Low Power Search Engine using Non-volatile Memory based TCAM with Priority Encoding and Selective Activation of Search Line and Match Line

Cheol Kim*†, Rak-Joo Sung*, Sung-Gi Ahn*, Jisu Min* and Kee-Won Kwon*
Email: {kim.c,nockjoo,todrl1783,jschic1305,keewkwon}@skku.edu
*Department of Information and Communication Engineering, Sungkyunkwan University, Suwon, Korea
†Memory Division, Samsung Electronics Co., LTD., Hwaseong, Korea

***Abstract*—Ternary Content Addressable Memory (TCAM) is emerging as a hardware solution for high performance router by supporting an internal routing table for fast search despite the burden on dynamic power and speed loss due to Longest Matching Prefix (LMP) selection. In this paper, we propose a Higher Priority First Out (HiPFO) memory architecture that outputs only the LMP in a single cycle using a nonvolatile TCAM (nvTCAM), and power reduction schemes that selectively activates search lines and cell array. Thanks to these techniques, encoding power and dynamic power is improved by 86% and 68% respectively and latency is 47% better than the conventional nvTCAM. A 64bitx32 nvTCAM fabricated using 180nm-CMOS technology demonstrates HiPFO operation.**

Keywords—TCAM; nonvolatile TCAM (nvTCAM); High Priority First Output (HiPFO)

I. INTRODUCTION

Along with the development of ultra-high-speed internet service, the core of a high performance router is to retrieve quickly the Internet Protocol (IP) address from the look up table and transfer the packet to the destination. Since Ternary Content Addressable Memory (TCAM) executes high-speed search using an internal routing table, it is emerging as a hardware solution for high performance router. However, as shown in Fig. 1, the router based on TCAM needs to select the Longest Matching Prefix (LMP) in the lookup table which has the restrictions on ordering, so the power loss and processing time delay due to frequent update is increasing.

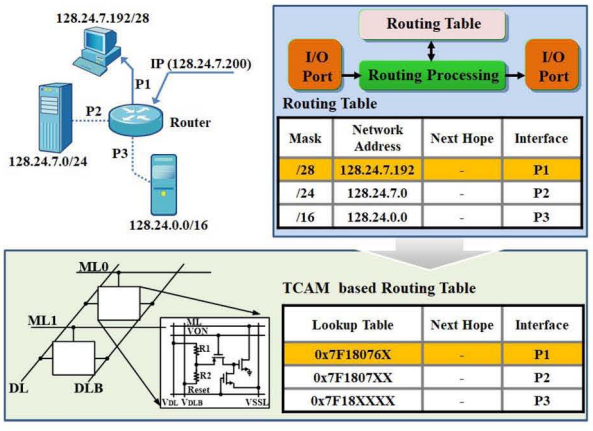


Fig. 1. Router configuration using TCAM lookup table which maintains the order of pattern save.

In connection with this issue, lookup table and the priority encoder (PE) is integrated in the TCAM cell [1], thereby it can save the energy required for update across the different update scenarios around 36% to 67%, resulting in better performance. However, it still becomes burdensome in terms of cell complexity, latency degradation and dynamic power loss of the search operation.

To overcome these drawbacks, we propose a Higher Priority First Out (HiPFO) memory architecture that outputs only LMP in a single cycle using the previously proposed cell structure [2] without area penalty, and achieves 86% better in encoding power consumption compared to prior art [1]. In addition to this, several techniques are utilized to reduce dynamic power: Selective Search Line Control (SSC) and Selective Match Line Control (SMC). The former activates the search line (SL) driver selectively to reduce unnecessary dc current by up to 21% with R-ratio. In the latter case, only the match line (ML) expected to match is activated reducing array power by up to 31% during search operation.

II. RELATED WORK

A. Background of TCAM

Unlike the Random Access Memory (RAM) where the operating system provides memory addresses to sequentially access data, Content Addressable Memory (CAM) executes parallel searches in a single cycle using an internal lookup table which allows fast searching speed than RAM. Furthermore, TCAM enhances the search engine throughput by extending the search range by storing '0', '1', and don't care state ('X') which means wild card operation.

