

High Performance Area Efficient Low Power CAM Architecture Design

Remya .M.M¹, Amrutha.E²

PG Scholar Dhanalakshmi Srinivasan College of Engineering, Coimbatore, India
Asst Professor Dhanalakshmi Srinivasan College of Engineering, Coimbatore, India

Abstract: Content addressable memory (CAM) is a memory that implements the look up table function in a single clock cycle using dedicated comparison circuitry. CAM's are composed of conventional semiconductor memory (usually SRAM) with added comparison circuitry that enables a search operation to complete in a single clock cycle. Content addressable memory offers high-speed search function in a single clock cycle. Due to its parallel match-line (ML) comparison, CAM is power-hungry. The main CAM design challenge is to reduce power consumption associated with the large amount of parallel active circuit without sacrificing speed or memory capacity. In this paper, an effective gated-power technique is applied to 9T CAM cell to reduce the peak and average power consumption is been designed and it enhances the robustness against process variations. A feedback loop is employed to auto-turn off the power supply to the comparison elements and hence reduces the average power consumption. The design can work at a reduced supply voltage.

Index Terms: CMOS, content addressable memory (CAM), match-line.

I. Introduction

Content-Addressable Memory (CAM) is a type of solid-state memory in which data are accessed by their contents rather than physical locations. It receives input search data, i.e., a search word, and returns them address of a similar word that is stored in its data-bank^[1]. Content addressable memory (CAM) compares input search data against a table of stored data, and returns the address of the matching data. CAMs have a single clock cycle throughput making them faster than other hardware- and software-based search systems. CAMs can be used in a wide variety of applications requiring high search speeds. In general, a CAM has three operation modes: READ, WRITE, and COMPARE, among which "COMPARE" is the main operation as CAM rarely reads or writes^[4]. Figure.1 shows a simplified block diagram of a CAM core with an incorporated search data register and an output encoder.

II. Cam Architecture Design

Content-addressable memory (CAM) is a special type of computer memory used in certain very high speed searching applications. It is also known as associative memory, associative storage, or associative array. Unlike standard computer memory (random access memory or RAM) in which the user supplies a memory address and the RAM returns the data word stored at that address, a CAM is designed such that the user supplies a data word and the CAM searches its entire memory to see if that data word is stored anywhere in it. If the data word is found, the CAM returns a list of one or more storage addresses where the word was found. Thus, a CAM is the hardware embodiment of what in software terms would be called an associative array.

Content-addressable memories (CAMs) are hardware search engines that are much faster than algorithmic approaches for search-intensive applications. CAMs are composed of conventional semiconductor memory (usually SRAM) with added comparison circuitry that enables a search operation to complete in a single clock cycle. The two most common search-intensive tasks that use CAMs are packet forwarding and packet classification in Internet routers

It starts a compare operation by loading an n-bit input search word into the search data register. The search data are then broadcast into the memory banks through n pairs of complementary search-lines (SL) and directly compared with every bit of the stored words using comparison circuits. Each stored word has a matched line (ML) that is shared between its bits to convey the comparison result. Location of the matched word will be identified by an output encoder, as shown in Figure 1. During a pre-charge stage, the MLs are held at ground voltage level while both SL and \sim SL are at V_{DD} . During evaluation stage, complementary search data is broadcast to the SLs and \sim SLs. When mismatch occurs in any CAM cell (for example at the first cell of the row $D = "1"; \sim D = "0"; SL = "1"; \sim SL = "0"$), transistor P3 and P4 will be turned on, charging up the ML to a higher voltage level.

