

Review of Low Power ML Sensing Schemes of Ternary Content Addressable Memory

¹Yashaswini H G, ²Aswatha Kumar M, ³B N Shobha

¹M tech Student, ²Principal, ³Associate Professor, Sapthagiri College of Engineering, Bangalore, India

Abstract: A CAM compares input search data in parallel against a table of stored data and returns the address of the matching data. A BCAM stores and searches only “0”s and “1”s. A TCAM can store and search an additional state, called “don’t care” so that a TCAM can perform partial matching. CAM offers high search speed in a single clock cycle, so CAMs are much faster than other search systems. Therefore CAMs are used in a wide variety of applications. Due to its parallel comparison 46% of power is consumed by the ML sensing structure and 21% of power is consumed by SL sensing Structure, so TCAM is power hungry. As power consumption and search speed are the critical challenges faced by TCAM designs. In this paper different circuit level techniques for ML sensing schemes and SL driving approaches are reviewed.

Keywords: Ternary Content Addressable Memory (TCAM), Binary Content addressable Memory (BCAM), Random Access Memory, NOR-type TCAM, NAND-type TCAM, Hybrid-type TCAM, Match-Line (ML), Search-Line (SL).

I. INTRODUCTION

High speed networks are gaining huge popularity in bandwidth hungry real time applications such as packet classify and forwarding. The internet is a mesh of routers and switches, which process data packets and forwards them towards their destination. Each router maintains a routing table and forwards incoming packets based on the information stored in the routing table. As the capacity of routing table increases the speed of table lookup drops off. Software methods such as radix tree and hash function for lookup operation are slow relatively; they do not scale well with table size. Therefore, software methods are now being replaced by hardware solutions to meet performance requirements. An efficient hardware solution to perform table lookup is the CAM. CAM is an outgrowth of RAM. It supports both search and store operations. CAM stores a number of data words and perform parallel comparison of search data words with all the stored entries. If a match is found, the corresponding memory location is retrieved. In case of multiple matches, a priority encoder resolves the highest priority match. CAMs are divided into BCAM and TCAM. A BCAM can only search and store either ‘0’ or ‘1’. Thus, BCAMs are suitable for applications that require only exact match searches. But, TCAMs can store and search ternary state called ‘X’ i.e. don’t care. It can be used as a wild card entry to perform partial matching. However, high power dissipation is one of the major disadvantages of the TCAM.

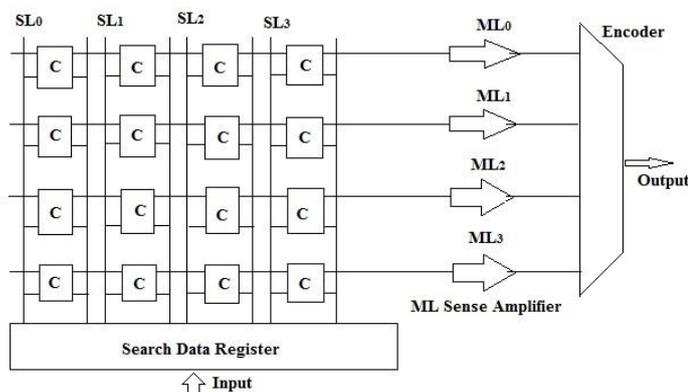


Fig.1 Basic Schematic of CAM.

The fig.1. Shows 4x4 CAM consists of CAM core cells, horizontal MLs and vertical SLs, ML sense amplifier (MLSA) and encoder. The input search data is applied to search data register which is broadcasted onto the SLs. MLs are pre-charged to high equal to supply voltage. Each cell is connected to ML which indicates whether the search data and stored data are identical. The MLSA then detects whether its ML has a matching condition or Miss condition. If the data is identical, it is called Match condition otherwise the condition is Miss. If the condition is Miss then the ML will be pull down. The MLSA are fed to an encoder that generates a binary match location to its encoded address corresponding to the ML of highest priority.

