

LOW-POWER PRIORITY ENCODER AND MULTIPLE MATCH DETECTION CIRCUIT FOR TERNARY CONTENT ADDRESSABLE MEMORY

Nitin Mohan, Wilson Fung and Manoj Sachdev

Department of Electrical and Computer Engineering, University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
(nitinm@ieee.org; wifung@ati.com; msachdev@uwaterloo.ca)

ABSTRACT

Multiple match detection (MMD) circuits and priority encoders (PEs) are employed in ternary content addressable memory (TCAM) chips to detect multiple matches and to resolve the highest priority match. This paper presents novel PE and MMD circuits. Measurement results of the proposed circuits, fabricated in 0.18 μ m CMOS technology, show significant (up to 70%) speed and energy improvements over the existing designs.

I. INTRODUCTION

Ternary content addressable memories (TCAMs) employ priority encoders (PEs) to resolve the highest priority match when multiple words match with the search key. Many recent applications also require partial matching and sequential output of all the matched addresses in prioritized order. Thus, most modern TCAMs also employ multiple match detection (MMD) circuits. Since PE and MMD both are in the TCAM critical path, their delays directly affect the speed and latency of TCAMs. A large fan-in makes these circuits slow and power hungry. In this paper, we present PE and MMD circuits that are attractive for TCAMs due to their regular structure, high speed and low power consumption.

II. PRIORITY ENCODER

A PE resolves the highest priority match and encodes this match location into binary format, which is used by an off-chip SRAM to retrieve the corresponding data. Typically, a PE is designed in two stages: (i) multiple match resolver (MMR), and (ii) match address encoder (MAE).

A. Multiple match resolver

An MMR is an n -bit input, n -bit output datapath circuit. Assuming the active-high logic convention

and highest priority for the lowest address, an MMR can be described by the Boolean expressions in (1).

$$\begin{aligned} Out_0 &= In_0 \\ Out_1 &= In_1 \cdot \overline{In_0} \\ &\vdots \\ Out_n &= In_n \cdot \overline{In_{n-1}} \cdots \overline{In_1} \cdot \overline{In_0} \end{aligned} \quad \dots (1)$$

A folding scheme reduces the worst-case delay (Out_n in (1)) by providing lookahead signals from the highest priority cell to the lowest priority cell, and the second highest to the second lowest, and so on [1]. However, this scheme is not suitable for integration with a TCAM array due to too many interconnect routings and excessive clock loading (and jitter) if the circuit is laid out in a single column. Thus, a regular cell-based design using a pass-transistor chain is more suitable for a TCAM.

Fig. 1 shows the block diagram, floorplan and circuit schematic of the proposed 64-bit MMR designed in two levels. The first-level (L1) is divided into eight macro-blocks (Fig. 1(a)). A wired-OR circuit is built into each macro-block for sensing if at least one match exists at its inputs. The output of this wired-OR gate is a lookahead (LA) signal for interfacing with the second-level (L2) MMR, which identifies the block that contains the highest priority match. The resolved L2 signals act as block enable (BE) signals for the L1 MMRs. In order to layout both levels of MMRs in one column, the MMR cells in the L2 are distributed between the L1 blocks.

The proposed MMR circuit is based on the "match token" concept [2][3]. Initially, the clock signal precharges the internal nodes of the pass transistor chain. The precharging at node C relies on the '1 \rightarrow 0' transition at the input. If there are two matches at words 1 and N, In_1 and In_N are raised to V_{DD} . The LA signal is applied to the input of a delay element for generating the strobe (SS) signal. Since

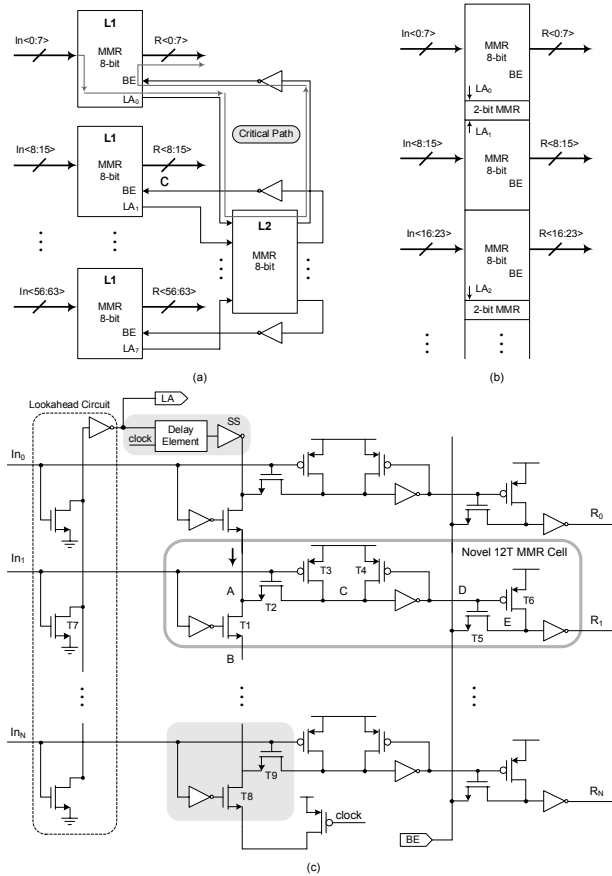


Fig. 1: (a) Block diagram, (b) floorplan, and (c) circuit schematic of the proposed 64-bit MMR.

