

AN ANALOG CONTENT ADDRESSABLE

MEMORY

By

SHAHRIAR ROKHSAZ

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1990

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1992

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 1998

1998

1998

1998

1998

1998

1998

1998

1998

1998

1998

1998

1998

1998

1998

Thesis
1998D
R742a

AN ANALOG CONTENT ADDRESSABLE

MEMORY

Thesis Approved:



Thesis Advisor



Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to my major adviser, Dr. Louis G. Johnson for his outstanding supervision, guidance, inspiration and friendship. My sincere appreciation extends to my other committee members, Dr. Chriswell Hutchens, Dr. Keith Teague and Dr. Jack Cartinhoure whose guidance, assistance and encouragement were also invaluable.

I would like to give my special appreciation to my brother, F. Rokhsaz, my mother S. Z. Mirfakhraii and my father M. Rokhsaz for their help, support and guidance throughout my live.

Finally, I would like to extend my special appreciation to my friend, companion and wife, Parisa Lamei for her precious suggestions to my research, her strong encouragement at the time of difficulty and understanding through this whole journey.

TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION	1
Neural Network Classifier	1
Chapter Description	2
II. LITERATURE SURVEY	4
Introduction	4
Mixed Mode Approach	4
All Analog Approach	23
Floating Gate Structure	27
III. ANALOG CAM	33
Introduction	33
A Pre-charge Based Content Addressable Memory	34
Large Signal Analysis	37
Closed-Form Solution	38
IV. THE LEARNING PROCESS	56
Introduction	56
Architecture	56
Programming Process	59
Controller	65
Initial Clocking	73
More Mismatch	74
V. THE HIGH VOLTAGE CIRCUIT	77
Introduction	77
High Voltage Transistor	78
A Complete High Voltage Transistor	80
Silicon Results	83

CHAPTER	PAGE
Field Transistor's Threshold Voltage	84
Layout Precautions	86
High Voltage Driver	88
 VI. THE WINNER TAKE ALL CIRCUIT	 92
Introduction	92
Winner Take All	93
WTA's Closed-Form Solution	97
Evaluation Mode	108
Closed-Form Versus Simulation	111
 VII. SILICON RESULTS	 116
Introduction	116
Silicon Results	116
 VIII. CONCLUSION	 123
 REFERENCES	 126
 APPENDIX A	 131
 APPENDIX B	 143

LIST OF FIGURES

FIGURE	PAGE
II-1. A Self-Organized Neural Network with an Analog Winner Take All circuit	6
II.2. A Charge Based Hamming Classifier	10
II.3. A Capacitive Neural Network	16
II.4. A Capacitive Comparator circuit	19
II.5. The WTA circuit	19
II.6. The Relaxative Content Addressable Memory (RCAM)	21
II.7. The RCAM and the WTA network model	22
II.8. The Two-Quadrant Multiplier	24
II.9. The Differential Pair Two-Quadrant programmable analog Multiplier circuit	25
II.10. The simplified EEPROM Model	28
II.11. The differential floating gate synapse	32
III.1. The Charge Base Analog Content Addressable Memory (CAM)	34
III.2. An Alternative Configuration of the Analog CAM	36
III.3. The CAM with parasitics	39
III.4. Large Signal Model of the Analog CAM	43
III.5. CAM's Output VS. Input Voltage ($V_{ref}=1.5$ Volts)	51

FIGURE	PAGE
III.6. CAM's Output VS. Input Voltage ($V_{ref}=2$ Volts)	52
III.7. CAM's Output VS. Input Voltage ($V_{ref}=2.5$ Volts)	52
III.8. CAM's Output VS. Input Voltage ($V_{ref}=3$ Volts)	53
III.9. CAM's Output VS. Input Voltage ($V_{ref}=3.5$ Volts)	53
III.10. CAM's Output VS. Input Voltage ($V_{ref}=4$ Volts)	54
III.11. The Analog CAM with Analog Memory incorporated	55
IV.1. The Recognition Engine	58
IV.2 The Winner Take All Circuit	61
IV.3. An equivalent engine's diagram while programming a single CAM	62
IV.4. The Analog CAM under program	63
IV.5. CAM's Output voltage and the corresponding match line voltage	64
IV.6. The Recognition Engine and the Controlling Circuitry in the programming mode	66
IV.7. The Programming Pulse Controller	69
IV.8. Clocking Sequence for programming each pixel	70
IV.9. The Sampler	71
IV.10. The Complete Clocking Scheme	74
IV.11. CAM's output vs. input voltage	75
V.1. An Extended Drain Transistor	79
V.2. A Complete High Voltage Transistor (Field Transistor)	81
V.3. Top View Layout of the complete High Voltage Transistor	82
V.4. The Current vs. Drain to Source Voltage Potential Curve	83

FIGURE	PAGE
V.5. Oscilloscope plot of the Drain to Source Current vs. the Gate to Source Voltage Potential	85
V.6. Top View of the High Voltage Transistor With Ground Shield	87
V.7. A High Voltage Driver	88
V.8. An n-channel transistor incorporated with an EEPROM cell	90
VI.1. The WTA circuit in the evaluation mode	94
VI.2. Theory vs. Simulation for the winner Match Line voltage	106
VI.3. Theory vs. Simulation for the Drain to source voltage of the Feed Back Transistor	106
VI.4. Theory vs. Simulation for the loser match line voltage	107
VI.5. The Analog Recognition Engine in the Evaluation mode	110
VI.6. Input Voltage vs. Match Line Voltage ($V_{ref}=1.5$ Volts)	112
VI.7. Input Voltage vs. Match Line Voltage ($V_{ref}=2$ Volts)	113
VI.8. Input Voltage vs. Match Line Voltage ($V_{ref}=2.5$ Volts)	113
VI.9. Input Voltage vs. Match Line Voltage ($V_{ref}=3$ Volts)	114
VI.10. Input Voltage vs. Match Line Voltage ($V_{ref}=3.5$ Volts)	114
VI.11. Input Voltage vs. Match Line Voltage ($V_{ref}=4$ Volts)	115
VI.12. Comparison between the winner and the loser match lines ($V_{ref}=2.5$ Volts)	115
VII.1. Fabricated 2x2 recognition engine	118
VII.2. Winner match line voltage ($V_{ref_{00}} = 1.5$ Volts)	120
VII.3. Winner match line voltage ($V_{ref_{00}} = 2$ Volts)	120
VII.4. Winner match line voltage ($V_{ref_{00}} = 2.5$ Volts)	121

FIGURE	PAGE
VII.5. Winner match line voltage ($V_{ref00} = 3$ Volts)	121
VII.6. Winner match line voltage ($V_{ref00} = 3.5$ Volts)	122

CHAPTER I

INTRODUCTION

Neural Network Classifier

Neural networks are parallel interconnected systems consisting of many simple processing units which interact among one other via links with adaptive weights. There are a variety of models proposed to approximate the neural network concept. Each model defines a learning process and its accuracy. The simplest yet most powerful model is implemented by a competitive classifier. This classifier is divided into two subnets, a lower subnet and an upper subnet. The lower subnet (qualifier) defines the correlation between the input and the stored patterns by evaluating each exemplar vector versus the input pattern and determining the number of mismatches encountered. The upper subnet is the competition layer selecting the reference pattern with the highest correlation (or highest score). This layer can be implemented by a winner take all (WTA) or any other discriminator type architecture capable of selecting the reference pattern which most closely matches the input.

The objective of this thesis is to propose a low power and high speed competitive classifier architecture. This classifier is process independent, reliable and manufacturable in a low cost fabrication process.

To pursue this objective, a novel low power qualifier cell is introduced and fully

characterized followed by a detailed discussion on the learning process. For this learning process, the system is required to handle high voltage; therefore, a novel high voltage driver is proposed and developed. Furthermore, a discriminator architecture [Jalaleddine, 1992] is presented and fully analyzed when used in conjunction with the proposed qualifier subnet. Then, a 2x2 recognition engine is manufactured in the 2.0u ORBIT process and fully characterized. The characterization includes, programming each cell and evaluating the input pattern. Finally, the steady state results found using simulation and the closed-form solution are compared against the silicon results.

Chapter Description

Chapter II covers a variety of existing techniques to implement a neural network classifier.

Chapter III introduces and analyzes a novel content addressable memory (CAM) structure. The analysis includes finding a closed-form solution for the CAM cell and comparing them to simulation results.

Chapter IV describes the learning process for the content addressable memory while examining the effects of a non-ideal CAM on the learning process.

Chapter V introduces a high voltage transistor which constitutes the basis of a proposed high voltage driver.

Chapter VI investigates the behavior of the Winner Take All circuit (WTA) used in conjunction with the proposed CAM. This includes finding a closed-form solution for the WTA circuit and relating its output to the input of the CAM cell.

Chapter VII contains the data from the fabricated silicon and compares the closed-form, simulation and the silicon's results.

Chapter VIII discusses the goals achieved in this thesis.

CHAPTER II

LITERATURE SURVEY

Introduction

Due to the number of the neurons and their required connections, research on neural net implementation has been evolved with key concepts such as area, learning efficiency, resolution, and power consumption. Traditionally, one way to reduce area and total power consumption is to implement most of the architecture using analog circuitry; so that, a large number of components can be monotonically integrated [Joongho, May, 1993]. To further reduce the power consumption, purely capacitive methods were introduced [Ugur, Feb. 1991, Jan., Mar., May, 1993]. These methods use precharge and charge sharing concepts in order to manipulate information. Finally, to allow longer storage ability and lower total power consumption, a combination of analog circuitry and analog memory is suggested. This section introduces these different architectures while briefly discussing their pros and cons.

Mixed Mode Approach

As mentioned a competitive classifier architecture is expected to identify the closest match with respect to the input pattern. This is accomplished in two stages. First

the qualifier indicates how close the input pattern is to the stored patterns, then based on the qualifier's result the discriminator determines the winner.

Figure II-1 depicts a possible hardware implementation of a competitive classifier architecture [Joongho, May, 1993].

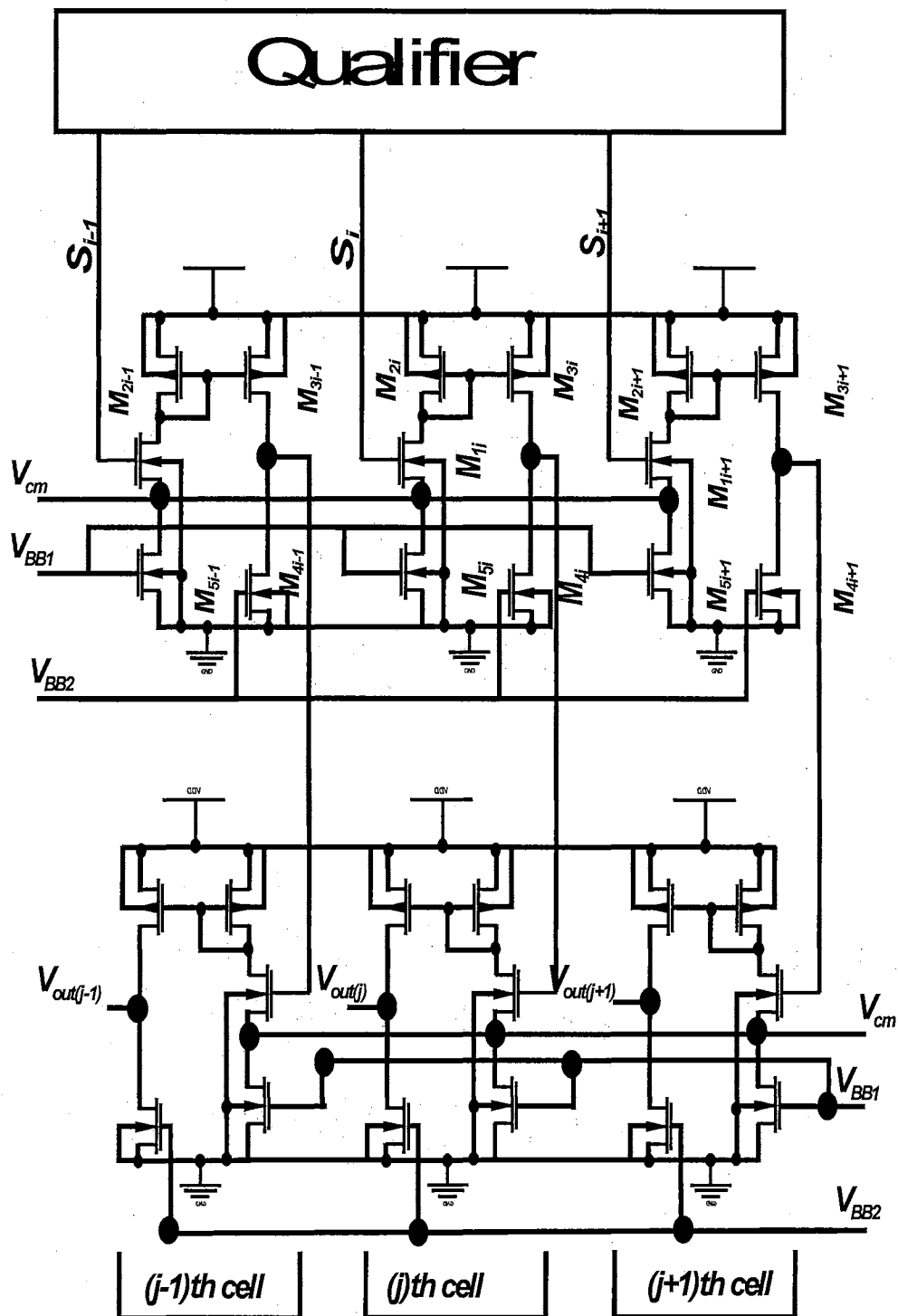


Figure II-1. A Self-Organized Neural Network with an Analog Winner Take All circuit

The qualifier determines the score between the synapse weights (stored pattern) and the input signals as follow [Ugur, Jan., 1991]

$$S_i = A' - A \cdot \sum_{j=1}^m (x_{ij} - T_{ij})^2 \quad \text{Equation II-2}$$

$$S_i = A' - A \cdot \sum_{j=1}^m x_{ij}^2 + 2 \cdot A \cdot \sum_{j=1}^m T_{ij} \cdot x_{ij} - A \cdot \sum_{j=1}^m T_{ij}^2 \quad \text{Equation II-3}$$

where

- S_i is the qualifier's output (score)
- x_{ij} is the input at the i^{th} row and the j^{th} column
- T_{ij} is the pattern stored at the i^{th} row and the j^{th} column
- A' is an arbitrary constant
- A is an arbitrary constant

The output of each cell (S_{i-1} , S_i , S_{i+1}) is then compared to one other via a discriminator (Winner-Take-All circuitry) to determine the winner [figure II-1]. Each score (S_i) is fed in to a follower configuration, securing a proportional current from the total pull down current. This total current is provided by equally divided transistors (M_{5i-1} , M_{5i} , M_{5i+1}) whose drains are connected to a common node V_{cm} . The cell with the largest score (S_i) will conduct the most current, since it forces the largest gate to source potential voltage (V_{gs}) across the corresponding follower; therefore, it takes up the majority portion of the total bias current, forcing less current through the competing cells. Each current is then mirrored to the next stage and gets converted to a proportional voltage. To quantify this voltage a relationship between each inputs and output must be found.

Total current (the sink current) through V_{cm} is

$$I_{tot} = N \cdot I_{bias} \quad \text{Equation II-3}$$

$$I_{bias} = \frac{\beta_n}{2} \cdot (V_{BB2} - V_{SS} - V_T)^2 \quad \text{Equation II-4}$$

where

I_{bias} is the tail bias current of each cell

N is the number of cells in the architecture

V_{BB1} is a DC voltage

V_{BB2} is a DC voltage

β_n is a constant proportional to the size and process variables

The current through each cell (I_i) can be determined by its corresponding score

$$I_i = \frac{\beta_n}{2} \cdot (S_i - V_{cm} - V_T)^2 \quad \text{Equation II-5}$$

This current is converted to a voltage as follow

$$V_{out(j)} = \frac{1}{\lambda_n} \cdot \left(2 \cdot m \cdot I_i \cdot \frac{1}{\beta_4} \cdot (V_{BB2} - V_{SS} - V_T)^2 - 1 \right) + V_{SS}$$

Equation II-6

where

m is the gain between M_2 and M_3

Equation II-5 and II-6 indicate how the largest score secures the most current and how it is converted to a proportional output voltage.

This architecture occupies a relatively large amount of area due to the number of interconnections between each cell. It is also designed such that more gain can be used to further separate the winner from the losers; however, this increases both area and power consumption. Therefore, one has to trade between speed, power consumption and area, a classic bottle neck for Neural Networks.

A more efficient design has been implemented using purely capacitive Hamming Classifier shown in figure II-2 [Ugur, Jan., 1993]. This architecture also consists of two major layers, the qualifier and the discriminator. The qualifier is sub-divided into two layers, synaptic matrix and normalization matrix. The discriminator is an inhibiting circuit whose outputs are directly fed in to the inputs of tri-state buffers.

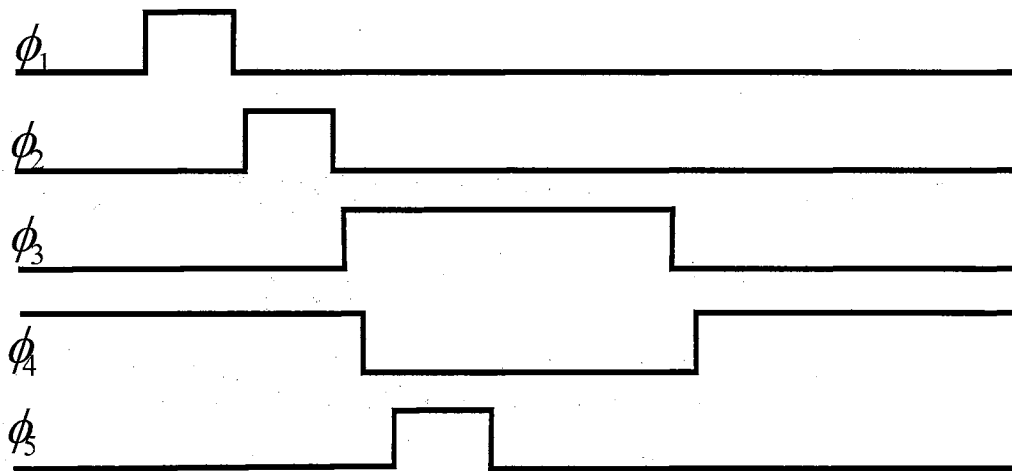
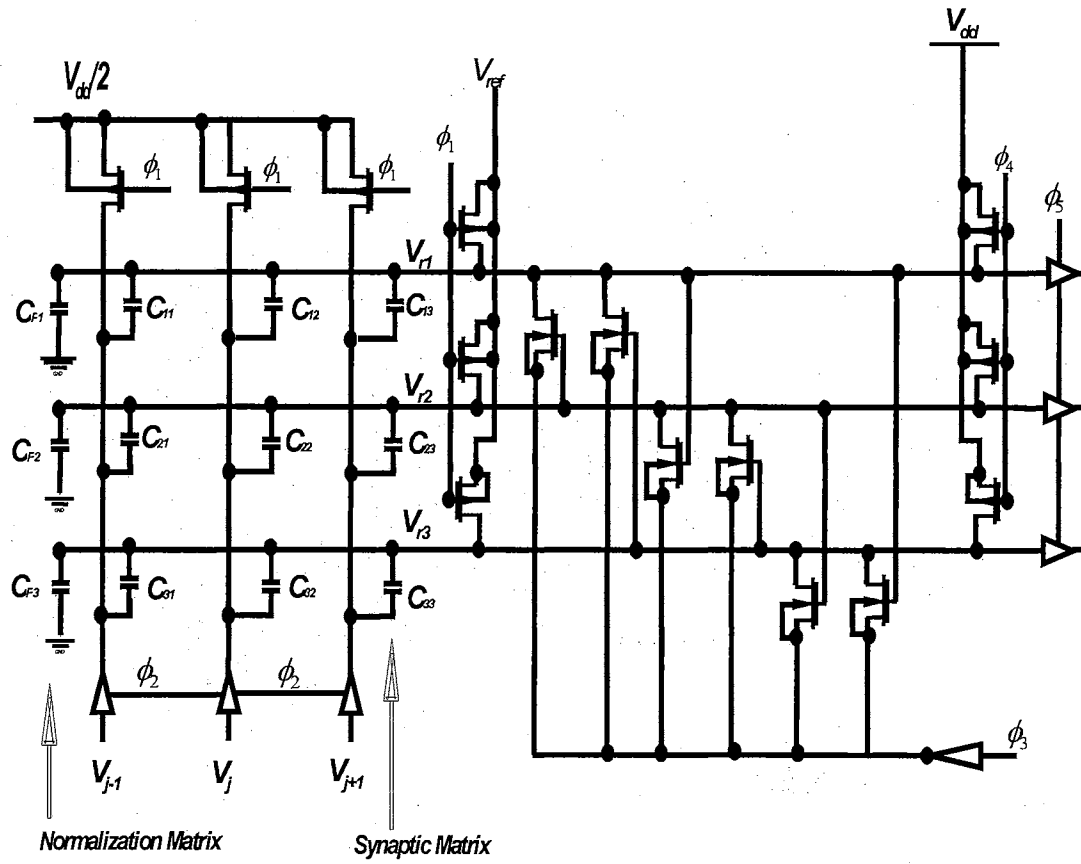


Figure II-2 . A Charge Based Hamming Classifier

As indicated in figure II-2, this architecture requires four non overlapping clocks.

At phase one (ϕ_1), all of the rows are set to a reference voltage V_{ref} , while the columns are

set to $\frac{V_{dd}}{2}$. At the end of this phase the total stored charge on each row is

$$Q_{i|\phi_1} = \left(V_{ref} - \frac{V_{dd}}{2} \right) \cdot \sum_{j=1}^m C_{ij} + V_{ref} \cdot C_{pi} \quad \text{Equation II-7}$$

where

$$C_{tot} = \sum_{j=1}^m C_{ij} + C_{pi}$$

m is the number of capacitors on each row

C_{pi} is normalizing capacitor on the i^{th} row

C_{min} is the smallest capacitance achievable on chip

C_{ij} is a capacitance between i^{th} row and j^{th} column (synaptic capacitance)

At phase 2 the input voltages (input pattern) are applied to the qualifier's columns. At this point, the final charge on each row becomes

$$Q_{i|\phi_2} = \sum_{j=1}^m (V_{ri} - V_j) \cdot C_{ij} + V_{ri} \cdot C_{pi} \quad \text{Equation II-8}$$

where V_{ri} is the line voltage. Since the charge must be conserved, equation II-7 and II-8 must be equal; therefore, at the end of phase 2 the line voltage V_{ri} becomes

$$V_{ri} = V_{ref} + \frac{1}{C_{tot}} \cdot \sum_{j=1}^m C_{ij} \cdot \left(V_j - \frac{V_{dd}}{2} \right) \quad \text{Equation II-9}$$

where

V_j is a normalized input at the j^{th} column

To relate equation II-9 to equations II-1 and II-2, inputs and the weights are defined as follow.

$$x_j = \frac{V_j}{V_{dd}} \quad \text{Equation II-10}$$

$$T_{ij} = \left(\frac{V_{dd}}{2} \cdot A \right) \cdot \frac{1}{C_{tot}} (C_{ij} - C_{min}) \quad \text{Equation II-11}$$

where

A is an arbitrary positive voltage

T_{ij} is the stored pattern at the i^{th} row and the j^{th} column

rewriting V_{ri} in equation II-9

$$V_{ri} = V_{ref} + \left(\frac{C_{min} \cdot V_{dd}}{2} \right) \cdot C_{tot} \cdot \sum_{j=1}^m (2 \cdot x_j - 1) + A \cdot \sum_{j=1}^m T_{ij} \cdot (2 \cdot x_j - 1)$$

Equation II-12

In equation II-12 input vectors are normalized while stored pattern (T_{ij}) is programmed in terms of synaptic capacitance value. Equation II-11 suggests T_{ij} is zero once $C_{ij} = C_{min}$, and it is one if $C_{ij} = \frac{2 \cdot A}{V_{dd}} \cdot C_{tot} + C_{min}$. Therefore, a binary one is presented by any capacitance larger than C_{min} , while a binary zero is realized by a minimum capacitance.

The inhibiting circuit is enabled when phase 3 (ϕ_3) clock turns on. In this circuit, all of the match lines are interconnected to one other in a fashion such that each match line controls the gate of a single pull down transistor connected between each of the competing match lines and ground. As a result, each row is connected to a series of pull down transistors whose gates are connected to each of the competing match lines. The line with the highest correlation forces the highest V_{gs} on the rest of the pull downs, causing them to decay at a higher rate while it decays at a normal rate. Eventually, this row will force all of the competing voltages to zero.

This architecture has gone to a great extent to reduce power consumption; however, the main problem still exists with respect to the connectivity and area. As the number of match lines increases, the complexity of the inhibiting circuit increases in a more rapid manner. Therefore, to accommodate N match lines NxN transistors are required; hence, the size and inter-connectivity becomes another bottle neck.

[Ugur, Feb., 1991] proposed a similar capacitive Hamming architecture (figure II-3). This architecture consists of coupled linear capacitors, an inverter per row, series of switches, two reference voltages, and a tri-state buffer.

The recognition process is divided into three different phases, storage, evaluation, and discrimination; therefore, three non overlapping clocks are required. The process starts by first charging each row voltage to the threshold of its inverter while applying reference V_{RI} and input voltage to C_{ij} and $\overline{C_{ij}}$ respectively. At this point the charge proportional to the input and the reference voltage is

$$Q_{i\phi_1} = V_T \cdot C_p + \sum_{j=1}^m (V_T - V_j) \cdot \bar{C}_{ij} + \sum_{j=1}^m C_{ij} \cdot (V_T - V_{R1}) \quad \text{Equation II-13}$$

where

- C_p is the total non synaptic parasitic associated to each row
- C_{ij} and \bar{C}_{ij} are synaptic capacitors between the i^{th} row and j^{th} colon
- V_{R1}, V_{R2} are the reference voltages.

During the next phase the input voltage is transferred to C_{ij} and the reference voltage V_{R2} is imposed on \bar{C}_{ij} . At this point of time the total charge is changed to

$$Q_{i\phi_2} = (V_T + V_{ri}) \cdot C_p + \sum_{j=1}^m (V_T + V_{ri} - V_{R2}) \cdot \bar{C}_{ij} + \sum_{j=1}^m C_{ij} \cdot (V_T + V_{ri} - V_j)$$

Equation II-14

At the end of the second phase, according to the conservation of energy, equations II-13 and II-14 must remain the same. This results in a total deviation (V_{ri}) from V_T . Therefore, at the end of the second phase the output voltage of each inverter (V_{oi}) goes high or low depending on the polarity of V_{ri} , where

$$V_{ri} = \left(\frac{\sum_{j=1}^m (C_{ij} - \bar{C}_{ij}) \cdot V_j - \left[V_{R1} \cdot \sum_{j=1}^m C_{ij} - V_{R2} \cdot \sum_{j=1}^m \bar{C}_{ij} \right]}{C_p + \sum_{j=1}^m (C_{ij} + \bar{C}_{ij})} \right) \quad \text{Equation II-15}$$

Finally, during ϕ_3 the final output U_{oi} can be written as

$$U_{oi} = V_{dd} \cdot H \left(\sum_{j=1}^m (C_{ij} - \overline{C_{ij}}) \cdot V_j - \left[V_{R1} \cdot \sum_{j=1}^m C_{ij} - V_{R2} \cdot \sum_{j=1}^m \overline{C_{ij}} \right] \right)$$

Equation II-16

Where H is approximating the transfer characteristics of each row's comparator. Equation II-16 has the same form as the generic neural function

$$Y_i = H \cdot \left(\sum_{j=1}^m T_{ij} \cdot X_j - \psi_i \right)$$

Equation II-17

Where

- T_{ij} is the bipolar connection weights
- X_j is the unipolar input ($0 \leq X_j \leq 1$)
- ψ_i is a neural threshold

Therefore, the capacitive network can operate as a neural network (equation II-17) under the following transformation.

$$X_j = \frac{V_j}{V_{dd}}$$

$$T_{ij} = \frac{C_{ij} - \overline{C_{ij}}}{K}$$

$$\Psi = \frac{V_{R1}}{V_{dd}} \sum_{j=1}^m \frac{C_{ij}}{K} - \frac{V_{R2}}{V_{dd}} \sum_{j=1}^m \frac{\overline{C_{ij}}}{K}$$

Where

K is a positive scale factor

$$\overline{C_{ij}} = C_{\min} \text{ and } C_{ij} = K \cdot T_{ij} + C_{\min}$$

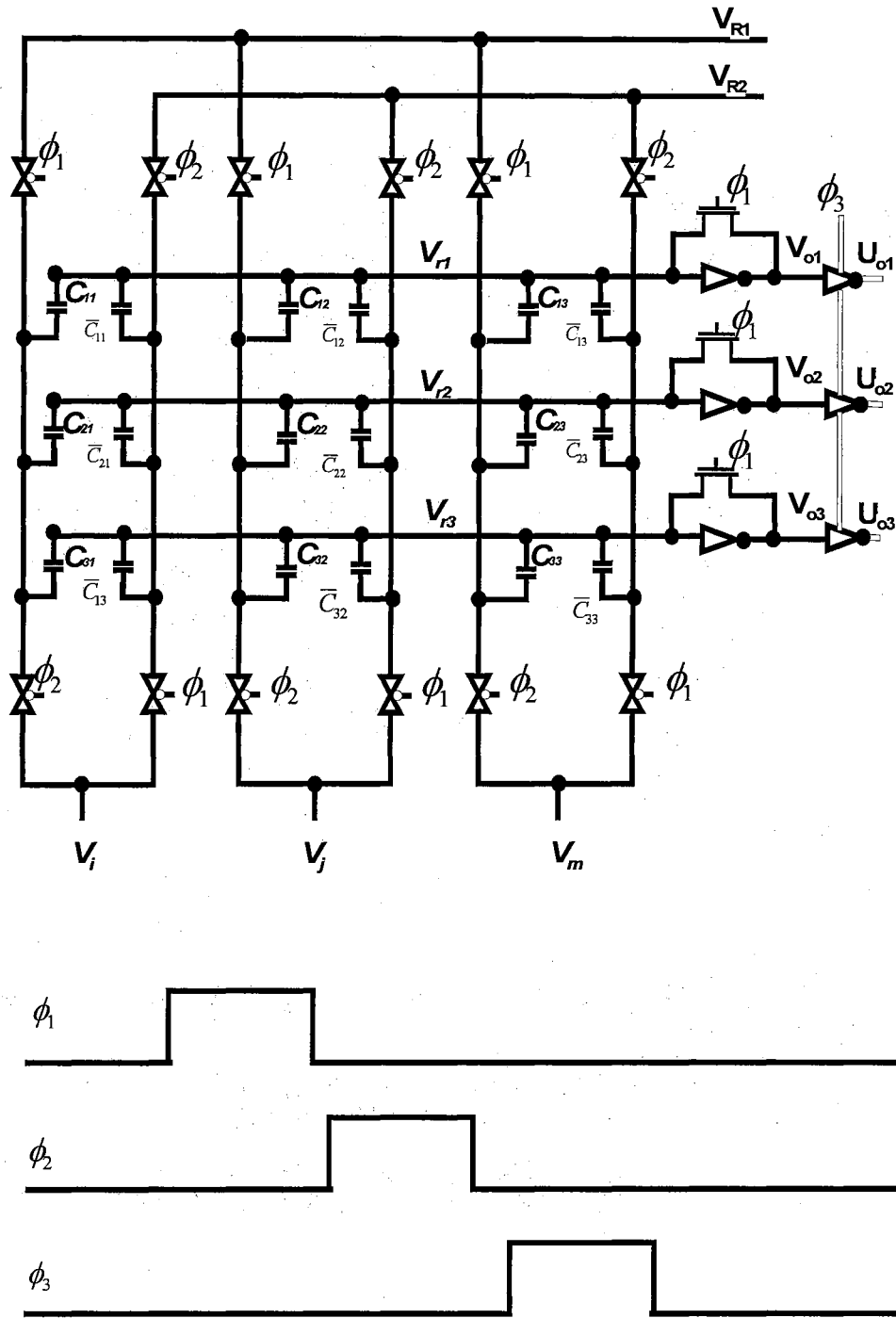


Figure II-3. A Capacitive Neural Network

[Yuping, May, 1993] proposed yet another low power and high speed solution. This technique uses a charge base comparator as the main qualifier. This comparator determines the score (S_i) based on the Euclidean distance defined by

$$S_i = C \cdot \sum_{j=1}^m [1 - |w_{ik} - x_k|] \quad \text{Equation II-18}$$

where

C is a constant

w_{ik} is the stored pattern at the i^{th} row and the k^{th} column

x_k is the input voltage at the k^{th} column

j is the number of columns

According to the above scoring mechanism, if there is a match between the stored and the input pattern, the score becomes $S_i = C \cdot \sum_{j=1}^m [1]$; otherwise, $S_i = 0$.