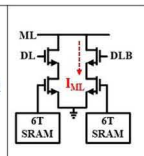
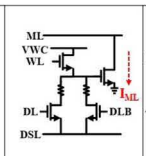
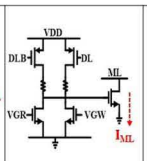
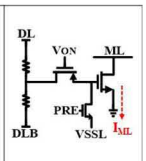
As shown in Fig. 2(a), TCAM is classifed according to memory type. Typically there are SRAM based TCAM (sTCAM) and nonvolatile TCAM (nvTCAM). In particular, nvTCAM has been reported as a suitable solution for a high end network router by reducing cell cost and standby power because nvTCAM uses a nonvolatile memory (NVM) element instead of SRAM. Also nonvolatility of nvTCAM enbles high-speed and low power wakeup operation [5].

978-1-5386-4881-0/18/$31.00 ©2018 IEEE
Authorized licensed use limited to: National Taiwan University. Downloaded on March 16,2025 at 06:45:22 UTC from IEEE Xplore. Restrictions apply.

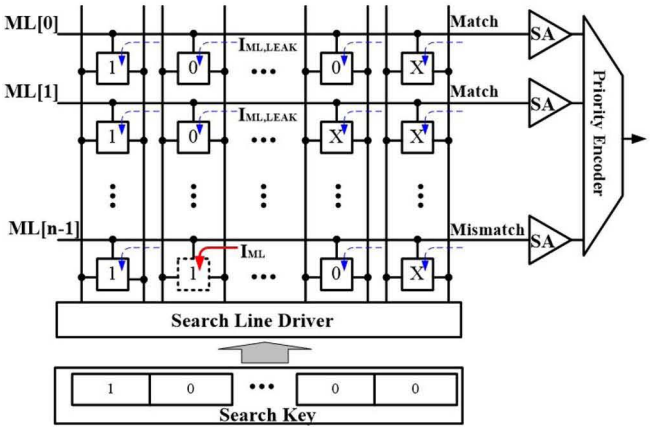
B. *Challenges TCAM faced*

Conventional nvTCAM has limitations due to the characteristics of the device and cell structure: small on/off ratio, large ML capacitance and the restrictions of priority encoding to select LMP among multiple matched results caused by wild card operation. Recently, various nvTCAM cells have been developed to overcome these drawbacks [1-7]. However, problems such as area overhead, power and latency loss due to priority encoding still remain.

Wild card operation makes TCAM be a very important component in a Classless Inter-Domain Routing (CIDR) environment to increase the flexibility of the IP address domain. As shown in Fig. 2(b), the current flowing through the ML driver transistor (I_{ML}) is only leakage current ($I_{ML,LEAK}$) in match or don't care cell (x-cell), but the mismatch cell generates enough current to discharge ML to VSS. Even if the number of x-cells connected to the ML is different, if there is no mismatch cell, all MLs return the matched result. That is, TCAM makes multiple matched results due to wild card and the router has burdensome of selecting and returning the highest priority matched result in the look up table. To return the highest priority result among a plurality of matched results, it is necessary to add an additional functional block such as PE or to frequently update the lookup table to maintain the aligned ordering. The resulting power consumption, area increase and processing delay are challenges TCAM faced as a lookup table for high-speed router.

TCAM cell				
Type	16T SRAM	4T-2R NVM [5]	5T-2R NVM [6]	3T-2R NVM [2]
Cell Size (Normalize)	12.9	9.7	-	8.61

(a)



(b)

Fig. 2. Conventional (a) TCAM cell structure (b) TCAM architecture.