the L1 pass-transistor chain is not the critical path of a multi-level MMR (Fig. 1), this delay reduces the LA node capacitance without slowing down the critical path. Switching SS ('1 → 0') allows the discharging of the pass-transistor chain down to the highest priority match, which is analogous to a '0' ("match token") traversing down the pass-transistor chain. The internal nodes C and D of the highest priority cell are inverted, and T5 is turned 'ON'. Upon the arrival of the BE signal, node E is discharged to '0', which switches the output R_1 to '1' to indicate that word 1 is the highest priority match. Fig. 2 shows the energy-delay simulation results (with and without the clock energy) comparing the proposed MMR with the existing designs [1][2]. For the same delay, the proposed MMR consumes 50% less energy than the existing "match token" MMR [2]. The energy reduction is even more dramatic when the clock energy is taken into consideration.

The proposed MMR achieves high-speed and low-power operation due to several circuit improvements over the previous design [2]. First of

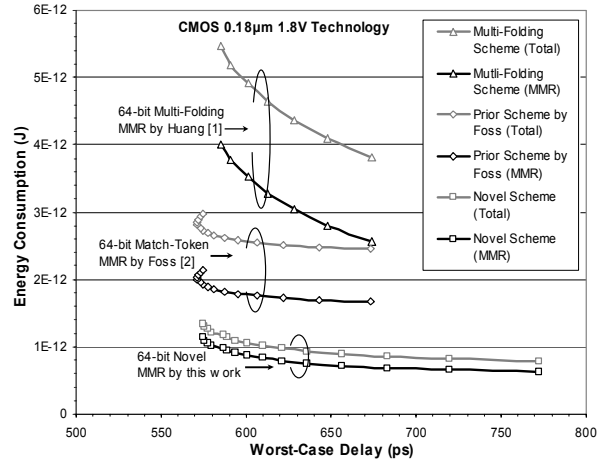


Fig. 2: Energy-delay curves (with and without the clock power) comparing the proposed MMR with the existing MMRs.

all, BE is shielded from the output inverter capacitance by transistor T5. Thus, the size of a macro-block is not limited by the BE capacitance. Secondly, our scheme precharges node C using the corresponding MMR input. This eliminates the clock power consumed in precharging these nodes. Thirdly, since the clock signal is gated by LA (Fig. 1(c)), the proposed MMR generates the "match token" only if there is at least one match in a macro-block. Finally, if the lowest priority cell receives the "match token", this cell must be the only match in the macro-block. Thus, the transistors T8 and T9 can be removed to reduce the worst-case delay.

B. Match address encoder

In the presence of multiple matches, the MMR always favors the highest priority match (lowest physical address). We designed the MAE to take advantage of this property. Fig. 3 illustrates this idea using a dynamic CMOS encoder. In this circuit, the worst-case power consumption corresponds to the least likely condition ($R_N = '1'$). In our design, we chose a pseudo-NMOS based MAE (clock = '0') because it consumes less power for high-speed TCAM operations due to the absence of the clock power. In addition, the impact of static power is minimized by the MMR action.

III. MULTIPLE MATCH DETECTION

The MMD circuit normally operates in parallel with the MMR. It is a ternary decision circuit whose 2-bit output indicates the presence of 0, 1, or multiple matches. A current-race MMD scheme

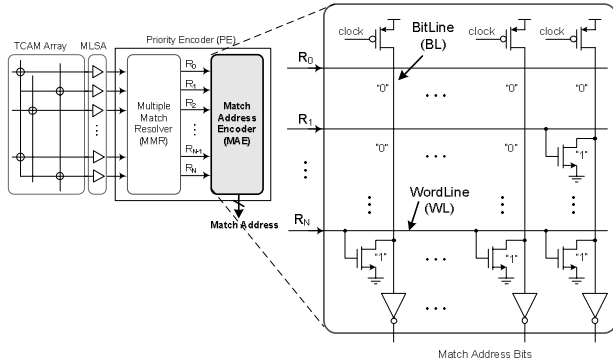


Fig. 3: CMOS dynamic MAE favoring the highest priority address

(shown in Fig. 4) is attractive due to the absence of static power for $MLSO_i = '1'$ [4]. It compares the rising voltage rate of the MML against that of a reference line (RMML). A conducting reference transistor (TR), sized same as the other transistors (TN), emulates a “single match” on RMML.