II. LITERATURE SURVEY

Many works on low-power design techniques for TCAMs have been reported, e.g., [1]–[10]. Effective power-reduction techniques for both ML structures and SL structures that are important for designing a TCAM are reviewed. Different core cells are analyzed and comparison done in terms of performance and power. Concluded that hybrid structure consume less power and provide better performance which combine the advantages of both NAND-type TCAM and NOR-type TCAM. Different ML sensing techniques are surveyed such as Pre-charge High ML sensing scheme, Low swing scheme, selective pre-charge scheme, Pipe-line Scheme and Butterfly ML scheme. Pre-charge High ML sensing scheme used NOR ML for sensing the state by first pre-charging ML to high and then after evaluation the MLs are pull down in case of miss or else remained high for match condition. Low swing scheme is used to address various low swing voltage challenges to reduce power consumption linearly with potentially increasing speed. Selective Pre-Charge Scheme used to save power on ML by performing match operation on first few bits say 3bits out of 144-bit word before activating search of the remaining bits which achieved uniform data distribution and eliminating further stages if there is miss in initial stage. In Selective pre-charge, ML is divided into two segments but in Pipe-line Scheme it is divided into multiple segments which are connected in parallel and evaluated serially. If the output is miss in the initial stage then subsequent stages are shut off resulting in power saving but the drawback is increased latency and area overhead. Butterfly ML scheme increases degree of parallelism to reduce the critical path and decrease the number of stages and when one TCAM segment output is miss then it turns off more TCAM segments connected in butterfly style. Part of SL power consumption depends on ML schemes; different SL sensing techniques are Conventional approach, Eliminating SL pre-charge, Hierarchical SL and Self timed charge recycling SL. In Conventional approach, during search cycle SLs are driven to pre-charge level by cascade inverters but the SL power consumption depends on the SL capacitance. In Eliminating SL pre-charge technique the toggling of SLs is reduced, ML sensing schemes that pre-charge the ML low eliminate the need for SL pre-charge but this scheme directly activate SLs with their search data without going through an SL pre-charge phase which reduce 50% of SL power than the pre-charge high ML sensing scheme. Hierarchical SLs are built on top of Pipe-lined MLs; the hierarchical SL scheme divides the SLs into a two-level hierarchy of global SLs (GSLs) and local SLs (LSLs). The GSLs are active every cycle, but the LSLs are active only when necessary, hence overall power consumption of SLs is reduced. Self timed charge recycling SL scheme reduces critical path and delay overhead of the control logic, hence addresses charge sharing problem [1]. In [2], the authors proposed a current- race ML sensing scheme which uses 4T static storage TCAM for increasing density to reduce power consumption by minimizing switching activity of SLs and shrinking voltage swing of MLs by achieving a match time of 3ns with a minimum supply voltage of 1.2V. In paper [3], Hybrid-type TCAM architecture is used which utilize the benefits of both NOR and NAND-type TCAM cells. A Hidden bank selection scheme is used to turn on fewer cells during search operation by reducing delay; ML repeaters and sub ML scheme are used for fast NAND search operation. The scheme achieves a match time of 2.2ns and 144-bit data search with 0.7 fJ/bit/search energy efficiency. In Prefix Segregation Scheme [4], addresses longest prefix matching problem for internet routers, TCAM act as an efficient hardware solution for IP packet forwarding engine with lower complexity, cost and shorter search latency. In [5], an adaptive two-level tree-style ML was proposed to reduce the power and improve the performance of Compare operation where the scheme implements 256X128bit TCAM macro which achieves 1.56ns search time using 1.42 fJ/bit/search of energy. In [6], an AND-type pseudo-footless clock-and-data pre-charged dynamic gate with butterfly routing scheme was proposed to design a low-power TCAM. Some low-power techniques use the concept of ML partitioning to reduce the power consumption of MLs [7]–[9]. In [7], a selective-pre-charge ML scheme was proposed. An NOR-type ML is partitioned into two parts. Then, the comparison result of the first part determines whether the second part is pre-charged or not such that the power consumption of the ML is reduced. The low energy tag storage [8], for caches uses mixed serial parallel comparison which achieves four-fold improvement in energy per access, compared to a standard parallel CAM. The proposed CAM checks the first four LSBs and evaluates the remainder of the tag only if they match; which supports parallel operating mode without a speed loss but with reduced energy savings and