A sense amplifier (MLSA) is used to detect the voltage change on the ML and amplifies it to a full CMOS voltage output. If mismatch happens to none of the cells on a row, no charge up path will be formed and the voltage on the ML will remain unchanged, indicating a match. Since all available words in the CAMs are compared in parallel, result can be obtained in a single clock cycle. Hence, CAMs are faster than other hardware- and software-based search systems [1]. They are therefore preferred in high-throughput applications such as network routers and data compressors. However, the full parallel search operation leads to critical challenges in designing a low-power system for high-speed high-capacity CAMs [1]. The power hungry nature due to the high switching activity of the SLs and the MLs; and a huge surge-on current (i.e., peak current) occurs at the beginning of the search operation due to the concurrent evaluation of the MLs may cause a serious IR drop on the power grid, thus affecting the operational reliability of the chip [1]. As a result, numerous efforts have been put forth to reduce both the peak and the total dynamic power consumption of the CAMs [2]-[8]. A parity-bit is introduced to boost the search speed of the parallel CAM with less than 1% power and area overhead. A power-gated ML sense amplifier is used to improve the performance of the CAM ML comparison in terms of power and robustness. It also reduces the peak turn-on current at the beginning of each search cycle.

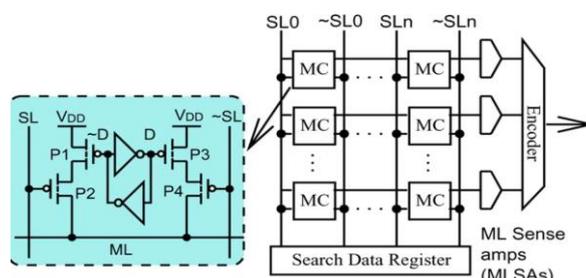


Fig 1: Block diagram of a conventional CAM.

A CAM is designed to search its entire memory in a single operation. It is much faster than RAM in virtually all search applications. There are cost disadvantages to CAM however. Unlike a RAM chip, which has simple storage cells, each individual memory bit in a fully parallel CAM must have its own associated comparison circuit to detect a match between the stored bit and the input bit. Additionally, match outputs from each cell in the data word must be combined to yield a complete data word match signal. The additional circuitry increases the physical size of the CAM chip which increases manufacturing cost. The extra circuitry also increases power dissipation since every comparison circuit is active on every clock cycle. Consequently, CAM is only used in specialized applications where searching speed cannot be accomplished using a less costly method.

Content-addressable memory is often used in computer networking devices. For example, when a network switch receives a data frame from one of its ports, it updates an internal table with the frame's source MAC address and the port it was received on. It then looks up the destination MAC address in the table to determine what port the frame needs to be forwarded to, and sends it out on that port. The MAC address table is usually implemented with a binary CAM so the destination port can be found very quickly, reducing the switch's latency.

III. Search Speed Boost Using A Parity Bit

An auxiliary bit is introduced to boost the search speed of the CAM at the cost of less than 1% area overhead and power consumption. This is similar to the existing Pre-computation schemes but in fact has a different operating principle.

Pre-Computation CAM Design

The pre-computation CAM uses additional bits to filter some mismatched CAM words before the actual comparison. These extra bits are derived from the data bits and are used as the first comparison stage. For example, in Fig. 2(a) number of "1" in the stored words are counted and kept in the Counting bits segment. When a search operation starts, number of "1"s in the search word is counted and stored to the segment on the left of Figure.2 (a). These extra information are compared first and only those that have the same number of "1"s (e.g., the second and the fourth) are turned on in the second sensing stage for further comparison.

The mentioned pre-computation and all other existing designs share one similar property. The ML sense amplifier essentially has to distinguish between the matched ML and the 1-mismatch ML. This makes CAM designs sooner or later face challenges since the driving strength of the single turned-on path is getting weaker after each process generation while the leakage is getting stronger. This problem is usually referred to as I_{on}/I_{off} . The auxiliary bit that can concurrently boost the sensing speed of the ML and at the same time improve the as I_{on}/I_{off} of the CAM by two times.

