

Design Considerations and Benefits of Three-Dimensional Ternary Content Addressable Memory

Eun Chu Oh and Paul D. Franzon
 ECE Dept., North Carolina State University
 2410 Campus Shore Drive, Raleigh, NC, USA 27606

Abstract-Three dimensional (3D) ternary content addressable memory (TCAM) has been designed in a 0.18 μm fully depleted silicon on insulator (FD SOI) 3D IC process. This paper demonstrates that a 3D TCAM with three tiers can achieve 40% matchline capacitance reduction and 21% power reduction compared to a TCAM in a conventional single-tier process. This paper also discusses design considerations of 3D TCAM including partitioning methods for multiple tiers and layout methods of interconnects.

I. INTRODUCTION

As device feature size scales down, on-chip metal interconnects become longer and more closely packed with high parasitic capacitances, emerging as a significant limiting factor to the power consumption and performance of the chip [1]. This paper proposes three-dimensional integrated circuit (3D IC) design to improve power consumption by reducing the parasitic capacitance of interconnects. We have designed a ternary content addressable memory (TCAM) in 3D IC to demonstrate the benefits of 3D IC.

MIT Lincoln Laboratory offers a 3D IC technology with three tiers of 0.18 μm fully depleted silicon on insulator (FD SOI) process, as shown in Fig. 1. In this process, the base tier (Tier 1) is on a FD SOI substrate while Tier 2 and 3 are flipped and stacked on the top of the base tier with the substrate removed. The 3D vertical inter-tier vias, formed by tungsten deposition, provide a direct interconnection between wafer tiers replacing highly capacitive interconnects.

Despite of its advantages, the thermal issue in 3D IC is a challenging problem and on-going research topic. The upper tiers do not have easy access to heat conduction paths resulting in a tight thermal budget. In general, high power consuming circuits are placed on the first tier, close to a heat sink while less power consuming circuits are placed on the top tier. Thick back-metal layers on upper tiers also help to extract the heat generated on those tiers.

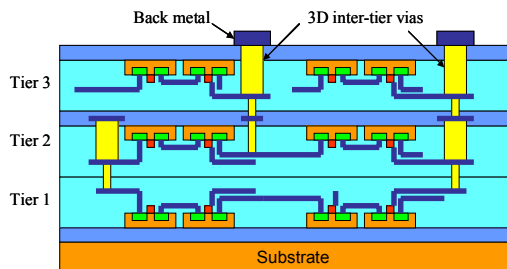


Fig. 1. MITLL FD SOI CMOS 3D IC process [2].

II. TERNARY CONTENT ADDRESSABLE MEMORY

Ternary CAM was chosen for a 3D IC design because of its regular structure as well as its characteristics of fast access and high power consumption. TCAMs are widely used in applications such as networking and processor caches for lookup tables. The write operation in TCAM is done in a similar way as in SRAM. In the search operation, TCAM tries to find a match with the comparand data by comparing it with all the stored words simultaneously.

A general TCAM architecture consists of a memory core, an address decoder, a comparand data driver, a bitline data driver, a searchline pre-discharger, a matchline precharger, a ML sense amplifier, and a priority logic encoder, as shown in Fig. 2. TCAM cells are composed of two storage circuit parts and a comparison circuit, used in the search operation as shown in Fig. 3 (a). Due to the high power consumption of the horizontally running matchlines during the search operation, low power TCAM still remains a challenging part of TCAM design [3][4]. The matchlines (MLs) have high parasitic capacitance as each ML is shared by every memory cell in a word as shown in Fig. 3 (b). Also, most MLs toggle every search cycle as MLs gets precharged every search cycle and discharged if a mismatch is found, which is the case for most words. By reducing the ML capacitance using 3D vias, low power TCAM can be achieved.

