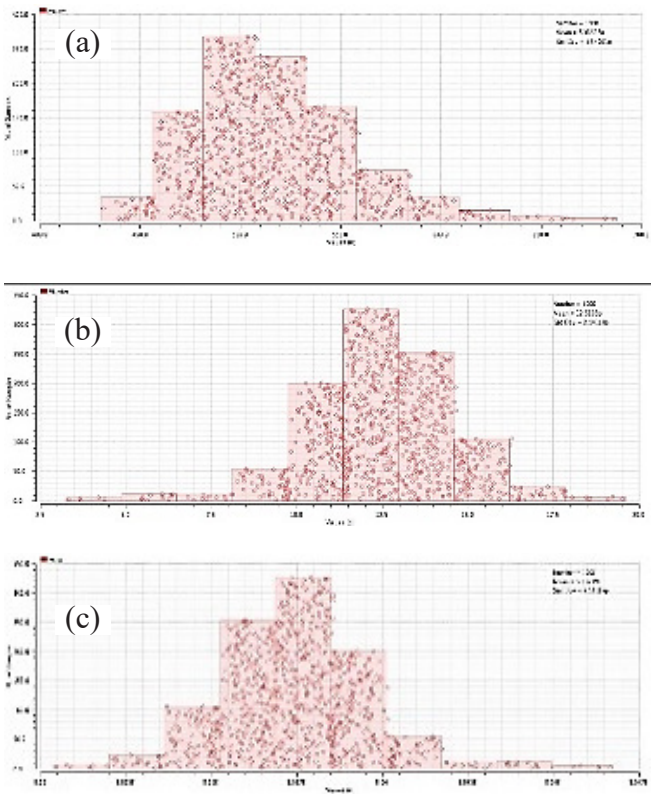
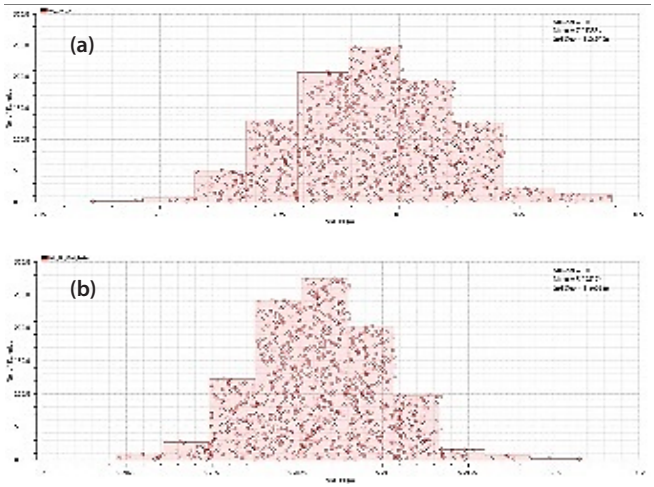


Figure 13: Monte Carlo simulation for proposed single CAM Cell (a)Average Power, (b) Delay between ml_pre and ML, (c) Delay between SL and ML



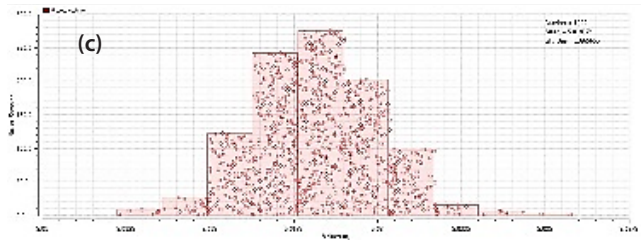


Figure 14: Monte Carlo simulation for array (1*16) of proposed CAM Cell: (a) Average Power, (b) Delay between ML_Eval and ML, (c) Delay between ml_pre and ML.