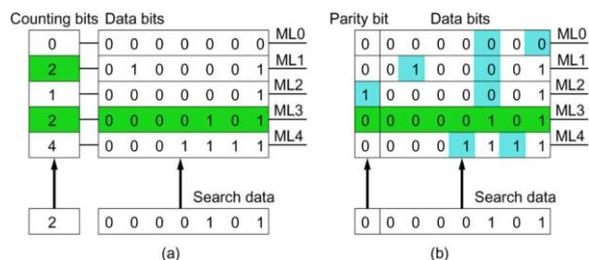


Fig 2: Conceptual view of (a) pre-computation CAM and (b) parity-bit based CAM

Parity Bit Based CAM

The parity bit based CAM design is shown in Figure 2(b) consisting of the original data segment and an extra one-bit segment, derived from the actual data bits. We only obtain the parity bit, i.e., odd or even number of “1”s. The obtained parity bit is placed directly to the corresponding word and ML. Thus the new architecture has the same interface as the conventional CAM with one extra bit. During the search operation, there is only one single stage as in conventional CAM. Hence, the use of this parity bits does not improve the power performance. However, this additional parity bit, in theory, reduces the sensing delay and boosts the driving strength of the 1-mismatch case (which is the worst case) by half. In the case of a matched in the data segment (e.g., ML3), the parity bits of the search and the stored word is the same, thus the overall word returns a match. When 1 mismatch occurs in the data segment (e.g., ML2), numbers of “1”s in the stored and search word must be different by 1. As a result, the corresponding parity bits are different. Therefore now we have two mismatches (one from the parity bit and one from the data bits). If there are two mismatches in the data segment (e.g., ML0, ML1, or ML4), the parity bits are the same and overall we have two mismatches. With more mismatches, we can ignore these cases as they are not crucial cases. The sense amplifier now only has to identify between the 2-mismatch cases and the matched cases. Since the driving capability of the 2-mismatch word is twice as strong as that of the 1-mismatch word, the design greatly improves the search speed and the Ion/Off ratio of the design. Figure 3 shows the 1-mismatch ML transient waveforms of the original and the proposed architecture during the search operation.

IV. Gated-Power MI Sense Amplifier Design

The 10T CAM architecture is depicted in Figure 3. The comparison circuitry of the 10T CAM cell consists of 4 transistors. The CAM cells are organized into rows (word) and columns (bit). Each cell has the same number of transistors as the conventional P-type NOR CAM and use a similar ML structure. However, the “COMPARISON” unit, i.e., transistors M1-M4, and the “SRAM” unit, i.e., the cross-coupled inverters, are powered by two separate metal rails, namely VDDML and the VDD, respectively. The VDDML is independently controlled by a power transistor (Px) and a feedback loop that can auto turn-off the ML current to save power. The purpose of having two separate power rails of VDDML and VDD is to completely isolate the SRAM cell from any possibility of power disturbances during COMPARE cycle.

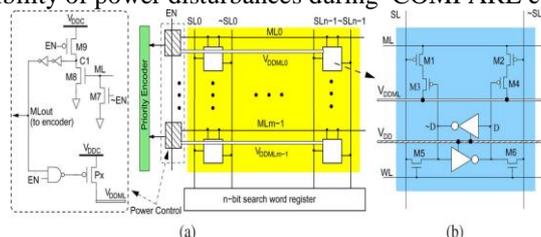


Fig3: (a) Gated power CAM architecture with 10T structure. (b) Each CAM cell is powered by two power rails, VDDML for the compare transistors, VDD for the SRAM transistors

The rail of a row is connected to the power network via a p-MOS device, which is used to limit the transient current. All the cells of a row will share the limited current offered by the transistor, despite whatever number of mismatches. As shown in Figure 4(a), the gated-power transistor Px, is controlled by a feedback loop, denoted as “Power Control” which will automatically turn off Px once the voltage on the ML reaches a certain threshold. At the beginning of each cycle, the ML is first initialized by a global control signal EN. At this time, signal EN is set to low and the power transistor Px is turned OFF. This will make the signal ML and C1 initialized to ground and VDD, respectively. After that, signal EN turns HIGH and initiates the COMPARE phase. If one or more mismatches happen in the CAM cells, the ML will be charged up. Interestingly, all the cells of a row will share the limited current offered by the transistor Px, despite whatever number of mismatches. When the voltage of the ML reaches the threshold voltage of transistor M8 (i.e., Vth8), voltage at node C1 will be pulled down. After a certain but very minor delay, the NAND2 gate will be toggled and thus the power transistor Px is turned off again. As a result, the ML is not fully charged to VDD, but limited to some

voltage slightly above the threshold voltage of M8, V_{th} . Thus the new CAM architecture offers both low-power and high-speed operation.

