A SURVEY PAPER OF TCAM ARCHITECTURES

CH.V.D Ashok Kumar, M.tech , ECE Department, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru

Y. Sri Chakrapani, Assoc. Proff, ECE Department, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru

Abstract:

Ternary content-addressable memory (TCAM) is very important in forwarding and classification of packets in every network. To improve the power efficiency and reduce the delay, researchers designed multiple architectures on TCAM using multiple memories. TCAM architectures are more usually found in software defined networks, which can be controlled and programmable using a centralized controller. This centralized controller acts as a brain power of the whole network, which directs and regulates the flow of data between the switches and routers, that are located in control plane of the network. The control plane is responsible for routing and serves as central intelligence. In modern FPGAs, single search operation is missing to activate the entire circuit, which is used for implementing TCAM. Hence, different architectures are analyzed to find the better TCAM architecture that gives sufficient results, based on networking parameters

Keywords: TCAM, FPGA, Delay, Power, SDNs

I. INTRODUCTION

Ternary content-addressable memory (TCAM) is the fastest memory address locations where the input data is propagated in one clock cycle only. The critical path for the transmission and receiving of the data is very less in the TCAM architectures. Hence, these TCAM architectures are widely used in the multiple applications. There is small difference between the general RAM architectures and the TCAM architectures. In RAM the data processing is done in two ways. Initially the data has to be stored based on the address location provided by the input data and the data is stored in the specific location either the data should be '1' or '0', and some cases the data is stored in the don't care condition and high impedance states. The high impedance is represented by 'z' and don't care conditions are represented by 'x'.

The storing of the input data is quite different in TCAM architectures compares with the other architectures. In TCAM the address locations provided by the input data is stored by the contents of the data word instead of address locations and output address locations. In this the data is searched based on the matched algorithms and search algorithms. Here, the data is searched parallel to reduce the computation time and to reduce the power consumption. When the data is found in the particular address location it returns the input data with a list of multiple address locations. This search operation is also used in the RAM architectures to improve the speed of the design, while reducing the iteration time and comparing the read and write enable signals for every search request. Hence, the time consuming for search operation in RAM is more when compared with the CAM architectures, due to this these TCAM architectures are widely used to improve the speed of the operations.

These CAM architectures are used in multiple applications to reduce the critical path of the design and with the use of parallel processing is one more advantage for the usage of CAM architectures. These CAM architectures are used to accelerate the applications mainly on the network applications. The memory used here is the parallel processing data and these data is transmitted and received at the same time and these operations are done by the search algorithms. These are used for the storage of data and these are applied in the fields of radar signal tracking, artificial intelligence, image processing and many other applications.

These are used in the FPGA's for the acceleration of the hardware architectures and the parallel processing of the data is done in the large look up tables and mainly for the low power applications. The data processing is mainly done in the networking area and here the data classification is done based on the packets and the transmission of these packets is done based on the forwarding of the TCAM architectures. These are designed in such a way that these are less configurable and the cost of the design is too expensive. Due to the speed of the operating time these are used in the network routers using the look up table and the address locations used here are done based on the search algorithms. When the digital world is upgrading towards the software defined networking (SDN), which can be flexible for the design and yet to achieve high performance in terms of hardware usage and the configurability. Hence, FPGA's are the devices where the data can be processed in a high speed manner, hence these TCAM architectures are widely used in the FPGA's and ASIC based architectures.

The FPGA's are used for testing the multiple modules and give the results in the hardware manner. Hence the processing time for the FPGA's need to be very fast and the critical path of processing the data should also achieve low power consumption. Then there is a need of implanting the TCAM memories in the FPGA's to improve the performance of the FPGA's. Hence there is a need of replacing the existing memory with the high speed TCAM memory architectures.

II. LITERATURE SURVEY:

CAM architecture is designed based on the hashing technique, which is evaluated in the [9] which is having the disadvantage of the hashing collisions and the overflow of the data. The data stored in the memory is done based on the address locations which can improve the performance of the CAM architectures. While increasing the number of elements in the storage capacity and the performance of the hashing method is degradable. The design is modified to improve the performance which can emulate the binary CAM instead of the TCAM.

