

A Novel Low-Power Matchline Evaluation Technique for Content Addressable Memory (CAM)

TELAJALA VENKATA MAHENDRA¹, SHEIKH WASMIR HUSSAIN¹,
SANDEEP MISHRA² AND ANUP DANDAPAT¹

¹*National Institute of Technology Meghalaya
Shillong, 793003 India*

²*Indian Institute of Information Technology Pune
Pune, 411048 India*

*E-mail: telajalahendrah@nitm.ac.in; wasmir_vlsi17@nitm.ac.in;
sandeepmishra@iiitp.ac.in; anup.dandapat@nitm.ac.in*

Content addressable memory (CAM) is an outgrowth of static random access memory (SRAM) to execute search functions, which is essential in variety of high-speed applications. An array in CAM system is characterized by matchline (ML) rows which are essential to be pre-charged to default voltage levels prior to every search. During search, mismatching MLs discharge to ground and hence cause huge switching power dependant on the level of pre-charge and amount of discharges. In this paper, a novel pre-charge technique is presented to reduce the pre-charge voltage leading to a low-voltage ML transition between pre-charges and searches. The CAM operation using this scheme offers not only the energy-efficiency but also improved performance because of lower loading capacitance and reduced ML voltage swing. A 256×64-bit NOR-based CAM has been designed with the incorporation of this scheme using a predictive 45-nm CMOS technology and post-layout simulations are carried out in SPECTRE at 1-V supply. Energy dissipation in the proposed macro reduces by 38% while search speed improves by 4.7% compared to the traditional pre-charge scheme. The proposed low-power matchline evaluation approach can be utilized in related existing works on binary as well as ternary CAMs for better power and delay reductions.

Keywords: content addressable memory (CAM), energy efficient, low-power, low matchline (ML) voltage swing, pre-charge mechanism, ternary CAM (TCAM)

1. INTRODUCTION

Random access memory (RAM) is most popular and commonly used memory from a long decades. Through RAM, the information as words can be either stored or read from the memory by providing proper address location [1]. The RAM can also be called as location-addressable memory. Because RAM works on the basis of address/location, it keeps limitation on the effectiveness in data processing systems. On the other hand, content addressable memory (CAM) is a hardware search engine in which the access is performed on the basis of contents rather than physical address. The CAM functionality clearly differentiates from RAM operation in the following approaches: (1) There is no

Received August 8, 2019; revised November 12, 2019; accepted March 23, 2020.
Communicated by Lenin Gopal.

need of address location to be known for writing data into the memory and retrieving data from the memory; (2) During retrieval of data, the information is searched and sensed in parallel from all locations and processed further to find whether the search data is associated with pre-stored data or not. Although accessing of pre-stored data in a memory does not require address locations in CAM architecture, it can operate as RAM by integrating with an address system. The searching can also be carried out in typical RAM by periodical reading and comparing with all stored entries in every search operation, however, it results in longer search time requirement than utilization of CAM for the same operation. A CAM compares the provided search data with the list of stored words in parallel and returns address location of the matched word within a single clock cycle [2, 3]. CAMs are also familiar as associative memories distributed in high-speed processing systems, which are able to perform extensive range of numeric and non-numeric computational tasks. Moreover, the arrangement and design of CAM system attracted memory designers and found suitable for high-performance VLSI implementations. Therefore, the designers have been showing interest and developed many associative memory designs for incorporating in high-speed processing applications [4].

CAMs are widely used in various network applications for packet classification, table look-up in routers, long prefix matching, branch prediction tables and cache memories in processors. Primarily, CAMs are utilized over the decades for packet classification and forwarding in high-speed network systems [5]. Various software as well as hardware methods have attempted to improve performance for single packet classification but resource constraints have limited the implementation of these methods for multiple tasks and hence become expensive. A ternary CAM (TCAM) can integrate these tasks and also allow parallel rule matching in achieving the packet tasks. However, various emerging applications also use CAMs in their systems for high-speed data processing in electronics and communication related devices [6]. A new emerging application of CAM is transactional memory (TM) which is used to identify how the data is accessed and shared among various cores and threads in a multi-processor environment. Several software transactional memories (STMs) were already existed for some time where as hardware transactional memories (HTMs) are emerged recently in enterprise and consumer electronic devices [7, 8]. Other CAM applications also includes networking, neuromorphic associative memory, reconfigurable computing, analytics and data mining, wearable/implantable systems and pattern matching. So, there is a high demand for high-speed and low-power CAM designs to meet the next generation hardware and software applications in the growing applications. To meet requirements of faster search and large capacity storage cells, CAM becomes more power hungry [9] due to parallel access and association of many driver circuits. Since the search bits of a search word are broadcasted to every CAM cell while performing search operation, every cell requires minimum amount of energy, otherwise large fan-out problem turns unconquerable. Fig. 1 shows a basic CAM architecture embedding all necessary input/output (I/O) drivers where CAM cell acts as core element. Search-lines and matchlines are also the main sources of higher power dissipation due to high switching activities across these components [10]. Higher power dissipation sometimes nullifies the benefit of high-speed search operation and prohibits the utilization of CAM as an efficient search engine. Since, for every search operation the entire structure is activated, it results in high energy consumption [11]. Energy dissipation is an essential parameter to be considered in CAM

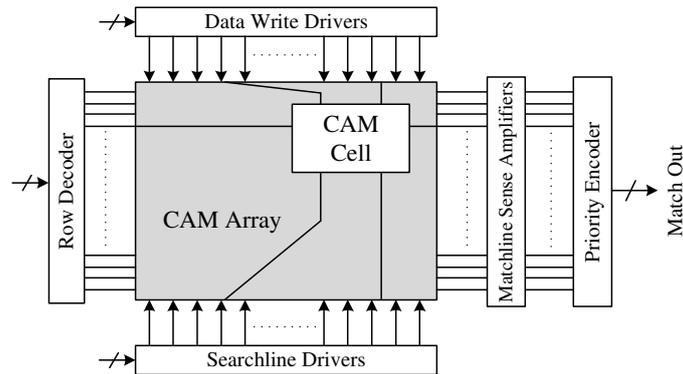


Fig. 1. CAM architecture along with necessary I/O peripherals.

design. Therefore, many designers have focused on energy-efficient CAM designs. It can be achieved by improving circuit level and structural level approaches. NAND-type CAMs are appropriate for low-power applications [12] while NOR-type structures are suitable for high-speed applications [13].

The organization of this paper is as follows: Section 2 provides the preliminaries of CAM, types and ML operations. Section 3 discuss about various types of existed CAM designs for achieving low-power and high-speed goals. Section 4 describes the novel low-power matchline evaluation technique followed by performance comparison results derived from the post layout simulations carried out on 256×64 -bit array in section 5 and section 6 concludes the paper.

2. CONTENT ADDRESSABLE MEMORY (CAM): CIRCUITS AND OPERATIONS

A CAM array is an collection of binary or ternary core cells arranged in the form of ML stacks to allow parallel access. The type of CAMs are chosen based on applications. Matchline types differ in the circuit connections and the discharge characteristics during search. All these are discussed in detail in the following sub-sections.