V. Cam Architecture With 9t Cam Cell

The 9T CAM cell variants shown in fig 4 uses only 9-transistors compared to the previous 10-T CAM cell. The bitcomparison uses pass transistors. The cells use SRAM-based data-storage cells. The SRAM storage and access transistors account for six of the cell transistors.

The cells are shown using SRAM-based data-storage cells. For simplicity, the figure omits the usual SRAM access transistors and associated bitlines. The SRAM storage and access transistors account for six of the cell transistors.

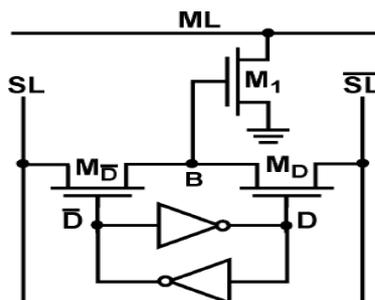


Fig 4: 9T CAM cell

However, the NOR property of this cell is apparent when multiple cells are connected in parallel to form a CAM word by shorting the ML of each cell to the ML of adjacent cells. When the bit comparison succeeds in this cell, one of the transistor paths between and is ON. Thus, when multiple cells are shorted together these transistor paths appear in series just as in the pulldown network of a CMOS NAND gate.

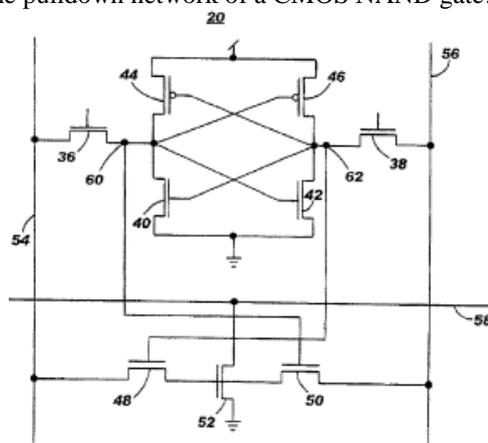


Fig 5: Schematic of 9T CAM Cell

The CAM cell with 9T as shown in fig.5, includes six transistors T1-T6 denoted as reference numerals 36, 38, 40, 42, 44 and 46, which form a six-device SRAM (static random access memory) cell. The SRAM can be written to or read from in the conventional manner. In addition, three transistors, T7-T9, designated as 48, 50 and 52, form a compare portion of the CAM cell.

The transistors T1 and T7, 36 and 48, are coupled to a first search line 54, and the transistors T2 and T8, 38 and 50, are coupled to a second search line 56. The first and second search lines 54 and 56 may also be referred to as true and complement signals, respectively. The transistor T9 52 is coupled to ground and a match line 58. The transistors T1, T3, T5 and T8, 36, 40, 44, and 50, are coupled together at node "A" 60. The transistors T2, T4, T6 and T7, 38, 42, 46 and 48, are coupled together at node "B" 62. During a CAM search operation, the match line 58 is pre-charged high, while the first and second search lines 54 and 56 are pre-charged low. Data applied to the search lines 54 and 56 will drive either the first search line 54 or the second search line 56 high in accordance with the search data. Depending on the data contained in the particular cell, the corresponding match line 58 will either remain in its pre-charged state (e.g., search data matches the cell contents), or discharge to ground (e.g., search data mismatches the cell contents). For example, if the CAM cell 20 contains a logical "0," node "A" 60

Will be loW and node “B”62 Will be high prior to a search of the cell. If a logical “1” is applied to the searchlines 54 and 56, the first search line 54 Will drive high, Whilethe second search line 56 remains in its pre-charged low state.

