

Power and Failure Analysis of CAM Cells Due to Process Variations

Mahmoud Bennaser and Csaba Andras Moritz
 Department of Electrical and Computer Engineering
 University of Massachusetts
 Amherst, MA 01003
 Email: {mbennase, andras}@ecs.umass.edu

Abstract—Process variations arise due to processing and masking limitations, and result in random or spatially varying deviation from the designed parameter values. Changes in these parameters cause electrical parameters to vary, such as effective channel length and threshold voltage. These mismatches modify the strength of individual devices resulting in various failures. In this paper, we present a failure analysis of CAM cells under process variation in 32-nm CMOS technology. We investigate the effects of variations in L_{eff} and V_t on the performance and power consumption of CAM cells. Finally, we discuss techniques at circuit and architecture levels to mitigate the effect of process variation on CAM access.

I. INTRODUCTION

As we move forward into the smaller geometry devices, it's likely that manufacturing process will be more difficult to control. The manufacturing process causes variations in many different parameters in the device. These variations increase as technology scales due to the difficulty of fabricating small structures consistently across a die or a wafer [2].

In this paper we evaluate the impact of process variations on CAM-tag cache; CAM-tag caches have been employed in many low-power processors. For example, the Intel XScale processor employs 64-way set-associative CAM tags [5] and other ARM processors [4] use highly associative (typically 32-way) CAM-RAM caches for their instruction and data caches.

The CAM cell design evaluated in this work is a ten-transistor cell as shown in Fig. 1. The cell incorporates an SRAM cell to store a bit of data and a dynamic XOR gate used for comparison. The match line is precharged high and conditionally discharged when there is no match. The CAM cell design uses separate bit line pairs to reduce the load capacitances on bit lines as well as to eliminate the cross-talk effect between long parallel wires.

In this paper, we present a failure analysis of CAM cells under process variations in 32-nm CMOS technology and explore the impact of process variations on power consumption. To our knowledge, this is the first work to consider the impact of process variation on the different failure mechanisms in CAM cells at 32-nm CMOS technology. In addition, we propose new circuits and architectures that take into consideration the effect of process variations.

The rest of the paper is organized as follows. In Section II, we analyze different failure mechanisms in CAM cell due to process variation. The results of the analysis are presented in

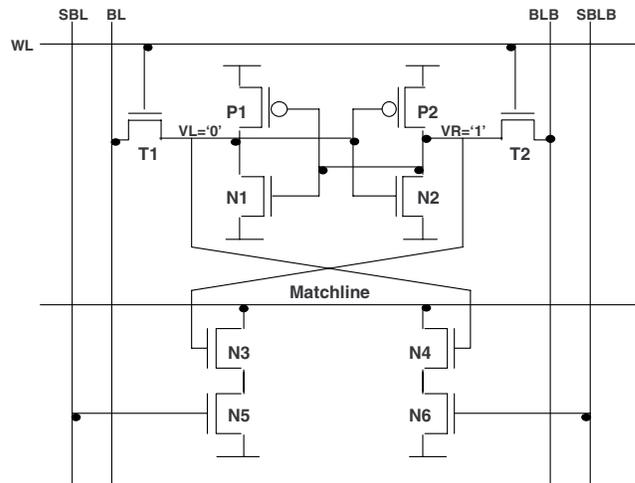


Fig. 1. Schematic of a CAM cell.

Section III. In Section IV, we discuss circuit and architecture techniques to mitigate the impact of process variations. Finally, concluding remarks are presented in Section V.

II. FAILURE ANALYSIS OF CAM CELL DUE TO PROCESS VARIATION IN A CACHE

Intra-die variations result in mismatches in device parameters (effective channel length (L_{eff}), and threshold voltage (V_t)). These mismatches modify the strength of individual devices resulting in various failures. Process variations in CAM cells may result in:

A. Search Time Failure

The CAM cell search time (T_{search}) is defined as the time required to produce a pre-specified voltage difference ($\Delta_{MIN}=0.1 \times V_{dd}$) between the Matchline and the reference source (V_{ref}) in the single ended sense amplifier (see Fig. 2). In case of mismatch, a decrease in the current (due to slow, i. e. high V_t , in the pull down transistors N3, N4, N5 and N6 in Fig. 1) that discharge the Matchline during the search operation will lead to less voltage difference between the Matchline and the reference source when the sense amplifier samples them, which may result in wrong evaluation. This failure is referred to as *search time failure*.