To improve hashing methods are evaluated in [8] which show the different TCAM architectures. Based on the hashing technique, there is overflow of the data and collision of data while transmitting and receiving the data. Different search algorithms are introduced to reduce the drawbacks introduced in the hashing methods. In search algorithms, the stored data can be operated with the don't care conditions and these don't care conditions are replaced with the large memory locations based on the address locations. The memory utilized is not efficient for the search algorithms and used to improve the performance of the design.

These hashing techniques are not effective the data in which potential collisions which are inefficient in handling the data without providing any deterministic performance. In the hashed CAM architectures, the new TCAM achieves better deterministic searching algorithm which can perform effectively and reduces the memory utilization. SRAM based CAM architectures which are designed by using pipeline structure which can take multiple clock cycles to achieve the output data at the desired address location. The memory utilization of these architectures is not effective. Hence TCAM architectures achieve better throughput on a single clock cycle with better utilization of memory.

The shortcomings which cause unavoidable conditions in the SRAM based CAM's. The size of the memory in TCAM is represented with the number of bits in TCAM word and these are parallel in structure. The arrangement of memory is allocated in the matrix format which is available in the rows and columns. The size of the memory increases exponentially along with increasing the number of bits propagating in the packets. This method is practically induced for a large bit pattern. The design is suitable for the portioning of the data and it supports the arbitrary of the large words.

A new method is for the TCAM architecture to improve the performance in [15]. With the increase in the number of bits in CAM architecture to reduce the delay of the architecture. RAM based CAM in works in the descending order, on the arrangement of the data, which is used in the real application where data is completely random in structure. The original data is arranged in the ascending order and the complete data is preserved, in this method. The design is mainly focused on the delay and power consumption to improve the performance. This TCAM is evaluated for propagation of large bit data and to preserve the original addresses, which is suitable for portioning methodology.

The performance of the CAM architectures is improved with the integration of the CAM and RAM to reduce the disadvantages in the CAM architecture. The arrangement of the TCAM done by separating into multiple groups based on some distinguishing bits in the input data. Each group is separated with the most possible matches those are available in the TCAM words. The data is propagating in the real applications which are totally random, while reducing the time delay. This method achieves a good TCAM architecture which are used in the SRAM, to emulate the overall performance of the TCAM functionality.

The TCAM architectures in [10] for the processing of input data in the FPGA boards in which the TCAM are introduced in the memory blocks with some logic blocks those require pre-processing of the CAM input data which is stored in the RAM blocks. These block based RAM architectures are designed for the specific size and those are configured for the TCAM architectures in the memory compared with the other CAM architectures. The static RAM cells are useless because the data processed in the block ram of the FPGA boards in the size of the design. In this case the ram cells are not used for the declaration of the block based ram cells. In fact for the large memory distributed RAM is not efficient in the TCAM architectures.

The FPGA based TCAM architectures are mostly designed based on the RAM based architectures to reduce the distributed memory resources for the propagation delay. The RAM based TCAM architecture which can achieve the better performance compared with the previous CAM architectures. The memory storage in the TCAM architectures are processed with the address locations and this address location is separated into multiple sets of BRAM units and the memory usage for these address locations is inefficient. The configuration of the TCAM architectures are remapped with the time consuming and the hardware overhead is more in the FPGA boards.

Different CAM architectures are evaluated in the binary CAM and TCAM architectures while using the BRAM structures and the shift registers are evaluated in the FPGA's in different applications. The SRAM based CAM architectures are evaluated in the storage inefficient architectures [9]. which can employ the LUT utilization for the shift registers, which can reduce the resource efficiency.

The TCAM architectures are presented in a new method to improve the performance of the architecture which can reduce the delay and improves the accuracy of the design. In the new TCAM architecture the throughput is improved and the searching algorithm achieves better results for the propagation of the input word. The TCAM architecture is designed based on the multiplexer to reduce the switching activity and the power consumption. The resources are utilized for the LUT RAM architectures which are evaluated using the comparison circuitry to reduce the programmable FPGA boards. The outputs are compared based on the and cascaded and these are evaluated in the multiple bits of the TCAM word. The TCAM architecture achieves better throughput and the contents of the TCAM are evaluated based on the search algorithm to reduce the complexity of the design. The SRAM based TCAM is presented in updating the new mechanism. The latency of the input word depends on the don't care conditions and those bits are updated on the sub words. The TCAM approach is evaluated by using the BRAM resource allocations with the worst case update latency based on the clock cycles.

