A Comparative Study of Low-Power Cam Match-Line Sense Amplifier Designs

Becky Elfreda.J
Department of Electrical & Electronics Engineering
Dr.Mahalingam College of Engineering & Technology
Pollachi,Tamilnadu,India

Nandhakumar.A
Department of Electrical & Electronics Engineering
Dr.Mahalingam College of Engineering & Technology
Pollachi,Tamilnadu,India

ABSTRACT
Robust, high-performance and low-power match-line sense amplifier designs are urgently required to catch up with the new requirements of large-scale CAMs in nano-scale CMOS technologies. In this paper we evaluate the performance of four state-of-the-art match-line sense amplifier designs in terms of power, delay and robustness against temperature, supply voltage and process variations. Our results show that the pre-charge low match-line sensing schemes suffers from process variations. Despite featuring low power consumption, these designs can hardly be scaled down to operate in low-voltage sub-65 nm CMOS process. On the other hand, the conventional and the charge-injection designs are much more robust and hence more suitable for low-voltage sub-65 nm CMOS implementations.

Keywords
Low-power, CMOS memory, Content addressable memory (CAM)

1. INTRODUCTION
Content Addressable Memory (CAM) is extremely power-hungry, due to its parallel search nature. To make things worse, CAM power consumption is almost linearly dependent on the rapidly increasing CAMs capacity [1]. Fig. 1 shows the basic block diagram of a CAM, consisting of an array of CAM cells, a search-word register, a column of sense amplifiers and a priority encoder. Each row of the array has n CAM cells (i.e. an n-bit CAM word) and one associated match-line (*ML*). Each CAM cell has one SRAM storage element and two sets of NOR-type comparison circuits ($N1$-$N4$) as shown in Fig. 1. In this work, we use the conventional 10T NOR-type CAM cell (Fig. 1 - the two NMOS access transistors of the SRAM storage element are not shown for the sake of simplicity) as a benchmark to demonstrate the operation of the CAM and the *MLSA*.

Its operation is as follows: A CAM search operation starts by loading the search word into the search-word register. The search data are then broadcast to the array through the n-pairs of differential *SL*s. The search data on the *SL*s will be compared directly with the stored data within each CAM cell by the comparison circuits. If at least one mismatch occurs on a row one of the compare branches in the mismatched cell is turned on. This will discharge the *ML* to ground, indicating a miss. If all of the stored bits on a row are identical to the search bits, none of the compare branches on the row will be turned on hence the *ML* voltage remains unchanged, indicating a match. A sense amplifier is used for each row to improve the speed by digitizing

the voltage transition on the *ML*, as shown in Fig. 1. The priority encoder receives the search results from the sense amplifiers and returns the address of the highest priority row that has a match.
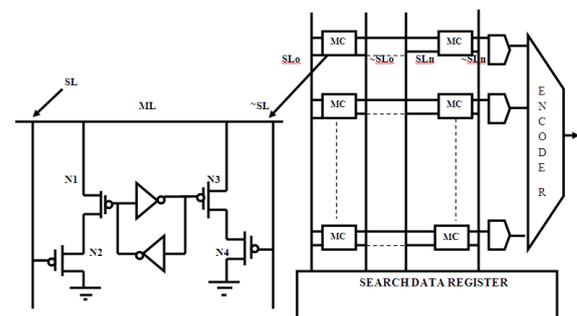

Figure 1: A generic CAM architecture consists of an array of CAM cells, a search data register, a column of MLSAs and a priority encoder.

Since CAM compares all of the stored words concurrently, its search speed is high and so is its power consumption [1]. Thus, lots of works have been proposed to reduce the power consumption of the CAM by reducing either the switching activity or the voltage swing of the *ML*s [1] [2–8]. Among these, designs in [2] and [3–6] are the most attractive designs because of their single-clock, high-speed and low power operations. In this work, a comparative study of four state-of-the-art *MLSA*s implemented in 65 nm CMOS process is presented to access the negative impacts of the process and environmental variations on the operation of the *MLSA*s. From this point onwards, these designs are referred to as the conventional [9], the charge-injection [2], the stability [6] and the positive feedback [5] designs.

