Low-Power Ternary Content-Addressable Memory Using Power Reduction of Match-Line and Search-Line

Jeong-Su Kim[1]   **Jeong-Beom Kim[2]***

[1] *Medical Convergence (Electronics Engineering), Kangwon National University, Republic of Korea*
[2] *Department of Electronics Engineering, Kangwon National University, Republic of Korea*
* Corresponding author's Email: kimjb@kangwon.ac.kr

Abstract: The challenge to solve in large capacity ternary content addressable memory (TCAM) is the large power consumption of search-line and match-line. In this paper, we propose a circuit that uses match-line selective-charging scheme and low-swing search-line scheme simultaneously to reduce power consumption of search-line and match-line which occupies a large ratio in power consumption of TCAM. The circuit was verified by Hspice simulation using a 1.8V supply voltage in a CMOS 0.18 um CMOS process. The proposed 128x128 bit TCAM shows 45.4% power savings in match-line and 74.7% power savings in search-line compared to the conventional precharge high scheme in search cycle. Also, the minimum search cycle time is 3 ns (333MHz).

Keywords: Ternary content-addressable memory, Low-swing scheme, Selective-charging scheme, Low-power memory circuit.

1. Introduction

Content-addressable memory (CAM) is a special memory used in search applications that require very high speeds and it is also called associative memory. CAM compares input search data against a table of stored data, and returns the address of the matching data [1]. Applications requiring high search speed of CAM include parametric curve extraction [2], Hough transformation [3], and image coding [4], etc. Today, CAMs primary commercial application is to classify and forward Internet protocol [5 - 10].

A binary CAM (BCAM) is a general form of CAM that stores "0" and "1" in a storage cell to perform search operation and comparison with search data. Ternary CAM (TCAM) can store "X" (don't care) in addition to "0" and "1" by extended from of BCAM, so that it can perform search operation more efficiently. For example, if the TCAM searches for "11X0", it searches for two types of stored data, "1110" and "1101".

Today, the development of communications technology and the appearance of high-performance applications require high capacitive CAMs of larger size. As the size of the CAM increases, the number of storage cell arrays and match-line sensing amplifiers (MLSA) increases. As the array of storage cell increases, the number of match-lines (MLs) and search-lines (SLs) and the parasitic capacity increase, and therefore the problem of search time and power consumption can't be avoided. This problem becomes even more serious as the size of the CAM becomes larger, so designers need a proper design considering the trade-off between speed and power. For the reasons, there have been many schemes to reduce the power consumption of the SL and ML at the circuit level. The previous methods to reduce the power consumption of the ML are low-swing scheme [11], selective precharge scheme [13], pipelining scheme [14], and current-saving scheme [17]. Furthermore, previous methods for reducing the power consumption of the SL include the eliminating SL precharge scheme [11], and hierarchical SL scheme [14 - 15]. Therefore, this paper describes the basic

structure and operation of TCAM and explains the previous structures that reduce power in the ML and SL, which are the main causes of power consumption in TCAM. In addition, we describe the operation of the proposed structure and see how much power can be saved compared to conventional TCAM based on simulation results. This paper is organized as follows; Section 2 provides CAM basics. Section 3 presents proposed TCAM. Simulation results and comparison are presented in Section 4. Finally, we make conclusions in Section 5.

2. CAM basics

This section contains a general description of the CAM to help you understand the TCAM proposed in this paper. First, the structure and operation of conventional CAM and TCAM are described. We also describe previous techniques that have improved performance by changing the match-line and search-line structure of the CAM.

2.1 Conventional CAM structure and operation

Fig. 1 shows the structure of a BCAM cell composed of 10 transistors. The top four NMOSs are the search part, which compares the SL input with the data stored in the cell to turn on/off and form a path from ML to ground. Search part can be composed of XOR type depending on the design, and can be designed flexibly according to the number of transistor or desired structure [18]. The SRAM cell consisting of the bottom six transistors is a storage part and the data node is connected to the search part.

Fig. 2 shows the structure of a TCAM cell consisting of 16 transistors. The difference with BCAM is the use of two SRAM cells in the storage part. Thus, a TCAM can have two more states than a BCAM that stores "0" and "1" using one storage cell. Therefore it has four data storage states but does not allow both storage cells to be "0" (always mismatch) and only allow the remaining three states. A state where "1" is stored in both storage cells becomes an "X" (don't care) state, which is an unconditional match regardless of search data during a search operation.

