A Ternary Content-Addressable Memory (TCAM) Based on 4T Static Storage and Including a Current-Race Sensing Scheme

Igor Arsovski, *Student Member, IEEE*, Trevis Chandler, *Student Member, IEEE*, and Ali Sheikholeslami, *Senior Member, IEEE*

***Abstract*—A 256×144-bit TCAM is designed in 0.18-μm CMOS. The proposed TCAM cell uses 4T static storage for increased density. The proposed match-line (ML) sense scheme reduces power consumption by minimizing switching activity of search-lines and limiting voltage swing of MLs. The scheme achieves a match-time of 3 ns and operates at a minimum supply voltage of 1.2 V.**

***Index Terms*—Associative memories, content-addressable memory (CAM), high density, low power, match-line sensing, storage, ternary.**

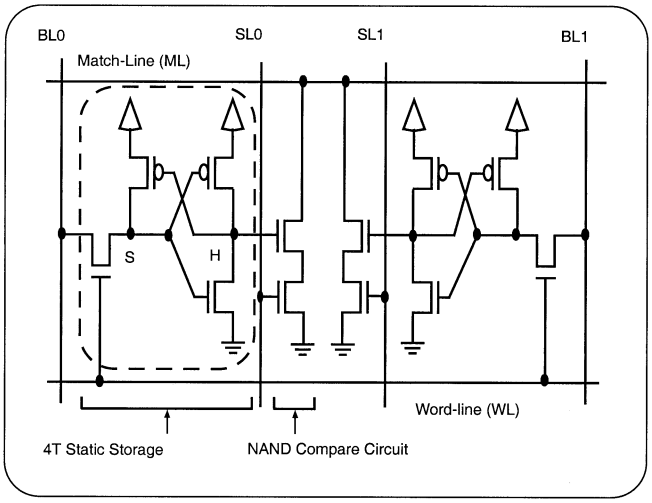
I. Introduction

CONTENT Addressable Memory (CAM) is an application specific memory that allows its entire contents to be searched within a single clock cycle. Binary CAM performs exact-match searches, while a more powerful Ternary CAM (TCAM) allows pattern matching with the use of "don't cares." Don't cares act as wildcards during a search, and are particularly attractive for implementing longest-prefix-match searches in routing tables [1]. Dynamic storage of ternary data [2] requires refresh operation and an embedded DRAM process, while static storage of ternary data requires considerable layout area. This paper features: 1) a compact TCAM cell based on a novel 4T static storage cell and 2) a match sensing scheme that increases speed and reduces power consumption by limiting the voltage swing of the match-lines (MLs) and minimizing the switching activity of the search-lines (SLs).

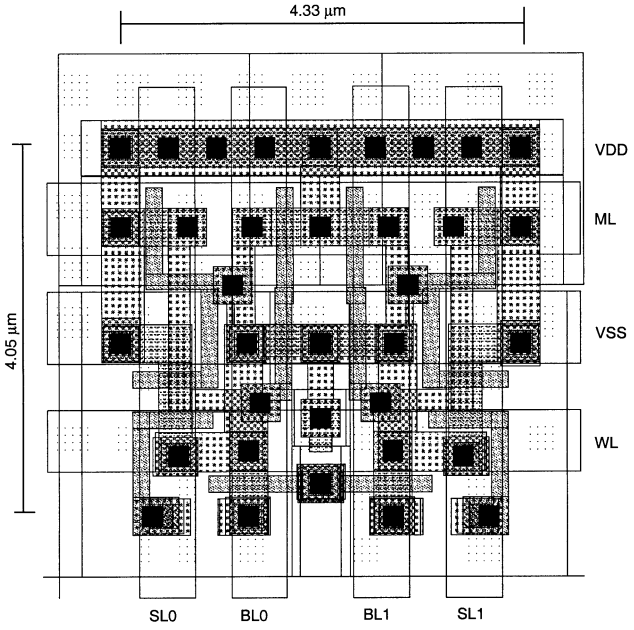
II. TCAM Cell

Fig. 1 highlights an asymmetric 4T cell which is the basic storage element in the TCAM. Unlike the loadless 4T SRAM cell previously reported [3], the proposed asymmetrical arrangement provides a hard node, "H," that stores a rail-to-rail logic signal and is well-suited for interrogation by the NAND compare circuit. Two such cells, plus two NAND compare circuits, are used to form the TCAM cell. The cell layout occupies 17.54 μm^2, the same area as previously reported for a "binary" CAM cell [4], providing the ternary implementation at no extra cost.

