On Estimation of Static Power-Peformance in TCAM

Dhireesha Kudithipudi
Dept. of Computer Engineering
Rochester Institute of Technology
Rochester, USA
dxkeec@rit.edu

Eugene John
Dept. of Electrical and Computer Engineering
University of Texas – San Antonio
San Antonio, USA
Eugene.john@utsa.edu

***Abstract*—High demand on network security, and Quality of Service (QoS) along with increasing line rates made Ternary Content Addressable memories more attractive for fast searching in network routers switches for forwarding and filtering IP (Internet Protocol) signals. While high performance through parallel search is achieved in the TCAM circuits, the power dissipated in these systems is high compared to other storage devices (eg. Binary CAM). Past research on TCAM cells, mainly concentrated on optimizing dynamic power by reducing the switching activity or matchline swing. At present the power dissipation in TCAM cells is exacerbated with rapid scaling of CMOS technology, where subthreshold leakage is becoming dominant in the total power. In this research, we estimate the TCAM static power at 65nm and 45nm and propose solutions to optimize static power. A case study of four most popularly used TCAM cells is also presented with a discussion on relative performance impact.**

I. INTRODUCTION

A CAM (***C***ontent ***A***ddressable ***M***emory) is a special type of storage device. Unlike in traditional storage devices (SRAM's), here the operating system provides data and the CAM returns a list of addresses where the data is stored [8, 9]. It searches the entire memory in one operation and is much faster than the SRAM cells. Ternary CAM (TCAM) has one additional state than the Binary CAM (1, 0, and X), where Don't-Care "X" is used for masking the bits. TCAM's are commonly used in network routers (e.g.: Intel IXP cards) and switches for forwarding and filtering IP (Internet Protocol) signals. A TCAM cell performs searches for the destination address of an incoming IP packet with all the prefixes in parallel. Several prefixes can match the destination address. Priority encoder logic then selects the longest matching prefix [8]. Each of these high-density TCAM chips consumes around 12 to 15 W of power, when all the entries are enabled for search [8,9]. Previous research on TCAM mainly concentrates on alleviating the dynamic power and thereby the access frequency (switching activity power). Panigrahy et al propose the concept of paged TCAM, in which the prefixes in the routing table are partitioned into smaller sets to reduce the power consumption]. Zukowski et al, identify a potential solution to reduce the switching activity power by reducing the number of transitions in the match line. Some research was on two stage lookup process for power efficient operation of TCAM cell. In [11], Mohan et al proposed a static power reduction technique, which lowers the supply voltage of the storage portion of the TCAM. However, there is no comprehensive study of leakage power characteristics of a TCAM cell available in the literature.

A conventional TCAM cell is shown in Fig.1, in which each entry of the array relates to the content of the cells in a particular row. Every row has a corresponding match line, which acts as a status indicator. A ternary bit stored in the TCAM cell is in three different states, 0, 1, or "X" and are encoded. The TCAM cell stores these encoded bits in two of its 6-transistor SRAM cells, present on either side of the TCAM cell. In order to store the extra state ("X") in a TCAM, a supplemental SRAM cell is used. The resultant data values in these two SRAM cells d and $\overline{d}$ are independent of each other. However, they are not mutually exclusive bits. In order for the query to match a value of '1', we set d = 1 and $\overline{d}$ = 0 and to match a value of '0', we set d = 0 and $\overline{d}$ = 1. To store an "X", we set both d = 0 and $\overline{d}$ = 0, which forces the cell to match irrespective of the input data (for a NAND gate). The TCAM cell asserts the query values (and its complement) through the Search Lines (SL and $\overline{SL}$**).** If there is a mismatch with the query, the Match Line (ML) is asserted, which creates a path from supply rail to the ground [1]. Depending on a match or a mismatch, the Match Line (ML) of the corresponding row is asserted or de-asserted respectively. The comparison between input address bits and the TCAM array contents is performed using NAND or NOR logic. Prior to every match comparison operation, the ML is precharged to a logic high and remains at logic high or discharges to logic low depending on the output of the logic gates. Finally, in the SRAM cell data is stored or retrieved using the bit lines (BL) which are in turn activated by the Word Line (WL).