III. ARCHITECTURES OF DIFFERENT TCAM's

3.1 ARCHITECTURE OF UE-TCAM:

The layered UE-TCAM architecture is shown in the Fig.1.In this architecture each layer is represented by the layer architectures. The layer architecture is shown in the Fig.2 . The CAM architecture has L layers which are supposed to match the address lines which are fed to CPE, which can selects the matching address among PMA's. The components are in the TCAM is

included in the N bit SRAM units. The operations performed in the layers are and operations and a layer is used in the priority encoder.
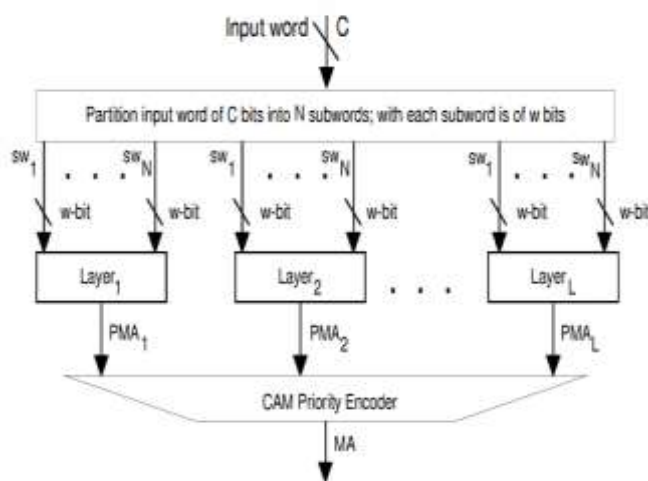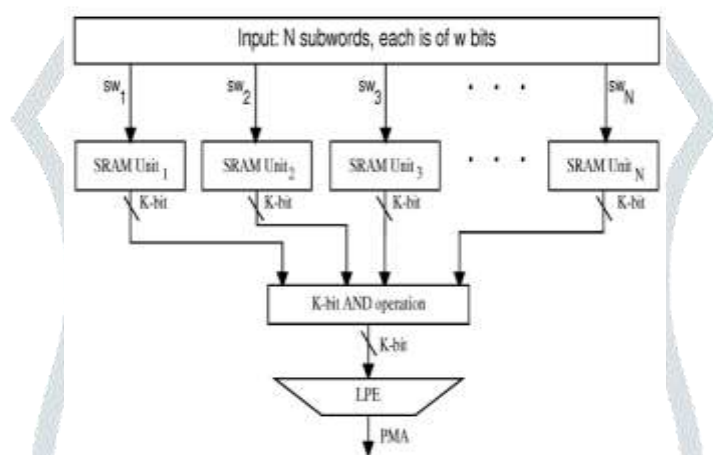

Fig. 1. Layered Architecture of UE-TCAM.

Fig. 2. Architecture of a layer of UE-TCAM using SRAM

SRAM unit: The SRAM unit consists of the size width specific bit length and the subset of the original address locations from the conventional TCAM architectures are shown in Fig 2. The maximum possible combinations of fixed width of bits are multiple bits where each combination represents the sub set of word length and those are presented in the TCAM architectures. The new address locations are invoke the address data corresponding the rows and columns of the bit length. The multiple rows are read by the corresponding sub words, those are the bit wise and operation and the results are then forwarded by the processing of the input data.

Layer priority encoder: These LPE and PMA are used to select the output of the input bit length.