Fig. 3 shows a simplified block diagram of conventional CAM. Each storage cell is connected and arrayed with ML and SL, and the TCAM has the same structure except that another storage cell is added. In the search operation SL is connected to search data resistors/drivers and receives search data and compares input search data with data stored in storage cell. All MLs are high precharged prior to

the search operation. All MLs are high precharged prior to the search operation. When both the search data and the data stored in the storage cell are matched, the ML remains high because there is no path to the ground. If any cell is mismatched, a path to the ground is formed and the ML is discharged.
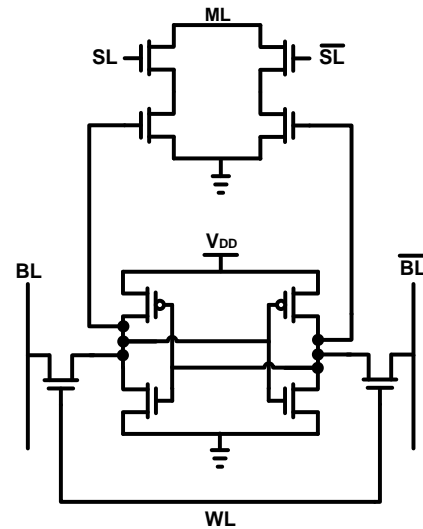

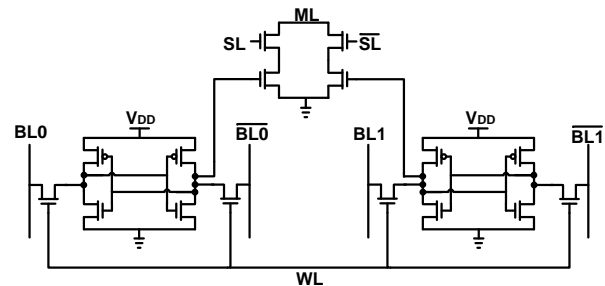Figure. 1 Internal circuit of conventional NOR-type BCAM cell and

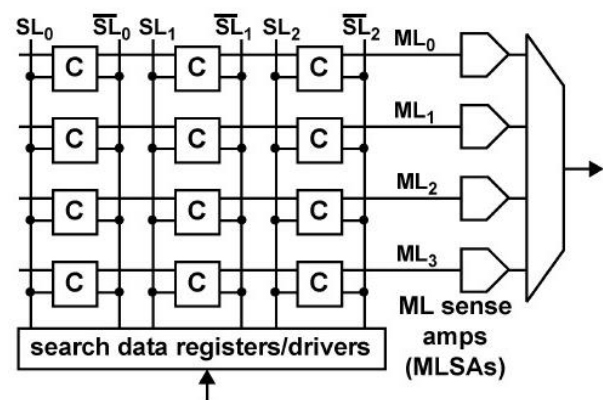Figure. 2 Internal circuit of conventional NOR-type TCAM cell

Figure. 3 Simple schematic of a model CAM with 4 word having 3 bit each. The schematic shows individual core cells, differential search-lines, and match-line sense amplifiers (MLSAs).
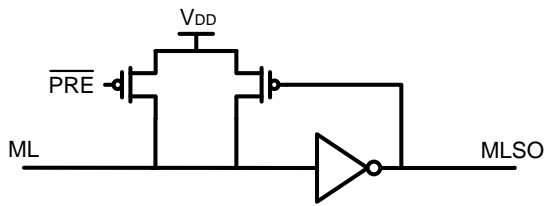
Figure. 4 Schematic of conventional match-line sense amplifier

Fig. 4 shows the conventional match-line sense amplifier (MLSA). The ML of the CAM is connected to the MLSA, and the MLSA inverts the value of the unstable ML after the search operation and outputs it with a definite high or low. As a result, in the ML evaluation of conventional CAM, MLSA inverts the value of ML and output "0" in match condition and "1" in mismatch condition.

