Match Line Sense Amplifiers with Positive Feedback for Low-Power Content Addressable Memories

N. Mohan, W. Fung, D. Wright and M. Sachdev
Department of Electrical and Computer Engineering, University of Waterloo, Ontario, Canada - N2L 3G1

***Abstract*-Match line sense amplifiers (MLSAs) consume a significant portion of the total power in ternary content addressable memories (TCAMs). In this paper, we present two MLSAs that employ positive feedback techniques to reduce the power consumption. Energy measurement results of the two MLSAs, fabricated with a 144x144 TCAM array in 0.18µm CMOS technology, show a reduction of 56% and 48% respectively over the conventional current-race MLSA.**

I. Introduction

Ternary content addressable memories (TCAMs) are attractive for high-speed lookup-intensive applications such as packet classification and forwarding in network routers. The parallel search operation in large TCAMs results in high power consumption. A significant portion of the total TCAM power is consumed by match line sense amplifiers (MLSAs) due to frequent switching of highly capacitive match lines (MLs). The Current-Race MLSA (CR-MLSA), shown in Fig. 1(a), is an attractive scheme that achieves power savings by reducing ML voltage swing and search line (SL) switching activity [1]. Initially, MLs are discharged to GND, and MLSA outputs are reset to '0'. The search operation starts by enabling the ML current sources (I_{BIAS}). If a TCAM word matches with the search key, its ML does not have a current discharge path. Thus, it charges faster than the MLs with 1- or multiple-bit mismatch conditions (Fig. 1(b)). In the remaining paper, we'll denote matching MLs by ML_0 and MLs with k-bit mismatch by ML_k (as shown in Fig. 1(b)), where $k \geq 1$. A dummy word, which matches in every search operation regardless of the search key, generates a signal ($\overline{MLOFF}$) that turns off the current sources indicating the completion of the search operation.

The conventional CR-MLSA charges all MLs (ML_0 and ML_k) with the same current. Since the MLSA outputs of ML_k's remain at '0' even after enabling their current-sources, this current is wasted. This unnecessary power consumption is amplified by the fact that most TCAM words do not match with the search key. An ideal CR-MLSA will provide the maximum current to ML_0 (maximizing the speed) and the minimum current to ML_k (minimizing the power). Supplying a smaller current to ML_k (particularly ML_1) also improves the robustness of the MLSA by making it easier to detect the difference between ML_0 and ML_1. A mismatch-dependent MLSA (MD-MLSA) has been published with simulation results showing 40% energy reduction over the conventional CR-MLSA [2]. However, the MD-MLSA consumes static power, which becomes significant in TCAMs employing architecture-level techniques for reducing the chip activity.

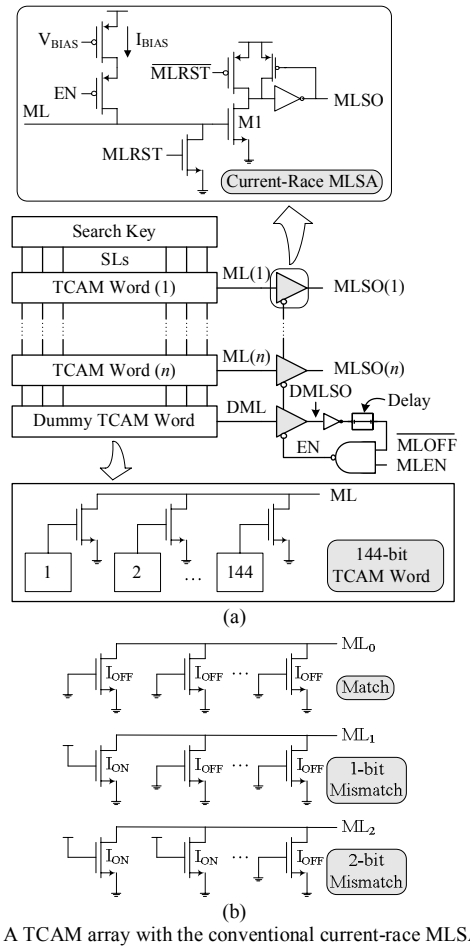

Fig. 1. (a) A TCAM array with the conventional current-race MLSA, and (b) ML current discharge paths for different match/mismatch conditions

In this paper, we present two modified CR-MLSAs, which achieve power reduction by applying positive feedback technique in ML sensing. The proposed MLSAs do not consume any static power, and they outperform the MD-MLSA in speed, energy, area and robustness.