The hardware implementation of this scheme is shown in figure II-4. This cell is consisted of four transistors, M_1 , M_2 , M_3 , and M_4 whose parasitics at each node are lumped in to a single capacitor as below

$$C_d = C_{db1} + C_{db2} + C_{gd1} + C_{gd2} \quad \text{Equation II-19}$$

$$C_p = C_{db3} + C_{gs1} + C_{gd3} + C_{sb1} \quad \text{Equation II-20}$$

$$C_s = C_{gs3} + C_{sb3} \quad \text{Equation II-21}$$

Where C_s is controlled via the gate voltage and must be designed to be the dominant capacitor. Once the input voltage (x) is high and the weight (w) is zero, M_1 and M_4 are on

while M_2 and M_3 are off; therefore, the total capacitance contributed by the cell is

$$C_{toti} = C_d + C_p \quad \text{Equation II-22}$$

However, if the input and the weight are both high, M_1 and M_3 are on while M_2 and M_3 are off, the total capacitance contributed by this cell is

$$C_{toti} = C_d + C_p + C_s \quad \text{Equation II-23}$$

In general the two above cases can be summarized as

$$C_{toti} = (C_d + C_p) + C_s \cdot [1 - |w - x|] \quad \text{Equation II-24}$$

which is indeed the Euclidean distance. For n cells in a row, the total capacitance becomes

$$C_{toti} = \left((C_d + C_p) \cdot n + C_s \cdot \sum_{j=1}^n [1 - |w_{ik} - x_k|] \right) \quad \text{Equation II-25}$$

It is easy to see, the more exact match there is, the more capacitance (C_s) is added to the row, storing more charge; therefore, longer decay time.

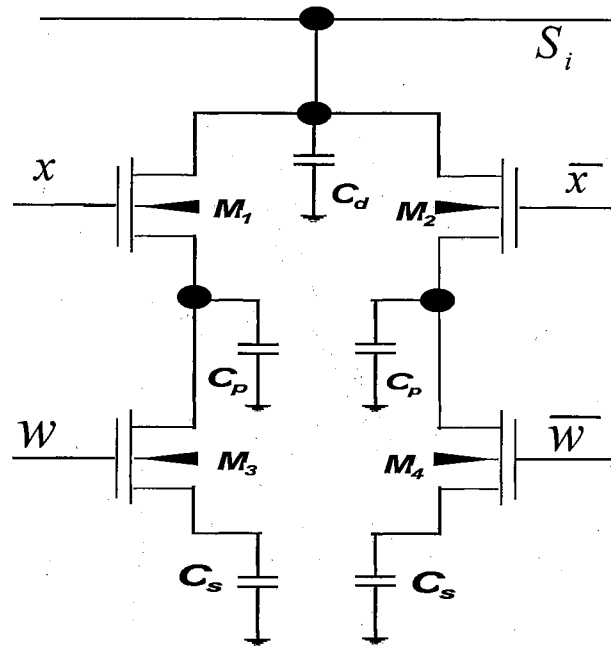


Figure II-4. A Capacitive Comparator circuit

To determine the winner a discriminator is used. This circuit is an inhibiting WTA circuit shown in figure II-5.

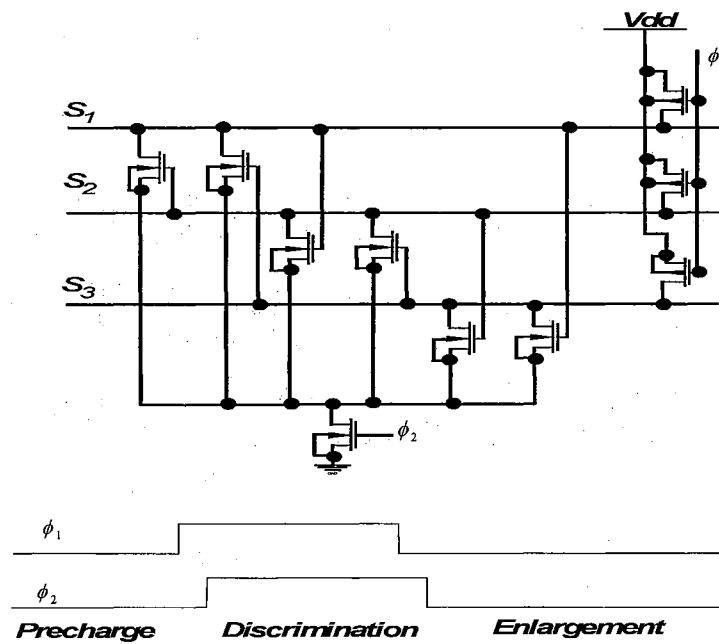


Figure II-5. The WTA circuit

The system works as follows: first precharge clock is asserted (ϕ_1 is low), next comes the discrimination or competition phase when both ϕ_1 and ϕ_2 are high. At this point the winner has a voltage higher than the threshold voltage of the inhibiting transistors. [Yuping, May, 1993] shows that the time it takes for the winner to force the loser rows to a lower voltage is proportional to $n \cdot L \cdot C_s / g_{ds}$ where n is the number of comparators in a row, L is the number of equal bits and g_{ds} is output conductance of an NMOS transistor.

[Johnson, Sep., 1991] proposed a digital circuitry which is based on the ability to recall by association using Content Addressable Memory (CAM) as neurons. CAM is a memory in which data is acquired on the basis of content rather than address; therefore, information is retrieved from a memory location with content matching the input. Unlike other CAM implementations used in pattern recognition, this architecture is not searching for an exact match; instead, the closest match is acquired; hence, the name Relaxative Content Addressable Memory (RCAM) [Jalaliddine, Jun., 1992]. RCAM finds a correlation between input and the pattern (qualifier) and lets the second stage determine the winner (discriminator). Figure II-6 depicts such architecture where the neuron is basically a CAM cell composed of two layers, RAM and an XOR circuit. The output of the CAM controls the gate voltage of pull down transistors M_{wij} called the weight transistors. These transistors operate in a digital fashion since their gates are either high or low. The inhibition strength that each word receives is proportional to the total current that sinks through all of the weight transistors ($I_{M_{wij}}$). This current can be expressed as

$$I(\text{inhibit})_i = \sum_{j=1}^m I_{M_{wij}}$$

Equation II-26

The process starts by precharging all of the lines to V_{dd} . Then the input path to each cell is enabled, causing the weight transistors to turn on or stay off depending on whether or not there is a match between the input vector and the stored pattern. The more mismatch, the more weight transistors will turn on; therefore, stronger inhibition current will flow through the corresponding row. Naturally, the word with the least inhibition current will have the slowest discharge rate as opposed to the rest. As a result, the inhibiting transistors (M_{cij}) in the winning word will be turned on weaker than the rest of the M_c transistors. As a result, all of the competing words will eventually decay to zero while the winner remains at a higher voltage (at least a threshold above ground).

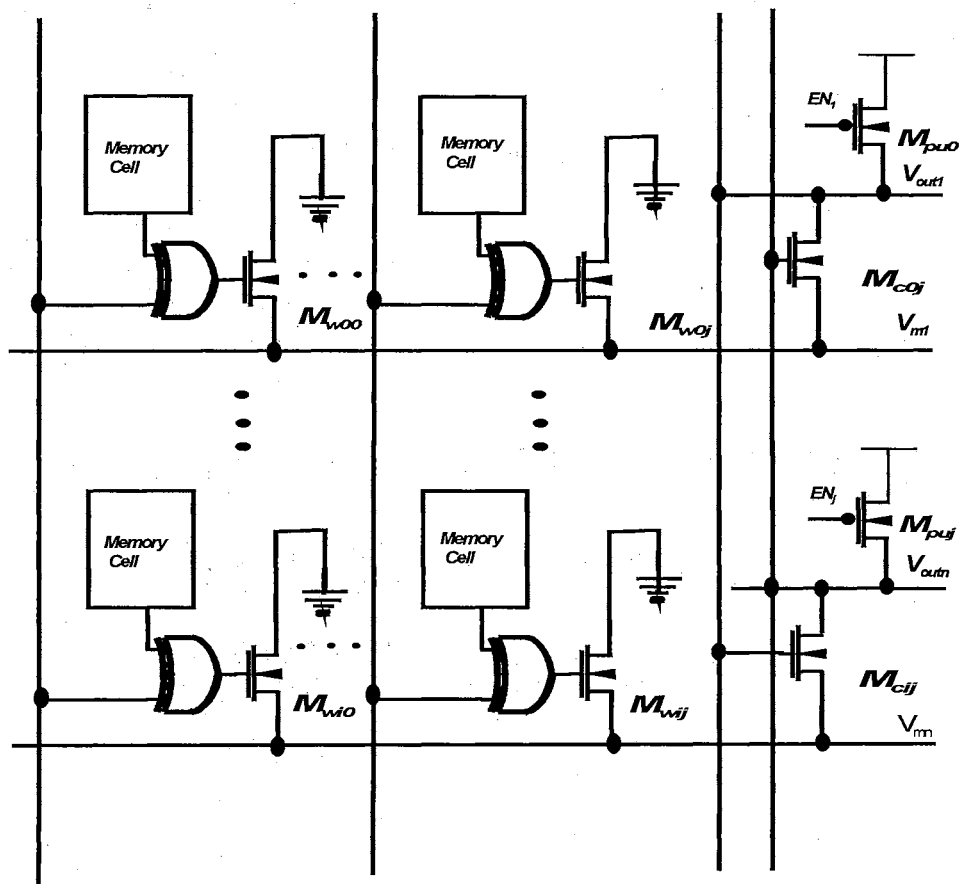


Figure II-6. The Relaxative Content Addressable Memory (RCAM)

All of the discriminators discussed above use some kind of inhibiting transistor whose gates are controlled via the competing lines. As a result the number of inhibiting transistors is proportional to the square of the number of the match lines. Due to the importance of the area, a new WTA circuit was proposed [Lazzaro, 1989]. This architecture has a complexity of $O(N)$, as opposed to the $O(N^2)$. However, the ability of this circuit is limited in finding the best match even with perfect devices. [Jalaleddine and Johnson, Jun., 1992] introduced an improved RCAM architecture based on the same WTA concept. The basic architecture is shown in figure II-7.

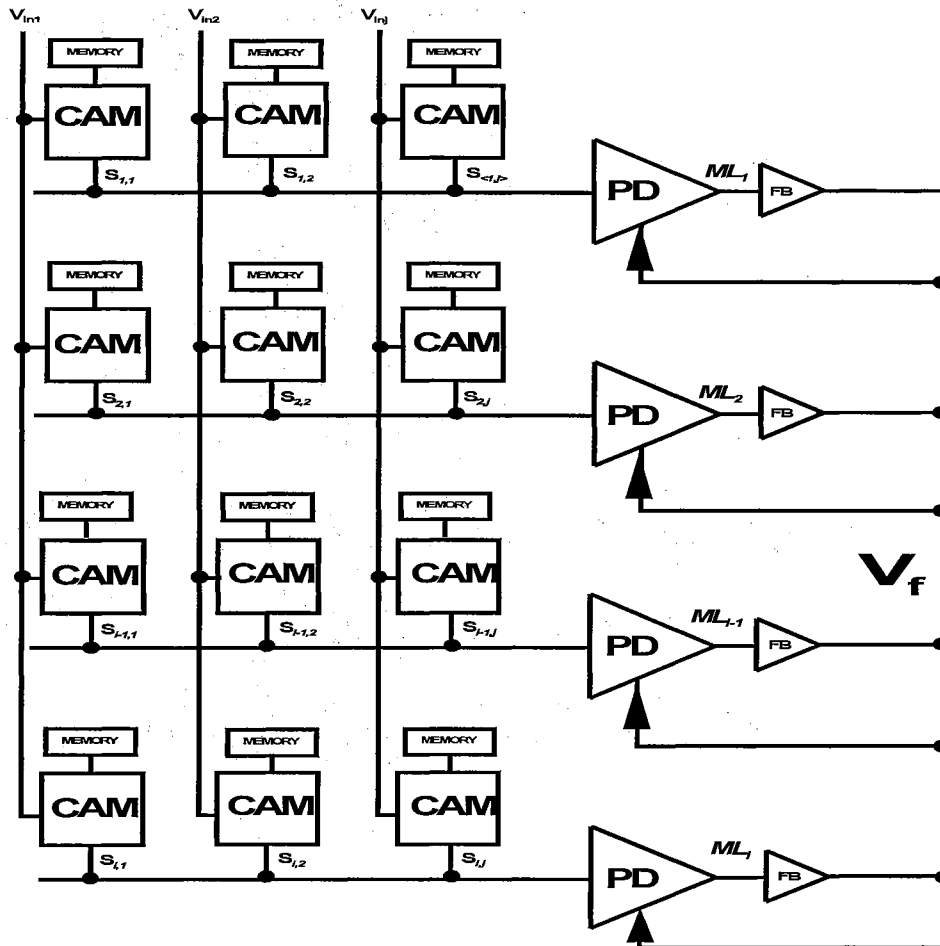


Figure II-7. The RCAM and the WTA network model

In this architecture, each word line is connected to the gate of a follower whose source is connected to a common node which is connected back to the gate of the pull down transistors. Once the evaluation starts and the RCAM evaluated the input versus the stored pattern, the match line with the smallest mismatch will have the fewest pull down transistors on; therefore, a higher voltage is established on this match line (the winner match line). Based on this voltage, the feed back voltage regulates the pull down currents throughout the whole network. Since, this feed back voltage is higher than the required regulating voltage for the rest of the competing match lines; hence, their corresponding pull down transistors tend to sink more current than normal. Eventually, these lines decay to a small voltage while the line with the highest voltage (the line with the fewest mismatch) stays at a higher level.

All Analog Approach

The multiplication operation is a function used in the vast majority of neural networks algorithms to determine the relationship between the input and the stored pattern (qualifier). The most suitable analog cell implementing this operation is the Gilbert multiplier. The multiplication is accomplished by multiplying two known voltages in order to produce a proportional output current.

A basic differential multiplier circuit is shown in Figure II-8. Using a square law model and assuming all of the transistors are in saturation, the output current can be expressed as

$$I_{out} = I_1 - I_2 \quad \text{Equation II-27}$$

$$I_{out} = k \cdot (V_{in}) \sqrt{\left(\frac{2 \cdot I_d}{k}\right) - V_{in}^2} \quad \text{Equation II-28}$$

Where

V_{in} is the differential input voltage

I_d is the tail current

k is a process dependent constant

For a small input voltage, the term V_{in}^2 can be neglected

$$I_{out} = k \cdot (V_{in}) \sqrt{\left(\frac{2 \cdot I_d}{k}\right)} \quad \text{Equation II-29}$$

Substituting for I_d , where $I_d = k \cdot (V_{gs} - |V_T|)^2$ yields

$$I_{out} = \sqrt{(2 \cdot k^2)} \cdot (V_{in}) \cdot (V_x - |V_T|) \quad \text{Equation II-30}$$

Equation II-27 shows how the output current is related to a proportion of the input voltage (V_{in}) multiplied by the tail bias voltage (V_x).

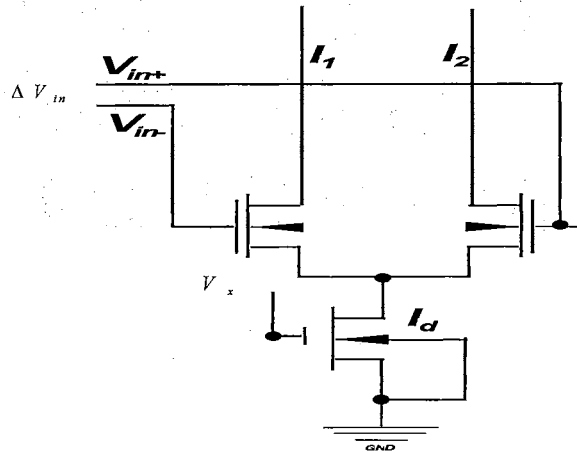


Figure II-8. The Two-Quadrant Multiplier

This cell can be used to implement a simple yet efficient single synapse cell as shown in figure II-9 [Francis, Feb., 1990].

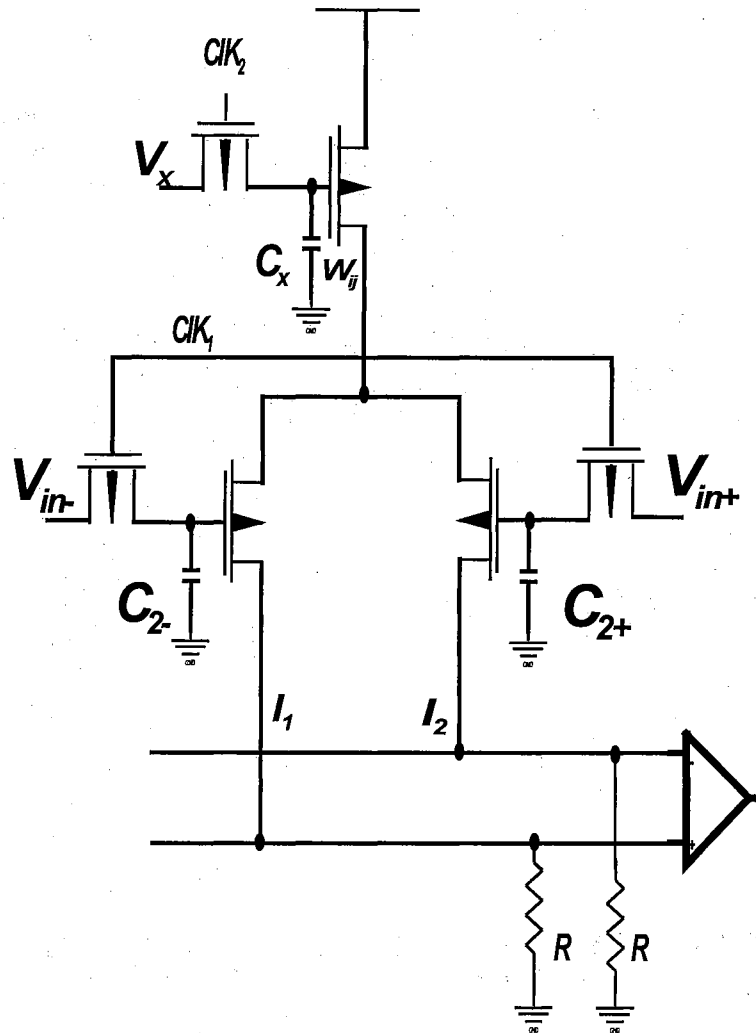


Figure II-9. The Differential Pair Two-Quadrant programmable analog Multiplier circuit

This circuit stores the input and weight vectors on capacitance, C_{2-} , C_{2+} and C_x . Moreover, since the output is current, the result of multiple stages can be summed simply by connecting all of these synaptic cells' drains to a common node

$$I_{tot} \propto \sum_{j=1}^n W_{ij} \cdot V_{x_{ij}}$$

Equation II-31

Where

W_{ij} is the stored weight at the i^{th} and the j^{th} column

$V_{x_{ij}}$ is the differential input voltage on the cell at the i^{th} and the j^{th} column

This simple architecture suffers from two major flaws, limited input range and finite charge retention. The range can be improved by design techniques; however, it increase both area and power consumption. On the other hand, there is a finite time that charges can be held on each node; therefore, refreshing is required and must be done frequently. This requires constant clocking; hence, a noisy environment which is undesirable for any analog system. To avoid this problem analog memory, also known as floating gate device, introduces an easy solution since it increases charge retention to an equivalent of 100 years.

The following section introduces the concept of analog memory. Such device has found its way in to hardware implementation of neural network systems due to its ability to store high resolution data. This property can result in an area and power efficient system. The following chapters, will show how an analog device is employed as a memory cell in the classifier proposed in this thesis; but first, a detail functionality of this analog device is necessary.

Floating Gate Structure

The first floating gate device was proposed by Khang and Sze [Khang, 1967]. In this structure, the charge is transported from the silicon substrate across the oxide to a floating metal electrode. This injection is accomplished by the electrons with the excess energy acquired from a high source to drain channel electric field. Injection of electrons in this manner is known as a hot electron injection. The hot electrons with sufficient energy will conduct across the oxide barrier and charge the gate. The drawback of this device is that removing electrons from the floating gate is not controllable. Therefore, to erase these electrons, the floating gate must be exposed to UV light.

[Johnson, 1980] introduced the first electrically erasable programmable read only memory (EEPROM). This topology gave a better control over charge transfer in and out of the floating gate. Eventually, this architecture led to the floating gates with tunneling injector [Yong-Yoong, Dec., 1994].

Figure II-10 depicts a simple model of such EEPROM device where C_{PP} is the capacitance formed by floating Poly and a second Poly, C_{ox} is the capacitance between the floating gate and substrate, C_{inj} is the second double-poly capacitance and C_g is the capacitance formed by the floating gate and the channel of the measuring transistor whose purpose will be explained in the later chapters.

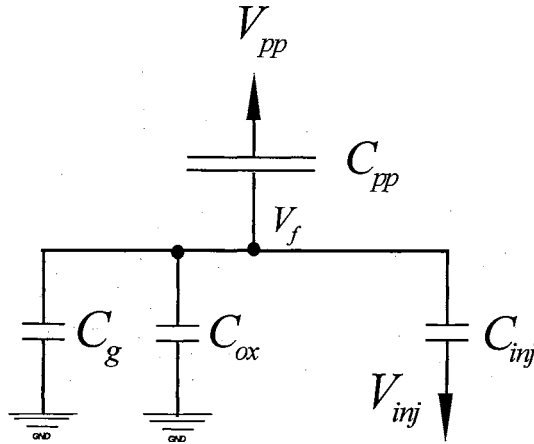


Figure II-10. The simplified EEPROM Model

The charge flow and storage in the EEPROM with the tunneling injection structure shown in figure II-10 can be explained by Fowler - Nordheim tunneling phenomena. Without getting into too much physics, Fowler - Nordheim is explained as follows; there exists an energy barrier of approximately 3.2V between poly silicon and silicon dioxide which prevents electrons flow between the two layers. At room temperature; however, electrons have enough kinetic energy to tunnel approximately 5nm in to the SiO_2 [Kolodny, Jun. 1986]. At this point, if the potential within this distance (5nm) is below 3.2V, these electrons will return; hence, no net current flow has been established. However, if there exists a strong electric field in SiO_2 ($> 3.2\text{V}/5\text{nm}$), some of these electrons will be carried away by the electric field, causing a net current to flow. Therefore, more electric field results in more current flow. The current density defining this phenomena is

$$I_{tun} = \alpha \cdot E^2 \cdot e^{-\left(\frac{\beta}{E}\right)}$$

Equation II-32

$$E = \frac{V_{tun}}{t_{ox}}$$

Equation II-33

where

- E is electric field in the silicon dioxide
- V_{tun} is the voltage across the silicon dioxide
- t_{ox} is the thickness of silicon dioxide
- α and β are parameters determined experimentally

Before getting into the actual charge and voltage relationships it is important to understand different ways of programming an EEPROM. There are two ways to alter the amount of charge on an EEPROM, a write and an erase operation. A write operation is done when more negative charge is transferred into the floating gate via Fowler-Nordheim tunneling. To perform a write operation a high voltage is applied to the V_{pp} terminal (figure II-10) while grounding V_{inj} . Doing so causes a strong electric field with a direction from the floating gate to the injector causing a negative charge flow in the opposite direction of the electric field. If this electric field is strong enough negative charge penetrates into the floating gate across the injector oxide.

An erase operation is accomplished by grounding the control gate (V_{pp}) and applying a high voltage to the V_{inj} terminal. As a result, a strong electric field is imposed across the tunneling terminal (from V_{inj} to the floating gate) causing a negative charge movement from the floating gate to the V_{inj} terminal; hence, removing electrons from the floating gate.

Now that the basic functionality of an EEPROM has been explained, a more in

depth characterization is in order.

As shown in figure II-10, EEPROM is consisted of a multi-conducting bodies in an isolated system. It should be obvious that presence of charge on one of the conductors will affect the potential of the others; therefore, the total charge at the floating gate can be written as

$$Q_{float} = (V_f - V_{pp}) \cdot C_{pp} + V_f \cdot C_g + V_f \cdot C_{ox} + (V_f - V_{inj}) \cdot C_{inj} \quad \text{Equation II-34}$$

rearranging above and solving for the total voltage at the floating gate terminal

$$V_f = \frac{Q_{float}}{C_{tot}} + \frac{V_{pp} \cdot C_{pp}}{C_{tot}} + \frac{V_{inj} \cdot C_{inj}}{C_{tot}} \quad \text{Equation II-35}$$

where

$$C_{tot} = C_{pp} + C_g + C_{ox} + C_{inj}$$

V_f is the voltage at the floating gate

Q_{float} is the total charge stored at the floating gate

For the write operation, $V_{inj} = 0$

$$V_{tun} = V_f - V_{inj} = V_f \quad \text{Equation II-36}$$

$$V_f = \frac{Q_{float}}{C_{tot}} + \frac{V_{pp} \cdot C_{pp}}{C_{tot}} \quad \text{Equation II-37}$$

$$\frac{dQ_{float}}{dt} = -I_{tun} = -\alpha \cdot \left(\frac{V_{tun}}{t_{inj}}\right)^2 \cdot e^{-\left(\frac{\beta}{\frac{V_{tun}}{t_{inj}}}\right)} \quad \text{Equation II-38}$$

Where

t_{inj} is the thickness of the oxide in the injector capacitor

The same procedure results in a similar expression for the erase operation.

$$V_f = \frac{V_{inj} \cdot C_{inj}}{C_{tot}} + \frac{Q_{float}}{C_{tot}} \quad \text{Equation II-39}$$

$$V_{tun} = V_{inj} - V_f \quad \text{Equation II-40}$$

$$V_{tun} = V_{inj} \cdot \left(1 - \frac{C_{inj}}{C_{tot}}\right) - \frac{Q_{float}}{C_{tot}} \quad \text{Equation II-41}$$

$$V_{tun} = \left(\frac{C_{pp} + C_{ox} + C_g}{C_{tot}}\right) \cdot V_{inj} - \frac{Q_{float}}{C_{tot}} \quad \text{Equation II-42}$$

$$\frac{dQ_{float}}{dt} = I_{tun} = \alpha \cdot \left(\frac{V_{tun}}{t_{inj}}\right)^2 \cdot e^{-\left(\frac{\beta}{V_{tun}}\right)} \quad \text{Equation II-43}$$

Comparing Equations II-39 and II-43 it is easy to see that in a given programming process, as the charge trapped in the floating gates varies, the tunneling current also varies in a nonlinear fashion due to the amount of charge already trapped on the floating gate.

Figure II-11 combines the two ideas presented in this chapter (Gilbert multiplier and EEPROM) to create a programmable neuron with a finite resolution and long term charge retention [Holler, Aug., 1990]. In this architecture the tail voltage of each current sink is programmed through a close loop programming circuit. It was shown [Holler, Aug., 1990] that the output current is

$$\Delta I_{out} = I^+ - I^- \quad \text{Equation II-44}$$

$$\Delta I_{out} = k \cdot \Delta V_{in} \cdot \Delta V_{fg}$$

Equation II-45

where

$$\Delta V_{fg} = \frac{\Delta Q_{fg}}{C_{tot}} \text{ (assuming } V_{inj} = V_{pp} = 0)$$

ΔV_{fg} is the differential weight stored in each neuron

ΔQ_{fg} is the charged written/removed to/from the floating gate

k is a positive constant

ΔV_{in} is the differential input voltage

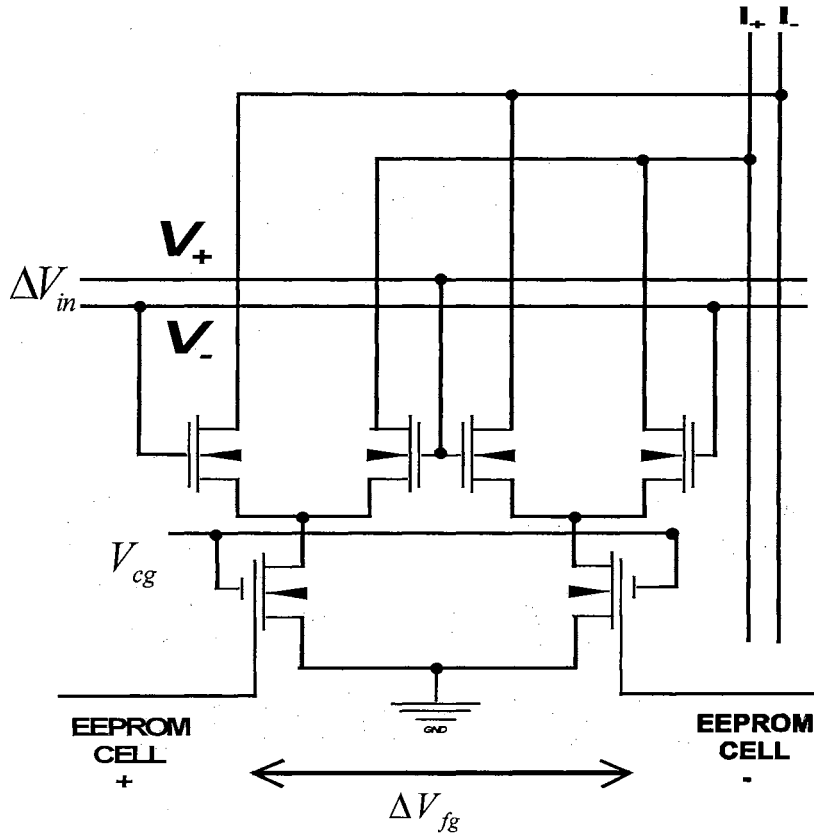


Figure II-11. The differential floating gate synapse

Although, the above technique is an effective method of implementing a programmable Neuron, it consumes power and tend to be area consuming.

CHAPTER III

ANALOG CAM

Introduction

Traditionally, recognition engines realized exact match by searching for a perfectly matched pattern [Grant, Sep., 1994, Perfetti, Oct., 1990]. Then came the neural inspired system which decided on an output based on the closest match [Johnson, Jalaleddine, Jun., 1992]. In this architecture the patterns were stored in a digital fashion; therefore, the content addressable memory (CAM) cells could only represent the patterns with a one or zero value (a binary pattern). To improve such system, this chapter proposes an ultra low power Content Addressable Memory device (CAM) which allows a greater information storage by storing the patterns in an analog fashion (analog pattern).

First the function of the CAM cell is discussed. Next, a piece-wise linear transistor model is used to derive a theoretical closed-form solution for the CAM. Furthermore, to prove the validity of the model and the closed-form solution, the theoretical results are compared with the simulation results. Finally, it will be shown how an EEPROM can be integrated into this proposed CAM architecture as a long term memory device.

A Pre-charged Based Content Addressable Memory (CAM)

Figure III-1 depicts the novel analog CAM cell. This architecture is based on two paths competing to reach a state which best represents the closeness of two input voltages (the input voltage V_{in} and the pattern voltage V_{ref}). Therefore, the main operation of this circuit is to determine how close the input voltage is to the stored pattern.