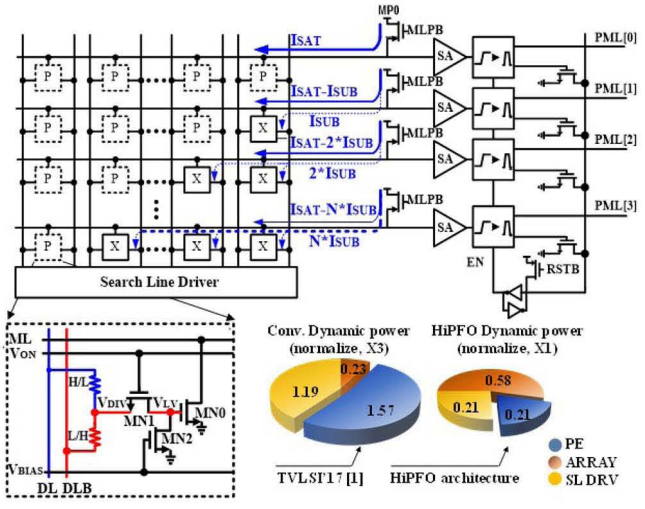
III. PROPOSED DESIGN

A. *Proposed HiPFO architecture based on 3T-2R nvTCAM*

Fig. 3(a) illustrates HiPFO architecture which executes the PE function in parallel during the cell search operation. In the TCAM cell [2] as a lookup table, V_{DIV} is divided according to the resistance ratio and the switching operation of the data line between the high ($V_{DL,H}$) and low ($V_{DL,L}$) level. The NMOS pass transistor (MN1) is disposed between the resistance divider and ML driver, MN0. It transfers V_{DIV} to V_{LV} during pre-charge period to determine overdrive voltage (V_{OV}) of MN0 depending on the matched result. During the evaluation period, voltage level of ML (V_{ML}) is intimately linked with ML pull up transistor (MP0) and determines current driving capability of MN0.

In the mismatched cell, V_{LV} is high enough to exceed the threshold voltage (V_{TH}) of MN0 and interfere with charge accumulation of ML. On the other hand, both matched cell and x-cell are vice versa. Meanwhile, x-cell takes on a very special function. V_{LV} in the x-cell is not high enough to overcome the V_{TH}, but ML develop delay (T_{ML}) can be adjusted according to the number of x-cell with subthreshold leakage. That is, N x-cells slow down the charge accumulation speed of ML, and as N increases, T_{ML} is also increased as follows (1). The HiPFO controller receives the result of MLs which is generated with a regular time interval. The first received ML restrains the subsequent received ML from being processed in the controller, so that it can output only the search result with the highest priority. The simulation result indicates that the encoding power reduced by 86% compared with [1] as shown in Fig. 3(b).

$$\Delta T_{ML} = \frac{C_{ML}}{I_{SAT}} \int_0^{V_{ML}} \left(\frac{N \times I_{SUB}}{I_{SAT} - N \times I_{SUB}} \right) dV \qquad (1)$$



(a) (b)

Fig. 3. (a) Proposed HiPFO architecture (b) Comparision of dynamic power.

978-1-5386-4881-0/18/$31.00 ©2018 IEEE
Authorized licensed use limited to: National Taiwan University. Downloaded on March 16,2025 at 06:45:22 UTC from IEEE Xplore. Restrictions apply.

HiPFO technique is more effective than the prior art [1] in terms of power consumption, but there is still room for saving dynamic power of the used TCAM cell. In particular, dc current flowing between complementary data lines (DL/DLB) constraints application of nvTCAM in terms of power consumption.

B. *Proposed Selective Search Line Control*

Fig. 4 introduces several power reduction techniques adopted in our study. SSC extends prior art based on SRAM [8] to nvTCAM. Proposed SSC consisting of segmented cell array, Local SL (LSL) driver and vertical compare logic operates as follows. During search operation, V_{OV} margin of MN4 (MN5) is set by V_{DIV} and adjustable bias voltages (V_{OFF}, V_{PASS}) to determine whether to discharge VEN or not. The LSL driver is controlled based on the level of VEN. If all cells sharing the LSL are x-cells, LSL driver is turned off selectively. Since the dual SL structure controls the LSL driver independently for each segment, it is possible to remove restrictions on ordering.