II. MLSA with Resistive Feedback

Fig. 2(a) shows the proposed MLSA with resistive feedback. It uses an NMOS transistor (M2) in the triode region to decouple the ML and its MLSA. For a fixed V_{RES}, as the ML voltage increases, the gate-to-source voltage of M2 (V_{GS_M2}) reduces and its threshold voltage increases due to the body effect, which becomes stronger with the increasing source-to-body voltage (V_{SB_M2}). Both these effects increase

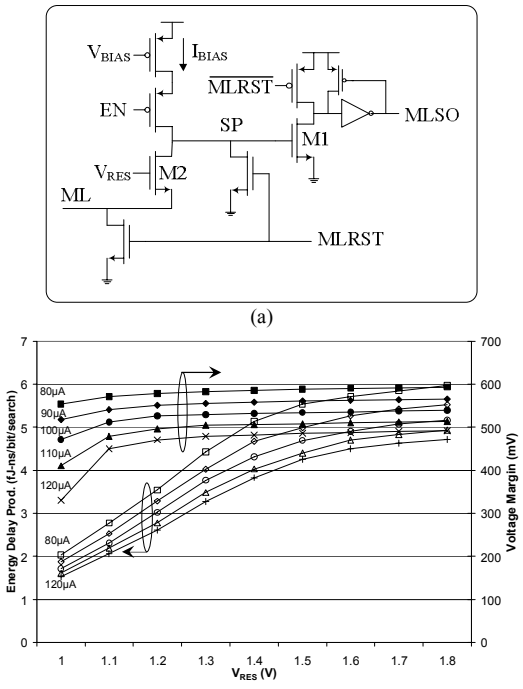

Fig. 2. (a) Proposed MLSA with resistive feedback, and (b) effect of V_{RES} on energy delay product and voltage margin between ML_0 and ML_1.

M2-resistance with the ML voltage. Since less current is now being diverted to the ML, the node SP charges much faster reaching the threshold voltage quickly. Faster sensing of ML_0 also reduces energy consumption because the ML current sources are shut-down sooner in this case. The charging current of an ML_k is less affected by M2 because V_{GS_M2} is larger and V_{SB_M2} is smaller in this case. Energy and delay can be further reduced by decreasing V_{RES}. Although the positive feedback action results in a large voltage margin between ML_0 and ML_1, a combination of small V_{RES} and large I_{BIAS} may reduce the voltage margin causing a false match for ML_1. Fig. 2(b) shows the effect of V_{RES} on the energy delay product (EDP) and the voltage margin for different I_{BIAS} (word-size=144bit). The effectiveness of this scheme is reinforced by the fact that a reduction in V_{RES} decreases the EDP more rapidly than the voltage margin. In addition, for a small V_{RES}, a reduction in I_{BIAS} improves the voltage margin significantly without making much difference in the EDP.

III. MLSA WITH ACTIVE FEEDBACK

In order to reduce the EDP without sacrificing the voltage margin, we developed a CR-MLSA with active feedback. The proposed MLSA is shown in Fig. 3(a). Transistor M3 operates as a current source (I_{FB}) to bias the feedback circuit. MLEN signal (shown in Fig 1) enables the MLSA by activating EN, I_{BIAS} and I_{FB} (Fig. 3(a)). Initially, all the MLs receive the same current from the current sources (I_{BIAS}). As ML_0 charges at a faster rate than ML_k, its M2 overdrive voltage, and thus V_{CS}, becomes smaller than that of ML_k. As a consequence, ML_0 receives higher current and charges more