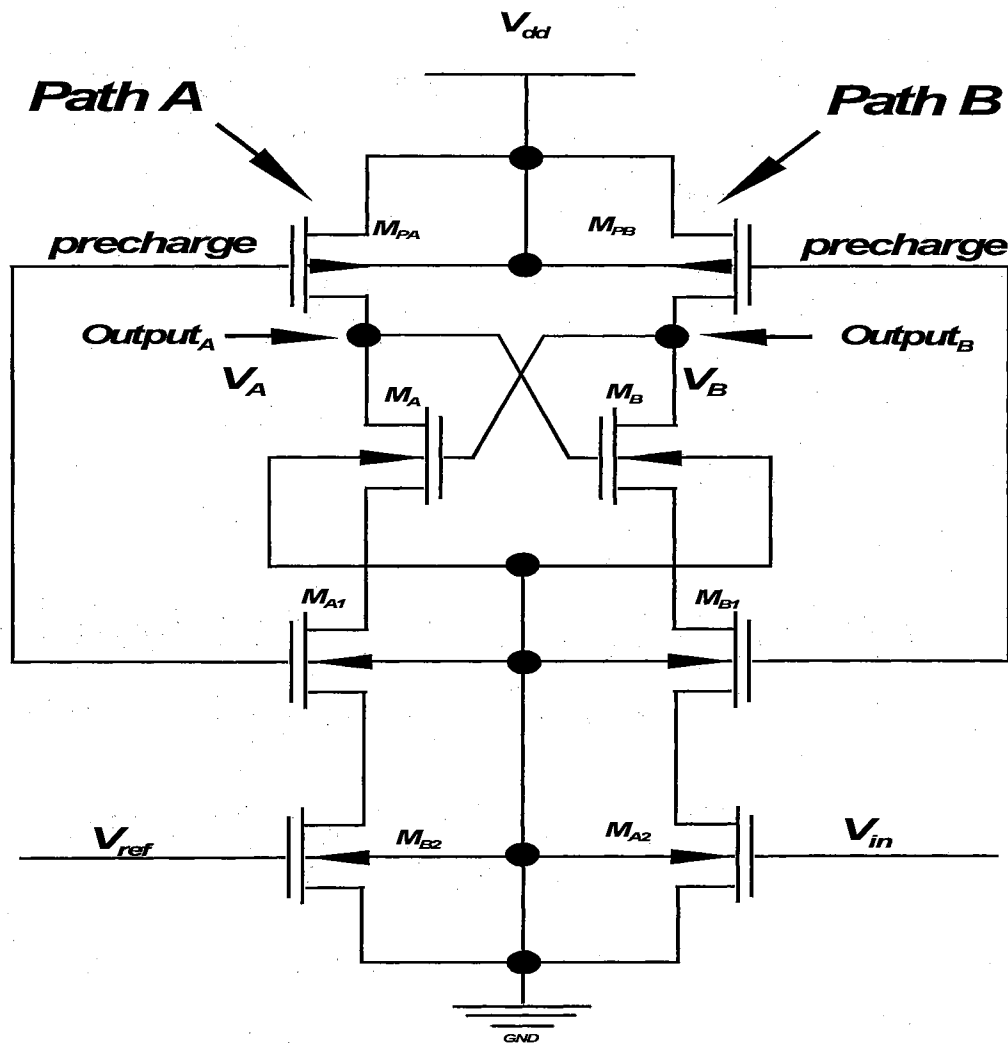


Figure III-1. The Charge Based Analog Content Addressable Memory (CAM)

There are two operating modes in this architecture, precharge and evaluation. The process starts with precharging nodes V_A and V_B of the two competing paths. Once the precharge clock is low, the transistors M_{PA} and M_{PB} (PMOS) are both conducting while transistors M_{A1} and M_{B1} are both off, preventing any DC current from the positive power supply (V_{dd}) to Ground; hence, charging nodes V_A and V_B to V_{dd} . The second phase is the evaluation phase when the precharge clock is high. At this time transistors M_{PA} and M_{PB} are both turned off while transistors M_{A1} and M_{B1} are on. At this point, a discharge path now exists from both nodes V_A and V_B to ground.

Since the current through a transistor is primarily proportional to its gate to source potential (V_{gs}), the discharge current in each path is initially set by the input voltages V_{in} and V_{ref} (figure III-1). The higher these voltages the higher the discharge current through the corresponding paths would be. To further enhance the difference between the two paths, a pair of negative feed back transistors (cross coupled or inhibiting transistors) are included in between each path (M_A and M_B). Assuming, all of the devices match perfectly, if one of the paths initially conducts more current due to a higher input voltage (for instance, V_{ref} in path A), the corresponding precharged node (V_A) tends to fall at a faster rate than its counter part (V_B in path B). As a result the fast falling node V_A asserts a lower V_{gs} on the inhibiting transistor (M_B) in the competing path (path B). Therefore, less current flows through this path (path B) than the faster falling path (path A). Therefore, node V_B falls at an even slower rate, asserting more V_{gs} on the crossed couple transistor M_A . This condition forces even more conduction in the path A while gradually turning M_B transistor in path B off. Thanks to this negative feed back mechanism, the path with higher input voltage (path A) falls at a faster rate while inhibiting further

conduction in the competing path (path B). This process feeds itself until V_A falls below the threshold of the inhibiting transistor in the path B (M_B). At this point the final steady state is reached.

Figure III-2 depicts another possible way of implementing the CAM cell. The difference between figure III-2 and figure III-1 is the location of the input transistors (M_{A2} and M_{B2}). Constructing the CAM as shown in figure III-2 introduces a body effect due to the voltage drop across the drain to source of M_{A1} and M_{B1} . This body effect is variable and changes as the discharging current changes in the corresponding path. This nonlinear effect causes the input transistors to mismatch which is not desired since the comparison between the inputs will not be fair. Constructing the CAM as shown in figure III-1 eliminates this body effect issue, making it more desirable.

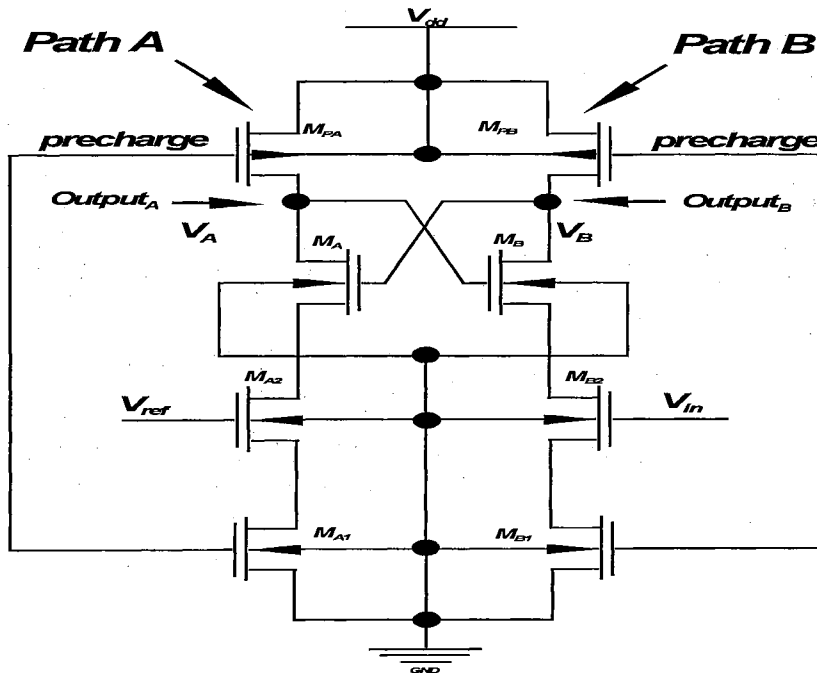


Figure III-2. An Alternative Configuration of the Analog CAM

Large Signal Analysis

Now that an overview of the CAM architecture has been described, a theoretical analysis is in order, but first a proper transistor model must be introduced. Generally, transistor models are presented in two ways, linear and nonlinear. The linear models assume a current linearly related to the drain to source voltage potential of a transistor in the triode region and a constant current in the saturation region (velocity saturation). These models tend to predict the first degree dependence of the transistor's current I_D , the drain to source potential (V_{ds}) and gate to source voltage potential (V_{gs}). Although, there is a discontinuity between these regions, this type of model works well for fast short channel digital circuit design. The second group models the non-linearities, such as the transistor behavior under electric field exerted by V_{gs} and V_{ds} . Such models are used in precision analog circuit designs where these non-linearities influence the behavior of the overall circuit. For analytical calculation, a simple linear model is sufficient, since a more accurate result can be computed via numerical manipulations. The main objective of the analytical derivation is to provide the designer an insight on the behavior of the CAM cell by expressing the final result in a simplified manner. It will be clear that although the model used is largely simplified, yet an excellent agreement with the numerical results (HSPICE) is achieved.

The model used is a piece-wise linear model introduced by [Johnson, Sep., 1991]

$$I_D = G_{sat} \cdot (V_{gs} - V_T)$$

Equation III-1

$$G_{ohm} \cdot V_{ds} > G_{sat} \cdot (V_{gs} - V_T) \text{ and } V_{gs} > V_T$$

$$I_D = G_{ohm} \cdot V_{ds} \quad \text{Equation III-2}$$

$$G_{ohm} \cdot V_{ds} < G_{sat} \cdot (V_{gs} - V_T) \text{ and } V_{gs} > V_T$$

$$I_D = 0 \quad \text{Equation III-3}$$

$$V_{gs} < V_T$$

Equations III-1 and III-2 imply no dependence of I_D to V_{ds} in the saturation region and a completely linear dependence of I_D to V_{ds} in the ohmic region.

Closed-Form Solution

It was pointed out that the CAM cell is a highly non-linear system. The incentive behind finding a closed-form is to approximate the CAM's behavior by a linear approximation method such that the system's behavior is still modeled the same, yet simplified enough to be able to predict the CAM's behavior under any parameter variations.

The key to a successful closed-form solution is to simplify the circuit as much as possible without jeopardizing the accuracy of the final result. Figure III-3 depicts the CAM cell with its parasitics. All of the parasitics to ground are lumped at the drain and source of each transistor, while the coupling capacitors are drawn as they appear in the real circuit. These parasitics are assumed to be linear and not varying with the voltage.

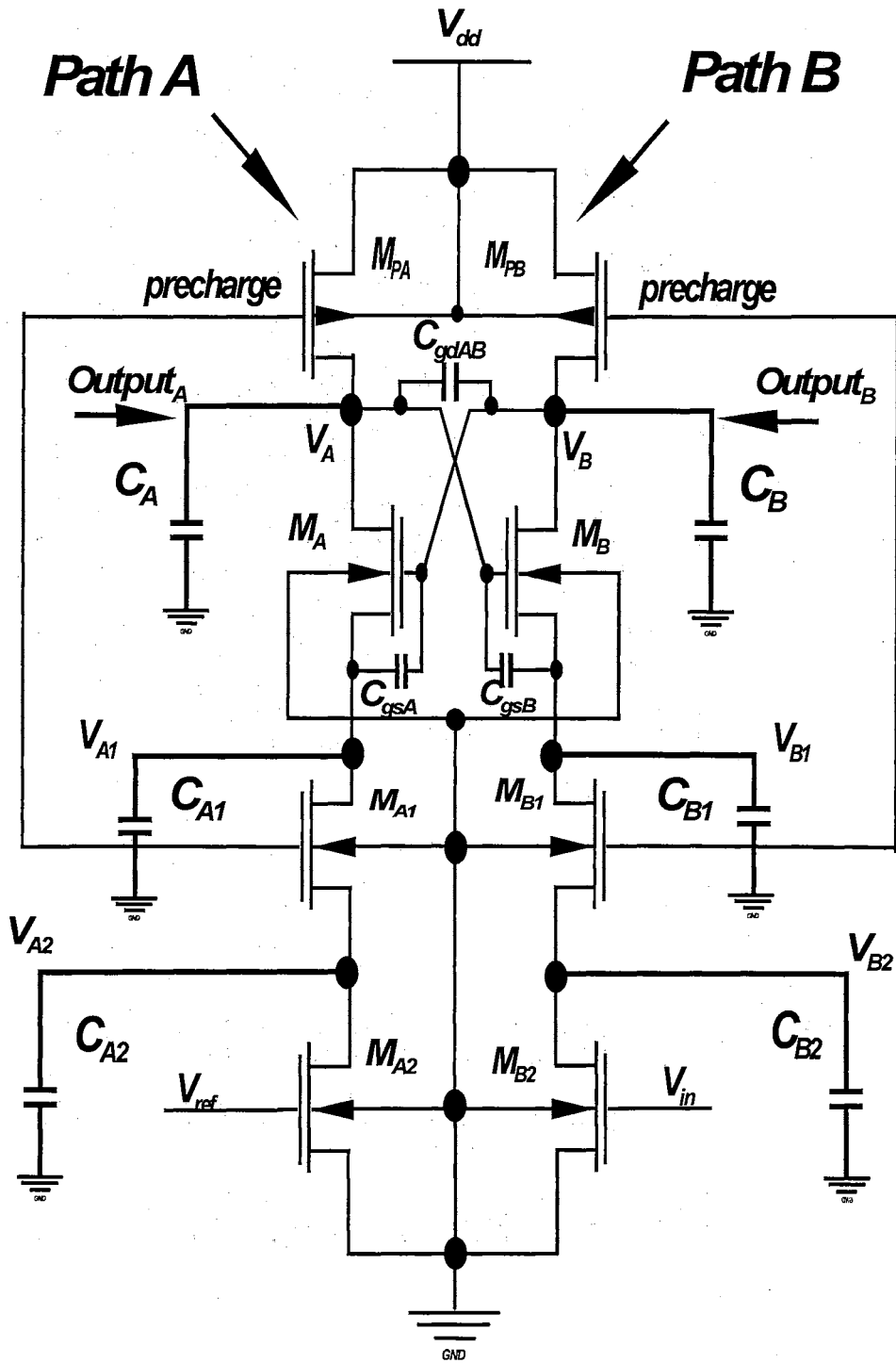


Figure III-3. The CAM with parasitics

Where

$$C_A = C_{gd_{MPA}} + C_{bd_{MPA}} + C_{db_{MA}} + C_{load} \quad \text{Equation III-4}$$

$$C_{A1} = C_{gs_{MA}} + C_{bs_{MA}} + C_{db_{MA1}} + C_{gd_{MA1}} \quad \text{Equation III-5}$$

$$C_{A2} = C_{gs_{MA1}} + C_{bs_{MA1}} + C_{db_{MA2}} + C_{gd_{MA2}} \quad \text{Equation III-6}$$

$$C_B = C_{gd_{MPB}} + C_{bd_{MPB}} + C_{db_{MB}} + C_{load} \quad \text{Equation III-7}$$

$$C_{B1} = C_{gs_{MB}} + C_{bs_{MB}} + C_{db_{MB1}} + C_{gd_{MB1}} \quad \text{Equation III-8}$$

$$C_{B2} = C_{gs_{MB1}} + C_{bs_{MB1}} + C_{db_{MB2}} + C_{gd_{MB2}} \quad \text{Equation III-9}$$

C_{load} is the total load capacitance seen at the output of the CAM cell (It will be shown in chapter IV that this load capacitance is the total gate capacitance of an n-channel device connected to V_A and V_B)

The coupling capacitor C_{gdAB} which is composed of the two drain to gate capacitors of M_A and M_B ($C_{gdAB} = C_{gdMA} + C_{gdMB}$), can be ignored since before the steady state voltage has reached, the two nodes V_A and V_B fall nearly at the same rate which implies $\frac{dV_A}{dt} \approx \frac{dV_B}{dt}$. Therefore, the total charge transfer between the two nodes is

minimal. This is written as $\Delta Q = C_{dAB} \cdot \left(\frac{dV_A}{dt} - \frac{dV_B}{dt} \right) \approx 0$. This assumption is valid since

the region during which the final voltage is set, occurs before one of the coupling transistors (M_A or M_B) turn off. During this period, V_A , V_{B1} , V_B and V_{A1} nodes fall at the same rate, causing the charge transfer between any of these two nodes negligible. Minimal charge transfer with time means minimal current through the corresponding

capacitor; therefore, any current through these paths can be ignored without effecting the accuracy of the final voltage. Similar argument can be made for the C_{gsB} and C_{gsA} coupling capacitors.

Further assumption is in regards to the regions of operation of each transistors. Using equations III-1 through III-3, it is possible to predict the exact region of operation of each transistor. It is assumed $G_{ohm}=G_{sat}$, since this model used already grossly approximates the transistor's behavior in the ohmic region (Johnson, Sep., 1991); therefore, the boundaries become

$$V_{ds} > (V_{gs} - V_T) \quad (\text{saturation case}) \quad \text{Equation III-10}$$

$$V_{ds} < (V_{gs} - V_T) \quad (\text{triode case}) \quad \text{Equation III-11}$$

In the following analysis, it is assumed that each transistor will stay entirely in the region that it spends the majority of its time during the evaluation period. For instance, initially M_{A2} and M_{B2} transistor's drain voltage are discharged to ground given the input voltages are greater than the threshold voltage of each input transistor. Once the evaluation starts and the transistors M_{A1} and M_{B1} start conducting, each drain voltage rises to some voltage prior to discharging back to ground. Depending on the input voltage, M_{A2} and M_{B2} transistor can feasibly enter the saturation voltage for a short period of time, then enter the ohmic region again. As a result, these transistors operate in the ohmic region for the majority of the operating time. Therefore, in this analysis, M_{A2} and M_{B2} are assumed to be in the ohmic region over the entire evaluation process, since they can only be in the saturation region for a small period of time.

Furthermore, C_{A1} is charged to a threshold voltage below the precharged node V_B

($V_{A1} = V_{dd} - V_T$) and the node V_{A2} is discharged to zero (assuming the input voltage is higher than the threshold voltage of the input transistor). This implies

$$V_{A1} \succ V_{pre} - V_T \quad \text{Equation III-12}$$

where

$$V_{pre} = V_{dd}$$

$$V_{A1} = V_{dd} - V_T$$

Substituting for V_{pre} and V_{A1} in equation III-12 yield $V_{dd} - V_T \succ V_{dd} - V_T$. This implies that the transistor M_{A1} is marginally in saturation even at the instant the evaluation mode commences. It is obvious that this transistor will fall out of saturation, much earlier than the rest of the transistors. Therefore, it is safe to assume this transistor is in the ohmic region for the entire evaluation period. The same argument is valid for the M_{B1} transistor.

The last set of transistors to consider are the M_A and M_B transistors. After the precharge period, both M_A and M_B are in saturation. During the evaluation process, one of the transistors starts to turn on harder than the competing transistor (assuming V_{in} and V_{ref} are different). This process continues until one of the transistors is turned on very hard while the competing cell is only marginally on. Eventually, the marginally on transistor turns off while the competing transistor stays on, until its drain decays to ground. The moment one of the transistors starts to turn off is the time when the two voltage V_A and V_B start separating and not fall at the same rate. Prior to this time, the gate and drain of these transistors fall at a rate such that the gate voltage of neither transistor exceeds its drain

voltage by more than a threshold voltage (i.e. $V_B > V_A - V_T$). This implies that both transistors stay in saturation until one of them starts to turn off. This insures that these two transistors are for the most part in the saturation region. Furthermore, since the parasitic capacitors C_{A1} , C_{B1} , C_{A2} and C_{B2} are smaller than C_A and C_B , and the falling rate of the voltage V_A and V_B is dominated by the combination of the larger capacitance C_A , C_B and the high impedance looking in to the drain of each transistor M_A and M_B , C_{A1} , C_{B1} , C_{A2} , and C_{B2} can be ignored without effecting the final result.

Figure III-4 depicts a large signal representation of the analog CAM.

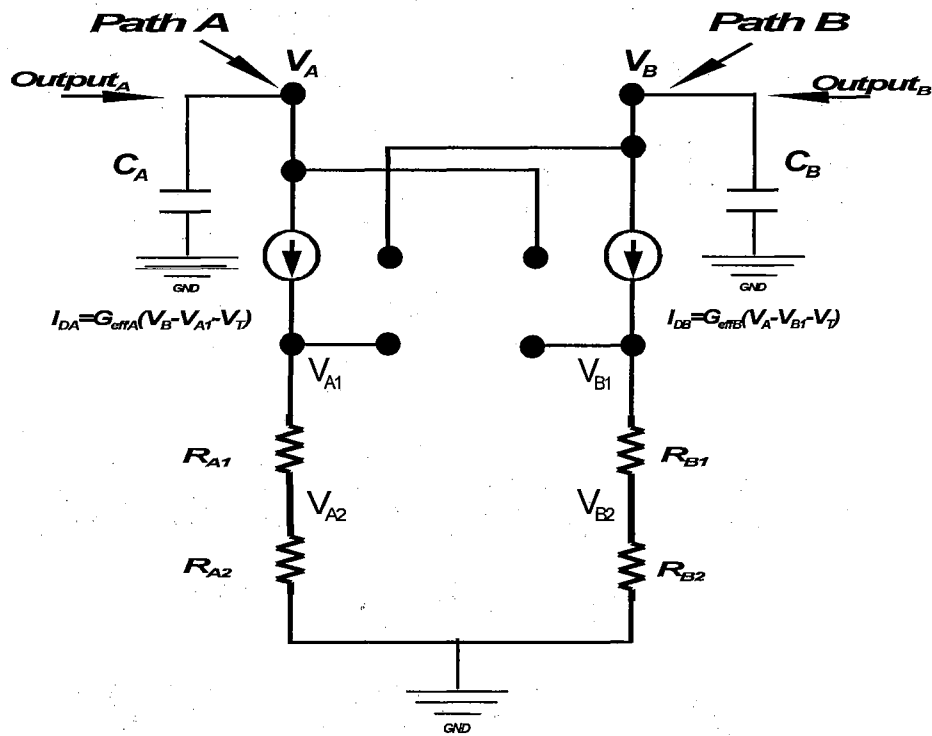


Figure III-4. Large Signal Model of the Analog CAM

In this figure, transistors M_A and M_B are presented as voltage controlled current sources (VCCS) whose currents are controlled via the drain voltage of the competing

transistor (negative feedback). Transistors M_{A1} , M_{B1} , M_{A2} and M_{B2} are modeled as resistors since these transistors spend most of their time in the ohmic region. Current through each of these devices is presented by a piece-wise linear model discussed at the beginning of this chapter (equation III-1 through equation III-3).

Once the evaluation mode begins, V_A and V_B are discharged at the rate set by the I_{DA} and I_{DB} .

Where

$$I_{DA} = C_A \cdot \frac{dV_A}{dt} \quad \text{Equation III-13}$$

$$I_{DB} = C_B \cdot \frac{dV_B}{dt} \quad \text{Equation III-14}$$

$$I_{DA} = G_{effA} \cdot (V_B - V_{A1} - V_T) \quad \text{Equation III-15}$$

$$I_{DB} = G_{effB} \cdot (V_A - V_{B1} - V_T) \quad \text{Equation III-16}$$

$$V_{A1} = I_{DA} \cdot R_A \quad \text{Equation III-17}$$

$$V_{B1} = I_{DB} \cdot R_B \quad \text{Equation III-18}$$

$$R_A = \frac{1}{G_{A1}} + \frac{1}{G_{A2}} \quad \text{Equation III-19}$$

$$R_B = \frac{1}{G_{B1}} + \frac{1}{G_{B2}} \quad \text{Equation III-20}$$

G_{effA} is the effective conductance for the M_A transistor

G_{effB} is the effective conductance for the M_B transistor

- G_{A1} is the effective conductance for the M_{A1} transistor
- G_{B1} is the effective conductance for the M_{B1} transistor
- G_{A2} is the effective conductance for the M_{A2} transistor
- G_{B2} is the effective conductance for the M_{B2} transistor

These parameters will be discussed in detail later in this chapter.

Substituting equation III-13 and equation III-14 into the equation III-15 and equation III-16 for I_{DA} and I_{DB} yields

$$C_A \cdot \frac{dV_A}{dt} = G_{effA} \cdot (V_B - V_{A1} - V_T) \quad \text{Equation III-21}$$

$$C_B \cdot \frac{dV_B}{dt} = G_{effB} \cdot (V_A - V_{B1} - V_T) \quad \text{Equation III-22}$$

Substituting for V_{A1} and V_{B1} (equations III-17 and III-18) in equations III-21 and 22 and arranging the result gives

$$\begin{bmatrix} C_L & 0 \\ 0 & C_L \end{bmatrix} \cdot \begin{bmatrix} \frac{dV_A}{dt} & 0 \\ 0 & \frac{dV_B}{dt} \end{bmatrix} + \begin{bmatrix} 0 & G_A \\ G_B & 0 \end{bmatrix} \cdot \begin{bmatrix} V_A \\ V_B \end{bmatrix} = \begin{bmatrix} G_A \cdot V_T \\ G_B \cdot V_T \end{bmatrix} \quad \text{Equation III-23}$$

Where

$$G_A = \frac{G_{effA}}{1 + G_{effA} \cdot R_A} \quad \text{Equation III-24}$$

$$G_B = \frac{G_{effB}}{1 + G_{effB} \cdot R_B} \quad \text{Equation III-25}$$

Solving the above differential equations for the time the current through the M_A transistor is equal to zero, yields the final steady state value of the V_A voltage. This voltage is expressed as follows (Appendix A)

$$V_{Af} = \sqrt{(V_{A0} - V_T)^2 - \frac{G_A}{G_B} \cdot (V_{B0} - V_T)^2} + V_T \quad \text{Equation III-26}$$

Using the same procedure, the steady state output voltage at the output VB is

$$V_{Bf} = \sqrt{(V_{B0} - V_T)^2 - \frac{G_B}{G_A} \cdot (V_{A0} - V_T)^2} + V_T \quad \text{Equation III-27}$$

Where

V_{A0} is the initial precharged voltage of the V_A node

V_{B0} is the initial precharged voltage of the V_B node

Equation III-26 and III-27 are the closed-form solutions indicating the final steady state voltage of each path. These expressions are simple yet able to predict the effect of any parameter variation on the final steady state voltage.

To see how these equations predict the final steady state voltages, consider increasing the conductance of M_{A2} . This decreases R_A (equation III-19), which results in

an increase of the G_A parameter (equation III-24). Assuming all the other parameters are constant, equation III-26 and III-27 predict that V_{Af} decreases while V_{Bf} is not a valid solution. Hence, a closed-form solution which predicts the CAM's behavior based on any parameter variation has been derived.

Now that a closed-form solution has been found, proper conductance values need to be found. The idea is to adjust the conductance so that the simplified current model approximates the more accurate non-linear current equation. This is done by equating the linear current equation to a more accurate non-linear equation. The conductance found using this method results in a first approximation. To fully predict this parameter a fitting technique is suggested. This method will be discussed in detail, later in this chapter.

First the input transistors' conductances are considered. It was concluded that these transistors are in the ohmic region; therefore, equation III-2 is set equal to the ohmic model [Johnson, ECEN 5263, Jan. 1997] as follow

$$I_{DA} = G_{A2} \cdot V_{ds} = \beta \cdot (V_{gs} - V_T) \cdot V_{ds} \quad \text{Equation III-28}$$

Where

$$\beta = \frac{k \cdot w}{l} \quad \text{Equation III-29}$$

k is the process parameter

w is the width of the device

l is the length of the device

According to the equation III-29, the conductance of M_{A2} transistor (G_{A2}) can be approximated by relating it to $\beta \cdot (V_{gs} - V_T)$. This is accomplished by taking the average gate to source voltage of this transistor during the evaluation region. Therefore,

$$G_{A2} = \beta \cdot (V_{gs_av} - V_T) \cdot k_2 \quad \text{Equation III-30}$$

Where

$$V_{gs_av} = \frac{V_{gs}(0) + V_{gs}(final)}{2} \quad \text{Equation III-31}$$

$V_{gs}(0)$ is the initial gate to source voltage potential on the transistor

$V_{gs}(final)$ is the final gate to source voltage potential on the transistor

k_2 is the fitting constant for the input transistors (M_{B2} and M_{A2})

which will be discussed later in this chapter

Equation III-31 suggests the averaging process is based on two points, initial voltage and final voltage. This technique results in a first guess approximating for the input transistors' conductance. Further manipulation will be conducted later in this chapter to get the best fit. For the input transistors M_{A2} and M_{B2} , the $V_{gs}(0)$ and $V_{gs}(final)$ stay constant. Therefore, according to the equation III-31, $V_{gs_av} = V_{ref}$ for M_{A2} and $V_{gs_av} = V_{in}$ for M_{B2} . Substituting these parameters in to the equation III-30 implies that

$$G_{A2} = \beta \cdot (V_{ref} - V_T) \cdot k_2 \quad \text{Equation III-32}$$

$$G_{B2} = \beta \cdot (V_{in} - V_T) \cdot k_2 \quad \text{Equation III-33}$$

Similar procedure is considered for the transistors M_{A1} and M_{B1} . The initial gate to source of these transistors are at zero volt, while their final voltage remains at five volts;

therefore, $V_{gs-av} = \frac{V_{dd}}{2}$. Substituting this average voltage in the equation III-30 yields

$$G_{B1} = \beta \cdot \left(\frac{V_{dd}}{2} - V_T \right) \cdot k_1 \quad \text{Equation III-34}$$

$$G_{A1} = \beta \cdot \left(\frac{V_{dd}}{2} - V_T \right) \cdot k_1 \quad \text{Equation III-35}$$

The same procedure can be used for transistors M_A and M_B . Using the averaging concept, an average gate to source over the entire evaluation period must be found. As discussed earlier in this chapter, the initial gate to source voltage across these cross coupled transistors (M_A or M_B) is equal to V_T ($V_{gs}(0) = V_T$). However, finding the average gate to source current in this case is not an easy task. The difficulty is finding the final gate to source voltage, $V_{gs}(final)$, in equation III-31. Since this voltage can vary from V_T to V_{dd} , there is no single best value to choose for the final voltage. If the final voltage is left as a parameter, equation III-23 becomes a nonlinear differential equation. In case of a nonlinear set of equations, finding a closed-form solution for the steady state voltage requires a numerical solution which defeats the purpose of deriving a simplified closed-form expression; therefore, some other technique must be used. For the first approximation, the final voltage is chosen to be $V_{gs}(final) = \frac{V_{dd}}{2}$. To approximate the

conductance, the linear model in equation III-1 is set equal to the traditional saturation equation. This is an adequate approach for the first approximation since to fully define this parameter, a fitting technique is used. This method is described later in this chapter.

$$I_D = G_{effA} \cdot (V_{gs} - V_T) = \frac{\beta}{2} \cdot (V_{gs} - V_T)^2 \quad \text{Equation III-36}$$

Equations III-31 and III-36 imply

$$V_{gs_av} = \frac{V_{dd}}{4} + \frac{V_T}{2} \quad \text{Equation III-37}$$

$$G_{effA} = \frac{\beta}{2} \cdot (V_{gs_av} - V_T) \cdot k_3 \quad \text{Equation III-38}$$

$$G_{effB} = \frac{\beta}{2} \cdot (V_{gs_av} - V_T) \cdot k_3 \quad \text{Equation III-39}$$

The fitting method employed, involves choosing the best conductance value which best fits the outcome of the closed-form solution to the simulation result (Matson, Oct. 1990). The fitting process starts by varying each of the fitting parameters in the equations III-32 through III-35, III-38 and III-39, until the best fit between the closed-form solution's steady state result (V_{Af} , V_{Bf}) and the simulation steady state result is achieved. Once these fitting parameters are found, the closed-form must yield similar results as the numerical solution (HSPICE), over all of the input range. The following are the parameters that best fit the closed-form to the HSPICE simulation

$$G_{A1} = G_{B1} = 4 \cdot \beta$$

$$G_{A2} = (V_{ref} - V_T) \cdot \frac{\beta}{2} \quad \text{Equation III-40}$$

$$G_{B2} = (V_{in} - V_T) \cdot \frac{\beta}{2} \quad \text{Equation III-41}$$

$$G_{effA} = G_{effB} = 5 \cdot \beta \quad \text{Equation III-42}$$

Above indicates that the averaging process is only used as an initial guess to the fitting process. Once the best fitting constant is achieved, that value is used in equations III-40 through III-42 to estimate the corresponding transistor's conductance. Furthermore, these conductances will remain the same over the entire CAM process.

To verify the validity of the above derivations, the closed-form solution's results are compared with simulation results. Figure III-5 through III-10 compare the simulation's results which take all of the non-linearity of the system in to account, versus the close-form's results just derived from a simple linear model.