2.1 Binary CAM vs Ternary CAM

CAM is of two types: Binary CAM (BCAM) and ternary CAM (TCAM). BCAM can store and search only data bits of '0's and '1's where as, TCAM provide additional don't care (X) bit in storage as well as comparison. Search through BCAM matches with an exact data pattern for which many possible combinations with single data entry in not possible with it. With the ternary approach, a CAM can perform association with the range of data words and hence this provides benefit to perform association with more possible data (it is not possible with BCAM designs). The advantage of TCAM over BCAM can be seen from Fig. 2 where the BCAM matches with the fixed entries of data while ternary allows the association of single entry with different possible data due to possibility of don't care storage in TCAMs. Both global and local masking traits are

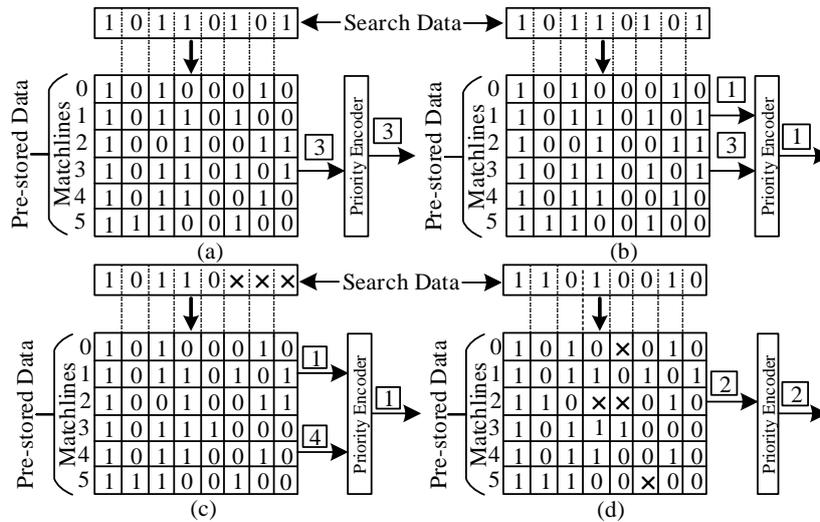


Fig. 2. Illustration of BCAM and TCAM operation by supplying data; (a) Single match of BCAM; (b) Multiple entry match in BCAM; (c) Global masking in TCAM; (d) Local masking in TCAM.

shown in Figs. 2 (c) and (d) with the help of different search key examples. The masking of search pattern is known as global masking while masking of storage data is familiar as local masking. The illustrations shows pre-stored data in CAM table and search data fed through search registers. Each digit of a stored data represents individual CAM cell. However, the possibility of more matches is possible with search data as same data pattern can be stored in different locations. If two or more data words match, then a priority encoder (PE) at output stage resolves and produces the top matched ML as output since all MLs are connected to PE with highest priority criteria.

2.1.1 CAM cell operation

Contrary to data accessed random access memory, CAM allows location accessibility with the addition of simultaneous data comparison and address evaluation. A static CAM cell [13] utilizes the same storage block of 6T SRAM cell as shown in Fig. 3. A comparison circuit is connected to the cell to feed the search bit. Bits stored in the cell function as the gate input for the evaluation logic part. Comparison of the stored bit and the search bit is then reflected on the ML of the cell. The inputs to evaluation logic create two discharge paths (M_7 - M_8 and M_9 - M_{10}) with the comparison circuit. In this 10T CAM cell, match of comparison is implemented based on the XNOR function of stored-bit and search-bit (SL) while the mismatch leads the ML to discharge through either of the active discharge paths. In another configuration (XOR), the connection of SL and \overline{SL} can be altered and so that ML discharges to GND in match scenario via either of the paths.

2.2 CAM Matchlines and Operation

A group of bits forming a word is assigned an address which is designated by the ML corresponding to the word. The ML in functional block can be viewed as the arrangement

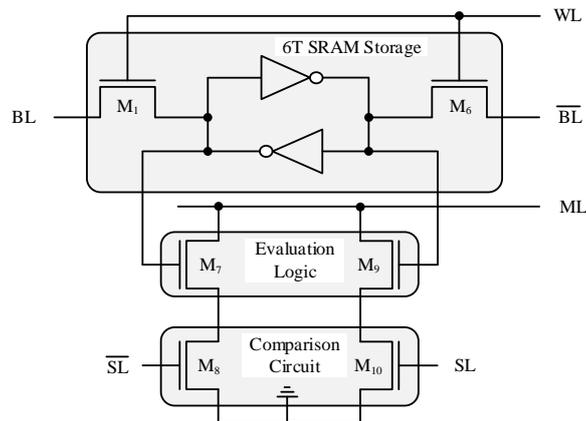


Fig. 3. 10T BCAM cell consisting of storage, evaluation and comparison units.

of CAM cells per word according to a series or parallel connection. Broadly, CAM architecture can be designed using NAND and NOR types of ML shown in Figs. 4 and 5. In order to store data word into a row represented by specific ML, word-line (WL) fed from row decoder is made HIGH and a row of bits are passed into 6T SRAM storage nodes. As the initial stage of ML evaluation begins, WL is pulled down to LOW logic and ML is pre-charged to a default state (usually a HIGH logic). In the next stage also WL is held LOW and a sequence of search bits are applied to the comparison circuits. Procedure of write and pre-charge in both the architectures are same but search differs according to ML types.

2.2.1 NAND-type ML architecture

In NAND-ML structure as shown in Fig. 4, comparison circuits of all cells except the first are modified such that they are connected to the ML of the preceding cells. There are two possible outcomes of a ML evaluation: (i) match; and (ii) mismatch. In case of match, evaluation logic parts provide a ML discharge path to GND via first CAM cell to N-bit CAM cell in N-bit word. In case of mismatch, for example in the second cell, the series path between the MLs of the first and second cell is disconnected and hence ML could not be discharged to GND. A LOW ('0') logic on the ML indicates match state while a HIGH ('1') value indicates mismatch.

2.2.2 NOR-type ML architecture

As shown in Fig. 5, ML of all cells are shorted together to form a NOR-ML. Operational behavior of evaluation logic in NOR-type architecture differs from that in NAND counterpart. In case of match in all cells, evaluation logic disconnects their paths from GND and therefore ML retains the pre-charged state (match). Considering the same example of mismatch in second cell, evaluation logic creates a discharge path from ML at second cell to GND via the comparison circuit and consequently ML changes its state from pre-charge ('1') to GND ('0'). Although, the discussion is fixed on the example of second-bit mismatch, number of mismatching cells vary in different MLs. Based on the