Writing to a cell consists of applying data to the bit-line (BL), followed by pulsing the word-line (WL). When the cell is idle, or being searched, the BL is held low at VSS. A "1" stored at node "S" is maintained by the turned-on PMOS. A "0" stored at node "S" is maintained by the sub-threshold leakage current of the NMOS ($I_{\text{off}-n}$), which is larger than that of the PMOS ($I_{\text{off}-p}$). Fig. 2 shows the simulation results confirming that $I_{\text{off}-n}$ is generally larger than $I_{\text{off}-p}$ except in the slow-NMOS process corners. To maintain $I_{\text{off}-n} \gg I_{\text{off}-p}$ over all process corners, an off-chip-generated bias voltage of 0.2 V ($V_{\text{WLB}} = 200$ mV) is



(a)



(b)

Fig. 1. Ternary content-addressable memory cell. (a) Schematic. (b) Layout.

Manuscript received April 2, 2002; revised June 27, 2002. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

The authors are with the University of Toronto, Toronto, ON M55 3G4, Canada (e-mail: arsovsk@eecg.utoronto.ca).

Digital Object Identifier 10.1109/JSSC.2002.806264

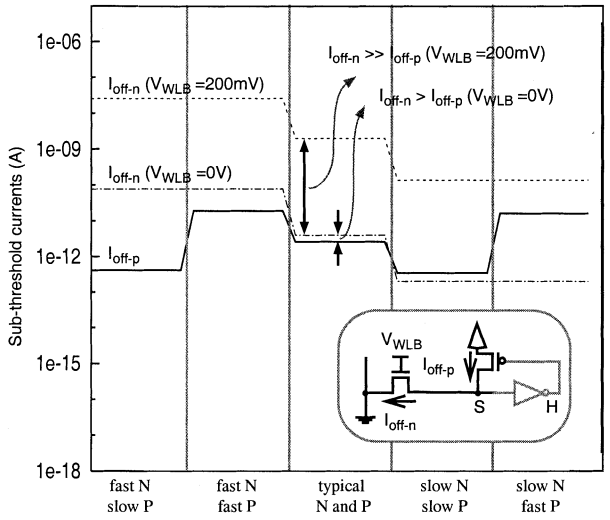


Fig. 2. Comparison of NMOS and PMOS sub-threshold currents over various process corners $V_{\rm DS}({\rm NMOS}) = 0.05$ V, $V_{\rm DS}({\rm PMOS}) = 1.8$ V.

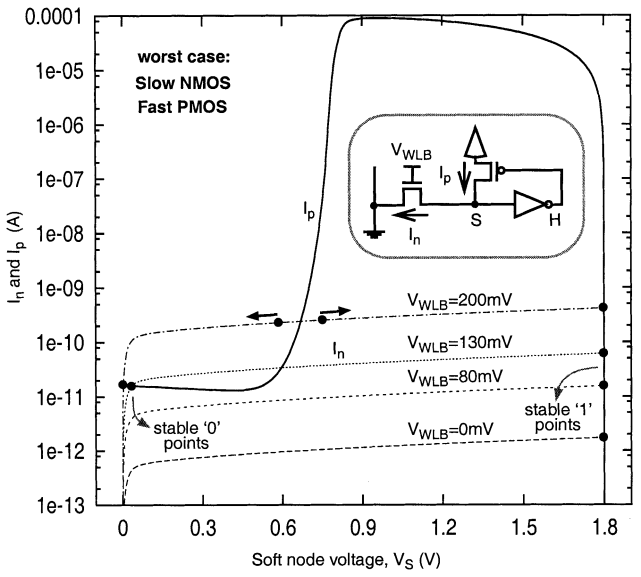


Fig. 3. Worst-case I_n and I_p versus V_s for four values of $V_{\rm WLB}$.

