



LOW POWER CONCEPT FOR CONTENT ADDRESSABLE MEMORY (CAM) CHIP DESIGN

Dejan Georgiev

PhD Student, Faculty of Electrical and Information Technologies- Skopje, Macedonia

ABSTRACT: A Content Addressable Memory (CAM) is a memory unit that performs single clock cycle content matching instead of addresses. CAM's are vast used in look-up table functions , network routers and cache controllers. Since basic lookups are performed over all the stored memory information there is a high power dissipation. In reality there is always trade-offs between power consumption, area used and the speed. Here is presented an conceptual abstraction for content addressable memory chip at architecture level with reduced power requirements based on combination and modifications of power saving techniques.

Keywords: power, architecture, XOR port, CAM cell, chip, design

I.INTRODUCTION

Content Addressable Memories (CAM) are fast data parallel search circuits. Unlike standard memory circuits , for example Random Access Memory (RAM) data search is performed against all the stored information in single clock cycle. In fact CAM is outgrowth of RAM. While CAM's are widely used in many applications like memory mapping, cache controllers for central processing unit, data compression and coding etc. it primary application is fast Internet Protocol (IP) package classification and forwarding at high speed network routers and processors. IP routing is accomplished by examination of the protocol header fields i.e. the originating and destination address, the incoming and outgoing ports etc. against stored information in the routing tables. If a match is registered the package is forwarded towards the port(s) defined in the table. On very high speed networks and huge traffic volume the task is to be performed in fast and massive parallelism. However, managing high speeds and large lookup tables requires silicon area and power consumption. The power dissipation , silicon area and the speed are three major challenges for designers. Since there is always trade-off between them , reducing one without sacrificing the others is a is the main treat in recent research for large CAMs [2]. Finding a solution could be approached on circuit or architectural level. Here is presented the later.

The rest of the paper is organized as follows. Section 2 refers to a related work in the scope. Section 3 states the problem to be solved and introduction to the techniques that are implement. Also an architecture design of proposed CAM chip design is presented. Section 4 is for evaluation the results and discuss the issues. The conclusion of the work is given in the last section 5.

II.RELATED WORK

CAM hardware has been available for decades and many research are addressed to development of high capacity and effective CAM designs at circuit, architectural and application level. A lot of projects are leaning towards "real live" application for effective algorithms for package forwarding based on CAMs and it extended version i.e. Ternary CAM (TCAM) [1]. CAM memories enhanced with "don't care" states are used for more complex project like hardware based Network Intrusion Detection and Prevention Systems (NIDPS) [9]. At "lower" level designs many papers introduce methodologies and optimization to speed , power and physical circuit resources. Authors of [2] in detail describe the principle of CAM functions at transistors and circuits level including core cells, match line and search line structures and power consumption formulation. Also power and area reducing techniques are presented on the circuit level. Practical design on architecture level is presented by [4]. The proposed CAM chip design is based on modification to the RAM chip circuit explained in [5]



III.CAM CHIP DESIGN

A basic CAM cell function could be observed as twofold: *bit storage* as in RAM and *bit comparison* which is unique to CAM. At transistor i.e. circuit level CAM structure implemented as NAND-type or NOR-type and its variants has been explained by [2]. But at architectural level bit storage uses simple (S)RAM cell and comparison function is equivalent to XOR i.e. XNOR logic operation. Thus our elementary chip cell design is abstracted as a cross product of SRAM and XNOR circuits. Figure 1 represents the logical symbol and the circuit compilation.