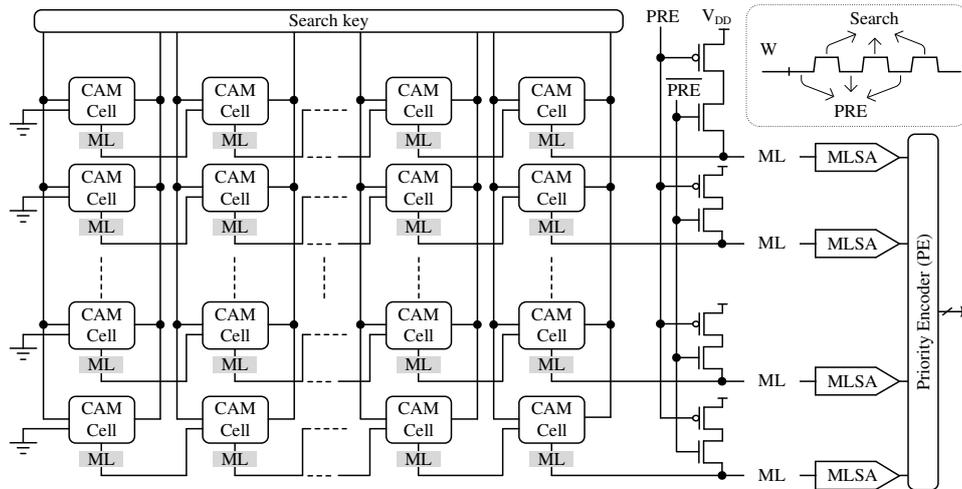


Fig. 4. NAND-type CAM structure incorporating the proposed matchline evaluation technique.

variation, the number of disconnected paths in NAND-ML and connected discharge paths in NOR-ML increases or decreases. And, pre-charge of MLs to a full level HIGH logic is an attractive feature to be targeted for reducing power consumption in CAM search operations. Pre-charge between initial write and first search and also between subsequent searches is responsible for the ML to result high switching activity.

3. BACKGROUND: LOW-POWER AND HIGH-SPEED CAM DESIGNS

The NAND-type CAM implementations follow the discharge of a long chain, hence search speed is limited. The NOR-type design features “pre-charge all matchlines (MLs) and discharge all except the matched ML”. The frequent charging and discharging due to repeated pre-charge and search increases the switching action in CAM, which contributes to large amount of energy dissipation [14]. Most energy-efficient CAM structures may lead to increase overall delay and it affects the overall performance. Therefore, designing such CAM architecture without degrading the performance is an important task. One of the major energy dissipation sources is induced by pre-charging of MLs in every evaluation phase [15]. NAND-type ML exhibits low dissipation but charge sharing causes faulty hit/miss of search in long word lengths, while NOR-type ML features high-speed search and is preferable in most applications but at the cost of heavy power consumption. All MLs of NOR-CAM except one are needed to be pre-charged between successive searches, this leads to challenging task in reduction of ML power. To reduce power dissipation, segmented NOR-type CAM structures presented in [16, 17] divides the NOR-ML, the design in [18] combines both NAND-type ML and NOR-type ML where as the design in [19] accommodates various types of CAM arrays using recycling mechanism to recycle the charge on matchlines. A low-power data-relocation scheme based TCAM (DR-TCAM) is presented to reduce searchline power by relocating the data presented

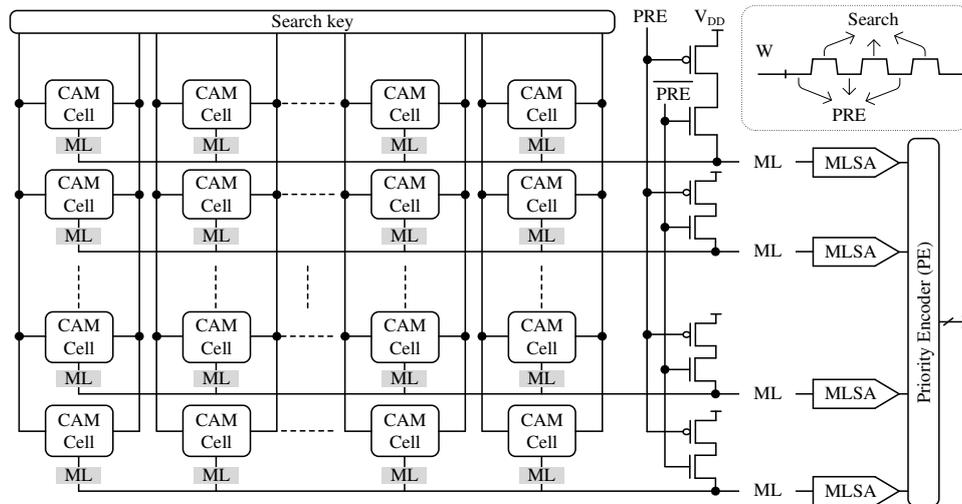


Fig. 5. NOR-type ML architecture with proposed low-swing pre-charge technique.

in the prefix region [20]. A configurable memory based TCAM by using push-rule 6T SRAM cells can improve energy efficiency of search as it allows to reconfigure SRAM as CAM [21]. Dynamic reconfigurable CAM is also developed for achieving low-power and energy efficiency with an additional area overhead [22] for various packet classification applications. The don't care reduction based TCAM is presented for high density applications by embedding additional bypass transistors and decoders [23], which performs search at an energy efficiency of 0.41 fJ/bit/search (this can be improved further with the incorporation of proposed scheme). The adaptive matchline discharging based TCAM [24] and TCAM using early termination of ML precharge [25] are presented to improve the power efficiency. A single ended storage TCAM using split-controlled mechanism and triple margin sensing scheme is presented to achieve low-power and lower energy along with low cell area [26].

However, all these described designs are also use traditional pre-charge for charging matchlines. Designers have presented some efficient pre-charge techniques to control precharge voltage levels such as precomputation pre-charge [27], selective pre-charge [28] and butterfly mechanism [29] to reduce dissipation with the traditional pre-charge mechanism. As the number of words increases, the significance of ML power dissipation increases. Many works are existed in literature to improve energy/power efficiency without affecting the speed. Mostly, all available CAM structures in the literature are using single pMOS transistor controlled with a pre-charge signal (PRE) to pre-charge the MLs. Pre-charge free works are reported recently [30-32], which eliminates necessity of ML pre-charge but provides only binary association. The self controlled pre-charge free CAM [30] works without precharging but dissipates more power due to diode structure and weak logic passage in evaluation logic. In pre-charge free dynamic CAM (PF-DCAM) [31], the data has to be refreshed before performing every search as it uses dynamic storage (it cost additional cycle time). The refresh circuit used to refresh the data periodically leads to area overhead. The energy efficient precharge free CAM (EPPF-CAM) [32] dissipates

less power compared to SCPF-CAM but speed of search to be improved further. In this paper, we proposed a low-swing technique for matchline precharge which can be applied on most of the available CAM (both binary and ternary) structures instead of traditional precharge for improving speed and energy efficiency further with negligible area overhead.