With both a gate and source of transistor T7 48 high, the matchline 58 Will be discharged through the transistor T9 52 to indicate a mismatch. On the other hand, When a logical “0” is applied to the first and second search lines 54 and 56, the first search line 54 remains loW and the second search line 56 drives high When searched. Under these conditions, transistor T.sub.9 52 cannot be turned on and the match line 58 will remain high, indicating a match has occurred

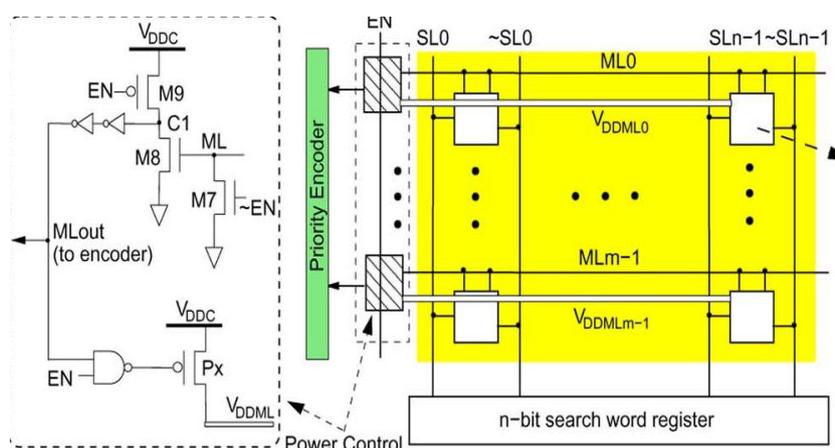


Fig6 (a) Gated power CAM architecture with 9T structure. (b) Each CAM cell is powered by two power rails, VDDML for the compare transistors, VDD for the SRAM transistors

The comparison circuitry of the Gated power CAM architecture with 9T structure as shown in fig.6 consists of 3 transistors. The CAM cells are organized into rows (word) and columns (bit). Each cell has the same number of transistors and use a similar ML structure. However, the “COMPARISON” unit, i.e., transistors M1-M4, and the “SRAM” unit, i.e., the cross-coupled inverters, are powered by two separate metal rails, namely VDDML and the VDD, respectively. The VDDML is independently controlled by a power transistor (Px) and a feedback loop that can auto turn-off the ML current to save power. The purpose of having two separate power rails of VDD ML and VDD is to completely isolate the SRAM cell from any possibility of power disturbances during COMPARE cycle.