2.2 Previously technique

In the search operation of the CAM, the number of words in the match state is on or several, and thus a significant amount of power is consumed when precharging and discharging a large number of MLs or SLs in the conventional CAM structure. In addition, as the array size of the CAM cell increases, the parasitic capacitance of ML and SL increases and the power consumption in ML and SL become serious. For these reasons, a large part of the power consumption in large capacity CAM is occupied by ML and SL. Therefore, some of the methods proposed previously to solve the power consumption problem of match line and search line are introduced.

The low-swing scheme reduces power consumption in the ML by reducing the voltage swing of the ML. The reduction of power consumption is linearly proportional to the reduction of the voltage [11-12]. The main challenge addressed by low-swing implementations is using a low-swing voltage without resorting to an externally generated reference voltage [18].

The selective-precharge scheme is a method of unequally allocating power to ML. This method divides the ML into two segments and evaluates only some of the upper bits of the word. In the case of the word matched by the upper bits, the remaining lower bits are evaluated [13]. However there are actually two overhead sources that limit power saving. First, to maintain speed, the initial match implementation may draw a higher power per bit than the search operation on the remaining bits. Second, an application may have a data distribution that is not uniform, and, in the worst-case scenario, the initial match bits are identical among all words in the CAM, eliminating any power saving [18].

The current-race scheme starts the search cycle by low precharging the ML and charges the current supplied by the current source to evaluate the ML. Since the ML is low precharged, this scheme does not require a separate SL precharge phase required by the ML precharge high scheme. Thus, additional power savings can be achieved in the SL [16]. However, there is a disadvantage of allocating the same amount of current to all MLs.

The current-saving scheme compensates for the disadvantages of the current-race scheme. The current-race scheme improves the disadvantage of allocating the same amount of current to all MLs, so that a larger current is assigned to the match ML and a lower current is assigned to the mismatch ML [17]. The disadvantage of this scheme is that it requires an additional current control block.

The pipelining scheme uses more segments as an extension of the selective-precharge scheme that uses two segment [14 - 15]. It can compensate for the disadvantages of the selective- precharge scheme, which can not the save power when the initial bits match, and has the advantage of being able to use the hierarchical search-line scheme to be described later. However, this scheme has drawbacks of increasing latency and area overhead due to several pipeline stages [18]. The schemes described above are previously proposed schemes to reduce ML power consumption. Since the SL scheme differs according to the operation method of the ML scheme, a method for saving power of SL based on the ML scheme is introduced.

The eliminating search-line precharge scheme saves SL power by eliminating the precharge phase of SL in the search operation. Thus, this scheme skips the SL precharge phase and directly activates the SL with the search data. Since, in the typical case, about 50% of the search data bits are toggled per cycle, SL power is reduced by 50% compared to the precharge high ML sensing scheme with SL precharge phase [17]. This scheme can only be used when low precharging the ML.

The hierarchical search-line scheme is implemented in the pipelining match-line scheme described above. Since the cell array is segmented, the SL is divided into global SL (GSL) for the initial segment and local SL (LSL) for the subsequent segment. GSL is active every cycle and LSL is active only when the upper bits are match. Since most MLs are mismatching, most LSLs are not active, thus saving power of the SL, which can significantly reduce the power of the SL. However, this scheme significantly increases the implementation complexity.

In this section, we have looked at a variety of previously presented schemes for reducing the ML and SL power of TCAM. However, SL scheme for power saving of SL is limited to use according to structure and operation method of ML scheme. Therefore, if a proper scheme to reduce the power consumption of the ML and the SL together is applied to the TCAM, the power consumption of the TCAM can be considerably reduced.

3. Proposed TCAM

As we have seen, it is obvious that the power consumption of ML and SL increases as the size of TCAM increases. In addition, many schemes for reducing power consumption of ML and SL have limited use of SL scheme according to ML scheme. Therefore, we propose a TCAM that reduces power consumption of ML and SL by using match-line selective charging scheme and low-swing search-line scheme that can be applied to TCAM.

3.1 Match-line selective charging scheme

Fig. 5 shows the structure of the match-line selective charging scheme used in the paper [19]. This scheme pre-discharges all of the MLs and selectively charges the ML in the match state during the evaluation phase. Thus, this scheme is similar to the current saving scheme, which assigns current differently to ML. Each charging controller block controls the charging and discharging process of the respective ML and the sense amplifier (SA) senses the ML voltage and gives the final match/miss decision [19]. Also, since the ML is low precharged in the operation of the charging controller as in the current saving scheme, an eliminating search-line precharge scheme can be used.