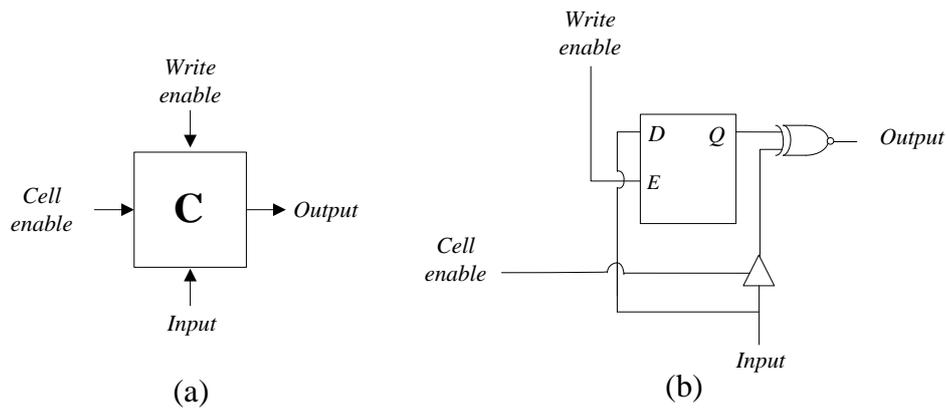


Figure 1. CAM cell : (a) logic symbol; (b) circuit

The *input* signal is one bit value from the *search data register* i.e. the input word to be compared against all the values stored in CAM arrays or the value to be stored in the CAM cell. *Cell enable* signal allows or prevent comparison i.e. matching process meaning XOR-ing the stored bit value in the Flip-flop and the input bit. It should be mentioned the extended truth table of three-state buffer presented on Table 1 where *x* represents the input signal and *y* is the output signal. "Z" denotes high-impedance or practically disconnected line or switch on/off.

TABLE 1. Truth table for three-state buffer

| E | y |
|-------------------------|---|
| 1 | x |
| Others (0,Z,X,'-' etc.) | Z |

The power dissipation in every single cell is present either on matching state or even at missing state since the comparison process is performed in both cases. The purpose of *Cell enable* signal i.e. three-state buffer is to "disconnect" the cell from the matching process and thus from the power consumption process.

A. One-cell segmentation power reducing scheme

As was mentioned earlier at regular content addressable memories the data search process is uniformly performed along all the cells of the array, thus producing heat in every single cell. Though some proposed techniques can allay the issue. When performing a search if the first few bits do not match there is no point in checking the remaining bits. *Selective pre-charge* initially searches only first *n* bits and only searches the remaining bits for words that matches first *n* bits. With uniform random data it only have to search $(1/2)^n$ of the rows. For *n*=3 this will save about 88% of the match line power [2],[3]

Selective pre-charge scheme basically divides the mach line in two segments. In general following the same concept it can be divided in many number of segments thus forming a *pipeline*. If any stage is miss the subsequent stages are shut off resulting in power saving. The drawback of this scheme are the increased latency and area overhead due to the pipeline stages. Here is shown a power saving design scarifying the speed i.e. increased delay but retaining the same circuit area. The basic idea behind the concept is the segmentation in the mach line in a manner that every CAM cell form a segment for its own as it is presented on Figure 2.

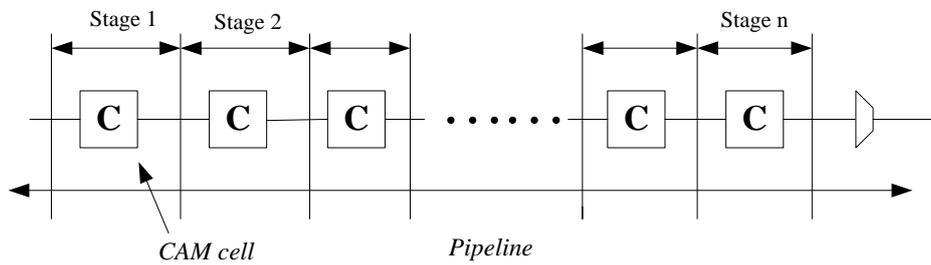


Figure 2. One-cell pipeline segmentation

The main benefit for the proposed scheme comes from implementation with CAM cells shown on Figure 1. Namely, the output of a cell is simply the cell enable signal for the successive bit comparison thus mitigating extra gates to transfer the results from the cells. The disadvantage is increased propagation delay that comes from the three-state buffer and XNOR gate at each cell. Typical CAM consist words length ranging from 36 to 144 bits and in practice it should be acceptable delay value. It should be noted that one cell segmentation approach presented here is a conceptual view rather than real power saving scheme that can be achieved on circular level.

B. Parity check pre-computation power reducing scheme

Content addressable memories are widely used at network routers for IP package forwarding or at firewalls and NIDPS systems for package filtering. For IPv4 the basic filter set is 5-tuple defined over header fields {Source IP, Destination IP, Source port, Destination port, protocol} each of which equal to {32 bits, 32 bits, 16 bits, 16 bits, 8 bits} long respectively. Follows that 104 bits words CAM are required. On the other hand most of the filter and NIDPS rules are defined over port ranges e.g. [1024:2048]. Implemented with CAM without "don't cares" it will acquire word fields repetition where the port bits are increased by 1. It not only requires huge memory area but unnecessary power lost in bit comparisons even if pre-charge or pipeline method is used. For example, for fixed source IP and destination IP addresses computation will be performed over all the first 64 bits regardless of residual bit fields. In that regards an improvement can be achieved by statistical pre-computation along word's bits.