In additional, pulse controlled SL minimizes dc current flowing through two resistances in series. A replica cell is modeled with x-cell which guarantees SL driver pulse to ensure sufficient ML development time. In accordance with the prefix length distribution of IP address [8], these sophisticated

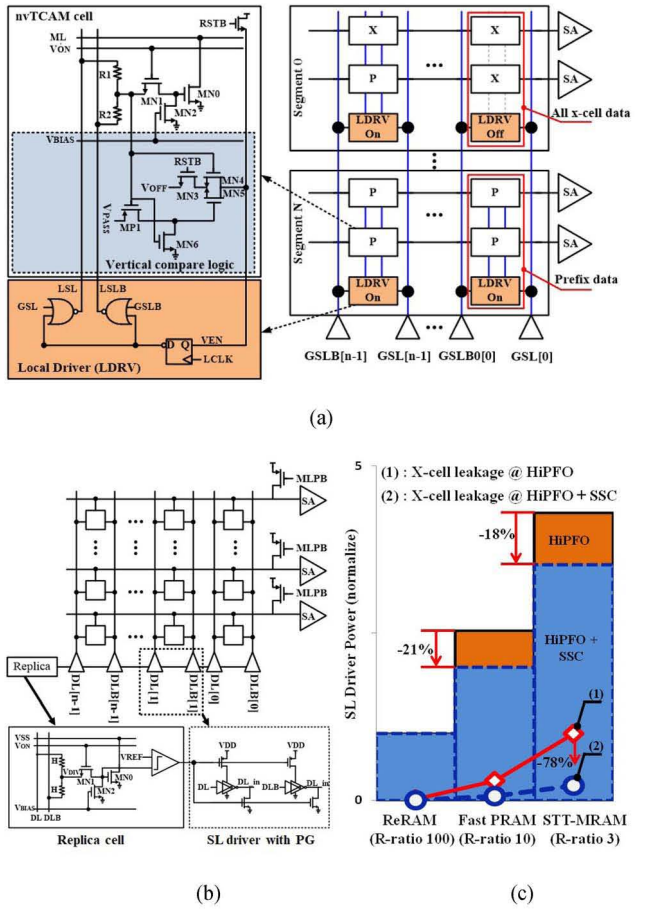


Fig. 4. Dynamic power reduction technique: (a) Selective SL control scheme (b) Pulse controlled SL scheme (c) Comparison of SL driver power consumption.

controls reduce the dynamic power by up to 21% at small R-ratio (R-ratio is 10 or less), and help to extend the application of nvTCAM [2] to emerging memory with small R-ratio such as Phase Change Random Access Memory (PCRAM) or Spin-Torque-Transfer Magnetic Random Access Memory (STT-MRAM).

C. *Proposed Selective Match Line Control*

Using SMC further reduces dynamic power. In SMC algorithm, the search field is defined according to the previous search result and current search key. It compares previous search key with current search key and judges whether they match or not. If those addresses match, only the matched MLs of the previous search result activated and vice versa. Therefore, it decreases dynamic power of ML by reducing the number of cells activated. Different from the existing nvTCAM structure, SMC needs 3 additional registers as shown in Fig. 5(a).

Fig. 5(b) illustrates the functions of additional registers in SMC. Firstly there is a mask bit register (M) consisting of one word-size (w) in the lookup table. Depending on the type of data in cell (prefix or x-cell), '1' or '0' is stored in the register. If cells connected in same SL are prefix cell, '1' is stored in the register whereas '0' is stored if they are x-cell. The vertical compare logic introduced in SSC is used to judge whether x-cell or not. Secondly, there is a search key register (S) storing previous search key. It compares search key register to current search key. Finally there is a 1bit matching register which stores the matched result of ML in each sense amplifier. When (K+1) th search key is applied, an XOR operation is performed with prefix bit in (K) th search key stored in the register S as shown in (2). After that it can be checked whether or not the