4. A NOVEL LOW-POWER MATCHLINE EVALUATION TECHNIQUE

The CAM functionality undergoes mainly through two phases: (i) write phase (ii) evaluation phase. The write process is minimal because this is required while storing the data. Once storage of pre-defined data is completed, the modification of data is a rare probability. The evaluation is a repetitive process as the content is to be verified/fetched/associated among various modules in a system. However, the traditional evaluation follows two sequential cycles as: (a) ML pre-charge and (b) search. Before performing content search, all the matchlines (MLs) are to be pre-charged to HIGH logic in all search cases. If the search data matches with the pre-stored data, then the ML stays at its pre-charged value by indicating match (HIT) in Fig. 5. Otherwise, even a single mismatch bit in a word discharges the ML in the circuit thereby indicating mismatch (MISS). The unnecessary power dissipation and delay occurs due to unwanted switching activities during every evaluation due to requirement of ML pre-charge. High switching activity of ML pre-charge prior to every search in evaluation can be minimized through the proposed technique by precharging the MLs to voltages lower than the supply. The power dissipation is dependent mostly on the supply, ML swing, mismatch capacitance and switching frequency. Out of these parameters, control over the ML swing is achievable efficiently. Figs. 4 and 5 represent NAND-type and NOR-type ML structures of a $M \times N$ word with the incorporation of proposed technique. A 10T CAM cell shown in Fig. 3 has been used in this paper to verify the improvement of the proposed pre-charging and searching scheme. In this paper, low ML swing and small ML voltage discharge during search from the low pre-charged swing are achieved to improve performance efficiency. The proposed ML voltage minimization technique is simple yet works effectively to improve performance of CAM. The evaluate operation of the scheme is explained logically through one phase of pre-charge and search as follows:

1. During pre-charge phase, PRE signal is made LOW ($\overline{\text{PRE}} = \text{HIGH}$) so that a low voltage is passed to ML instantly from the supply because of the threshold drop at the nMOS.
2. To initiate the search operation, PRE signal is raised to HIGH ($\overline{\text{PRE}} = \text{LOW}$). This isolates the path from supply to ML as the pair of pre-charge MOS's are in cut-off. When search key matches with stored data of an entry, the matched ML remains at pre-charged voltage, else discharges to GND. The requirement of matchline and sensible voltage levels of matchline sense amplifier (MLSA) are discussed by considering match and mismatch scenario as follows:
 - **Match:** During match state, the ML holds the matchline voltage (V_{MLswing}), there after keeps output of MLSA at HIGH logic. The MLSA must be designed in such

a way that, it should be able to sense both LOW and HIGH voltage levels present on MLs without any error for stable functionality. Usually, mono-port sense amplifier circuits are able to sense both LOW and HIGH voltage levels by having good stability.

- **Mismatch:** The associated matchline voltage " $V_{MLswing}$ " discharges to GND through the available discharge paths in mismatched CAM cells whenever a mismatch occurs in single or more CAM cells during search operation. The MLSA starts sensing the LOW voltage on MLs whenever the $V_{MLswing}$ is lower than the MLSA input resolution. Consequently, the output state of MLSA changes its state from HIGH to LOW and indicates a mismatch.

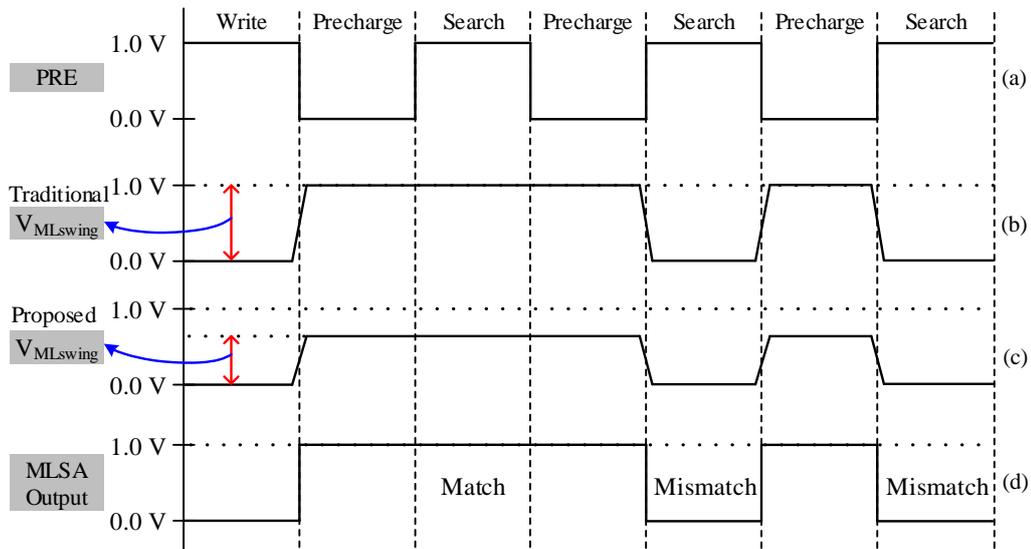


Fig. 6. Matchline transition states during precharge and search: (a) PRE signal; (b) Traditional ML precharge scheme; (c) Proposed ML precharge scheme; and (d) Sense output of both the schemes.

Fig. 6 shows the timing diagram of the traditional and proposed ML schemes with PRE as the signal to pre-charge under the consideration of 1 V supply. From Fig. 6 (b), it can be concluded that through traditional precharge mechanism, the matchlines charges/discharges up to full supply during precharge and search phases at both match and mismatch states. This leads longer charge as well as discharge duration causing huge amount of power dissipation. Through the proposed mechanism, the voltage swing is minimized along with the reduction in charge/discharge time as it charges and discharges lower than the applied supply voltage as shown in Fig. 6 (c). Since most of the MLs in CAM system mismatch with the searched data, the charge/discharge level (difference between HIGH to LOW) can be minimized by introducing the low-swing pre-charge scheme to meet low-power applications demand by maintaining its high-speed advantage. From Fig. 6 (d), it can also be observed that the output of MLSA gets full rail-to-rail level in both schemes during match as the supply nodes of the sense amplifier are connected to the

applied voltage (V_{DD}). Based on post-layout simulations over 1000 MC runs, the SPICE waveform validation of the proposed scheme is shown with ML states in Fig. 8 (a) and ML discharge delay in Fig. 8 (b).