2. STATE-OF-THE-ART MLSA DESIGNS
In this section we will cover the basic operating principles of the four *MLSA* designs in consideration.

2.1 Conventional design
Fig. 2 shows the schematic of the conventional design. It consists of two *P MOS*s (*P* 1-*P* 2), one *NMOS* (*N*1) and one output inverter (P3-N2). It operates as follows: During pre-charge, the *MLP RE* signal deactivates the *SA* by turning on *P* 1

and $N1$, pre-charging node $MLso$ to V_{DD}. No DC current is allowed to flow within the SA. During evaluation, the SA is activated by asserting the $MLPRE$ signal low to turn off transistor $P1$ and $N2$. If there is at least one mismatch, the ML voltage, V_{ML}, will be discharged to a lower level and eventually to ground. The output inverter will therefore returns a "1" at node $MLso$, indicating a miss. If no mismatch occurs, the V_{ML} will stay at V_{DD}, indicating match.
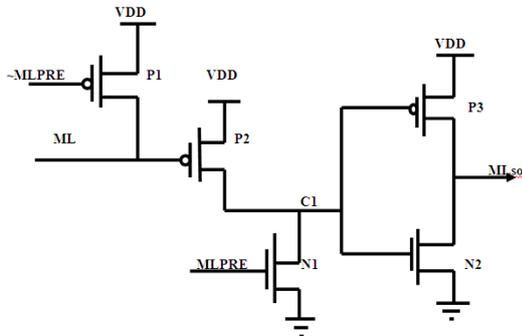

Figure 2: The conventional *MLSA* [9] is the simplest design with two PMOSs, one NMOS and one inverter.

2.2 Charge injection design

To reduce the voltage swing of the MLs, a charge injection sensing scheme was proposed [2]. It consists of one Injection Capacitor, two reset transistors $P1$ and $N1$, one injection transistor $N2$ and one asymmetric latch-type comparator, as shown in Fig. 3. Its operation can be divided into three phases: Injection Capacitor and ML pre-charge, charge injection and evaluation. During pre-charge, the Injection Capacitor and the ML are reset to V_{DD} and ground, respectively. Meanwhile, the SAequalizer is asserted high to reset the sense amplifier. During the charge-injection phase, the ChargeIn signal is triggered high for a short while to share the charge from the Injection Capacitor to the ML, bringing the ML potential from ground to a pre-determined voltage much lower than V_{DD}, followed by the evaluation phase. Similar to the conventional design, during the evaluation phase the matched ML will remain unchanged while the missed ML is discharged to ground. An asymmetric latch-type comparator is then enabled by the SAequalizer signal to determine the compare result. Its detailed operation can be found in [2]
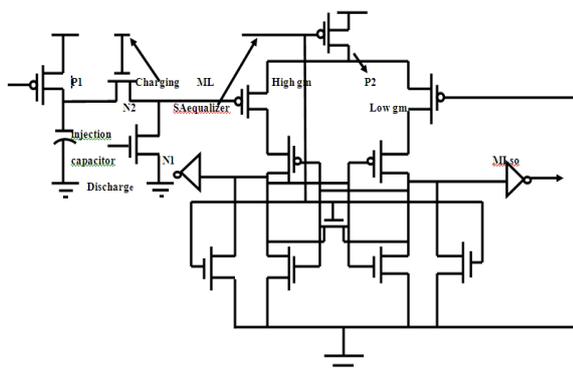

Figure 3: The charge injection *MLSA* [2] uses an explicit *Injection Capacitor* to limit the *ML* voltage swing and includes a latch-type comparator to amplify the sensed *ML* result.

