

Rochester Institute of Technology

RIT Digital Institutional Repository

Theses

8-1-2008

Novel low power CAM architecture

Ka Fai Ng

Follow this and additional works at: <https://repository.rit.edu/theses>

Recommended Citation

Ng, Ka Fai, "Novel low power CAM architecture" (2008). Thesis. Rochester Institute of Technology.
Accessed from

This Thesis is brought to you for free and open access by the RIT Libraries. For more information, please contact repository@rit.edu.

Novel Low Power CAM Architecture

by

Ka Fai Ng

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Computer Engineering

Supervised by

Dr. Kenneth Hsu
Department of Computer Engineering
Kate Gleason College of Engineering
Rochester Institute of Technology
Rochester, NY
August 2008

Approved By:

Dr. Kenneth Hsu – Professor
Primary Advisor – R.I.T. Dept. of Computer Engineering

Dr. Dhireesha Kudithipudi – Assistant Professor
Secondary Advisor – R.I.T. Dept. of Computer Engineering

Dr. Muhammad Shaaban – Associate Professor
Secondary Advisor – R.I.T. Dept. of Computer Engineering

Thesis Release Permission Form

Rochester Institute of Technology
Kate Gleason College of Engineering

Title: Novel Low Power Content Addressable Memory Architecture

I, Ka Fai Ng, hereby grant permission to the Wallace Memorial Library to reproduce my thesis in whole or part.

Ka Fai Ng

Date

Dedication

This thesis is dedicated to my parents for their constant motivation and support to push me through.

Acknowledgments

I would like to give thanks to Dr. Ken Hsu for his support, patience, and help in motivating me to complete this thesis. I would also like to give thanks to Dr. Dhiresha Kudithipudi for her technical support and guidance. Finally, I would not have been able to complete this thesis without the help from Dr. Muhammad Shaaban.

Abstract

One special type of memory use for high speed address lookup in router or cache address lookup in a processor is Content Addressable Memory (CAM). CAM can also be used in pattern recognition applications where a unique pattern needs to be determined if a match is found. CAM has an additional comparison circuit in each memory bit compared to Static Random Access Memory. This comparison circuit provides CAM with an additional capability for searching the entire memory in one clock cycle. With its hardware parallel comparison architecture, it makes CAM an ideal candidate for any high speed data lookup or for address processing applications. Because of its high power demand nature, CAM is not often used in a mobile device. To take advantage of CAM on portable devices, it is necessary to reduce its power consumption. It is for this reason that much research has been conducted on investigating different methods and techniques for reducing the overall power. The objective is to incorporate and utilize circuit and power reduction techniques in a new architecture to further reduce CAM's energy consumption. The new CAM architecture illustrates the reduction of both dynamic and static power dissipation at 65nm sub-micron environment.

This thesis will present a novel CAM architecture, which will reduce power consumption significantly compared to traditional CAM architecture, with minimal or no performance losses. Comparisons with other previously proposed architectures will be presented when implementing these designs under 65nm process environment. Results show the novel CAM architecture only consumes 4.021mW of power compared to the traditional CAM architecture of 12.538mW at 800MHz frequency and is more energy efficient over all other previously proposed designs.

Table of Contents

Dedication	iii
Table of Contents	vi
List of Figures	ix
List of Tables.....	xii
List of Tables.....	xii
Glossary	xiii
Chapter 1 Introduction	1
1.1. Thesis Objective.....	1
1.2. Background: Power Model.....	2
1.3. CAM Basic	4
1.3.1 CAM Architecture.....	6
1.3.2 Supporting Works	11
1.3.3 Match Line Sensing Scheme.....	12
1.3.4 Search Line Driving Scheme.....	17
1.3.5 Power-Saving Techniques in the Architectural Level.....	19
1.4. Dynamic Power Reduction Techniques	21
1.4.1 Circuit Optimization.....	21
1.4.2 Multiple Supply Voltages.....	22
1.4.3 Dual Threshold Voltage	22
1.5. Static Power Reduction Techniques.....	23

1.5.1	Stack Effect	23
1.5.2	Sleep Transistors	24
1.6.	Other Low Power Novel CAM Architecture	25
1.6.1	SDW Match Line Architecture	25
1.6.2	Butterfly Match Line CAM Architecture.....	28
Chapter 2	Design and Implementation	31
2.1.	Modified CAM Cell	31
2.2.	Parallel Segmented Architecture.....	34
2.3.	Static and Dynamic Reduction.....	38
Chapter 3	Simulations and Results	41
3.1.	Comparison with Traditional CAM Architecture	42
3.1.1	Power Consumption	43
3.1.2	Performance/Speed.....	46
3.1.3	Area.....	47
3.2.	Comparison With Butterfly CAM Architecture.....	48
3.2.1	Power Consumption	50
3.2.2	Performance/Speed.....	52
3.2.3	Area	53
3.3.	Comparison with SDW Match Line Scheme Architecture	53
3.3.1	Power Consumption	55
3.3.2	Performance/Speed.....	57
3.3.3	Area	58
3.4.	Proposed CAM Architecture Simulation	58

3.4.1	Power Consumption	61
3.4.2	Performance/Speed.....	62
3.4.3	Area	63
3.5.	Benchmark All Architectures.....	64
3.5.1	Power.....	64
3.5.2	Performance/Speed.....	68
3.5.3	Area	69
Chapter 4	Conclusion.....	70
Chapter 5	Future Work	72
Bibliography	73

List of Figures

Figure.1.1: CAM Operation Block Diagram.....	5
Figure 1.2: Simplified CAM Operation	6
Figure1.3: NOR Type Match Line CAM Cell Configuration	7
Figure 1.4: NAND Type Match Line CAM Cell Configuration.....	8
Figure 1.5: Simple Voltage Sensing Differential Amplifier	10
Figure 1.6: WTA Current Sense Amplifier Uses in This Work.....	11
Figure 1.7: Low-Swing Match Line Scheme Configuration.....	14
Figure 1.8: Selective Pre-Charge Transistors Configuration	14
Figure 1.9: Current Race Transistor Configuration	15
Figure 1.10: Pipelined CAM Architecture Into Stages	16
Figure 1.11: Using Current Control Mechanism in Current Saving Scheme	17
Figure 1.12: Hierarchical Search Line Scheme Combined with Pipelined Configuration	19
Figure 1.13: Divided Entire Memory into Different Sub-Modules or Banks	20
Figure 1.14: Block Diagram for SDW Match Line Scheme.....	26
Figure 1.15: Two Divided Stage Comparison Process for SDW Match Line	27
Figure 1.16: NAND/AND Type CAM Cell Use in SDW Scheme.....	27
Figure 1.17: NOR/OR Type CAM Cell Used in SDW Scheme	27
Figure 1.18: Different Types of Butterfly Match Line Connection	29
Figure 2.1: Modified Proposed CAM Cell Configuration	32
Figure 2.2: 16-Bit Memory Module Configuration	35
Figure 2.3: One Segment of the 16-Bit Memory Data.....	36

Figure 2.4: Match Line Computation Unit.....	37
Figure 2.5: Overview of the Parallel Segmented Architecture for 64-bit Memory	38
Figure 3.1: Traditional CAM Architecture Simulation Under Nominal Matching Condition.....	42
Figure 3.2: Traditional CAM Architecture Simulation Under Maximum Matching Condition.....	43
Figure 3.3: Simulation Measurement of Traditional CAM Output Delay	46
Figure 3.4: Butterfly CAM Architecture Simulation Under Nominal Matching Condition	48
Figure 3.5: Butterfly CAM Architecture Simulation Under Maximum Matching Condition.....	49
Figure 3.6: Simulation Measurement of Butterfly CAM Output Delay	52
Figure 3.7: SDW CAM Architecture Simulation Under Nominal Matching Condition ..	54
Figure 3.8: SDW CAM architecture simulation under maximum matching condition	54
Figure 3.9: Simulation Measurement of SDW Match Line CAM Output Delay	57
Figure 3.10: Proposed CAM Architecture Simulation Under Nominal Matching Condition at 800MHz.....	59
Figure 3.11: Proposed CAM Architecture Simulation Under Maximum Matching Condition at 800MHz.....	59
Figure 3.12: Proposed CAM Architecture Simulation Under Nominal Matching Condition at 1GHz	60
Figure 3.13: Proposed CAM Architecture Simulation Under Maximum Matching Condition at 1GHz	60

Figure 3.14: Simulation Measurement of Proposed CAM Output Delay at 800MHz Clock	63
Figure 3.15: Simulation measurement of Proposed CAM output delay at 1GHz clock...	63
Figure 3.16: Summary of Static Power Consumption for All Architectures	64
Figure 3.17: Summary of Dynamic Short-Circuited Power Consumption for All Architectures	65
Figure 3.18: Summary of Dynamic Transitional Power Consumption for All Architectures	66
Figure 3.19: Summary of Average Block Power Consumption for All Architectures	67
Figure 3.20: Summary of Speed Performance for All Architectures.....	68
Figure 3.21: Summary of the Total Number of Transistors Used for All Architectures .	69

List of Tables

Table 3.1: Power Comparison for Traditional CAM and Proposed CAM.....	44
Table 3.2: Dynamic Power Dissipation Breakdown for Traditional CAM Architecture..	44
Table 3.3: Performance Comparison for Traditional CAM and Proposed CAM	46
Table 3.4: Total Number of Transistors Used in Traditional and Proposed CAM.....	47
Table 3.5: Power Comparison for Butterfly CAM and Proposed CAM.....	50
Table 3.6: Dynamic Power Dissipation Breakdown for Butterfly CAM Architecture.....	50
Table 3.7: Performance Comparison for Butterfly CAM and Proposed CAM.....	52
Table 3.8: Total Number of Transistors Used in Butterfly and Proposed CAM	53
Table 3.9: Power Comparison for SDW CAM and Proposed CAM	55
Table 3.10: Dynamic Power Dissipation Breakdown for SDW CAM Architecture	56
Table 3.11: Performance Comparison for SDW CAM and Proposed CAM	57
Table 3.12: Total Number of Transistors Used in SDW and Proposed CAM.....	58
Table 3.13: Power Comparison for Proposed CAM at 1GHz Clock Speed	61
Table 3.14: Dynamic Power Dissipation Breakdown for Proposed CAM Architecture ..	62
Table 3.15: Performance Comparison for Proposed CAM at Different Frequency.....	63

Glossary

BPTM	Berkley Predictive Technology Model
CAM	Content Addressable Memory
CPU	Central Processing Unit
GSL	Global Search Line
HSL	Hierarchical Search Line
LSL	Local Search Line
MLSA	Match Line Sensing Amplifier
NMOS	N-type Metal-Oxide Semiconductor
PDA	Personal Device Assistants
PMOS	P-type Metal-Oxide Semiconductor
SDW	Static Divided Word
SRAM	Static Random Access Memory
TLB	Translation Lookaside Buffer
WTA	Winner Takes All

Chapter 1 Introduction

1.1. Thesis Objective

Low power consumption has become the new metrics for determining the performance of an electronics device. Many electronics systems such as laptop computers, PDA, and mobile phone have become a commodity and a necessity. As the demand for portable applications increases, the demand for longer utilization and higher processing speed of the system also increases. Content Addressable Memory (CAM) is a specialized type of memory used in very high speed search applications, mostly used as a Translation Lookaside Buffer (TLB). The TLB allows the translation of the virtual address of a CPU to a physical address used in cache memory. Both CAM and Static Random Access Memory (SRAM) share many similarities and functionalities. However, unlike SRAM, CAM has an ability to search within the memory against an input vector. Each memory bit in CAM has its own comparison circuit to determine whether a match is found; SRAM does not. This provides CAM a unique searching capability, with results in a larger area as a tradeoff. In addition to a larger overall area, energy consumption also increases because each cell performs comparison simultaneously. As the energy consumption increases, it becomes difficult to use in a portable system. Hence, to utilize the advantages of CAM in a mobile device, power reduction is a required task. The objective of this thesis is to investigate power reduction techniques and power-saving circuit architecture to reduce the total power consumption. Furthermore, a different CAM architecture will be proposed to reduce the energy consumption in the architectural level. Combining both architectural and circuitry level power reduction techniques, the new

CAM will minimize the total power consumption while maintaining its high speed searching performance.

1.2. Background: Power Model

The power of any given system or device can be separated into dynamic and static power. Dynamic power consists of transistors switching activity and short-circuited current. Static power dissipation contributes by leakage current. The power usage of a system can be modeled by the following equations[23]:

$$P_{Total} = P_{dynamic} + P_{static} = (P_{transition} + P_{shortcircuit}) + P_{leakage}$$

$$P_{transition} = \alpha * f_{CLK} * \frac{CV^2}{2}$$

α = Activity factor

f_{CLK} = Clock frequency

V = Power supply voltage

$$\frac{CV^2}{2} = \text{Energy dissipated for each transition}$$

The transitional power of a device can be calculated based on its operating frequency, the probability of being active, load capacitance, and the power supply that charges the capacitors.

Dynamic power can be reduced tremendously if power supply voltage is minimized. Furthermore, lowering the clock frequency and lowering the amount of activity can also help reduce the amount of dynamic power consumption. Ironically, the clock speed has been increasing in many of today's applications. By lowering the power supply, it can reduce the dynamic power consumption significantly with a trade off of

decreasing the noise margin and performance. Similarly, P type and N type transistors may both be active simultaneously and in turn create a short-circuited scheme[2][9].

$$P_{shortcircuit} = \alpha * f_{CLK} * \sum E_{SC}$$

α = Activity factor

$\sum E_{SC}$ = Sum of short circuit energy produced by rise time and fall time

$$E_{SCf} = t_f (V_{DD} - |V_{TP}| - V_{TN}) \frac{I_{SCMax_fall}}{2}$$

$$E_{SCr} = t_r (V_{DD} - |V_{TP}| - V_{TN}) \frac{I_{SCMax_rise}}{2}$$

The short circuit energy can be decreased by reducing the rise and fall time. This can eliminate the time for either the NMOS or PMOS to be turned on at the same time. Increasing the clock frequency or decreasing the load capacitance can reduce the short circuit rise and fall time, but it will increase the overall clock rate and result in higher switching power.

Leakage power is contributed by many different factors. Factors include sub-threshold conduction, reverse bias PN junction conduction, gate induced drain leakage, drain-to-source punch-through, and gate tunneling. The major contributor to the leakage power is the sub-threshold conduction. The sub-threshold conduction can be modeled as:

$$I_{Sub} = \alpha \frac{1}{L_{eff}} \exp\left(\frac{q(V_G - V_T)}{kT}\right)$$

α = Activity factor

L_{eff} = Effective channel length

$$\frac{q}{kT} = \text{Thermal voltage}$$

$$V_G = \text{Gate voltage}$$

$$V_T = \text{Threshold voltage}$$

As shown in the equation above, the sub-threshold voltage is contributed by the temperature, the effective channel length, and the voltage threshold. If the channel length is reduced, then the sub-threshold will increase because of their inverse relationship with each other. Hence, reducing the effective channel (from 130nm \rightarrow 90nm \rightarrow 65nm \rightarrow 45 nm) can greatly increase the sub-threshold current. By reducing the voltage threshold due to lowering the supply, voltage will also increase the sub-threshold current exponentially.

System power consumption can be estimated using this model. Furthermore, it provides an insight to the type of power that dominates and estimates the result when parameters changed.

1.3. CAM Basic

The operation of a CAM is very similar to a hash table. It searches through the memory against the input search bits. A signal will be generated if a complete match occurs, and no signal otherwise. Figure 1.1 below shows how a typical CAM operates when an input search bits are sent to CAM for comparison.

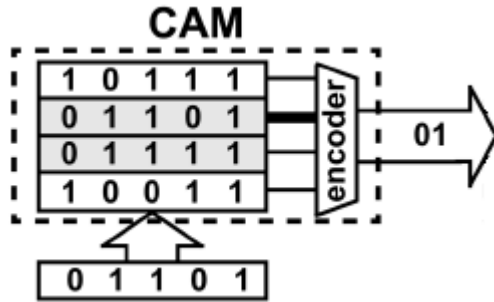


Figure.1.1: CAM Operation Block Diagram[17]

The output of CAM is followed by an encoder to translate the result from the comparison into an address that can be deciphered by the receiving side. Content addressable memory utilizes the same basic structure as an ordinary memory device. Each cell within the CAM has the same basic structure of a 6 Transistor (6T) SRAM cell. It has two cross-coupled inverters to retain its state and two other transistors that act as a switch or gate used in writing and reading. The unique part of CAM is its additional circuitry that has the capability of comparing the bit that is stored within the cell. The comparison circuitry varies depending on the designer and the implementation. Some use an XNOR operation and others use an XOR operation. Furthermore, there will be two dedicated input signals connected into the cell for searching operation. The result of this circuitry will determine if the search bit matches with the bit that is stored within the cell.

This is done for each of the cells within the memory. A simplified structure of a CAM is shown in Figure 1.2.

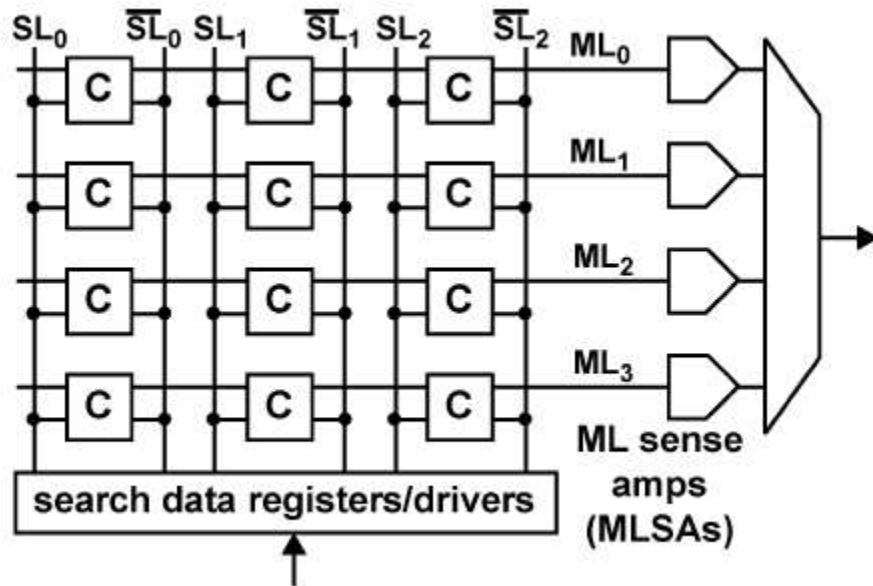


Figure 1.2: Simplified CAM Operation[17]

Each cell has two search lines, merely the complement of each other. If all the cells within the same row are matched, the match line is triggered. The match line can be logic high or logic low for match indication depends on the implementation. The output of the match line is typically connected to a sense amplifier, which amplifies the signal to the output.