25% increase in search time. In paper [9], uses Pulsed NAND-NOR ML to reduce ML power by activating only a few MLs and Charge recycling SL driver reduces the SL power by recycling the charge of SLs without pre-charging; a ML is divided into two parts, the first part is an NAND-type ML and the second part is an NOR-type ML. Similarly, the comparison result of the first part determines whether the second part is pre-charged or not. In [10], a ripple-pre-charge ML scheme which is similar to an NAND-NOR ML scheme was proposed to reduce the power dissipation of a TCAM. AND-type ML scheme [11], used to realize a high performance energy efficient CAM which achieves a 2.1ns search time with stored and search data, measured with energy of 2.33fJ/bit/search. In Hybrid TCAM [12], CAM word is divided into two segments, and all the CAM cells are decoupled from ML. By minimizing both the ML capacitances and switching activities largely reduce the power consumption because CAM provides a fast pull down path to speed up the light weight ML discharge it roughly reduces 89% of energy consumption. Low-Capacitance and Charge-Shared ML scheme [13], cell-level comparison logic and a charge-shared ML scheme both schemes reduce search time and energy in TCAMs. Measurement results of the above schemes show a search time reduction of 42% and 11%, and a search-energy reduction of 25% and 9%, respectively. Dual ML TCAM [14], the power consumption of fully paralleled search operation at 250 MHz is 9.3 W. This paper realizes high-speed, low-power and robust large-scale TCAM. This contributes to reducing the power of network systems. In Two level ML segmentation scheme [15], the first segment of several bits and the second segment of the others. The proposed two-level segmented word circuit is used to determine the word lengths of the segmented sub-words with minimizing the power dissipation by minimizing the number of active cells. In hybrid IP lookup architecture [16], is based on an efficient trie partitioning scheme that divides the FIB into two prefix sets: a large disjoint leaf prefix set mapped into an external TCAM-based lookup engine and a small overlapping prefix set mapped into an on-chip SRAM-based lookup pipeline. The implementation shows a throughput of 250 million lookups per second equivalent to 128 Gb/s with 64-B packets. The update overhead is significantly lower than that of previous work, the memory consumption is reasonable and the utilization ratio of external TCAMs is up to 100%. In paper [17], ML was partitioned into four segments and selectively pre-charged to reduce the match-line power consumption. In [18], the ML was decoupled and a CAM word was divided into NOR-type CAM and NAND-type CAM to hold the advantage of performance and reduce the CAM power. Some TCAM power optimization methods are developed in algorithm level. In [19], various TCAM architectures and proposed a method to generate optimal prefix set for the given routing table, so that the power consumption could be reduced. Similar concepts are seen in [20] and [21]. Some other methods are also proposed for reducing the dynamic power of TCAM. In this paper [22], a ripple pre-charge match-line scheme similar to a NAND-NOR match-line scheme that has been proposed for reducing the power dissipation of a TCAM. [23], [24] papers explain don't care gating scheme for reducing both search-line switching activity and the average switching power, a don't care gating design could minimize the TCAM search-line power consumption. Unfortunately, of previous research focused on reducing dynamic power of TCAM. In [25], the authors discuss a 90 nm TCAM cell was designed to reduce both the cell leakage power and search-line dynamic power. The author exploiting the feature of prefix data was redundant in the wild match of TCAM. The key idea is to destroy the prefix data to reduce TCAM leakage power when the cell state was "don't care." In paper [26], two TCAM cells are implemented, that are designed to reduce leakage power by eliminating one of the sub-threshold leakage paths. The authors also described techniques to minimize sub-threshold and gate leakages in conventional and proposed TCAM cells. In [27], the leakage suppressed technique using dynamic power source is proved, which minimizes TCAM leakage power. Connecting the prefix power source to the mask data, the dynamic power source technique could reduce not only the TCAM cell leakage power but also the SL dynamic power without any search performance penalty. In [28], cut-off and multi-mode data-retention power gating technique on don't care cells and storage cells are employed to reduce leakage current of TCAM. The architecture is expensive since an extra control circuit is needed for every segment. In paper [29], Dynamic power source (DPS) technique used to suppress dynamic power and leakage power by connecting the prefix power source to the mask data. The DPS technique improves search performance by 11% and reduces 18% of the TCAM dynamic power dissipated in the SL switching activity. In paper [30], a Pai-Sigma ML to reduce the compare power of TCAMs. The proposed ML does not incur the issues of charge sharing and short circuit current. Moreover, the switching activity of the search lines of a TCAM with the proposed MLs is low.