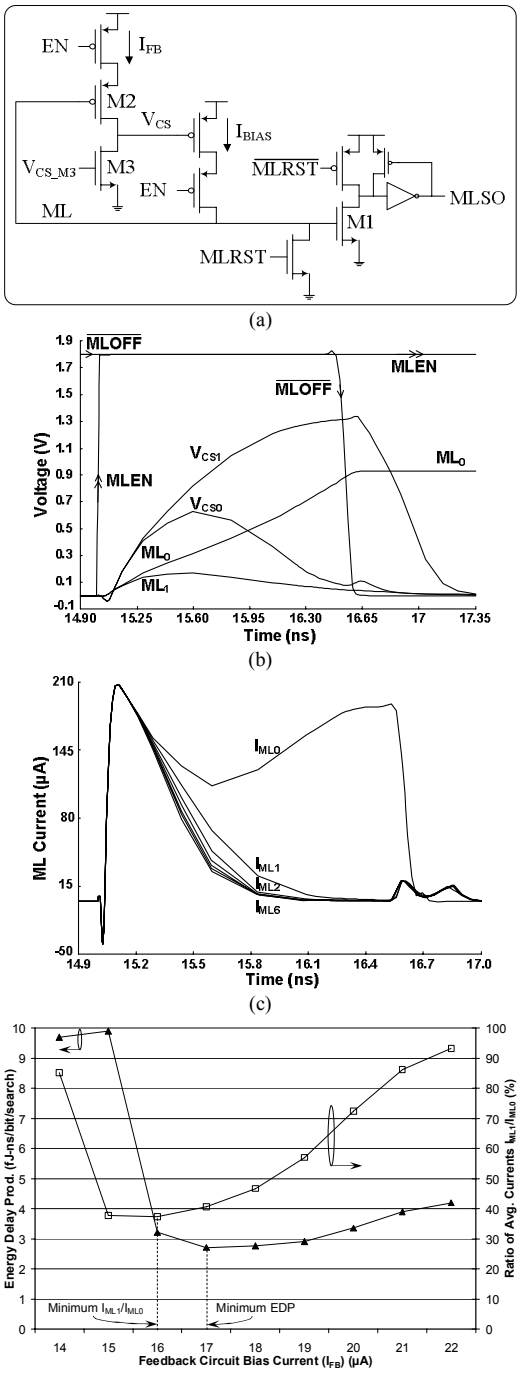

Fig. 3. (a) Proposed MLSA with active feedback with (b) voltage waveforms, (c) current waveforms, and (d) energy-delay product and I_{ML1}/I_{ML0} with I_{FB}

rapidly than ML_k. This positive feedback action continues until $\overline{MLOFF}$ turns off the current sources. Fig. 3(b) shows the voltage waveforms of ML_0 and ML_1. The positive feedback action starts around 250ps after enabling the current sources. Subsequently, ML_0 and ML_1 diverge significantly from each other. Fig. 3(c) shows the current waveforms for ML_0 and ML_k (for k=1 through 6) highlighting two main features of the proposed scheme: (i) I_{ML0} is significantly larger

than I_{ML1}, and (ii) I_{MLk} is a weak function of 'k'. The former confirms that the scheme has a large sense margin between ML_0 and ML_1, and the latter suggests that this scheme is equally applicable to TCAM applications where the average value of 'k' is small. Fig. 3(d) shows the variations in EDP and the ratio of average currents flowing into ML_1 and ML_0 with I_{FB}. As mentioned in section I, the ratio I_{ML1}/I_{ML0} should be minimized to improve the robustness of the ML sensing. For $I_{FB}\geq16\mu A$, the EDP does not change significantly. However, I_{ML1}/I_{ML0} increases rapidly for $I_{FB}>19\mu A$. Thus, choosing $I_{FB}=16\mu A$ or $17\mu A$ provides a good trade-off between the MLSA's performance and its robustness.

It can be noticed in Fig. 3(b) that the ML voltage always remains below 0.9V. Thus, the body effect can be exploited in favor of the positive feedback technique by connecting the ML to the body (substrate) of M1. For example, the higher voltage of ML_0, if connected to the body of M1, reduces the threshold voltage of M1, which results in lower ML voltage swing and faster switching ('0'→'1') of the corresponding MLSO. It also expedites the arrival of $\overline{MLOFF}$ that turns off the current sources quickly and saves energy (Fig. 1(a)). However, M1 body also has some capacitance with the surrounding n-wells, which are connected to V_{DD}. Therefore, this technique increases the ML capacitance while decreasing the ML voltage swing. In addition, it requires a process technology that allows an NMOS body connection other than GND. The resulting energy reduction is also highly dependent on the process technology.

Fig. 4 shows the energy simulation results comparing the proposed MLSAs with the conventional CR-MLSA for an MLSA sensing time (T_{MLSA}) of 1.5ns and a word-size of 144bit. As expected, the resistive feedback MLSA consumes less energy for a smaller V_{RES} and remains functional for $V_{RES}\geq1V$. The active feedback MLSA shows a 43% reduction in ML sensing energy even for 1-bit mismatch (k=1). The energy reduction reaches 51% for higher numbers of mismatches. Moreover, the body-bias technique further improves the energy efficiency of the active feedback MLSA.