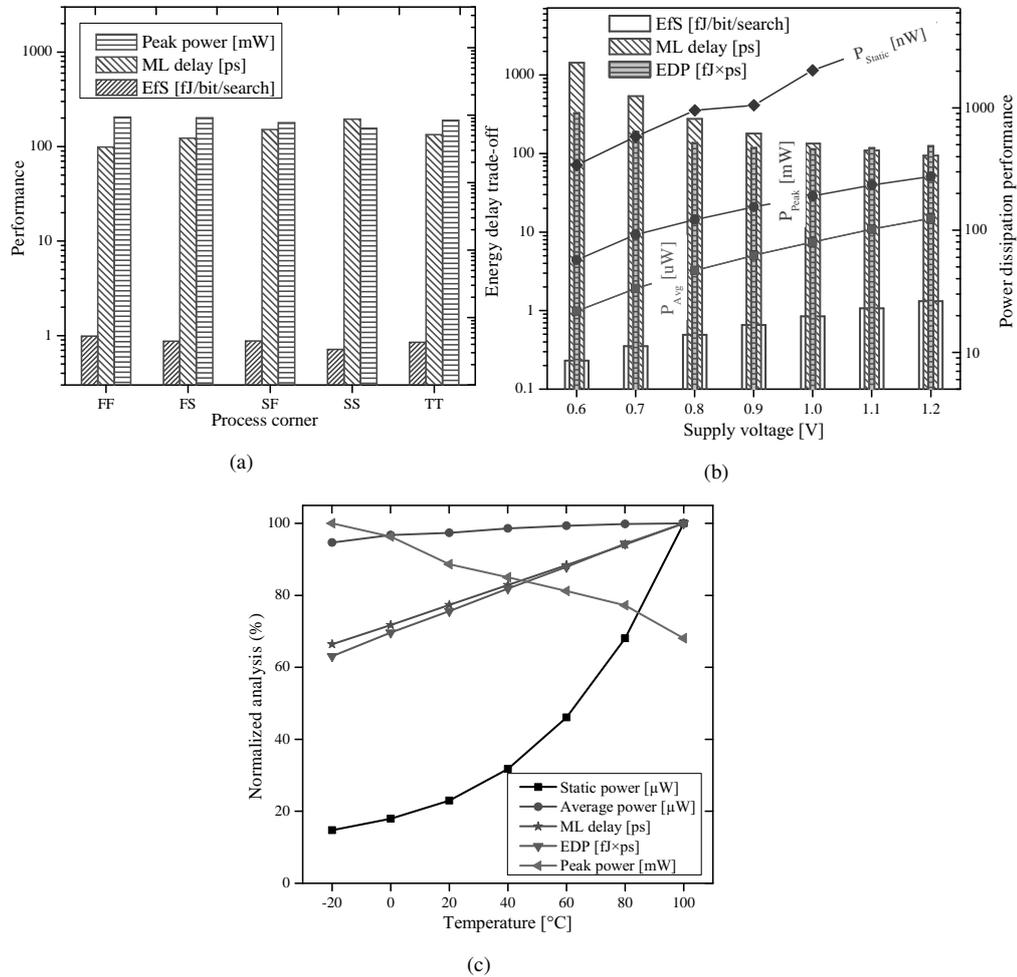


Fig. 7. Performance analysis at PVT variations: (a) Power, delay and energy at process corners; (b) Energy-delay analysis for voltage scaling; (c) Normalized performance against temperature scaling.

The advantage of the proposed mechanism over the traditional pre-charge technique (single pMOS controlled by PRE signal) is the reduction of pre-charge level voltage on MLs through the proposed evaluation technique. This mechanism minimizes overall power dissipation and energy for per search (EfS) due to the contributions from both reduction in discharge voltage and the small discharge time elapsed in evaluation phase. During the pre-charge operation, the drain terminals of both the MOS's are shorted, acting as a virtual capacitor. The minimization of power dissipation is directly proportional to

the reduction of supply, leads to reduction in matchline voltage. The matchline power can be calculated from the following expression:

$$P_{ML} = WC_{ML}V_{DD}V_{ML\ swing}f \tag{1}$$

According to Eq. (1), more reduction in matchline power (P_{ML}) can be achieved by minimizing supply (V_{DD}) and matchline voltage swing ($V_{ML\ swing}$). If the supply voltage is dropped down to a lower supply then in turn $V_{ML\ swing}$ also reduces. Through the proposed mechanism, we can have less $V_{ML\ swing}$ on matchlines compared to traditional scheme. Thus, employing the low-swing ML precharge technique for pre-charging purpose can achieve better power/energy efficiency over traditional mechanism as $P_{ML} \propto V_{ML\ swing}$. During search operation, the mismatched MLs in a CAM array discharges its pre-charge voltage from $V_{ML\ swing}$ to “0” slowly. If $V_{ML\ swing}$ is more than the time it takes to discharge to “LOW” voltage also increases, which is also affects speed and energy efficiency. The time taken by a high $V_{ML\ swing}$ to discharge to GND level will be more than that taken by a lower $V_{ML\ swing}$. Therefore, higher $V_{ML\ swing}$ level affects both the ML discharge time and energy consumption. Lowering $V_{ML\ swing}$ along with supply gives best performance metrics in most of the CAM designs. The proposed technique can be used in most of the existing BCAM and TCAM designs for precharging purpose to minimize power and improve speed.

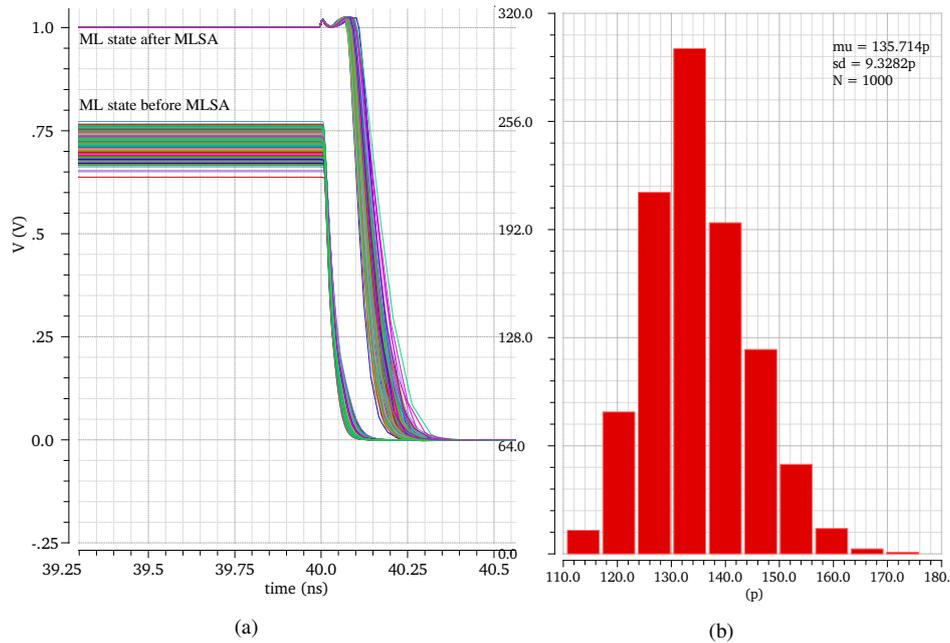


Fig. 8. Performance over 1000 runs using MC sampling method; (a) ML state; (b) ML delay.

Table 1. Performance comparison summary.

Parameter	Traditional	Proposed
Configuration	256×64	256×64
Technology	45-nm	45-nm
Supply Voltage [V]	1.0	1.0
Average power [μ W]	126	80.2
Peak power [mW]	238.2	190.1
EfS [fJ/bit/search]	1.36	0.84
ML delay [ps]	140.9	134.2
Energy-delay [fJ×ps]	191	112
Minimum supply [V]	0.5	0.5
Maximum frequency [GHz]	1.1	1.1

5. RESULTS AND ANALYSIS

256×64-bit NOR-based CAM macros are designed with proposed and traditional pre-charge technique using 45-nm CMOS technology. Post-layout simulations are performed in SPECTRE simulator to verify search performance and advantages of the proposed scheme. The possibility of ML discharging is high in CAM search engine because many stored words would not match with the search key. The same advantages can be obtained through traditional mechanism also by operating at lower supply but utilization of very lower supply may affect the functionality of other peripherals in CAM system and utilization of two power rails makes difficulty while routing. So, presented ML pre-charge technique is an alternative scheme to be used in designing an energy-efficient hardware search engine. To verify the robustness of the novel mechanism, the PVT (process, voltage and temperature) variations are carried out. Fig. 7 (a) shows the performance of the proposed CAM macro at different process corners and energy-delay trade-off along with power performance against voltage scaling from 0.6-V to 1.2-V are plotted in Fig. 7 (b). ML delay and significant power/energy dissipation with respect to temperature scaling are plotted in Fig. 7 (c) and energy-delay-product (EDP) emphasizes the overall performance improvement.