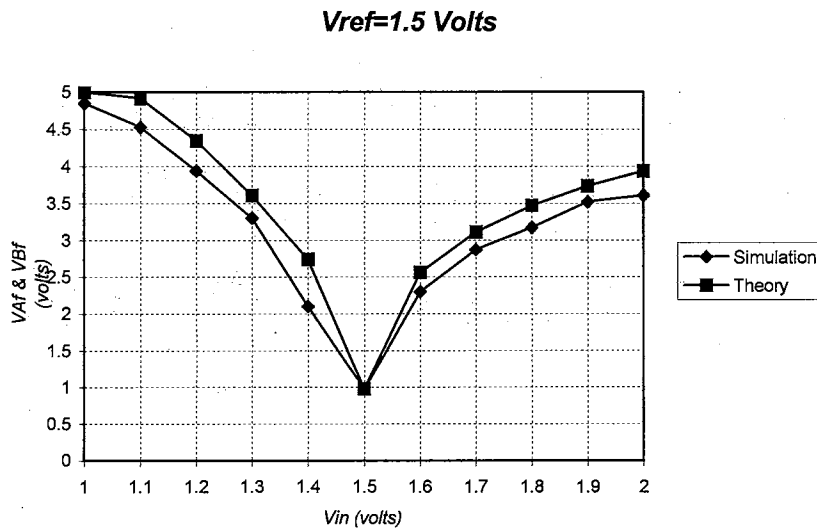


Figure III-5. CAM's Output VS. Input Voltage ($V_{ref}=1.5$ Volts)

Vref=2 Volts

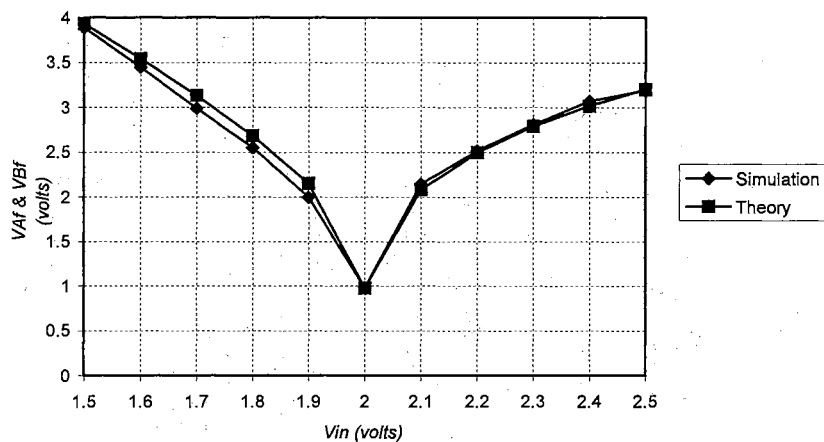


Figure III-6. CAM's Output VS. Input Voltage ($V_{ref}=2$ Volts)

Vref=2.5 Volts

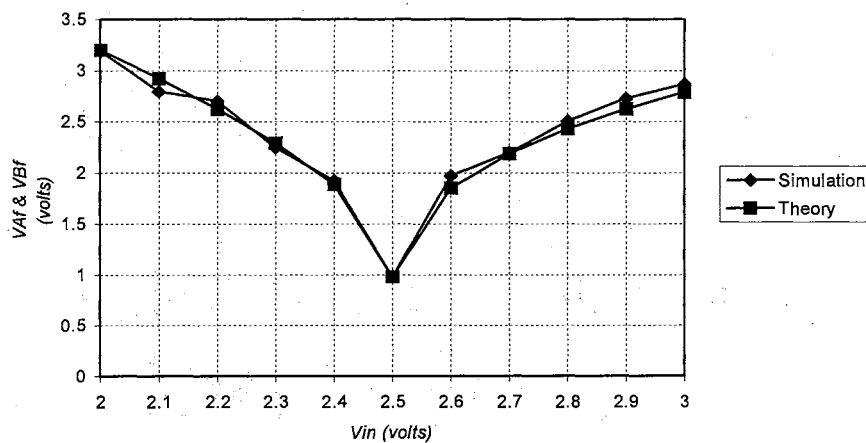


Figure III-7. CAM's Output VS. Input Voltage ($V_{ref}=2.5$ Volts)

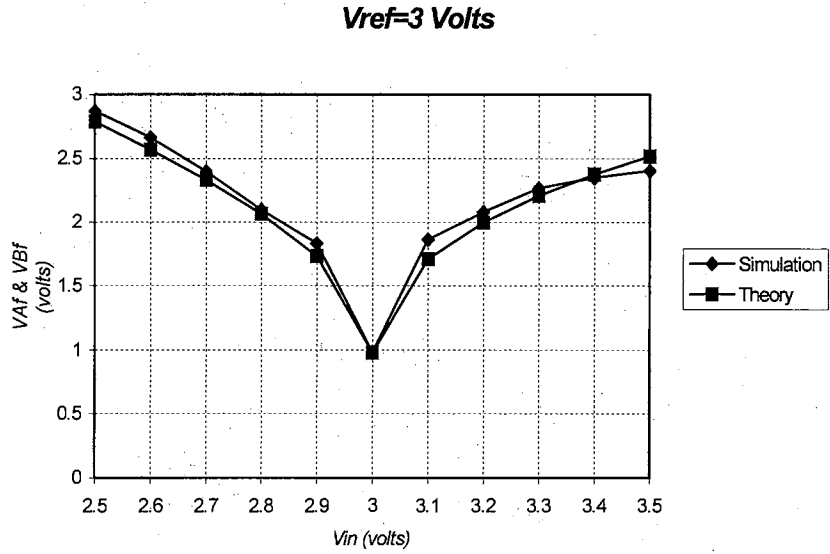


Figure III-8. CAM's Output VS. Input Voltage ($V_{ref}=3$ Volts)

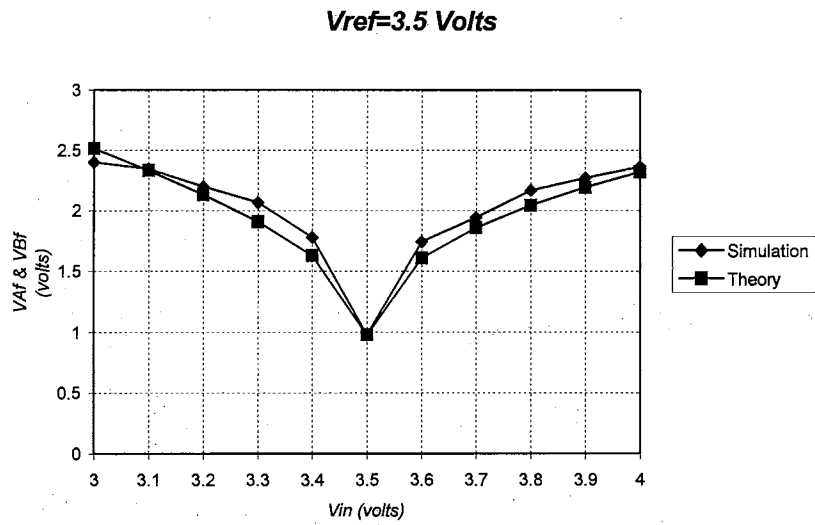


Figure III-9. CAM's Output VS. Input Voltage ($V_{ref}=3.5$ Volts)

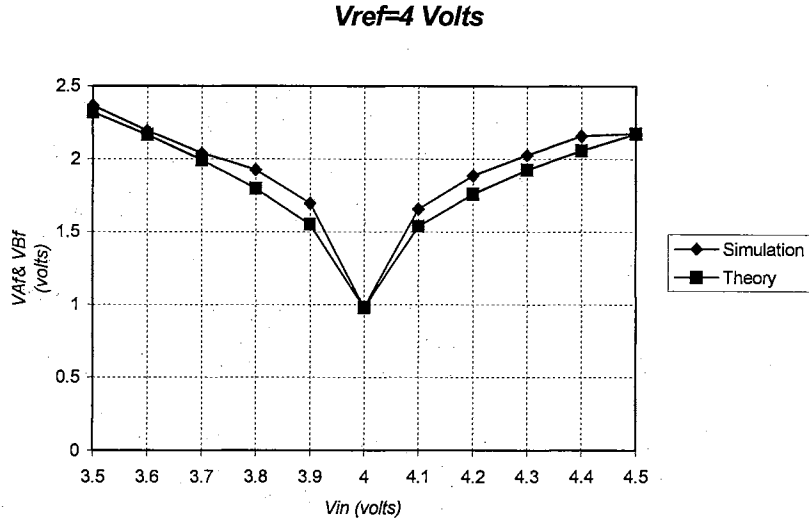


Figure III-10. CAM's Output VS. Input Voltage ($V_{ref}=4$ Volts)

The comparison is done as follow; first a reference voltage is set, then the input voltage is swept from 0.5 volts below the reference voltage to 0.5 volts above it. As the input voltage gets closer to the reference voltage, the output falls to a lower value, until the input voltage is exactly equal the reference voltage. Once that happens the absolute minimum steady state output voltage has been reached. As the input voltage deviates away from the reference voltage towards a more positive voltage, the output voltage gets larger. As shown in these figures, there is an excellent correlation between HSPICE and the closed-form solution's results despite all of the simplifications made.

Above proves that the CAM is a nonlinear cell which recognizes whether or not its input is getting closer to the pattern stored. This is very similar to a neuron's behavior. Neurons are in nature nonlinear cells which recognize the patterns once they learn what they are. The learning process and memory retention are what determine how well the neuron is able to recognize the patterns. If the learning is efficient, chances that the

neuron makes a wrong choice is far less than an inefficient learning process. The CAM cell can be trained by storing charge on a floating device connected to one of the input transistors (M_{A2} or M_{B2}) as shown in figure III-11. The learning efficiency is determined by a combination of the amount of the charge transfer in and out of the EEPROM and the method of programming each cell. Programming each EEPROM has already been discussed; however, the method of programming each cell remains to be discussed in the next chapter.

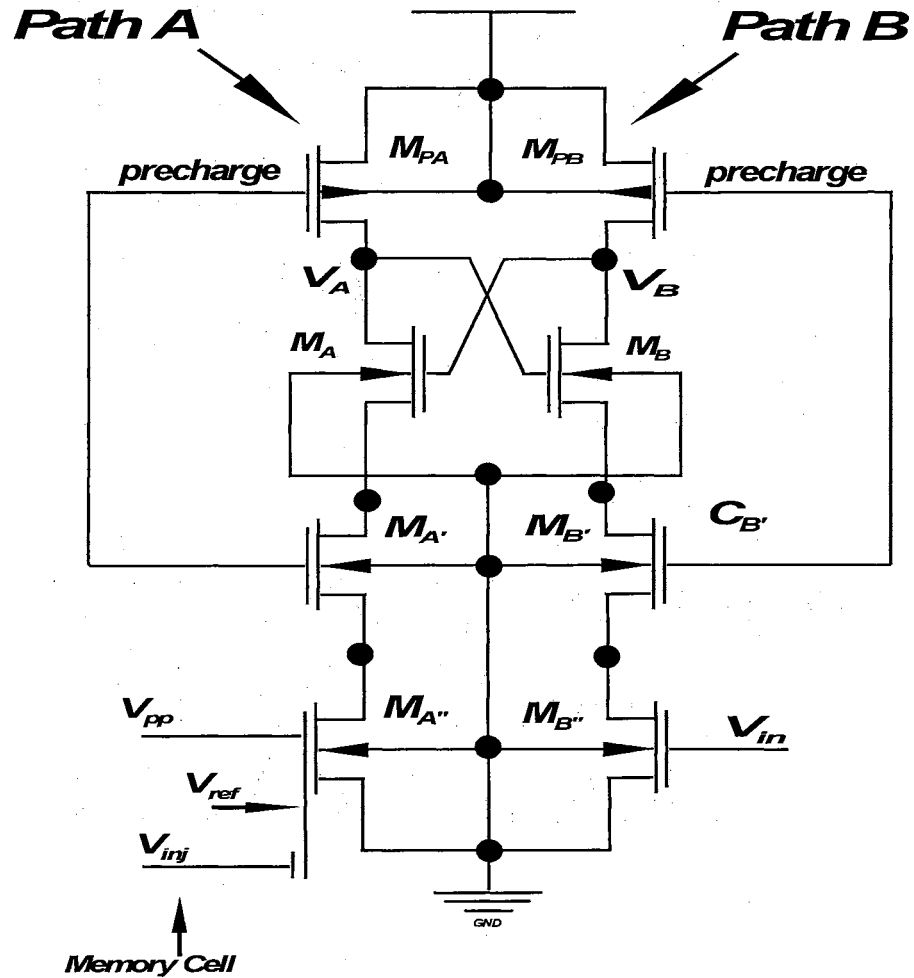


Figure III-11. The Analog CAM with Analog Memory incorporated

CHAPTER IV

THE LEARNING PROCESS

Introduction

So far the main components of the pattern recognition engine has been discussed. First, the old WTA circuit was introduced [Johnson, Aug., 1992] in chapter II, then the CAM cell was analyzed in chapter III. However, to structure the pattern recognition architecture, more controlling circuitry needs to be added. The objective of this chapter is to combine all of the pieces required to build a pattern recognition engine, and describe the programming and evaluation cycles in a detailed fashion.

Architecture

It is quite obvious that any transistor mismatch in the CAM cell will result in an erroneous result since this cell is fundamentally based on RC delay of each competing path. Therefore, an architecture insensitive to process variations such as parasitic mismatch, edge effects, and threshold mismatch is required. The idea is to be able to program each cell so that device and parasitic mismatch have no effect on the overall result. Hence, a topology which programs the EEPROM through the same path as it evaluates from is essential. [Yong-Yoong, Jun., 1996] introduced such concept by

programming different EEPROM cells using a single amplifier. Through this same amplifier, the value of the floating gate voltage just programmed is sent off-chip. Using this technique, during a read operation, the output voltage becomes independent of any process variations such as the input-offset voltage. Similar concept will be used to program information on each CAM cell. It will be shown that this topology manifests these process variations as a DC offset associated with each cell. Therefore, as far as the end user is concerned, there is no effect from process variation. This concept will be explained in detail once the programming and evaluation modes are clearly understood.

As discussed in chapter III, an analog memory (EEPROM) is incorporated into each CAM cell. To program these analog memories, high voltage circuitry [Yang-Yoong, Dec., 1994] must be used. Figure IV-1 depicts a 2x2 array of CAM cells with their corresponding high voltage circuitry. To program a single cell, one has to select the row and the column where the memory cell of interest resides. Doing so allows a high voltage potential across the floating gate of interest and its injecting gate. The rest of the cells have mid-voltage on at least one of their terminals; therefore, no programming is done on these cells since the electric field does not get high enough for any of the Fowler current to flow. The complete programming process of an EEPROM is explained in detail in [Yong-Yoong, Dec., 1994]; however, in this thesis, the programming procedure of the CAM cells will be explained in detail. Moreover, a new approach is taken in implementing the high voltage circuitry.

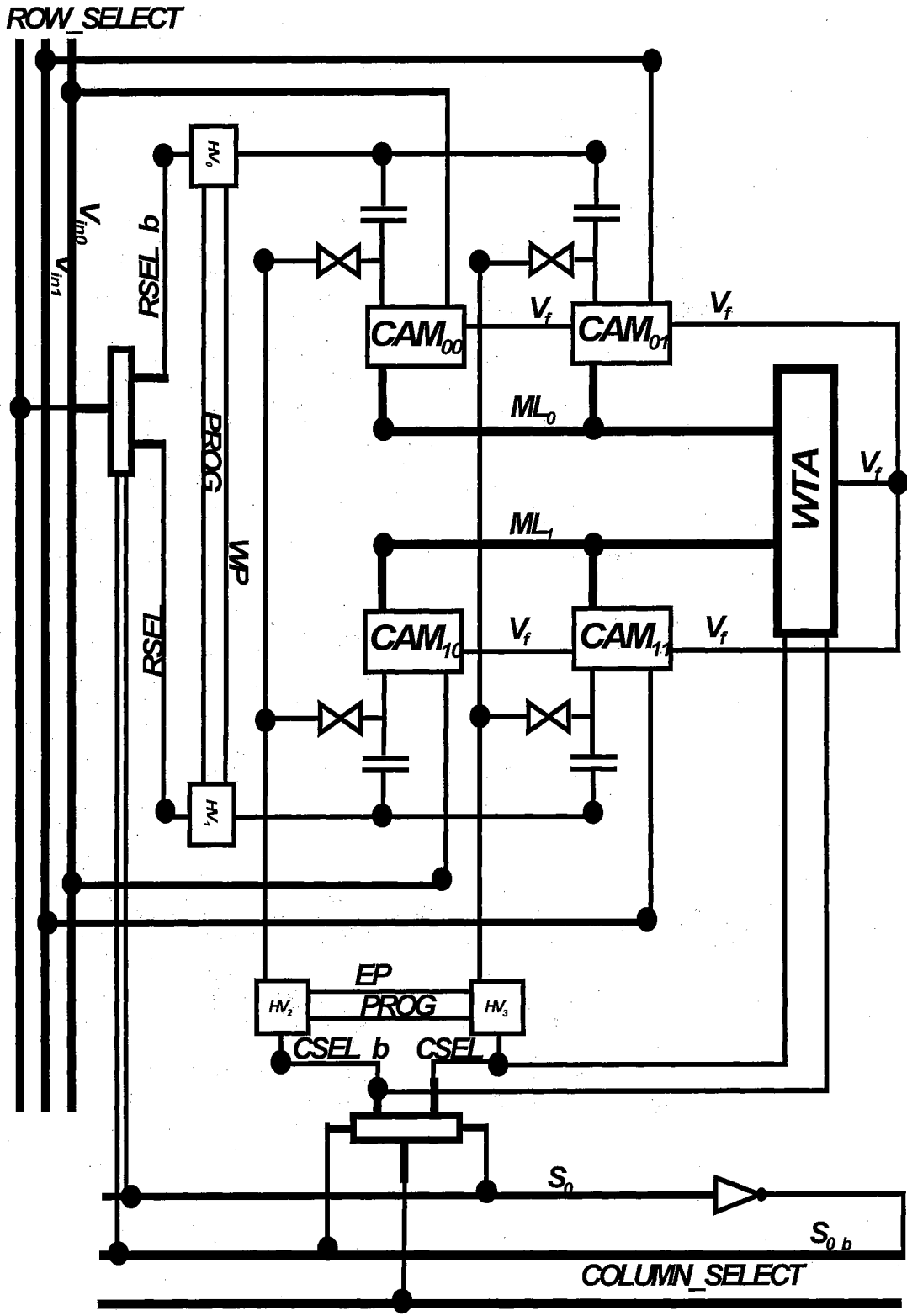


Figure IV-1. The Recognition Engine

Programming Process

It was discussed in chapter II that any competitive classifier requires a discriminator to determine the closest match. It was also explained in chapter III that the CAM's performance is based on an RC delay concept; therefore, the capacitance loads seen by the two output of each CAM cell (V_A and V_B) is important to remain the same while in the programming and the evaluation mode. Furthermore, to program each CAM cell to the desired pattern, extra circuitry besides the discriminator circuit is required. To eliminate the need for this additional circuitry and to make sure that the CAM's output sees the same capacitive loading whether in the programming mode or in the evaluation mode, the discriminator circuit is converted to a new circuit by applying an external signal called *start_prog*. Forcing this signal to V_{dd} indicates that the recognition engine is in the programming mode allowing the programming process to be monitored through the match line, where final evaluation is done from. This section describes how the discriminator circuit is set in to the programming mode, and how each CAM cell gets programmed.

A winner take all circuit was chosen as the discriminator in this thesis. Figure IV-2 depicts this circuit used both to evaluate and to program each EEPROM cell. The programming process starts with the *start_prog* signal set high. Then, the corresponding row and column where the EEPROM cell resides need to be selected. For instance for demonstration purposes, lets assume the CAM cell located in the first row and first column is to be selected and a write operation is to be performed on it. To do so, the EEPROM connected to the CAM_{00} is selected by both column and row signals set to zero,

while S_0 is set high. Doing so enables HV_0 to pass the high voltage [Yong - Yoong, Jun., 1996]. Selecting the first column turns the pull downs on the first column on (through the M_{n1} and M_{n2} transistors) while the out put of the high voltage driver (HV_2) is set to ground. This forces a high voltage potential across the floating gate connected to the CAM_{00} while the rest of the cells have mid-voltage potential across their terminals [Yong - Yoong, Jun., 1996]. This causes the first match line (ML_1) to be controlled only by the selected CAM cell (CAM_{00}). Furthermore, while programming, it is essential that non of the gate to source potential voltage of the transistors in the selected path vary except the M_c transistors. Doing so ensures that the match line is only controlled by the output of the CAM. As a result, while programming, any change in the value of the match line indicates a change in the charge stored on the EEPROM connected to the CAM. Therefore, it is important that the feed back voltage does not contribute to the programming process. This is done by disconnecting the feed back regulators from the feed back voltage (M_{n1} , M_{n2} in figure IV-2) and connecting their gates V_{dd} which forces the feed back transistors in to the ohmic region. Once in this region, these transistors act like resistors connected to the source of M_{c1} and M_{c2} , implementing a follower. To disconnect the feed back voltage from the regulating circuit, the *start_prog* signal (figure IV-2) needs to be asserted.

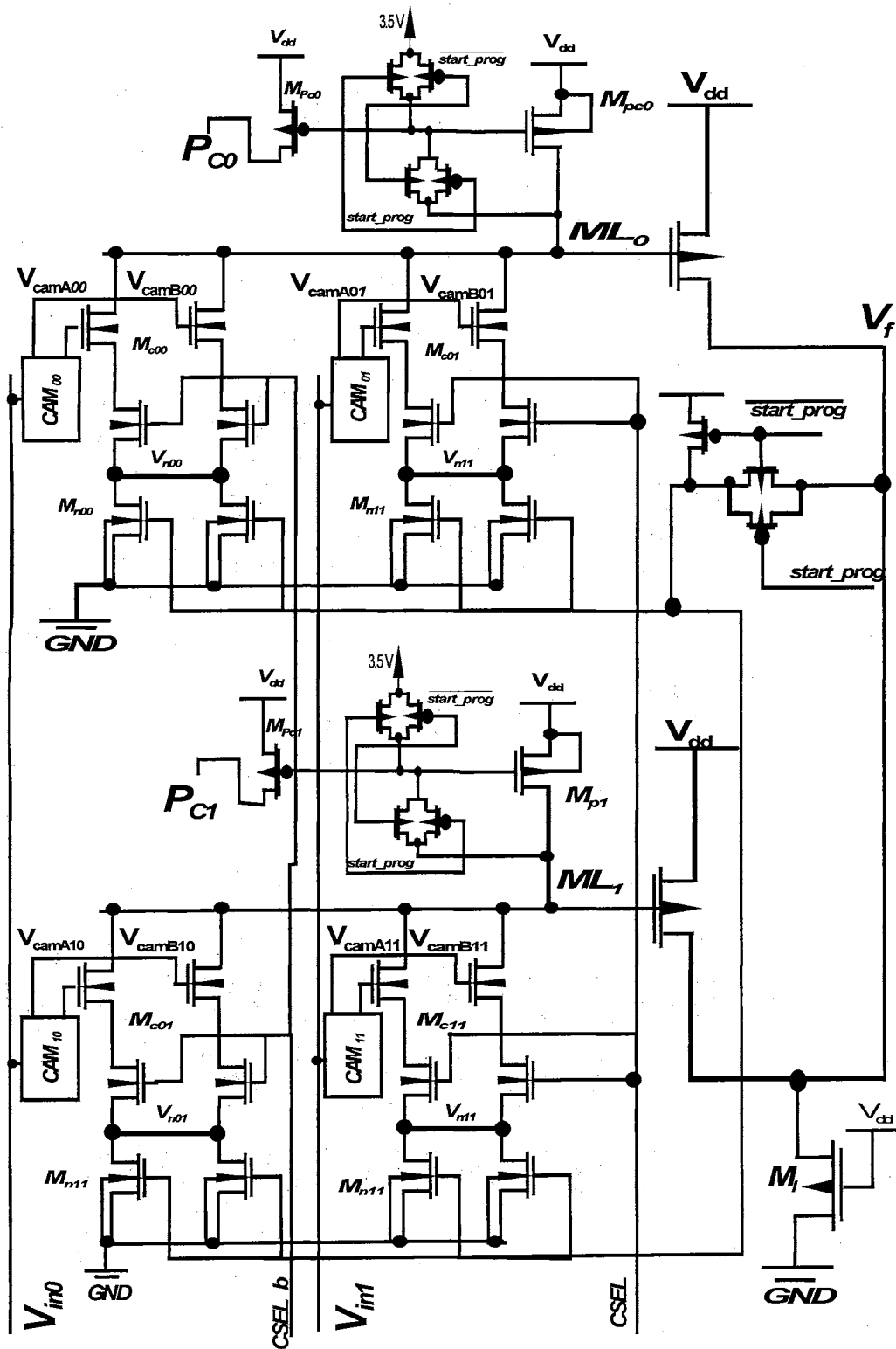


Figure IV-2. The Winner Take All Circuit

Once the CAM has reached its final steady state value, the corresponding CAM's output voltage minus a gate to source voltage (V_{gs}) drop of M_{cij} is reflected on the drain of M_{nl} ($V_{nij} = V_{Bij} - V_{gs_M_{cij}}$) which sets the current drawn from the match line. To be able to measure this current, the p-channel current source (M_{pci}) is converted to a diode connect device, mirroring this current to another p-channel (M_{poi}) device for measurement. This conversion is done once the *start_prog* signal is set high. This whole process converts the CAM's output voltage to a proportional current. Figure IV-3 shows a single CAM cell under program with the WTA circuit in programming mode.

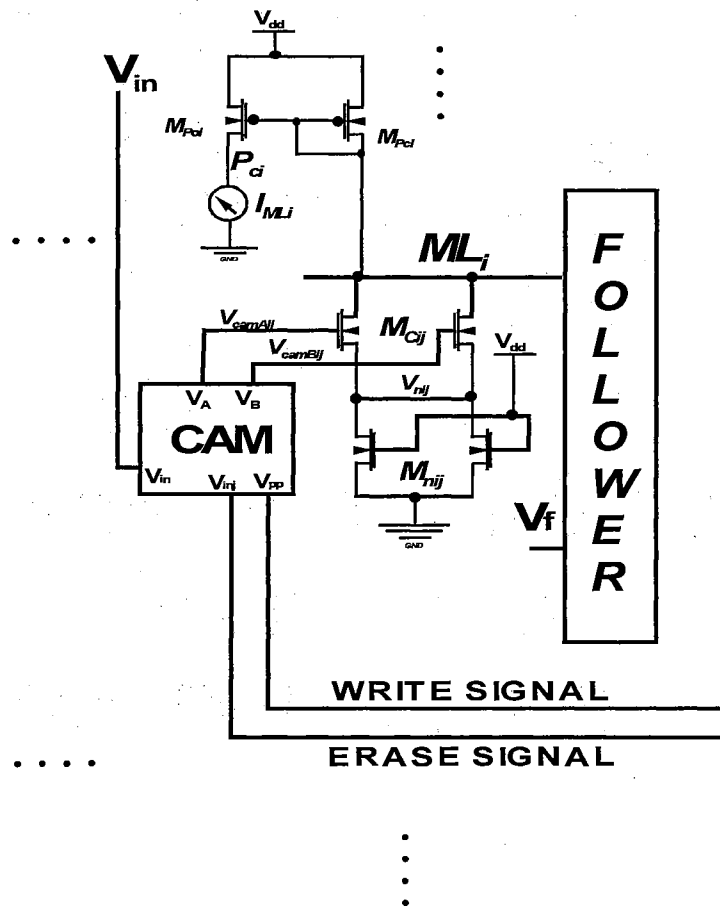


Figure IV-3. An equivalent engine's diagram while programming a single CAM

For illustration purposes, suppose the pattern voltage that needs to be programmed, is some voltage V_{ref} . Furthermore, it is assumed that there is no charge on the EEPROM. To program this cell from zero volts to V_{ref} (a positive voltage) an erase operation must be conducted. To do so, high voltage pulses are applied to the injecting gate while the controlling gate is grounded. As more pulses are applied to the cell, the stored value increases to a more positive value. As this voltage increases towards the reference voltage, the CAM cell's output voltage drops. This process continues until the minimum output value is encountered, corresponding to a perfect match. Further programming causes the CAM's output voltage to deviate from its minimum to a higher value. Figure IV-4 shows this process.

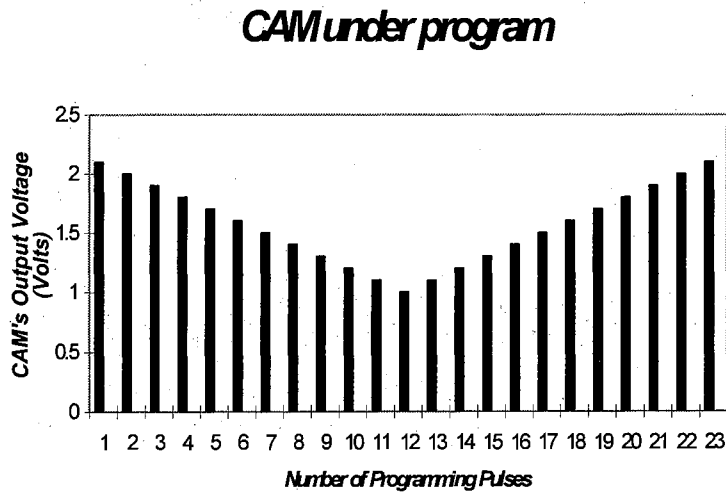


Figure IV-4. The Analog CAM under program

As the programmed voltage on the EEPROM cell gets closer to V_{ref} , the gate to source potential difference of the pull down device in the winner take all circuit (M_{cij} , figure IV-3) also decreases. This voltage is converted to a proportional current thanks to

this follower configuration (M_{cij} and M_{nij}). Figure IV-5 superimposes an example of the output of the CAM cell and the corresponding current generated through the WTA circuit.

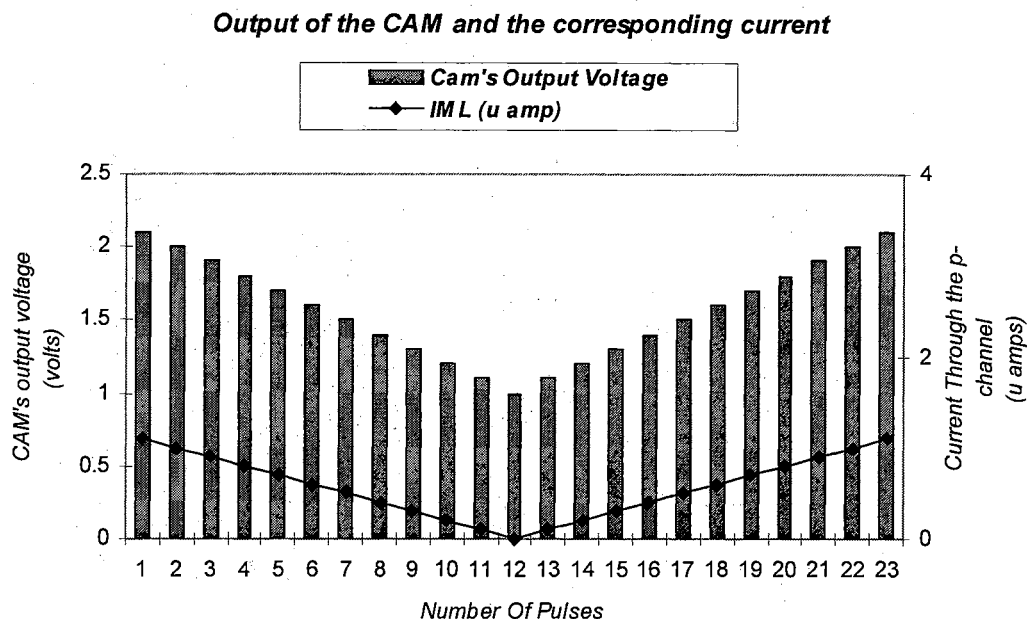


Figure IV-5. CAM's Output voltage and the corresponding match line voltage

Looking at figure IV-5 from the 0th pulse to the 12th pulse, it is clear that as the programmed voltage gets closer to the reference voltage, the match line current decreases. Therefore, with every pulse, the present current measured is smaller than the previous one. As long as this is true an erase operation ought to be performed. The moment the minimum current is reached, the programming process needs to stop.

Based on this behavior, the controller is to decide whether or not the CAM needs more programming while indicating the programming direction based on the behavior of the CAM. Once that has been established, this circuit will signal the high voltage circuitry to program the selected cell, in the appropriate direction [Yong-Yoong, Jun., 1996]. The

detail implementation of this procedure will be explained in the following chapter.

Controller

Knowing how the CAM responds to each programming pulse, a programming control circuit is proposed. During the programming process, the proposed controller is able to determine the programming duration and direction. This is done by recognizing the previous state at which the CAM had been in, and compare it with the present state, then decide whether or not the CAM cell needs to be programmed.

To built this controlling circuit, first the match line current generated by the configuration shown in figure IV-5, needs to be converted back to a voltage. This is accomplished by forcing this current through a resistor as shown in figure IV-6. Since this voltage is analog, to store the present and previous states, a sample and hold circuit is used. Then, these analog voltages are converted to digital level voltage using a high gain comparator (figure IV-6). These states are then stored in digital storage devices which are then fed in to a digital circuitry in order to decide the direction and duration of the programming process. A fix duration programming signal is asserted as long as the last state and the present states of the digital storage devices are the same. As soon as the two states are different, the programming signal is set low, indicating a stop in the programming process. [Yong-Yoong, Jun., 1996] explains in detail how the direction of programming is determined; therefore, no more explanation is given in this thesis.