VI. Results And Discussion

Output Waveforms

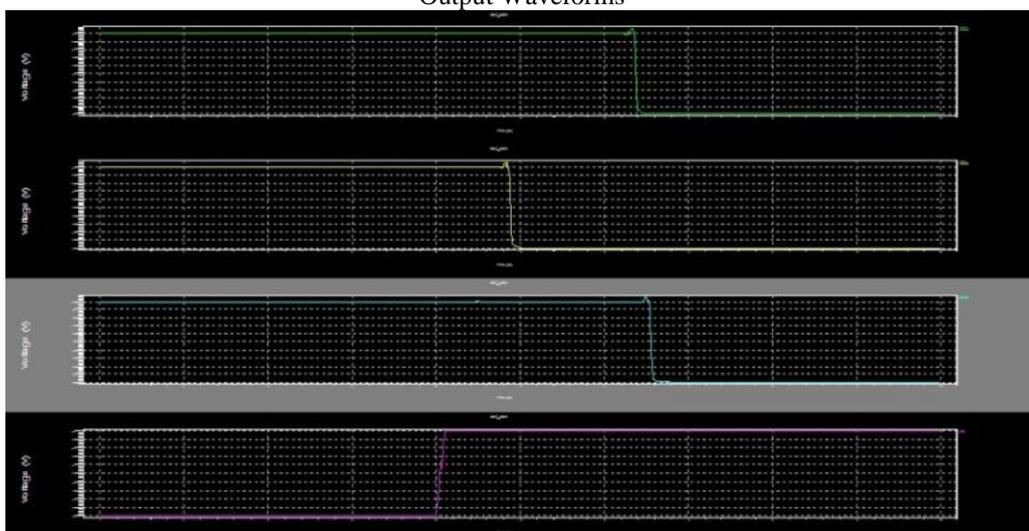


Figure 7: Output waveform of pre-computation CAM design

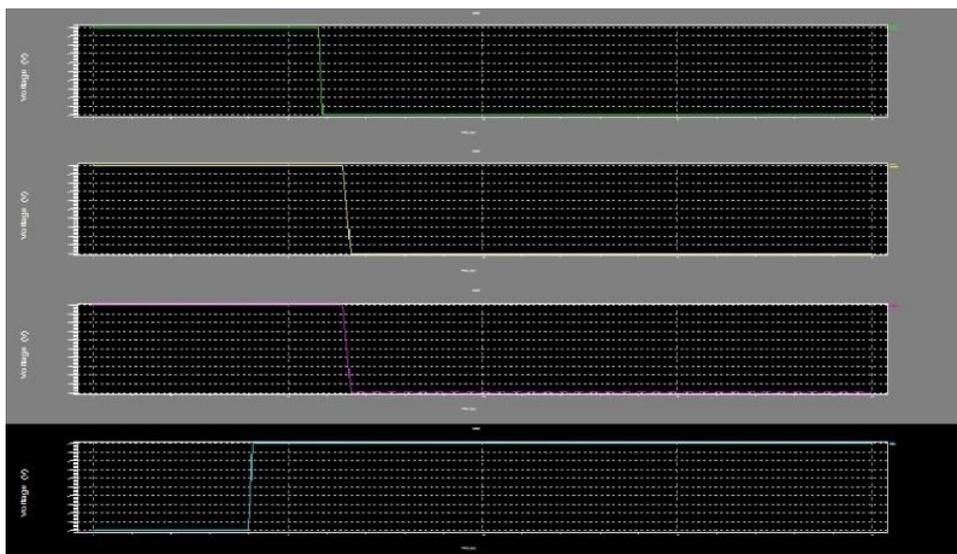


Figure 8: Output waveform of gated power CAM design with 10T structure

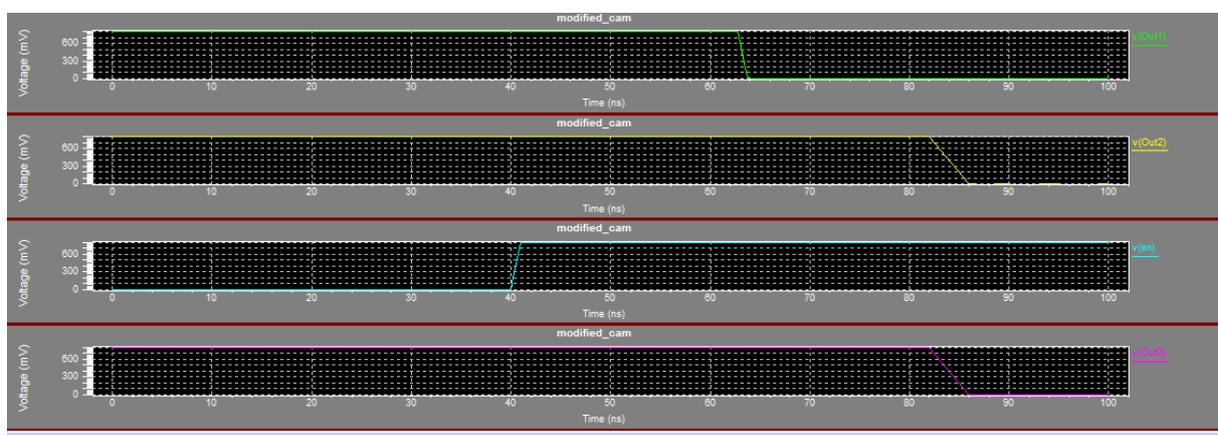


Figure 9: Output waveform of gated power CAM design with 9T structure

TABLE 1 :Power analysis of CAM designs

The power analysis shows that the Gated power CAM is more efficient than pre-computation CAM.