IV. Test Chip Measurement Results

We implemented the proposed MLSAs on a test chip (1mm x 2mm) in 0.18µm 1.8V CMOS technology (chip micrograph shown in Fig. 5). It contains a 144x144 TCAM array divided into 4 blocks. Blocks 1 and 2 contain 64 words each with the conventional and the proposed active feedback CR-MLSAs respectively. Blocks 3 and 4 contain 8 words each with the proposed resistive feedback and body-bias active feedback CR-MLSAs. In order to perform exhaustive testing of the test chip, we also included other peripheral components such as priority encoders, address and column decoders, data multiplexers, scan chains, etc. The test chip was designed to perform only write and search operations, and bit line sense amplifiers (for read operations) were not included due to die-area constraints. CMOS 0.18µm bulk technology also provides a deep n-well (DNW) layer. In order to implement the body-bias technique, we used the DNW layer to isolate

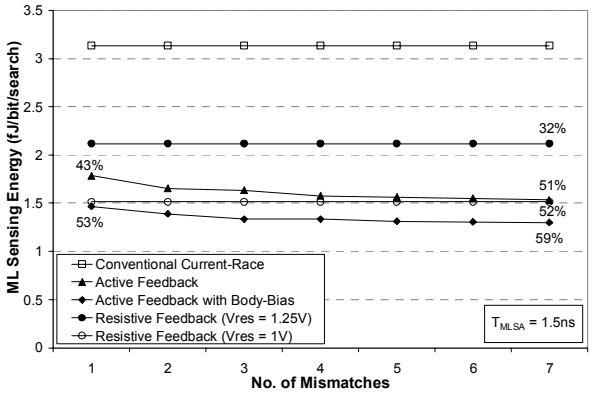

Fig. 4. Simulation results of the proposed and conventional CR-MLSAs

Fig. 5. Micrograph of the test chip including a 144x144 TCAM

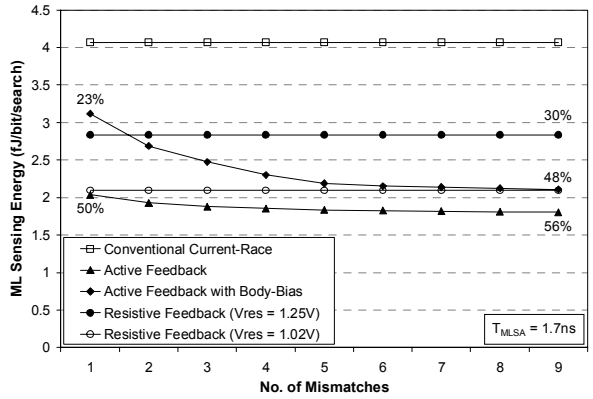

Fig. 6. Measurement results of the proposed and conventional CR-MLSAs

M1 body (shown in Fig. 3(a)) from rest of the p-substrate. However, this choice results in a fourfold increase in the MLSA area in this technology due to large DNW design rules. Fig. 6 shows the test chip measurement results comparing the energy consumption of the proposed MLSAs with that of the conventional CR-MLSA for T_{MLSA}=1.7ns at V_{DD}=1.8V. The measured results are reasonably close to the simulation results except the body-bias active feedback MLSA, which shows degradation in energy savings possibly due to a large body capacitance (connected to MLs), which has not been modeled accurately in the simulations. Therefore, the body-bias technique is not suitable for the given process technology. However, this technique can be beneficial in other process technologies such as silicon-on-insulator (SOI), where the body capacitance is much smaller.

The resistive feedback MLSA remains functional for $V_{RES} \geq 1.02V$. $V_{RES}=1.02V$ ($I_{BIAS}=79.69\mu A$ and $T_{MLSA}=1.7ns$) gives an energy reduction of 48% over the conventional CR-MLSA. If an additional design margin of 0.25V is given ($V_{RES}=1.25V$, $I_{BIAS}=114.39\mu A$ and $T_{MLSA}=1.7ns$), this MLSA still achieves an energy reduction of 30%. The active feedback MLSA shows an energy reduction of 50% even for 1-bit mismatch. The energy consumption further reduces to 56% for higher number of mismatches. It should be noted in Fig. 6 that the energy consumption is a weak function of the number of mismatches. Hence, the active feedback MLSA is equally attractive for TCAM applications where the average number of mismatches is small. The MLSA sensing time $T_{MLSA}=1.7ns$ is obtained at $I_{FB}=16.58\mu A$ (shown in Fig. 3(a)) by EDP minimization. The body-bias active feedback MLSA and the conventional CR-MLSA achieve the same speed at $I_{FB}=14.76\mu A$ and $I_{BIAS}=141.37\mu A$ respectively.

V. Discussion

In order to compare the proposed active feedback MLSA with the existing MD-MLSA, we used the conventional CR-MLSA as the reference design due to two main reasons: (i) the complicated circuit of MD-MLSA was difficult to reproduce in the current technology, and (ii) it was difficult to control the energy and delay of MD-MLSA for comparison purposes. Table I compares the active feedback MLSA with MD-MLSA with reference to the conventional CR-MLSA.

