Design of a power-efficient CAM using automated background checking scheme for small match line swing

Do, Anh Tuan; Yin, Chun; Yeo, Kiat Seng; Kim, Tony Tae-Hyoung

2013

Do, A. T., Yin, C., Yeo, K. S., & Kim, T. T. H. (2013). Design of a power-efficient CAM using automated background checking scheme for small match line swing. 2013 Proceedings of the ESSCIRC (ESSCIRC).

<https://hdl.handle.net/10356/100930>

<https://doi.org/10.1109/ESSCIRC.2013.6649109>

© 2013 IEEE. This is the author created version of a work that has been peer reviewed and accepted for publication by Proceedings of the ESSCIRC 2013, IEEE. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: <http://dx.doi.org/10.1109/ESSCIRC.2013.6649109>

Downloaded on 26 Apr 2023 22:56:50 SGT

Design of a Power-Efficient CAM Using Automated Background Checking Scheme for Small Match Line Swing

Anh Tuan Do, Chun Yin, Kiat Seng Yeo and Tony Tae-Hyoung Kim
VIRTUS, IC Design Centre of Excellence, Nanyang Technological University, Singapore
atdo@ntu.edu.sg

Abstract— This work reports a fully parallel match-line (ML) structure with an automated background checking (ABC) scheme. MLs are pre-charged by a pulsed current source to minimize power. The proposed ABC scheme monitors the ML sensing using two dummy rows. It digitally adjusts the pulse width and the delay of the search control signals of the CAM without disturbing the normal operation. Therefore, it can continuously track the optimum operating point, making the CAM tolerant to fabrication variations. Additionally, multi-V_t transistors are used in the CAM cell to reduce the leakage by 15× while improving the ML discharge speed by 2× when compared with the standard-V_t devices at 1.2V, 80 °C. The test chip was prototyped using a standard 65 nm CMOS process. The average energy consumption is 0.77 fJ/bit/search at 500 MHz /1.2 V.

Keywords— CAM, small match line swing, match-line, variation tolerant design.

I. INTRODUCTION

Content Addressable Memory (CAM) is a special type of solid-state memory in which data are accessed by contents rather than physical locations. It receives input search data, i.e. a search word, and returns the location of the similar word in its data banks [2]. Preferably, this search operation completes in one clock cycle. Therefore, CAM has been utilized in a wide range of applications such as network routers [4], associate cache and image processing [5] that require high-speed search operations. However, because the entire memory is searched at the same time, CAM consumes a huge amount of dynamic power. Hence reducing the search power is the topmost concern of the CAM design.

Match line (ML) sensing consumes a dominant amount of total power in high-speed CAM operation. Various circuit techniques have been proposed [6-8] to lower the sensing power. In [6], an additional capacitor is employed in each ML to limit the amount of charges per search transaction for energy reduction (Fig. 1). This approach requires more area overhead and suffers from serious row-to-row capacitor variations which have a negative impact on the sensibility of the latch-based comparator. *Tyshchenko et al.* injects a small current to each ML prior to the sensing phase which acts as an excitation charge to observe the stability of the R-C ML model [7]. However, a level shifter required at each ML comparator makes its sensing speed inferior to that of the latch-based comparator. Furthermore, in addition to the initial charges, each ML consumes extra power depending on the sensing outcome. Since CMOS scaling increases systematic and random device variations, the aforementioned techniques will be more vulnerable in deep sub-micron technology.

Figure 1. Conventional small ML swing scheme.

In this work, we propose a power efficient ML sensing technique that achieves both low power consumption and process variation tolerance. On top of that, we also address the leakage current issue by employing multi-V_t CMOS devices in the CAM cell. Section II explains the details of the proposed CAM cell, followed by the proposed ML sensing design in section III. The proposed automated background checking (ABC) scheme is introduced in Section IV. Section V evaluates the performance of the test chip. Finally, Section VI concludes the paper.

II. MULTI-V_t CAM CELL FOR LEAKAGE REDUCTION AND SEARCH SPEED ENHANCEMENT

With the rapid increase of memory depth due to the vast amount of search data, CAM is usually partitioned into banks to improve speed and reduce power. As a result, most of the CAM cells are kept in the idle state, continuously consume leakage current. As technology scales down to sub-65nm CMOS technology nodes, this leakage current increases exponentially and thus has a profound impact on the overall power consumption of the system.