applied to all the inactive WLs. Further simulations (not shown) confirm that with $V_{\rm WLB} = 200$ mV, $I_{{\rm off}-n} \gg I_{{\rm off}-p}$ over a temperature range of 0 °C–75 °C, ensuring proper cell operation. A temperature-and-process-tracking $V_{\rm WLB}$ can be generated using an on-chip replica bias circuitry [5]. However, for test purposes, we rely only on an off-chip $V_{\rm WLB}$.

To verify the worst-case (slow NMOS, fast PMOS) cell stability, we have simulated (as shown in Fig. 3) the NMOS and PMOS currents (I_n and I_p) as functions of the soft node voltage (V_S) for a set of four values of $V_{\rm WLB}$. The steady-state stable points for a stored "0" and a stored "1" are marked in the figure at the far ends of the V_S axis. A stored "1" is stable with a relatively large noise margin (about 0.9 V), independent of $V_{\rm WLB}$. A stored "0" is stable only if $V_{\rm WLB}$ is larger than 130 mV. In fact, at $V_{\rm WLB} = 200$ mV there is a noise margin of 0.6 V for V_S. A higher $V_{\rm WLB}$ would increase the "0" noise margin at the cost of increased static power dissipation of the memory array.

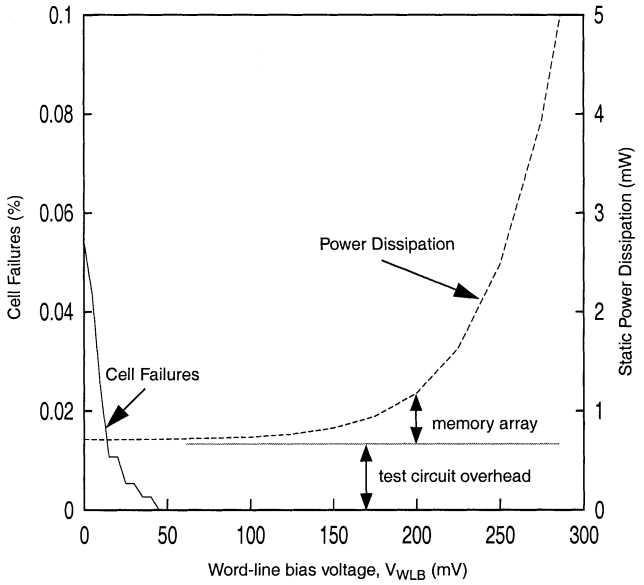


Fig. 4. Measurement results. Cell failures and total static power dissipation versus WL bias voltage.

The test chip measurement results, shown in Fig. 4, depict the percentage of cell failures and the total static power dissipation as functions of $V_{\rm WLB}$. All the cells in the memory array reliably store a "0" at $V_{\rm WLB} > 41$ mV, which is far below the worst-case simulated value of 130 mV. The measured power dissipation includes contributions from both the memory array and test circuits on the chip. Test results show that the power dissipation of the memory array increases substantially for $V_{\rm WLB}$ larger than 200 mV. Combining the measurement results of Fig. 4 with the worst-case simulation results of Fig. 3, results in an $V_{\rm WLB}$ optimum range of 130–200 mV. This range ensures a noise margin above 0.5 V and a memory array power dissipation of less than 470 μW.

III. Current-Race Sensing Scheme

The search operation consists of comparing the data applied to the search-lines (SL) with the data stored in the memory cells through the NAND compare transistors (Fig. 1). Exact-matching is implemented by storing and subsequently searching for complementary data on the two "H" nodes in each TCAM cell. Pattern matching is implemented by using local and global masking. Local masking is implemented by storing "0" on both "H" nodes of an individual TCAM cell. Global masking is implemented by searching for "00" on a pair of SLs.