Pre-computation stores some extra bits derived from the stored word and it is used in the initial search before the search of main word. If the initial search fail, the main word search is aborted and thus saving power. The schematic concept is presented on Figure 3.

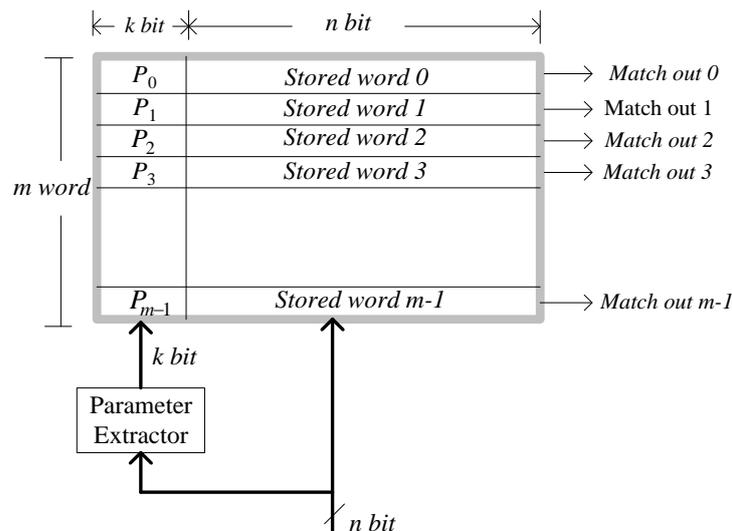


Figure 3. Conceptual view of CAM memory with pre-computation (adopted from [8])

One method uses pre-computation circuits to count the number of ones and stores this data along with the word in binary format. The number of bits reserved for pre-computation is $\log_2(n + 2)$ [8]. As first step pre-computation bits



are compared for every stored word and for those that match the process continues with the data search. One possible solution at circular level for one's count parameter extractor consist parallel and serial connected Full Adders (FA) [7]. To perform the computation the data word bits are grouped in three bit segments. It worth noting that simple FA at circuit level is implemented with two AND, two XOR and one OR gates [5]. The main drawback for 1's count pre-computation CAM raises for long data word. The first issue is the complex computation scheme created as enormous number of full adders assemble. The second issue we grumble is the lost of silicon area reserved for storing the pre-computation bits.

As a compromise here is proposed novel and simple pre-computation algorithm. This solution suites well in the previously promoted one-cell segmentation pipeline power reduction scheme and is consisted of follow: instead of counting the number of 1's (or 0's) we only check the parity of 1's in the data word. For uniform distribution a half power reduction is achieved. Compared with one's count scheme we have less power save, but in term of complexity and area used this scheme provides improvement. More over bit parity computation requires only one bit ($k=1$) for storing the result. Indirectly, reduced computation complexity has speed increment implications. Simple logical circuit for bit parity computation is implemented with only XOR gates as shown on Figure 4. The outcome result is bit '1' for odd number of one's in data word and '0' for even number of ones.

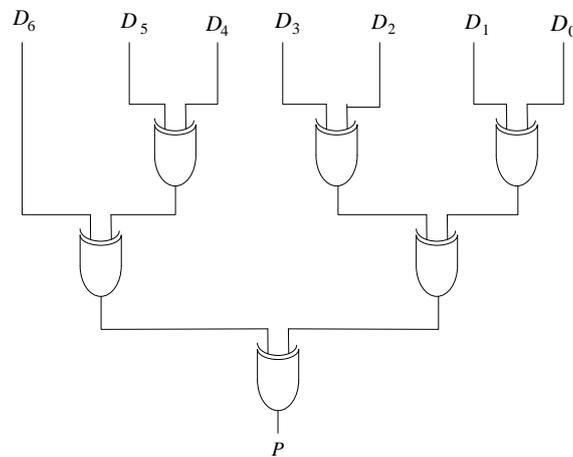


Figure 4. Logical circuit for bit parity parameter extractor

In is interesting that XOR-ing can be performed over arbitrary positioned bits instead of adjacent ones. Similar circuit has been used at so called Block-XOR computation block to achieve uniform comparison [7].