3.2 G-AETCAM ARCHITECTURE:

The G-AETCAM architecture is a matrix of the multiple cells shown in Fig.3. which are known as the gate based and area efficient architectures. Those architecture are the cell consisting for masking the elements which are the storage element when the comparison gate and masking are the multiple cells in the architecture. This architecture consists of 16 ram cells and each consists of four different rows and columns which are in the below design.
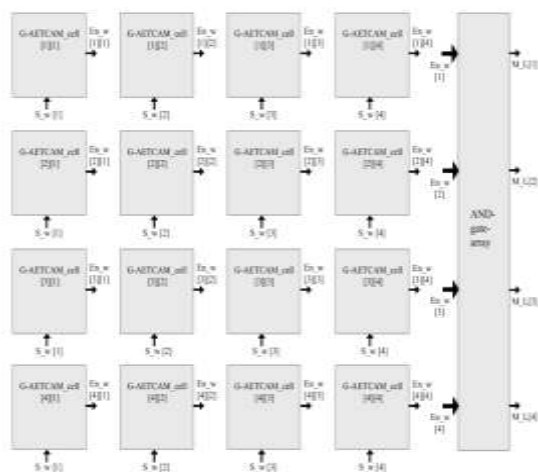
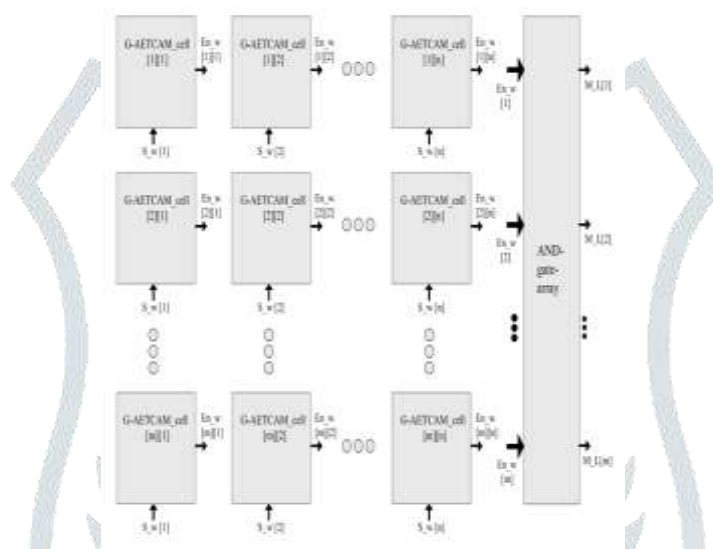**Fig .3 . A 4×4 G-AETCAM Architecture.**

Fig.4 .Generalized (m×n) G-AETCAM Architecture

The generalized architecture of G-AETCAM is shown in the Fig.4. This architecture is represented by the number of the word length and each input word data is splatted into the number of bits in each word. The TCAM memory of the input data width which can show the multiple rows and columns.The number of cells which are evaluated in the product of the m and n rows and columns of the design. The and gates which are used in the number of the multiple bit lengths. The encoded words which are propagated to the input word length and those are provided by the priority encoder which is processed by the search algorithm yields the high performance and reduce the memory of the new architecture.