2.3 ML stability design

The ML stability design (Fig. 4) analyzes the properties of the MLs in the s domain [6]. By shunting a negative resistance of -$2R_{cell}$ to a ML (where R_{cell} models the equivalent resistance of a one mismatch ML), a matched ML becomes an unstable system while the missed MLs remain stable. Thus, if excited by an initial energy, the matched ML will grow to V_{DD} while the missed ML will decay to zero [6]. It consists of a level shifter, a threshold sensor and a -$2R_{cell}$ realization circuit, also shown in Fig. 4. The replica biasing circuit is shared among all SAs. Its operation is as follows:
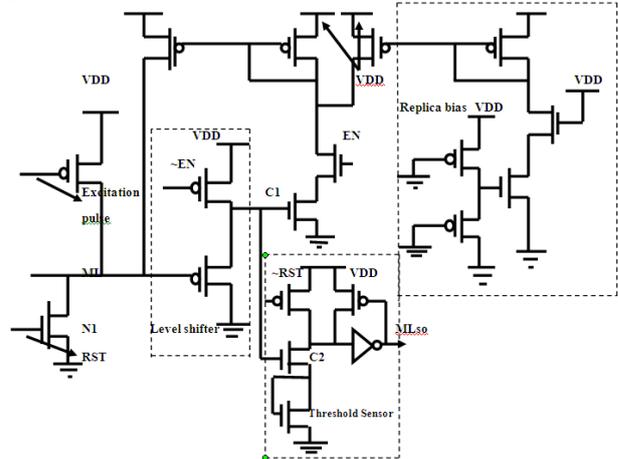

Figure 4: The stability *MLSA* [6] consists of a biasing circuit to implement a negative resistance, a level shifter and a threshold sensor.

During standby, the EN signal is set at zero to turn off any DC current while the RST signal is kept high to pre-charge node $C2$ to V_{DD}. Meanwhile the ML is reset to ground by a reset transistor $N1$. During evaluation, an excitation pulse is used to supply an initial energy to the ML. If it is a match, the ML voltage will rise to a high level, indicating a match through level shifter and a amplifier. On the other hand, if at least one miss occurs, the ML will decay to zero, indicating a miss.

2.4 Positive feedback design

The positive feedback circuit [5] is shown in Fig. 5, which consists of six $P MOS$s, four $NMOS$s and one inverter. Its operation is as follows: During pre-charge, the ML and node $C1$ is reset to ground and V_{DD}, respectively.
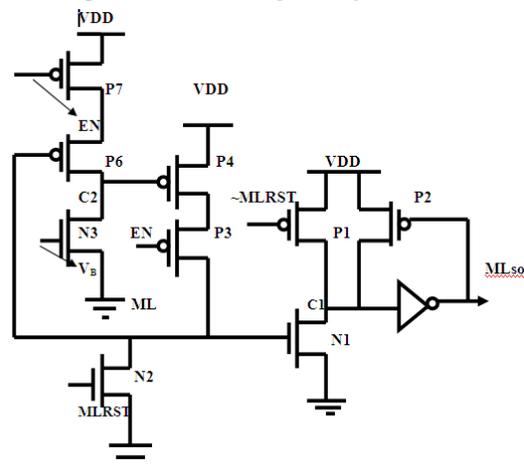

Figure 5: The positive feedback *MLSA* [5] implements a feedback loop to quickly sense the *ML* results where the matched *ML* is supplied with more current while the current to the other *ML*s are gradually reduced to save power.