C. CAM chip architecture

Based on approaches discussed so far , more specifically with combination of proposed one-cell segmentation pipeline and parity check power reduction mechanisms we can design complete architecture of a content addressable memory as is shown on Figure 5. It actually represents small 4x4 CAM chip where each row contains the stored word including the one bit of parameter memory. The 2-to-4 decoder is used for row selection only for write function. It is important to note that CAM chip does not require clock signal, except for the data register where the new stored value for comparison is synchronized with a global system clock. Since the three-state buffer at each memory cell acts as a switch for a bit comparison CAM function , the CAM enable (CE) signal can be assumed as global switching signal for the CAM chip.

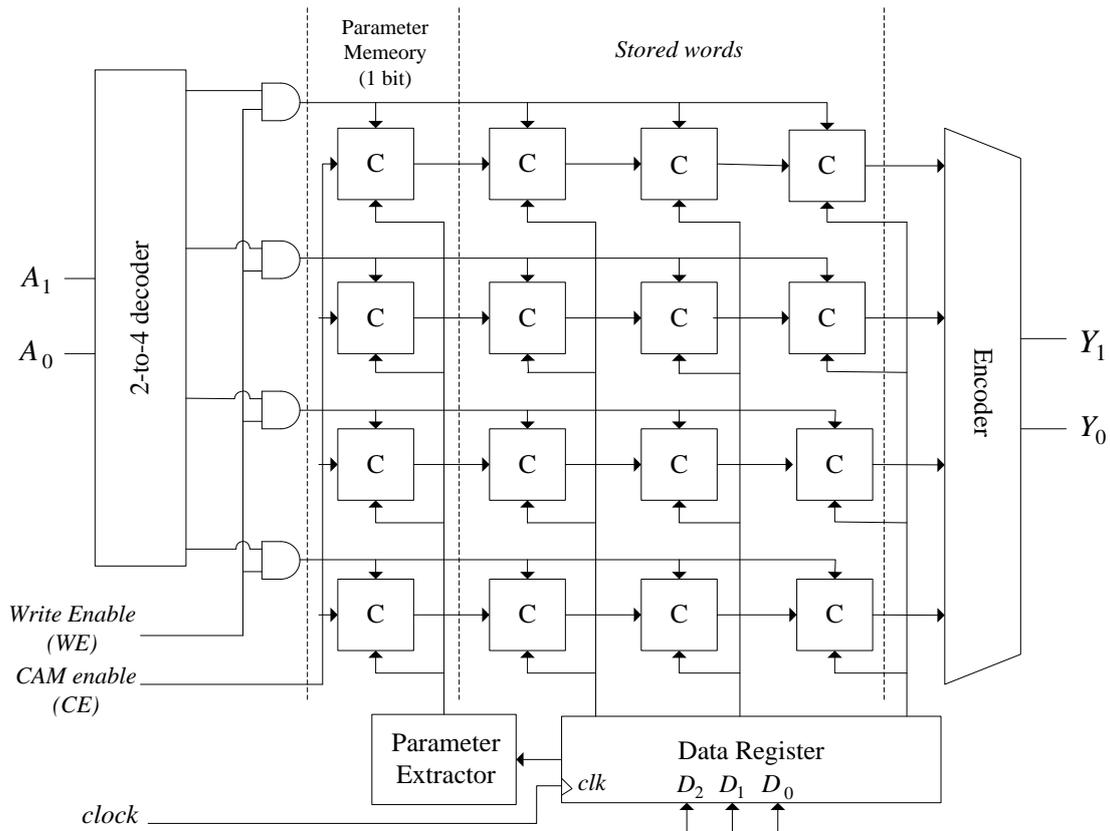


Figure 5. CAM chip circuit

When CE=0 all parameter memory cells are off yielding to disabled searching processes for remaining cells. The output of 4 rows CAM memory is 4-to-2 encoder generating binary representation of the row number where match occurred. Most often priority encoders have been used.

IV. PERFORMANCE EVALUATION AND DISCUSS

Power reducing techniques are always in counterbalance to the logic area or the processing speed. Here presented approach compared to the previous designs requires equal or less area resources, especially the parameter extractor. All the power reduces are attained on increased timing delays. Since the match process should be accomplished in a single clock cycle the overall system delay is to be projected to fit within the clock period as shown in Figure 6. The overall delay T for a complete match is given as a sum of all the successive delay of each component

$$T = t + n\tau$$

where t represents the delay from parameter extractor and τ is the delay of a CAM cells.