III. CONCLUSION

In this paper, difference between BCAM and TCAM are shown. And different TCAM core cells are reviewed such as NOR, NAND and Hybrid-type TCAM Cell. And comparison is performed based on power and speed. Different circuit level techniques for ML sensing schemes and SL driving approaches are reviewed. It also describes ML power

consumption and reviewed techniques in terms of power savings they provide. And also focused on SLs to examined SL power consumption.

REFERENCES

- [1] K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory circuits and architectures," *IEEE J. Solid-State Circuits*, vol. 41, No. 3, Mar. 2006, pp. 712–727.
- [2] I. Arsovski, T. Chandler, and A. Sheikholeslami, "A ternary content addressable memory based on 4 T static storage and including a current-race sensing scheme," *IEEE J. Solid-State Circuits*, vol. 38, No. 1, Jan. 2003, pp. 155–158.
- [3] S. Choi, K. Sohn, and H. J. Yoo, "A 0.7 fJ/bit/search 2.2-ns search time hybrid-type TCAM architecture," *IEEE J. Solid-State Circuits*, vol. 40, No. 1, Jan. 2005, pp. 254–260.
- [4] M. J. Akhbarizadeh, M. Nourani, and C. D. Cantrell, "Prefix segregation scheme for a TCAM-based IP forwarding engine," *IEEE Micro*, vol. 25, No. 4, Aug. 2005, pp. 48–63.
- [5] J. S. Wang, C. C. Wang, and C. W. Yeh, "High-speed and low-power design techniques for TCAM macros," *IEEE J. Solid-State Circuits*, vol. 43, No. 2, Feb. 2008, pp. 530–540.
- [6] P. T. Huang, S.W. Chang, W. Y. Liu, and W. Huang, "A 256 x 128 energy-efficient TCAM with novel low power schemes," in *Proc. IEEE Int. Symp. VLSI Design, Autom., Test*, 2007, pp. 1–4.
- [7] C. A. Zukowski and S. Y. Wang, "Use of selective pre-charge for low power content-addressable memories," in *Proc. IEEE Int. Symp. Circuits Syst.*, 1997, pp. 1788–1791.
- [8] A. Efthymiou and J. D. Garside, "A CAM with mixed serial-parallel comparison for use in low energy caches," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 12, No. 3, Mar. 2004, pp. 325–329.
- [9] B. D. Yang and L. S. Kim, "A low-power CAM using pulsed NAND-NOR match-line and charge-recycling search-line driver," *IEEE J. Solid-State Circuits*, vol. 40, No. 8, pp. 1736–1744, Aug. 2005.
- [10] D. S. Vijayarathi, M. Nourani, M. J. Akhbarizadeh, and P. T. Balsara, "Ripple pre-charge TCAM: A low-power solution for network search engines," in *Proc. IEEE Int. Conf. Comput. Design*, 2005, pp. 243–248.
- [11] H. Y. Li, C. C. Chen, J. S. Wang, and C. W. Yeh, "An AND-type match-line scheme for high-performance energy-efficient content addressable memories," *IEEE J. Solid-State Circuits*, vol. 41, No. 5, May 2006, pp. 1108–1119.
- [12] Yen Jen Chang and Yuan Hong Liao, "Hybrid-type CAM design for both power and performance efficiency," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 16, No. 8, Aug. 2008.
- [13] Nitin Mohan and Manoj Sachdev, "Low-Capacitance and Charge-Shared Match Lines for Low-Energy High Performance TCAMs," *IEEE Journal of Solid-State Circuits*, vol. 42, No. 9, Sep. 2007.
- [14] N. Mohan and M. Sachdev, "Low power Dual ML Ternary Content addressable memory," in *Proc. IEEE Int. Symp. Circuits and systems*, vol. 2, pp. 633- 636.
- [15] Shoun Matsunaga and Akira Katsumata, "Design of a Low Energy Nonvolatile Fully Parallel TCAM using a Two Level Segmented ML scheme," *IEEE International Symp. on Multiple valued logic*, 2011.
- [16] N. Mohan, W. Fung, D. Wright and M. Sachdev, "Design techniques and test methodology for low power TCAMs," *IEEE Trans. VLSI Syst.*, vol. 14, No. 6, June 2006, pp. 573- 586.
- [17] S. Baeg, "Low power ternary content-addressable memory design using a segmented match line," *IEEE Trans. Circuits Syst.*, vol. 55, No. 6, Jul. 2008, pp. 1485–1494.
- [18] Y. J. Chang, Y. H. Liao and S. J. Ruan, "Improve CAM power efficiency using decoupled match line scheme," in *Proc. IEEE/ACM Des. Autom. Test Eur.*, vol.16–20, Apr 2007, pp. 1–6.
- [19] T. B. Mishra and S. Sahni, "PETCAM—A power efficient TCAM architecture for forwarding tables," *IEEE Trans. Comput.*, vol. 61, No. 1, Jan. 2012, pp. 3–17.
- [20] W. Wu, J. Shi, L. Zuo and B. Shi, "Power-efficient TCAMs for bursty access patterns," *IEEE Micro*, vol. 25, No. 4, Jul. 2005, pp. 64–72.