1.3.1 CAM Architecture

1.3.1.1 CAM Cell

1.3.1.1.1 NOR Type Match Line CAM

The most common type of CAM architecture is the NOR Type Match Line Architecture shown in Figure 1.3. The NOR Type CAM architecture indicates a mismatch by pulling the match line to ground. The advantage of a NOR Type CAM architecture is its performance. When a mismatch for a bit occurs, the entire match line signal is pulled to ground, indicating a mismatch. Since all comparison units are parallel, each comparison does not depend on the previous cell's match result. Therefore, it

delivers very quick comparison speed. However, the major disadvantage is the power it consumes during charge and discharge. The probability of discharging the match line is $\frac{(N-1)}{N}$ for total of N search lines. It is almost certain there will be a discharge during a comparison using the NOR Type Match Line Architecture. Hence, the match line needs to charge and discharge whenever a comparison is needed.

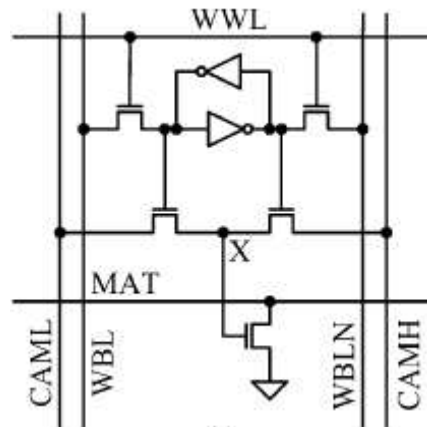


Figure1.3: NOR Type Match Line CAM Cell Configuration[4]

1.3.1.1.2 NAND Type Match Line CAM

Another type of CAM architecture proven to consume less power is the NAND Type Match Line shown in Figure 1.4. The advantage of the NAND type match line CAM architecture comes from its matching method. Instead of discharging for a mismatch, the NAND Type Match Line architecture discharges for a match. The probability of a match for the NAND type during a CAM comparison operation would be $\frac{1}{N}$ for the total of N search lines. This indicates that the chance for discharge of a match line is minimal. Therefore, the NAND type match line greatly reduces the amount of discharge energy.

The chance for the transistors to discharge is almost certain in a NOR type match line. Hence, the NAND type match line is a good choice to be used in a low power device. However, with the low power saving advantage, comes the disadvantage of poor performance. The disadvantage of a NAND type match line is the delay time to indicate a match increased linearly as the number of bits increased. In order to indicate a match, all NMOS transistors need to be turned ON and try to pull the signal to the ground. However, as the number of bits increases, the time it takes the ground signal to travel through all memory bits would be enormous. Hence, even with the advantage of reducing power consumption, the poor performance is not applicable in high speed application.

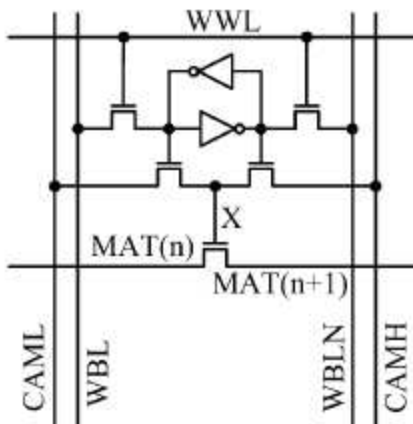


Figure 1.4: NAND Type Match Line CAM Cell Configuration[4]

1.3.1.2 Address Decoder

The address decoder is a very common component used in almost every memory device. The essential purpose is to translate the input address into an address, which corresponds to the address of the memory matrix. An example is the input address of 0101 is translated into an address of 5 in decimal, with the output to the memory matrix in the binary form of 0000 0000 0010 0000. This translation of address essentially indicates the particular row of the memory matrix this operation is interested in.

1.3.1.3 Controller

The controller of a memory array is the center of determining which operation the memory matrix is required to perform. In most common memory systems, writing to and reading from a memory location is essential. In the case of CAM, a search operation is added. Hence, incorporating the two common operations of reading and writing; CAM's controller would also need to implement the search operation to instruct the memory to go into its matching mode. This additional operation increases the complexity and area of the controller component.

1.3.1.4 Sense Amplifier

A sense amplifier is another essential component of a memory device during a read operation. A sense amplifier is an analog circuitry rather than a digital circuitry like other components. The essential function is to sense the voltage across the bit line of each cell and to quickly amplify the differential signal. As shown in Figure 1.3 and Figure 1.4, each cell contains two cross-coupled inverters and has two complementary bit lines. The sense amplifier detects the two complementary bit lines for voltage difference. The voltage difference is amplified by the sense amplifier to the read output. A traditional sensing amplifier is usually a voltage sensing differential amplifier shown in Figure 1.5.

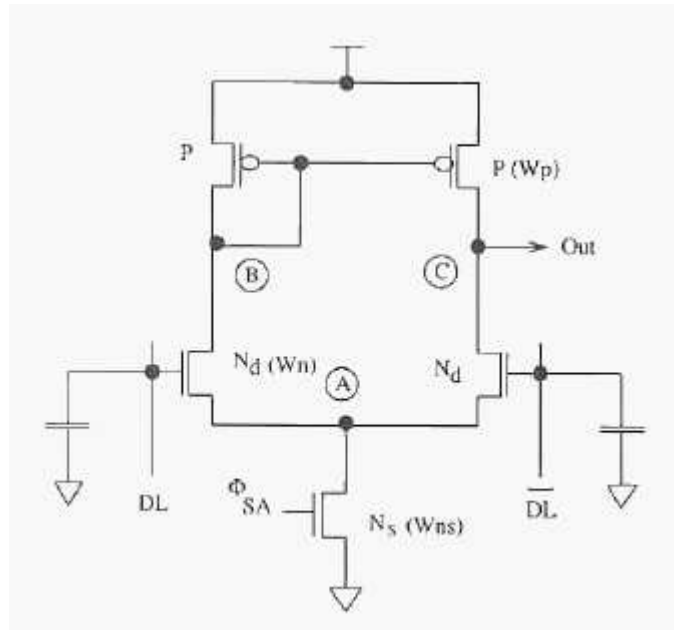


Figure 1.5: Simple Voltage Sensing Differential Amplifier[1]

This type of amplifier is simple to implement, but requires a large differential voltage between the two complementary lines to amplify the signal. Hence, it is not an ideal candidate as the technology continues to scale down to sub-micron level. Furthermore, the power consumption of the voltage sensing differential amplifier can consume a large amount of power. The current sense amplifier is another type of sensing amplifier introduced to compensate for the slow response time and large power consumption of a voltage sensing amplifier. Instead of detecting the differential voltage, which can be degraded by the capacitance as the size of the memory column increases, it senses the differential current. By sensing the differential current, it reduces the effect of large capacitance from the increasing column size. Hence, it achieves a quick response time. Furthermore, it only requires a very small current difference to activate the amplifier. A Winner Takes All (WTA) current sense amplifier is used in this work to provide fast reading time. The current sense amplifier is illustrated in Figure 1.6.

1.3.3 Match Line Sensing Scheme

Many different match line sensing schemes are proposed with the idea of reducing the power consumed by the match line. Some of the match line schemes are discussed below:

1.3.3.1 Conventional (Pre-Charge High) Matching Scheme

The conventional matching scheme involves pre-charging high to the match lines. In the NOR match line scheme, the match line remains high until a mismatch occurs, which pulls the match line to the ground. Hence, at the end of the match operation, logic low indicates a mismatch and a high indicates a match. In the NAND match line scheme, the match line pre-charges high, but when it is a matched word, the match line is pulled to the ground or logic low. Hence, it is considered a match when the match line indicates a zero and VDD otherwise.

The power consumed for a mismatch is due when the rising edge for pre-charging and falling edge for evaluation. Hence, the power consumed when a miss occurs is:

$$P_{miss} = C_{ML} V_{DD}^2 f$$

C_{ML} Represents the match line capacitance

V_{DD} Represents the power supply

f Represents the search operation frequency

The power consumption associated with a single match line depends on the previous state. Consider there are n number of match lines in the CAM; the match line power is then illustrated by the following equation:

$$P_{ML} = nP_{miss} = nC_{ML} V_{DD}^2 f$$

If the previous state is a miss, it requires a pre-charge for the current state equivalent to P_{miss} . At the same time, if the previous state is a match with the current comparison indicating a mismatch, then it also consumes the same amount of power as P_{miss} . If the previous state and the current state are equivalent, then the power consumed is negligible. This indicates the change of state is where power dissipated.

1.3.3.2 Low-Swing Schemes

The low-swing match line scheme provides a method to reduce the power consumed in the match line in the case of a match. This technique requires the match line to significantly reduce the voltage level; nevertheless, it is still able to indicate whether a match occurs. With the driving voltage reduced, the equation for the match line can be rewritten as:

$$P_{ML} = nC_{ML}V_{DD}V_{MLSwing}f \text{ where } V_{DD}^2 \text{ becomes } V_{DD}V_{MLSwing} .$$

The major drawback to this scheme is the difficulty in producing a lower voltage drive to the match line without the use of an external power supply or buck converter. Instead, additional circuitries are added to the existing CAM architecture to reduce the voltage drive of the match line. A tank capacitor is added to drive the match line. By inserting the tank capacitor, the match line voltage, if it were to pre-charge, would be:

$$V_{MLpre} = V_{DD} \frac{C_{\tan k}}{C_{ML} + C_{\tan k}}$$

Furthermore, since the match line voltage is reduced, it would not be capable of driving the match line to indicate a match (high). Hence, a sensing amplifier needs to be added to the match line to amplify the signal. This scheme greatly reduces the driving

voltage in the match line, at the price of increasing the overall area and the complexity of the circuit.

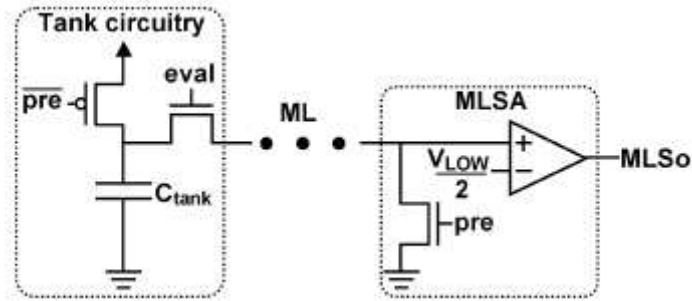


Figure 1.7: Low-Swing Match Line Scheme Configuration[17]

1.3.3.3 Selective Pre-Charge Scheme

The selective pre-charge scheme is a slightly different method of reducing the match line powers. The method of reducing the match line power is to selectively pre-charge certain match lines evaluated as having the potential of matching. The Selective pre-charge performs the match operation on the first few bits of a word. Once all the bits in a word matched, the match line is charged and compared with the rest of the bits. For the first few bits not matched, the match line is not pre-charged; hence, power is saved. This method works very well when data distributions are uniformly distributed. However, the power would be equivalent to an ordinary pre-charge scheme when all the first few bits of the word are identical. A simple implementation of the selective pre-charge scheme is shown below:

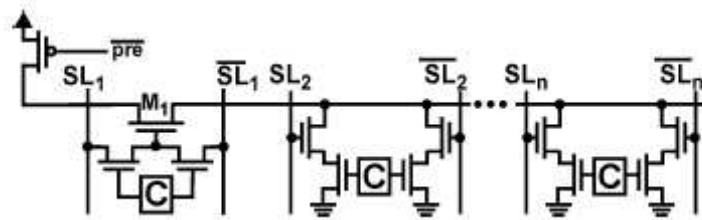


Figure 1.8: Selective Pre-Charge Transistors Configuration[17]

1.3.3.4 Current Race

The Current Race scheme is another technique for reducing the power consumed in the match line. In this scheme, the match line pre-charges low and evaluates the match line states by charging the match line with a current I_{ML} supplied by a current source. The configuration is shown in Figure 1.8.

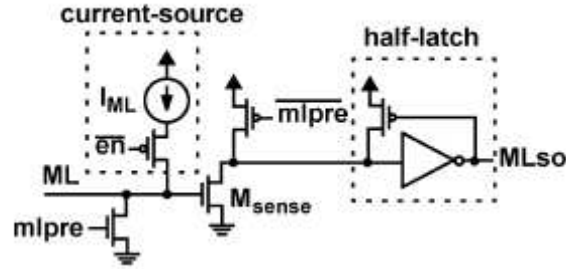


Figure 1.9: Current Race Transistor Configuration[17]

In the pre-charge phase, the match line is pulled down to ground when $mlpre$ is high. During the evaluation phase, $mlpre$ is low and \overline{en} connects the current source to the match line. If there is a match, then the match line charges linearly to a high voltage. This turns on M_{sense} with the half-latch outputting ML_{so} high as the indication of a match. In the case of a mismatch, the match line charges to a voltage of $\frac{(I_{ML} * R_{ML})}{N}$. N represents the number of bits that are a mismatch (pulls to the ground) and R_{ML} indicates the resistance for the transistor that pulls the match line to ground. The M_{sense} transistor trips the latch with a threshold of V_{TH} . In the case of matching, the match line is charged above V_{TH} , pulling the input of the half-latch to low and output ML_{so} to high. For a mismatch, the match line has less voltage and leaves the latch with the initial state. The power consumption is very similar to the case of low-swing scheme, except the match

line voltage is slightly above V_{TH} . This is also the case when a mismatch occurs. Hence, the power consumed using this scheme is:

$$P_{ML} = nC_{ML}V_{DD}V_m f$$

1.3.3.5 Pipelined Scheme

The pipelined scheme is just another variation of the selective pre-charge scheme. Instead of dividing into two segments, the pipelined scheme divides into multiple segments. The pipeline scheme needs to add additional latches to the output of each segment. If a segment matches with the search bits pattern, it is latched and evaluated during the next segment in the next cycle. If a segment is determined to be mismatched, then the next segment is not turned ON; hence, power is saved. However, the pipelined scheme does not save a significant amount of power in general. Furthermore, it increases the complexity and area overheads. A simplified pipelined scheme is shown below:

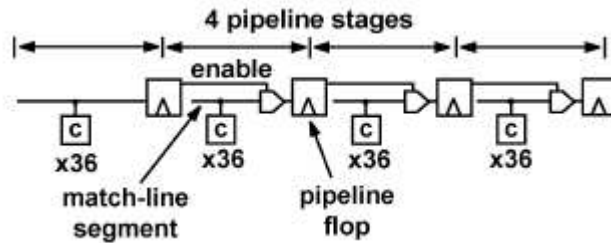


Figure 1.10: Pipelined CAM Architecture Into Stages[17]

1.3.3.6 Current Saving

The current saving scheme is an improved version of the current race scheme. In the current race scheme, the current source is supplying a constant current whether a match or miss. However, in the current saving scheme, the current is reduced when a miss has occurred and is maintained at the same current when it is a match.

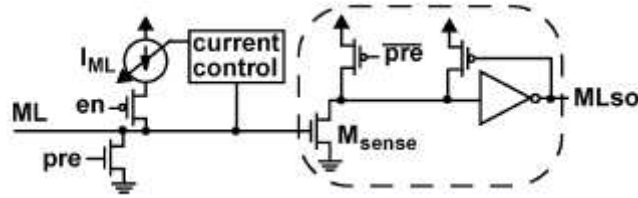


Figure 1.11: Using Current Control Mechanism in Current Saving Scheme[17]

However, this design greatly increases the complexity of the circuit. The employment of the current control unit also increases the overall area of the circuit.

1.3.4 Search Line Driving Scheme

Another part of the CAM architecture consuming power is the search line driving scheme. There are two search lines per cell; one line is the actual search bit and the other is its complementary bit. For every search operation, the search lines pre-charge or discharge to indicate the search bit value. Hence, the power consumption is also very high and not power efficient. To compensate for the large energy consumption, many different schemes are proposed to reduce the total power dissipated when driving the search lines. Some of the schemes are discussed below:

1.3.4.1 Eliminating Search Line Pre-Charge

The most direct way of reducing the search line power is to eliminate the pre-charge phase for the search line. The total power consumed by the search line pre-charges and discharges is:

$$P_{SL} = nC_{SL}V_{DD}^2f$$

n Represents the total number of search line pairs

$2n$ Represents the total number of search lines

C_{SL} Represents the capacitance of the search line

V_{DD} Represents the power supply

f Represents the searching frequency

If the pre-charge search line phase is eliminated, and directly drives the search line when matching, the power consumed is reduced into:

$$P_{SL} = \frac{1}{2}nC_{SL}V_{DD}^2f$$

Without the pre-charge phase, only 50% of the search lines are charged. Therefore, reduces the search line power by a half.

1.3.4.2 Hierarchical Search Line (HSL)

The hierarchical search line scheme is built on top of the pipelined scheme. This method tries to shut off the search line when it determines it is not necessary to perform the search operation. Based on the pipelined scheme, it splits the search line into the Global Search Line (GSL) and the Local Search Line (LSL). Each GSL activates more than one LSL. Also, the GSL is activated at all times, but the LSL is only activated when it is triggered by a match line. The idea is in the pipelined scheme; some of the match lines are not activated when the previous segment determines a mismatch. The matching lines trigger the LSL. Otherwise, the LSL is disabled for that particular search cycle, thus saving power.

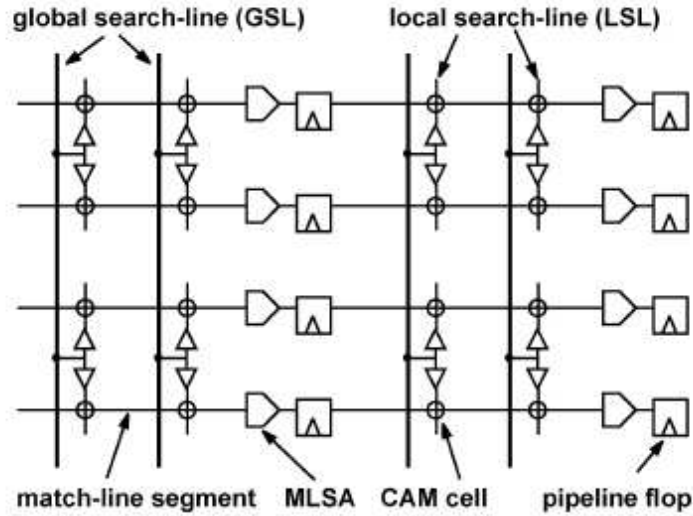


Figure 1.12: Hierarchical Search Line Scheme Combined with Pipelined Configuration[17]

The total power consumed using this method is:

$$P_{SL} = (C_{GSL} V_{DD}^2 + \alpha C_{LSL} V_{DD}^2) f$$

C_{LSL} does not turn ON all the time; the rate by which C_{LSL} turns on is represented by α . Hence, as the number of α decreases, or the rate of which C_{LSL} is activated, the less power is consumed.