In this scheme, TCAM cells use NOR-type cells. The NAND type cell has the advantage of less power consumption than the NOR type because the ML of the cell is connected in series. However, as TCAM size increases, there is a disadvantage that an unacceptably long search time is required [20]. In addition, NOR-type is preferred over NAND-type because of the less susceptibility to failure due to the variation of process, temperature and supply voltage [21].

Figs. 6 and 7 show the internal circuit and timing diagram of the charging controller used in the match-line selective charging scheme. When the search cycle starts, the charging controller operates in three phases: precharge phase, test-charging phase and selective charging phase.
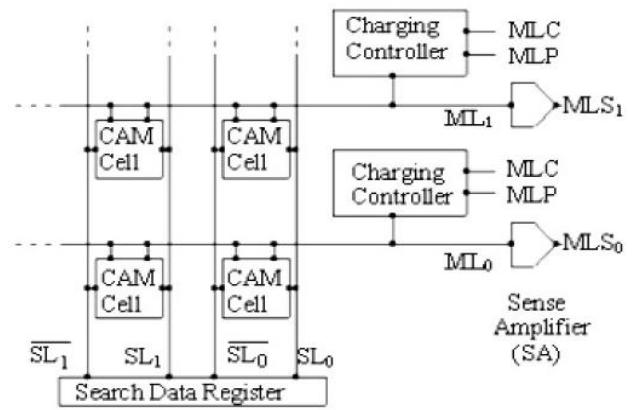

Figure. 5 Structure of the CAM array with the match-line selective charging scheme
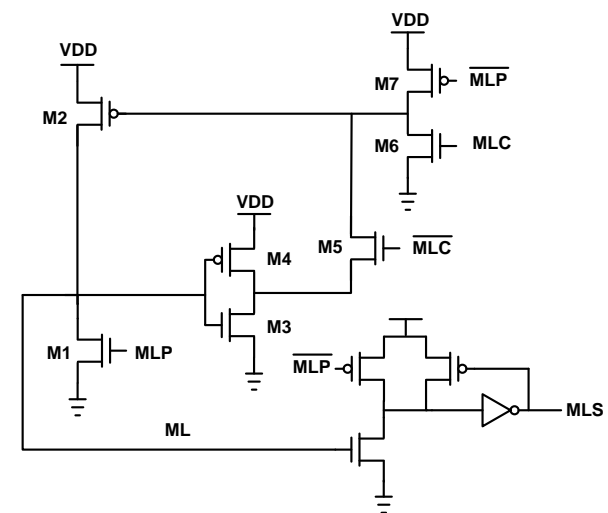

Figure. 6 Internal circuit of match-line charging controller
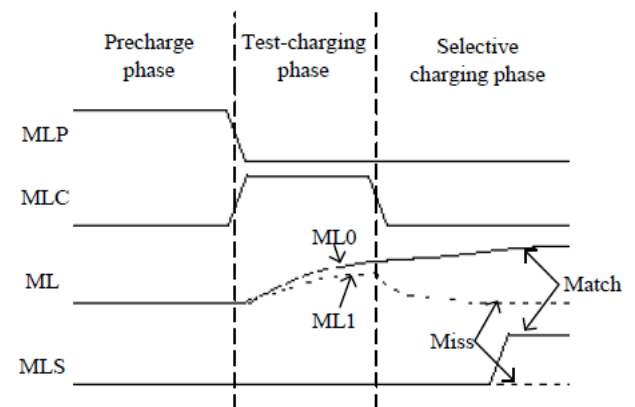

Figure. 7 Timing diagram for a single search cycle

In the precharge phase, MLP is high and MLC is low, so transistors M1, M5 and M7 are turned on and the ML is pre-discharged through M1. Since the ML is discharged, the inverter composed of M3 and M4 outputs high, and M7 is also turned on by the MLP, so that high is inputted to the gate of M2. Therefore, M2 is turned off and charging of ML does not occur.