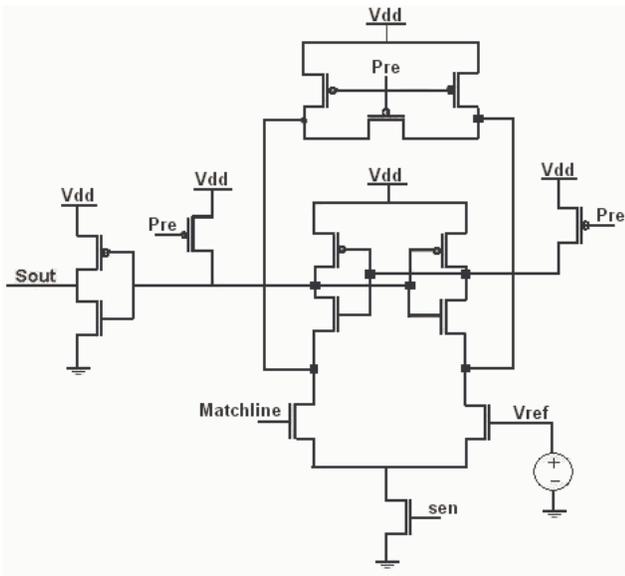


Fig. 2. Schematic of single-ended sense amplifier.

For example, in a CAM cell storing $VL="0"$, $VR="1"$, and search bitlines $SBL="1"$, $SBLB="0"$ (see Fig. 1), the Matchline will discharge through pull down transistors N3 and N5 during search operation. Any variation in L_{eff} or V_t of these transistors changes T_{search} and will cause a search time failure.

B. Match Failure

During a search operation, if a CAM cell is storing $VL="0"$, $VR="1"$, and search bitlines $SBL="0"$, $SBLB="1"$, then the Matchline will stay in logic "1". Any increase in the leakage current (due to low V_t or short L_{eff} , in the pull down transistors N4, and N5 in Fig. 1) will leads to voltage drop in the Matchline. This mismatch in transistors leads to voltage difference between the Matchline and the reference source when the sense amplifier samples them, which may result in wrong evaluation. We refer this failure as *match failure*.

C. Write Stability Failure

Unsuccessful write operation occurs due to the deviation of the strength of the access transistors and the trip point of the crossed inverters. This failure is referred to as *write failure* in SRAM cell [3] and the same thing is applied to CAM cell. For example, writing a "1" to a CAM cell storing "0", the node VR (see Fig. 1) get discharged through BLB to a low value determined by the voltage division between P2 and the access transistor T2. As shown in Fig. 3, if the node VR can not be reduced below the trip point of the inverter P1 and N1, when the wordline is high, then a *write failure* occurs.

III. CIRCUIT DESIGN AND RESULTS

In order to examine delay and power consumption under process variations, we choose to setup an experiment involving a tag array of 32 rows with each row of 23 CAM cells

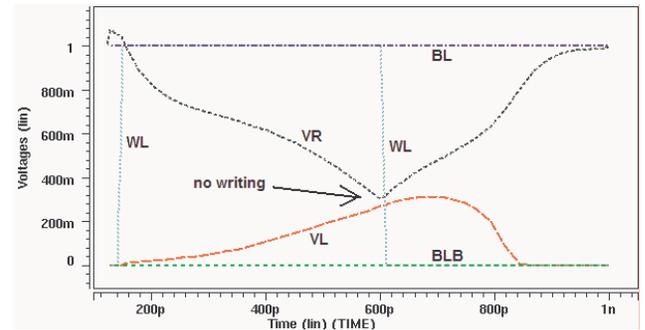


Fig. 3. Write stability failure.

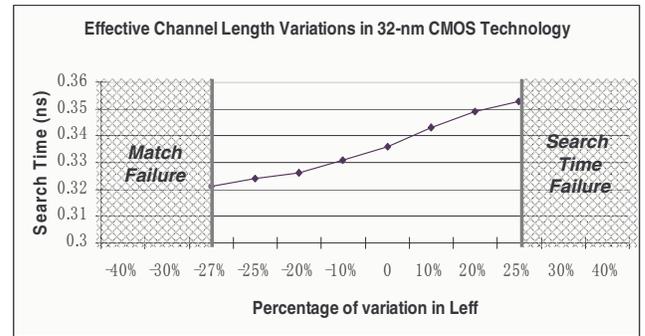


Fig. 4. Effect of L_{eff} variation on search time.

and a single-ended sense amplifier circuit (see Fig. 2). We use HSPICE circuit simulator in 32-nm PTM technology [1] with supply voltage of 1V. A nominal value of 0.2V for the threshold voltage and 13.7-nm for an effective channel length is used and given by the Predictive Technology Model.