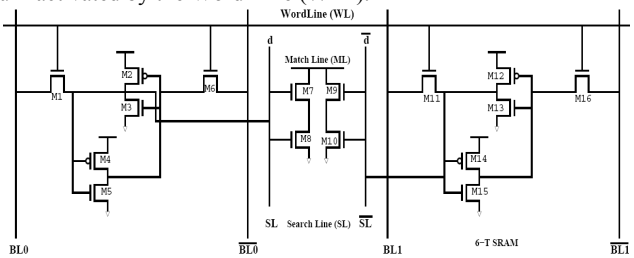

Figure 1. Conventional TCAM cell with two 6 transistor SRAM memory devices and a NOR gate

Many variations have been proposed to this conventional TCAM cell. A TCAM built using an asymmetrical 4 transistor SRAM

storage cell is proposed by Arsovski et.al. [5]. The asymmetrical SRAM cell holds logic '1' using the PMOS transistors. A logic '0' is sustained by the subthreshold leakage current of the NMOS transistors which is ideally larger than that of the PMOS transistors. The compare logic circuitry used is same as the conventional model.

In [6], Noda et.al, proposed a loadless four transistor SRAM cell, which can also be incorporated in the TCAM module. In the loadless SRAM cell the access transistors are PMOS and the drive transistors used are NMOS. In a standby mode of operation, the bitlines are precharged to the supply rail voltage and the PMOS transistors act as load elements. As opposed to the conventional 4 transistor SRAM circuit with load resistors, the high-node level rises to full supply voltage swing right after a Read or a Write operation. To restore the data in the SRAM cell, the off state current of the PMOS should be less than that of the NMOS.

In [7], Choi et. al proposed hybrid type TCAM architecture,. exploiting the benefits of both the NAND and the NOR gates. An additional mask bit at node 'M' will modify the NAND gate to a ternary storage. Above discussed models are commonly used in TCAM designs. Since most of the previous study for these cells is concentrated on dynamic power consumption, we solely surveyed on the effects of static power consumption in each of these cells.

The rest of the paper is organized as follows: Section 2 and 3 describe the static power dissipation in TCAM cells; Section 4 presents experimental results on TCAM cells total leakage for four different implementations and the delay metrics for the TCAM cells and conclusions are presented in Section 5.

II. LEAKAGE POWER ESTIMATION

Subthreshold current, prominent at sub-65nm, is the current flowing from the drain to source of a MOS transistor for a PMOSFET $V_{gs} > V_{sat}$ and for an NMOSFET $V_{gs} < V_{sat}$. When $V_{gs} = 0V$, we have the sub-threshold current (I_{sub}) given by the following equation [].

$$I_{sub} = \mu_0 C_{ox} \frac{W}{L}(n-1)(v_T)^2 e^{\frac{Vgs-Vth}{nVt}}\left(1 - e^{\frac{-Vds}{v_T}}\right) \quad (1)$$

where μ_0 is the zero bias mobility, n is the body effect coefficient, and v_T is the thermal voltage. The gate oxide leakage is measured as a combination of gate to source and gate to drain currents. The flow of both these currents for different modes of operation is discussed below: In a WRITE operation the TCAM cell loads the data on to either of the bitline pairs ($\overline{BL0}, BL0, BL1, \overline{BL1}$). To write the data into the SRAM cells, the Word Line (WL) is asserted high which also turns the access transistors "ON". Once the data is written to the SRAM cells the Word Line is de-asserted. During this stage there will be a subthreshold leakage path from the WL. In a READ operation, the contents from the SRAM cells are read on to the bitline pairs, by asserting the WL to logic high. In a SEARCH operation, the search data is retrieved by the Search Line (SL) pair when the pre-charged Match Line (ML) is set to logic high. If there is no match with the stored data contents, then the ML is discharged to ground. However, if there is a match, then the ML remains in its original state. This charging and discharging operation in the ML's consume maximum switching activity power in a TCAM cell. If the ML is discharged to ground and in off state the leakage current is high.