At the same time, the *EN* signal is kept high to turn off *P* 7 and *P* 3, blocking any potential DC current to flow from V_{DD} to ground. When an evaluation cycle starts, the *EN* and *MLRST* signals are triggered low to activate the sense amplifier. The current source *P* 3−*P* 4 will supply a current i_{ML} to the *ML* and thus the *ML* voltage V_{ML} will rise gradually. Depending on the number of mismatches on the *ML*, this voltage will response differently. In the case of a matched, the *ML* voltage will rise quickly. A high V_{ML} will reduce the source-to-gate voltage of transistor *P* 6 ($V_{GSP\ 6}$), pushing its drain voltage to go low so that the same current can be maintained. Correspondingly, $V_{GSP\ 4}$ becomes larger, allowing more current to flow to the *ML*. This in turns causes the V_{ML} to rise more quickly, forming a positive feedback effect. Once the V_{ML} reaches the threshold voltage of transistor *N*1, i.e. V_{thN1}, *N*1 is turned on and eventually node *C*1 is pulled to ground. The MLSA then outputs a high V_{MLso} value, indicating a match.

On the other hand, in the case of a miss, at least one compare branch in the CAM cells is turned on, prohibiting the V_{ML} to rise quickly. Eventually, node *C*2 will turn off transistor *P* 4 and hence V_{ML} will be discharged completely to ground by the compare circuit within the mismatched cell.

3. PROPOSED ML SENSE AMPLIFIER DESIGN

3.1 Search Speed Boost Using A Parity Bit

We introduce a versatile auxiliary bit to boost the search speed of the CAM at the cost of less than 1% area overhead and power consumption. This newly introduced auxiliary bit at a glance is similar to the existing Pre-computation schemes but in fact has a different operating principle. We first briefly discuss the Pre-computation schemes before presenting our proposed auxiliary bit scheme.
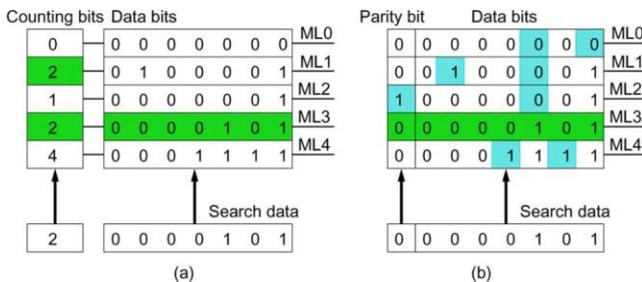

Figure 6: Conceptual view of (a) conventional pre computation CAM and (b) proposed parity-bit based CAM

1) Pre-Computation CAM Design: The pre-computation CAM uses additional bits to filter some mismatched CAM words before the actual comparison. These extra bits are derived from the data bits and are used as the first comparison stage. For example, in Fig. 6(a) number of "1" in the stored words are counted and kept in the Counting bits segment. When a search operation starts, number of "1"s in the search word is counted and stored to the segment on the left of Fig. 6(a). These extra information are compared first and only those that have the same number of "1"s (e.g., the second and the fourth) are turned on in the second sensing stage for further comparison. This scheme reduces a significant amount of power required for data comparison, statistically. The main design idea is to use additional silicon area and search delay to reduce energy consumption. The previously mentioned pre-computation and all other existing designs shares one similar property. The ML sense amplifier essentially has to distinguish between the matched ML and the 1-mismatch ML This makes CAM designs sooner or latter face challenges since the driving strength of the single turned-on path is getting weaker after each process generation while the leakage is getting stronger. This problem is usually referred to as I_{on}/I_{off}. Thus, we propose a new auxiliary bit that can concurrently boost the sensing speed of the ML and at the same time improve the I_{on}/I_{off} of the CAM by two times.

2) Parity Bit Based CAM: The parity bit based CAM design is shown in Fig. 6(b) consisting of the original data segment and an extra one-bit segment, derived from the actual data bits. We only obtain the parity bit, i.e., odd or even number of "1"s. The obtained parity bit is placed directly to the corresponding word and ML . Thus the new architecture has the same interface as the conventional CAM with one extra bit. During the search operation, there is only one single stage as in conventional CAM. Hence, the use of this parity bits does not improve the power performance. However, this additional parity bit, in theory, reduces the sensing delay and boosts the driving strength of the 1-mismatch case (which is the worst case) by half, as discussed below. In the case of a matched in the data segment (e.g.ML3,), the parity bits of the search and the stored word is the same, thus the overall word returns a match. When 1 mismatch occurs in the data segment (e.g.,ML2), numbers of "1"s in the stored and search word must be different by 1. As a result, the corresponding parity bits are different. Therefore now we have two mismatches (one from the parity bit and one from the data bits). If there are two mismatches in the data segment (e.g.,ML0,ML1,or ML4), the parity bits are the same and overall we have two mismatches. With more mismatches, we can ignore these cases as they are not crucial cases. The sense amplifier now only have to identify between the 2-mismatch cases and the matched cases.