In the test-charging phase, MLP is low and MLC is high, so transistors M1, M5 and M7 are turned off and M6 is turned on. Since M2 is input low and turns on, it begins to charge the ML. When the ML is charged in the test-charging phase, if the word is mismatch, a path from ML to the ground is formed and charged more slowly than the ML of match word. The MLC remains high until ML in the match state turns on M3.

In selective charging phase, both MLP and MLC remain low. Therefore, transistor M1, M6 and M7 are turned off and M5 is turned on. In the test-charging phase, if the voltage of the ML is charged enough to turn on M3, M2 remains turned on ML continues charging. Since ML is connected to the gate of NMOS transistor connected with conventional MLSA, high is output to MLS when match and low when it is mismatch.

As a result, in the selective charging phase, the ML continues to charge when it is a match, and stops charging and discharges when it is a mismatch, thus saving power by selectively charging the ML. In this operation, except for the fully matched word, the ML of the word which is a mismatch of 1 bit is charged the fastest. Therefore, the MLP must have a reasonable charge time to ensure that the ML of this condition does not output an incorrect result.

3.2 Low-swing search-line scheme

The basic idea of a low-swing search-line scheme is to reduce the SL power consumption by reducing the swing voltage of the SL. This scheme is an Eliminating search-line precharge scheme that directly activates the search data without performing a low precharge of the SL described above. The ML scheme used in this paper can be used because it is a scheme for low precharge ML in the search operation. This scheme eliminates the SL precharge phase and low swings the SL so that the SL can save even more power. Fig. 8 shows the search-line voltage generator for generating the low-swing voltage used in this paper [22]. The search-line voltage generator consists of two comparators using operational amplifier (OPAMP), a resistor R_1-R_4 to generate the reference voltage, and a PMOS transistor and NMOS transistor to hold the low-swing voltages V_H and V_L. Resistors R_1-R_4 generate reference voltages V_{REF_H} and V_{REF_L} according to the resistance ratio from supply voltage V_{DD}. For example $V_{REF\_H} = 0.9V$ and $V_{REF\_L} = 0.3V$ when the resistance ratio is the same at a supply voltage of 1.8V. OPAMP receives the reference voltage and low-swing voltage as input. If V_H becomes lower
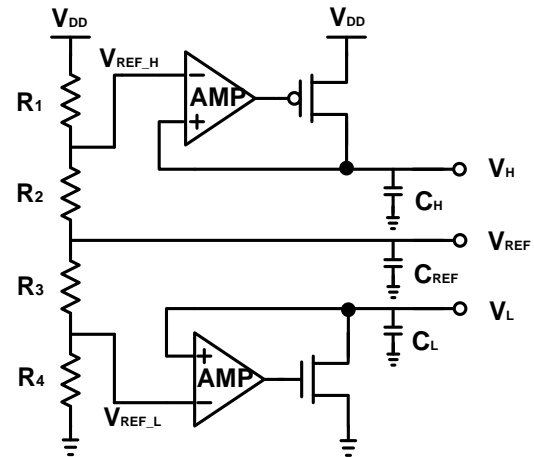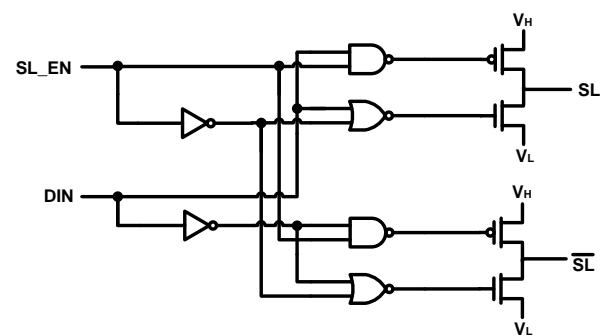

Figure. 8 Reference voltage generator

Figure. 9 Low-swing search-line driver

than V_{REF_H}, it turns on PMOS to raise V_H. Also, if V_L is higher than V_{REF_L}, NMOS is turned on to lower V_L to keep the voltage constant. Since C_{REF} is only used for the reference voltage, a small capacitance may be used, but C_H and C_L must be large enough to supply current to V_H and V_L.

Fig. 9 shows the low-swing search-line driver. SL enable (SL_EN) and data in (DIN) are input to assign the voltage to the SL. By supplying V_H to the driver's supply voltage and V_L to the ground, the SL operates in low-swing and power consumption in the SL is reduced.