- [21] W. Lu and S. Sahni, "Low-power TCAMs for very large forwarding tables," *IEEE/ACM Trans. Netw.*, vol. 18, No. 3, Jun. 2010, pp. 948–959.
- [22] D. S. Vijayasarithi, M. Nourani, M. J. Akhbaizadehand P. T. Balsara, "Ripple-precharge TCAM: A low-power solution for network search engines," in *Proc. IEEE Int. Conf. Comput. Design*, 2005, pp. 243–248.
- [23] Y. J. Chang, "Don't-care gating TCAM design used in network routing table," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 18, No. 11, Nov. 2010, pp. 1599–1607.
- [24] Y. J. Chang, "A high-performance and energy-efficient TCAM design for IP-address lookup," *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 56, No. 6, Jun. 2009, pp. 479–483.
- [25] Y. J. Chang, "90 nm TCAM cell design with leakage suppression technique," *Electron. Lett.*, vol. 45, No. 6, Mar. 2009, pp. 300–302.
- [26] N. Mohan and M. Sachdev, "Low leakage storage cells for ternary content addressable memories," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 17, No. 5, May 2009, pp. 604–612.
- [27] P. T. Huang, S. W. Chang, W. Y. Liuand, W. Hwang, "Green microarchitecture and co-design for ternary content addressable memory," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2008, pp. 3322–3325.
- [28] N. S. Kim, T. Austin, D. Blaauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemirand V. Narayanan, "Leakage current: Moore's law meets static power," *IEEE Comput.*, vol. 36, No. 12, 2007, pp. 68–75.
- [29] Yen Jen Chang, "Using the Dynamic power source technique to reduce TCAM leakage power", *IEEE Trans. on circuits and systems*, vol. 57, No. 11, Nov 2010.
- [30] Shun Hsun Yang and Yu Jen Huang, "A Low-Power Ternary Content Addressable Memory with Pai-Sigma MLs", *IEEE Transaction on VLSI Systems*, vol. 20, No. 10, October 2012.