1.3.5 Power-Saving Techniques in the Architectural Level

Power saving does not seem to be very significant at the circuitry level. Most often, the level saving the most power of the device is at the architectural level. It is at this level, that major power saving is achieved because it often disables or turns off the unit not being used. Hence, achieving major power saving. Some of the architectural power-saving techniques are discussed below:

1.3.5.1 Bank-Selection

In most cases, CAM is usually divided into different banks or subsets. Each subset contains a specific set of addresses. Each search word has additional bits indicating which bank to select from. Hence, given an address, there is only one bank activated at one time. The inactivate banks save the significant amount of power.

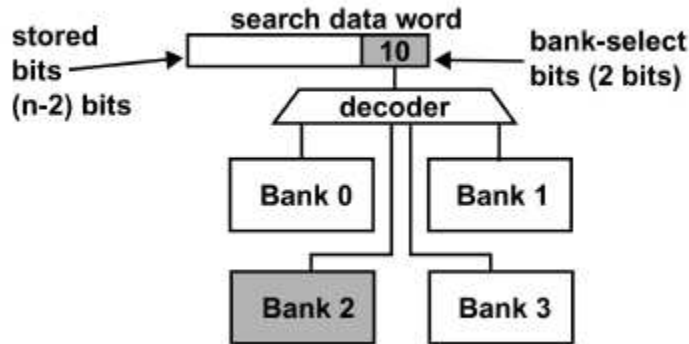


Figure 1.13: Divided Entire Memory into Different Sub-Modules or Banks[17]

The drawback of this scheme is the banks overflow. In a CAM, there are many more input combinations than storage locations; hence, the storage locations in the bank can quickly overflow. An example is a CAM with 72-bit words with additional bank selection bits and 32K entries divided into two banks. Each bank has 16K entries. While each bank has 16K entries, with 2^{72} possible combinations, it is more common to have more entries than it can fit in the assigned bank. This often requires extra circuitries to periodically re-partition the banks and forces multiple banks to activate at once. The CAM may consume more power as the result of re-partitioning and adding extra circuitries.

1.3.5.2 Pre-Computation

This scheme is very similar to the selective pre-charge match line scheme. However, instead of checking the first few bits, it uses other means of computation to

determine the probability of the word line being activated or disabled. The major difficulty of the pre-computation scheme is to find a mechanism to filter out most of the words determined to be a mismatch. Hence, with the pre-determination of a word not matching, it can be put to sleep with reduced power consumption. The disadvantage is when a pre-computational unit is poorly designed. It may increase power consumption due to more power dissipated in the pre-computational unit. Hence, it is important to ensure the pre-computational unit does not consume too much power, thus, filtering out many known mismatches before the second comparison phase occurs.

Overall, a CAM can be constructed using multiple techniques mentioned above to reduce overall power consumption. Furthermore, many of the techniques required a tradeoff of performance when reducing the energy consumption. Also, as the technology emerged into the submicron level (<90nm), static power consumption likely dominates the overall power consumption. Hence, the proposed CAM not only reduces the dynamic power, but also the static power. The proposed CAM limits the switching activities to reduce the dynamic power whereas the leakage feedback gating[25] technique reduces the static power. Lastly, the proposed CAM architecture not only reduces overall power, but also maintains its high performance.

1.4. Dynamic Power Reduction Techniques

1.4.1 Circuit Optimization

The idea behind this technique is to exam the circuit and tries to rearrange or optimize the configuration to reduce or balance the circuit, so the amount of switching is evenly distributed across the circuit. Some transistors in a circuit would likely to switch more often than other transistors within the same circuit. Hence, an actively switching

transistor consumes more power and generates more heat compared to an idle transistor. Through determining the probability a particular part of the circuit is active, it is possible to re-configure the circuit organization. Although, not all circuit configurations can be re-arranged and achieve power reduction. Some re-configuration or optimization might lead to worse power consumption.

1.4.2 Multiple Supply Voltages

Another technique reducing power consumption of a circuit is to utilize multiple supply voltages. Certain parts of the circuitry might not require to completely utilize the full supply voltage compared with the others. In such a case, it is sufficient to use a different voltage to supply the power for that part of the circuitry instead of using the global power supply. Given the total power a circuit consumes is inversely proportional to the power supply squared, utilizing a lower power supply for one part of the circuit would mean reducing the overall total power consumption. Furthermore, some part of the circuit might require only a very small amount of voltage to maintain its state; such that the part can use a lower power supply sufficient enough to maintain its value. One drawback of this approach is the additional circuitries required to have multiple power supply within the same circuit. Furthermore, it increases the complexity of the circuit and the overall area.

1.4.3 Dual Threshold Voltage

Another technique used to reduce the power consumption is the dual threshold voltage technique. This technique not only reduces the dynamic power considerably, but is also capable of reducing static power consumption. The concept of dual threshold is to use two different threshold type transistors; one type of transistor might have a higher

threshold than the other. The higher threshold voltage transistor has a slower response and higher power consumption. The lower threshold voltage transistor has a fast ON-OFF response and less power consumed. The dual threshold voltage configuration is to mix the two types together. To reduce the power consumption, it is likely to use lower threshold transistors entirely. However, because it is lower in threshold voltage, it increases the chances for static leakage current to pass through the transistor when it is idle or during an unknown state. Hence, lower threshold voltage transistors are likely to be placed in the critical path of the circuit where timing is crucial and require a fast ON-OFF response. To compensate for the large leakage current, the higher threshold voltage transistor is to be placed on the non-critical part of the circuit to reduce the amount of leakage current because of its high threshold voltage characteristic.

1.5. Static Power Reduction Techniques

1.5.1 Stack Effect

The stack effect is the simplest static power reduction technique used to reduce the leakage current of a circuit. The configuration of this technique is to “stack” the transistors on top of each other. This organization can help the leakage current because the stacking has created a large resistance when connecting transistors in series. A transistor can be modeled as a resistor at the linear operating region when the gate is driving with a voltage that barely turns on the transistor. This series connection can be imaged as many resistors connected together in series. The result is a large resistance that reduces the amount of current passing through. This technique is particularly useful when speed is not a critical requirement. Due to its series connection, the signal at one end

requires additional time to transmit to the other end of the connection. This technique is applicable when there are only a few transistors in series and timing is not critical.

1.5.2 Sleep Transistors

Another technique reducing the static power dissipation is to utilize a sleep transistor. A sleep transistor, as the name states, is used when a particular part of the circuit is idle or is placed in stand-by mode. In a large scale integrated circuit, there are a few hundred thousands of transistors and circuitries. Not all circuit elements are operating at the same time. There are certain parts of the circuitry that are idling or inactive when some conditions are not met. The inactive parts are consuming power even when they are not performing any operations. Therefore, it can reduce the power when cutting off power to those inactive devices. A sleep transistor acts as a gate for connecting and disconnecting a certain part of the circuit to its power supply and ground. While in nominal operation, the sleep transistor connects the circuitry to its power supply and ground to perform the requested operation. When it is not active, the sleep transistor disconnects the power supply and ground, leaving the circuit in a floating state. While the circuit is isolated from the power source and ground, all state information before the sleep phase is lost, but there is no leakage current flowing out while it remains disconnected. Isolating the power supply and ground from the main circuit is useful for any circuitry that is combinational, since the output does not depend on the previous state to produce the correct result. Extra circuitries are added to keep the main circuit floating while remembering the state it previously held before entering the sleep phase.

1.6. Other Low Power Novel CAM Architecture

The traditional CAM architecture is simply connected in series or in parallel. For high processing and performance, parallel or NOR Type of CAM architecture is used. However, such a connection produces a large amount of static leakage wasted current and consume a large amount of power. For series or NAND Type CAM architecture, the power dissipation is very minimal. It reduces a lot of static leakage current and dynamic power consumption. However, it is not applicable for any high performance application because the search time for the NAND type increases when the number of data to compare increases. Two other novel CAM architectures proposed are the Static Divided Word (SDW) match line scheme[10] and the Butterfly Connection Style CAM architecture[19].

1.6.1 SDW Match Line Architecture

The principle idea of this architectural scheme is to divide the memory into two comparison processes instead of one long comparison process. The first comparison process consists of partial bits from the actual stored data. Hence, the partial bits of each stored data compares the input data partial bit to the stored data partial bits. If the first comparison process produces an indication that all the partial bits matched, then the second comparison process performs its final comparison. However, if it fails at the partial bits comparison, then it does not generate the match signal to the second stage. It saves power by preventing more comparison is being done. The architecture of SDW is illustrated in the figure below:

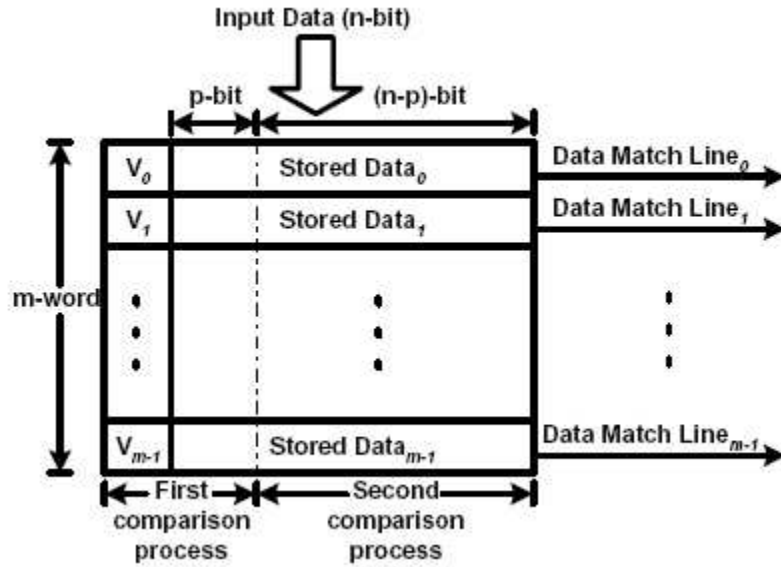


Figure 1.14: Block Diagram for SDW Match Line Scheme[10]

As the first stage fails, there is no driving signal from the first comparison process for activating or continuing the drive for the second stage comparison. Furthermore, to compensate when stored bits become performance overhead, the comparison circuit in the second process is a NOR type configuration to decrease the matching delay. On the other hand, the first process usually has fewer bits to compare, and is active in all searching operation. Therefore, a NAND type comparison circuitry is used to reduce the static power consumption.

This architecture provides both dynamic and static power reduction. For the first comparison stage, it utilizes the stacking effect of transistors to reduce the static leakage current. In the second stage, where it is not active as often as in the first stage, it utilizes the NOR type comparison circuit. The NOR type comparison circuit compensates the delay the series transistor causes in the first stage. The configuration is shown in the following figures.

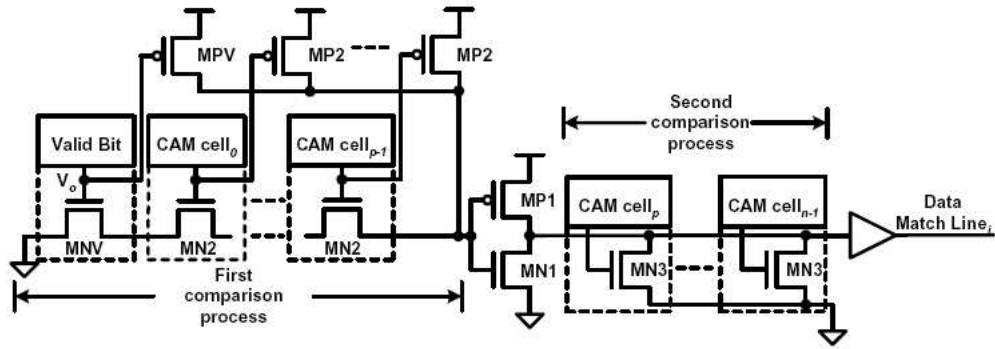


Figure 1.15: Two Divided Stage Comparison Process for SDW Match Line[10]

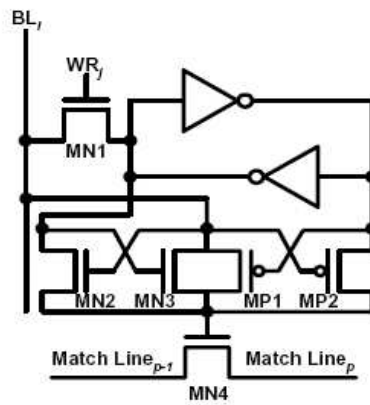


Figure 1.16: NAND/AND Type CAM Cell Use in SDW Scheme[10]

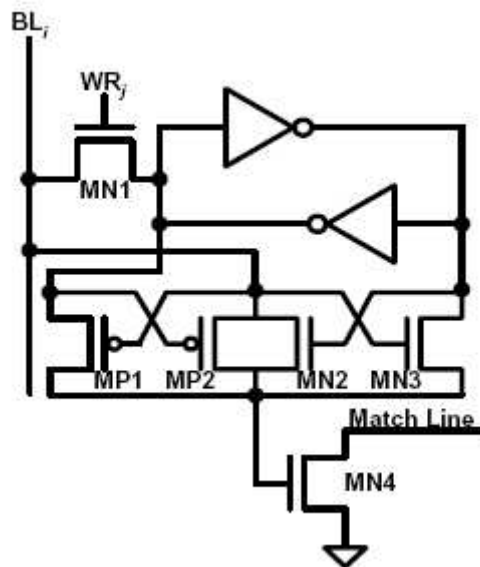


Figure 1.17: NOR/OR Type CAM Cell Used in SDW Scheme[10]

For the SDW match line architecture, the possible drawback of this design is when comparing all the bits and only when the last bit is incorrect, causing both stages to operate and be active. This uses the same amount of power consumption as if it is a fully parallel or fully series. Furthermore, by reducing one complementary side of the memory cell, it reduces dynamic power as less switching occurs. However, it increases the resistance and capacitance for the bit line as it becomes the line for the read and write operation. Without the complementary side, a different sense amplifier is utilized to read the bit out correctly. Moreover, the signal received at the sense amplifier might be distorted or corrupted due to the high impedance of the wire.

1.6.2 Butterfly Match Line CAM Architecture

The Butterfly Match Line CAM architecture is recently proposed for low power CAM operation. As the author suggested, there are three types of butterfly match line connection. Depending on the number of bits in the system, a different style can be implemented for the specific situation. The operation of the butterfly match line architecture is very similar to the static divided word match line described earlier. Nevertheless, instead of dividing comparison into two large sets, it divides into multiple sets of smaller bits to compare. Furthermore, it relieves some of the bottleneck of the static divided match line scheme by parallelizing the comparison process. As shown in Figure 1.18, the entire data is divided into multiple stages. For example, for Type C, the first comparison is only required for comparing the first four bits of the stored data. If the initial four bits are matched, the next stage is triggered to compare. Such comparison propagates into the next stage when the previous stage is a match. Notice that if any of the previous stage indicates a mismatch, the mismatch signal transmits through all four

parallel comparison groups and continues to spread along the later part of the stored data. Hence, it provides the same power-saving mechanism when at any previous stage, for any bits is a mismatch, the later part of the circuit does not perform the comparison because it has been determined a mismatch.

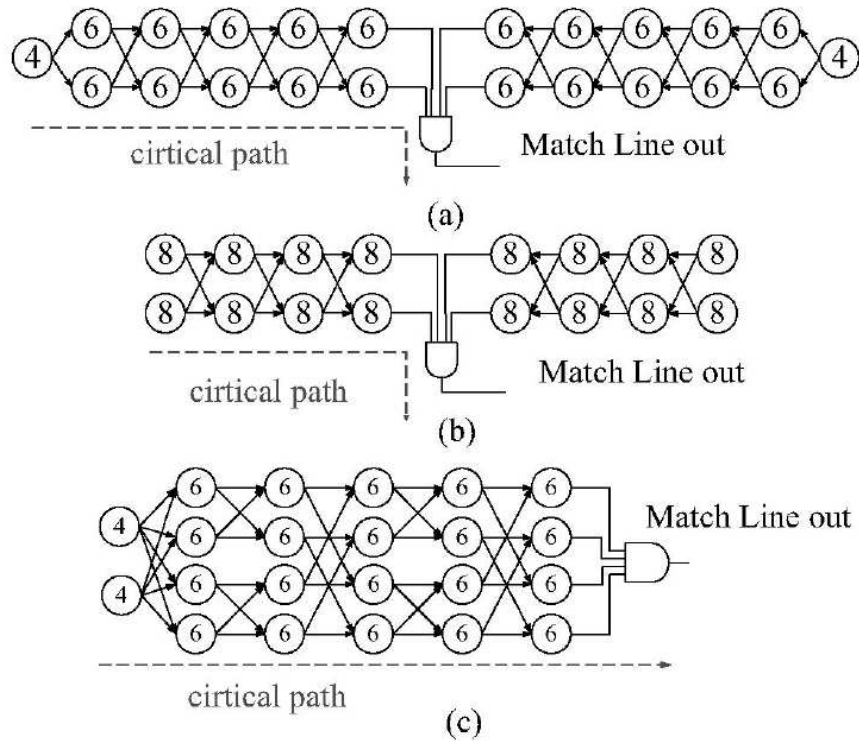


Figure 1.18: Different Types of Butterfly Match Line Connection[19]

Finally, all of the independent comparison part results in four signals performing a final AND operation to indicate whether this particular row matches with the input search data. One potential drawback is the area, because it requires an additional circuit for connecting each stage. For example, the third stage (from Figure 1.18(c) above), requires the comparison result from two other comparison groups. One of many signals might require additional time to compute and the others might be easily processed. The final result cannot be generated because not all signals have arrived. Hence, the output signal that needs to be sent out to the next stage takes additional time. Another drawback

is decreasing in match performance as each stage requires a signal before processing; hence, creating a large amount of delay to finish processing all the bits.

Many techniques have been explored and used in CAM architecture to reduce the overall power. Even though each technique described above reduces the power in dynamic, static or both forms of power consumption, it also contains drawbacks that either reduces the performance or increases the area. This work has developed ideas from other proposed architectures and power reduction techniques together to produce a novel architecture that can significantly reduce the power consumption while maintaining its high performance.