4. Simulation results

The simulation of the proposed TCAM is 128bit×128bit array and the supply voltage is 1.8V. CMOS 0.18um technology was used and verified using Hspice simulation.

Fig. 10 shows SL swing of full-swing search-line driver and SL swing of low-swing search-line driver proposed in this paper. Since the resistance ratio is the same, the low-swing SL swing to about 0.45V to 1.35V.

Fig. 11 shows the simulation results of the search operation in the proposed TCAM. In the top

panel, ML0-ML3 represents the match-line voltage of the word from the 0th word to the 3rd word respectively, and below ML3 shows the voltage of the remaining match-line. Also, MLSO0 shows the output of the match-line sense amplifier connected to the match-line of the 0th word in fully match state (ML0). Finally, the MLP and MLC show two control inputs for the operation of the match-line selective-charging scheme. The lower panel shows the voltage swing change of the search-line for the same time. The MLP maintains high for 0.84ns, at which time the discharge of the MLs is complete. When MLP changes to low, the MLC remains high for 0.48ns, where ML0, which is a fully matched condition, can be charged to 0.7V. After that, the MLP also changes to low, and when it is the selective charging phase, only the match-line (ML0)
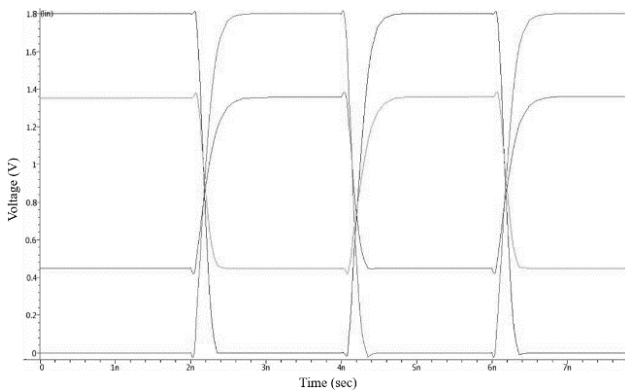

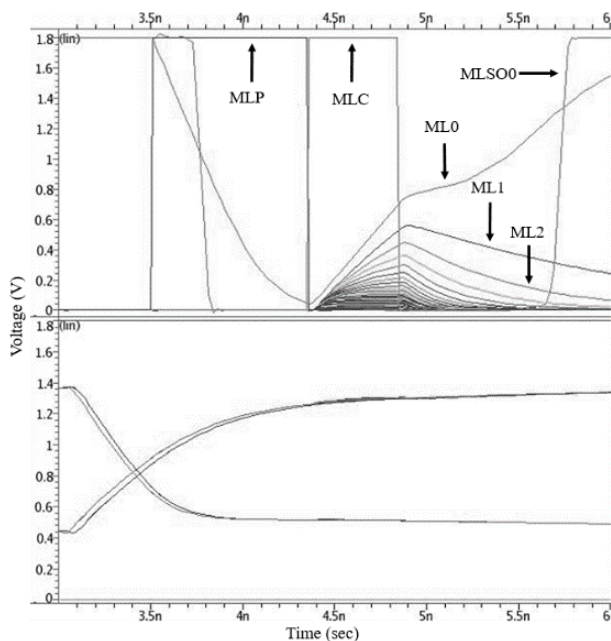Figure. 10 Low-swing and full-swing operation of search-line driver

Figure. 11 The search operation waveform of the proposed TCAM

of the fully match condition word remains high and the remaining MLs are discharged.

In the simulation shown in Fig. 11, word0 (ML0) is set to a fully match condition, and the miss-match bit is incremented by 1bit until word1-word127 (ML1-ML127). Therefore, it can be seen that ML1 is the highest charge except ML0 in the selective charging phase, and the remaining becomes gradually less charged. Therefore, it can be seen that ML1 is the highest charge except ML0 before the end of the selective charging phase. Generally, since the number of words in the mismatch condition is larger than the word in the match condition in the TCAM, the structure that can reduce a considerable amount of power consumption.

Tables 1 to 3 show the comparison of the conventional CAM, conventional match-line selective charging TCAM, and the proposed TCAM, 32bit×32bit, 64bit×64bit and 128bit×128bit array respectively.