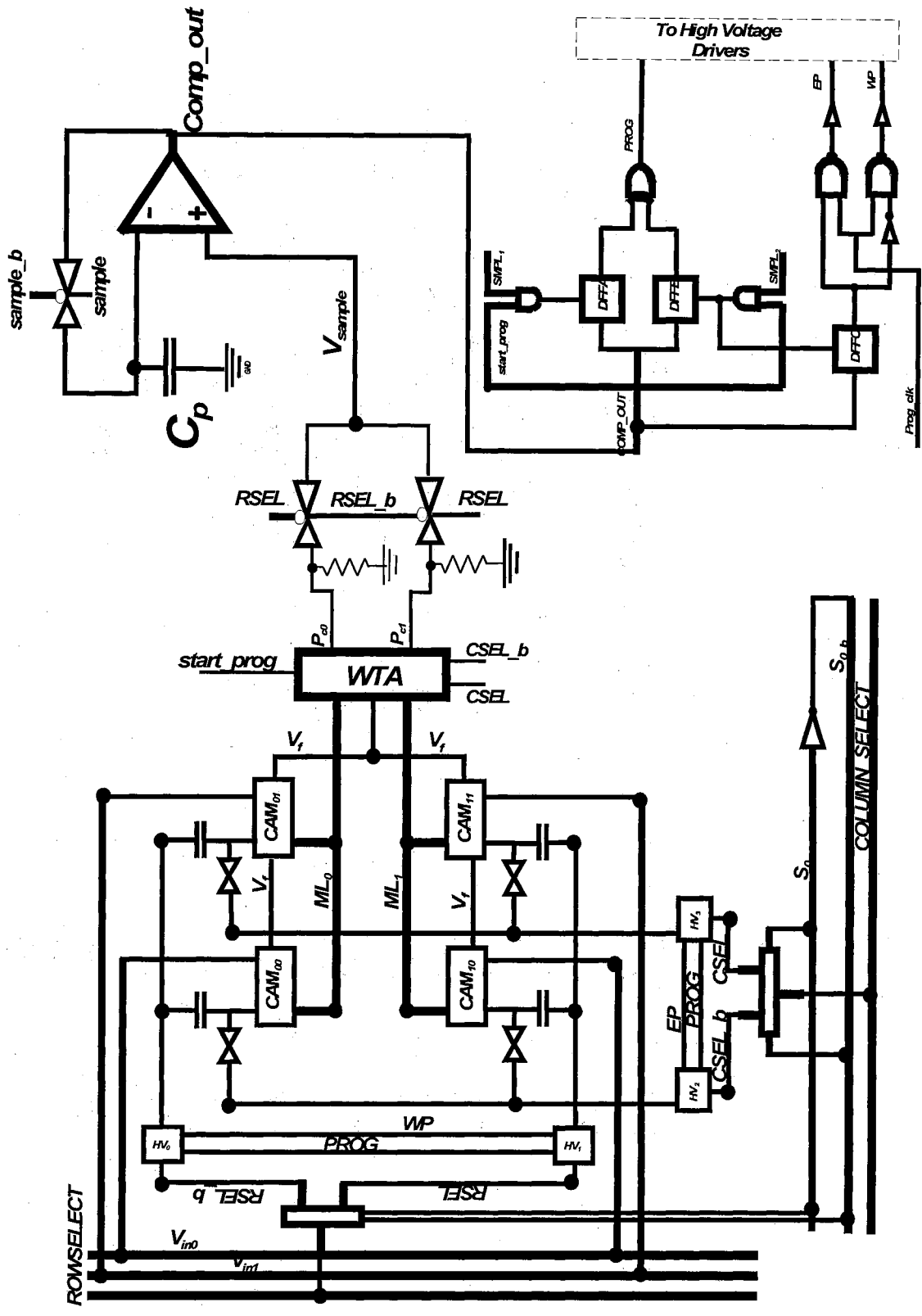


Figure IV-6. The Recognition Engine and the Controlling Circuitry in the programming mode

As shown in figure IV-6, this architecture consists of four major sub-blocks, high voltage circuitry, an analog sampler, an analog to digital sample converter, and a digital logic circuit, determining the direction and duration of the programming process. The novelty of this architecture is in the samplers and the high voltage circuitry. To start the programming process it is important to know the starting state. The best state to start with is when there is no charge or a charge equivalent less than the threshold voltage of the floating gate transistor (M_{B2} in figure III-10) exists on the floating gate. Once in this state, there is only one direction that the EEPROM needs to be programmed which is in the erase direction. However, in silicon implementation of EEPROMs, there might be positive charge trapped on the floating gate. To ensure this charge has been removed, a write operation on the EEPROM must be conducted until the floating gate transistor is turned off.

Assuming the EEPROM has no charge stored on it, the floating gate transistor is off; therefore, a precharge and evaluation operation will result in some high voltage at V_B node while node V_A discharges to zero. This ensures that one of the pull downs (M_{cij}) in the WTA circuitry is on; which sets the current through the corresponding match line. This current is converted to voltage (V_{sample}) which is compared with the stored voltage on the capacitor at the negative terminal of the comparator (C_P in figure IV-6). If V_{sample} is larger than the stored value on C_P , the output of the comparator goes to a high state; otherwise, the output is low. Finally, the comparator's output is sampled at the end of every other evaluation cycle, respectively via sampling clocks $smp1_1$ and $smp1_2$, on the two digital storage elements DFF_A or DFF_B . This allows a comparison between the previous and present states. The program signal is asserted high if the two states are the

same; however, if the two are different, this signal is set low indicating a stop in the programming process. Prior to programming the CAM further, it is necessary to transfer the present state (V_{sample}) to the sampling capacitor C_P . Doing so makes V_{sample} the previous state. This is done by an extra sampling cycle while not asserting a programming pulse to the EEPROM under program.

To understand the complete process it is helpful to go through a complete programming cycle. Suppose the programming mode is an erase operation and the CAM has just been programmed with the 10th pulse (figure IV-5). The corresponding voltage (V_{sample}) is compared with the value stored on the sampling capacitor (the output of the 9th pulse) and is converted to the appropriate digital level using the high gain comparator. Prior to applying the next pulse, this digital state is stored by a digital storage device such as a *DFF* (DFF_A in figure IV-7) using $smpl_1$ clock. Next, the analog voltage corresponding to the 10th pulse is stored on the sampling capacitor (C_P) prior to applying the 11th programming pulse. Once the 11th pulse is complete, the old state's voltage value (the 10th pulse) and the new one (the 11th pulse) are compared at the input of the comparator. If the previous voltage is larger than the present one, the comparator switches low; otherwise, the output is high. The outcome at this point is stored on the second *DFF* (DFF_B , in figure IV-7) using $smpl_2$ clock, indicating the present digital state. Since the two digital states are the same, the programming process continues. The programming direction is determined by a separate circuitry which samples the output of the comparator every other cycle by the $smpl_2$ clock on DFF_C (figure IV-7). It is easy to see that if the stored digital voltage is low on DFF_C (when the previous analog state is higher than the present one), the erase signal operation is asserted; otherwise, a write signal is

enforced. The same procedure is continued until the 13th pulse is applied. At this point the present state is higher than the previous state; hence, the output of the comparator is set low. This value is stored on the DFF_A (using $smpl_1$) which causes the program signal to go low. This stops the programming process, indicating a programmed cell. A detailed clocking of the programming scheme is shown in figure IV-8.

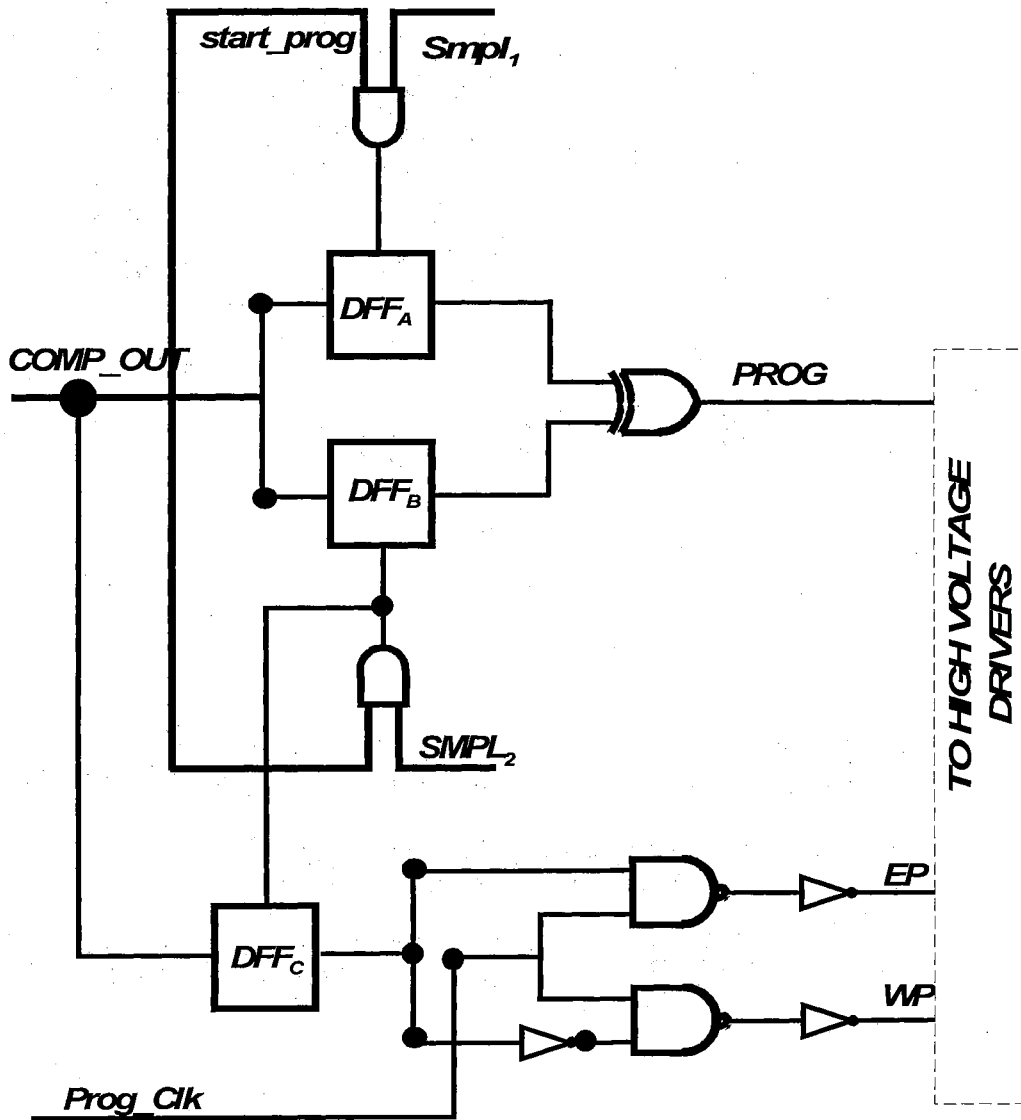


Figure IV-7. The Programming Pulse Controller

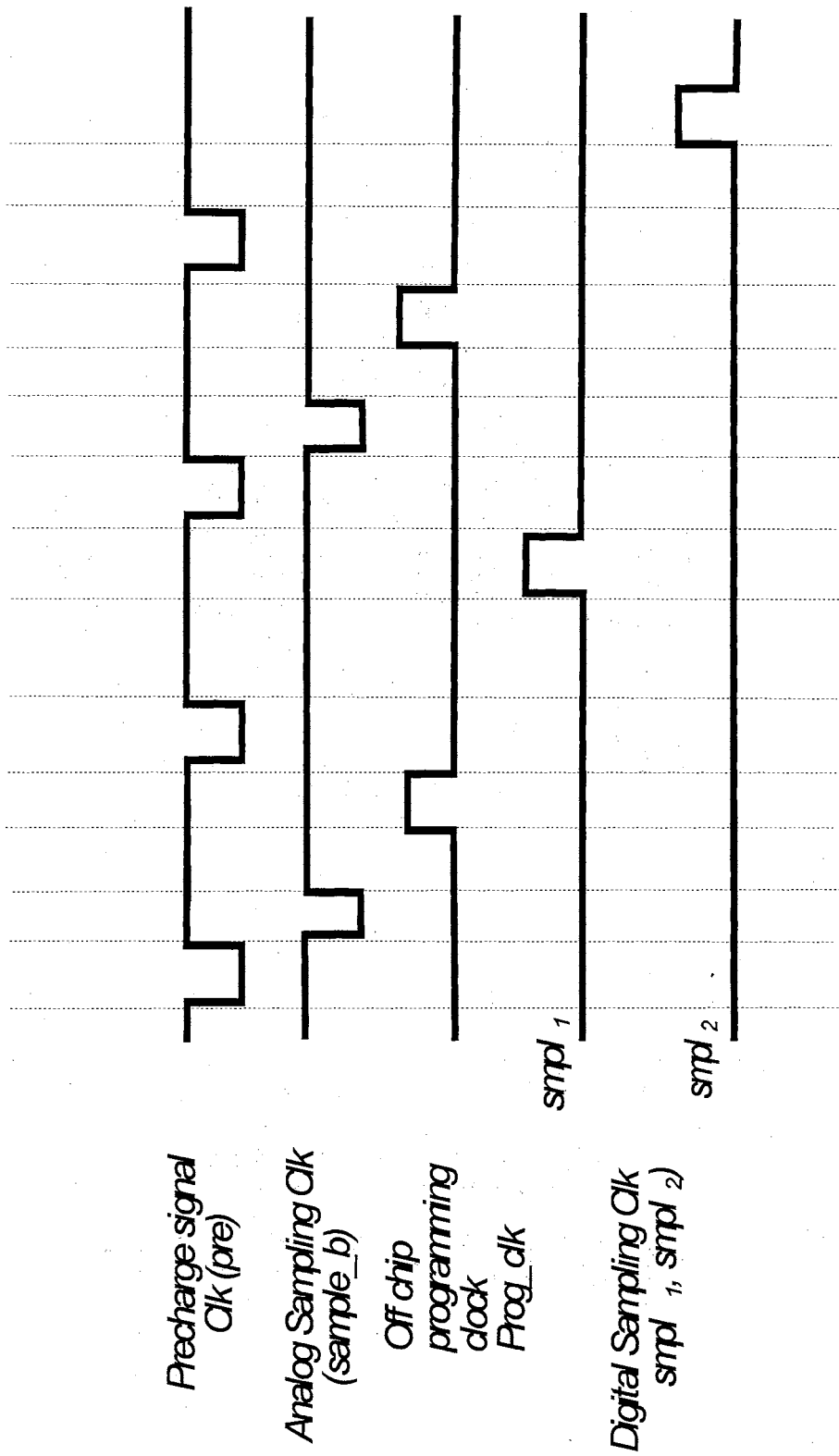


Figure IV-8. Clocking Sequence for programming each pixel

The sampling method that was just discussed must be offset free since any offset that exists in this portion of the path will be directly stored on the EEPROMs. However, it is well known that having a circuit free of any physical offset is impossible; therefore, techniques must be used to eliminate the effect of any offsets. These offsets attributed to many different sources, such as threshold mismatch, oxide variation, and edge effects. To eliminate such offsets, the technique shown in figure IV-9 can be used. This technique uses the well known auto-zeroing technique concept without an extra clock cycle. Once a voltage is sampled using this technique, any offset referred to the input is also stored on the capacitor. Once the next value is compared to the stored value, the stored offset cancels the input referred offset.

Figure IV-9 depicts a comparator with an input referred offset modeled as a voltage source at its positive terminal. Please note that this offset is completely random and it could have a positive or a negative value and it does not effect the concept.

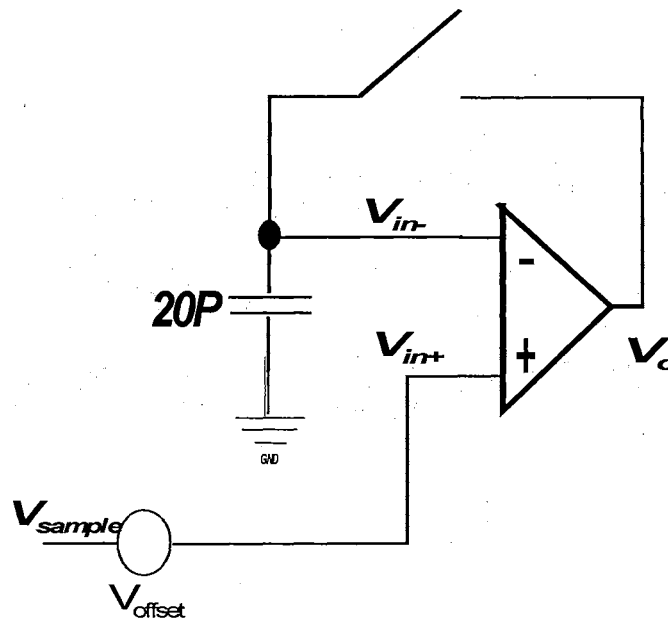


Figure IV-9. The Sampler

The sampling process starts with closing the switch. Once the switch is closed and the comparator is put in the unity feed back configuration, the input voltage is equal to

$$V_{in+} = V_o \cdot \left(\frac{1}{AV} + 1 \right) \quad \text{Equation IV-1}$$

where

V_{in+} is the virtual ground of the comparator

AV is the open loop of the comparator

Assuming $AV \gg 1$ and $\frac{1}{AV} \approx 0$

$$V_{in+} = V_o \quad \text{Equation IV-2}$$

Writing V_{in+} in terms of the input voltage (V_{sample}) and the offset voltage,

$$V_{in+} = V_{sample} - V_{offset} \quad \text{Equation IV-3}$$

Substituting above in equation IV-2 and solving for V_o gives

$$V_o = V_{in-} = V_{in+} = V_{sample} - V_{offset} \quad \text{Equation IV-4}$$

This is the voltage stored on the sampling capacitor (C_P) at the end of each cycle. Now assume the switch is open, the charge remains stored on this capacitor until the switch closes. In reality this capacitor leaks through junction leakage; however, its time

constant is much lower than sampling cycle; therefore, it is ignored.

Now assume that a new voltage (V_{sample_new}) is applied to the input of the comparator while the sampling switch is open. At this time, the output voltage can be written as

$$V_o = [(V_{sample_new} - V_{offset}) - (V_{sample_old} - V_{offset})] \cdot AV \quad \text{Equation IV-5}$$

$$V_o = (V_{sample_new} - V_{sample_old}) \cdot AV \quad \text{Equation IV-6}$$

Above indicates that offset voltage does not effect the final state of the comparator. This is an extremely important property of this architecture which allows programming each CAM cell without worrying about offset and mismatch issues in the sampling circuit.

Initial Clocking

As pointed out earlier in this chapter, it is important to ensure that the programming procedure starts from a known state; otherwise, cells can be programmed in the wrong direction or always be stuck in the starting state. It was also indicated that the programming sequence starts in the erasing mode. To do so, first the sampling capacitor C_P is discharged to zero, then the CAM cell under program is precharged. During the evaluation cycle this cell sets V_{sample} to a voltage higher than ground voltage causing the output of the comparator to go high. At this point both clocks $sample_1$ and $sample_2$ are applied simultaneously. Doing so causes the input of the XOR gate to have the same state causing the program voltage to be asserted (figure IV-7). At the same time, the output of

the $DFFC$ is set to a high voltage enforcing the erase signal. From here on regular programming is assumed. Figure IV-10 shows the first period clocking scheme along with the regular clocking.

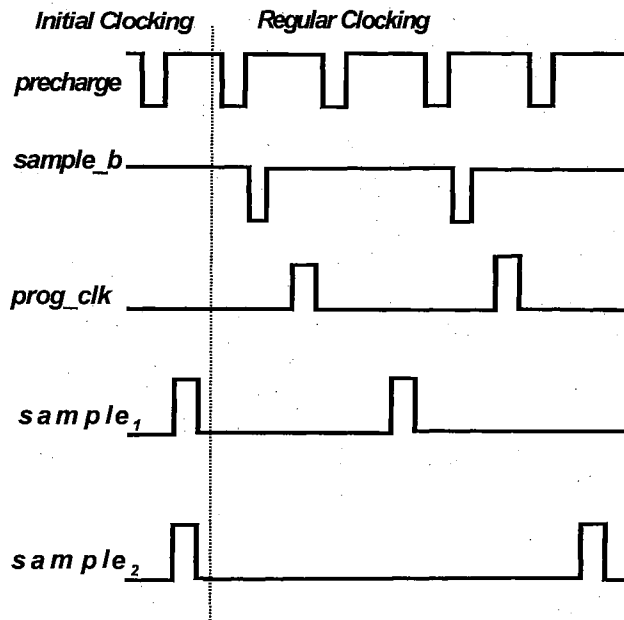


Figure IV-10. The Complete Clocking Scheme

More Mismatch

So far all of the programming and results were based on ideal and well matched devices. However, it is clear that in real life there is a variety of mismatches to be accounted for. These mismatches should be considered and quantified. Previously, it was assumed that the two paths in the CAM cell are matched. Furthermore, from the discussion in chapter III, it is obvious that any mismatch in the competing paths results in an erroneous result. It was also pointed out, while programming, the CAM's response is

measured from the match line the point at which the CAM evaluates the pattern stored against the input voltage; hence, the error becomes irrelevant. The key point to remember is that it does not matter what the stored value is, as long as the CAM cell results in the best match when compared with the input pattern.

As an example, assume the parasitic capacitance on one of the output nodes (C_A or C_B) is smaller than the other, and the reference voltage (V_{ref}) is set equal to the input voltage. In an ideal CAM cell, both paths decay at the same rate to about a threshold above ground (when both cross coupled transistors turn each other off). Now that there is a mismatch between the two paths, the path with the smaller capacitance decays at a faster rate. This difference gets enhanced in a hurry, thanks to the feed back mechanism (M_A and M_B). Therefore, the outputs do not decay with same rate anymore; rather, the path with the smaller capacitance decays to ground leaving the other path at some higher value. Figure IV-11 depicts a possible case that might happened.

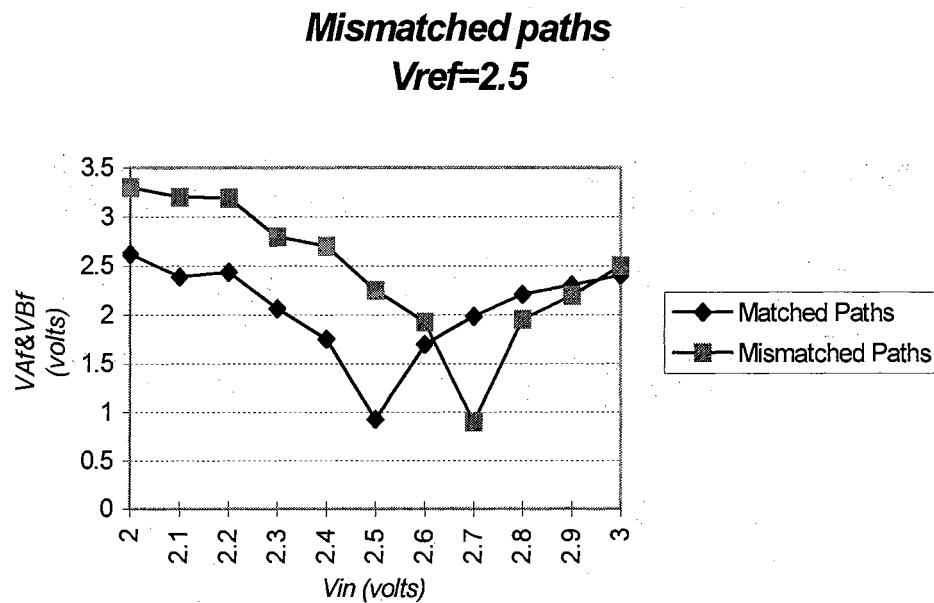


Figure IV-11. CAM's output vs. input voltage

Figure IV-11 assumes, due to the mismatch between the two paths, to store 2.5 volts using this architecture, a 2.7 volts needs to be stored on the EEPROM. Obviously, this is not the intended value that was suppose to be stored. Now suppose the same location where the EEPROM had been programmed from, is used to evaluate the stored value against the input value. As far as the CAM is concerned, a 2.5 volts matches 2.7 volts since a 2.5 volts on one side and a 2.7 volts on the other side results in a perfect match. That means, once the input is at 2.5 volts an exact match is encountered. This implies no matter what the offset is, the exact match is realized if the stored value had been programmed through the same location (as long as the nature of the offset does not change with voltage change). Therefore, this mismatch can be viewed as a mapping function that is transparent to the end user. This is the most important characteristic of this architecture which gives the designer the freedom to chose a very inexpensive process.

CHAPTER V

THE HIGH VOLTAGE CIRCUIT

Introduction

The amount of charge transferred in and out of the floating gate of the EEPROM must be controlled while programming. Since thickness of the oxide varies across the die and from one wafer to the next, different voltage amplitudes might need to be applied to get the same charge flow in and out of different floating gates. This voltage can range from 14 to 25 volts which is quite high and in some cases (voltages higher than 16 volts) surpass the physical process limitations in most of the standard process. For instance, in the process under use (2.0u ORBIT Process), junction breakdown is around 14 to 16 volts; therefore, regular circuitry with regular MOSFETs can not be used. Hence, techniques must be used to create a high voltage tolerant circuit.

There have been numerous papers published on how to implement high voltage circuits [Donly, Pasternak, Dec., 1986]. Some [Parpia, Dec., 1986] use special processes which can be quite costly; hence, a more cost effective solution is in order. Some [Parpia, Salama, Oct., 1988] use layout or circuit techniques to implement these devices. This chapter introduces a unique method to implement a high voltage n-channel MOSFET using a regular process with no special doping or additional mask layers. First the existing layout technique is discussed, then the novel transistor is introduced. Finally, a

true high voltage driver is introduced and characterized.

High Voltage Transistor

A regular transistor built in the 2.0u ORBIT process can only handle voltages as high as 14 volts on its source and drain and not more than 15 volts on its gate. It is well known that the drain/source of a regular transistor is created by implanting highly doped N+ into a lightly doped substrate. Where these materials contact each other, majority carriers cross the boundary leaving behind ionized atoms in a region depleted of free carriers. This simple charge depletion model gives a breakdown voltage of

$$BV = \frac{\epsilon_{si}}{2q} E_c^2 \left(\frac{1}{N_p} + \frac{1}{N_n} \right),$$
 where N_p , N_n are the doping concentration on the P and N

sides of the junction, E_c is the maximum electric field that can exist across the depleted junction, and ϵ_{si} is the dielectric constant of the silicon. To get high break down voltage (BV), low doping on both sides of the junction is best. However, changing doping is not an appropriate solution since it requires a special mask in the fabrication process and adds cost.

The next issue to consider is oxide break down. In a regular transistor the oxide formed between the gate and the channel is a thin oxide with maximum gate break down voltage of approximately 15 volts. Moreover, if the voltage potential difference between the gate and any other terminal (drain, source or body) exceeds this voltage the oxide will break down.

Figure V-1 depicts the cross section of a high voltage n-channel transistor introduced in [Parpia, Dec., 1986]. This transistor is essentially the same as a regular transistor with a different drain implementation. The transistor employs a layout technique to increase both the PN junction break down voltage as well as increasing the gate to drain oxide breakdown voltage. The PN junction breakdown is increased by surrounding the highly doped diffusion by a lightly doped N type material (N-well) which increases the effective resistance from drain to channel and substrate. Furthermore, this device is able to handle a high voltage potential between gate and drain thanks to the thick oxide in between these two terminals.

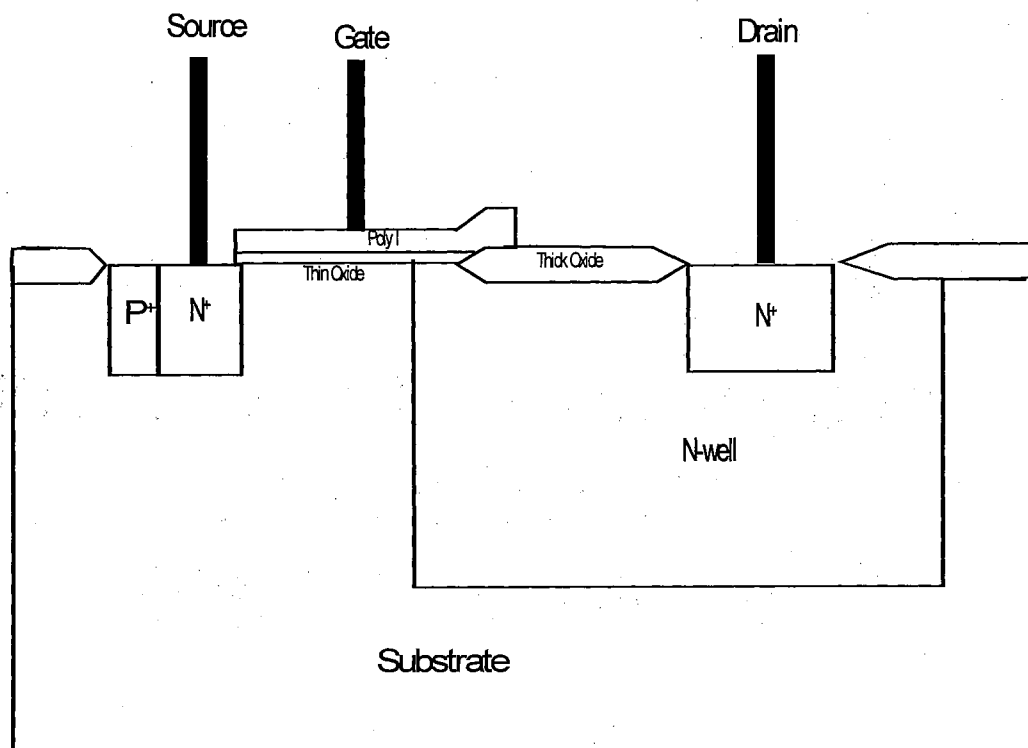


Figure V-1. An Extended Drain Transistor

The above topology is adequate yet does not implement a true high voltage transistor. To use this transistor as a high voltage pass transistor, one has to make sure the gate to source voltage does not exceed the thin oxide break down voltage; therefore, extra circuitry is required to control the gate to source voltage. A true high voltage transistor must endure high voltage both on the source and drain side while being able to sustain high voltage on the gate. Such transistors can be used to implement high voltage EEPROM drivers as discussed in [Yong-Yoong, Dec., 1994] with much higher break down voltage.

A Complete High Voltage Transistor

Figure V-2 depicts the side view of the proposed transistor. This transistor is designed so that all of its terminals can tolerate a high voltage potential. Both drain and source are implemented by surrounding the highly doped material (N_+) with a lower doped material (N-well) in order to increase the breakdown voltage to a much higher voltage. In order to be able to sustain high voltage on the gate, thick oxide has been used. To create thick oxide underneath the gate, a clever layout trick is implemented. It is known that once a poly layer overlaps active area, a thin oxide is created. Therefore, to prevent creating thin oxide it is made sure that the gate never overlaps the active area (as shown in figure V-2). Now that the gate thickness is roughly 15 times more than a regular gate thickness a much higher voltage can be applied to the gate before the oxide breaks down.