If both the SRAM cells store a logic zero or logic high (depending on NAND or NOR compare circuitry), then a don't care condition occurs and the ML status remains unchanged. The total static power dissipation in a TCAM cell is sum of the leakage occurring from the Match Line and Search Line to the ground, static power dissipated through the NAND/NOR compare circuitry and the leakage current though the SRAM cells. This can adhoc expression representation is

$$I_{TCAM} = ML(Isub + Ig) + NAND/NOR(Ig + Isub) + BL(Isub + Ig) + (SL)(Isub) \quad (2)$$

where ML is the match line, $Isub$ is the subthreshold leakage current, Ig is the gate leakage current, SL is the search line, BL is the bit line and the compare circuitry can be either $NAND$ or NOR. Based on this equation, we estimated the leakage current through each of the individual blocks and summed up the total leakage current in a TCAM cell.

III. STATIC POWER DISSIPATION IN TCAM

The Match Line (ML) is built using a connection of NAND or NOR cells.

A. NAND/NOR Leakage Power

In a two input NAND circuit the leakage power is high when both the inputs are at logic high and less when both the inputs are at logic zero. The series connected NMOS transistors which make the pull down path are switched off and form a transistor stack. For a two input NOR gate the leakage current is the same for 01, 10, and 11 combinations and is high for 00 combination. The simulation results for three different input combinations of transistors are shown in Table 2. A NAND gate consumes slightly higher leakage current than it's NOR counterpart. This can be attributed to the stacking effect. The subthreshold leakage current is reduced when the transistors are stacked as compared to a single transistor. Also, a NOR circuit is much faster than its

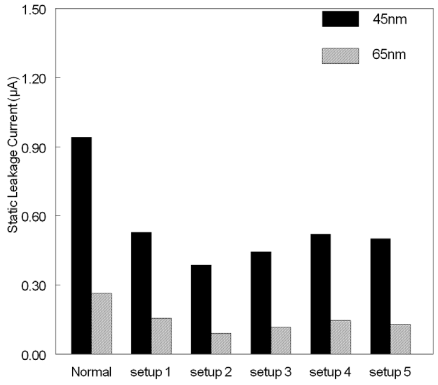

Figure 6. Static Leakage Current through the TCAM

NAND counterpart. Because, when there is a miss in a NOR Match Line (ML), the critical path of the cell will be through the series connected transistors. In a NAND cell the critical path will be through 8 transistors. This makes a NOR gate better choice for compare circuitry in a TCAM.

B. Total Leakage in TCAM

In this section we estimate the total leakage power through the conventional TCAM cell (Fig 1). To obtain savings in leakage power, we used combination of cell.high and low threshold voltage

devices for a 45 nm technology size. The high Vt devices yield slower logic but present lower leakage. Low Vt devices yield faster logic but present high leakage.

The static leakage current for different setups is shown in Fig 6, where the device settings for each setup are shown below.

- Normal:- No special devices
- Setup 1:- Transistors M2, M4, M12, M14- high V_t devices
- Setup 2:- All high V_t devices
- Setup 3:- Transistors M2, M4, M12, M14, M9, M10, M7, M8- high V_t devices
- Setup 4:- Access Transistors M1, M6, M11, M16- high V_t devices
- Setup 5:- Transistors M2, M4, M12, M14, M7, M9- high V_t devices

The leakage current is high when there are no special devices used in the circuit. Using all high threshold voltage devices reduces the leakage power of the TCAM cell, because the high- V_t devices increase the on-current and also reduce the over drive voltage of the ML, SL and the bitlines (BL). However, having only access transistors with high threshold voltage also shows considerable savings in the leakage power as compared to the normal setup. For all the setups, the leakage current at 45 nm is significantly higher than at 65 nm technology size, due to reduced threshold voltages (at 45 nm). As compared to employing all high threshold voltage devices, the dual threshold voltage TCAM cells will have worse stability. This is because the access transistors which are low V_t have increased read current, making them stronger relative to the pull down NMOS transistors (present in both the SRAM cells) of the TCAM. The regressive response of leakage current at smaller geometries is clearly represented in Fig.7. The percentage of static current with respect to the overall wasted current for 65 nm went to a maximum of 60%, whereas for the 45 nm it went up to 72%. Henceforth, exerting multiple V_t devices at smaller geometries shows further savings in leakage current. Incorporating combinations of these high V_t and low V_t devices in the TCAM reduced the static current and the associated wasted leakage current by ~20% in some of the cases. The average block power consumption for each case is shown in Fig 8.This power constitutes of the wasted current during active and inactive period of the TCAM. The difference in total power between the setups is not as apparent as the static leakage current, due to the dominance of switching activity power. At 65 nm the static leakage current