Fig. 2 shows the 10T NOR-type CAM cell topology in its standby state where BLs are pre-charged to V_{DD} and SLs are pre-charged to ground. Since ML is pre-charged low (i.e. V_{ML} = 0), the comparison circuit (i.e. N1-N4) consumes no leakage power. As a result, the CAM's total leakage is solely determined by the 6T SRAM element. To minimize the amount of leakage, we deployed high-V_t devices in the 6T SRAM part. The arrows in Fig. 2 depict the sub-threshold leakage currents when the cell data (Q) is '0'. Fig. 3 compares the total CAM

Figure 2. Multi-V_t CAM cell where the SRAM element uses high-V_t to reduce leakage and the comparison circuit uses low-V_t to achieve high-speed sensing.

Figure 3. SRAM leakage currents using different V_t devices.

Figure 4. I_{on}/I_{off} ratio of the single-miss ML at 80 °C when different V_t devices are used.

Figure 5. Waveforms of the matched and single-miss MLs during a search operation.

leakage using different V_t options at different operating voltage and temperature. Using high- V_t devices reduces the leakage by more than 15 $\times$ and 100 $\times$ when compared to standard- V_t devices and low- V_t devices at 1.2V, 80 °C, respectively. Note that the CAM is normally operated at nominal supply voltage, and read and write operations are not as critical as the search operation. Consequently, using high- V_t devices in the SRAM element has an insignificant impact on the overall performance of the CAM.

The device type for the comparison circuit (i.e. N1-N4) should be selected for better search performance. Fig. 4 shows the I_{on}/I_{off} ratios of three NMOS types. Since the sense amplifier should distinguish the ML level with single-miss from that with match, a larger I_{on}/I_{off} ratio is desirable. Fig. 4 indicates that high- V_t devices are more favorable than low- V_t devices due to the much higher I_{on}/I_{off} ratio. However, the transient waveforms of the MLs in Fig. 5 explain that using low- V_t devices is a better option in performance. It can be seen that while the matched ML using low- V_t indeed ‘leaks’ more charges when compared with standard- and high- V_t cases, its impact on the ML sensing is insignificant. This is because the ML sensing is very fast and thus the amount of leaked charges ($Q_{leak} = I_{leak} \cdot T_{sense}$) affecting the ML sensing is negligible. On the other hand, using low- V_t leads to a much faster discharge of the single-miss ML. Simulation shows that low- V_t devices improve the ML discharging speed by 2 $\times$ when compared to the high- V_t counterpart (Fig. 5). Consequently, we adopt high- V_t devices in the SRAM element while low- V_t devices are used in the comparison circuit as shown in Fig. 2. This combination allows us to concurrently minimize the cell leakage and achieve high speed ML sensing. Note that employing low- V_t devices in the comparison circuit have no impact on the leakage power of the CAM design due to grounded ML in the pre-charging phase.

III. PROPOSED SMALL SWING ML SENSING SCHEME

Fig. 6 illustrates the proposed small swing sensing scheme with a latch-based comparator. When ML sensing starts, an internally generated pulse is sent to the gate of P1 and P2 (Pulse in Fig. 6), charging ML to a certain level (Fig. 7). In a matched case, ML remains at the pre-charged level even after the pulse signal (Pulse) is turned off. The reference ML (i.e. V_{ref}) is a dummy matched ML with an additional parasitic capacitor (C_p). Hence, Pulse charges V_{ref} to a level which is lower than that of the matched ML. In a missed case, ML is discharged to ground by the comparison circuits in the CAM cells as shown in Fig. 7. After a short delay, the sense amplifier

Figure 6. Proposed low ML swing using short current pulse and latch comparator to achieve short delay and low power consumption.

Figure 7. Waveforms of the proposed ML sensing scheme during search.

(SA) at each ML is enabled by an enable signal (EN). The final result is generated after the delay of the latch comparator. In contrast to the conventional design, PMOS-input latched comparator is used since MLs are pre-charged to a very low level for implementing small ML swing.

Theoretically, the minimum allowable level for the matched ML (ML_{match}) is $2 \times V_{offset}$ where V_{offset} is the offset voltage of the comparator. Since the ML power consumption is proportional to the ML voltage swing, small ML swing is highly required for power reduction. However, due to process variations each comparator has an intrinsic offset voltage. Thus the following conditions must be satisfied when the SA is enabled to generate a correct ML output:

$$V_{in1} = V_{MLmatch} - V_{ref} > V_{offset} \quad (1)$$

$$V_{in2} = V_{ref} - V_{MLmiss} > V_{offset} \quad (2)$$

Figure 8. Monte-Carlo simulation of the pre-charged ML level.