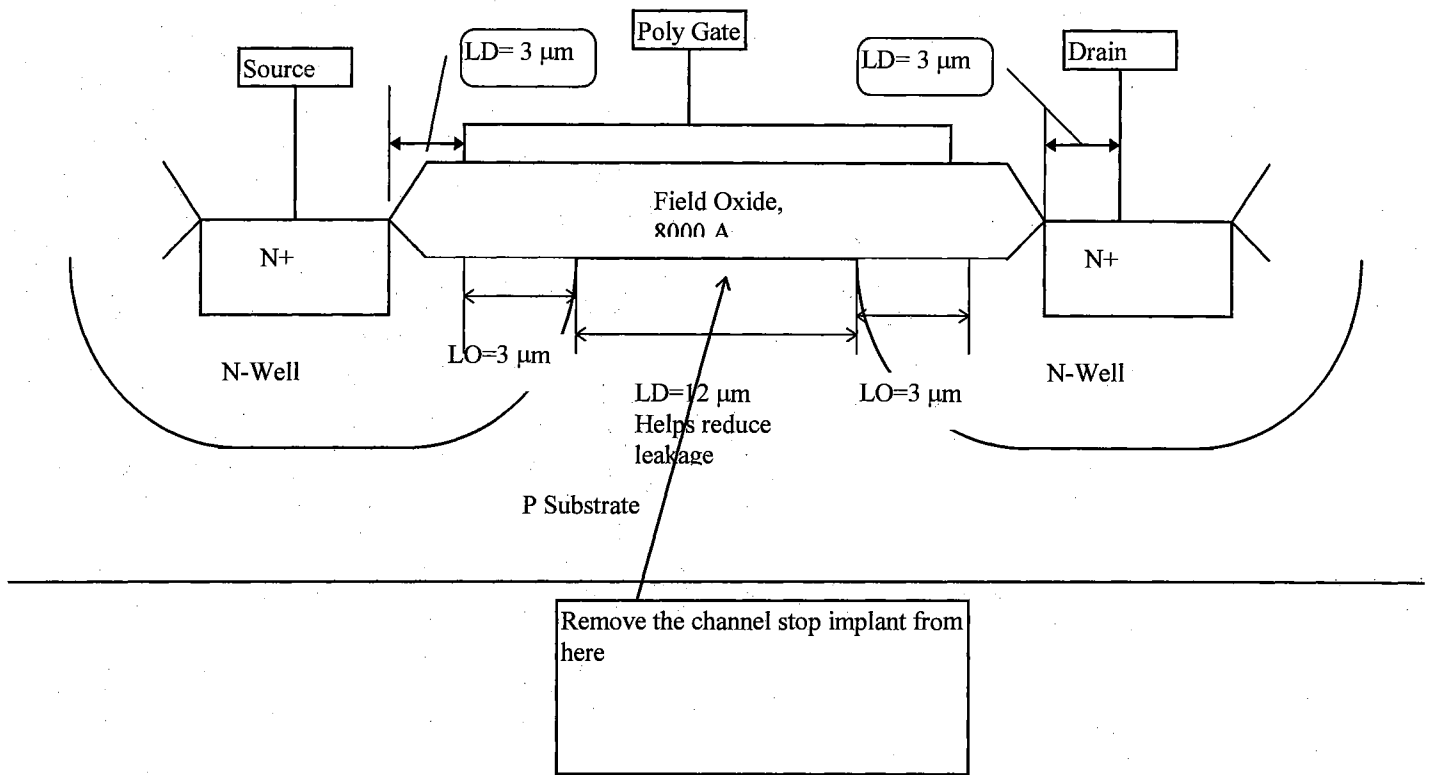


Figure V-2. A Complete High Voltage Transistor (Field Transistor)

So far all terminals can handle high voltage; however, the threshold voltage for the channel formation is still extremely high (15 to 20 Volts). The reason for this high threshold voltage is the field implant layer which increases the doping in the substrate area underneath any thick oxide layer. This is a necessary step during fabrication in order to prevent field inversion by any voltage applied over the thick oxide. As a result, an extremely high voltage is required to invert these areas. Blocking the field implant layer underneath selected areas reduces these region's doping. By selecting the intended channel area not to be heavily doped, the effective threshold drops to a much lower voltage; therefore, making it possible to form channel at much lower voltage. Figure V-3 depicts the top view layout of the proposed device.

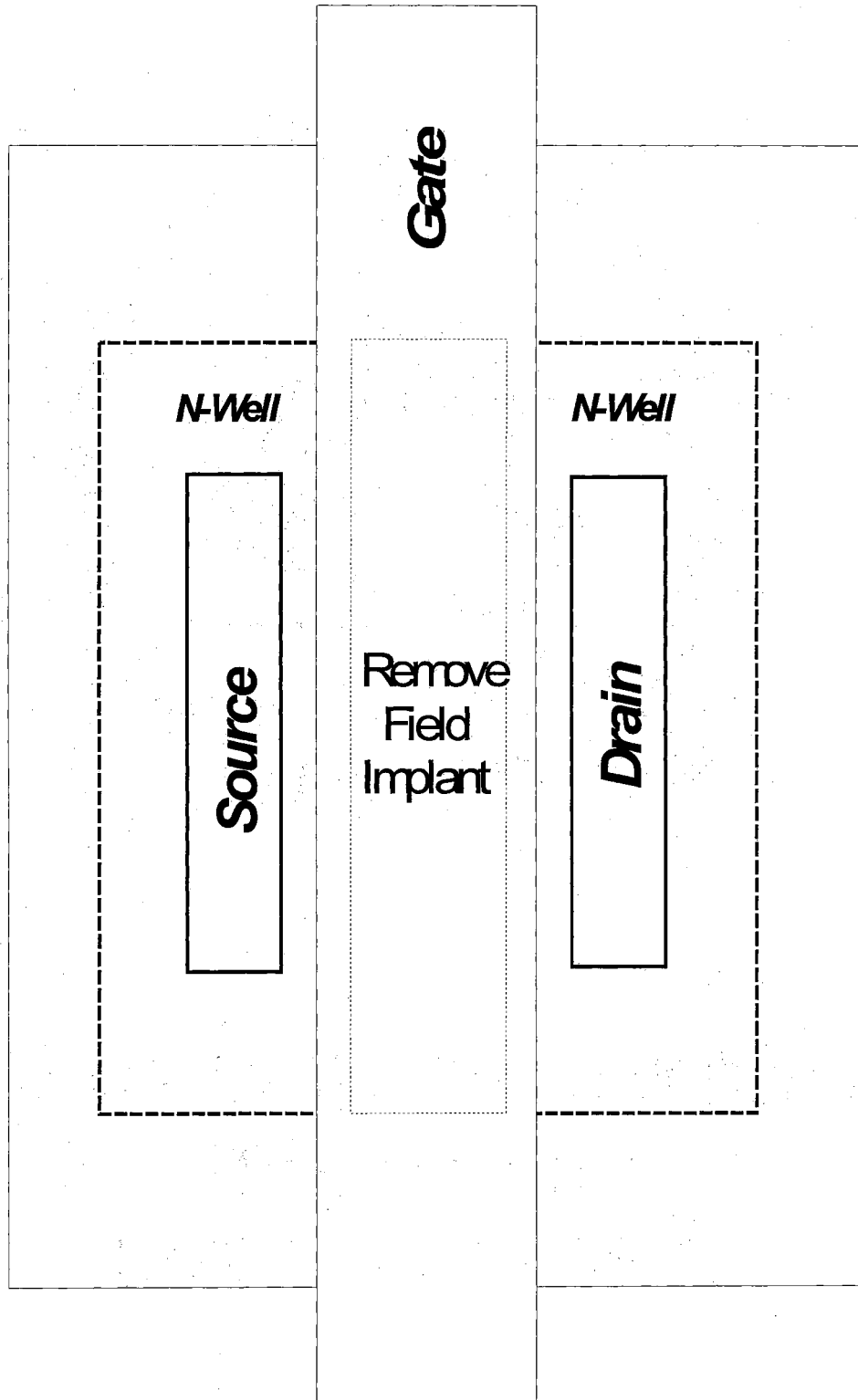


Figure V-3. Top View Layout of the complete High Voltage Transistor

Silicon Results

The proposed high voltage device was fabricated in the 2.0u ORBIT process. This device can sustain up to 100 volts on its drain or source, while handling up to 100 volts on its gate. Figure V-4 depicts current versus the drain to source (V_{ds}) voltage potential for different gate to source (V_{gs}) potential.

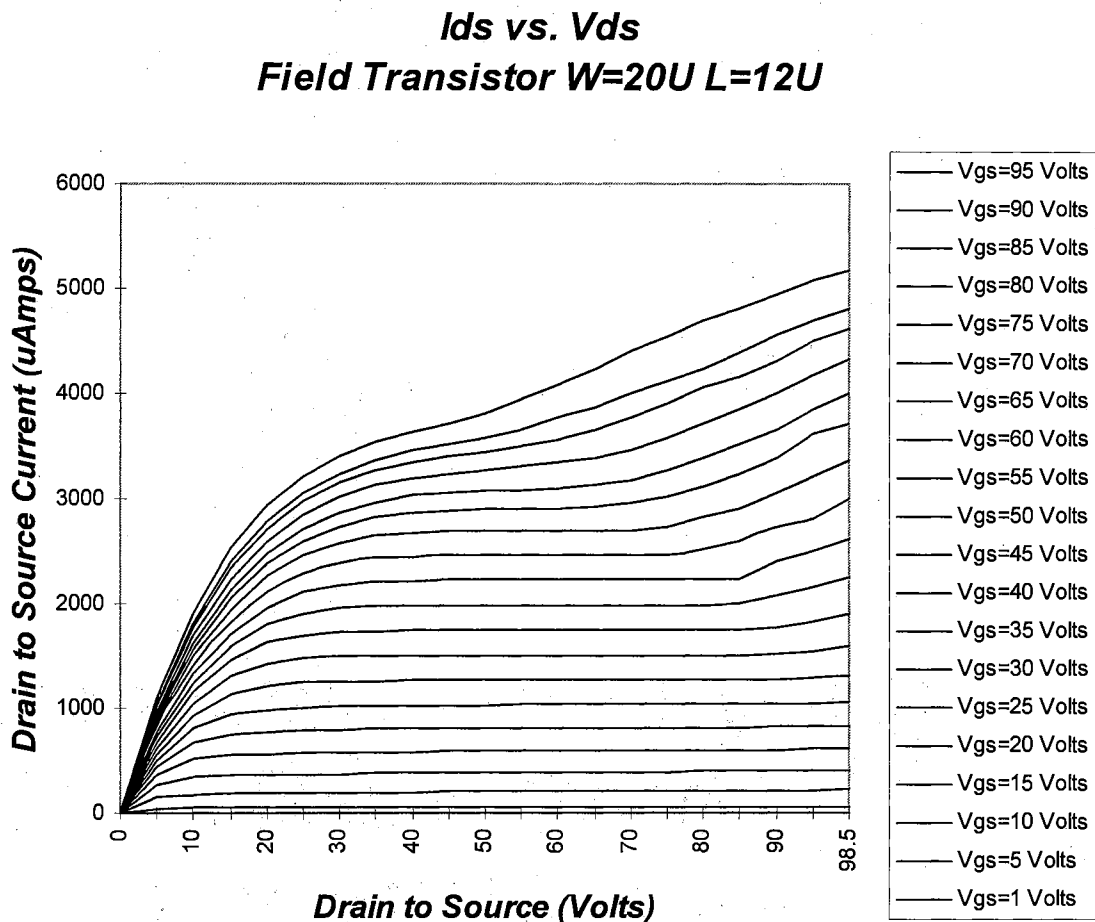


Figure V-4. The Current vs. Drain to Source Voltage Potential Curve

Field Transistor's Threshold Voltage

In a standard process, formation of the thin oxide is always followed by a threshold adjust mask to precisely control the doping underneath the thin oxide (the doping in the channel). This step is an extra channel doping step which sets the threshold to about one volt. Without this step, there is no channel doping, causing the threshold to be approximately equal to zero volt (the threshold voltage of a native transistor). In case of the extended drain device, as in the regular transistor's case, the gate is self aligned forming a thin oxide under the gate; hence, followed by the threshold adjust step. As a result, this device has a threshold equivalent to that of a regular transistor in the same process. However, as explained earlier in this section, the oxide formed underneath the gate of a field transistor is a thick oxide. It is also explained that the threshold adjust mask is applied only where thin oxide has been formed. Therefore, due to the lack of the thin oxide formation underneath the gate area of the field transistor, there is no channel doping adjustment to maintain a higher threshold for this transistor. As a result, the threshold voltage for this device is approximately equal zero volt (the threshold of a native transistor).

Figure V-5 depicts the drain to source current versus gate to source voltage potential of the field transistor with different body to source voltage potentials. It can be concluded that the threshold voltage of this device is approximately 0.1 volts for $V_{BS}=0$. Clearly, this threshold is well below that of a regular transistor; therefore, it requires careful biasing of the FET. Another way of increasing the effective threshold voltage for such device is to increase its body to source voltage potential.

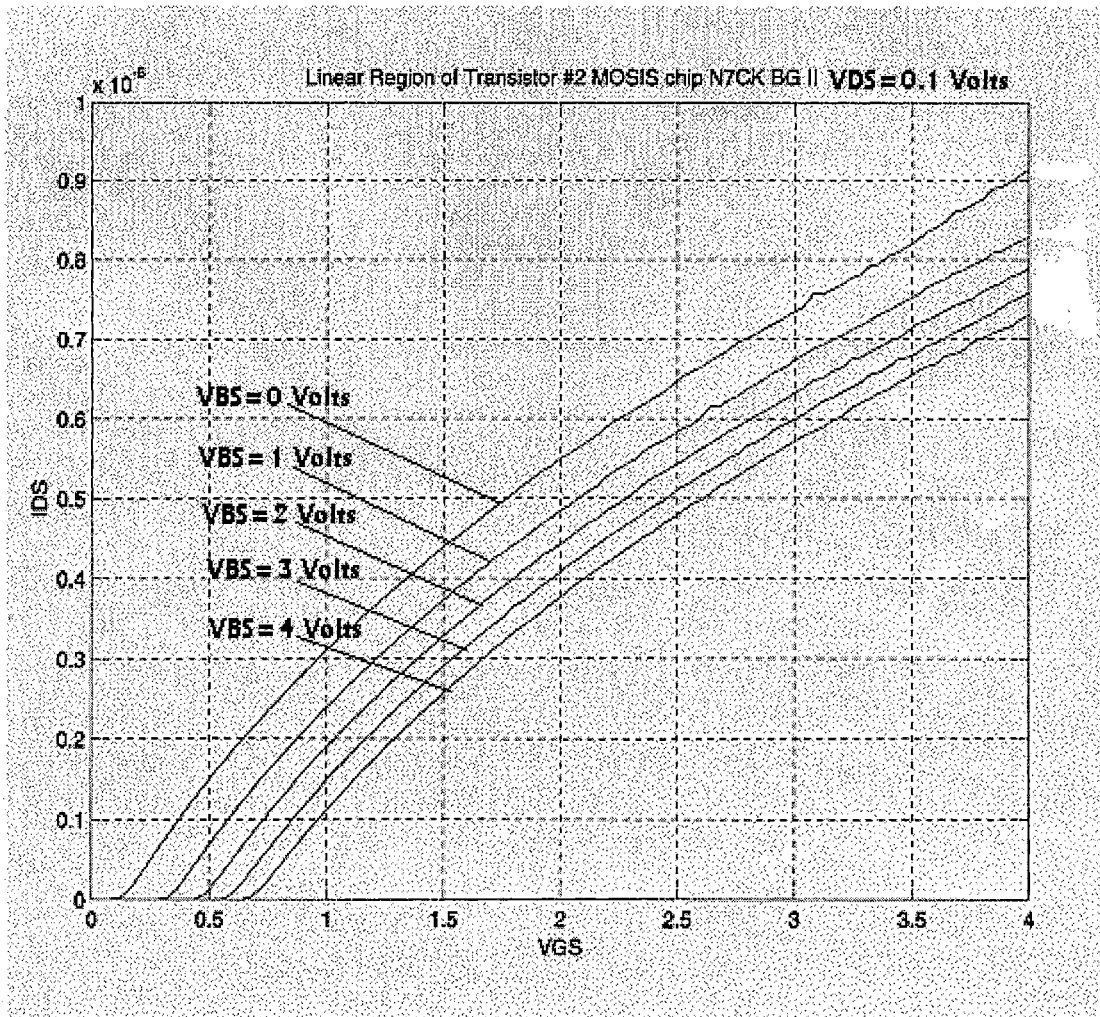


Figure V-5. Oscilloscope plot of the Drain to Source Current vs. The Gate to Source Voltage Potential

Numerous samples were tested to ensure that the field transistor is not leaky. This test was done by setting the gate to source potential to zero while increasing drain to source potential voltage to 100 Volts. The current through these field transistors were 10 to 15 pAmps. This implies that the proposed field transistors are not leaky even though the threshold is small.

Layout Precautions

The first precaution is to completely isolate this transistor from any adjacent device. That is done by creating a closed loop ground shield around the high voltage device. This ground shield helps to prevent any floating charge to get in to the substrate causing noisy environment or possible latch up. There are a number of ways to implement the ground shield. Guard ring is the best way of implementing such shield; however, this technique can increase the total area required by the transistor. To reduce the area required, a metal or poly II loop shield connected to ground can be used. Although this type of shielding is not as robust as the guard ring, it allows a significant decrease in the total area while effectively prevents any floating charge to escape the enclosed area.

Due to the low threshold in this device, extra care must be taken in order to completely control the channel behavior. Failing to do so could result in a leaky transistor. Specially, care should be taken when the gate is off (at ground potential) and the voltage potential difference between drain to source is a high voltage. In this scenario, due to the high electric field and low threshold voltage the channel could turn on if the gate does not completely block the channel. In order to make sure there are no leaky paths, the gate must completely cover the entire channel. This is done by extending the gate over the complete channel and over the shield as shown in Figure V-6. This ensures that the channel is completely off and no path is formed between drain and source due to extensive electric field across the channel.

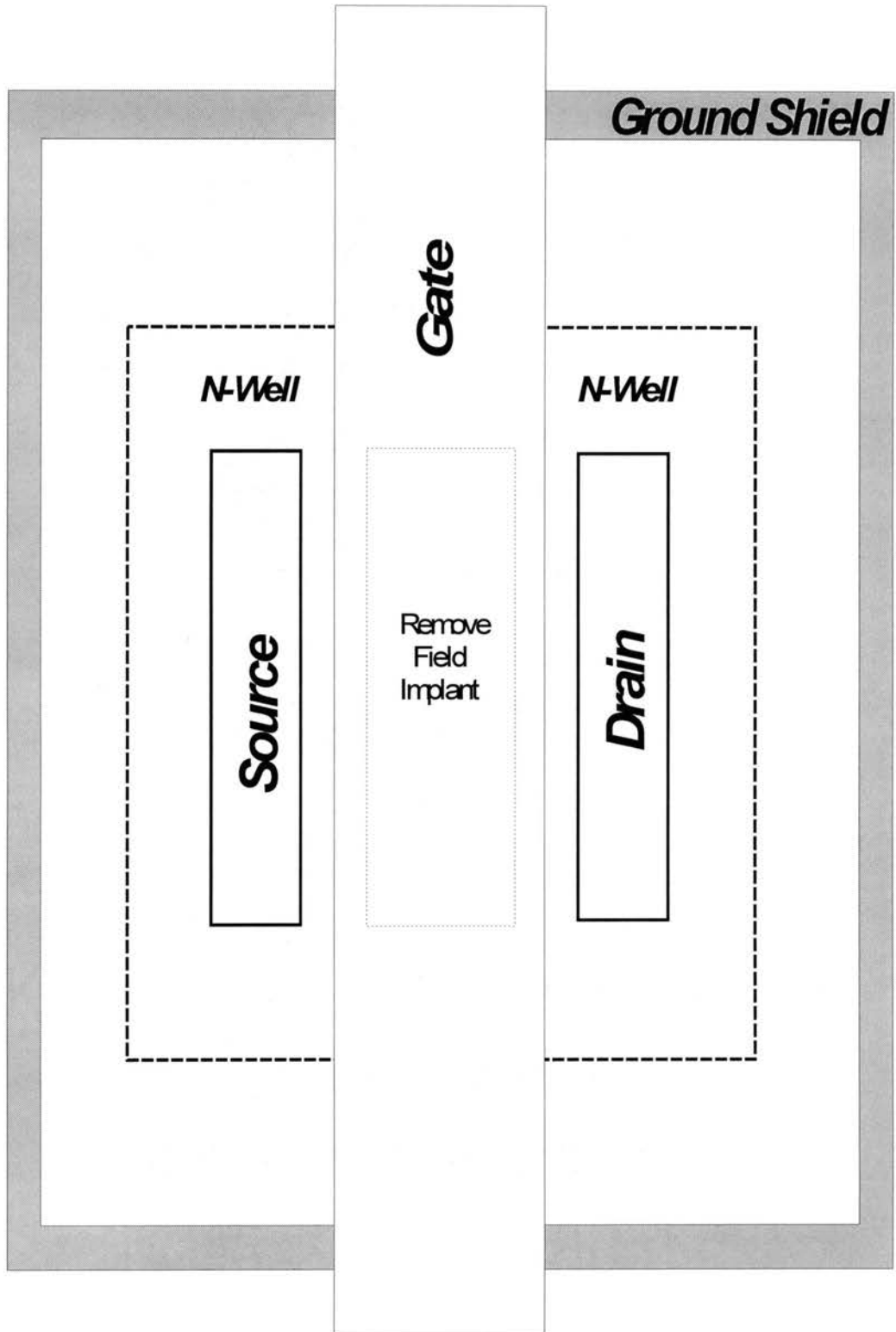


Figure V-6. Top View of the High Voltage Transistor with Ground Shield

High Voltage Driver

[Yong-Yoong, Dec., 1994] proposed a circuit capable of passing three different high voltages; however, it can not sustain voltages above the P-N junction break down voltage. To get around this problem, the circuitry in figure V-6 is proposed.

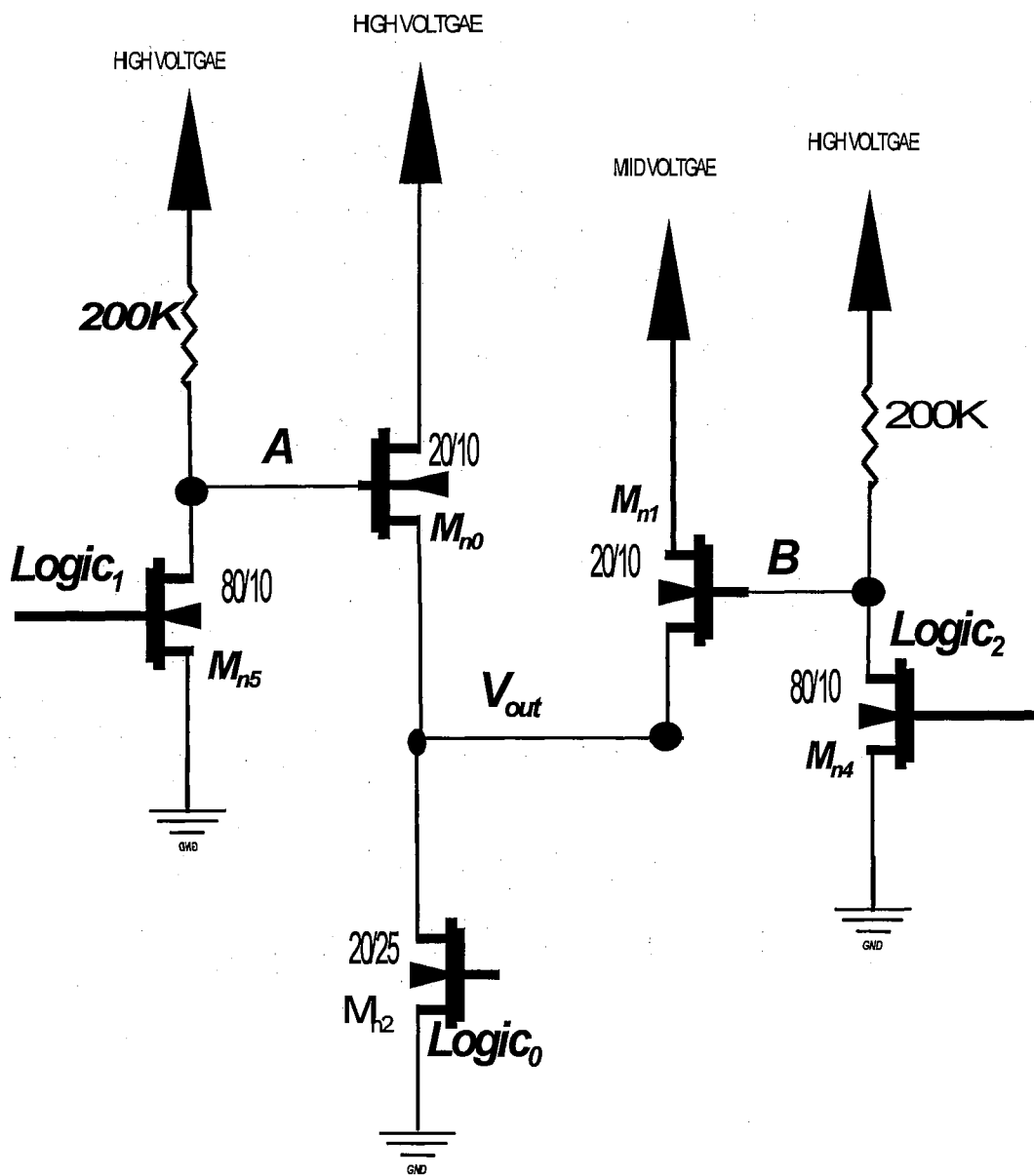


Figure V-7. A High Voltage Driver

Like [Yong-Yoong, Dec., 1994], this circuit is able to pass three different voltage levels, high, mid and zero volts. To sustain high voltage, both types of transistors discussed in this chapter are used. The pass transistors in the middle (M_{n0} and M_{n1}) must sustain high voltage both on their drains and sources; therefore, a field transistor must be used. The rest of the transistors need to handle high voltage only on their drains; hence, they are made of extended drain transistors. Similar to the high voltage transistors, the resistors also need to sustain the high voltage potential difference between any section of the resistor and the substrate. Therefore, the resistors are made of N-well material due to the high breakdown voltage between N-well and substrate .

To pass high voltage, $logic_0$ and $logic_1$ are set to ground and $logic_2$ is high. This allows node A to rise to the high voltage, allowing the node V_{out} to rise to the high voltage minus the M_{n0} 's threshold. To pass the mid-voltage $logic_0$ and $logic_2$ are forced to ground, while $logic_1$ is forced high. This drives node B to the high voltage while forcing output V_{out} to the mid-voltage. Finally, V_{out} is forced to ground if $logic_0$, $logic_1$ and $logic_2$ are set high.

This circuit was fabricated in a 2.0u ORBIT process. V_{out} was able to drive the scope probe load (15 pF) up to 80 volts. The average rise time was 5 Usec while the fall time was 1 Usec.

Although this circuit consumes DC power, it can reliably handle an extremely high voltage. The point to be careful about is to make sure transistors M_{n0} and M_{n1} are completely off when they are not suppose to be operational. This concern is risen due to the low threshold of these devices. While passing high voltage and mid-voltage, the low threshold is not an issue since V_{out} is at some high voltage, turning off M_{n1} or M_{n0}

accordingly. However, the problem arises when V_{out} is equal to zero. At this point, if any of the two field transistors leak, V_{out} tends to rise to a voltage higher than zero unless the pull down current is much larger than these leakage currents. Therefore, the pull down transistor M_{n2} is sized so that it sinks much larger current than any leakage current from the field transistors.

This driver was used to program several EEPROM cells fabricated in the 2.0u ORBIT process proposed by [Yong-Yoong, Dec., 1994]. As discussed in chapter II, the EEPROM's resolution is determined by the amount of charge transferred in and out of the floating gate. It was also discussed that the amount of charge transferred on each pulse depends on pulse amplitude and pulse duration. The shorter the pulse duration, the smaller amount of charge is added to or removed from the floating gate; hence, a higher resolution is achieved. The goal was to determine the pulse duration and amplitude which results in the minimum observable charge transfer in and out of the floating gate. To do so, the EEPROM cell was connected to the gate of an n-channel device as shown in figure V-8.

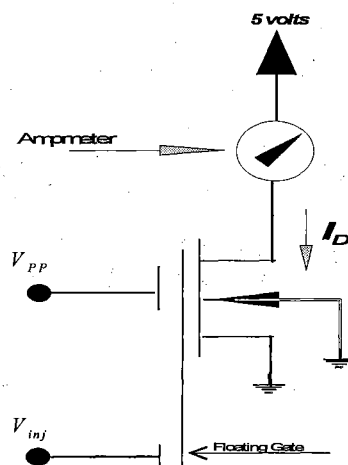


Figure V-8. An n-channel transistor incorporated with an EEPROM cell.

In order to program the EEPROM cell a high voltage pulse on either V_{pp} or the V_{inj} is applied while the rest of the terminals are grounded. After applying the high voltage pulse, all of the terminals are grounded. With these terminals at the ground potential, the effective voltage on the gate is the amount of charge trapped in the floating gate. This charge transfer can be viewed as a change in the effective threshold voltage of the n-channel transistor. Therefore, any shift in the effective threshold, will cause the drain current (I_D) through the n-channel to vary. For instance if a write operation has occurred, the effective voltage increases, causing less current to flow through the n-channel transistor. On the other hand if an erase operation has occurred, the threshold shifts in a positive direction, causing more current to flow through the n-channel transistor (chapter II). Therefore, by measuring the current flow through the n-channel transistor prior and after each high voltage programming pulse, it is possible to find the optimum high voltage programming pulse shape which results in the minimum observable change in this drain current. After programming several EEPROMs, the optimum pulse duration was found to be 200u seconds. Furthermore, for the erase operation with the pulse amplitude of $V_{inj}=11$ volts, the minimum observable current change of 0.8u amps is achieved. While for the write operation, and $V_{pp}=15$ volts, the minimum observable current change is equal to 1u amps. The significance of the minimum observable change will be discussed in the later chapters.

CHAPTER VI

THE WINNER TAKE ALL CIRCUIT

Introduction

There are numerous architectures proposed on how to implement winner take all (WTA) circuits. Some [Ugur, Mar., 1993, Yuping, Mar., 1993] suffer from the increasing size and the interconnect complexity as the number of match lines increase. This increase is usually of $O(N^2)$ complexity. An example of such architecture is the pull down inhibiting circuit discussed in chapter II. The increasing size clearly puts this architecture at a disadvantage.

To get around this problem a winner take all of $O(N)$ complexity was introduced by [Lazzaro, 1989]. However, this circuit is limited in its ability to find the best match even when perfectly matched transistors are used. A more elegant architecture followed [Johnson, Jalaeddine, Jun., 1992]; however, as used in this architecture, the WTA's performance was limited by the resolution of digital CAM cells. To utilize the fine resolution of this WTA architecture, it seems only logical to incorporate it with an analog device which distinguishes between the input and the stored pattern with a higher resolution than the standard digital CAM. The key objective of this chapter is to investigate the behavior of the WTA introduced by [Johnson, Sep., 1991] incorporated with the analog CAM described in chapter III.

First the functionality of the circuit is explained. Then a closed-form solution is found for the steady state voltage which enables the designer to predict the behavior of this circuit based on any parameter variation. Furthermore, the CAM's closed-form and the WTA circuit's closed-form will be combined to characterize the behavior of the whole system with respect to the input voltage difference. Finally, the closed-form results will be compared with the simulation results to prove the closed-form solution's validity.

Winner Take All

Figure VI-1 depicts the simplified WTA circuit in the evaluation mode. Each match line is connected to a follower circuit whose source is connected to a common node (V_f). This node is used as a global feed back which regulates all of the match lines using a feed back network. The feed back circuitry consists of a series of transistors (M_n) whose gates are controlled by the feedback node (V_f) and their drains are connected to the corresponding match lines through other pull down transistors (figure VI-1).