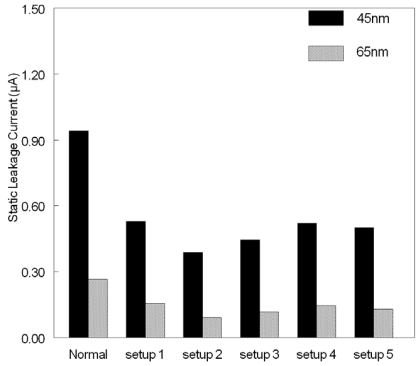

Figure 7. Percentage of Static Wasted Current with respect to the Total Wasted Current in a TCAM cel

contributes to ~45 % of the total power and the switching activity power contributes the rest. Since the supply voltage is higher at 65 nm by 0.1V, (and is quadratically proportional to the dynamic power) the switching power increases by a factor of ~23% compared to 45 nm.

C. Comparison of TCAM Models

To compare the leakage power consumed in each of these TCAM modules we performed analysis using three different setups along with the normal cell. The setup descriptions are as follows

- Case 1- Access Transistors – High V_t
- Case 2- Access + PMOS Transistors – High V_t
- Case 3- Compare Circuitry+ Access Transistors– High V_t

For both 65 nm and 45 nm, shown in Fig 9, the TCAM cell containing the asymmetric 4 transistor SRAM cell dissipated more leakage power. This is because of the fact that the bias voltages of the word lines are higher than the nominal operating range for this circuit (< 0.2V). The hybrid TCAM cell displayed the minimal leakage current than all the other models. In spite of its high transistor count, the conventional TCAM cell has savings only second to the hybrid TCAM circuit. Savings in leakage current are obtained by using high V_t access transistors for all the TCAM modules. This is consistent over both 45 nm and 65 nm technology nodes.

D. Performance of TCAM Models

The delay performance measure of the TCAM cells is a measure of the Match Line delay and Search Line delay. The Match Line delay is defined as the time to precharge the Match Line (10% to 90% rise time of the Match Line). The time to evaluate the Match Line depends on the time to charge the ML

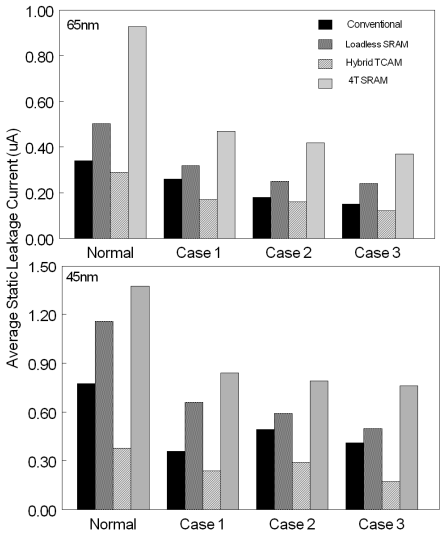

Figure 8. Average leakage current in four different TCAM cells

capacitance and the ML pulldown resistance. The delay for Match Line evaluate phase is defined as the time for the Match Line to fall to 50% of the precharge voltage.

TABLE I. MATCHLINE DELAY FOR CONVENTIONAL TCAM AT 45 NM

Cell Type	*Delays*	
	Precharge Delay (ns)	***Evaluate Delay(ns)***
TCAM Cell	2.84	1.67
Setup1	3.21	1.91
Setup2	3.46	2.10
Setup3	3.10	2.01
Setup4	3.08	1.96
Setup5	3.12	1.98

TABLE II. MATCHLINE DELAYS FOR FOUR DIFFERENT TCAM AT 45 NM

Cell Type	*Delays*	
	Precharge Delay (ns)	***Evaluate Delay(ns)***
Conventional TCAM Cell	2.84	1.67
Loadless SRAM	2.55	1.34
Hybrid TCAM	2.23	1.42
4T SRAM	2.29	1.51