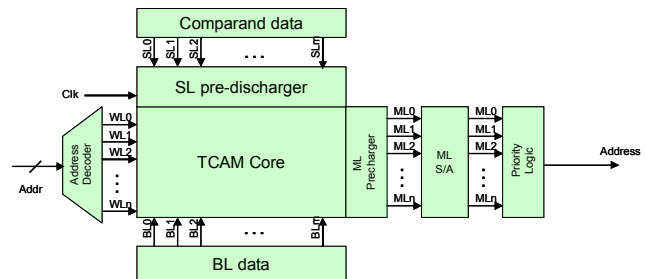


Fig. 2. TCAM Architecture

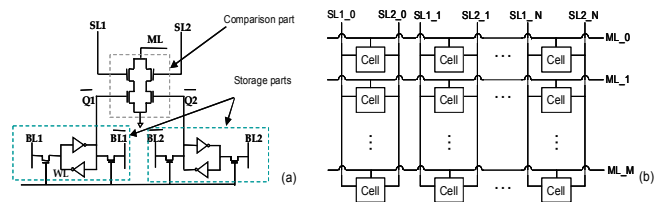


Fig. 3. (a) TCAM cell, (b) TCAM memory core

III. 3D DESIGN METHODOLOGY

A. Partitioning Methodology of Circuits into Multiple Tiers

The first stage to be considered in 3D circuit designs is how to partition the circuits into a number of design parts to be allocated on different tiers. The degree of advantage of 3D design greatly depends on the partitioning method. There are three ways to partition a system composed of arrays – 1) intra-cell, 2) inter-cell, and 3) inter-block partitioning.

Intra-cell partitioning is breaking a small basic building block or a cell into N parts to be distributed on N tiers. In the case of the MITLL 3D process with three tiers, each TCAM cell can be broken into three parts: e.g. a storage part, a comparison part and a second storage part. This would reduce the width of the cell by one third, from 20 μm to 7 μm , shortening the length of metal by one third when the dimension of 3D inter-tier via is ignored. However, a 3D via occupies as much as 3.5 μm (l) x 3.5 μm (w) x 8 μm (h) including the metal wire required for 3D via alignment making the reduction of ML less significant. Due to the large size of a 3D via compared to the cell size, the intra-cell partitioning method provides little benefit to TCAM arrays and is more suitable for designs with larger cells or a process with smaller 3D vias.

In inter-cell partitioning, the system is divided into N sections comprising multiple cells or small basic building blocks. More 3D vias can replace shorter interconnects compared to inter-block partitioning. Inter-cell partitioning is appropriate for a system with a regular structure such as a memory system where multiple short interconnects play a critical role in terms of power and performance. Inter-cell partitioning was selected for our 3D TCAM design.

In inter-block partitioning, the entire system is broken down into three parts where each part is composed of large functioning blocks. Using this partitioning method, less 3D vias replaces longer wires that interconnect two blocks at a distance. The benefit is maximal if the 3D vias replace critical paths or the most power consuming path and the benefit is minimal if the 3D vias replace non-critical paths or if the system is already optimized for interconnect length.

B. Matchline Layout Methodology

In 3D IC design, the parts of the interconnect actually replaced by 3D vias and the placement of the vias can be an important factor in terms of the interconnect capacitance. In this paper, various layout methods for matchlines, shown in Fig. 4, have been studied and the matchline capacitances (C_{ML} s) are evaluated for each case.

The simplest way to convert a ML in 2D design into a 3D design with three tiers is to partition the ML interconnect into three parts and interconnect them with a single set of 3D vias, as shown in Fig. 4 (b). However, this results in almost no improvement on the C_{ML} as the matchline is not reduced in size. Another way to connect the matchline of each cell in a 3D design is to interconnect the ML on every cell on the third tier with a metal wire and interconnect them to the underlying cells through the 3D vias, shown in Fig. 4 (c). This results in more than a 30% reduction in C_{ML} , according to the Ansoft

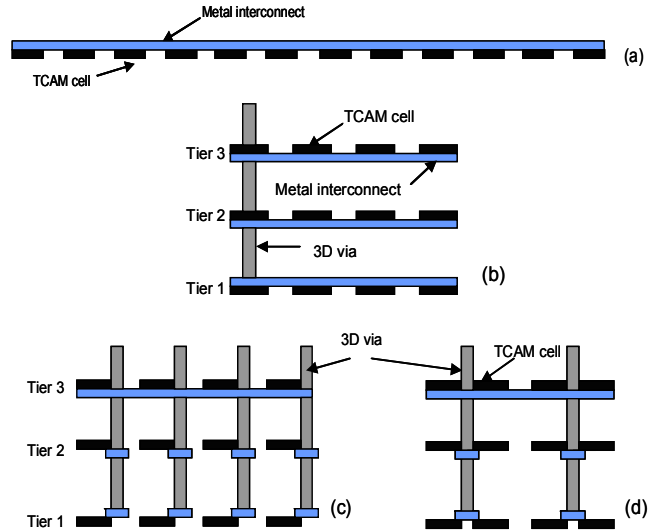


Fig. 4. Cross-sectional views of matchline layout methods for 12 bit word in (a) Conventional 2D with single tier, (b) 3-tier 3D design with 1 set of 3D vias per word, (c) 3-tier 3D design with 1 set of 3D vias per 3 cells, and (d) 3-tier 3D design with 1 set of 3D vias per 6 cells.