Since the driving capability of the 2-mismatch word is twice as strong as that of the 1-mismatch word, the proposed design greatly improves the search speed and the I_{on}/I_{off} ratio of the design. We are going to proposed a new sense amplifier that reduces the power consumption of the CAM.

3.2 Gated-Power Ml Sense Amplifier Design

Operating Principle

The proposed CAM architecture is depicted in Fig. 4. The CAM cells are organized into rows (word) and columns (bit). Each cell has the same number of transistors as the conventional P-type NOR CAM (shown in Fig. 1) and use a similar ML structure. However, the "COMPARISON" unit, i.e., transistors M_1-M_4 , and the "SRAM" unit, i.e., the cross-coupled inverters, are powered by two separate metal rails, namely V_{DDML} and the V_{DD}, respectively. The V_{DDML} is independently controlled by a power transistor(p_x) and a feedback loop that can auto turn-off the ML current to save power. The purpose of having two separate power rails of (V_{DD} and V_{DDML}) is to completely isolate the SRAM cell from any possibility of power disturbances during COMPARE cycle.

As shown in Fig. 8, the gated-power transistor p_x , is controlled by a feedback loop, denoted as "Power Control" which will automatically turn off p_x once the voltage on the ML reaches a certain threshold. At the beginning of each cycle, the ML is first initialized by a global control signal EN . At this time, signal EN is set to low and the power transistor p_x is turned OFF . This will make the signal ML and C1 initialized to ground and V_{DD} , respectively. After that, signal EN turns *HIGH* and initiates the COMPARE phase. If one or more mismatches happen in the

CAM cells, the ML will be charged up. Interestingly, all the cells of a row will share the limited current

Figure 7: Proposed CAM architecture.

Offered by the transistor p_x , despite whatever number of mismatches. When the voltage of the ML reaches the threshold voltage of transistor M_8(i.e.V_{th8},), voltage at node C1 will be pulled down. After a certain but very minor delay, the NAND2 gate will be toggled and thus the power transistor p_x is turned off again. As a result, the ML is not fully charged to V_{DD}, but limited to some voltage slightly above the threshold voltage of M 8, V_{th8}.

With the introduction of the power transistor P_x, the driving strength of the 1-mismatch case is about 10% weaker than that of the conventional design and thus slower. However, as we combine this sense amplifier with the parity bit scheme, the overall search delay is improved by 39%. Thus the new CAM architecture offers both low-power and high-speed operation.

4. SIMULATION RESULTS AND COMPARISON

The software that is used for simulation is microwind. MICROWIND is truly integrated EDA software encompassing IC designs from concept to completion, enabling chip designers to design beyond their imagination. MICROWIND integrates traditionally separated front-end and back-end chip design into an integrated flow, accelerating the design cycle and reduced design complexities.

Process Variation Analysis
Process variation is a critical issue in nano-scale CMOS technologies. We simulate the performance of the proposed design against empirical process variation data from the foundry. It is worth mentioning here that the feedback loop to turn off the gated-power transistor P_x operates digitally and hence is almost insensitive to process variations. Similar to the conventional design, there are two scenarios where the proposed design may sense the results wrongly: 1) the sense amplifier is enabled too early, the 1-mismatch ML has not been pulled up to a voltage higher than the threshold value and thus trigger the output inverter and 2) the delay of the enable signal is too long, resulting in the matched ML to be pulled up by the leakage current, indicating wrong miss.