Prior to a search, the match-lines (ML) are reset to ground and the match-sense nodes of the sense amplifiers are precharged to VDD by asserting MLRST as shown in Fig. 5. Also during MLRST, the new search data is applied to the SLs. Next, by activating MLEN, the current sources attached to each ML begin charging the MLs. Since a path to ground is created by any unmasked mismatch between the 144-bit search and stored data, a ML with no mismatch (i.e., a full match) experiences a charge-up at a higher rate compared to any ML with at least one mismatch. A dummy ML (DML) is also included which is identical to a ML with no mismatch. During the ML charge-up

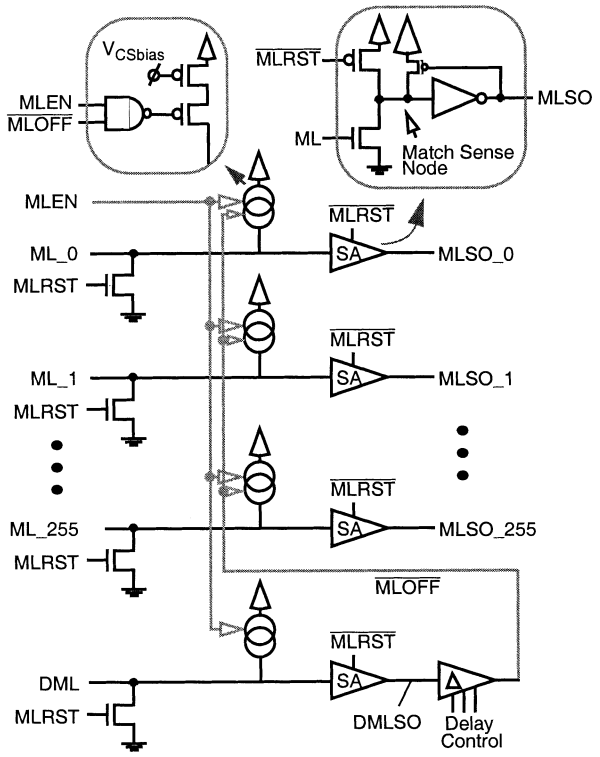


Fig. 5. Current-race ML sensing scheme.

process, all MLs are timed in a race against the DML, to reach the threshold voltage of their corresponding sense amplifier. Once the DML charges up beyond the threshold of the sense amplifier, its match-sense node is discharged, and the MLOFF signal is sent to all MLs to disable their current sources and to latch the match results. All MLs that have reached the threshold of the sense amplifier detect a match and all others sense amplifiers detect a mismatch. The timing of the MLOFF signal can also be adjusted by a variable delay element to ensure proper match operation.

This sensing scheme saves power in two ways. First, by cutting the current when DML reaches the threshold voltage of the ML sense amplifier, the voltage swing of all MLs is limited to 960 mV ($\sim VDD/2$). This reduces the ML power dissipation by a factor of two when compared to a full swing ML sensing scheme. Second, by precharging the MLs to ground (mismatch state), the SLs do not need to be reset between consecutive searches, hence minimizing SL switching activity and thereby reducing SL power dissipation by a factor of two.

Fig. 6 illustrates how the voltage on DML, ML with no miss (ML0), and ML with one miss (ML1) develop as a function of time for the maximum programmable current. With programmable delay set to 1 ns, the voltage on ML1 is at least 140 mV below the voltage of DML, and hence guaranteed to produce a mismatch. At the maximum current setting, the resulting match-time, defined as the time from when the search data is valid to when the match-result is latched, is 3 ns. A less-than-maximum programmable current can be used to increase the voltage difference between ML1 and DML at the cost of lowering search speed. At the minimum programmable