Chapter 2 Design and Implementation

2.1. Modified CAM Cell

The proposed CAM architecture utilizes a different type of CAM cell. This cell operates in a XOR operation. However, instead of utilizing traditional CAM cell configuration, the modified CAM cell contains ten transistors with only four transistors (P2, P3, N4 and N5) used as XOR comparison shown in Figure 2.1. The XOR comparison circuit is constructed so only two transistors are operational for any comparison. For example, if the data bit from the memory is a 1 and matched with the search bit, N4 and P3 turns ON while P2 and N5 turns OFF, and vice versa if the data bit is a 0 and there is a mismatch against the search bit. Furthermore, the XOR comparison only utilizes one of the complementary memory bits to drive the comparison circuit. Since only one of the transistors (P2 or N4) is turned ON at any comparison, it is needless to utilize another memory bit to drive another transistor. This in turn reduces an additional drive from the memory and reduces power consumption.

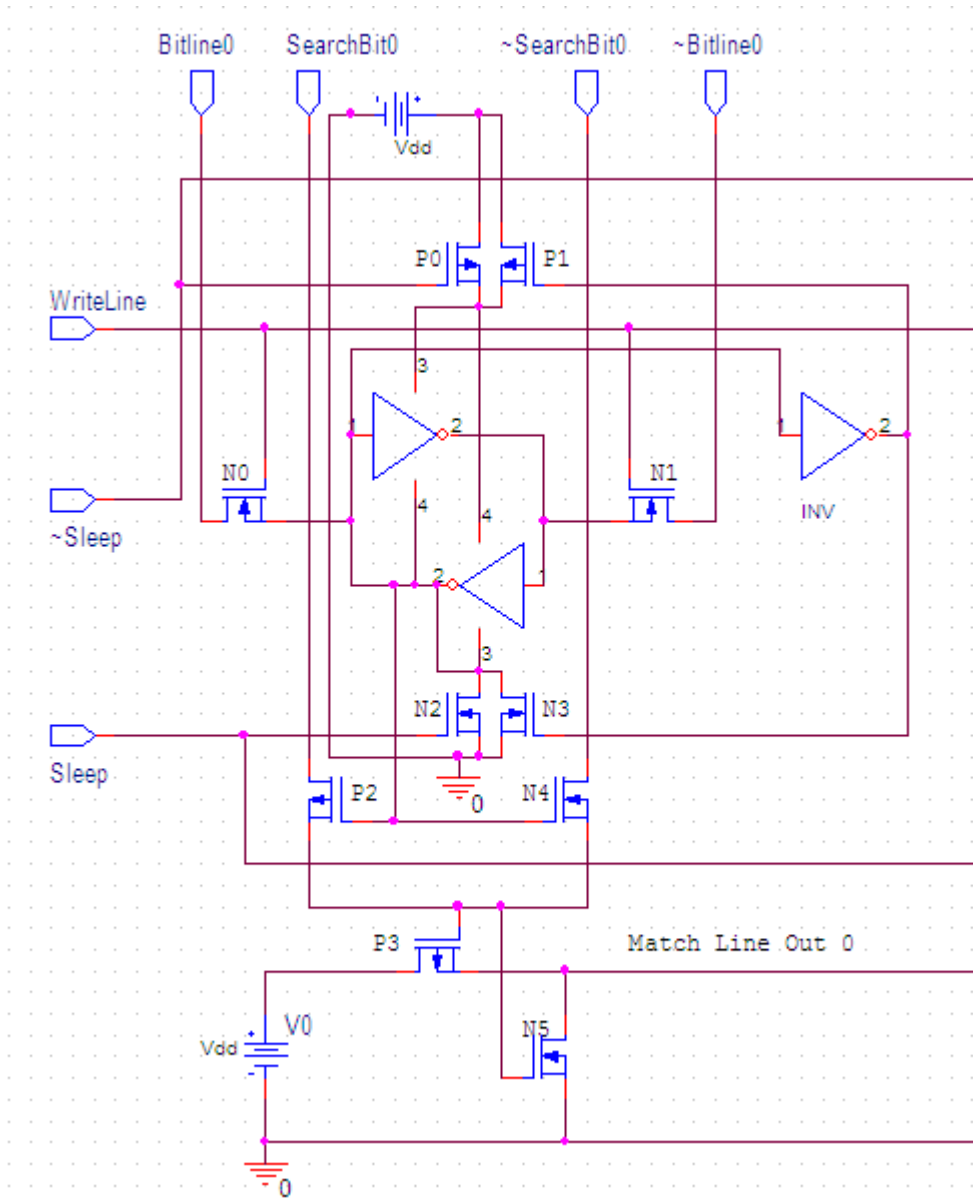


Figure 2.1: Modified Proposed CAM Cell Configuration

The CAM cell is designed to balance between comparison speed and power consumption. With this XOR comparison circuit, each comparison is delayed by two transistors (P2 & P3, P2 & N5, N4 & P3 or N4 & N5). With only two transistors delay, the configuration reduces the comparison time. Furthermore, the NAND type match line scheme is employed to reduce the power consumption. Additional transistors are added along with the match line to provide a gate to disconnect for a match or connect for a

mismatch. The searching operation is as follow: If a match occurs, the P3 transistor turns ON and connects the match line with the power source. On the other hand, the N5 transistor turns OFF to prevent a path from the power source to the ground. If a mismatch occurs, The P3 transistor turns OFF and disconnects the match line from the power source. The N5 transistor turns ON and connects the ground to the match line. Therefore, no path is created between the power source and the ground either for a match or a mismatch. In addition to a slightly modified comparison circuit, there is also a storage unit built-in for each memory bit to temporarily store the data bit during sleep mode. The sleep mode is added into the functionality of CAM architecture to provide further power reduction. For this work, the sleep mode is configured to be active low. Hence, sleep mode is enabled when it is 0 and disabled when it is 1. The sleep transistors (P0 and N2) are different from other transistors by their higher threshold voltage. Transistor with a higher threshold voltage is capable of reducing the amount of sub-threshold current while the CAM is inactive. While the CAM cell is inactive, the sleep transistors receive a signal to enter sleep mode. During sleep mode, both sleep transistors turn OFF to disconnect the inactive cell from the power source and the ground. Concurrently, the data stored within the memory is floating. The storage unit is used in this case to temporarily store the data. The storage unit composes of an inverter and two other transistors, the configuration is known as the Leakage Feedback Gate. The data within each cell is driving an inverter. The output of the inverter turns ON one of the transistors (P1 or N3). The turned ON transistor provides the minimum drive to circulate the data within the storage unit. For example, if the memory data is a 1, the output of the inverter would be a 0. This turns ON the P1 transistor that connects to the power source. The power source provides a

minimum drive to maintain the data and circulate through the inverter while other parts of the cell are inactive. Once the cell exits the sleep mode, the sleep transistors turns ON to provide power to the cell. The storage unit returns the stored data back to the cell.

The power saving configuration comes from the NAND type match line scheme. As described, the NAND type match line reduces power consumption tremendously. However, the NAND type match line still provides a path between the power source and the ground during a match. With the modified NAND type match line scheme, it eliminates the discharge path even during a match. With no path to discharge during a match or a mismatch, the total match line power consumption is reduced enormously.

2.2. Parallel Segmented Architecture

The proposed CAM architecture combines many of the previously researched architecture's advantages together and modified them to reduce the overhead it bears. The proposed CAM architecture is similar to the static divided word match line scheme, however, instead of being divided into two separate stages, there is an additional stage added. The idea is to parallelize the comparison process, yet maintain the high speed comparison while using a NAND type CAM match line scheme. To achieve parallelization, every bit, segment or group needs to be operated independently regardless of any value from the previous stage. The proposed CAM architecture is illustrated below. This idea utilizes the NAND type or the stacking effect of the transistor to reduce the overall power consumption. However, as mentioned in other literatures, the stacking effect greatly reduced the performance as the number of comparison increases. Hence, the idea of this new architecture is to segment or modulate the NAND type comparison.

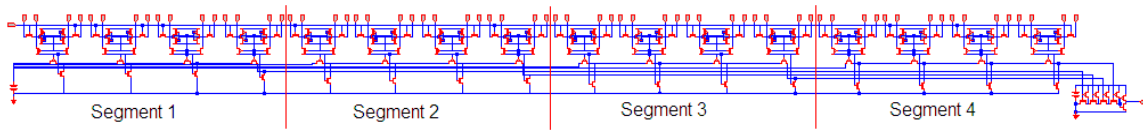


Figure 2.2: 16-Bit Memory Module Configuration

The 16 bits memory block is divided into four segments. Each segment contains four cells stacked in a NAND type comparison connection. If all of the cells are stacked, the comparison is very slow because the signal takes a longer time to travel from the first cell to the last cell. Hence, by dividing the memory block into four segments, each segment performs its comparison simultaneously and acts like an independent comparison block. Instead of waiting for 16 memory cells to compare and output the result, it would take approximately eight memory cells for comparison. Each segment compares four bits against its corresponding search bits. All four segments output their results to the second comparison units and perform the final comparison. For instance, as the search operation begins, four segments are divided from the 16 bits memory block, each contains four memory bits. Each segment compares their corresponding search bits, and generates a signal to indicate either all four bits match or mismatch. The output signals from all four segments gather at the intermediate comparison unit. The intermediate comparison unit further compares if any of the signals are mismatched. If no mismatch is detected, then the intermediate comparison unit produces a logic high signal to indicate all 16 bits match against the search data; otherwise, a logic low signal is issued to indicate a mismatch.

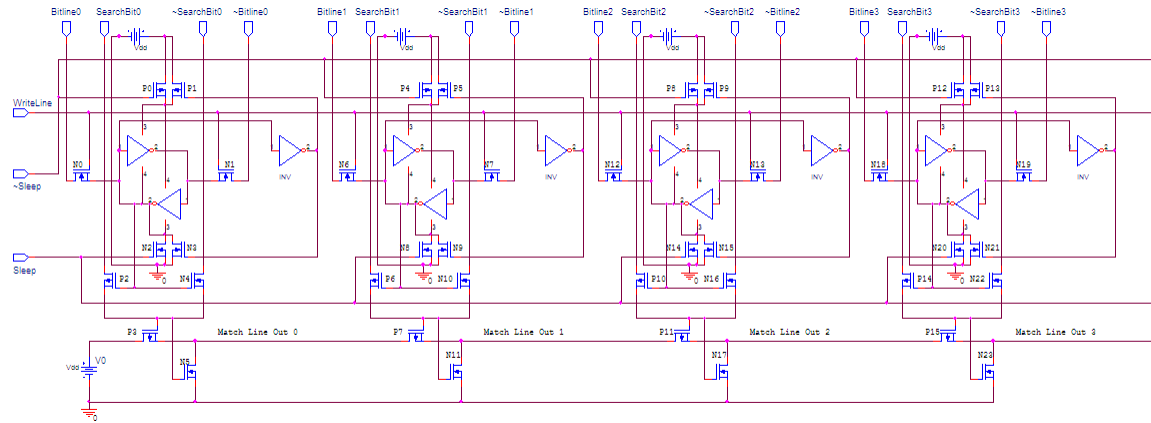


Figure 2.3: One Segment of the 16-Bit Memory Data

Figure 2.3 illustrates one segment from the 16 bits memory. As shown, there are only four comparisons with the results sent out to the intermediate comparison circuit shown in Figure 2.4. The intermediate comparison unit is constructed similarly to the NAND type match line scheme used for the memory cell. The difference is instead of providing a logic high when all are matched, it provides a logic low when all are matched. Therefore, it also provides power saving because of the stack effect configuration. An inverter is added to the intermediate unit to inverse the logic low to logic high signal and provides a stronger drive for the next stage. Moreover, the inverter helps to eliminate some switching activity if the current and the previous stage is the same.

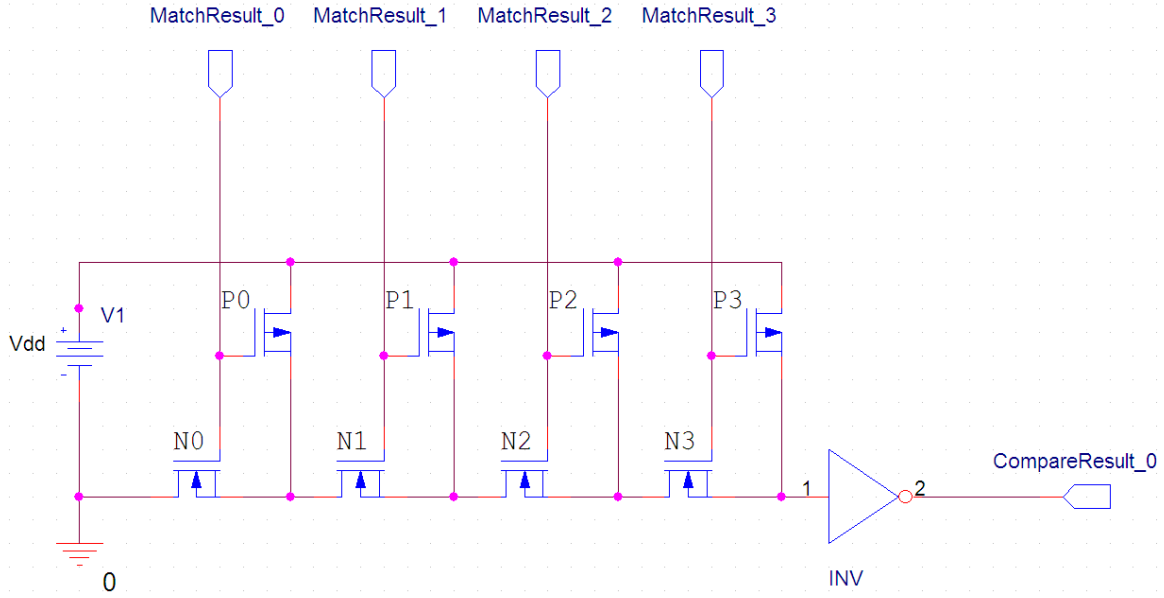


Figure 2.4: Match Line Computation Unit

For the 64 bits memory, it takes 64 memory bits comparisons in the traditional stacked comparison method. For this current method, it only takes about 12 memory bits comparisons. Divide the 64 bits into sixteen 4-bit independent memory segments, with each segment simultaneously comparing the search bits. This results in 16 signals from each of the 16 segments. From these 16 signals, further separate these signals into four groups and compare the signals in the intermediate comparison circuit. This results in four outputting signals. The final four outputting signals converge to a final comparison circuit to perform the last comparison for determining a match or a mismatch.

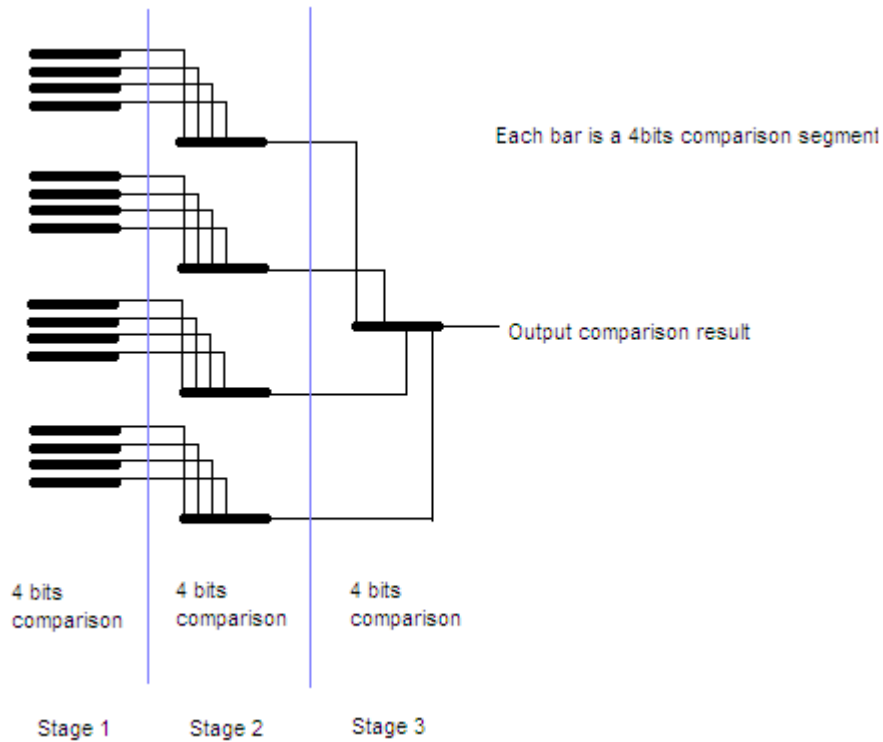


Figure 2.5: Overview of the Parallel Segmented Architecture for 64-bit Memory

Through this parallel segmented architecture, the 64 bit comparison can be greatly reduced its comparison speed, and with a NAND type match line scheme, static power dissipation can be reduced significantly.

2.3. Static and Dynamic Reduction

To further improve performance and reduce static power consumption, the dual threshold voltage technique is utilized on the architecture. The dual threshold voltage technique is a power reduction technique capable of reducing power dissipation and simultaneously improving performance. The concept is utilizing lower threshold voltage transistors in the critical path and higher threshold voltage transistors in the non-critical path. With a lower threshold voltage transistor, the delay of the transistor is reduced, and provides a faster response time. On the other hand, a higher threshold voltage transistor is

used in a non-critical or less switching environment because of its capability of reducing sub-threshold leakage current. Hence, the critical path of the circuits, specifically the match line, uses lower threshold voltage transistors to enhance the speed. Other transistors not part of the critical path, they would be used a higher threshold voltage transistor. Leakage current can easily flow through a lower threshold voltage transistor. Therefore, to reduce the amount of leakage current from going into the ground, a high threshold voltage transistor is used as a block. Furthermore, the Leakage Feedback Gate technique is also used to further reduce the static power dissipation. The purpose of the Leakage Feedback Gate technique is to reduce the leakage current while some of the circuits are not active. This technique becomes invisible while the circuit is active. When activated, it disconnects the power source and the ground line from the main circuit. The storage section turns on and preserves the stored data while the rest of the circuit is “sleeping”. When it is time to wake up, it reconnects the power source and the ground to activate the circuit, and restore the value before it went into sleep from the storage area.

The new architecture also provides some power and performance improvement. An independent segment reduces the amount of comparison units and dependency for each segment. This would eliminate extra circuitry to be added for the comparison, and reduce dynamic power consumption. Furthermore, performance increases due to the reduction in additional circuits.

Even though CAM has the capability of match indication, basic memory operations must be compliant with the matching speed. Traditionally, the voltage sensing amplifier increases the reading speed in any memory architecture. However, as the density of the memory module increases, wire capacitance and resistance also increase

because of the longer length. For this work, a current sensing amplifier is employed. The advantage of using a current sensing amplifier is its independency of wire length. As compared to a voltage sensing amplifier, the source voltage drops across the wire due to wire resistance. The voltage at the receiving point might be too low or too noisy to be detected. On the other hand, a current sensing amplifier detects the differential current instead of the differential voltage. As the current does not change regardless of the length of the wire, it eliminates the voltage drop problem seen in the voltage sensing amplifier. Furthermore, current sensing provides a higher detecting speed compared with voltage sensing because it only requires a small differential current to amplify the signal.