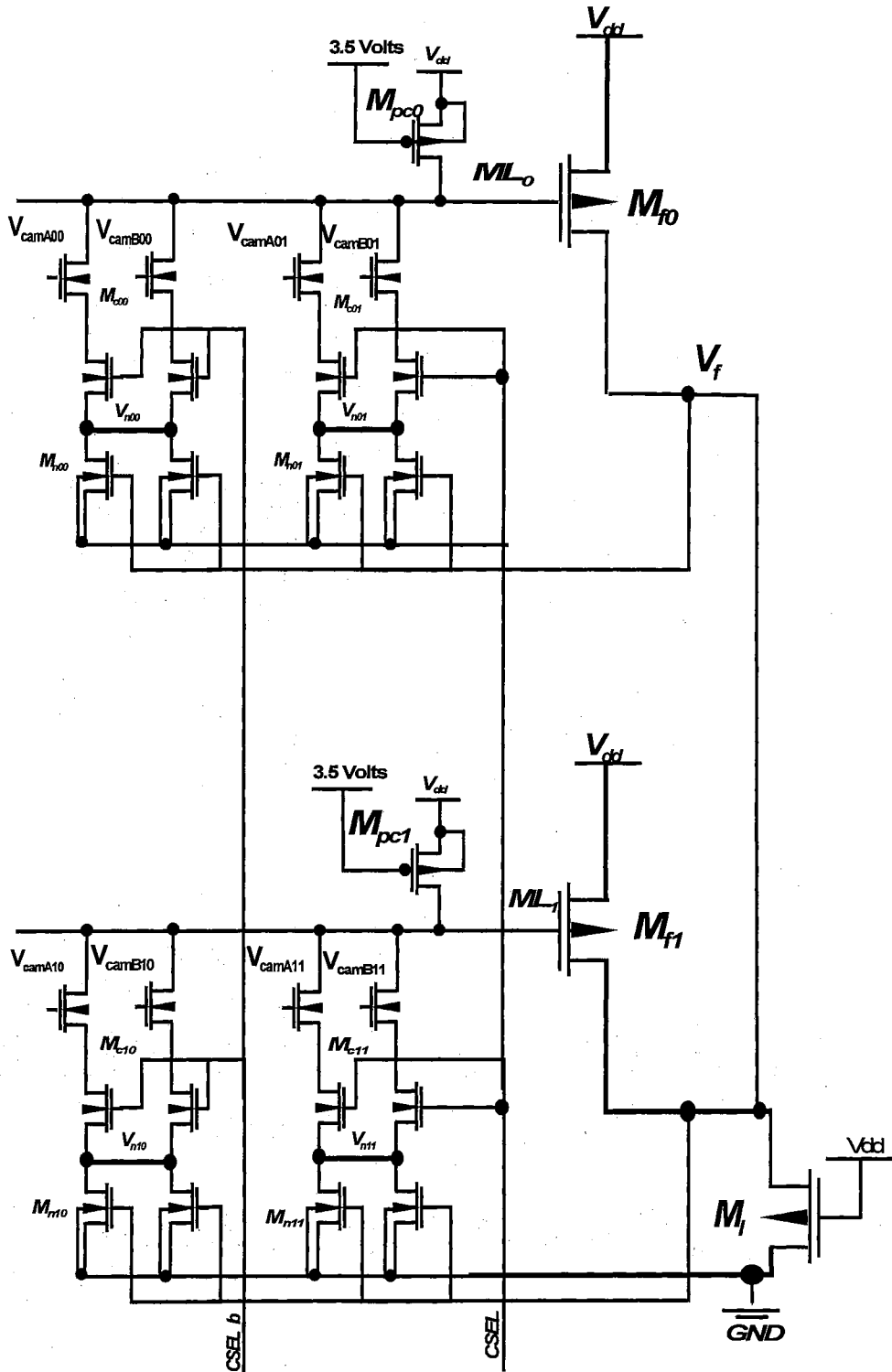


Figure VI-1. The WTA circuit in the evaluation mode

In this architecture, the match line with the least pull downs (M_c transistors) turned on, holds the highest voltage compared to the rest. This line which is connected to the gate of a follower will cause the common node's voltage (V_f) to rise. For the already losing match lines (the lines with more pull down transistors on), this feed back voltage is higher than normally required; therefore, pulling more current through the corresponding pull down transistors, causing these match lines to fall towards ground. Eventually, depending on the degree of the mismatch between the pattern and the input voltage, the match line with the highest voltage (winner) causes all of the competing voltages to reduce to a low voltage (the losers) while it converges to a higher value. This scenario happens only if there is a clear winner, that is if there is much more mismatch in the competing lines versus the winner.

To accurately predict the behavior of the circuit and its overall performance, a closed-form solution is required for the steady state output voltage. To get a closed-form solution a model different from the one used in chapter III is used. It was shown in chapter III that the crossed couple transistors dominate the final steady state value result. Furthermore, it was concluded that these transistors are in the saturation region for the entire evaluation process while only the input transistors were in the ohmic region. Therefore, it was most important that the model characterizes the saturation region better than the ohmic region. However, in deriving the a closed-form solution for the WTA circuit, it will be shown that the region of operation for the feed back circuit has a great influence in determining the final steady state of the match line. Therefore, it is essential to predict the behavior of this circuit in each of the operating regions; hence, one can not get away with the crossly simplified model used for the linear region in chapter III.

Unlike the model used in chapter III, this model includes the effect of the gate to source potential on the drain current of a transistor in the ohmic region. While in the saturation region, it is assumed that the transistor has effectively reached velocity saturation limit where the drain current is no longer dependent on the drain to source voltage. Furthermore, while in the ohmic region, the drain to source voltage is small enough so that the $\frac{V_{ds}^2}{2}$ term in the ohmic region equation can be neglected. These assumptions are necessary to be able to get a closed-form solution since the conventional nonlinear equations result in complex equations that are either too complex or are only solvable through numerical method (HSPICE). The goal is to derive a simple closed-form solution that will predict the behavior of the circuit as close as one can achieve without using any numerical methods.

The ohmic and velocity saturation limit model used are as follow (Johnson, ECEN 5263, Jan. 1997):

$$I_D = \beta \cdot (V_{gs} - V_T) \cdot V_{ds} - \frac{V_{ds}^2}{2} \quad V_{ds} < V_{Dsat} \quad \text{Equation VI-1}$$

$$I_D = \beta \cdot V_c \cdot (V_{gs} - V_T) \quad V_{ds} \geq V_{Dsat} \quad \text{Equation VI-2}$$

where

$$K = \mu \cdot C_{ox}$$

$$\beta = K \cdot \frac{w}{l}$$

$$V_{Dsat} = V_c$$

C_{ox} is the oxide capacitance per unit area

V_c is the carrier saturation voltage and is equal to $v_{max} \cdot \frac{l}{\mu}$

v_{max} is the maximum carrier velocity

μ is the low field inversion layer mobility and assumed to be constant

V_{Dsat} is the drain to source voltage at which the velocity saturation limit is achieved

WTA's Closed-Form Solution

To find the closed-form solution, it is assumed that there is a single “winner” match line which is the closest match and all of the competing “loser” match lines are pulled low enough so that there is no contribution from these lines to the follower voltage. Furthermore, the stored pattern in the winner matches the input pattern except for a single pixel (one CAM). As shown in the figures III-5 through III-10, the steady state output of the CAM can vary between one to five volts. In case this voltage is much larger than the threshold of M_c (figure VI-1), the transistor M_n is entirely in saturation region. However, if this voltage is approximately equal to the M_c 's threshold voltage, M_n is in the ohmic region. Therefore, the feed back transistor M_n can potentially operate in two different regions depending on the output of the CAM. The p-channel transistor (M_{pc}) which acts as a current source during the regular operation, is assumed in the ohmic region when the CAM's output voltage is approximately the same as the threshold voltage of the M_c transistor. This is an appropriate assumption since the pull down

current is much less than the pull up current provided from the p-channel current source. This forces the match line voltage to approximately 5 volts; therefore, forcing M_{pc} in the ohmic region. The follower transistor M_f and transistor M_c are always in the saturation region since their gate voltage never exceeds their drain voltage by a threshold. The transistor M_l is always in the ohmic region by design.

To start, it is assumed that the CAM's output is approximately equal to the M_c transistor's threshold; therefore, M_p , M_n and M_l are in the ohmic region while the rest of the transistors are in the saturation region. Using this as a starting point, the closed-form solution will predict the change of operating region as the drain to source voltage of each transistor passes the boundary regions (V_{Dsat}).

Summing all of the currents in and out of the winner match line (figure VI-1) and the common node

$$C_{nw} \cdot \frac{dV_{nw}}{dt} = \beta_p \cdot (V_{dd} - V_{bias} - V_{Tp}) \cdot (V_{dd} - V_{nw}) - \sum_{i=1}^n \beta_n \cdot V_{cn} \cdot (V_{cam_i} - V_{nw_i} - V_{Tf})$$

Equation VI-3

$$C_{nw} \cdot \frac{dV_{nw}}{dt} = \beta_n \cdot V_{cn} \cdot (V_{cam} - V_{nw} - V_{Tf}) - 2 \cdot \beta_n \cdot (V_f - V_{Tn}) \cdot V_{nw}$$

Equation VI-4

$$C_f \cdot \frac{dV_f}{dt} = \beta_f \cdot V_{cf} \cdot (V_{mw} - V_f - V_{Tfo}) - \beta_l \cdot (V_{dd} - V_{Tn}) \cdot V_f$$

Equation VI-5

Where

V_{cam_i}	output of the i^{th} CAM cell at the steady state (V_{Af} or V_{Bf})
n	is the number of mismatch in the winner match line
V_{mw}	is the voltage of the winner match line
V_{nw}	is the drain voltage of the feed back transistor in the winner match line
V_f	is the global feed back voltage
C_{mw}	is the total capacitance on the winner match line
C_{nw}	is the total capacitance on the drain of the feed back transistors in the winner match line
C_f	is the total capacitance on the feed back node
V_{bias}	is the bias voltage on the gate of the p-channel transistors
β_p	is the β of the M_{pc} transistor
β_n	is the β of the M_c transistor
β_l	is the β of the M_l transistor
β_f	is the β of the M_f transistor
V_{Tn}	is the threshold voltage of the transistors M_n and M_l (transistors with no body effect)
V_{Tfo}	is the threshold voltage of the transistor M_c (transistor with body effect)
V_{Tf}	is the threshold voltage of the follower transistor M_f (transistor with body effect)
V_{cn}	is the carrier saturation voltage for all of the pull down

n-channel devices

V_{cf} is the carrier saturation voltage for follower

n-channel devices

Let

$$G_a = \beta_p \cdot (V_{dd} - V_{bias} - V_{Tp})$$

$$G_b = \beta_l \cdot (V_{dd} - V_{Tn})$$

Once the steady state is reached, the AC current no longer exists resulting the following

$$0 = G_a \cdot (V_{dd} - V_{mw}) - \sum_{i=1}^n \beta_n \cdot V_{cn} \cdot (V_{cam_i} - V_{nw_i} - V_{Tf}) \quad \text{Equation VI-6}$$

$$0 = \beta_n \cdot V_{cn} \cdot (V_{cam} - V_{nw} - V_{Tf}) - 2 \cdot \beta_n \cdot (V_f - V_{Tn}) \cdot V_{nw} \quad \text{Equation VI-7}$$

$$0 = \beta_f \cdot V_{cf} \cdot (V_{mw} - V_f - V_{Tfo}) - G_b \cdot V_f \quad \text{Equation VI-8}$$

Manipulating the above equations and solving for V_f , V_{nw} and V_{mw} in terms of the CAM's output voltage (V_{cam}) results (Appendix B).

$$V_f = \frac{1}{2 \cdot A} \cdot \left[-B - \sqrt{B^2 - 4 \cdot A \cdot C} \right]$$

Equation VI-9

Where

$$A = -2 \cdot G_a \cdot (\beta_f \cdot V_{cf} + G_b)$$

$$B = \left[-2 \cdot \beta_n \cdot V_{cn} \cdot V_{cam} + (2 \cdot \beta_n \cdot V_{Tf} - G_a) \cdot V_{cn} + 2 \cdot G_a \cdot (V_{dd} - V_{Tfo} + V_{Tn}) \right] \cdot V_{cf} \cdot \beta_f \\ + (2 \cdot V_{Tn} - V_{cn}) \cdot G_a \cdot G_b$$

$$C = 2 \cdot V_{cf} \cdot \beta_f \cdot \beta_n \cdot V_{Tn} \cdot (V_{cam} - V_{Tf}) \cdot V_{cn} + (V_{cf} \cdot \beta_f \cdot G_a) \cdot (V_{cn} - 2 \cdot V_{Tn})$$

$$V_{mw} = \left(1 + \frac{G_b}{\beta_f \cdot V_{cf}} \right) \cdot V_f + V_{Tfo}$$

Equation VI-10

$$V_{nw} = \frac{(G_a \cdot (V_{mw} - V_{dd}))}{2 \cdot \beta_n \cdot (-V_f + V_{Tn})}$$

Equation VI-11

Equation VI-9 suggests as the V_{cam} increases (in parameters B), the feed back voltage V_f drops. At this point, M_n is in the triode region and looks like a variable resistor whose value is controlled via its gate to source voltage (V_f). As V_f drops M_n 's impedance gets larger; therefore, less current is being pulled from the match line. This causes the drain voltage of this transistor to rise.

To better illustrate how well the WTA circuit works, it is essential to find the loser voltage. As mentioned earlier in this section, it is assumed that the loser voltage does not contribute to the feedback voltage (V_f) since the follower transistor connected to the loser match line assumed to be in the cut off region. Therefore, Equation VI-9 remains

the same while deriving an expression for the loser voltage. In the case of the loser voltage, the p-channel current source is in the saturation region since it is assumed that the loser match line is already at a voltage below the threshold voltage of the follower transistor M_{fl} . Furthermore, the input transistor to the WTA circuit, M_c , is also assumed to be in the ohmic region. The last transistor to consider are the feed back transistors which can operate in both operating regions. Like the derivation for the winner match line, it is further assumed that only one of the pull downs in the loser match line is on. This corresponds to a single CAM mismatch in the overall loser pattern. It is also assumed that this mismatch (V_{caml}) is much larger than the mismatch in the winner match line. With these assumptions, the currents through the loser match line are

$$0 = G_a - \sum_{i=1}^{n_l} \beta_n \cdot (V_{camli} - V_{nli} - V_{Tf}) \cdot (V_{mli} - V_{nli}) \quad \text{Equation VI-12}$$

$$0 = \beta_n \cdot (V_{caml} - V_{nl} - V_{Tf}) \cdot (V_{ml} - V_{nl}) - 2 \cdot \beta_n \cdot (V_f - V_{Tn}) \cdot V_{nl} \quad \text{Equation VI-13}$$

Where

- n_l is the number of the mismatch in the loser match line
- V_{nli} is the drain voltage on the i^{th} feed back transistor in the loser match line
- V_{mli} is the voltage of the i^{th} loser match line
- V_{camli} is the input to the i^{th} pull down in the loser match line

Manipulating above equations, the loser voltage is as follow

$$V_{nl} = \frac{G_a}{2 \cdot \beta_n \cdot (V_f - V_{Tn})} \quad \text{Equation VI-14}$$

$$V_{ml} = V_{nl} + \frac{G_a}{(V_{caml} - V_{nl} - V_{Tf}) \cdot \beta_n} \quad \text{Equation VI-15}$$

Equations VI-14 and VI-15 implies that the loser match line voltage is predominately controlled by the global feed back voltage (V_f) which is what is expected.

When the drain to source voltage of the feed back pull down transistor (M_n) crosses its V_{Dsat} voltage, it changes its operating region from the ohmic to the saturation region. The input voltage at which this happens can be quantified by solving for the input voltage to the WTA circuit (V_{caml}) when the drain to source potential on the feed back voltage equals the V_{Dsatn} . This voltage is equal (Appendix B)

$$V_{caml} = \frac{-V_{fc}^2 \cdot A - d - V_{fc} \cdot b}{c + V_{fc} \cdot a} \quad \text{Equation VI-16}$$

$$a = -2 \cdot V_{cn} \cdot V_{cf} \cdot \beta_n \cdot \beta_f$$

$$b = (2 \cdot V_{Tn} - V_{cn}) \cdot G_a \cdot G_b + (V_{dd} - V_{Tfo} + V_{Tn}) \cdot 2 \cdot \beta_f \cdot V_{cf} \cdot G_a \\ + 2 \cdot \beta_f \cdot V_{cf} \cdot V_{cn} \cdot (G_a + \beta_f \cdot V_{Tf})$$

$$c = -a \cdot V_{Tn}$$

$$d = (2 \cdot V_{Tn} - V_{cn}) \cdot \beta_f \cdot G_a \cdot V_{cf} \cdot V_{Tfo} \\ + ((-2V_{Tn} + V_{cn}) \cdot V_{dd} \cdot \beta_f \cdot G_a - 2 \cdot V_{cn} \cdot \beta_n \cdot V_{Tf} \cdot V_{Tn} \cdot \beta_f) \cdot V_{cn}$$

Where

V_{camc} is the CAM's steady state voltage which the feed back transistor

leaves the ohmic region and enters the saturation region

V_{fc} is the feed back voltage exerted on the gate to source of the feed back voltage at the point where it leaves the ohmic region and enters the saturation region

Below is the set of equations when the feed back transistor changes its operating region.

$$0 = G_a \cdot (V_{dd} - V_{mw}) - \sum_{i=1}^n \beta_n \cdot V_{cn} \cdot (V_{cam_i} - V_{nw_i} - V_{Tf}) \quad \text{Equation VI-17}$$

$$0 = \beta_n \cdot V_{cn} \cdot (V_{cam} - V_{nw} - V_{Tf}) - 2 \cdot \beta_n \cdot V_{cn} \cdot (V_f - V_{Tn}) \quad \text{Equation VI-18}$$

$$0 = \beta_f \cdot V_{cf} \cdot (V_{mw} - V_f - V_{Tfo}) - G_b \cdot V_f \quad \text{Equation VI-19}$$

Equations VI-17 indicates that M_n is operating in the saturation region while the rest of the transistors remain in their previously assumed regions. Solving for the V_f , V_{mw} , and V_{mw} in terms of V_{cam} yield the following

$$V_{fs} = \frac{\left[(V_{dd} - V_{Tfo}) \cdot G_a + 2 \cdot \beta_n \cdot V_{cn} \cdot V_{Tn} \right] \cdot \beta_f \cdot V_{cf}}{\left[(\beta_f \cdot V_{cf} + G_b) \cdot G_a + 2 \cdot \beta_n \cdot V_{cn} \cdot \beta_f \cdot V_{cf} \right]}$$

Equation VI-20

$$V_{mw} = \left[1 + \frac{G_b}{\beta_f \cdot V_{cf}} \right] \cdot V_{fs} + V_{Tf0}$$

Equation VI-21

$$V_{nw} = V_{cam} - V_{Tf} - 2 \cdot V_f + V_{Tn}$$

Equation VI-22

Equations VI-20 and VI-21 indicate, once the M_n transistor is in the saturation region, the outcome of the WTA (V_{mw}) is no longer dependent on the output of the CAM cell (V_{cam}). At this point, the pull down current is set by the gate to source voltage potential (V_f) of the feed back transistor (M_n).

Equation VI-16 suggests when V_{cam} exceeds V_{camc} , the feed back transistor leaves the ohmic region and enters the saturation region. Therefore, as long as V_{cam} is less than or equal V_{camc} , equations VI-9 through VI-11 are valid, as V_{cam} passes V_{camc} , equations VI-19 through VI-21 become valid.

Now it is time to check the validity of the closed-form solution by comparing the theoretical against the simulation results (HSPICE). This will indicate how closely the closed-form solution predicts the actual behavior of the system.

As indicated while deriving the closed-form solution, it is assumed that there is only a single mismatch between the input and the stored pattern. Like the closed-form, HSPICE suggests that the circuit starts with the M_n transistor in the ohmic region. As V_{cam} increases, more current sinks through this path as long as M_n is in the ohmic region. This causes the match line to fall and the feed back voltage to follow, causing the V_{gs} of the M_n transistor to drop while its drain to source potential increases.

Once M_n starts operating in the saturation region, the M_c transistor is no longer setting the pull down current directly. This current is now set by the gate to source of M_n (V_f) and is independent of V_{cam} ; therefore, the match line does not vary with V_{cam} any more. This concept is shown in figure VI-2 through VI-3; however, the point where this voltage becomes flat (when M_n is in saturation) is different from the closed-form solution. This is expected since the closed-form solution ignores the pinch off region and models transistors with drain to source voltage greater than V_{Dsat} in the velocity saturation region. Despite this assumption the closed-form solution correlates with HSPICE result.

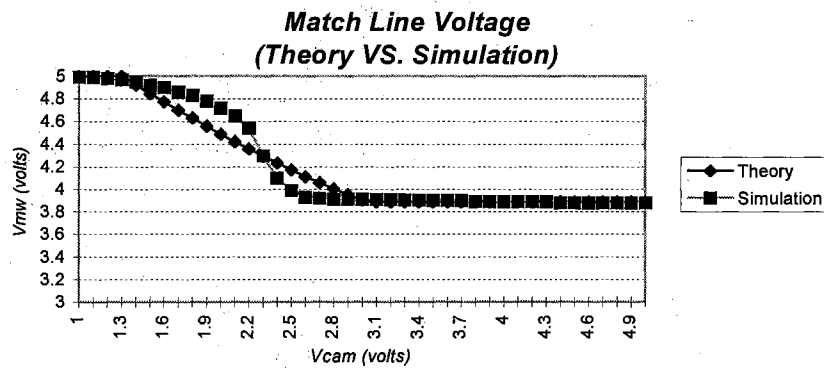


Figure VI-2. Theory vs. Simulation for the winner match line voltage

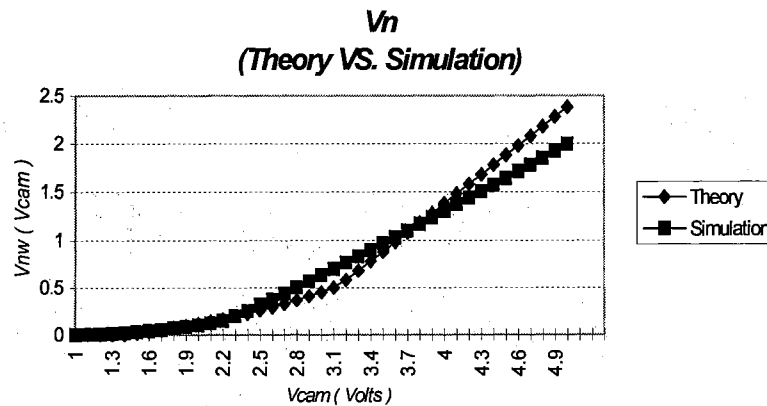


Figure VI-3. Theory vs. Simulation for the Drain to source voltage of the Feed Back Transistor

Figures VI-2 and VI-3 clearly show the nonlinear characteristics of the results found by the simulator versus the closed-form solution. Besides the obvious difference between the model used in this chapter and the conventional model [Meta-Software], issues such as dependence of the mobility on the gate to source potential, accurate estimation of the threshold and its body effect are some of the major contributors to these nonlinearities. In deriving the closed-form solution, these effects were ignored for simplicity; otherwise, obtaining a close-form solution would be extremely difficult if not impossible.

To further characterize the closed-form solution, the loser match line voltage found using the closed-form solution is compared against the results found from the simulation. As shown in figure VI-4, the closed-form solution also agrees with the simulation results.

**Loser Match Line Voltage
(Theory VS. Simulation)**

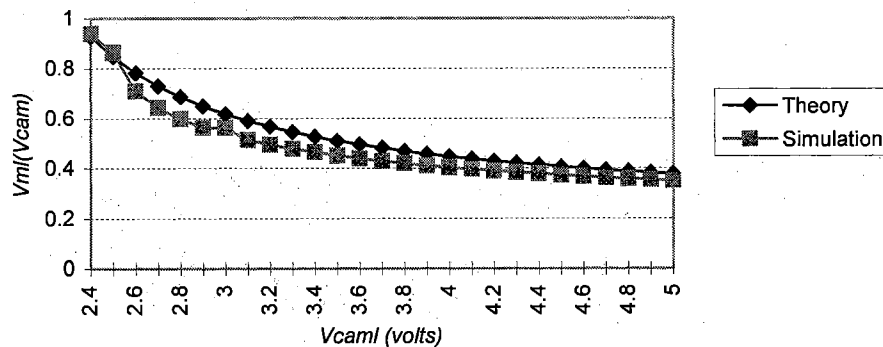


Figure VI-4. Theory vs. Simulation for the loser match line voltage

So far all of the pieces of the recognition engine has been introduced. Furthermore, it was shown how different patterns are stored on each CAM cell while the recognition engine

is in the program mode. After storing the desired patterns on each CAM cell, the engine is ready to evaluate any input test pattern against the stored patterns. Next section describes this evaluation process

Evaluation Mode

Figure VI-5 depicts the complete recognition engine while in the evaluation mode. The evaluation mode is performed once all of the CAM cells have been programmed with the desired patterns. This process compares all of the stored patterns against the input pattern in a parallel fashion. At the end of a this cycle, the match line with the highest score eliminates the competing match lines (given there is enough mismatch in loser patterns).

Prior to the evaluation cycle, the *start_prog* signal is set low which puts the WTA circuit in the evaluation mode. Meanwhile, all of the columns are selected enabling all of the pull down transistors (M_n). At this point the WTA circuit acts as a discriminator. Next, the test pattern is applied to the inputs, then the precharge signal is applied globally to the chip precharging each of the CAM cell's output nodes (V_A and V_B) to V_{dd} . The evaluation process commences, once the precharge clock is set high. During this evaluation cycle, any mismatch between the pattern and the input value, causes the corresponding pull down transistor to be on. For instance if there is a mismatch between the pattern in the CAM cell in the first row and the first column (CAM_{00}), the corresponding pull down transistor M_n will be on. Moreover, the strength of each pull down is determined by the degree of the mismatch between the stored and the test pattern

since the gate voltage of each pull down is controlled by the output of the corresponding CAM cell (V_A and V_B). In the above example, if the input and the stored pattern in CAM_{00} are mismatched by 0.5 volts, the output of this CAM cell (V_A or V_B) settles to a higher voltage than if the inputs mismatched by 0.1 volts (V_{Af} and V_{Bf} in figure III-5 through 10). Therefore, the latter case exerts a smaller gate to source voltage (V_{gs}) on the corresponding pull down transistor (M_{n00}). Hence, the closer the pattern is to the input value, the lower the output voltage of the CAM is; as a result, the weaker the corresponding pull down transistor is turned on.

Furthermore, the more number of mismatch there is between the test and the store patterns in the entire match line (more mismatch CAMs), the more pull down transistors are on, causing more current to sink from the corresponding match line. For instance, if the pull down current through the match line exceeds the current being provided by the pull up transistor M_{pc} , this match line falls towards ground. On the other hand, if the match line has less or no mismatch, it will settle to some voltage close to V_{dd} , since there is little or no pull down current compared to the current provided by the pull up p-channel transistor. This voltage dominates the feedback voltage and eventually pulls down all of the competing match lines to a lower voltage.

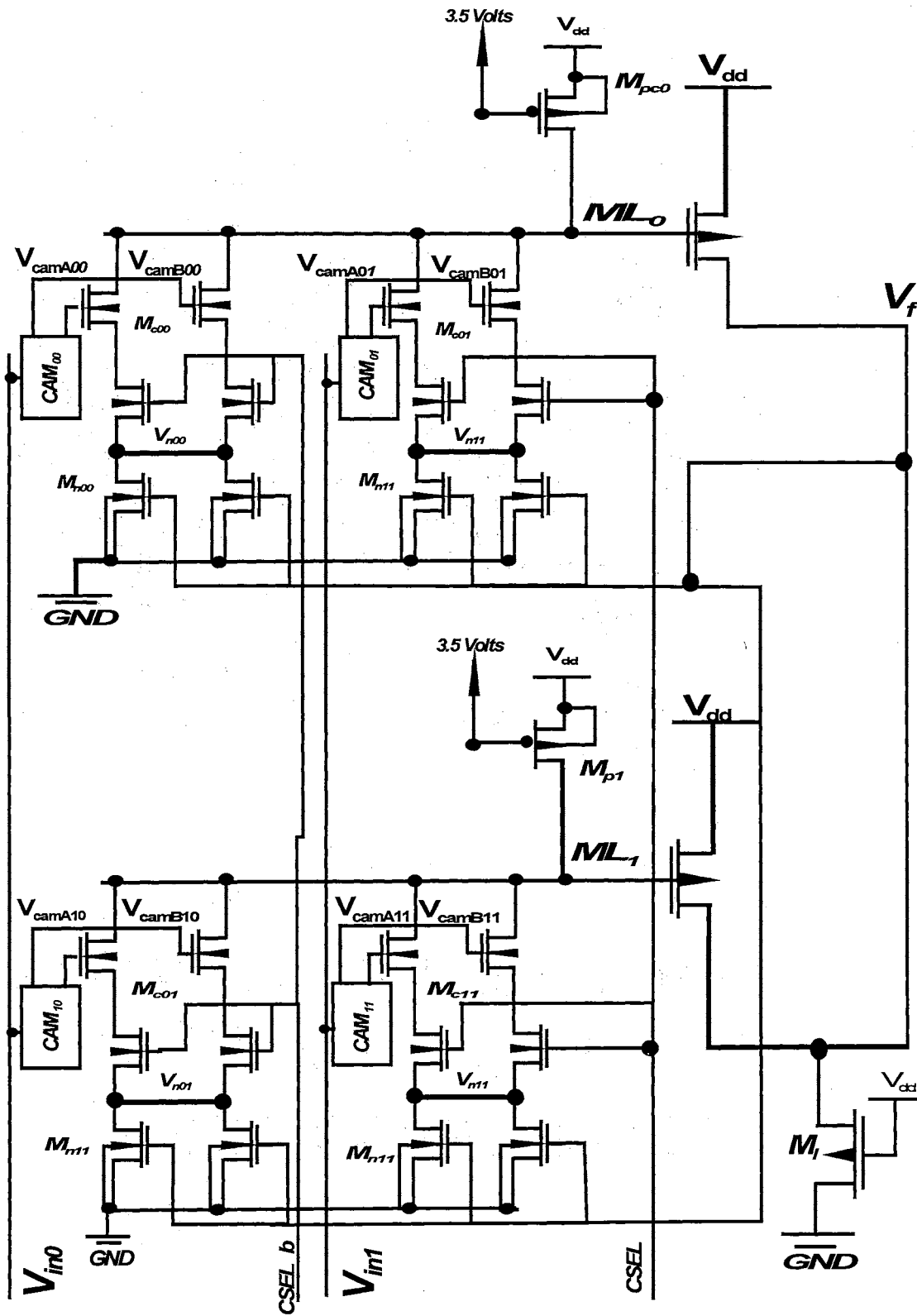


Figure VI-5. The Analog Recognition Engine in the Evaluation mode

Closed-Form Versus Simulation

So far a closed-form solution is found for both the CAM and the winner take all circuit separately. By putting these results together, the overall closed-form solution of the complete architecture while the evaluation mode is found. This is possible since the final result of interest is the steady state of the whole system; therefore, combining the steady state value of each individual stage (equations VI-9 through VI-22 and equations III-26 and III-27), results in the characterization of the input versus output voltage of the overall system. The goal is to get a tabulated result relating the stored and the input pattern, to the match line voltage. As shown in figure VI-5, the output of each CAM cell is connected to the gate of the pull down transistors M_c . As already indicated in the chapter III, depending on the inputs, one of the CAM's output voltage decays to ground leaving the second path at a higher voltage. This causes only one of the pull down nodes to be high. For instance if the V_{ref} equals 1.5 and V_{in} equals 1.3 volts, according to the closed-form solution found in chapter III, V_A decays to ground while V_B stays at 3.5 volts. These voltages are now considered the input voltage to the WTA circuit. Based on the closed-form solution found in this chapter, the corresponding match line voltage is 4.2 volts.

To compare the closed-form solutions results against the simulation results, the 2x2 array shown in figure VI-5 was considered. The pattern in the first match line was set equal to the input pattern. Then the input to the first CAM (CAM_{00}) cell was varied +/- 0.5 volts from the stored voltage while the inputs to the second CAM (CAM_{01}) were

perfectly matched. On the other hand, the reference pattern on CAM_{10} was purposely mismatched to its input voltage by at least 2 volts while the inputs to CAM_{11} were perfectly matched. The goal was to see how the engine picks the winner, and how close are the results found from simulation and closed-form solution.

Figure VI-6 through VI-11 tabulate the simulation results versus the closed-form's results during the evaluation process. These figures indicate that the results follow the same pattern and are close despite the difference in the model used in chapter III, this chapter and the model used by the simulator [Meta-Software]. In each figure, V_{ref} is set to the pattern voltage on the CAM_{00} , while the input pattern voltage deviates ± 0.5 volts from this pattern voltage.