3.3 DURE TCAMARCHITECTURE:

The DURE algorithm is shown in the Fig.5. This architecture utilized the onchip memory allocations which are distributed in the RAM cells which are available in the state of the FPGA's. The FPGA consists of the LUT RAM cells which are configured on the read and write signals based on the DURE algorithm. The LUT's are evaluated in the multiple FPGA structure to reduce the delay and the same cells are the write and read signals those are the cells using data ports to read and write the signals. The address locations of the each port are calculated based on the input LUT's and these are compared based on the same locations to read and write the signals. The read and write signals are processed in the data and these read ports are in multiple LUT's in which the data is processed parallel to reduce the delay and reduce the power consumption. The data is processed to improve the performance of the design and these are propagated in multiple address locations to reduce the switching activity of the TCAM architecture.
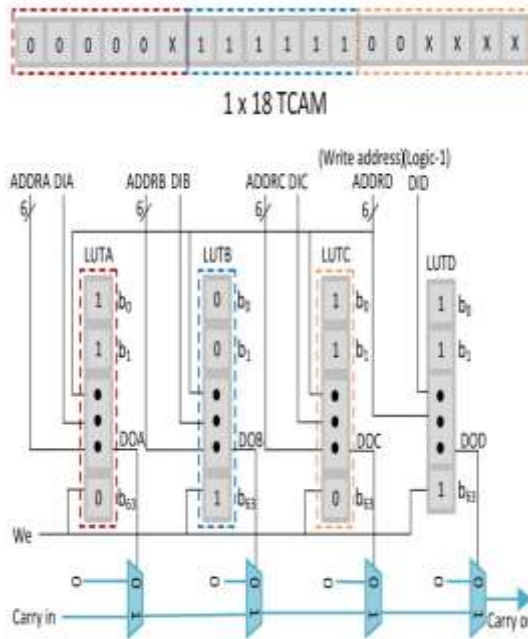
Fig. 5. Architecture of the BM block of DURE a 1×18 TCAM on the FPGA.
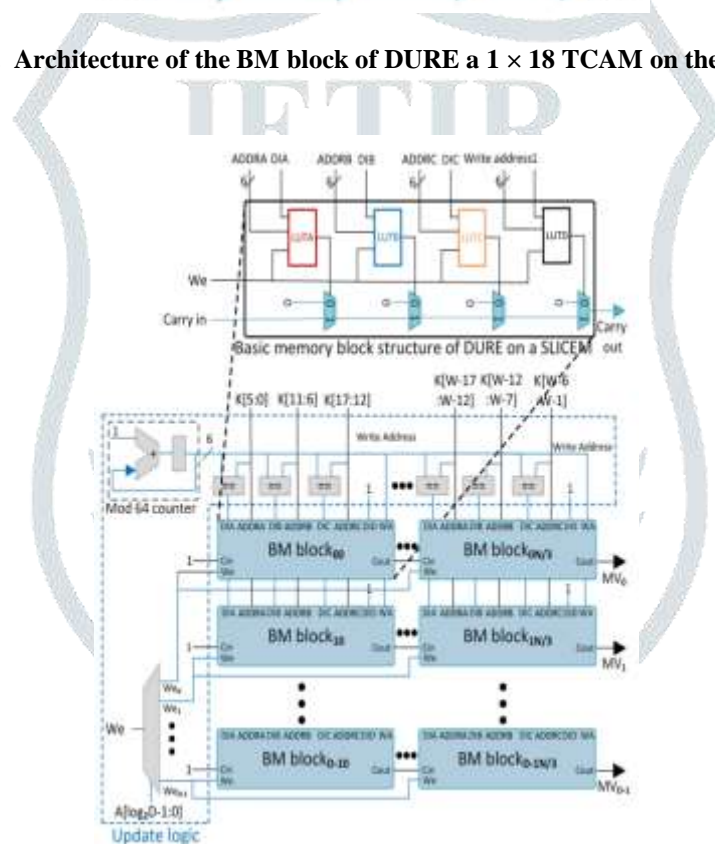

Fig .6. Architecture of the extended DURE.

The above architecture in which DURE algorithm is shown in Fig. 6. In this architecture the data processed in the array of the binary RAM architectures in which the data is processed in cascade manner to reduce the complexity of the design. The input data word is splatted and processed in the FPGA architecture and these are processed in the sub blocks in which the basic blocks are mapped into multiple address locations in which the data is processed into multiple distributed memory locations. The processed data is passed into the multiplexers and the data is processed throughout the multiplexer with multiple address locations.

3.4. RPE-TCAM ARCHITECTURE:

The RPE TCAM architecture is shown in the Fig. 7. This architecture is based on the storage of the flip flops in which the data is compared with the help of comparator and those are processed to improve the performance of the design. The input data word is stored in the memory allocation to reduce the complexity based on the address locations. The masking of the data is processed with the logic '1' and logic '0' while storing the data with the don't care conditions. The new extended structure is divided into the multiple array of the input word to improve the performance with the reduction of the power consumption.
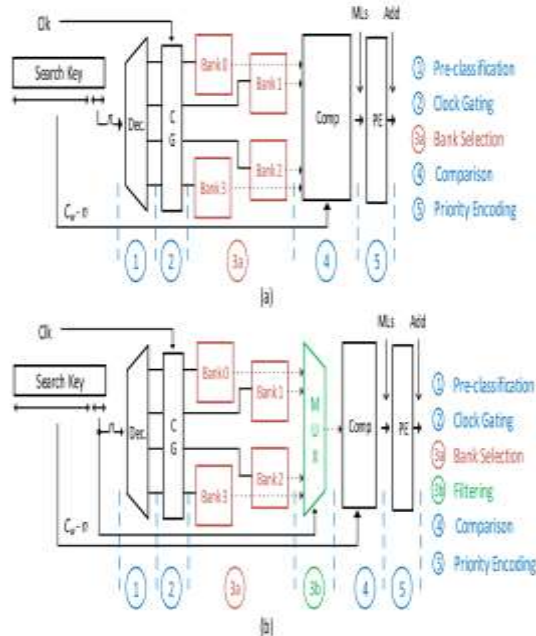
Fig . 7. **Architecture of RPE-TCAM**