CAM Designs	POWER IN WATTS
Pre-computation CAM	5.642259e-001
Gated power CAM with 10T structure	6.760075e-003
Gated power CAM with 9T structure	2.683e-005

VII. Conclusion

The gated power technique based 9T CAM architecture offer several major advantages, namely reduced peak current, average power consumption, boosted search speed and improved process variation tolerance, reduced sensing delay. It is stable and maintains their low-power consumption property. At normal operating conditions, both pre-computed CAM and power gated CAM designs are equally stable. The power analysis shows that the Gated power CAM design is more efficient than pre-computation CAM design. It is therefore the most suitable design for implementing high capacity parallel CAM in sub-45-nm CMOS technologies.

References

- [1]. Anh-Tuan Do, Shoushun Chen, Zhi-Hui Kong, and KiatSeng Yeo,(2013),‘A High Speed Low Power CAM With a Parity Bit and Power-Gated ML Sensing,’ IEEE Transactions On Very Large Scale Integration (VLSI) Systems, Volume. 21, No. 1, pp. 151-156.
- [2]. Arsovski. I and Sheikholeslami. A,(2003),‘A mismatch-dependent power allocation technique for match-line sensing in content-addressable memories,’ IEEE J. Solid-State Circuits, Volume. 38, No. 11, pp. 1958–1966.
- [3]. Do.A.T, Chen.S.S, Kong.Z.H and Yeo.K.S,(2011),‘A low-power CAM with efficient power and delay trade-off,’ in Proc. IEEE Int. Symposium CircuitsSystem (ISCAS), pp. 2573–2576.
- [4]. Igor Arsovski, Trevis Chandler, A. Sheikholeslami,(2003),‘A ternary content-addressable memory (TCAM) based on static storage and including a current- race sensing scheme’, IEEE J. Solid-State Circuits, Volume. 38.
- [5]. Jose. G. Delgado and JabulaniNyathi,(2000),‘A VLSI High-Performance Encoder with Priority Lookahead’, IEEE Transactions on Computers, Volume. 30, No. 5, pp 565-570.
- [6]. Mohan.N, Fung.W, Wright.D and Sachdev.M,(2009),‘A low-power ternary CAM with positive-feedback match-line sense amplifiers,’ IEEE Trans. Circuits Syst. I, Reg. Papers, Volume. 56, No. 3, pp. 566–573.
- [7]. Mohan. N and Sachdev. M,(2009),‘Low-leakage storage cells for ternary content addressable memories’, IEEE Trans. Very Large Scale Integration.(VLSI) System, Volume. 17, No. 5, pp. 604–612.
- [8]. Pagiamtzis. K and Sheikholeslami. A,(2007), ‘Content-addressable memory’ Home Publications articles- Introduction and CAM application.
- [9]. Pagiamtzis.K and Sheikholeslami.A (2004), ‘A low-power content-addressable memory (CAM) using pipelined hierarchical search scheme’, IEEE J. Solid-State Circuits, Volume.39, No. 9, pp. 1512–1519.
- [10]. S. Baeg,(2008), “Low-power ternary content-addressable memory design using a segmented match line,” IEEE Trans. Circuits Syst. I, Reg.Papers, volume. 55, no. 6, pp. 1485–1494.
- [11]. Sheikholeslami. A and Pagiamtzis. K,(2006),‘Content-addressable memory (CAM) circuits and architectures:A tutorial and survey,” IEEE J. Solid-State Circuits, Volume. 41, No. 3, pp. 712–727.
- [12]. Tyshchenko. O and Sheikholeslami. A,(2008),‘Match sensing using matchline stability in content addressable memory (CAM)’, IEEE J. Solid-State Circuits, Volume. 43,No. 9, pp. 1972–1981.