Besides PVT variation, the mismatch analysis is also carried out to verify the maximum ML variation at different local mismatch conditions by performing Monte-Carlo (MC) simulations over 1000 runs on 64-bit word. The post-layout simulation results indicated that the affect on speed is minimum under several random mismatch conditions. The voltage on MLs does not have full swing after performing search operation due to low-swing pre-charging and discharging of MLs. An appropriate ML sensing amplifier (MLSA) is essential to sense and pass strong logic for ease of interface with internal and external peripherals of memory. A charge-up ML sensing scheme [2] have been employed in the memory design. The related ML voltages have been evaluated and plotted in Fig. 8 (a) prior to ML sensing scheme and after sensing to show the importance of utilizing the sensing scheme. The architecture with proposed mechanism performs search with the energy competence of 0.84 fJ/bit/search while traditional technique performs at the cost of 1.36 fJ/bit/search. The results indicate that the proposed scheme can improve ML delay by 4.7% and EDP by 41% compared to traditional pre-charge scheme. The

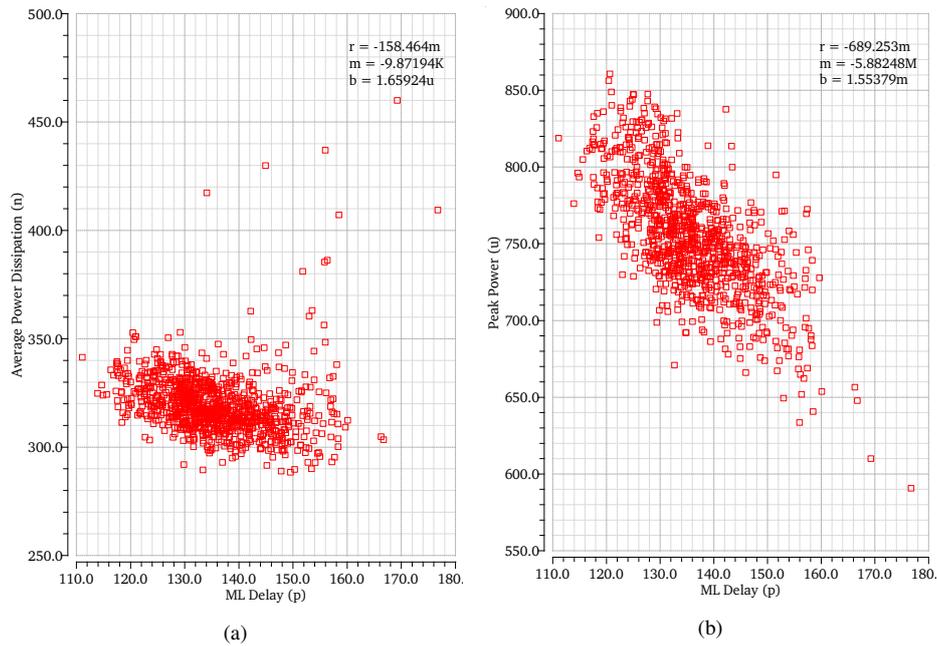


Fig. 9. Scatter plot over 1000 runs using Monte-Carlo sampling method. (a) Average power dissipation vs ML delay. (b) Peak power vs ML delay.

performance comparison of the proposed with traditional technique is presented in Table 1. The advantage of proposed mechanism increases as the number of searches increases.

5.1 Monte-Carlo (MC) Sampling Method

To test the stability of searching operations of the proposed pre-charge-level reduction mechanism, a statistical analysis is carried out for single 64-bit word. Monte-Carlo (MC) sampling method serves this analysis considering random sampling and characterization in the ML circuit. In CAM, even though a miss of search is expect, there may be chances of match due to the variations in charge amongst various nodes and device parameters of the circuit. A search key applied on the same row of ML and the difference in ML voltage during pre-charge as well as search phases are shown for 1000 runs [it varies ML state randomly by supplying mismatch and process variations over 1000 samples]. As shown in Fig. 8 (a), the ML pre-charges initially to a near low-swing around 700-mV which is approximately (Supply/1.43). During the search, the ML discharges to 0 V for all the runs confirming accurate results for all samples. Variable voltage of the ML in pre-charge phase ranges from 630-mV to 760-mV at 1 V supply which is able to drive the sense amplifier sufficiently during evaluation to indicate the correct result of CAM operation. From Fig. 8 (a), it can be observed that the ML is discharging perfectly while mismatch state over all 1000 runs and retained its ML state during write, pre-charge states [default match]. The importance of MLSA utilization in CAM design can be observed from SPICE waveform in Fig. 8 (a). The final ML state of a word is unable to retain

its full voltage levels due to cascade connection of individual CAM cells while forming a word. However, the logical states of the MLs are needed to be processed further to other modules in a system as input or terminates as output. Thus, the MLs shall have logical HIGH & LOW values which is ensured by the sense amplifier [MLSA helps to sense and boost the ML states to full voltage levels, as observed in Fig. 8 (a) before and after sensing through the MLSA].

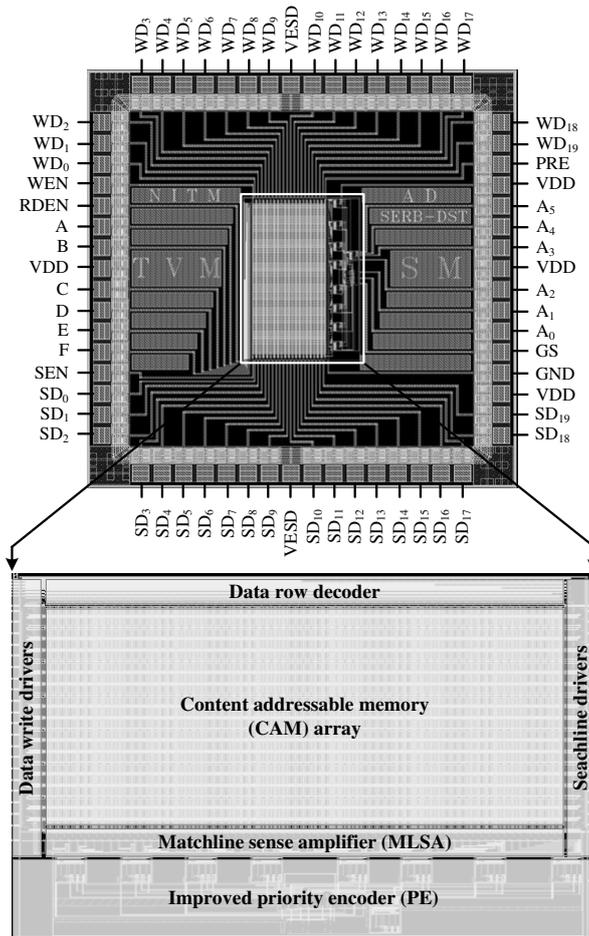


Fig. 10. Tapeout of 64×20-bit proposed matchline evaluation technique based CAM.

The ML state is achieved in few hundred pico-seconds (ps) for which the corresponding delay for 10 groups of near values of different sets (totaling 1000 samples) are shown in the histogram depicted in Fig. 8 (b). The mean value of ML delay is observed to be 135.71 ps leading to a standard deviation of 9.33 ps. Two performance metrics namely average power and delay are shown with the help of scatter plot in Fig. 9 (a). The power versus corresponding delay indicates sharp convergences in lower value combinations for maximum number of runs although few combinations of the two metrics diverge towards

Table 2. Feature comparison summary with prior CAM works.