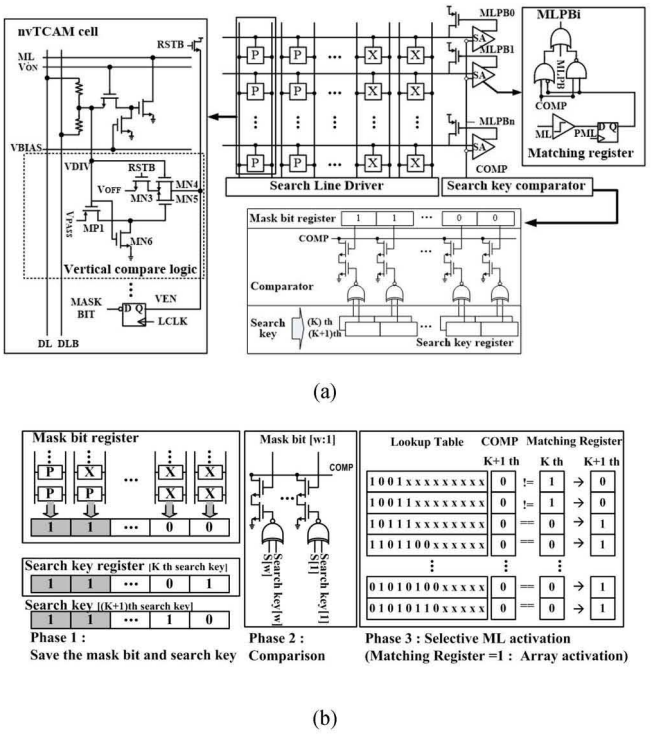


Fig. 5. Dynamic power reduction technique: (a) Selective ML control scheme (b) Function of additional registers in selective ML control.

978-1-5386-4881-0/18/$31.00 ©2018 IEEE
Authorized licensed use limited to: National Taiwan University. Downloaded on March 16,2025 at 06:45:22 UTC from IEEE Xplore. Restrictions apply.

result of XOR operation (COMP) matches. If search keys are matched, only the matched cells are activated in the (K) th search operation, and vice versa. Assuming K times of matches in 100 times search operations, the total power (Pt) can be expressed by frequency (K) of the match operation, the power in the matched region (Pm) and the power in the mismatched region (Pmm) as follows (3).

$$COMP = \sum_{n=1}^{w} \mathrm{M}[n] \times \{ S[n] \text{ XOR } (search\ key[n]) \} \quad (2)$$

$$Total\ Power\ (Pt) = \frac{K}{100} \times Pm + \frac{(100\text{-}K)}{100} \times Pmm \quad (3)$$

It was confirmed by simulation that the array power decreases by 31% with the continuous match operation in cell array which has 37.5% match region.

IV. Performance and Measurement

TABLE I summarizes simulation results and compares to recent TCAM based search engines. Thanks to HiPFO architecture and SSC scheme, the power dissipation is improved by 68% and the latency is 47% better than the prior art [1]. To verify the operation of the HiPFO architecture, a 64bit x32 3T-2R nvTCAM is fabricated using 180nm-CMOS technology. Fig. 6 represents the measurement waveform from the fabricated chip. In the sensing operation, the measured delay increases as the number of x-cells increases. When the number of x-cells is 32bit, ML was developed at a delay time interval of 600ps, which agrees with post-layout simulation.

TABLE I. Comparison of the recent TCAM

		TVLSI'12 [7]	*TVLSI'17 [1]*	*EL'17 [2]*	*This Work*
Cell Size	Core	X26.4	X9.7	X8.61	X8.05
	Ctrl	N/A	X8.08	N/A	X13.66
Search Power	Core	-	X1.21	X0.79	X0.64
	PE		X1.33	X0.21	X0.18
	Total		X2.54	X1.0	X0.82
tCYCLE		-	X1.96	X1.0	X1.07
Technology		180nm CMOS			

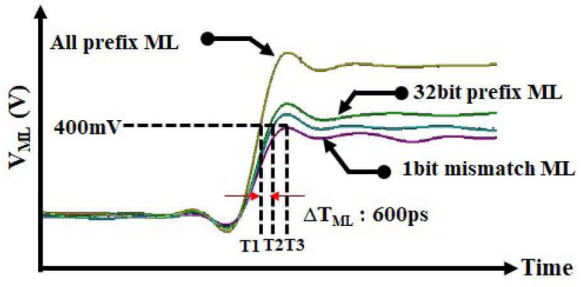


Fig. 6. Measurement waveform: 3T-2R nvTCAM using HiPFO.