Prior to the detection phase, the external signal EN2b is held at GND. The output nodes MMSO and RMMSO are either ‘1’ or ‘0’, depending on the result of the previous detection cycle. Another control signal EN1 is at V_{DD} to precharge both MML and RMML to GND. The circuit is initiated by the rising edge of EN2b, which sets MMSO, RMMSO and EN1 to GND, and turns on the current sources (I_{BIAS}). In the “no match” case, the rate of increase on the MML voltage would be faster than the rate on the RMML voltage. When the MML voltage is above the threshold voltage of $T4$ (V_{tn}), the output MMSO switches to ‘1’. The EN1 signal latches MMSO and RMMSO, and resets the MML and RMML back to GND. Table 1 summarizes the conditions and their encoded results. In this scheme, the speed is limited by the time of charging MML or RMML from 0V to V_{tn} . Increasing I_{BIAS} can improve the sensing speed by sacrificing the noise margin between the three conditions.

We propose a modified current-race scheme to improve the sensing speed without reducing the noise margin. Fig. 5 shows the circuit schematic and timing diagram of the proposed scheme for two matches. The MMD operation starts with the falling edge of MMRST, which turns on the current sources and initiates the race. Since RMML is emulating a “single match” condition, and there are two matches on MML as specified, the voltage at RMMSP increases at a faster rate. As a consequence, RMMSO switches from ‘0’ to ‘1’, which is latched by SCLK.

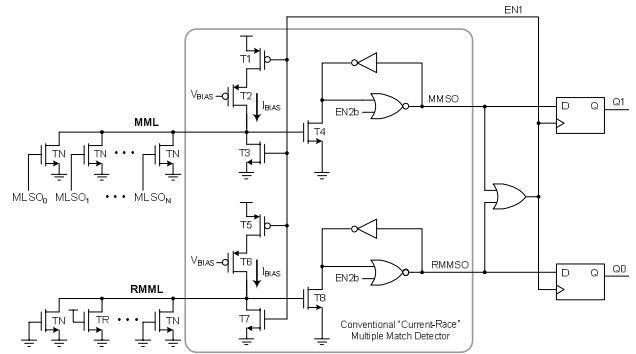


Fig. 4: Conventional current-race multiple match detection circuit

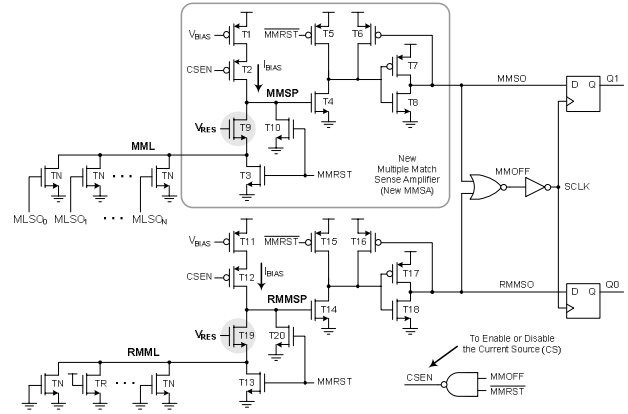


Fig. 5: Proposed multiple match detection circuit

Table 1: Interpretations of the current-race MMD circuit outputs

Condition	Q1	Q2	Interpretation
$V_{MML} > V_{RMML}$	1	0	No Match
$V_{MML} \approx V_{RMML}$	1	1	Single Match
$V_{MML} < V_{RMML}$	0	1	Multiple Matches

The introduction of transistors $T9$ and $T19$ improves the MMD circuit in three ways. First of all, an increase at the source voltage of $T9$ (or $T19$) during the detection phase would increase V_{tn} of $T9$ (or $T19$) due to the body effect. For the “no match” condition, this V_{tn} -shift is significant. For the “single match” condition, it is moderate. For the “multiple match” condition, it is minor or even not noticeable. With respect to the sensing point MMSP, this temporal V_{tn} -shift helps to increase the net pull-up current conditionally, and in turn helps to increase the overall sensing speed and widen the noise margins. Secondly, the resistance of pass-transistor $T9$ (or $T19$) shoots up when its source voltage (V_s) is approaching ($V_{RES} - V_{tn}$). This property also favors the “no match” condition because MMSP is rising at the fastest rate. Thus, V_{RES} must be chosen to maximize the noise margin between the “single match” condition and the “multiple match”

conditions. Finally, a current flow across T9 (or T19) creates an IR drop, which reduces the voltage swing on the highly capacitive MML (or RMML).