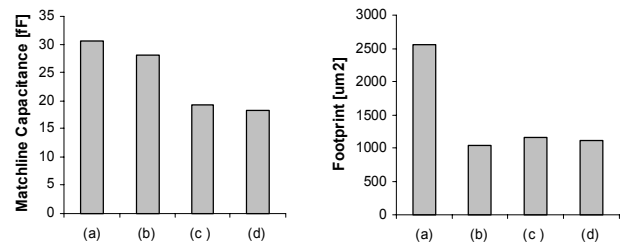


Fig. 5. Matchline capacitance and footprint with various matchline layout methods for 12 bit word (a) Conventional 2D with single tier, (b) 3-tier 3D design with 1 set of 3D vias per word, (c) 3-tier 3D design with 1 set of 3D vias per 3 cells, and (d) 3-tier 3D design with 1 set of 3D vias per 6 cells.

Q3D field software simulations, compared to the 2D design. By having the 3D vias shared by two adjacent cells, which reduces the 3D via need by a half, as shown in Fig. 4 (d), C_{ML} is further reduced to achieve 40% improvement compared to the 2D design. This ML layout method was selected for our 3D TCAM design with 3 tiers. More detailed comparison on the four designs is shown Fig. 5.

C. Cell Layout Methodology

The actual circuit layout is another factor that can influence the efficiency and benefits of 3D IC design. In the case of the TCAM cell, typical layout has two storage circuits and the comparison circuits in between them, which requires an extra ML interconnect to connect the comparison logic and 3D vias, as shown in Fig. 6 and Fig. 7. By simply rearranging the layout to have the compare logic located next to two storage cells for more efficient sharing of 3D vias, C_{ML} is reduced by 20%. This layout was selected for our 3D TCAM design. The resulting C_{ML} s from Q3D field simulations are shown in TABLE I.

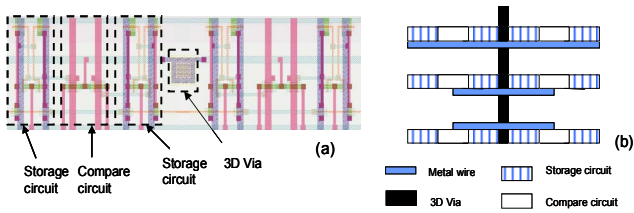


Fig. 6. Conventional layout of a 1w x 6b TCAM memory core in 3D IC (a) cell layout and (b) cross sectional view.

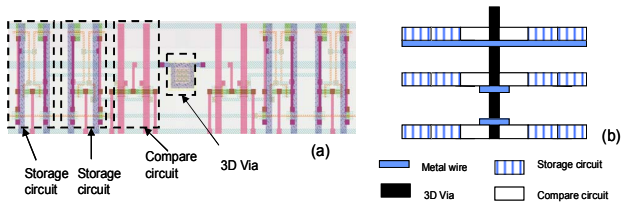


Fig. 7. Improved Layout of a 1w x 6b TCAM memory core in 3D IC with reduced ML interconnect length (a) cell layout and (b) cross sectional view.

TABLE I
COMPARISON OF VARIOUS TCAM CELL LAYOUTS

	Conventional Layout	Improved Layout
C _{ML} [fF]	23.8	18.3
Footprint [μm^2]	1109	1109

IV. SIMULATIONS AND RESULTS

A. Q3D simulations for Capacitances

For accuracy of their analysis, all parasitic interconnect capacitances are extracted from simulations by Ansoft Q3D Extractor, a quasi-static electromagnetic field simulation tool that computes capacitance using the method of moments and the finite element method [5]. For a fair comparison, TCAM memory in both conventional 2D IC structures and 3D IC structures with identical cell layouts are simulated, keeping in mind that 3D design allows more degrees of freedom for spacing between interconnects due to the area saved by going vertical. In fact, even with the sparsely packed interconnects in 3D design, the footprint area of the TCAM core in 3D IC using three tiers is about a half of that in 2D IC with a single tier. It is not one third due to the area occupied by 3D vias themselves.