The structure consists of reducing the filtering in which multiplexer consists of the output in which the output is stored in the banks for the reduction of the number of bits compared to the original input word. The data compared based on the reduction stages to improve the efficiency of the design in which the data is processed into multiple packets where the data is allocated to process the data through the multiplexer to reduce the complexity of the design. The extended architecture is designed based on the FPGA based TCAM architecture in which the data stored memory is replace with the TCAM architecture to reduce the power consumption.
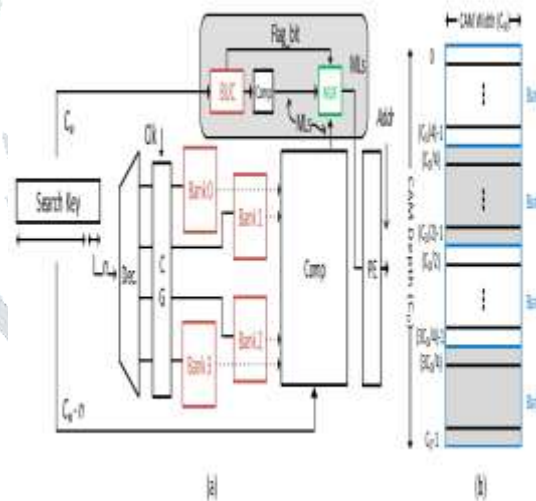

Fig.8.Extended Architecture of RPE-TCAM

The extended RPE TCAM architecture is shown in the Fig.8 .The size of BUC is half of the banks in which the extended CAM is evaluated. The flag bits are indicated in which the data is forwarded to compute the address locations and the clock gating is used to reduce the power consumption. The power saving TCAM architecture is evaluated on the Xilinx board in which the results are compared with the previous architectures. Finally the extended design achieves better results compared with the previous TCAM architectures.

IV. DISCUSSIONS

A sample design of 512 x 36 UE-TCAM was evaluated on Xilinx Virtex-6 FPGA. The count of slice registers were reduced to a greater extent when compared to UE-TCAM in G-AETCAM architecture.Table- I below shows results of different TCAM architectures. Measured the power consumption using Xilinx Xpoweranalyzer.

In DURE-TCAM the count of slice registers And LUT'S were reduced to 672 and 3067is achieved with a speed of 358MHZ and it is having energy efficiency of 67% with higher performance per area is achieved and it's have smaller single search clock latency when compared to SRAM based TCAMs.

The performance of HP-TCAM is being evaluated on Virtex-6 FPGA ,HP-TCAM size of 512x36 and the power dissipated is 190 mw with update latency of 512 cc(clock cycles).

Table- I: Performance Evaluation Results of TCAM Architectures:

Design	Slice registers	LUT's	Speed	EDP
HP-TCAM [13]	2057	5326	118.1	865.07
UE-TCAM [7]	521	1583	201.78	209.73
DURE [11]	672	3067	358	290.25
Xilinx SDNet [8]	2495	6894	171	220
RPE-TCAM [12]	297	1093	319	201

V. CONCLUSION

In this survey, Different TCAM architectures which are evaluated in Xilinx Virtex 6 FPGA, were analyzed. The evaluated results show the comparison between slice registers, LUT's, power and speed. For the specific application, the corresponding TCAM architecture is selected to improve the performance in terms of above all parameters. The power dissipation is measured using Xilinx XpowerAnalyzer tool. The above architectures are specially used in SDNs and in modern networks virtualization which decreases the part of hardware based functionality in the network.