Figure 9: Output Waveform Of Conventional CAM

Figure 10: Output Waveform Of Proposed CAM

Figure 11. Total average energy consumption of the conventional and the proposed design

PARAMETERS	CONVENTIONAL CAM	PROPOSED CAM
Energy Dissipation	1.148fJ/bit/search	☐ ☐ ☐ ☐ fJ/bit/search
Operating voltage	Below 0.9 V	Below 0.9 V
Average power consumption	☐ ☐ ☐	☐ ☐ ☐

Table 1: Comparison specifications of conventional and proposed design

5. CONCLUSION

In this paper**,** we presented a comprehensive study of four state-of-the-art *MLSA* designs. The stability [6] and positive feedback [5] designs are sensitive to supply voltage, temperature and process variations, despite their merit of low power consumption. The charge injection design [2] demonstrates comprehensive figure of merits in terms of low-power, high-speed and immunity against supply voltage, temperature and process variations. We proposed an effective gated-power technique and a parity-bit based architecture that offer several major advantages, namely reduced peak current (and thus IR drop), average power consumption (36%), boosted search speed (39%) and improved process variation tolerance. At 1 V operating condition, both designs are equally stable with no sensing errors, according to our Monte Carlo simulations. Its area overhead is about 11%. It is therefore the most suitable design for implementing high capacity parallel CAM in sub-65-nmCMOS technologies.

6. REFERENCES

[1] Arsovski I. and Sheikholeslami A. (2003) "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

[2] A. T. Do, S. S. Chen, Z. H. Kong, and K. S. Yeo, "A low-power CAM with efficient power and delay trade-off," in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), 2011, pp. 2573–2576.

[3] I.Arsovski, et al., "A ternary content-addressable memory(TCAM) based on 4T static storage and including a current-race sensing scheme," IEEE Journal of Solid- State Circuits, vol. 38, pp. 155-158, 2003.

[4] Hyjazie J M. and Wang C. (2003) "An approach for improving the speed of content addressable memories," in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), vol. 5 pp. 177–180

[5] N. Mohan and M. Sachdev, "Low-leakage storage cells for ternary content addressable memories," *IEEE Trans. Very Large Scale Integr.(VLSI) Syst.*, vol. 17, no. 5, pp. 604–612, May 2009.

[6] Mohan N. and Sachdev M. (2009) "Low-leakage storage cells for ternary content addressable Memory", IEEE Trans. Very Large Scale Integration.(VLSI) Syst., vol. 17, no. 5, pp. 604–612

[7] Mohan N, Fung W, Wright D and Sachdev M. (2009) "A low-power ternary CAM with positive-feedback match-line sense amplifiers", IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 56, no. 3, pp. 566–573

[8] Lin P F. and Kuo J B. (2002) "A 0.8-V 128-kb four-way set-associative two level CMOS cache memory using two-stage wordline/ bitline-oriented tag-compare (WLOTC/BLOTC) scheme," IEEE J. Solid-State Circuits, vol. 37, no. 10, pp. 1307–1317

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

[10] Rabaey J M, Chandrakasan A. and Nikolic´ B. (2003) DigitalIntegrated Circuits: A Design Perspective, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall

[11] L. Perng-Fei et al., "A 1-V 128-kb four-way set-associative CMOS cache memory using wordline-oriented tag-compare (WLOTC) structure with the content-addressable-memory (CAM) 10-transistor tag cell," IEEE Journal of Solid-State Circuits, vol. 36, pp. 666-675, 2001.

[12] A. T. Do, et al., "A Low-Power CAM with Efficient Power and Delay Trade-off," IEEE International Sym-posium in Circuits and System, ISCAS 2011.

[13] A. T. Do, et al., "Low IR Drop and Low Power Par-allel CAM Design Using Gated Power Transistor Tech-nique," IEEE Asia Pacific Conference Circuits and Sys-tems 2010, APCCAS 2010