With the novel architecture and power reduction techniques employed into the memory module, the overall power consumption of the new CAM architecture does illustrate significant power reduction and performance improvement over previous research. The following chapter illustrated the simulation results of the design and comparison with other recent research results.

Chapter 3 Simulations and Results

The simulation is performed using Nanosim simulator by Mentor Graphic. This simulator not only provides SPICE level simulation for the BSIM4 65nm technology transistor, but its built-in power measurement features provide insight about the power consumption of the design. The simulator outputs the statistics of the design based on the configuration file. The simulation level is divided into seven different levels. The simulation results can be affected by using a different level of simulation, model, and net list. Overall, as the level increases, the more accurate the result becomes. However, the simulator consumed a lot more processing time as the level of accuracy increases. For the following simulations, all simulations are done using a simulation level of six, a modeling level of six, and a net list level of six. The highest level of accuracy is not needed because it is targeted for a small design, typically less than 1,000 elements. As for all models used in this work, all designs are well above 10,000 elements. Hence, using level six provides a good estimate of the model performance.

The size of the CAM used in the simulation is 64 bits x 64 bits. The transistors model is using the 65nm BSIM4 model card for bulk CMOS v0.0. The circuit configuration is written by HSPICE. HSPICE is used instead of other flavors of SPICE not because Nanosim is configured to use HSPICE, but because of its capability of simulating faster and producing more accurate results for a larger design. All simulations run the same test vectors, clock frequency, and under default temperature (25°C or 77°F) environment.

3.1. Comparison with Traditional CAM Architecture

The traditional CAM architecture is compared with the new proposed architecture. Both designs are performed under the same test vectors. One vector is simulating a scenario where a least amount of matches occurs with a large amount of idle period. This type of testing vector estimates the performance of the architecture for static power dissipation. The second scenario is where a large amount of matches occurs with small amounts of idle time. This type of testing vector estimates the performance of the architecture for dynamic power dissipation. The simulations of the traditional CAM architecture for both the nominal and maximum matching scheme are shown:

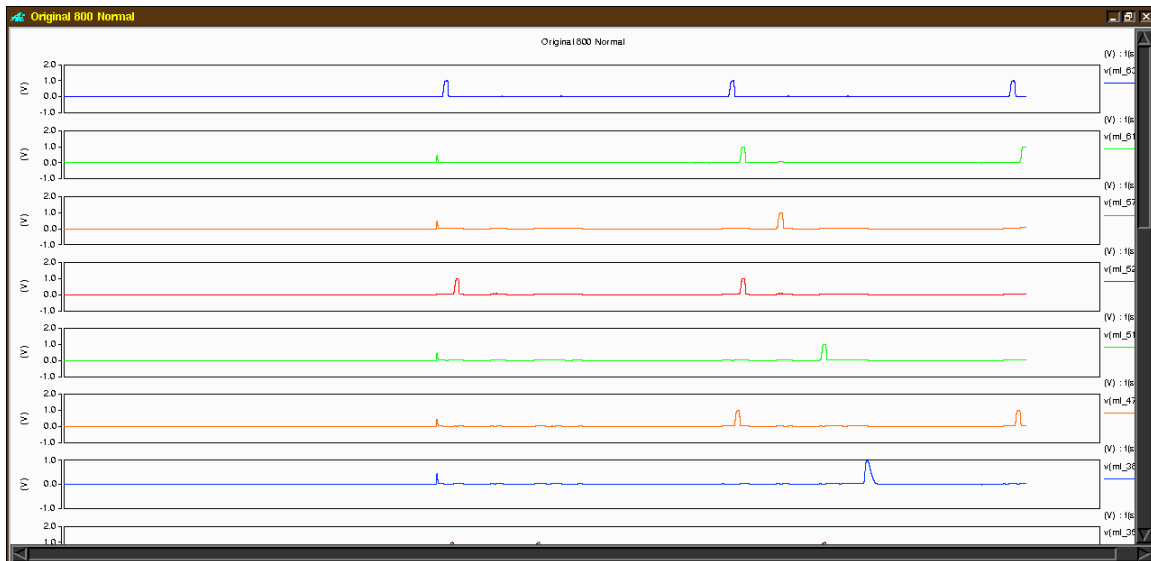


Figure 3.1: Traditional CAM Architecture Simulation Under Nominal Matching Condition

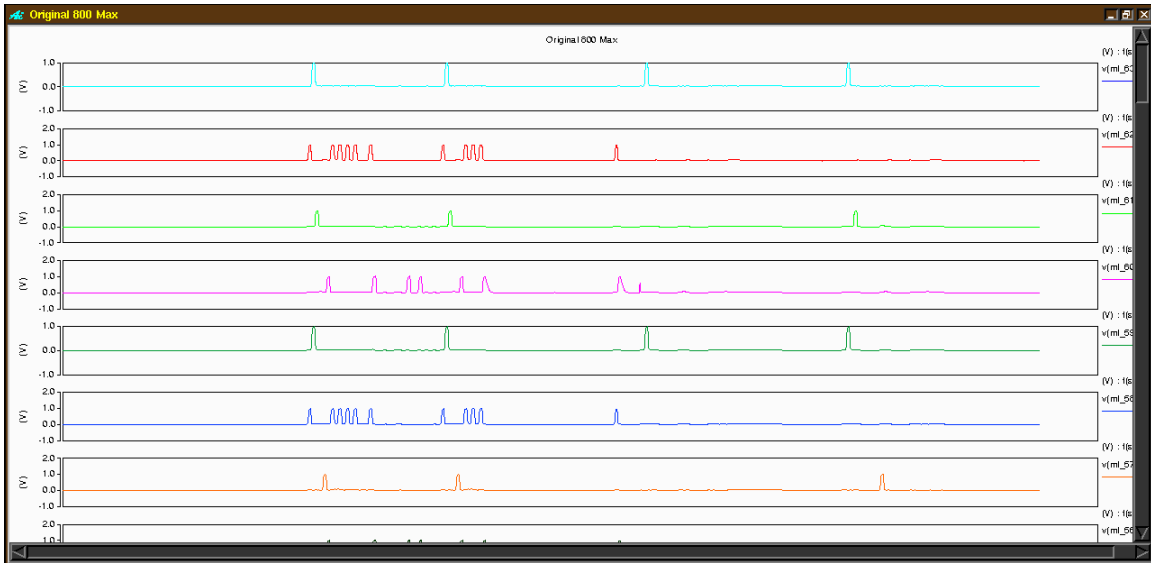


Figure 3.2: Traditional CAM Architecture Simulation Under Maximum Matching Condition

Each of the lines indicates the match line between address 0 to address 63 shown in both Figure 3.1 and 3.2. Noticeably, Figure 3.1 illustrates the CAM architecture with a lower number of matching indications compared to Figure 3.2. With more switching activities between 0 and 1 shown in Figure 3.2, it is expected that the dynamic power dissipation would be greater in the maximum matching condition than the nominal matching condition. Moreover, there are more match lines being idle or inactive in the nominal condition, therefore, expecting static power dissipation to increase due to sub-threshold current.

3.1.1 Power Consumption

After the simulations are run successfully, Nanosim generates a power dissipation report for the Traditional CAM and proposed CAM architecture. The power consumed in Traditional CAM architecture is illustrated below along with the proposed design.

Table 3.1: Power Comparison for Traditional CAM and Proposed CAM

Simulation: 65 nm Technology, CLK frequency = 800MHz, Nominal Matching Scenario			
Architecture	Total Static Wasted Power (μW)	Total Dynamic Wasted Power (μW)	Total Power (μW)
Traditional CAM	6777.913688	2392.273073	9170.186761
Proposed CAM	54.447974	1809.809307	1863.953981
Simulation: 65 nm Technology, CLK frequency = 800MHz, Maximum Matching Scenario			
Architecture	Total Static Wasted Power (μW)	Total Dynamic Wasted Power (μW)	Total Power (μW)
Traditional CAM	6972.03236	5566.558284	12538.59122
Proposed CAM	13.846385	4007.242598	4021.086448

Table 3.2: Dynamic Power Dissipation Breakdown for Traditional CAM Architecture

Simulation:65 nm Technology, CLK frequency = 800MHz, Nominal Matching Scenario			
Architecture	Dynamic Wasted Short-Circuited Power (μW)	Dynamic Wasted Transitional Power (μW)	Total Dynamic Wasted Power (μW)
Traditional CAM	1603.498885	788.774188	2392.273073
Proposed CAM	679.535563	1129.970444	1809.809307
Simulation: 65 nm Technology, CLK frequency = 800MHz, Maximum Matching Scenario			
Architecture	Dynamic Wasted Short-Circuited Power (μW)	Dynamic Wasted Transitional Power (μW)	Total Dynamic Wasted Power (μW)
Traditional CAM	3467.960754	2098.59753	5566.558284
Proposed CAM	1058.128829	2949.113769	4007.242598

Illustrated above, it is clear that simulated at the 65nm submicron level, the static wasted current is dominating the total power consumption. This is due to the shorter channel length causing leakage current to pass through even though the transistor is inactive. Moreover, the transistor might not completely turn ON or OFF with the shorter channel length. Therefore, this created a short circuit situation when both the NMOS and PMOS transistors were turned on at the same time. With reducing the static wasted current in mind, the proposed architecture significantly reduced the overall static current by almost 100 times. On the other hand, the total dynamic power dissipation is also reduced. The total dynamic power dissipated in the Traditional CAM architecture is the sum of both dynamic wasted short-circuited power and dynamic wasted transitional power. The Traditional CAM architecture consumed more in total dynamic power dissipation in both cases than the proposed CAM architecture.

For dynamic wasted transitional power, the Traditional CAM architecture utilized less power compared to the proposed CAM architecture. This may be due to the additional comparison circuitries within the new architecture where no intermediate comparison circuit is used in the Traditional CAM model. The proposed architecture reduces the short-circuited power significantly as well as the overall dynamic power dissipation.

3.1.2 Performance/Speed

Both designs are simulated utilizing the 800 MHz clock at 25°C. This only illustrate that both systems are able to operate at this frequency; it does not indicate the speed at which the memory responds to a look up sequence. The actual performance of the memory cell is determined by the time when the input signal is inputted to CAM until the time that the memory responds a match is found.

The match signal delays for both architectures are illustrated in the following table and figures.

Table 3.3: Performance Comparison for Traditional CAM and Proposed CAM

Architecture	Clock Speed (MHz)	Input signal to match signal delay (ps)
Traditional CAM	800	445.56
Proposed CAM	800	804.01

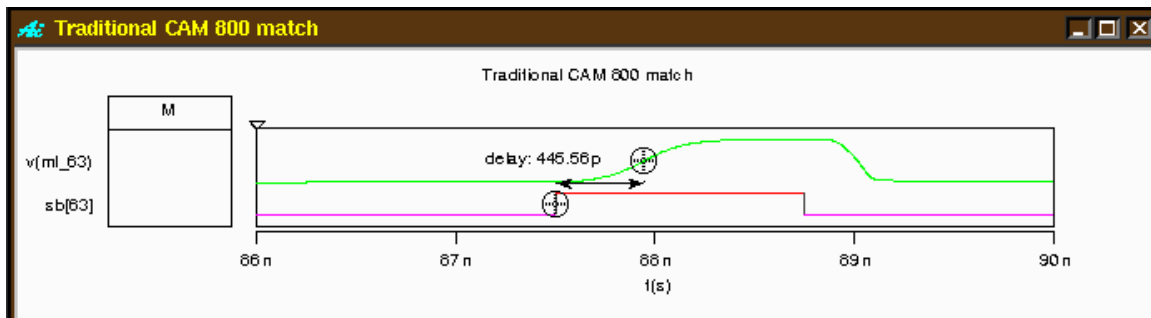


Figure 3.3: Simulation Measurement of Traditional CAM Output Delay

The speed of the traditional CAM is almost twice as fast as the proposed CAM. This is due to the NOR matching architecture, which provides a very fast comparison operation. However, a significant amount of static wasted current and short-circuited current is generated to achieve such fast comparison speed. Figure 3.4 illustrates the responds time when an input sequence is inputted into the CAM. In this case, since all

signals are inputted into the CAM at the same time, any input search bit signal can be used as the reference starting point for search operation. Using this, the delay time can be measured between the response time from the system and the reference starting time. This indicates the time the system takes to compare and determine whether a match exists within the memory matrix.

3.1.3 Area

The area of the memory can vary depending on the layout designer. However, one method of measuring the overall area is by measuring the total amount of transistors used in a design. For this, the following table summarizes the total number of transistors used in both the traditional CAM and proposed CAM architecture. The total number of transistors or nodes used in a design is generated along with the power dissipation report from Nanosim.

Table 3.4: Total Number of Transistors Used in Traditional and Proposed CAM

Architecture	Total number of transistors
Traditional CAM	39991
Proposed CAM	80519

The proposed CAM has almost three times as much transistor as the traditional architecture due to additional transistors added to the architecture to provide power reduction and signal rectification. Moreover, each cell in the proposed architecture utilizes a storage unit to store its data during sleep mode. Each of the storage units contributes the extra area for the new architecture. Sleep transistors and other gating

transistors are also used for power reduction. Hence, it is inevitable to have the overall area increase.

3.2. Comparison With Butterfly CAM Architecture

The static power dissipation of the Traditional CAM architecture dominates almost 75% of the overall power consumption. One type of power-saving architecture reducing the static current is the Butterfly CAM architecture. The Butterfly CAM architecture is simulated using the same test vectors, temperature, and clock speed as in the Traditional Architecture. Figure 3.4 illustrates the output waveform from the Butterfly CAM with nominal matching scenario and Figure 3.5 shows the output waveforms behaviors for maximum matching scenario.

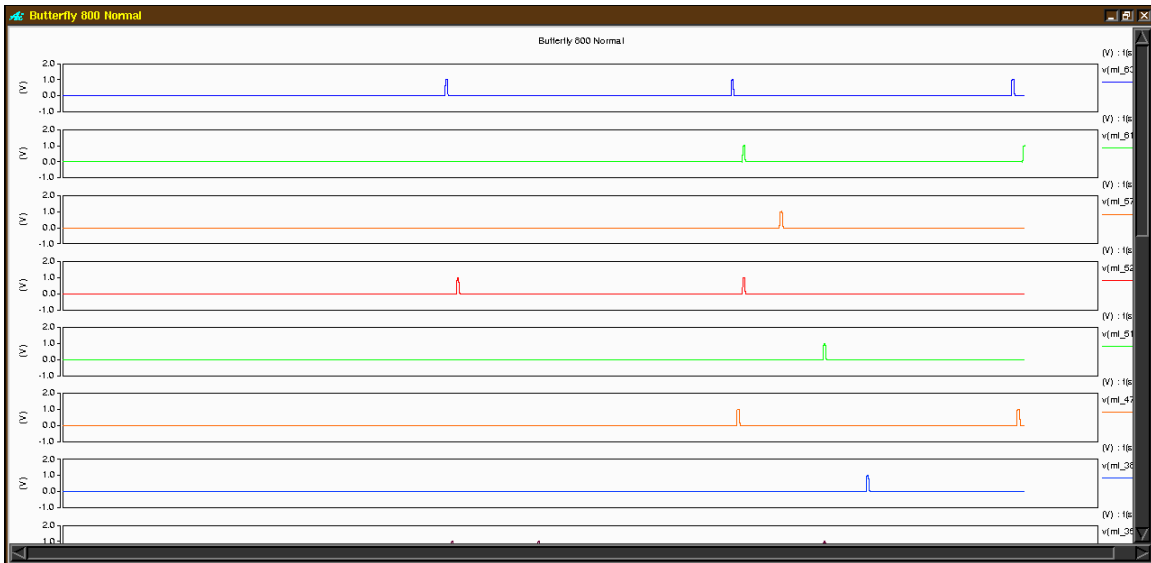


Figure 3.4: Butterfly CAM Architecture Simulation Under Nominal Matching Condition

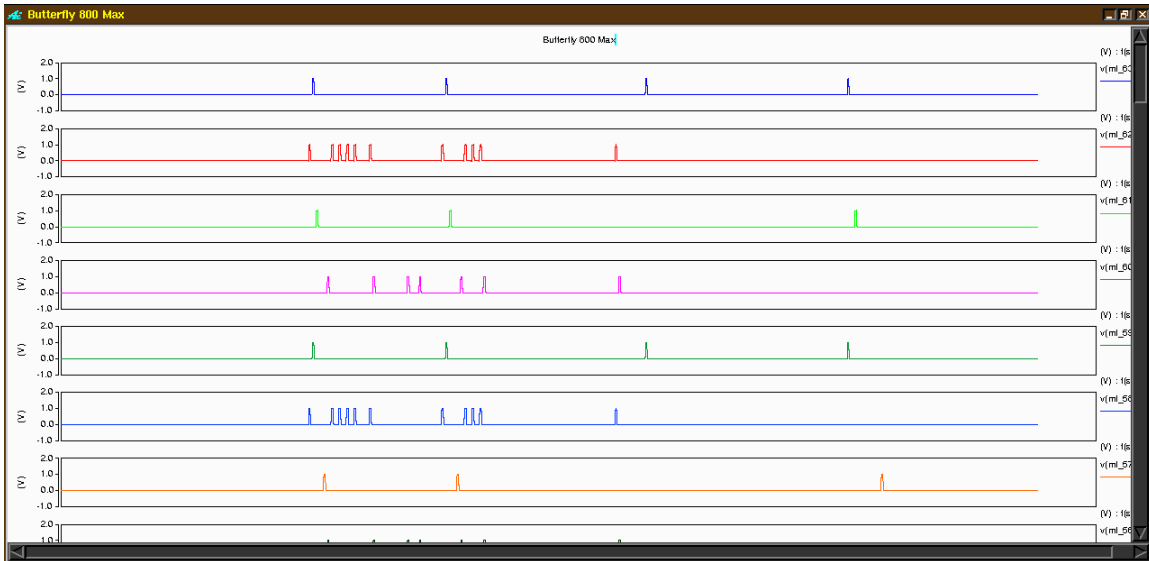


Figure 3.5: Butterfly CAM Architecture Simulation Under Maximum Matching Condition

The output waveform from the Butterfly CAM architecture is identical, if not similar, to the output waveform from the Traditional CAM architecture. This provides a verification check that the Butterfly CAM architecture is capable of identifying whether a match is found. Furthermore, the output waveform from the Butterfly CAM architecture illustrates a much cleaner match line output compared to the Traditional CAM architecture.

3.2.1 Power Consumption

The power consumption for the Butterfly CAM architecture compared to the proposed CAM architecture is illustrated below.