REFERENCES:

[1] R. Karam, R. Puri, S. Ghosh, and S. Bhunia, "Emerging trends in design and applications of memory-based computing and contentaddressable memories," Proc. IEEE, vol. 103, no. 8, pp. 1311–1330, Aug. 2015.

[2] M. S. Riazi, M. Samragh, and F. Koushanfar, "Camsure: Secure contentaddressable memory for approximate search," ACM Trans. Embedded Comput. Syst., vol. 16, no. 5s, p. 136, 2017.

[3] P. Reviriego, A. Ullah, and S. Pontarelli, "PR-TCAM: Efficient TCAM emulation on xilinx FPGAs using partial reconfiguration," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 27, no. 8, pp. 1952–1956, Aug. 2019.

[4] K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory (CAM) circuits and architectures: A tutorial and survey," IEEE J. SolidState Circuits, vol. 41, no. 3, pp. 712–727, Mar. 2006.

[5] N. Mohan, W. Fung, D. Wright, and M. Sachdev, "Design techniques and test methodology for low-power TCAMs," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 14, no. 6, pp. 573–586, Apr. 2006.

[6] M. Irfan and Z. Ullah, "G-AETCAM: Gate-based area-efficient ternary content-addressable memory on FPGA," IEEE Access, vol. 5, pp. 20785–20790, 2017.

[7] Z. Ullah, M. K. Jaiswal, R. C. C. Cheung, and H. K. H. So, "UE-TCAM: An ultra efficient SRAM-based TCAM," in Proc. IEEE Region Conf., Nov. 2015, pp. 1–6.

[8] Ternary Content Addressable Memory (TCAM) Search IP for SDNet, Xilinx Product Guide, San Jose, CA, USA, Nov. 2017.

[9] A. Ahmed, K. Park, and S. Baeg, "Resource-efficient SRAM-based ternary content addressable memory," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 25, no. 4, pp. 1583–1587, Apr. 2017.

[10] K. Locke, "Parameterizable content-addressable memory," Xilinx, San Jose, CA, USA, Appl. Note XAPP1151, 2011.

[11] I. Ullah, Z. Ullah, U. Afzaal, and J.-A. Lee, "DURE: An Energy- and resource-efficient TCAM architecture for FPGAs with dynamic updates," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 27, no. 6, pp. 1298–1307, Jun. 2019.

[12 M. Irfan, Z. Ullah, M. H. Chowdhury and R. C. C. Cheung, "RPE-TCAM: Reconfigurable Power-Efficient Ternary Content-Addressable Memory on FPGAs," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 28, no. 8, pp. 1925-1929, Aug. 2020, doi: 10.1109/TVLSI.2020.2993168.

[13] Z. Ullah, K. Ilgon, and S. Baeg, "Hybrid partitioned SRAM-based ternary content addressable memory," IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 59, no. 12, pp. 2969–2979, Dec. 2012.

[14] Z. Qian and M. Margala, "Low power RAM-based hierarchical CAM on FPGA," in Proc. Int. Conf. ReConFigurableComput., Dec. 2014, pp. 1–4.

[15] M. Irfan, Z. Ullah, and R. C. C. Cheung, "D-TCAM: A highperformance distributed RAM based TCAM architecture on FPGAs," IEEE Access, vol. 7, pp. 96060–96069, 2019.

[16] W. Jiang and V. K. Prasanna, "Scalable packet classification on FPGA," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 20, no. 9, pp. 1668–1680, Sep. 2012.

[17] F. Syed, Z. Ullah, and M. K. Jaiswal, "Fast content updating algorithm for an SRAM-based TCAM on FPGA," IEEE Embedded Syst. Lett., vol. 10, no. 3, pp. 73–76, Sep. 2018.

[18] I. Ullah, Z. Ullah, and J.-A. Lee, "EE-TCAM: An energy-efficient SRAM-based TCAM on FPGA," Electronics, vol. 7, no. 9, p. 186, 2018