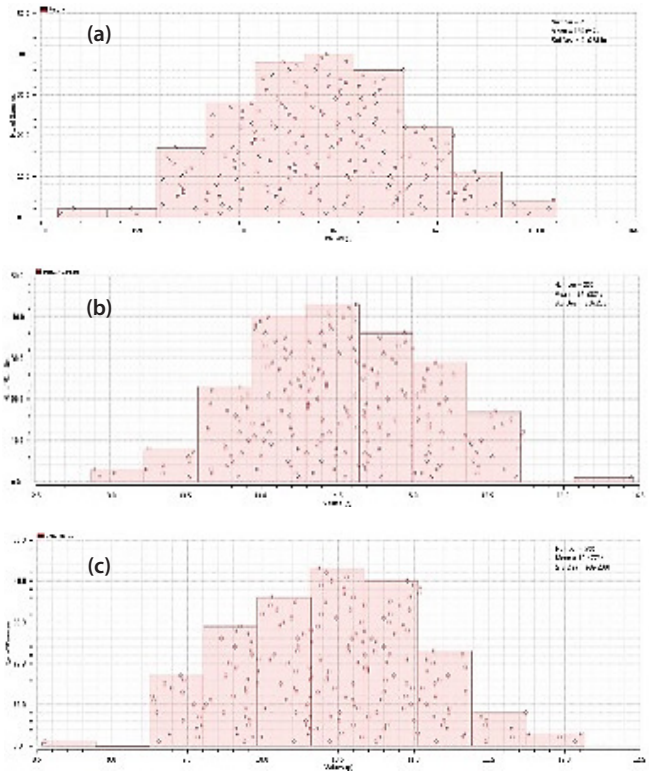


Figure 15: Monte Carlo simulation for array (16*16) of proposed CAM cell: (a) Average Power, (b) Delay between ml_pre and ML, (c) Delay between SL and ML.

Conclusion

Computing with memory has shown significant energy efficiency using Associative memory. CAMs are employed in a variety of applications like networking hardware, CPUs and can be integrated with the computing structures like multipliers and adders that are widely used in image processing applications. This paper proposes ultralow energy CAM that significantly reduces energy consumption. The Content Addressable Memory architecture is designed and simulated in Cadence Virtuoso IC 6.1.6 and analyzed in terms of power and delay incurred. It is observed that power consumed by the proposed CAM is reduced by 54.18% and delay is reduced by 88.0% when compared to the existing CAM (Pagiamtzis and Sheikholeslami, 2006). The power consumed by 1x16 and 16x16 array TCAM is also found to be 7.55 and 152.87uW, respectively. In conventional CAM architecture, Binary CAM and Ternary CAM is implemented in two circuits but in this proposed CAM structure, it is possible to configure the circuit to act as both Bi-CAM and TCAM with the additional transistor input pin Mask_Bar. Thus, the proposed CAM structure can be used to reduce the energy consumption in the associative memories.

References

Ghofrani, A., Rahimi, A., Lastras-Montaño, M. A., Benini, L., Gupta, R. K., & Cheng, K. T. (2016). Associative memristive memory for approximate computing in gpus. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 6(2), 222-234. DOI: 10.1109/JETCAS.2016.2538618.

Imani, M., Patil, S., & Rosing, T. Š. (2016). Approximate computing using multiple-access single-charge associative memory. IEEE Transactions on Emerging Topics in Computing, 6(3), 305-316. DOI: 10.1109/TETC.2016.2565262.

Irfan, M., Ullah, Z., & Cheung, R. C. (2019, December). High Performance Power-efficient Gate-based CAM for Reconfigurable computing. In 2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN) (pp. 327-331). IEEE. DOI: 10.1109/MSN48538.2019.00068.

Irfan, M., Ullah, Z., Chowdhury, M. H., & Cheung, R. C. (2020). RPE-TCAM: Reconfigurable power-efficient ternary content-addressable memory on FPGAs. IEEE Transactions on Very Large-Scale Integration (VLSI) Systems, 28(8), 1925-1929., DOI: 10.1109/TVLSI.2020.2993168.

Pagiamtzis, K., & Sheikholeslami, A. (2003, September). Pipelined match-lines and hierarchical search-lines for low-power content-addressable memories. In Proceedings of the IEEE 2003 Custom Integrated Circuits Conference, 2003. (pp. 383-386). IEEE., DOI: 10.1109/CICC.2003.1249423.

Pagiamtzis, K., & Sheikholeslami, A. (2006). Content-addressable memory (CAM) circuits and architectures: A tutorial and survey. IEEE journal of solid-state circuits, 41(3), 712-727. DOI: 10.1109/JSSC.2005.864128.