Reference	CAM Type	ML Scheme	Configuration	Technology	Supply [V]	EfS	Delay [ns]	EDP [fJ \times ns]
[2]	Ternary	NOR	128 \times 32	45-nm	1.0	1.72	0.43	0.74
[3]	Binary	NOR	256 \times 144	0.13- μ m	1.2	1.30	<2	–
[12]	Binary	Seg. NAND	128 \times 64	90-nm	1.0	0.77	0.39	0.30
[15]	Binary	Diff. NAND	128 \times 32	0.13- μ m	1.0	1.87	0.90	1.68
[16]	Binary	Seg. NOR	512 \times 72	0.18- μ m	1.8	6.57	–	–
[17]	Binary	Seg. NOR	128 \times 32	45-nm	1.0	0.73	0.11	0.08
[20]	Ternary	NAND	256 \times 128	130-nm	1.2	1.3	–	–
[23]	Ternary	NAND	256 \times 128	65-nm	1.2	0.41	–	–
[24]	Ternary	NOR	128 \times 64	65-nm	1.1	0.54	0.65	0.35
[25]	Ternary	NOR	128 \times 64	65-nm	1.1	0.58	0.68	0.39
[26]	Ternary	NOR	1024 \times 320	28-nm	0.9	0.422	0.71	0.30
[27]	Binary	NOR	128 \times 32	0.35- μ m	3.3	93	3.90	362.7
[30]	Binary	NOR	128 \times 32	45-nm	1.0	2.10	1.25	2.63
This work	Binary	NOR	256 \times 64	45-nm	1.0	0.84	0.13	0.11

*EfS is in [fJ/bit/search]; Seg.: Segmented; Diff.: Differential

higher values of either or both metrics. Compared to the average power vs ML delay scatter plot [Fig. 9 (a)], peak power vs delay [Fig. 9 (b)] shows wider scattering because of the higher dissipation during sudden change of phase from pre-charge to search. In overall, probabilistic determination of ML behavior (pre-charge and discharge capabilities) under MC method confirms the stability and robustness of the proposed scheme for CAM designs in case of any number of random variations. The Tapeout of 64 \times 20-bit NOR-based proposed CAM design is shown in Fig. 10 along with the compact layouts of all necessary I/O peripherals, it has been designed and tested using semiconductor laboratory (SCL) 180-nm technology in SPECTRE simulator.

5.2 Feature Comparison with Referred CAM Works

Performance summary of the proposed CAM macro and recently reported CAM macros are listed in Table 2. Quasi-static CAM [2] offers moderate EfS and delay indices by employing coupled bit-line and search-line for high-density ternary array. Current saving ML sensing [3] has low-EfS considering the larger macro size but the ML delay is higher. Energy index (EfS) is small in reordered overlapped search CAM (ROS-CAM) [12] at the cost of self pre-charge and global match circuits but ML speed is affected due to NAND-ML structure in self-disabled different NAND CAM [15]. In segmented ML architecture (SMA) [16], EfS is comparatively higher but it can be considered low relative to the high array size. Pre-processing of search keys prior to main search adds latency in pre-computation based CAM (PB-CAM) [27] causing a higher energy-delay-product (EDP). The proposed low-power matchline evaluation macro betters out most of the designs in both EfS and delay, in particular, it is 60% more energy efficient than our previous pre-charge-free scheme [30]. When compared with our recent segmented ML scheme [17], the energy and speed of the proposed scheme is almost comparable, however, the macro size and ML width is doubled in the proposed CAM macro. Therefore, proposed CAM would be better efficient for implementing higher CAM sizes. The EfS

of [23-26] is better than proposed design but in overall performance metric *i.e.* EDP is best in the proposed over most of the compared designs. However, the performance metrics (delay, EfS, EDP) of these designs can also be improved further by using proposed low-swing pre-charge technique in those matchline schemes as the proposed design can be used in most of the existing designs for ML pre-charge.

6. CONCLUSION

CAM features high-throughput searches and replaces software searching approaches. Hardware search engines designed using CAM are utilized in network, security, packet classification and various search-intensive applications. The matchline, which constitutes primary component in CAM array, consumes large amount of power due to parallel comparison. Designing energy efficient CAM architecture without performance drop is a challenging task. In this paper, a novel pre-charge technique is proposed to pre-charge MLs to voltages lower than the supply voltage. Consequently, the swing of mismatching MLs between pre-charge and search is reduced significantly to dissipate less power. The scheme offers dual benefits of low average and peak power of MLs. The 256×64 -bit macro using the proposed scheme dissipates only 0.84 fJ/bit/search at 1-V supply while achieving search time in 134.2 ps. Compared to the traditional pre-charge high scheme, the proposed low-swing pre-charge scheme improves the energy-delay by approximately 41%. This scheme can serve as an energy reduction technique in relevant ML structures of various CAM designs. The proposed scheme retains stable performance at process, voltage, temperature variations and also robust to random variation.

ACKNOWLEDGEMENT

This research was supported in part by the Science and Engineering Research Board under the Project YSS/2015/001198, Ministry of Electronics and Information Technology under the Project SMDP-C2SD 9(1)/2014-MDD, Young Faculty Research Fellowship (YFRF) and Ministry of Human Resource Development, Government of India.

REFERENCES

1. J. T. Koo, "Integrated-circuit content-addressable memories," *IEEE Journal of Solid-State Circuits*, Vol. 5, 1970, pp. 208-215.
2. S. Mishra, T. V. Mahendra, and A. Dandapat, "A 9-T 833-MHz 1.72-fJ/bit/search quasi-static ternary fully associative cache tag with selective matchline evaluation for wire speed applications," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 63, 2016, pp. 1910-1920.
3. I. Arsovski and A. Sheikholeslami, "A current-saving match-line sensing scheme for content-addressable memories," in *IEEE International Solid-State Circuits Conference, Digest of Technical Papers*, 2003, pp. 304-305.