One possible improvement in term of speed and complexity can be attained by segmentation of the pre-computation bits, meaning performing checkups only for some, for example the last have bits instead of all the bits in the data words.

The power saved in the one cell segmentation is better than regular pipelining, but the bit parity check pre-computation achieves less reduction compared to the previous approaches.

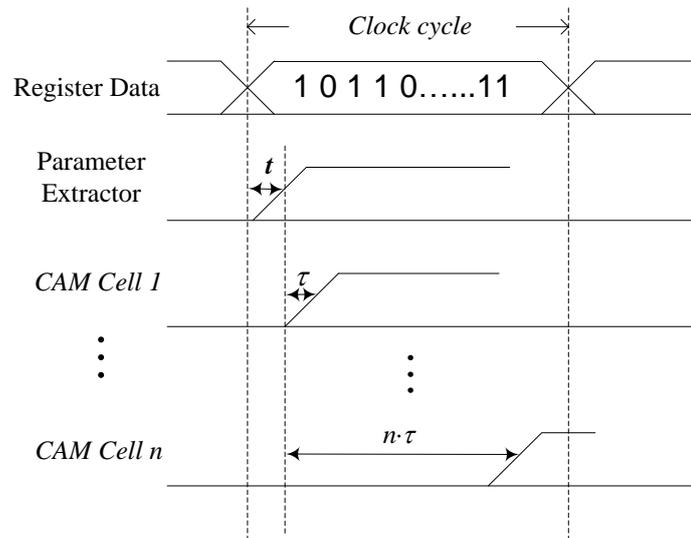


Figure 6. Timing diagram for overall delay

Another design consideration is the practical realization of three state buffer and XNOR gate with input signal 'Z' as an operand. For simulation purposes using VHDL using IEEE.STD_LOGIC_1164 package i.e. std_logic signal definition our CAM cell design suites well.

V. CONCLUSION

In this paper an overall CAM chip model designed at architecture level is presented. Also combination of two power reducing techniques: the pipelined power scheme and modified pre-computation based approach is introduced. First we segmented the pipeline to a single cell stage where each positive match in a given stage acts as switch on for the next state matching thus showing that there is no need for extra circularity required at each cell to gate the comparisons as was stated in some earlier research. The second improvement is proposed is by new pre-computation algorithm. Bit parity check requires less parameter memory space and perform quickly. Of course, it come with some drawbacks like less power reduction. The main challenge still remain: based on the proposed approach to project CAM grid at circuit level.

REFERENCES

- [1]. David E. Taylor , Edward W. Spitznagel "On using content addressable memory for package classification", Applied Research Laboratory, Washington University in Saint Louis, 2005
- [2]. Kostas Pagiamatzis, Ali Sheikholeslami "Content-addressable memory (CAM) circuits and architectures: a tutorial and survey" , IEEE Journal of Solid-State Circuits, Vol.41, No.3, March 2006
- [3]. Scott Beamer, Mehmet Akgul "Design of low power content addressable memory (CAM) ", Department of Electrical Engineering & Computer Science, University of California, Berkley
- [4]. Qutaiba Ibrahim "Design & implementation of high speed network devices using SRL16 reconfigurable content addressable memory (RCAM)" , International Arab Journal of e-Technology, Vol.2,No.2, June 2011
- [5]. Enoch O. Hwang "Digital logic and microprocessor design with VHDL" , La Sierra University, Riverside
- [6]. Jui-Yuan Hsieh, Shanq-Jang Ruan "Synthesis and design of parameter extractor for low-power pre-computation based content addressable memory using gate-block selection algorithm", Department of Electronics Engineering , Taipei , Taiwan
- [7]. M. Arun, A. Krishnan "Comparative power analysis of pre-computation based content addressable memory", Journal of Computer Science 7(4):471-474,2011
- [8]. Jinn-Shyan Wang "Low-power high-speed content addressable memories" , National Chung Cheng University, Taiwan
- [9]. Haoyu Song, John W. Lockwood "Efficient packet classification for network intrusion detection using FPGA", International Symposium on Field-Programmable Gate Arrays (FPGA'05) ,Monterey ,CA, Feb 20-22, 2005