A 1w x 12b TCAM array has been designed and laid out for single-tier 2D, 2-tier 3D, and 3-tier 3D structures in 0.18 μm MITLL FD SOI 3D process. TCAM arrays in the sizes of 1w x 6b and 1w x 24b have also been designed for all three structures mentioned above. For all the designs, the interconnect capacitances were extracted using Q3D Extractor. The results show approximately 28% and 40% matchline capacitance reduction for 2-tier and 3-tier 3D TCAM array designs, respectively, compared to the single-tier TCAM, as shown in Fig. 8 (a). This reduction is consistent for all three arrays, 1w x 6b, 1w x 12b, and 1w x 24b, showing independency on the word size. TCAM arrays with more than one word such as 2w x 12b and 4w x 12b were also designed, laid out, and had the interconnect capacitance extracted, to show the scalability of the table size. Fig. 108 (b) shows the

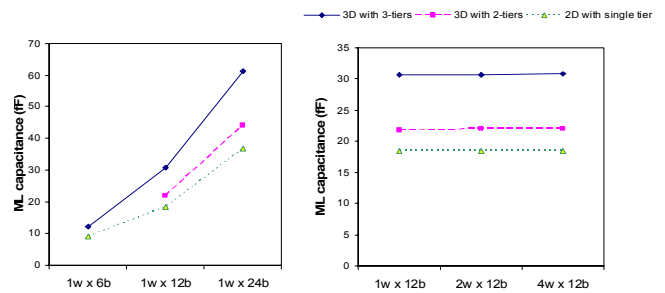


Fig. 8. ML capacitance from Q3D simulation results in sizes of (a) 1w x 6b, 1w x 12b, and 1w x 24b and (b) 1w x 12b, 2w x 12b, and 4w x 12b

ML capacitance being constant for all three sizes in all cases (single tier, 2-tier, and 3-tier design), showing the independency of ML capacitance on the number of words.

B. HSPICE simulations of TCAMs

With the parasitic capacitances extracted by the field simulator and the resistances calculated based on the sheet resistances given by MITLL, RC model for each interconnect were constructed and used for HSPICE simulation for TCAMs. To reduce the complexity of the model, parasitic capacitances less than 0.1 fF were neglected. The RC interconnect models for 1w x 12b in single tier process and three-tier process are shown in Fig. 9 and Fig. 10, respectively. In the single-tier 2D design, the parasitic capacitance due to the substrate dominates over the coupling capacitance with other interconnects, while coupling capacitance dominates over the parasitic capacitance with the substrate in three-tier 3D design. In 3D design, the interconnects on the Tier 2 and Tier 3 are a lot further from the substrate. Thus, only the interconnects on the bottom tier contribute to the parasitic capacitance with the substrate significantly, and ML couples strongly with the neighboring interconnects as shown in Fig. 10.

With the RC interconnect models described above, the TCAM designs in 2D and 3D are simulated in HSPICE. Fig. 11 shows that with the 3-tier 3D design, precharge time for matchline is reduced by 25% at 250 MHz. Fig. 12 shows the chip layout that has been submitted to MITLL for fabrication including 1w x 12b TCAM designs in single tier, 2-tiers, and 3-tiers as well as various test structures to characterize the 3D vias and other metal interconnects.

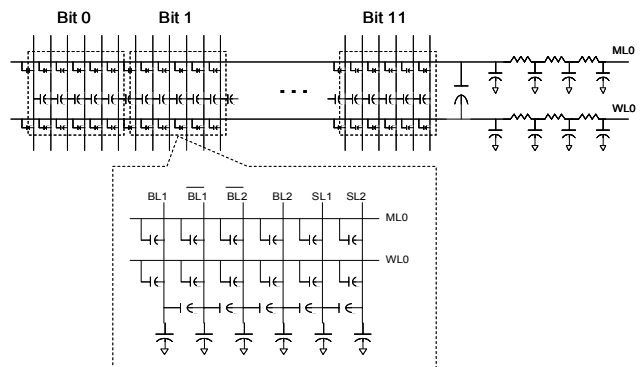


Fig. 9. RC model for interconnects in 1w x 12b TCAM array in 2D

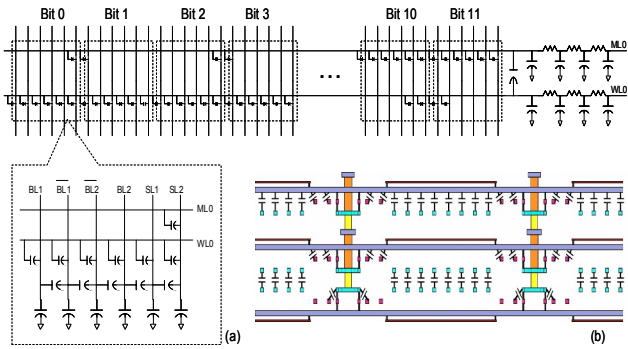


Fig. 10. RC model for interconnects in 1w x 12 b TCAM array in 3D with three tiers (a) top view and (b) cross sectional view with vertical coupling capacitances