Table 1. Comparison with previous TCAM schemes (32bit×32bit)

Scheme	Conventional	Match-Line Selective charging	This work
SL Power (pW)	5.391	2.360	1.249
ML power (pW)	6.128	2.727	2.746
Minimum cycle time (nS)	3	1.3	1.3
Speed, 1/T (MHz)	333	769	769

Table 2. Comparison with previous TCAM schemes (64bit×64bit)

Scheme	Conventional	Match-Line Selective charging	This work
SL Power (pW)	10.60	12.33	6.036
ML power (pW)	12.83	8.422	8.047
Minimum cycle time (nS)	5	1.7	1.8
Speed, 1/T (MHz)	200	588	555

Table 3. Comparison with previous TCAM schemes (128bit×128bit)

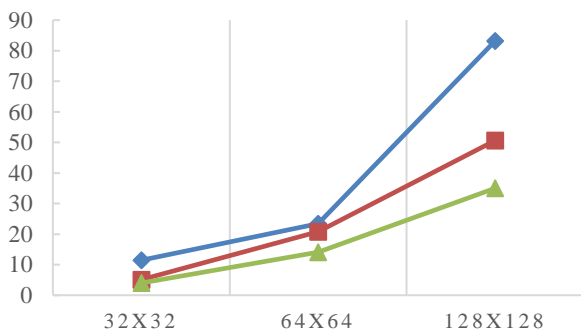| Scheme	Conventional	Match-Line Selective charging	This work
SL Power (pW)	35.45	24.49	8.966
ML power (pW)	47.75	26.12	26.03
Minimum cycle time (nS)	10	2.5	3
Speed, 1/T (MHz)	100	400	333

Figure. 12 Total power consumption of match-line and search-line according to array size change

Fig. 12 shows the variation in total power consumption in the match line and search line with the array size change of TCAM. Conventional schemes show the highest SL power consumption due to low precharging of SL. In the proposed scheme from 128bit × 128bit array, SL power reduced by about 74.7% compared with the conventional scheme, and the SL precharging phase is eliminated, and it is reduced by about 63.3% compared to full-swing ML-selective charging scheme. ML power was reduced by about 45.4% compared with the conventional scheme.

5. Conclusion

In this paper, we proposed a scheme to apply match-line selective-charging scheme and low-swing search-line scheme together to reduce power consumption of match-line and search-line which occupies a large proportion of power consumption in ternary content addressable memory.

In the proposed TCAM, the match line selection charging scheme reduces the match line power consumption by selectively charging the match line in the short selection charge stage without charging all the match lines. Further, by using a low precharge match line scheme, the power consumption of the search line can be significantly reduced. Therefore, the overall power consumption of the TCAM can be reduced and the power consumed by the TCAM in the application can be reduced.

The proposed scheme is simulated using Hspice in a 0.18um CMOS process with 128bit×128bit TCAM. As a result, the power consumption of the match-line is reduced by more than 45.4% and the power consumption of the search-line is reduced by more than 74.7% compared with the conventional scheme in the same simulation environment. The minimum cycle time was also reduced by 70%.

Acknowledgment

References

[1]   K. E. Grosspietsch, "Associative processors and memories: a survey", *IEEE Micro*, Vol. 12, No. 3, pp. 12–19, 1992.

[2]   M. Merbout, T. Ogura, and M. Nakanishi, "On using the CAM concept for parametric curve extraction", *IEEE Trans. Image Process*, Vol.9, No. 12, pp. 2126-2130, 2000.

[3]   M. Nakanishi and T. Ogura, "Real-time CAM-based Hough transform and its performance evaluation", *Machin Vision appl.*, Vol. 12, No. 2, pp. 59-68, 2000.

[4]   S. Panchanathan and M. Goldberg, "A content-addressable memory architecture for image coding using vector quantization", *IEEE Trans. Signal Process*. Vol. 39, No. 9, pp. 2066-2078, 1991

[5]   T.-B Pei and C. Zukowski, "VLSI implementation of routing tables:tries and CAMs", In: *Proc. IEEE INFOCOM*, Vol. 2, pp. 515-524, 1991.