Figure 9. Input voltages to the SA vs. difference pulse width (left) and EN delay (right).

Where V_{in1} and V_{in2} are the sense amplifier input voltage when sensing a matched and 1-miss ML, respectively. In addition, the pre-charged level of each ML also varies due to process variation as shown in Fig. 8. To reduce the amount of variation, the most straightforward methods are: (I) increasing the size of the input transistor pair to reduce V_{offset} ; (II) imposing a large margin on the pulse width for P1 and P2 in Fig. 6, and (III) providing an enough delay to the sense amplifier enable signal (EN) for enlarging V_{in1} and V_{in2} . Fig. 9 illustrates the relationship between V_{in1} and the pulse-width of Pulse (Fig. 6), and V_{in2} and the EN delay. While it is highly desirable to reduce V_{offset} as small as possible, there is a certain limit in the size of the input devices of the sense amplifier. Therefore, a large voltage margin must be set to cover possible variations, which leads to significant extra power consumption. To tackle this, we propose a circuit technique that can track the amount of margin, which will be discussed in the next section.

IV. AUTOMATED BACKGROUND CHECKING (ABC)

In this section, we explain the proposed automated background checking (ABC) scheme for setting a small margin on the above-mentioned ML sensing and adjusting it automatically. Its operation principle is illustrated in Fig. 10. The reference row is shared by all SA and has an additional parasitic C_p , which is formed by the device parasitic capacitance and the interconnection between the reference row and the SAs. As a result, the pre-charged level of V_{ref} is lower than that of matched MLs. The conditions described in (1) and (2) can be violated in the following two scenarios: (i) when the pre-charged voltage is too small, the condition in (1) is violated; (b) when the sense amplifier is enabled too early and hence V_{MLmiss} is not yet discharged to a low voltage, the condition in (2) is violated. The proposed ABC scheme uses two dummy MLs as the guard bands to protect the main array from these two sensing errors and simultaneously maintain the power consumption at the ML as low as possible.

If the condition in (1) is violated, the dummy ML_{match} (Fig. 10) creates a rising edge at its output to clock the shift-register (Fig. 11). This register shifts the “1” bit to a higher location,

Figure 10. Simplified architecture of the proposed design with adjustable SA delay and adjustable *Pulse* pulse-width for process variation tolerance. A dummy row is also used to generate the reference voltage to the SA.

Figure 11. Implementation of the pulse-width generator. Several switches are turned on/off accordingly by the output bits of the register to digitally control the strength of the delay path and thus the width of the *Pulse* signal.

Figure 12. Simulation waveforms. V_{ref} is purposely hold at 300 mV to create an initial wrong dummy-match output (middle) which clocks the shift-register (top) to increase the pulse width and thus the ML voltage (bottom).

weakening the strength of the delay path, resulting in a wider pulse width. Thus, in the next cycle, the MLs will be pre-charged to a higher level, ensuring a better V_{in1} . Similarly if the condition in (2) is violated, the dummy ML_{miss} in Fig. 10 will flag the adjustable delay generator to block and extend the delay of the EN signal to have a larger V_{in2} . When the chip is enabled, the first few search operations can be used to calibrate these two blocks. In this work, the pulse width and the EN delay is engineered so that the equivalent sensing margin of the main array is 10% higher than that of the dummy rows by sizing of the transistors. Hence, even if the condition of (1) or (2) is violated at the two dummy rows, the main array still

Figure 13. Simulated search delay of the proposed and the conventional designs at different supply voltages at 20°C and 100°C, respectively.

Figure 14. Measured search power per ML of the proposed and the conventional designs at different supply voltages.

works as normal. This means that errors are always detected at the dummy rows first and correction can be done in the background silently. Fig. 12 illustrates how the adjustable pulse width generator works. We force the matched dummy ML to a wrong state by holding the V_{ref} at 300 mV. As the offset voltage of the SA is assumed to be 100 mV, $V_{MLmatch}$ must be charged-up to 400 mV at least. However, initially, it is lower than that. Thus, the register keeps shifting its data (first 3 cycles) until no rising edge is sent back to its clock input. The adjustable EN delay generator works in a similar way.