A. Variations in Effective Channel Length L_{eff}

Variation in L_{eff} is caused by the lithographic process. These variations results in changes in device performance characteristics. A total of 40% variation in effective channel length (L_{eff}) is expected within a die [2]. We have found that the use of longer effective channel lengths (L_{eff}) tends to increase the search time as shown in Fig. 4 by as much as 7% and in worst-case leads to timing violations.

Our results show the variations in L_{eff} of CAM cells should be between the range of -27% and +25% from its nominal value so that our CAM cells work successfully. If the L_{eff} is out of this range of its nominal value, it will cause a failure. For example, based on our experiments, if the variation in L_{eff} is more than +25%, the CAM cell will fail due to *search time failure*. If the variation is less than -27% of its nominal value, then the CAM cell will fail due to *match failure*.

Fig. 5 shows the total power consumption for our CAM design. The power of the CAM array is the sum of the power in the match lines and search lines. The dynamic power is dominant because there is at most one match per access; all but one of the match lines discharge. A small variation in the L_{eff} value causes, however, a change in the leakage power

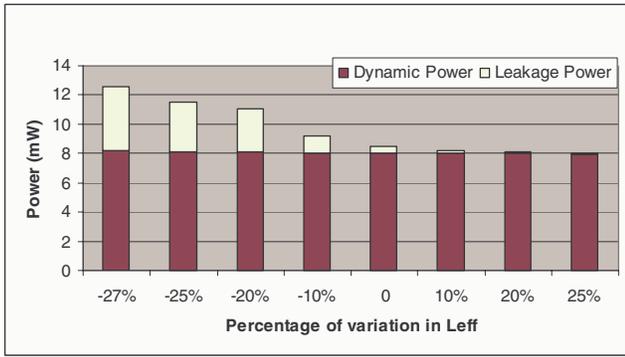


Fig. 5. Effect of L_{eff} variation on power consumption.

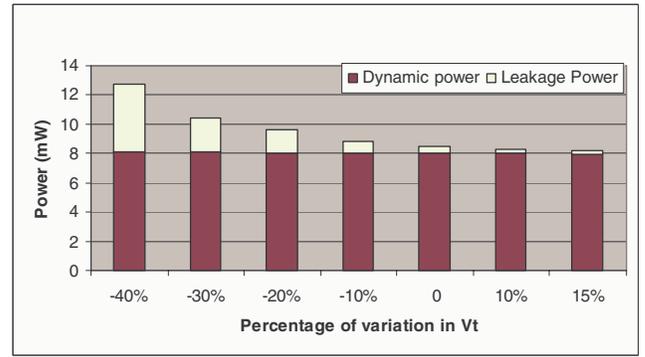


Fig. 7. Effect of V_t variation on power consumption.

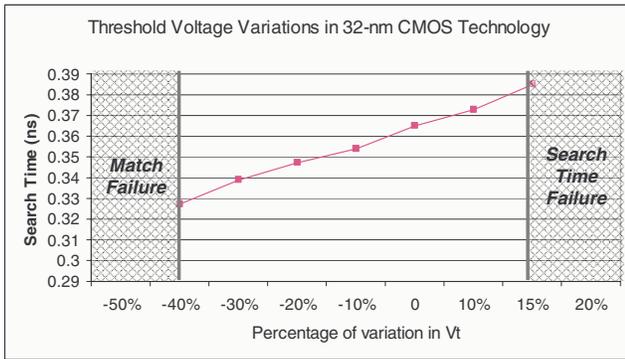


Fig. 6. Effect of V_t variation on search time.

by as much as 60X!

B. Variations in Threshold Voltage V_t

V_t can vary due to (1) changes in oxide thickness, (2) changes in the dopant levels in the substrate, polysilicon and implants, and (3) surface charge. Accurate control of V_t is very important for many performance and power optimizations and for correct execution.