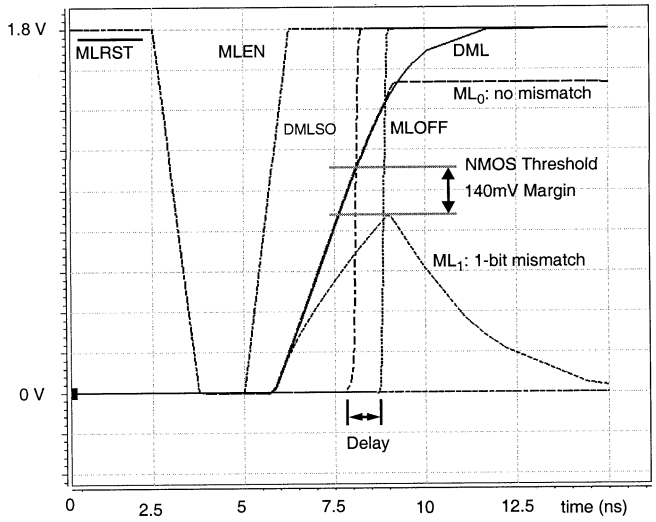


Fig. 6. ML timing: Voltage development of ML with a match (ML_0) against a ML with a one-bit miss (ML_1).

TABLE I
COMPARISON OF PROPOSED TCAM WITH PREVIOUS WORK

	This Work	Previous Work [4]
Process	0.18μm CMOS	0.18μm CMOS
Storage	Ternary	Binary
Cell Size	17.54μm^2	17.46μm^2
Match-Line Size	144 bits	20 bits
Match-Time	3ns @ 1.2V	2.7ns @ 1V

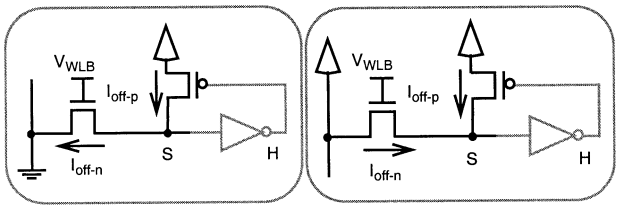
current, this voltage difference is simulated to be around 830 mV, while the match-time is increased to 11 ns.

IV. MEASUREMENT RESULTS

Extensive measurement at room temperature verifies successful storage in every cell across the memory array. The test chip successfully differentiates between an ML with a match and an ML with single or multiple mismatches. This includes a single mismatch in a cell farthest from the ML sense amplifier on a ML farthest from the DML.

The test chip features a range of programmable currents that can be used to measure the maximum match-time. At the maximum programmable current (260 μA), the test chip correctly detects a match in an estimated match-time of 3 ns at 1.2 V. This match-time could not be directly measured because our test equipment has a minimum time step of 6.25 ns. Instead, we have measured successfully a match-time of 6.25 ns that corresponds to 3/7 of the maximum programmable current, also at 1.2 V.

Comprehensive literature search has produced few published papers reporting CAMs of similar architecture for comparison purposes. Table I compares this work with a binary CAM used in a four-way, set-associative memory [4], also implemented in 0.18-μm CMOS technology. The proposed "ternary" CAM cell uses the same area as the previous "binary" CAM cell. Both



(a) (b)

Fig. 7. Sub-threshold currents (a) when BL is held low and (b) when BL is held high.

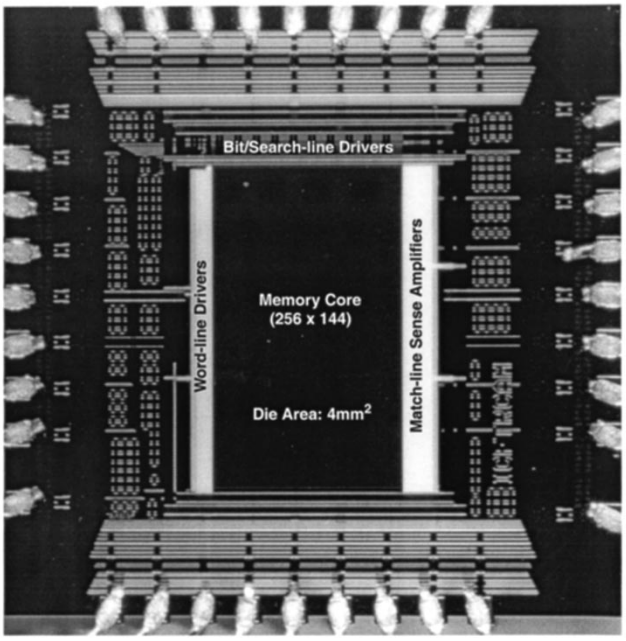


Fig. 8. Chip micrograph.