V. PERFORMANCE EVALUATION AND TEST CHIP IMPLEMENTATION

Fig. 13 and Fig. 14 compare the speed and power performance of the proposed design and the conventional one. At the nominal supply voltage, both designs have a similar search delay of 1.07 ns. However, as supply voltage reduces, the proposed design proves to be less sensitive to supply voltage scaling. Fig. 13 also shows that the proposed sense amplifier is less sensitive to temperature variations. Regarding the power consumption, the proposed design is 5.5x less power hungry when compared to the conventional design at 1.2 V operating voltage. As the supply voltage scales down, the full ML swing in the conventional design reduces significantly, making the benefit of the proposed scheme less attractive. Nevertheless, at 0.8V the power consumption of the new design is only one-third of the conventional design. The microphotograph and layout of the test chip is shown in Fig. 15. It occupies an active area of only 0.125 mm². The chip has been tested successfully at 1.2V/500 MHz with an average energy of 0.77 fJ/bit/search. Table 1 compares of the test chip with prior works.

Figure 15. Microphotograph and layout of the test chip.

TABLE I. PERFORMANCE COMPARISON OF PRIOR WORKS

	This work	[1]	[3]	[5]
Technology /Supply	65 nm /1.2V	180nm /1.8V	130nm /1V	32nm/1V
Delay (ns)	1.07	2.7	0.9	0.145
Energy (fJ/bit/search)	0.77	2.82	1.827	1.07
Frequency	500 MHz	210 MHz	250 MHz	N.A
Chip area (mm ²)	0.125	0.98	1.4	N.A

VI. CONCLUSION

A 128x128 binary CAM test chip is implemented using a standard 65 nm/ 1.2 V CMOS process. The SRAM element uses high- V_t devices to reduce leakage by 100× while the comparison circuits inside the CAM cell uses low-threshold devices to enhance the search speed. When compared to the conventional design, it reduces the power consumption by 5.5× while maintain a similar sensing delay. It achieves 1.07 ns delay, 0.77 fJ/search/bit at 500 MHz, 1.2 V supply voltage. Note that we focus on the main ML sensing and if pipeline or selective pre-charge architectures is used, its FOM can be brought down even further.

REFERENCES

- [1] Y. Byung-Do, L. Yong-Kyu, S. Si-Woo, M. Jae-Joong, O. Jae-Mun, and K. Hyeon-Ju, "A Low Power Content Addressable Memory Using Low Swing Search Lines," *IEEE Trans. on Circuits and Systems I: Regular Papers*, vol. 58, pp. 2849-2858, 2011.
- [2] K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory (CAM) circuits and architectures: a tutorial and survey," *IEEE J. of Solid-State Circuits*, vol. 41, pp. 712-727, 2006.
- [3] W. Chua-Chin, H. Chia-Hao, H. Chi-Chun, and W. Jun-Han, "A Self-Disabled Sensing Technique for Content-Addressable Memories," *IEEE Trans. on Circuits and Systems II: Express Briefs*, vol. 57, pp. 31-35, 2010.
- [4] R. Shanq-Jang, W. Chi-Yu, and H. Jui-Yuan, "Low Power Design of Precomputation-Based Content-Addressable Memory," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 16, pp. 331-335, 2008.
- [5] A. Agarwal, S. Hsu, S. Mathew, M. Anders, H. Kaul, F. Sheikh, and R. Krishnamurthy, "A 128x128b high-speed wide-and match-line content addressable memory in 32nm CMOS," in *Proceedings of the 2001 ESSCIRC* 2011, pp. 83-86.
- [6] G. Kasai, Y. Takarabe, K. Furumi, and M. Yoneda, "200MHz/200MSPS 3.2W at 1.5V Vdd, 9.4Mbits ternary CAM with new charge injection match detect circuits and bank selection scheme," in *IEEE Custom Integrated Circuits Conference*, 2003, pp. 387 - 390.
- [7] O. Tyschenko and A. Sheikholeslami, "Match Sensing Using Match-Line Stability in Content-Addressable Memories (CAM)," *IEEE J. of Solid-State Circuits*, vol. 43, pp. 1972-1981, 2008.
- [8] I. Arsovski, T. Hebig, D. Dobson, and R. Wistort, "1Gsearch/sec Ternary Content Addressable Memory compiler with silicon-aware Early-Predict Late-Correct single-ended sensing," in *VLSI Circuits (VLSIC), 2012 Symposium on*, 2012, pp. 116-117.