Table 3.5: Power Comparison for Butterfly CAM and Proposed CAM

Simulation: 65 nm Technology, CLK frequency = 800MHz, Nominal Matching Scenario			
Architecture	Total Static Wasted Power (μW)	Total Dynamic Wasted Power (μW)	Total Power (μW)
Butterfly CAM	1257.560433	8496.271087	9753.83152
Proposed CAM	54.447974	1809.809307	1863.953981
Simulation: 65 nm Technology, CLK frequency = 800MHz, Maximum Matching Scenario			
Architecture	Total Static Wasted Power (μW)	Total Dynamic Wasted Power (μW)	Total Power (μW)
Butterfly CAM	2499.480834	18272.155806	20771.63664
Proposed CAM	13.846385	4007.242598	4021.086448

Table 3.6: Dynamic Power Dissipation Breakdown for Butterfly CAM Architecture

Simulation: 65 nm Technology, CLK frequency = 800MHz, Nominal Matching Scenario			
Architecture	Dynamic Wasted Short-Circuited Power (μW)	Dynamic Wasted Transitional Power (μW)	Total Dynamic Wasted Power (μW)
Butterfly CAM	4944.956095	3551.314992	8496.271087
Proposed CAM	679.535563	1129.970444	1809.809307
Simulation: 65 nm Technology, CLK frequency = 800MHz, Maximum Matching Scenario			
Architecture	Dynamic Wasted Short-Circuited Power (μW)	Dynamic Wasted Transitional Power (μW)	Total Dynamic Wasted Power (μW)
Butterfly CAM	9898.494092	8373.661714	18272.155806
Proposed CAM	1058.128829	2949.113769	4007.242598

Shown in Table 3.5, the overall static wasted current is significantly reduced compared to the Traditional CAM architecture. However, the architecture does not provide enough protection against the dynamic transitional power. Illustrated in Table 3.6, the total dynamic wasted power is enormous. It consumed more power than the Traditional CAM architecture. The large amount of wasted dynamic power may be due to many additional intermediate units to perform signal comparison in order to determine whether next stage is activated. Dynamic wasted short-circuited power also contributed to the enormous overall dynamic wasted power. As described before, short-circuited power is when both P type and N type of transistors are turn ON and create a conductive path between the power source and the ground. The Butterfly CAM architecture may suffer the short-circuited power because of its fan-out effect. Referring to Figure 1.18(c), each stage needs to provide an output signal to drive the intermediate comparison unit. As the number of fan-out increases, the load capacitance increases. Therefore, the output signal from each stage might not have enough driving capability to provide the correct logic signal at the receiving side. The output signal might be a midpoint between 1 and 0, this unknown logic state cause both P-type and N-type transistors to be at their linear region. At the linear region, the transistor is able to conduct current between the drain and the source. As both P-type and N-type transistors are conducting current at the same time, it creates a short-circuited path between the power source and the ground. Therefore, short-circuited wasted power is generated and contributed to the overall dynamic power dissipation.

3.2.2 Performance/Speed

The performance of the Butterfly CAM architecture is measured and compared to the proposed CAM architecture. The result is illustrated in Table 3.7.

Table 3.7: Performance Comparison for Butterfly CAM and Proposed CAM

Architecture	Clock Speed (MHz)	Input signal to match signal delay (ps)
Butterfly CAM	800	1167.8
Proposed CAM	800	804.01

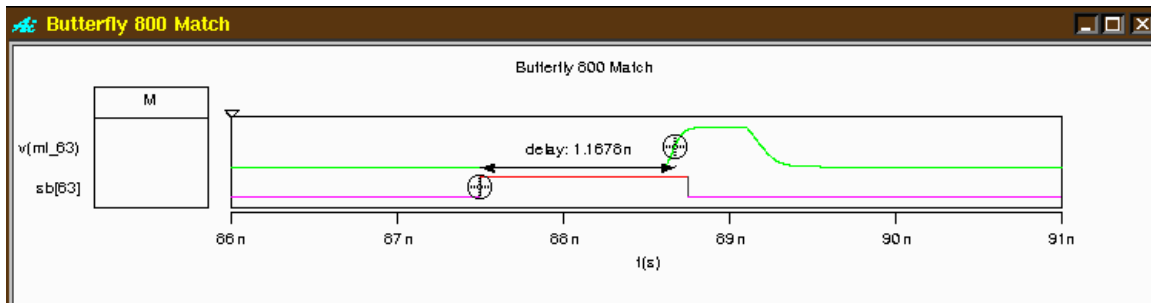


Figure 3.6: Simulation Measurement of Butterfly CAM Output Delay

Shown in Figure 3.6, the speed of the Butterfly CAM architecture is slower than the proposed CAM architecture and Traditional CAM architecture. The multiple stages and dependency check may contribute to the slow response time of the architecture. As discussed earlier, each stage required an enable or disable signal generated from the intermediate comparison unit, and each unit received the output signal from previous stages. This additional computation greatly reduces the overall comparison speed.

3.2.3 Area

The area is represented by the total amount of transistors used in the architecture.

The number of transistors used for the Butterfly CAM and proposed CAM is shown in Table 3.8.

Table 3.8: Total Number of Transistors Used in Butterfly and Proposed CAM

Architecture	Total number of transistors
Butterfly CAM	76627
Proposed CAM	80519

Shown in Table 3.8, the Butterfly CAM architecture utilized almost as many transistors as the proposed architecture. This greater number of transistors is contributed by the additional units that provide signal to the next stage in the architecture. Furthermore, additional buffering units were added for signal integrity and better driving capability.

3.3. Comparison with SDW Match Line Scheme Architecture

Static Divided Word Match Line architecture is designed to provide low power operation and fast comparison speed. The SDW CAM architecture is simulated using the identical test vectors, temperature, and clock frequency as in the other simulations. The output waveforms are generated to illustrate the validity of the design. The power dissipation, comparison performance, and area are to be compared with the proposed CAM architecture.

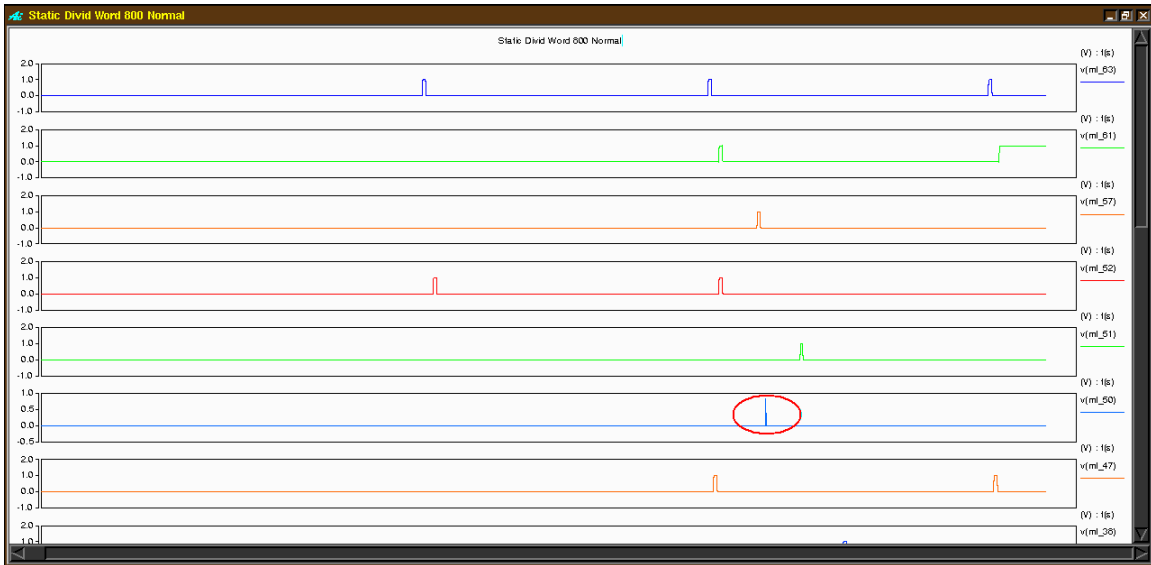


Figure 3.7: SDW CAM Architecture Simulation Under Nominal Matching Condition

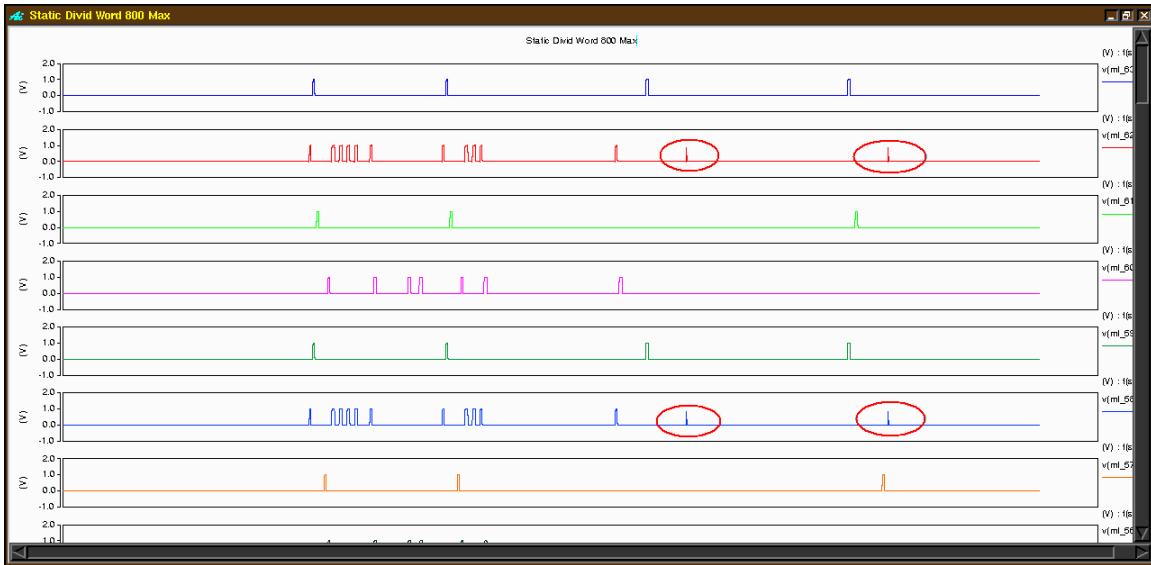


Figure 3.8: SDW CAM architecture simulation under maximum matching condition

Illustrated in Figure 3.7 and Figure 3.8, both simulations are identical with all other simulations using the same test vectors. However, there are glitches existing in both waveforms. In Figure 3.7, there is a very short glitch that appears, indicated by the red oval. Moreover, in Figure 3.8, there are four glitches that appear in the waveform. These glitches might be contributed by the different arrival time for some signals. There are computational units required multiple input signals in order to generate a signal result.

However, if one or more signals arrived differently, the output signal is incorrect because not all the input signals are of the same state. Therefore, it creates a small glitch as shown in the waveform simulation.

3.3.1 Power Consumption

The power consumption comparison for the SDW CAM and proposed CAM architecture is displayed in Table 3.9 and Table 3.10.

Table 3.9: Power Comparison for SDW CAM and Proposed CAM

Simulation: 65 nm Technology, CLK frequency = 800MHz, Nominal Matching Scenario			
Architecture	Total Static Wasted Power (μW)	Total Dynamic Wasted Power (μW)	Total Power (μW)
SDW CAM	282.743992	2273.687283	2556.431275
Proposed CAM	54.447974	1809.809307	1863.953981
Simulation: 65 nm Technology, CLK frequency = 800MHz, Maximum Matching Scenario			
Architecture	Total Static Wasted Power (μW)	Total Dynamic Wasted Power (μW)	Total Power (μW)
SDW CAM	390.258213	5066.565513	5456.823726
Proposed CAM	13.846385	4007.242598	4021.086448

Table 3.10: Dynamic Power Dissipation Breakdown for SDW CAM Architecture

Simulation: 65 nm Technology, CLK frequency = 800MHz, Nominal Matching Scenario			
Architecture	Dynamic Wasted Short-Circuited Power (μW)	Dynamic Wasted Transitional Power (μW)	Total Dynamic Wasted Power (μW)
SDW CAM	835.403742	1438.283541	2273.687283
Proposed CAM	679.535563	1129.970444	1809.809307
Simulation: 65 nm Technology, CLK frequency = 800MHz, Maximum Matching Scenario			
Architecture	Dynamic Wasted Short-Circuited Power (μW)	Dynamic Wasted Transitional Power (μW)	Total Dynamic Wasted Power (μW)
SDW CAM	1622.112447	3444.453066	5066.565513
Proposed CAM	1058.128829	2949.113769	4007.242598

The SDW architecture consumed almost the same amount of dynamic power as the proposed architecture. The reason for the overall lower power consumption is contributed significant by its simple and clean architecture design that does not utilized many additional computational units to compute signal for the next stages. However, as the architecture continues to utilize the NOR type match line architecture, the static power consumption continues to contribute to the overall power dissipation.

3.3.2 Performance/Speed

The performance of the SDW CAM architecture is measured and compared against the proposed CAM architecture. As discussed earlier, the SDW CAM architecture utilized a hybrid CAM cell match line scheme. The hybrid scheme mixed the use of a NAND type cell match line and a NOR type cell match line. The architecture is taking the advantage of the power saving NAND type match line, and the fast comparing speed of the NOR type match line. Table 3.11 illustrates the performance of the SDW match line architecture is not significantly impacted by incorporating the NAND type match line in the design.

Table 3.11: Performance Comparison for SDW CAM and Proposed CAM

Architecture	Clock Speed (MHz)	Input signal to match signal delay (ps)
SDW CAM	800	820.64
Proposed CAM	800	804.01

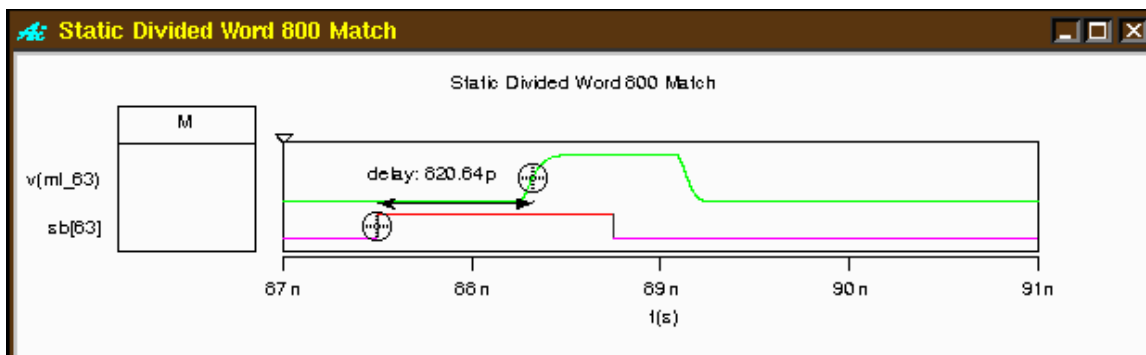


Figure 3.9: Simulation Measurement of SDW Match Line CAM Output Delay

The SDW CAM architecture provided minimum performance tradeoff because it utilized both the NAND and NOR type match line scheme. However, as the majority of the match comparison is performed using the NOR type match line configuration, the static leakage current would increase similar to the Traditional architecture.

3.3.3 Area

As mention earlier, the SDW architecture does not utilize many intermediate units to provide an enable or disable signal for the next stage. The extra transistors compared to the Traditional architecture are due to a slightly different sense amplifier and buffering units for signal integrity reason. Therefore, shown in Table 3.12, the total number of transistors used in the SDW CAM architecture are much less than the proposed CAM architecture.

Table 3.12: Total Number of Transistors Used in SDW and Proposed CAM

Architecture	Total number of transistors
SDW CAM	48453
Proposed CAM	80519

3.4. Proposed CAM Architecture Simulation

The proposed CAM architecture is designed to utilize the minimum amount of power to function while maintaining its high speed comparison performance. The proposed CAM is capable of operating at up to 1GHz speed, while other research designs were not able to operate at that speed. The major reason might be due to a long delay between intermediate signals or massive computation between units. Figures 3.11 through 3.14 below illustrate the simulation using the proposed architecture operating at 800MHz and 1GHz clock speed.