Based on our simulations as show in Fig. 6, CAM cells can handle variations in V_t between -40% and +15% from its nominal value. For threshold voltage variation more than 15% of its nominal value, the CAM cell will fail due to *search time failure*. If the variations in V_t is less than 40% of its nominal value, the CAM cell will fail due to *match failure*.

The subthreshold leakage current depends exponentially on transistor threshold voltage. Thus, lower threshold voltages lead to increased subthreshold leakage current and increased leakage power. As we can see in Fig. 7, lower transistor threshold voltages, significantly impact the leakage power by such as 25X.

IV. CIRCUIT AND ARCHITECTURE TECHNIQUES TO MITIGATE FAILURES IN CAM CELLS

This section describes possible approaches at the circuit and architecture levels to mitigate the effects of process variations and avoid failures. We address each failure mode and show

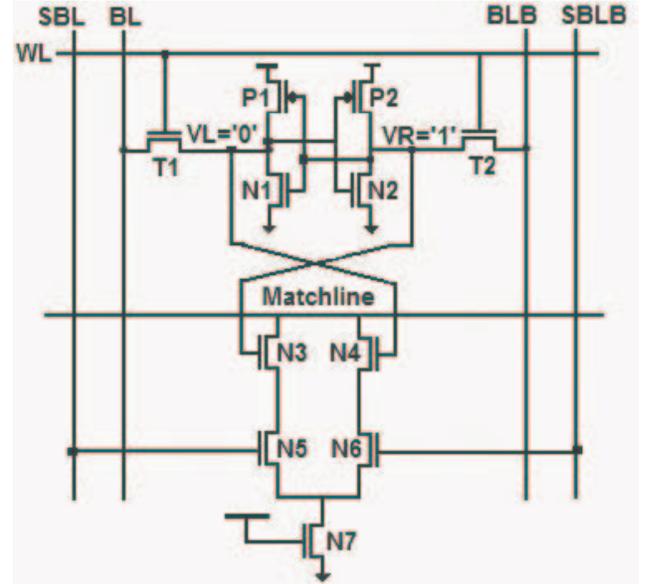


Fig. 8. Schematic of CAM cell circuit with extra NMOS transistor.

how a designer could combine simple modifications at the circuit level, detection techniques and testing techniques with minor architectural enhancements to make CAM designs more resilient against the failures presented in the earlier sections.

The main reason for the match failure is the leakage current from the precharged Matchline through the pull down transistors in the CAM cells during the search operation. In order to prevent the match failure, we propose to insert an extra resistance in the leakage path between the power supply and ground by adding an NMOS transistor (as shown in Fig. 8).

The extra transistor (N7) will reduce leakage due to stack effect but it will slightly increase the delay of the CAM cell. The area overhead of adding an extra transistor to each CAM cell can be reduced by sharing the extra transistor with all the CAM cells in each row.

To detect and mitigate the *search time* and *write stability* failures we can use double sensing technique [3] combined with an architectural add-on. The architectural add-on is a storage bit per each row called *Speed Storage*. We use testing

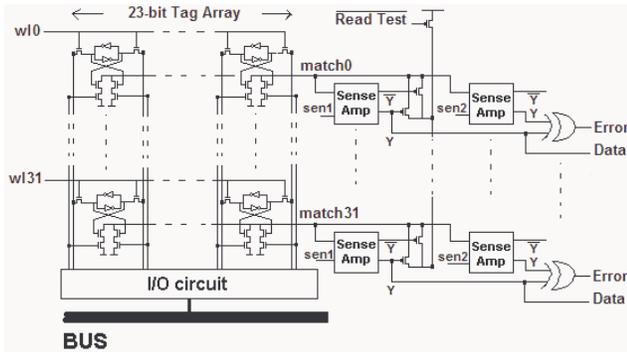


Fig. 9. Double sensing circuit for CAM tagged cache.

to set up the content of this storage and the double sensing technique to accommodate different access speeds during a CAM access.

However, to avoid *search time failure* we would need to be able to search at different speed. We use double sensing technique to accomplish that. The basic idea of double sensing is to have parallel single ended sense amplifiers to sample the Matchline twice during search cycle, as shown in our design in Fig. 9. This is needed to accommodate different speed across the CAM rows.