As mentioned in section I, MD-MLSA consumes static power which is around 2% of its dynamic power [2]. Since the average ML current in MD-MLSA is around 80μA, the feedback circuit consumes a static current of 1.6μA [2]. Many architecture-level techniques reduce the chip activity in TCAMs. For example, selective precharge and pipelined ML schemes divide each ML into two or more segments [3][4]. Each segment has a separate MLSA, and only one MLSA is activated in most of the words. Considering the large number of inactive segments, the static power in MLSAs can become a significant portion of the total TCAM power. The contribution of MLSA static power further increases in TCAMs that employ bank selection scheme [5].

The static power in MD-MLSA can be eliminated by adding a switch between its feedback bias current source and the remaining circuit. However, the turn-on time of the feedback circuit will be reasonably large due to a small feedback bias current (~1.6μA). The proposed MLSA turns-on faster due to relatively larger feedback bias current ($I_{FB}=16.58\mu A$). In addition, MD-MLSA uses a level-shifter (PMOS source follower) in the feedback loop, which reduces the feedback loop gain and bandwidth. The skewed sizing of two series-connected PMOS transistors also increases the settling time of the level-shifter. A smaller loop-gain reduces the sense margin between I_{ML0} and I_{ML1}, and also makes the circuit relatively slower. Table I confirms this deduction as the higher loop gain makes the proposed MLSA outperform the MD-MLSA both in speed and sense margin even though it is implemented in a larger feature size technology.

TABLE I
Comparison of MD-MLSA and Active Feedback MLSA

Feature	MD-MLSA [2]	Active feedback MLSA
1. Process technology	0.13μm CMOS	0.18μm CMOS
2. Type of results	Simulations	Measurements
3. ML sensing time (T_{MLSA})	2ns	1.7ns
4. Energy reduction (w.r.t. conventional CR-MLSA)	40%	56%
5. Sense margin: I_{ML0} and I_{ML1}	20-25%[a]	50%
6. Static power consumption	Yes	No
7. Number of transistors	5	3
8. FB circuit bias current (I_{FB})	~1.6μA[a]	16.58μA
9. Level shifter	Yes	No

[a] Deduced from [2]

The robustness of an MLSA is determined by its ability to detect the difference between discharge currents of ML_0 and ML_1. As illustrated in Fig. 1(b), the ML discharge current increases with the number of mismatches. Since ML_0 has no mismatch, it has no discharge path to GND except the leakage currents (I_{OFF}). Since the transistor leakage and TCAM word-size are increasing due to technology scaling and new applications (such as IPv6) respectively, the difference between discharge currents of ML_0 and ML_1 is decreasing. Thus, detecting the difference between ML_0 and ML_1 is becoming increasingly difficult. The proposed active feedback MLSA is more robust than MD-MLSA because it has a larger sense margin (I_{ML1} is 50% of I_{ML0}). The larger sense margin also helps in coping with the process variations, which are increasing with technology scaling.

VI. Conclusions

We presented two MLSAs that use positive feedback to reduce the power consumption in TCAMs. The proposed resistive feedback MLSA exploits the body effect to reduce the EDP without affecting the voltage margin. The proposed active feedback MLSA requires fewer transistors than the existing MD-MLSA. The proposed MLSA also improves energy savings and voltage margin without consuming any static power. The measurement results of the two MLSAs reduce ML sensing energy by 48% and 56% respectively.

References

[1] I. Arsovski, T. Chandler, A. Sheikholeslami, "A ternary content-addressable memory (TCAM) based on 4T static storage and including a current-race sensing scheme," *IEEE J. Solid-state Circuits*, vol. 38, no. 1, pp. 155-158, Jan. 2003.

[2] I. Arsovski, A. Sheikholeslami, "A mismatch-dependent power allocation technique for match-line sensing in content addressable memories," *IEEE J. Solid-state Circuits*, vol. 38, no. 11, pp. 1958-1966, Nov. 2003.

[3] C. Zukowski, S. Wang, "Use of selective precharge for low-power CAMs," *Proc. IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 745-770, Jun. 9-12, 1997.

[4] K. Pagiamtzis, A. Sheikholeslami, "A low-power content-addressable memory (CAM) using pipelined hierarchical search scheme," *IEEE J. Solid-State Circuits*, vol. 39, no. 9, pp. 1512–1519, Sep. 2004.

[5] G. Kasai, Y. Takarabe, K. Furumi, 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," *Proc. IEEE Custom Integrated Circuits Conference (CICC)*, pp. 387-390, Sep. 2003.