V. Conclusion

We suggest a Resistive Random Access Memory (ReRAM) based HiPFO memory architecture which outputs only the LMP in a single cycle based on 3T-2R nvTCAM and power reduction techniques that selectively activates SL and cell array. HiPFO architecture returns only the highest priority matched address value without additional hardware or software for priority encoding, resulting 86% better in encoding power consumption compared to prior art. SSC scheme optimizes the dynamic power by limiting the behavior of the meaningless cells during the search operation, thereby improving by 68%. Also SMC is a search technique that predicts a match region in advance and activates only cells existing in the region. These proposed techniques can free up the burden of priority encoding and can improve the performance of a high speed router by minimizing unnecessary dynamic power consumption. Furthermore these selective activation techniques enable the leakage control of the conventional 3T-2R nvTCAM [2], so that extends the applicability of nvTCAM with various emerging memory such as STT-MRAM.

Acknowledgment

This research was supported by Nano·Material Technology Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (No. 2016M3A7B4910575).

References

[1] H.-J. Tsai, K.-H. Yang, Y.-C. Peng, C.-C. Lin, Y.-H. Tsao, M.-F. Chang, et al., "Energy-Efficient TCAM Search Engine Design Using Priority-Decision in Memory Technology" IEEE Transactions on VLSI, vol. 25, no. 3, pp. 962-973, March 2017.

[2] C. Kim and K.-W. Kwon, "3T-2R Non-Volatile TCAM with Voltage Limiter and Self-Controlled Bias Circuit" IET Electronics Letters, vol. 53, no. 13, pp. 837-839, June 2017.

[3] M.-F. Chang, C.-C. Lin, A. Lee, Y.-N. Chiang, C.-C. Kuo, G.-H. Yang, et al., "A 3T1R Nonvolatile TCAM Using MLC ReRAM for Frequent-Off Instant-On Filters in IoT and Big-Data Processing", IEEE J. Solid-State Circuits, vol. 52, no. 6, pp. 1664-1679, June 2017.

[4] C.-C. Lin, J.-Y. Hung, W.-Z. Lin, C.-P. Lo, Y.-N. Chiang, H.-J. Tsai, et al., "A 256b-Wordlength ReRAM-based TCAM with 1ns Search-Time and 14× Improvement in WordLength-EnergyEfficiency-Density Product using 2.5T1R cell", IEEE Int. Solid-State Circuits Conference (ISSCC), pp. 136-137, Jan. 2016.

[5] M.-F. Chang, L.-Y. Huang, W.-Z. Lin, Y.-N. Chiang, C.-C. Kuo, C.-H. Chuang, et al., "A ReRAM-based 4T2R nonvolatile TCAM using RC-filtered stress-decoupled scheme for frequent-OFF instant-ON search engines used in IoT and big-data processing", IEEE J. Solid-State Circuits, vol. 51, no. 11, pp. 2786-2798, Nov. 2016.

[6] M.-F. Chang, C.-H. Chuang, Y.-N. Chiang, S.-S. Sheu, C.-C. Kuo, H.-Y. Cheng, et al., "Designs of Emerging Memory Based non-volatile TCAM for Internet-of-Things (IoT) and big-data processing: A 5T2R universal cell", IEEE Int. Symp. on Circuit and Syst. (ISCAS), pp. 1142-1145, May 2016.

[7] S.-H. Yang, Y.-J. Huang and J.-F. Li, "A Low-Power Ternary Content Addressable Memory With Pai-Sigma Matchlines", IEEE Transactions on VLSI, vol. 20, no. 10, pp. 1909-1913, Oct. 2012.

[8] C.-C. Wang, J.-S. Wang, C. Yeh, "High-Speed and Low-Power Design Techniques for TCAM Macros", IEEE J. Solid-State Circuits, vol. 43, no. 2, pp. 530-540, Feb. 2008.

978-1-5386-4881-0/18/$31.00 ©2018 IEEE
Authorized licensed use limited to: National Taiwan University. Downloaded on March 16,2025 at 06:45:22 UTC from IEEE Xplore. Restrictions apply.