IV. TEST CHIP MEASUREMENT RESULTS

We implemented a 64-bit PE and a 256-bit MMD on two test chips in 0.18 μm 1.8V CMOS technology (Fig. 6). For comparison purposes, the MMD chip also contains the conventional current race scheme. The measurement results of the PE chip are summarized in Table 2. MAE consumes significantly large energy in the worst-case situation (only address 64 has a match). However, this case is very unlikely under the normal operating conditions, and MAE consumes much smaller energy than the MMR. Table 3 compares the proposed PE with other recent designs. Although the other two schemes have been fabricated in a different process technology, the proposed PE significantly outperforms them. In addition, cell-based regular structure of the proposed PE is more suitable for the integration with the TCAM array.

Fig. 7 shows the MMD measurement results of the MMD chip comparing the worst-case energy and delay of the proposed MMD circuit with those of the conventional MMD for different values of V_{RES} . As explained in section III (Fig. 5), increasing V_{RES} reduces both energy and delay. The MMD remains functional for $V_{\text{RES}} \geq 1\text{V}$. At $V_{\text{RES}} = 1\text{V}$, the proposed MMD shows 70% improvement both in delay and energy consumption. In our design, we used an off-chip V_{RES} . In larger TCAMs (18Mb), on-chip V_{RES} generation becomes attractive as the power penalty due to the V_{RES} bias circuitry becomes negligible.

Table 2: Measurement results of the test chip with 64-bit PE

Feature	Matching address (es)	Result
Propagation delay	64	1.5ns
MMR energy per cycle	1	1.15pJ
	64	1.18pJ
MAE energy per cycle (@500MHz)	1 and 64	1.61pJ
	1	8fJ
PE energy per cycle	64	2.23pJ
	1	1.16pJ
PE energy per cycle	1	1.16pJ
	64	3.41pJ

Table 3: Proposed 64-bit PE along with the other recent designs

Feature	Proposed PE	Wang [5]	Huang [1]
Size	64-bit	32-bit	256-bit
Process technology	0.18 μm	0.6 μm	0.6 μm
V_{DD} (V)	1.8	3.0	3.0
Maximum speed	500MHz	100MHz	116MHz
Energy per cycle	1.16pJ-3.41pJ	90pJ	274pJ
TCAM compatibility	Yes	No	No

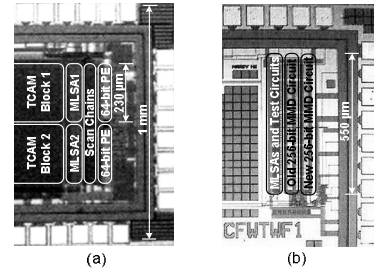


Fig. 6: Chip micrographs: (a) 64-bit PE (b) 256-bit MMD

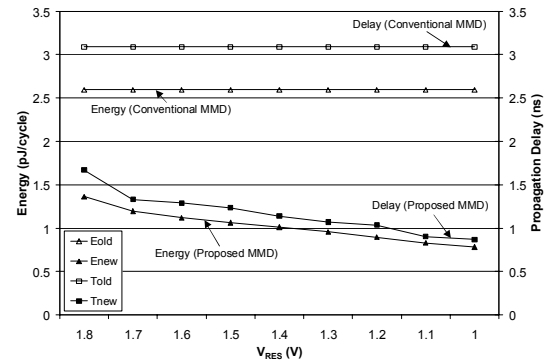


Fig. 7: Test chip measurement results comparing worst-case delay and energy of conventional and proposed MMD circuits

V. CONCLUSIONS

We presented a 64-bit PE and a 256-bit MMD for low-power TCAMs. The PE has a regular structure that can be easily integrated with the TCAM array. The MMD circuit exploits the body effect to reduce the energy and delay. The PE chip measurement results show significant reduction in energy and delay over the recently published designs. The MMD chip measurement results show up to 70% reduction in energy and delay over the conventional current-race scheme.

REFERENCES

1. Y. Huang, C. Huang and J. Wang, "Design of high-performance CMOS priority encoders and incrementor/decrementors using multilevel lookahead and multilevel folding techniques," *IEEE Journal of Solid-State Circuits*, Vol. 37, No. 1, pp.63-76, Jan. 2002.
2. R. Foss and A. Roth, "Priority encoder circuit and method for content addressable memory," Canadian Patent 2365891, Apr. 30, 2003.
3. W. Fung, "Low power circuits for multiple match resolution and detection in ternary CAM," MASc thesis, University of Waterloo, Waterloo, ON, Canada, pp. 22-42, 2004.
4. S. Ma and P. Ma, "Multiple Match Detection Circuit and Method," Canadian Patent 2310295, Nov. 30, 2001.
5. J. Wang and C. Huang, "High-Performance Encoder With Priority Lookahead," *IEEE Journal of Solid-State Circuits*, Vol. 35, No. 10, pp.1511-1514, Oct. 2000.