CAMs achieve a match-time of close to 3 ns. However, this work achieves this match-time for a 144-bit-long ML as opposed to only 20 bit-long ML of the previous work.

V. Discussion

A constraint with the use of the asymmetric 4T static storage cell is that the BL must not remain high for large periods of time. The BL is used to both write to a cell, and act as a ground to maintain stored data. Referring to Fig. 7(a), a "0" stored at node "S" relies on $I_{\mathrm{off}-n}$ to sink charge sourced by $I_{\mathrm{off}-p}$ through the grounded BLs. However, when writing a "1" to a neighboring cell in an adjacent row, the BL is high, reversing the direction of $I_{\mathrm{off}-n}$, hence helping $I_{\mathrm{off}-p}$ with charging up node "S" [see Fig. 7(b)], a phenomenon that could flip a stored "0" to a "1." Our measurement results indicate that a stored "0" will flip to a "1" if the BLs stay high for more than 10 μs. However, since $I_{\mathrm{off}-n} \gg I_{\mathrm{off}-p}$, a stored "0" is retained reliably if the BL is reset to ground at the end of every cycle.

In the test chip, SL and $\overline{\mathrm{SL}}$ control the bottom transistors while the stored data control the top transistors of the NAND compare circuits, as shown in Fig. 1(a). This choice results in a data-dependent ML capacitance, and hence a deviation from the data-independent capacitance of the DML. Nevertheless, the variable delay of the MLOFF signal in the ML sensing scheme allows correct matching for all ternary data patterns. To make the ML capacitance data-independent, similar to that of the DML, the search lines and the stored data must be swapped in their connections to the NAND compare circuits.

VI. Conclusion

The test chip shown in Fig. 8 demonstrates successful operation of the proposed TCAM cell and the ML sense amplifier. The TCAM cell achieves increased density in an all-logic 0.18-μm CMOS technology. The current-race sensing scheme achieves a 3-ns match-time and reduces power consumption by limiting ML voltage swing to 960 mV and by reducing the SL switching activity by 50%. The test chip remains fully functional at a minimum supply voltage of 1.2 V.

Acknowledgment

The authors would like to thank Mosaid Technologies Inc., Kanata, Canada, for contributions to this work: R. Mackenzie and D. Haerle for valuable suggestions and comments, D. Foss and A. Roth for the initial idea of the cell, P. Gillingham and D. Perry for the initial idea of match-line sensing scheme, and R. Colbeck and T. Wojcicki for their encouragement and support. They would also like to thank Canadian Microelectronics Corporation (CMC) for fabrication.

References

[1] M. Kobayashi, T. Murase, and A. Kuriyama, "A longest prefix match search engine for multi-gigabit IP processing," in *Proc. IEEE Int. Conf. Communications*, vol. 3, June 2000, pp. 1360–1364.

[2] V. Lines, A. Ahmed, P. Ma, S. Ma, R. McKenzie, H. Kim, and C. Mar, "66 MHz 2.3 M ternary dynamic content addressable memory," in *Proc. IEEE Int. Workshop Memory Technology Design and Testing*, Aug. 2000, pp. 101–105.

[3] K. Noda, K. Matsui, K. Tokashiki, K. Takeda, and N. Nakamura, "A loadless CMOS four-transistor SRAM cell in a 0.18-μm logic technology," *IEEE Trans. Electron. Devices*, vol. 48, pp. 2851–2855, Dec. 2001.

[4] P. Lin and J. Kuo, "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 J. Solid-State Circuits*, vol. 36, pp. 666–676, Apr. 2001.

[5] K. Takeda *et al.*, "A 16-Mb 400 MHz loadless CMOS four-transistor SRAM macro," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1631–1640, Nov. 2000.