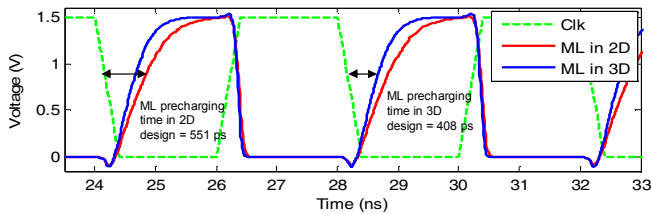


Fig. 11. Matchline waveforms of 1x12 TCAM in 2D and 3D with 3 tiers

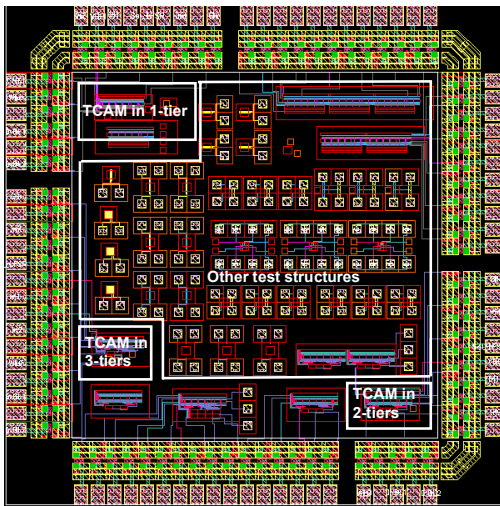


Fig. 12. Chip layout submitted to MITLL. It includes TCAM designs using a single-tier only, two tiers, three tiers, and various test structures.

Energy consumption and power dissipation have been analyzed for 1w x 12b TCAMs in both 2D structure and 3D structure with 3-tiers during the search operation according to (1). A number of different data patterns were written and searched for, and the value has been averaged over various combinations for a hundred clock cycles.

$$E = \int i \cdot v dt \quad \text{and} \quad P = \frac{\int i \cdot v dt}{T} dt, \quad (1)$$

where T is the total simulation time.

HSPICE simulation results show that power dissipation through ML is approximately 60% of the total power consumption. Energy consumption and power dissipation by ML only and the total power dissipation are shown in TABLE II. With 40% less ML capacitance, 32% ML power reduction was achieved and 21% total power reduction was achieved by using a three-tier 3D IC structure instead of the conventional 2D structure.

TABLE II
ENERGY CONSUMPTION AND POWER DISSIPATION OF 1 X 12 TCAM AT 250 MHZ

	Single Tier	Three-tier	Power reduction in 3D
Energy ML (pJ)	10.8	7.32	-
Energy total (pJ)	16.7	13.1	-
Power ML (μ W)	27.0	18.3	32 %
Power total (μ W)	41.8	32.8	21 %

C. Measurements

This chip was submitted for fabrication in November 2006. Measurement results will be included in the presentation.

V. CONCLUSION

Three identical TCAMs using a different number of tiers have been designed in MITLL FD SOI three-tier 3D IC process and are compared to show the benefits of 3D TCAM designs. One TCAM is designed using all three tiers, another TCAM is designed using two tiers only, and the last TCAM is designed using single tier only to emulate a conventional 2D process. By replacing matchlines with inter-tier 3D vias and using the inter-cell partitioning method, 28% and 40% matchline capacitance reduction was achieved for 2-tier design and 3-tier design, respectively, compared to the single tier design. The matchline capacitance reduction in TCAMs has been evaluated with various numbers of bits in a word and various numbers of words to show the scalability and consistency on larger designs. The three-tier 3D TCAM design shows 25% performance improvement in ML delay, 32% reduction in ML power dissipation and 21% reduction in total power dissipation, compared to the single-tier 2D design.

ACKNOWLEDGMENTS

The authors would like to thank Dr. S. Lipa and Dr. W. Davis for developing the Cadence design kit for MITLL process. We also thank E. Erickson for implementing the script for the Q3D conversion.

REFERENCES

- [1] J. Baliga, "Chips go beyond vertical [3D IC interconnection]," *Spectrum*, vol. 41, pp. 43-47, Mar. 2004.
- [2] MIT Lincoln Laboratory, "MITLL low-power FD SOI CMOS process: Design guide," revision 2006:7, Oct. 2006.
- [3] H. Miyatake, M. Tanaka and Y. Mori, "A design for high-speed low-power CMOS fully parallel content-addressable memory macros," *IEEE J. Solid state Circuits*, vol. 36, pp. 956-968, June 2001.
- [4] D. Perry and P. Gillingham, "Ternary CAM cell for reduced matchline capacitance," United States Patent Application Publication, Pub. No. US 2005/0276086, USPTO Class 365, Dec. 2005.
- [5] http://www.ansoft.com/products/si/q3d_extractor