The first sensing is performed in the same way as the conventional sensing, while the second one is a delayed sensing. The added second sense amplifier has to be fired as late as possible during a search cycle. During the search operation in the test mode, if a cell is affected by *search time failure*, the cell starts discharging the Matchline slowly. If the CAM cell can establish enough voltage differentials with the reference source by the time the second single ended sense amplifier is fired, the *search time failure* is detected. Therefore, comparing the outputs of the first and second sense amplifiers with the XOR gate, a *search time failure* is detected when a mismatch between the two outputs occurs.

Each CAM-tagged row is tested using double sensing technique and March C- test when the test mode signal is on. A row is considered slow even if a single CAM cell in a row is detected with *search time failure*. This information will be stored in a new component added to the CAM called *Speed Storage* (explained shortly).

A key idea of the proposed CAM architecture (see Fig. 10) is to enable the sense amplifiers at different speed (normal or slow) based on the content of the speed storage for each row, to avoid failure. We need to delay firing the sense amplifier due to search time failure, such that we give enough time to the CAM cells to discharge the Matchline.

As mentioned, in the proposed CAM architecture we also add a *Speed Storage*. It stores the speed information and is read along with the CAM tag on every search operation. The *Speed Storage* is implemented as small memory of one bit per CAM tag row. Therefore, each bit in *Speed Storage* stores the speed information of all the CAM cells in a row. The bit is “1” if the corresponding row in the cache is slow;

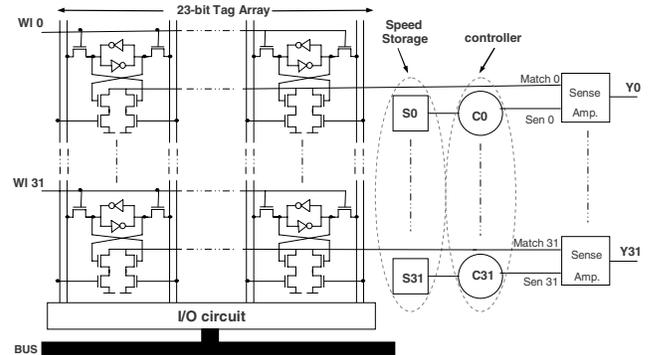


Fig. 10. The proposed CAM tagged architecture during the search operation.

otherwise, it is “0” for normal speed.

This bit is determined at the time of testing and stored by the double sensing circuit. *Speed Storage* provides the speed information to a controller unit per row; it is used by the controller to control speed of the sense amplifiers during regular search operations. We can build the *Speed Storage* using flip-flops that has a minimal read delay; we can build the controller using combinational circuits that will provide the enable signals to the sense amplifiers at different speeds.

Similarly, we can also use the double sensing technique at the bitlines; if the CAM cell is influenced by a *write stability failure*, both sense amplifier outputs are wrong (since the previous writing has not successfully written the correct value to the CAM cell). We can reduce the probability of write failure occurs by extending the time when the wordline (WL) is high during write operations.

V. CONCLUSION

Process variation introduces new delay and power tradeoffs that must be considered at circuit and architecture designs. In this paper, we presented failure and power analysis of CAM cells under process variation in 32-nm CMOS technology. We have shown that process variation can have a significant impact on delay, in worst-case leading to timing violations. In addition, power dissipation, especially leakage power has been shown to be affected by process variations by as such as 60X. The robustness of the CAM cells under variations in L_{eff} and V_t was discussed. We proposed a new circuit and architecture to mitigate the impact of process variations. In our future work we plan to implement these designs in a low power cache design.

REFERENCES

- [1] Predictive technology model. Nanoscale Integration and Modeling Group at ASU. Online available: <http://www.eas.asu.edu/ptm/>.
- [2] A. Chandrakasan, W. Bowhill, and F. Fox. *Design of High Performance Microprocessor circuits*. IEEE Press, 2001.
- [3] Q. Chen, H. Mahmoodi, S. Bhunia, and K. Roy. Modeling and testing of sram for new failure mechanisms due to process variations in nanoscale cmos. In *VLSI Test Symposium (VTS)*, May 2005.
- [4] S. Furber. *Arm system-on-chip architecture*. Addison-Wesley, 2000.
- [5] M. Zhang and K. Asanovic. Highly-associative caches for low-power processors. In *Kool Chips Workshop, 33rd International Symposium on Microarchitecture*, December 2000.