The Match Line delays for the conventional TCAM cell and the leakage optimized conventional TCAM cells are represented in Table 1. The maximum delay for the precharge and evaluate phase occurs when all high threshold voltage devices are used in the cell. The delay penalty is minimum when only access transistors are used as high threshold voltage devices. Since speed is a very important design criterion in TCAM cells, the best option would be to use the Setup 4, where only access transistors are high V_{th} devices. Compared to the leakage power savings obtained in the conventional TCAM cell the delay measurements are relatively high. Therefore these reduction techniques should be applied very cautiously. The delay measurements for different models of TCAM cells are shown in Table 2. The 4T SRAM based TCAM cell shows the best Match Line precharge delay while the conventional TCAM cell has the worst Match Line delay. The inherent architecture of the cells translates to these delays. Using all high V_t devices is non-intuitive and affects the delay significantly and therefore will not be a design choice for TCAM cells. Using high V_t devices for the access transistors will be a good choice to reduce static power with only 3.5% degradation in performance.

IV. CONCLUSIONS

In this paper, we estimated the static leakage current of a TCAM cell by a methodical approach. The gate oxide and subthreshold leakage current are estimated for the TCAM by categorizing in to Bit Line (BL) leakage, Search Line (SL) leakage, Match Line (ML) leakage, and the compare circuitry leakage. For our simulations, we extracted the parasitics from the layouts for each of the TCAM models and then evaluated the leakage current dissipated for 65 nm and 45 nm technology sizes. The TCAM cells using wired-NOR match line circuitry performed better than the wired-NAND compare circuitry. The savings increase further for higher bit lengths. The NOR Match Line circuit performs better in terms of speed too. The percentage of wasted static current is eminent at 65 nm as compared to 45 nm by almost ~20% in some TCAM designs. Using multiple V_t devices alleviates the wasted leakage current in the TCAM cell in contrary to using conventional TCAM setup. Incorporating high threshold voltage access devices dissipated an equal measure of leakage current as compared to using high threshold voltage devices for both the access and PMOS transistors. An asymmetrical SRAM storage cell expended high leakage current in the four designs explored.

REFERENCES

[1] V. De, and S. Borkar, "Technology and design challenges for low power and high performance," International Symposium on Low Power Electronics and Design, 1999, pp. 163-168

[2] D. Lee, D. Blaauw, and D. Sylvester, "Gate Oxide Leakage Current Analysis and Reduction for VLSI Circuits," IEEE Trans. on VLSI Systems, Feb, 2004, pp. 155-166.

[3] Y. Ye, S.Borkar, and V.De, "A new technique for standby leakage reduction in high-performance circuits," Symposium on VLSI Circuits, June, 1998.

[4] A.Agarwal, S. Mukhopadhyay, A. Raychowdhury, C. H. Kim,, and K.Roy, "Leakage Power Analysis and Reduction for Nanaoscale Circuits," IEEE Micro, pp. 68-80,March 2006.

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

[6] K. Noda, K. Matsui, K. Takeda, N. Nakamura, "A loadless CMOS four-transistor SRAM cell in a 0.18-μm logic technology", IEEE Tran. on Electron Devices,2001, vol. 48, pp. 2851 – 2855.

[7] S. Choi, K. Sohn, and H.J Yoo, "A 0.7-fJ/bit/search 2.2-ns search time hybrid-type TCAM architecture" IEEE J. of Solid-State Circuits, 2005, vol. 40, pp. 254 – 260.

[8] K. Pagiamtzis, and A. Sheikholeslami, "Content-addressable memory (CAM) circuits and architectures: a tutorial and survey" IEEE J. of Solid-State Circuits,2006, vol. 41, pp. 712 – 727.

[9] W. Wu, J. Shi, L. Zuo, and B. Shi, "Power-efficient TCAMs for bursty access patterns", IEEE Micro ,2005, vol. 25, pp. 64 – 72.

[10] B. Agrawal, T. Sherwood, "Modeling TCAM power for next generation network devices" Proceeding of the IEEE Symposium on Performance Analysis of Systems and Software,2006, pp. 120 – 129.

[11] N. Mohan, and M. Sachdev, "A static power reduction technique for ternary content addressable memories" Proceedings of the Canadian Conference on Electrical and Computer Engineering (2004), pp. 711 – 714.