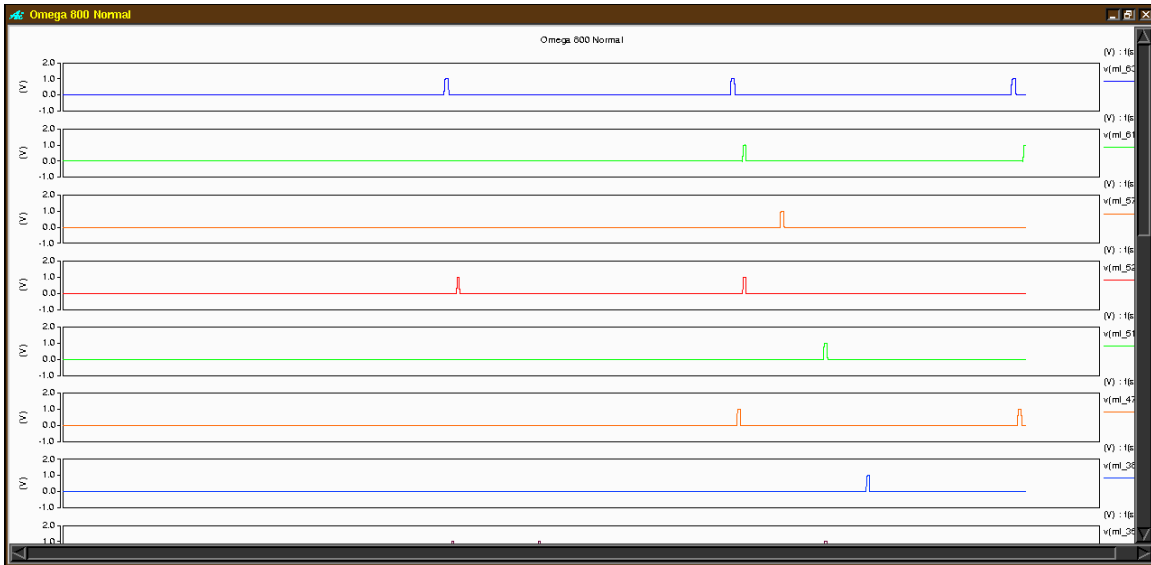


Figure 3.10: Proposed CAM Architecture Simulation Under Nominal Matching Condition at 800MHz

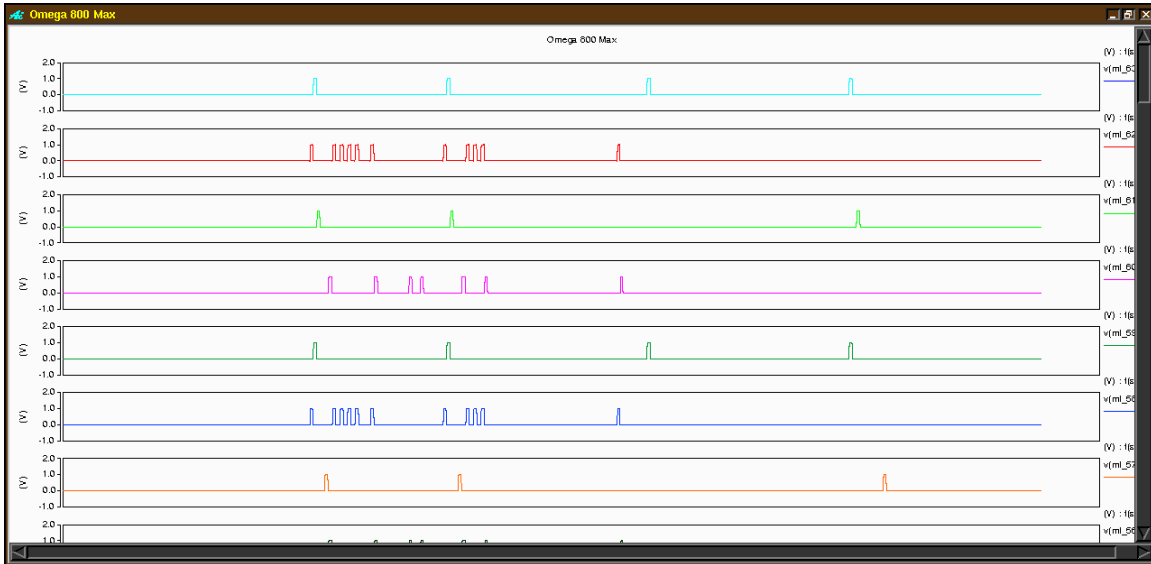


Figure 3.11: Proposed CAM Architecture Simulation Under Maximum Matching Condition at 800MHz

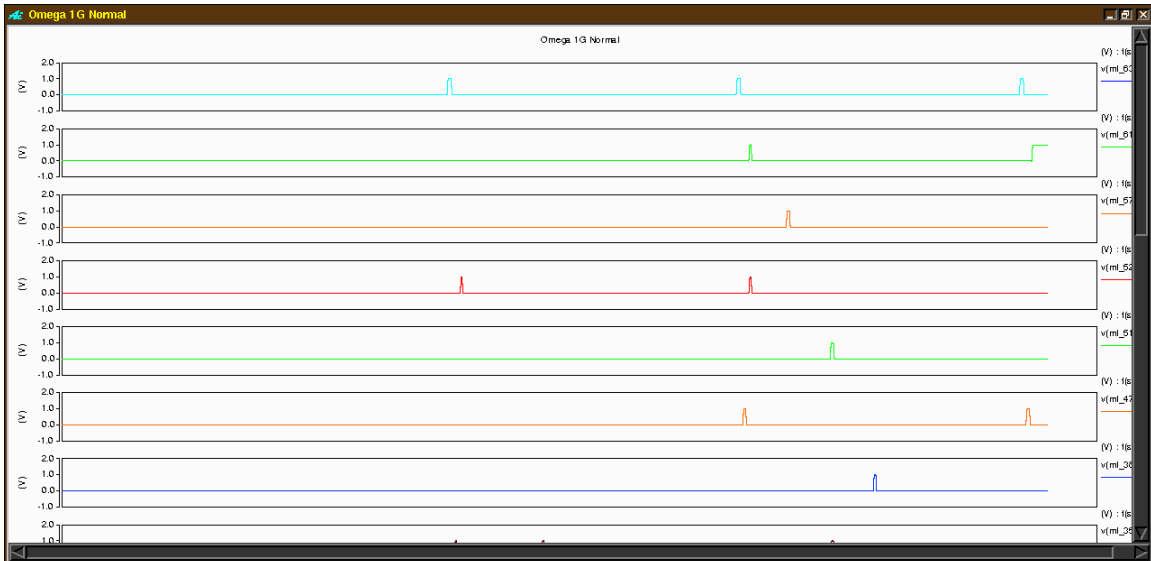


Figure 3.12: Proposed CAM Architecture Simulation Under Nominal Matching Condition at 1GHz

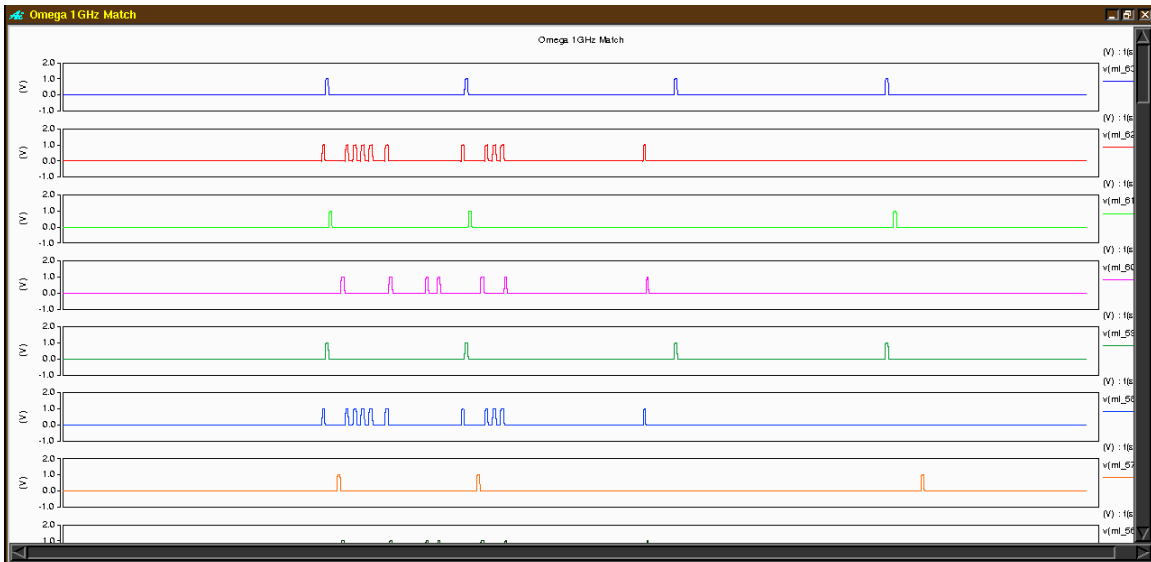


Figure 3.13: Proposed CAM Architecture Simulation Under Maximum Matching Condition at 1GHz

As shown above, the proposed CAM architecture produces identical output waveform when simulated using the same vectors with a different clock frequency. This shows that even when operating at a different frequency, the CAM produces the same result.

3.4.1 Power Consumption

The power consumption of the proposed CAM at 1GHz actually consumed less when simulated under maximum matching scheme. With a higher clock frequency, the time of inactivity or idle time is shorter; hence, leading to a reduction in static power dissipation. On the other hand, with a higher clock speed, it increases the dynamic transitional power dissipation as there are more switching activities. Table 3.13 and Table 3.14 summarize the static, dynamic and total average power consumption of the proposed CAM simulated at 1GHz clock frequency.

Table 3.13: Power Comparison for Proposed CAM at 1GHz Clock Speed

Simulation: 65 nm Technology, CLK frequency = 1GHz, Nominal Matching Scenario			
Architecture	Total Static Wasted Power (μW)	Total Dynamic Wasted Power (μW)	Total Power (μW)
Proposed CAM	0.138982	2120.545344	2120.684326
Simulation: 65 nm Technology, CLK frequency = 1GHz, Maximum Matching Scenario			
Architecture	Total Static Wasted Power (μW)	Total Dynamic Wasted Power (μW)	Total Power (μW)
Proposed CAM	19.927413	4721.877108	4741.804521

Table 3.14: Dynamic Power Dissipation Breakdown for Proposed CAM Architecture

Simulation: 65 nm Technology, CLK frequency = 1GHz, Nominal Matching Scenario			
Architecture	Dynamic Wasted Short-Circuited Power (μW)	Dynamic Wasted Transitional Power (μW)	Total Dynamic Wasted Power (μW)
Proposed CAM	739.534054	1381.01129	2120.545344
Simulation: 65 nm Technology, CLK frequency = 1GHz, Maximum Matching Scenario			
Architecture	Dynamic Wasted Short-Circuited Power (μW)	Dynamic Wasted Transitional Power (μW)	Total Dynamic Wasted Power (μW)
Proposed CAM	1131.762282	3590.114826	4721.877108

3.4.2 Performance/Speed

The performance of the proposed CAM cell remains relatively the same for both frequencies. There should not be a tremendous amount of difference in delay time with different clock speed. The proposed CAM architecture delay time between the input signal and the match response is about 774.01ps. Compared to the 804.01ps when simulated under 800MHz clock, there is only 30ps second of difference. The measurements are illustrated in Figure 3.14 and Figure 3.15, and summarized in Table 3.15.

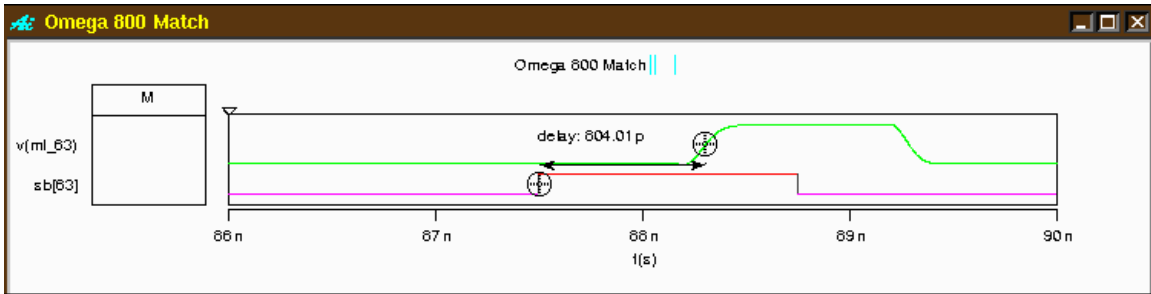


Figure 3.14: Simulation Measurement of Proposed CAM Output Delay at 800MHz Clock

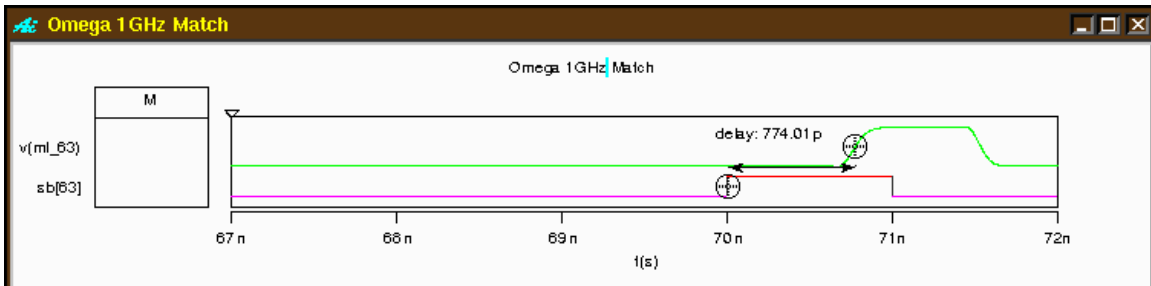


Figure 3.15: Simulation measurement of Proposed CAM output delay at 1GHz clock

Table 3.15: Performance Comparison for Proposed CAM at Different Frequency

Architecture	Clock Speed (MHz)	Input signal to match signal delay (ps)
Proposed CAM	800	804.01
Proposed CAM	1000	774.01

3.4.3 Area

The proposed CAM architecture is designed with high operating speed and power reduction in mind. The final design is capable of operating at 1GHz clock speed without adding or modifying existing units to achieve such operating speed. Therefore, the same amounts of transistors simulate the architecture at 800MHz or 1GHz clock speed.

3.5. Benchmark All Architectures

3.5.1 Power

The summary for the power consumption for all four types of architecture is illustrated below:

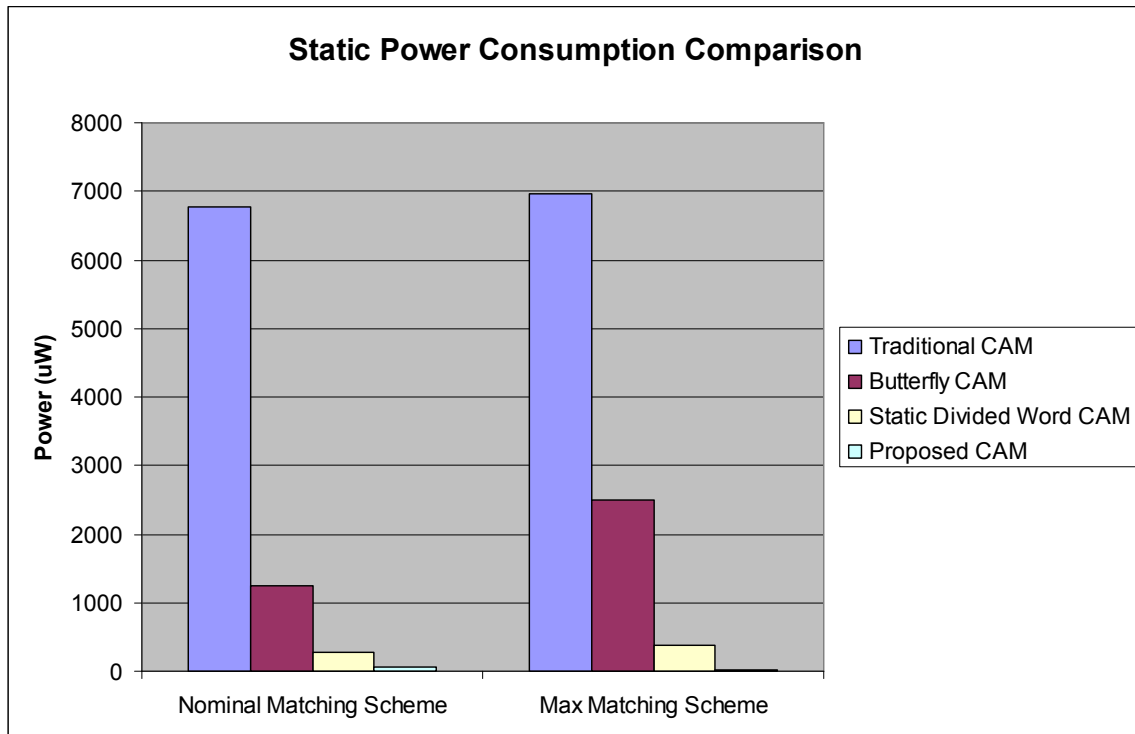


Figure 3.16: Summary of Static Power Consumption for All Architectures

The static power consumption is summarized in Figure 3.16, the proposed CAM architecture maintains as little static power dissipation in both nominal and maximum matching scheme.

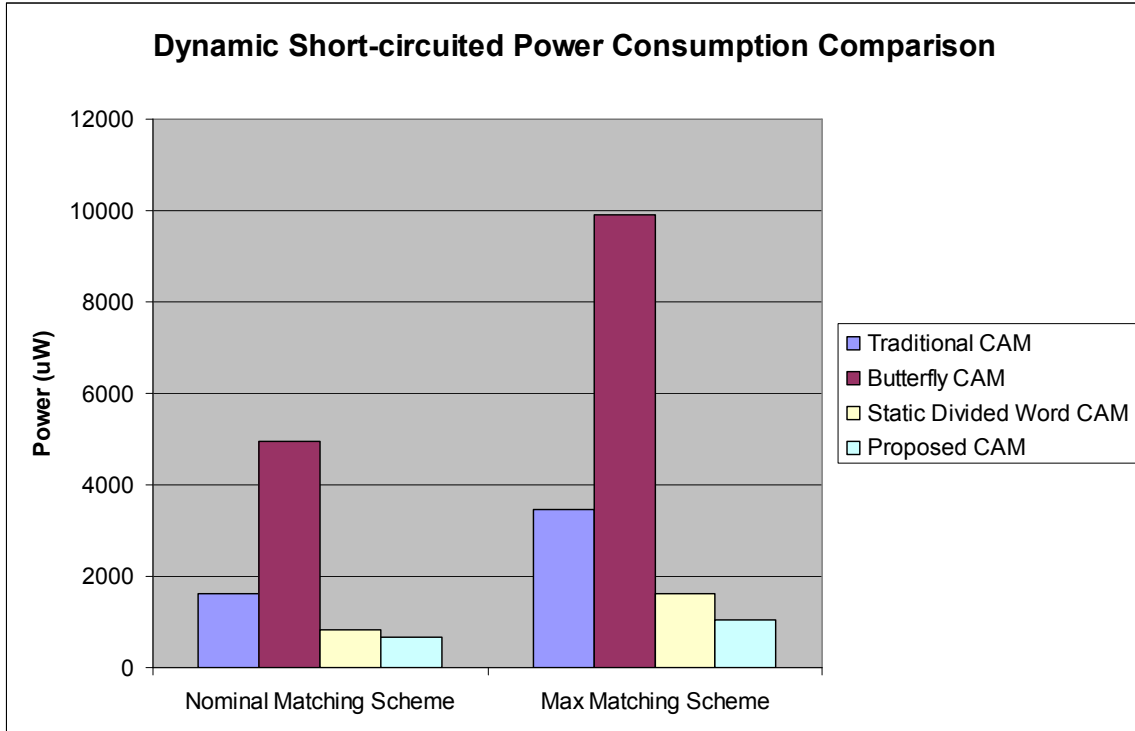


Figure 3.17: Summary of Dynamic Short-Circuited Power Consumption for All Architectures

The short-circuited power consumption of the proposed CAM architecture is minimized by reducing the load capacitance between stages and transistors size. This in turn diminishes the delay time, therefore reduces short-circuited power.