[6]   T.-B Pei and C. Zukowski, "Putting routing tables in silicon", *IEEE Network Mag*., Vol. 6, No. 1, pp. 42-50, 1992.

[7]   A. J. McAuley and P. Francis, "Fast routing table lookup using CAMs", In: *Proc. IEEE INFOCOM*, Vol. 3, pp.1282-1391, 1993.

[8]   N.-F. Huang, W.-E. Chen, J.-Y. Luo, and J.-M Chen, "Design of multi-filed IPv6 packet classifiers using ternary CAMs", In: *Proc. IEEE GLOBECOM*, Vol. 3, pp. 1877-1881, 2001.

[9]   G. Qin, S. Ata, I. Oka, and C. Fujiwara, "Effective bit selection methods for improving performance of packet classifications on IP routers", In: *Proc. IEEE GLOBECOM*, Vol. 2, pp. 2350-2354, 2002.

[10] H. J. Chao, "Next generation routers", In: *Proc. IEEE*, Vol. 90, No. 9, pp.1518-1558, 2002.

[11] G. Kasai,Y. Takarabe, K. Furumi, and M.Yoneda, "200 MHz/200 MSPS 3.2 W at 1.5 V Vdd, 9.4 Mbits ternary CAM with new charge injection match detect circuits and bank selection scheme", In: *Proc. IEEE Custom Integrated Circuits Conf. (CICC)*, pp. 387‑390, 2003.

[12] H. Miyatake, M. Tanaka, and Y. Mori, "A design for high-speed low power CMOS fully parallel content-addressable memory macros", *IEEE Journal of Solid-State Circuits*, Vol. 36, No. 6, pp. 956–968, 2001.

[13] C. A. Zukowski and S.-Y. Wang, "Use of selective precharge for low power content-addressable memories", In: *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Vol. 3, pp. 1788–1791, 1997.

[14] K. Pagiamtzis and A. Sheikholeslami, "Pipelined match-lines and hierarchical search-lines for low-power content-addressable memories", In: *Proc. IEEE Custom Integrated Circuits Conf. (CICC)*, pp. 383–386, 2003.

[15] K. Pagiamtzis and A. Sheikholeslami, "A low-power content-addressable memory (CAM) using pipelined hierarchical search scheme", *IEEE Journal of Solid-State Circuits*, Vol. 39, No. 9, pp. 1512–1519, 2004.

[16] I. Arsovski, T. Chandler, and A. Sheikholeslami, "A ternary content addressable memory (TCAM) based on 4T static storage and including a current-race sensing scheme", *IEEE Journal of Solid-State Circuits*, Vol. 38, No. 1, pp. 155–158, 2003.

[17] I. Arsovski and A. Sheikholeslami, "A current-saving match-line sensing scheme for content-addressable memories", In: *IEEE Int. Solid-State Circuits Conf. (ISSCC)* Dig. Tech. Papers, pp. 304–305, 2003.

[18] K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory (CAM) circuits and architecture: a tutorial and survey", *IEEE Journal of Solid-State Circuits*, Vol. 41, No. 3, pp. 712–727, 2006.

[19] A.B.M.H Rashid, M.M. Hasan, and A.B.M.H. Rashid, "A novel match-line selective charging scheme for high-speed, low-power and noise tolerant Content-Addressable Memory" *International Conf. On Intelligent and Advanced Systems (ICIAS)*, 2010.

[20] I. Arsovski and A. Sheikholeslami, "A mismatch-dependent power allocation technique for match-line sensing in content-addressable memories", *IEEE Journal of Solid-State Circuits*, Vol. 38, No. 11, pp. 1958-1966, 2003.

[21] A. Mupid, M. Mutyam, N. Vijaykrishnan, Y. Mie, and M. J. Irwin, "Variation analysis of CAM cells", In: *Proc. IEEE Int. Symp. Quality Electronic Design*, pp. 335-338, 2007.

[22] B. D. Yang, Y. K. Lee, S. W. Sung, J. J. Min, J. M. Oh, and H. J. Kang, "A Low Power Content Addressable Memory Using Low Swing Search Lines", *IEEE Transaction on Circuits And Systems*. I: Regular papers, Vol. 58, No. 12, pp. 2849-2858, 2011.