4. Z. Cai, Z. Wang, K. Zheng, and J. Cao, "A distributed TCAM coprocessor architecture for integrated longest prefix matching, policy filtering, and content filtering," *IEEE Transactions on Computers*, Vol. 62, 2013, pp. 417-427.
5. S. K. Maurya and L. T. Clark, "A dynamic longest prefix matching content addressable memory for IP routing," *IEEE Transactions on Very Large Scale Integration Systems*, Vol. 19, 2011, pp. 963-972.
6. R. Karam, R. Puri, S. Ghosh, and S. Bhunia, "Emerging trends in design and applications of memory-based computing and content-addressable memories," *IEEE Proceedings*, Vol. 103, 2015, pp. 1311-1330.
7. P. Bright, "IBM's new transactional memory: Make-or-break time for multi threaded revolution," *ARS Technica*, <https://arstechnica.com/gadgets/2011/08/ibms-new-transactional-memory-make-or-break-time-for-multithreaded-revolution/>, 2011.
8. P. Bright, "Transactional memory going mainstream with Intel Haswell," *ARS Technica*, <https://arstechnica.com/information-technology/2012/02/transactional-memory-going-mainstream-with-intel-haswell/>, 2012.
9. N. Onizawa, S. Matsunaga, V. C. Gaudet, and T. Hanyu, "High throughput low-energy content-addressable memory based on self-timed overlapped search mechanism," in *Proceedings of IEEE 18th International Symposium on Asynchronous Circuits and Systems*, 2012, pp. 41-48.
10. S. Mishra, T. V. Mahendra, J. Saikia, and A. Dandapat, "A low-overhead dynamic TCAM with pipelined read-restore refresh scheme," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 65, 2018, pp. 1591-1601.
11. B. Boguslawski, F. Heitzmann, B. Larras, and F. Seguin, "Energy-efficient associative memory based on neural cliques," *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 63, 2016, pp. 376-380.
12. N. Onizawa, S. Matsunaga, V. C. Gaudet, W. J. Gross, and T. Hanyu, "High-throughput low-energy self-timed CAM based on reordered overlapped search mechanism," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 61, 2014, pp. 865-876.
13. K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory (CAM) circuits and architectures: A tutorial and survey," *IEEE Journal of Solid-State Circuits*, Vol. 41, 2006, pp. 712-727.
14. T. V. Mahendra, S. W. Hussain, S. Mishra, and A. Dandapat, "Low match-line voltage swing technique for content addressable memory," in *Proceedings of the 7th International Conference on Smart Computing Communications*, 2019, pp. 1-5.
15. C. -C. Wang, C. -H. Hsu, C. -C. Huang, and J. -H. Wu, "A self-disabled sensing technique for content-addressable memories," *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 57, 2010, pp. 31-35.
16. S. Baeg, "Low-power ternary content-addressable memory design using a segmented match line," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 55, 2008, pp. 1485-1494.
17. S. W. Hussain, T. V. Mahendra, S. Mishra, and A. Dandapat, "Match-line division and control to reduce power dissipation in content addressable memory," *IEEE Transactions on Consumer Electronics*, Vol. 64, 2018, pp. 301-309.

18. Y. J. Chang and Y. H. Liao, "Hybrid-type CAM design for both power and performance efficiency," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 16, 2008, pp. 965-974.
19. K. L. Tsai, Y. J. Chang, and Y. C. Cheng, "Automatic charge balancing content addressable memory with self-control mechanism," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 61, 2014, pp. 2834-2841.
20. B.-D. Yang, "Low-power effective memory-size expanded TCAM using data-relocation scheme," *IEEE Journal on Solid-State Circuits*, Vol. 50, 2015, pp. 2441-2450.
21. S. Jeloka, N. B. Akesh, D. Sylvester, and D. Blaauw, "A 28 nm configurable memory (TCAM/BCAM/SRAM) using push-rule 6T bit cell enabling logic-in-memory," *IEEE Journal of Solid-State Circuits*, Vol. 51, 2016, pp. 1009-1021.
22. T. Chen, D. Lee, T. Liu, and A. Wu, "Dynamic reconfigurable ternary content addressable memory for openflow-compliant low-power packet processing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 63, 2016, pp. 1661-1672.
23. K. Woo and B. Yang, "Low-area TCAM using a don't care reduction scheme," *IEEE Journal of Solid-State Circuits*, Vol. 53, 2018, pp. 2427-2433.
24. W. Choi, K. Lee, and J. Park, "Low cost ternary content addressable memory using adaptive matchline discharging scheme," in *Proceedings of IEEE International Symposium on Circuits and Systems*, 2018, pp. 1-4.
25. K. Lee, G. Ko, and J. Park, "Low cost ternary content addressable memory based on early termination precharge scheme," in *Proceedings of IEEE International Symposium on Circuits and Systems*, 2019, pp. 1-4.
26. C. Xue, W. Zhao, T. Yang, Y. Chen, H. Yamauchi, and M. Chang, "A 28-nm 320-Kb TCAM macro using split-controlled single-load 14T cell and triple-margin voltage sense amplifier," *IEEE Journal of Solid-State Circuits*, Vol. 54, 2019, pp. 2743-2753.
27. S. -J. Ruan, C. -Y. Wu, and J. -Y. Hsieh, "Low power design of precomputation-based content-addressable memory," *IEEE Transactions on Very Large Scale Integration Systems*, Vol. 16, 2008, pp. 331-335.
28. W. Choi, J. Park, H. Kim, C. Park, and T. Song, "Half-and-half compare content addressable memory with charge-sharing based selective match-line precharge scheme," in *Proceedings of IEEE Symposium on VLSI Circuits*, 2018, pp. 17-18.
29. P. -T. Huang and W. Hwang, "A 65nm 0.165 fJ/bit/search 256×144 TCAM macro design for IPv6 lookup tables," *IEEE Journal of Solid-State Circuits*, Vol. 46, 2011, pp. 507-519.
30. T. V. Mahendra, S. Mishra, and A. Dandapat, "Self-controlled high performance precharge free content addressable memory," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 25, 2017, pp. 2388-2392.
31. T. V. Mahendra, S. W. Hussain, S. Mishra, and A. Dandapat, "Pre-charge free dynamic content addressable memory," *Electronics Letters*, Vol. 54, 2018, pp. 556-558.
32. T. V. Mahendra, S. W. Hussain, S. Mishra, and A. Dandapat, "Low discharge precharge free matchline structure for energy-efficient search using CAM," *Integration the VLSI Journal*, Vol. 69, 2019, pp. 31-39.



Telajala Venkata Mahendra received the B.Tech. degree in Electronics and Communication Engineering from JNTU, Kakinada, India, in 2013, and the M.Tech. degree in VLSI design from the National Institute of Technology at Meghalaya, Shillong, India, in 2016, where he is currently pursuing the Ph.D. degree with the Department of Electronics and Communication Engineering. His research interests include low-power VLSI designs, content addressable memories, volatile memories and CMOS integrated circuits.



Sheikh Wasmir Hussain received the Bachelor's degree in Electronics and Communication Engineering from the Vishvesvaraya Technological University, Belgaum, India, in 2013, and Master's degree in VLSI Design in 2017 from National Institute of Technology Meghalaya, Shillong, India, where he is currently working towards the Ph.D degree with the Department of Electronics and Communication Engineering. His research interests include high-performance memories and low-power VLSI design.



Sandeep Mishra received the B.Tech and M.Tech degrees in Electronics and Communication Engineering from the Biju Patnaik University of Technology, Rourkela, India, in 2011 and 2013, respectively, and the Ph.D. degree in VLSI design from the National Institute of Technology Meghalaya at Shillong, in 2018. He is presently an Assistant Professor with the Department of Electronics and Communication Engineering, Indian Institute of Information Technology Pune, India. His research area of interest covers low-power VLSI design, memory design, mixed signal circuits, analog-to-digital converters, and intelligent transportation systems.



Anup Dandapat received the Ph.D. degree in Digital VLSI Design from Jadavpur University, Kolkata, India, in 2008. He is presently an Associate Professor with the Department of Electronics and Communication Engineering, National Institute of Technology Meghalaya at Shillong, India. Dr. Dandapat has authored over 50 national and international journal papers. His current research interests include low-power VLSI design, low-power memory design, and low-power digital design.