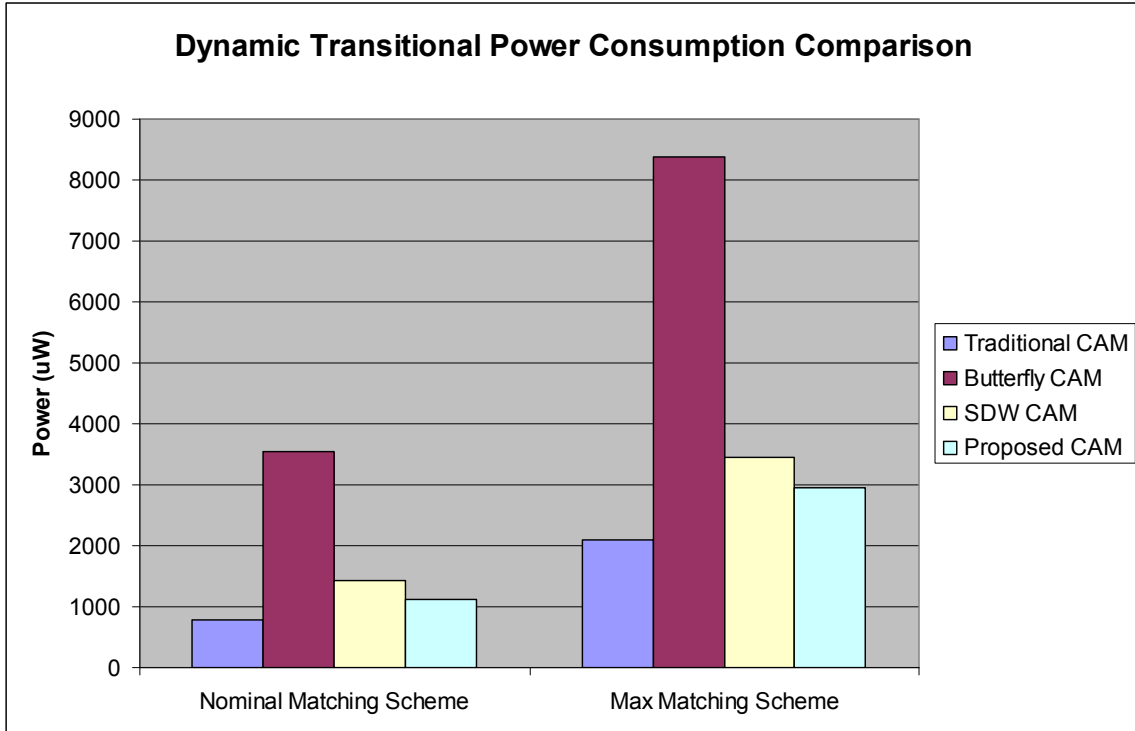


Figure 3.18: Summary of Dynamic Transitional Power Consumption for All Architectures

The dynamic transitional power dissipation, however, the proposed CAM architecture has a little increase of dynamic power consumption due to a larger number of switching activities in the maximum matching scheme.

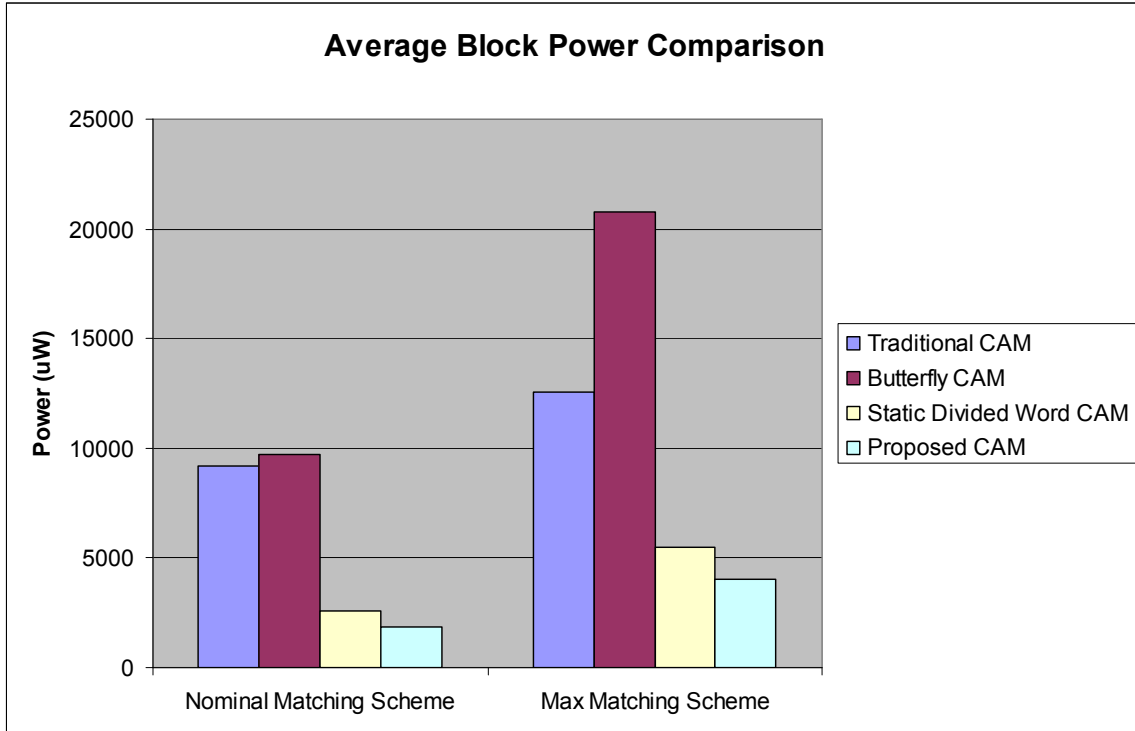


Figure 3.19: Summary of Average Block Power Consumption for All Architectures

The overall power dissipation of all architectures used in this work displays in Figure 3.19. The Traditional CAM architecture remained as one of the highest power consumed when simulated using 65nm technology. However, when simulated using the maximum matching scheme, the Butterfly CAM architecture consumed a greater amount of power due to its circuit complexity and frequently switching transistors.

3.5.2 Performance/Speed

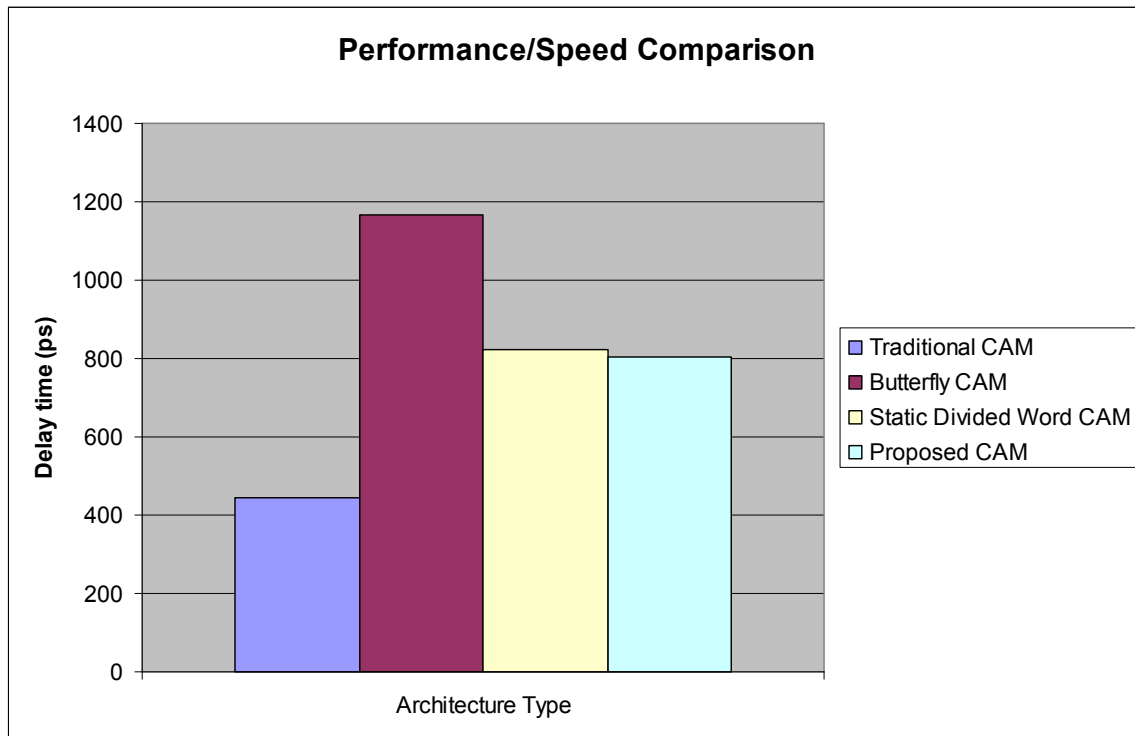
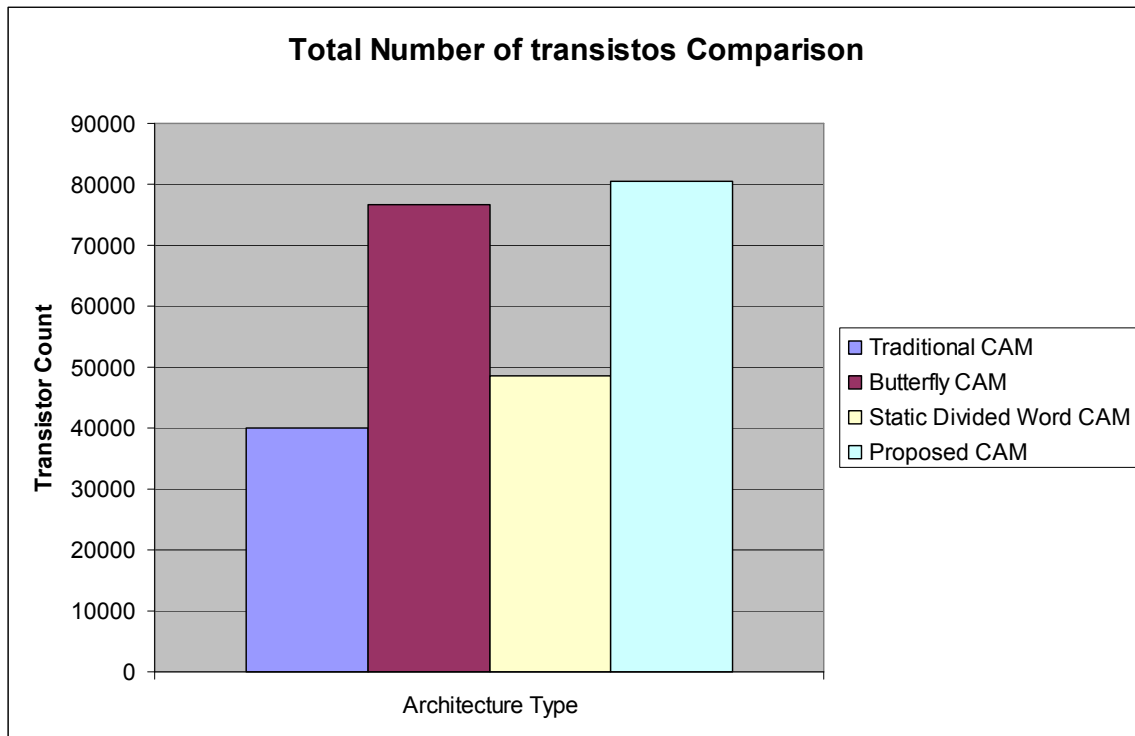


Figure 3.20: Summary of Speed Performance for All Architectures

Overall, all architectures are capable of operating with a clock frequency of 800MHz. However, with clock frequency of 1GHz, only the Traditional CAM and proposed CAM architecture are capable of operating. The SDW CAM architecture can only operate at 1GHz with some match lines incapable of producing a match signal because of internal structural delay.

3.5.3 Area



**Figure 3.21: Summary of the Total Number of Transistors
Used for All Architectures**

The drawback of the proposed CAM architecture is the overall area. It is larger than the Traditional CAM and other previously proposed architectures. This is inevitable because many different power reduction techniques had been implemented within the architecture.

Chapter 4 Conclusion

This work demonstrated that the proposed 64 x 64 bits CAM architecture consumed approximately 4mW as compared to the Traditional CAM architecture consuming approximately 12mW. Other proposed designs also demonstrated their advantages and disadvantages. Overall, the proposed architecture achieves the lowest total power consumption while maintaining the comparison performance. Furthermore, the proposed CAM architecture is capable of operating up to a 1GHz clock speed with minimal power increases. As the address and the data bits are increasing, expansion from 64 bits to 128 bits comparison is easily done by adding an additional comparison stage to the current architecture. Two additional comparison stages are added when expansion to 256 bits is desired.

Other previously proposed architectures provide many unique and alternative methods to reduce the power consumption. The Butterfly CAM architecture uses the butterfly connections in the comparison architecture. This method reveals that the static wasted current can be reduced dramatically. However, the dynamic power becomes dominant with extra circuitries added. The Static Divided Word Match Line CAM architecture combines both the NAND and the NOR type match line together to achieve power reduction and high speed comparison. Though the NOR type match line scheme continues to introduce wasted static current that consumed unnecessary energy. Nonetheless, the hybrid design is capable of reducing tremendous amount power consumption.

Though the proposed CAM architecture has reduced the overall power consumption to about $4\mu\text{W}$, but other units within the CAM has not been modified with low power techniques. It still provides much leeway for further power reduction.

Chapter 5 Future Work

For future work, there are ample of components within CAM that have not been explored, such as the sense amplifier, decoder, encoder and controlling unit. These components might not consume a large amount of power. However, with new configurations and designs for those units, CAM can be reduced even further compared to SRAM. Furthermore, an area reduction is needed for the proposed architecture to be used within a microprocessor. There is a limited amount of spaces allow for memory to be resided in the microprocessor, it is desired to have as much memory density as the area is allowed.

Bibliography

[1] Bellaouar, Abdellatif and ElmasryMohamed. Low-Power Digital VLSI Design: Circuits and Systems. Springer, 1st Edition, 1995.

[2] R. Ahmadi, "A Low Power Sense Amplifier Flip-Flop With Balanced Rise/Fall Delay," in *Electronics, Circuits and Systems, 2006. ICECS '06. 13th IEEE International Conference on*, 2006, pp. 1292-1295.

[3] R. G. Carvajal, J. Ramirez-Angulo, A. J. Lopez-Martin, A. Torralba, J. A. G. Galan, A. Carlosena, and F. M. Chavero, "The flipped voltage follower: a useful cell for low-voltage low-power circuit design," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 52, pp. 1276-1291, 2005.

[4] V. Chaudhary and L. T. Clark, "Low-power high-performance NAND match line content addressable memories," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 14, pp. 895-905, 2006.

[5] K. Dong-Sun, K. Jin-Tae, K. Ki-Won, and C. Duck-Jin, "Low power design using architecture and circuit level approaches," in *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on*, 2002, pp. 711-716 vol.2.

[6] A. Efthymiou and J. D. Garside, "A CAM with mixed serial-parallel comparison for use in low energy caches," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 12, pp. 325-329, 2004.

[7] P.-T. Huang, W.-K. Chang, and W. Hwang, "Low Power Pre-Comparison Scheme for NOR-Type 10T Content Addressable Memory," in *Circuits and Systems, 2006. APCCAS 2006. IEEE Asia Pacific Conference on*, 2006, pp. 1301-1304.

[8] H. Ilion Yi-Liang, W. Ding-Hao, and J. Chein-Wei, "Power modeling and low-power design of content addressable memories," in *Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on*, 2001, pp. 926-929 vol. 4.

[9] Sung-Mo Kang and Yusuf Leblebici, CMOS digital integrated circuits, Tata McGraw-Hill, 3rd edition-2003

[10] C. Kuo-Hsing, W. Chia-Hung, and J. Shu-Yu, "Static divided word matching line for low-power Content Addressable Memory design," in *Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on*, 2004, pp. II-629-32 Vol.2.

- [11] C. S. Lin, J. C. Chang, and B. D. Liu, "Design for low-power, low-cost, and high-reliability precomputation-based content-addressable memory," in *Circuits and Systems, 2002. APCCAS '02. 2002 Asia-Pacific Conference on*, 2002, pp. 319-324 vol.2.
- [12] N. Mohan, W. Fung, D. Wright, and M. Sachdev, "Match Line Sense Amplifiers with Positive Feedback for Low-Power Content Addressable Memories," in *Conference 2006, IEEE Custom Integrated Circuits*, 2006, pp. 297-300.
- [13] N. Mohan and M. Sachdev, "Low power dual matchline ternary content addressable memory," in *Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on*, 2004, pp. II-633-6 Vol.2.
- [14] A. Oruganti and N. Ranganathan, "Leakage power reduction in dual-Vdd and dual-Vth designs through probabilistic analysis of Vth variation," in *VLSI Design, 2006. Held jointly with 5th International Conference on Embedded Systems and Design., 19th International Conference on*, 2006, p. 4 pp.
- [15] K. Pagiamtzis and A. Sheikholeslami, "Pipelined match-lines and hierarchical search-lines for low-power content-addressable memories," in *Custom Integrated Circuits Conference, 2003. Proceedings of the IEEE 2003*, 2003, pp. 383-386.
- [16] K. Pagiamtzis and A. Sheikholeslami, "A low-power content-addressable memory (CAM) using pipelined hierarchical search scheme," *Solid-State Circuits, IEEE Journal of*, vol. 39, pp. 1512-1519, 2004.
- [17] K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory (CAM) circuits and architectures: a tutorial and survey," *Solid-State Circuits, IEEE Journal of*, vol. 41, pp. 712-727, 2006.
- [18] J. C. Park and V. J. Mooney Iii, "Sleepy Stack Leakage Reduction," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 14, pp. 1250-1263, 2006.
- [19] H. Po-Tsang, C. Shu-Wei, L. Wen-Yen, and H. Wei, "A 256 \times 128 Energy-Efficient TCAM with Novel Low Power Schemes," in *VLSI Design, Automation and Test, 2007. VLSI-DAT 2007. International Symposium on*, 2007, pp. 1-4.
- [20] K. Sami and M. Yehia, "Supply Voltage Adaptive Low-Power Circuit Design," in *Design, Applications, Integration and Software, 2006 IEEE Dallas/CAS Workshop on*, 2006, pp. 131-134.
- [21] K. Se Hun and V. J. Mooney, "Sleepy Keeper: a New Approach to Low-leakage Power VLSI Design," in *Very Large Scale Integration, 2006 IFIP International Conference on*, 2006, pp. 367-372.

[22] S. Sundaram, P. Elakkumanan, and R. Sridhar, "High speed robust current sense amplifier for nanoscale memories: a winner take all approach," in *VLSI Design, 2006. Held jointly with 5th International Conference on Embedded Systems and Design., 19th International Conference on*, 2006, p. 6 pp.

[23] Y. Taur and T. H. Ning, *Fundamentals of modern VLSI Devices*. New York: Cambridge Univ. Press, 1998, Ch-2, PP 94-95

[24] T. Yamagata, M. Mihara, T. Hamamoto, Y. Murai, T. Kobayashi, M. Yamada, and H. Ozaki, "A 288-kb fully parallel content addressable memory using a stacked-capacitor cell structure," *Solid-State Circuits, IEEE Journal of*, vol. 27, pp. 1927-1933, 1992.

[25] J. Kao and A. Chandrakasan, "MTCMOS sequential circuits," in *Solid-State Circuits Conference, 2001. ESSCIRC 2001. Proceedings of the 27th European*, 2001, pp. 317-320.