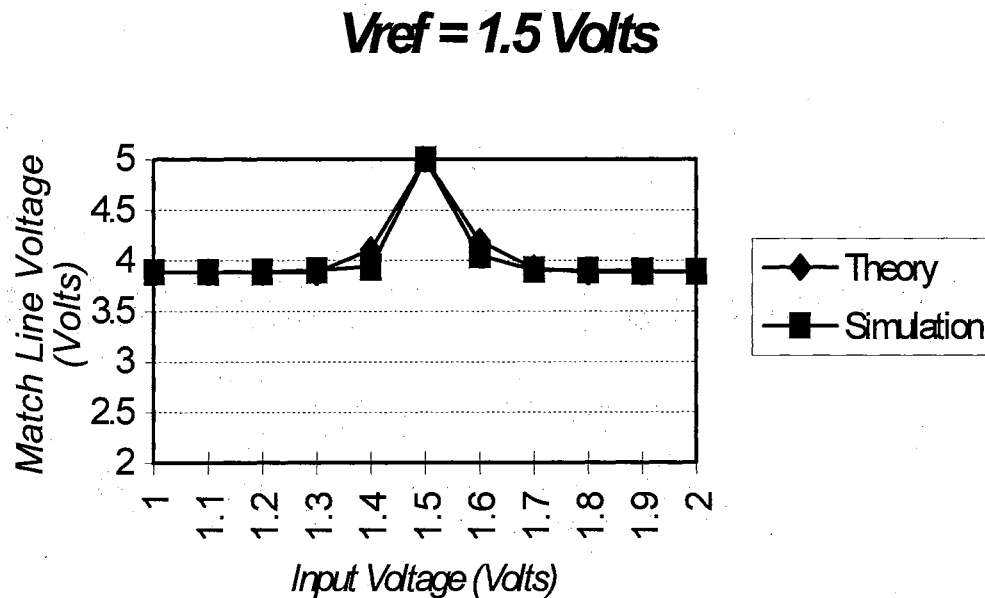


Figure VI-6. Input Voltage vs. Match Line Voltage ($V_{ref}=1.5$ Volts)

$V_{ref} = 2$ Volts

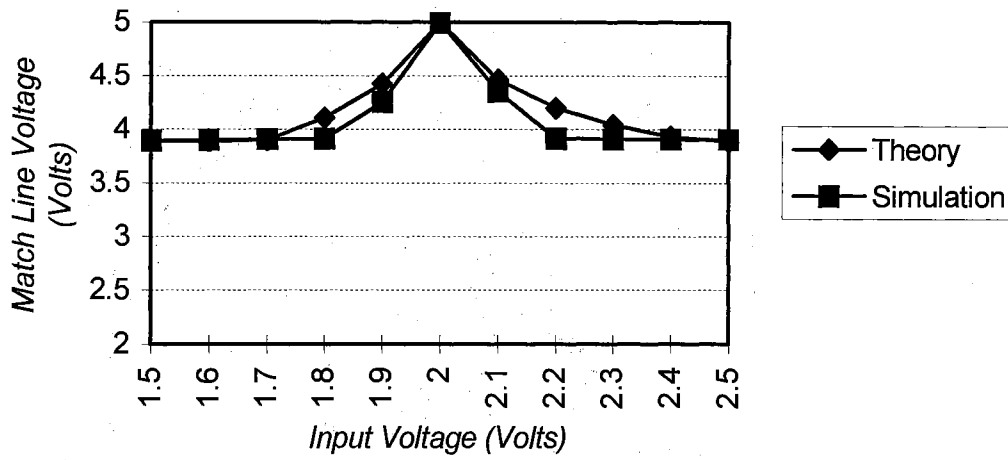


Figure VI-7. Input Voltage vs. Match Line Voltage ($V_{ref}=2$ Volts)

$V_{ref} = 2.5$ Volts

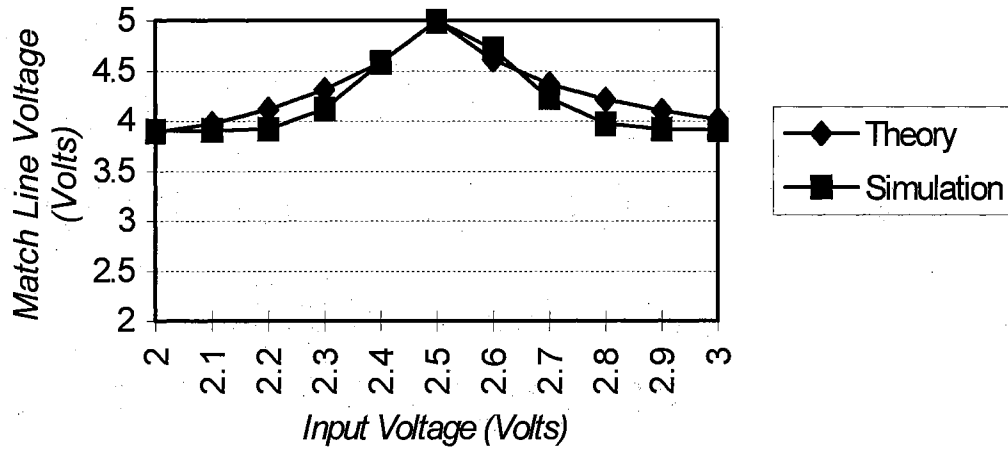


Figure VI-8. Input Voltage vs. Match Line voltage ($V_{ref}=2.5$ Volts)

$V_{ref} = 3$ Volts

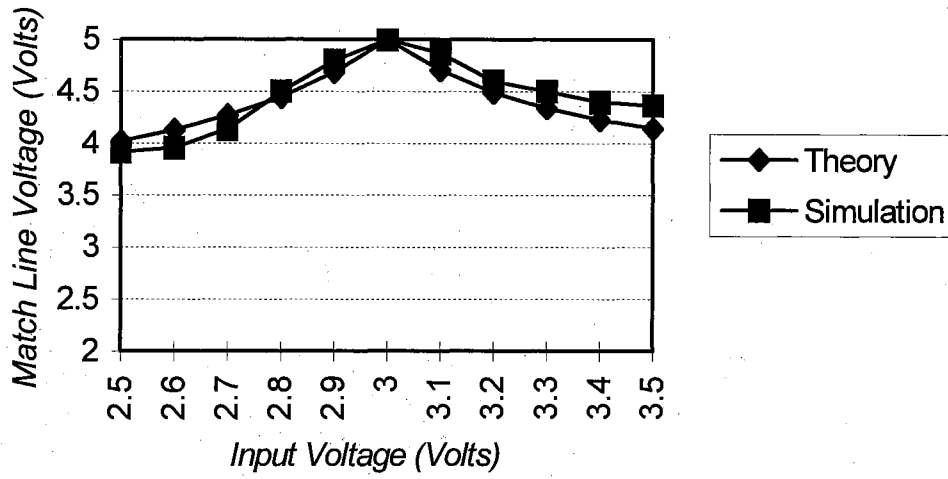


Figure VI-9. Input Voltage vs. Match Line Voltage ($V_{ref}=3$ Volts)

$V_{ref} = 3.5$ Volts

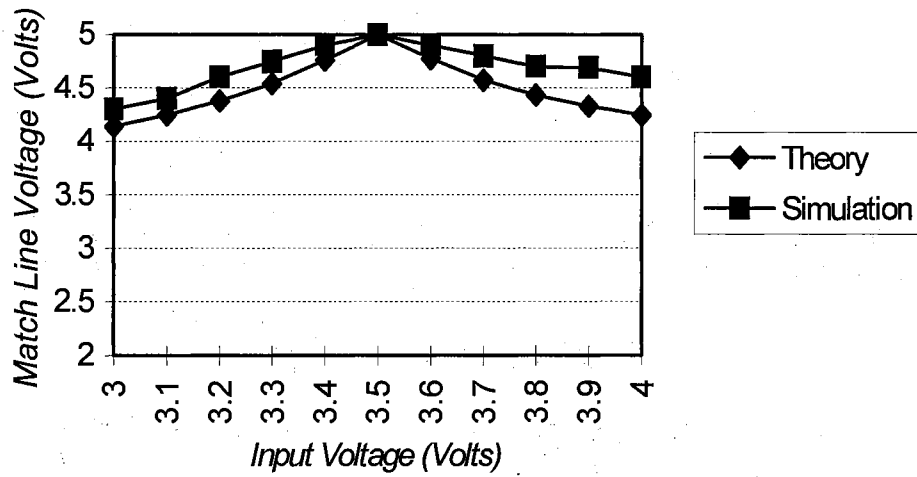


Figure VI-10. Input Voltage vs. Match Line Voltage ($V_{ref}=3.5$ Volts)

$V_{ref} = 4$ Volts

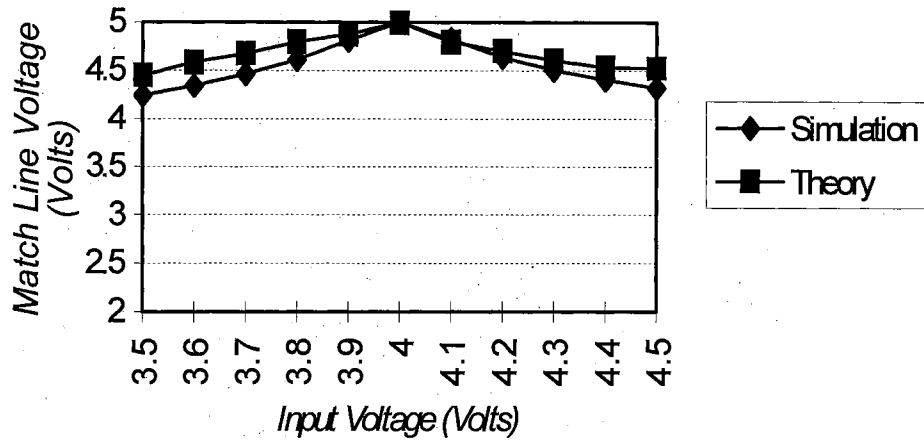


Figure VI-11. Input Voltage vs. Match Line Voltage ($V_{ref} = 4$ Volts)

Figure VI-13 shows an example of how the engine separates the winner from the loser despite the small difference between the patterns in these competing match lines. This figure indicates, although there is only one CAM mismatched between the two match lines, the recognition engine clearly picks the line with the closest pattern as the winner.

Winner match line voltage VS. Loser match line voltage

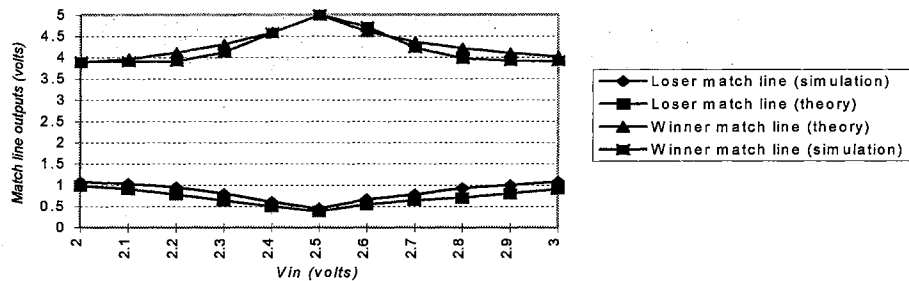


Figure VI-12. Comparison between the winner and the loser match lines ($V_{ref} = 2.5$ Volts)

CHAPTER VII

SILICON RESULTS

Introduction

This chapter characterizes the performance of the recognition engine manufactured in the 2.0u ORBIT process. The characterization includes programming and evaluation of each CAM cell. Furthermore, it investigates the validity of the simulation and the closed-form solution results against the results measured from silicon.

Silicon Results

In order to verify the functionality of the silicon, a very conservative approach was taken. The main objective was to show the functionality of the architecture as described in this thesis so far. This mainly includes, verifying the programming and the evaluation concept. To do so, a 2X2 array of the recognition architecture was fabricated in the 2.0u ORBIT process (figure VII-1). In this architecture, the EEPROM cells were used as the reference voltage storage devices along with four high voltage drivers to program each cell independently.

The programming process for the manufactured engine in the 2.0u ORBIT process

started by setting the *start_prog* signal to five volts and enabling the rows and columns of the CAM cell which was intended to be programmed. Then, the desired pattern voltage to be programmed in the memory, was applied to the input side of the selected CAM cell (V_{in}) while the programming direction was forced from outside the chip. Next, a high voltage programming pulse was applied followed by a precharge and evaluation clock. After each programming pulse, the current through the p-channel transistor (M_{pc}) was mirrored outside the chip where it was measured. This ensured that each cell was programmed in the appropriate direction. As the equivalent voltage in the EEPROM cell of the CAM under program got closer to the pattern voltage, the current through the p-channel device M_{po} (figure VII-1) got smaller. The programming continued until the measured current approached zero. At this point, the CAM was programmed and the programming cycle for the cell under program was complete. The same programming procedure was followed until all of the cells were programmed to the desired patterns.

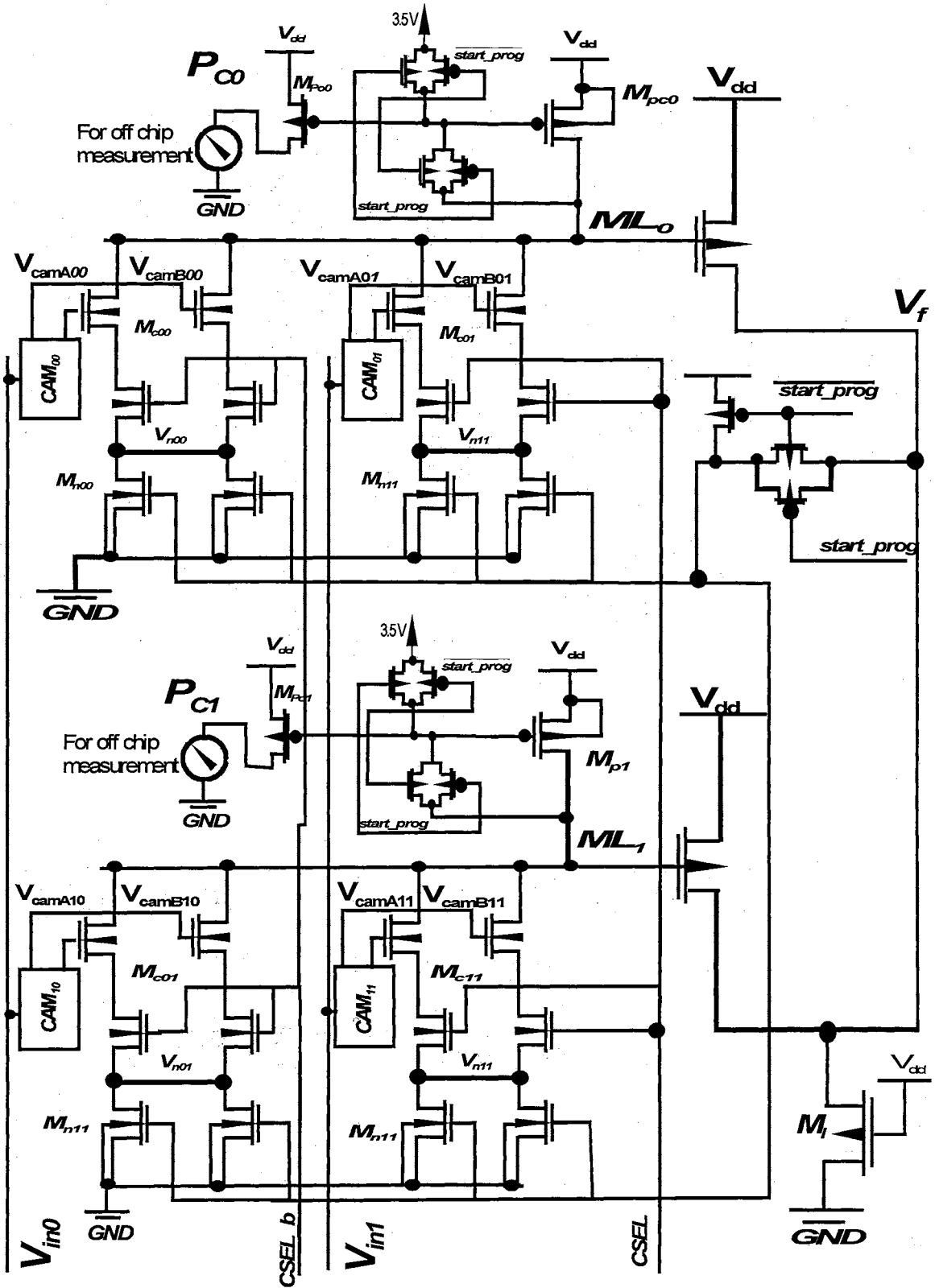


Figure VII-1. Fabricated 2X2 recognition engine

Once all of the cells were successfully programmed, the *start_prog* signal was set low, indicating that the engine is ready for the evaluation process. To investigate the functionality of the engine in the this mode, the input to one of the CAM cells (CAM_{00}) was purposely varied from -0.5 to + 0.5 volts of its stored reference voltage ($V_{ref_{00}}$). In the competing match line, the CAM (CAM_{10}) with its input connected to the input voltage varying +/- 0.5, was programmed so that the difference between its stored value ($V_{ref_{10}}$) and the input pattern was at least two volts. The input to the remaining CAMs (CAM_{01} and CAM_{11}) were well matched; hence, they did not contribute to any mismatch in the competing match lines. To summarize, the first match line (ML_0) had one CAM (CAM_{01}) with matched inputs ($V_{in_1} = V_{ref_{01}}$) and one CAM (CAM_{00}) with its input voltage a maximum of 0.5 volts below or above the stored pattern ($V_{ref_{00}} - 0.5 \leq V_{in_0} \leq V_{ref_{00}} + 0.5$). The competing match line (ML_1) on the other hand, had one CAM cell (CAM_{11}) with matched inputs ($V_{in_1} = V_{ref_{11}}$), and the second CAM cell whose input was different from the input pattern by at least two volts ($V_{ref_{10}} - 2 \geq V_{in_0} \geq V_{ref_{10}} + 2$). Therefore, there was one unmatched and one matched CAM in each line with the first line being the closest match compared to the second line. The goal was to investigate how the real silicon chooses the winner, and how the winning match line changes as one of the input patterns (V_{in_0}) varies from -0.5 volts of the reference pattern of CAM_{00} to 0.5 volts above it. These results were then compared with the results found using simulation and the closed-form solution.

Figures VII-2 through VII-6 depict these results. It is clear that there is a great accordance between the theory, the simulation and the actual silicon. Therefore, it is safe

to assume that the behavior of the system is very well defined using either simulation or theory.

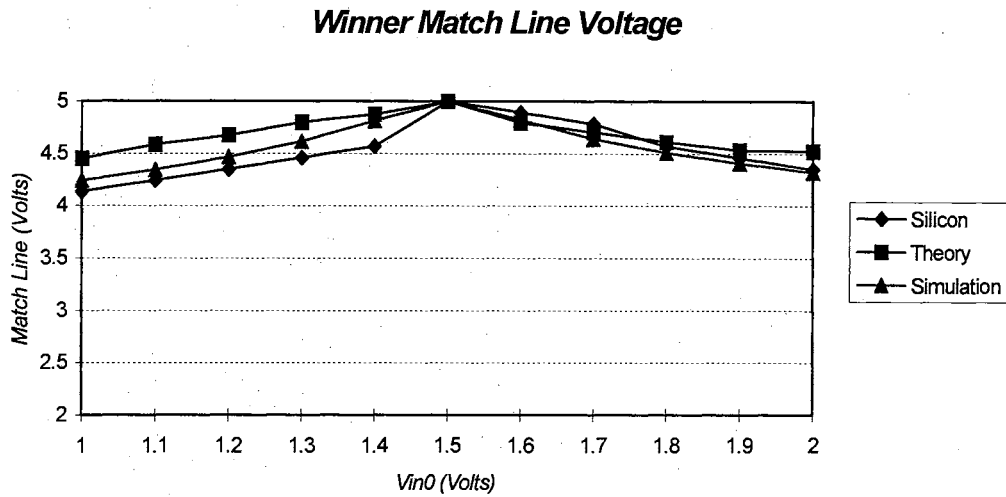


Figure VII-2. Winner match line voltage ($V_{ref00} = 1.5$ Volts)

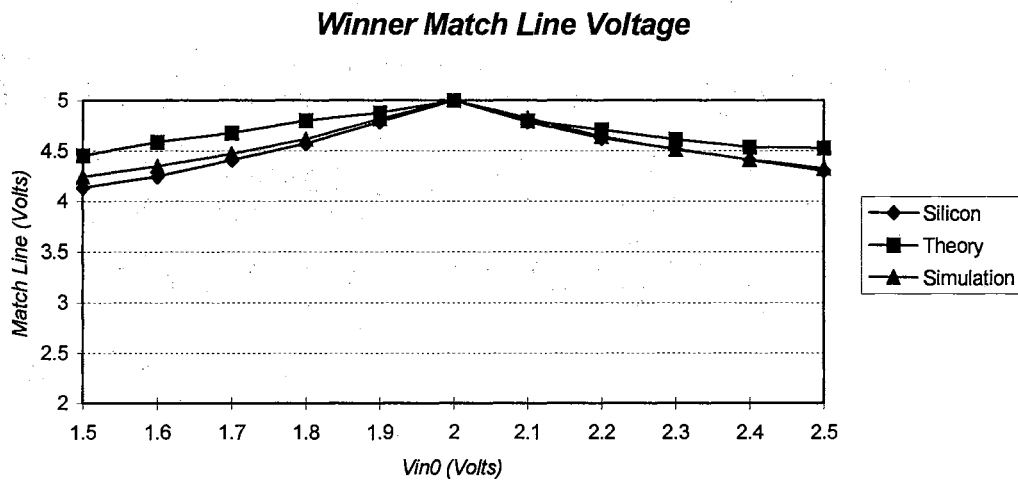


Figure VII-3. Winner match line voltage ($V_{ref00} = 2$ Volts)

Winner Match Line Voltage

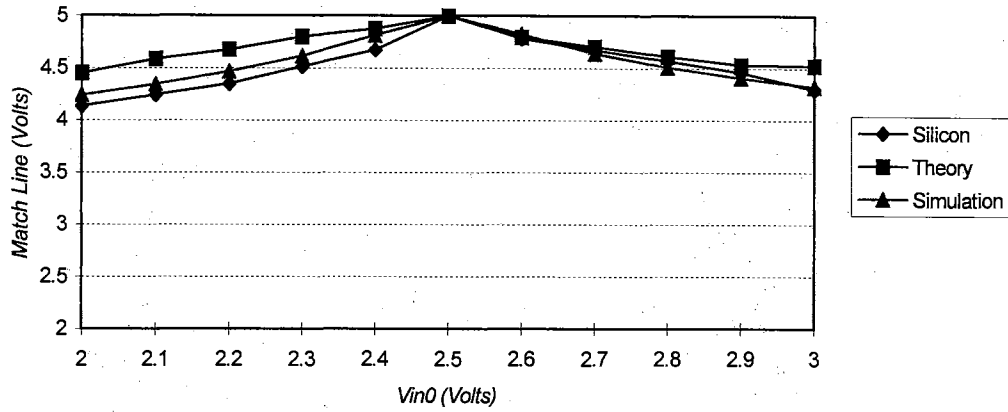


Figure VII-4. Winner match line voltage ($V_{ref00} = 2.5$ Volts)

Winner Match Line Voltage

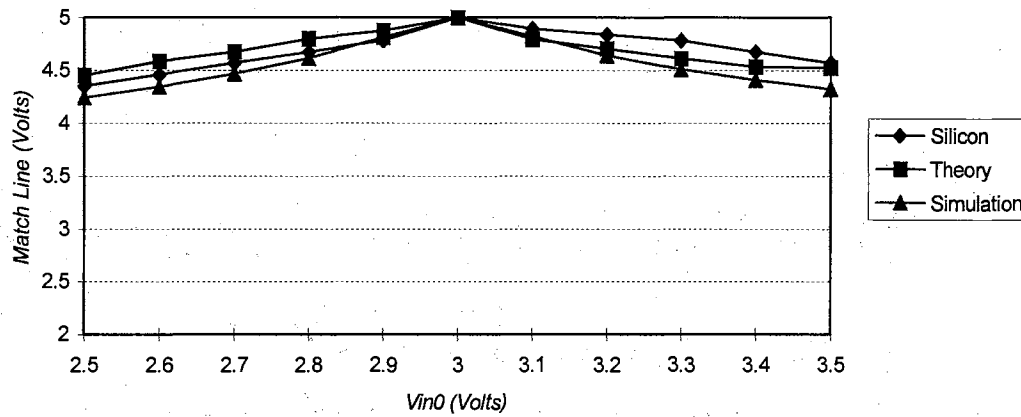


Figure VII-5. Winner match line voltage ($V_{ref00} = 3$ Volts)

Winner Match Line Voltage

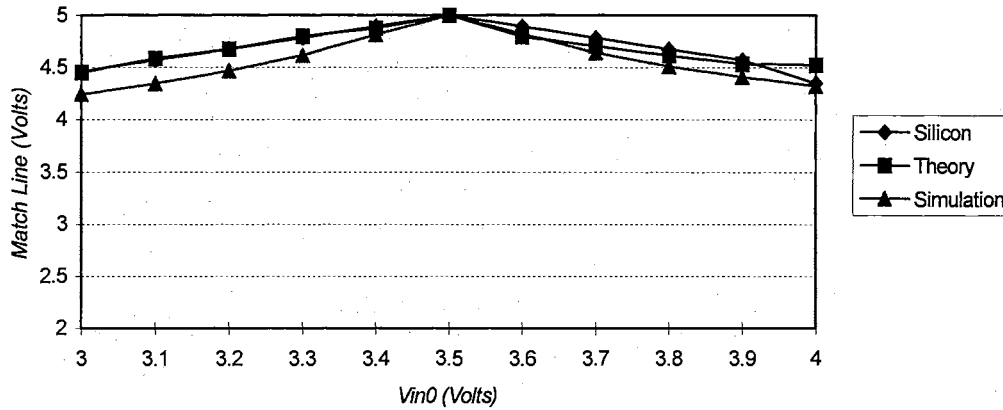


Figure VII-6. Winner match line voltage ($V_{ref0} = 3.5$ Volts)

It was also observed that the loser voltage was near the ground voltage while evaluating the above conditions. This implies that although there is one unmatched and one matched CAM in each line with the first line being the closest match compared to the second line, the proposed recognition engine clearly picks the closer match line as the winner.

CHAPTER VIII

CONCLUSION

The objective of this thesis was to propose a low power and high speed competitive classifier architecture. To pursue this goal a novel qualifier was proposed. This qualifier was implemented as an ultra low power analog content addressable memory (CAM) and was fully characterized. The characterization included a mathematical closed-form expression relating the CAM's steady state output to its input. Although the closed-form uses a simplified transistor model, it was demonstrated that its results were comparable with the results found using much more accurately modeled transistors (simulation results). To complete the qualifier, an analog memory (EEPROM) was used to constitute the long term memory for each CAM cell.

Next, the learning process was defined. This process involves programming the data pattern directly in the analog memory cells (a simple learning algorithm). Furthermore, it was demonstrated that the learning process is totally independent of any process variation and mismatch in the CAM cells. This important characteristic becomes inherent to the system by programming the qualifier cell through the match line that each cell uses to evaluate its inputs. Thanks to this learning technique, any additional offset in the programming path that is not common between the programming path and the evaluation path, appears as a DC offset associated with each single cell, causing the

outcome to be independent of any process mismatch in the system. As a result, the end user does not notice any process variation and mismatches.

Furthermore, it was indicated that the learning process requires high voltage pulses to move charge in and out of the analog memory cells. It was shown that there are two factors that effect the charge transfer during the programming process, amplitude and duration of the programming pulse. It was also indicated that the oxide thickness varies across the die and from one wafer to the next; therefore, effecting the amplitude voltage needed to get the minimum observable charge flow in and out of different floating gates (given the same pulse duration). This amplitude voltage can easily exceed the physical limitation of the standard process used in this project (2.0u ORBIT process). Therefore in this process, a regular MOSFET can not be employed to reliably generate the high voltage pulse required. As a result, a unique method to implement a high voltage n-channel MOSFET using a standard process with no special doping or additional mask layer was proposed and successfully manufactured. Results show this transistor has consistent characteristics when manufactured on the same wafer or on different dies.

Based on the proposed field transistor, a high voltage driver was proposed and manufactured. This circuit produces three different voltage level pulses necessary to program each analog memory.

To complete the architecture, a discriminator of $O(N)$ complexity [Johnson, Jun., 1992] was fully characterized. The characterization included a closed-form solution for the discriminator circuit (WTA). The closed-form solution was then combined with the closed-form solution found for the qualifier cell in order to relate the final steady state output of the discriminator to the inputs of the system (inputs of the qualifier). It was

shown that even though the transistor models were grossly simplified, the results obtained from the closed-form agreed with the result from the simulation.

Finally, a complete 2x2 recognition engine was fabricated in a 2.0u ORBIT process. It was proven that the silicon's behavior is well characterized either using numerical solution (HSPICE) or a closed-form solution derived in this thesis; therefore, a fast recognition engine has been designed and fully characterized.

REFERENCES

- Bleiker C. and Melchior H. (Jun. 1987), "A Four State EEPROM Using Floating Gate Memory Cells", IEEE JSSC, Vol. SC-22, NO. 3, PP: 460-463.
- Carley L. R. (DEC. 1989), "Trimming Analog Circuits Using Floating Gate Analog MOS Memory", IEEE JSSC, Vol. 24, NO.6, PP: 1569-1575. Chien G. (1995), "A Floating Gate Programmable MOSFET Using Standard Double-Poly CMOS Process", University Of California.
- Cioaca D., Lin T., Chan A., Chen L. and Mihnea A. (Oct. 1987), "A Million Cycle CMOS 256K EEPROM", IEEE JSSC, Vol. SC-22, NO. 5, PP: 684-691.
- Choi J. and Sheu B.J. (May 1993), "A High Precision VLSI Winner Take All Circuit for Self-Organizing Neural Networks", IEEE JSSC, vol. 28, NO. 5, PP:576-583.
- Dolny G. M., Schade O. H., Goldsmith B. and Goodman L. A. (Dec. 1986), "Enhanced CMOS for Analog-Digital Power IC Applications", IEEE Transactions On Electron Devices, Vol. ED-33, NO. 12, PP: 1985-1990.
- Ellis R. K. (Nov. 1982), "Fowler-Nordheim Emission from Non-Planar Surfaces", IEEE Electron Device Letters, Vol. EDL-3, NO.11, PP: 330-331.
- Francis Kub, Keith Moon, Ingham Mack, Francis Long, (Feb., 1990), "Programmable Analog Vector-Matrix Multiplier", IEEE JSSC, Vol.25, No.1, PP: 207-214.
- Fang S. (Oct 1983), "A Novel Nmos E2PROM Scheme", IEEE JSSC, Vol. SC-18, NO. 5, PP: 610-611.
- Guterman D. C., Rimawi I. H., Halvorson R. D. and McElroy D. (Apr. 1979), "An Electrically Alterable Nonvolatile Memory Cell Using a Floating Gate Structure", IEEE JSSC, Vol. SC-14, NO. 2, PP: 498-508.
- Grant, T., Taylor, J. and Housellander, P. (Sep. 1994), "Design, Implementation and Evaluation of a High-Speed Integrated Hamming Neural Classifier", IEEE JSSC, vol.29, No.9: PP1154-1157.
- Holler M., Tam S., Castro H. and Benson R. (Dec 1988), "An Electrically Trainable

- Artificial Neural Network (ETANN) with 10240 Floating Gate Synapses”, AIP Conference Proceedings, Snowbird, UT., PP. 11-191-11-196.
- Holler M., Tam S., Castro H. and Benson R. (Aug 1990),”Semiconductor Cell For Neural Network Employing A Four-Quadrant Multiplier”, Patent NO. 4950917.
- Hzuka H., Masuoka F., Sato T. and Ishikawa M. (Apr. 1976),” Electrically Alterable Avalanche-Injection-Type MOS Read-Only Memory with Stacked-Gate Structure”, IEEE Transactions On Electron Devices, Vol. 23, NO. 4, PP: 379,387.
- Johnson W. S., Perlegos G., Renninger A., Kuhn G., Ranganath T. (Feb., 1980), “A 16K electrically erasable non-volatile memory”, ISSCC Digest of Technical Papers, PP: 152.
- Joongho Choi, Bing J. Sheu (May. 1993),” A High-Precision VLSI Winner-Take-All Circuit for Self-Organizing Neural Networks”, IEEE JSSC, Vol.28, No.5, PP:576-583.
- Johnson L. G., Jalaeddine S. M. S. (Mar. 1991),”MOS Implementation of Winner-Take-All Network with Application to Content-Addressable Memory”, Electronics Letters.
- Johnson L. G., (Jan. 1997), “ECEN 5263 Digital VLSI System Design”, PP:1-15, Oklahoma State University.
- Johnson L. G., Jalaeddine S. M. S. (Sep. 1991),”Parameter Variations in MOS CAM with Mutual Inhibiting Network”, IEEE Transactions On Circuits And Systems, Vol. 38, NO. 9, PP: 1021-1028.
- Jalaeddine S. M. S., Johnson L. G. (Aug. 1992),”Associative Memory Integrated Circuit Based On Neural Mutual Inhibition”, IEE Proceedings, Vol. 39, NO. 4, PP: 445-449.
- Jalaeddine S. M. S., Johnson L. G. (Jun. 1992),”Associative IC Memories with Relational Search and Nearest-Match Capabilities”, IEEE JSSC, Vol.27, NO. 6, PP: 892-900.
- Kub F. J., Moon K. K., Mack A. I., and Long F. M. (Feb. 1990),”Programmable Analog Vector Matrix Multipliers”, IEEE JSSC, vol.25, No. 1: PP. 207-214.
- Khang D., S. M. Sze (Jul., 1967), “A Floating Gate and its Application to memory Devices”, Bell Syst. Tech., PP:1288-1295.
- Kolodny A., Nieh S. T. K., Eitan B. and Shappir J. (Jun. 1986),” Analysis and Modeling of Floating Gate EEPROM Cells”, IEEE Transactions On Electron Devices, Vol.

ED-33, NO. 6, PP: 835-844.

Kamigaki Y., Minami S. I., Hagiwara T. H., Furusawa K., Furuno T. Uchida K. Terasawa M. and Yamazaki K. (Dec 1989), "Yield and Reliability of MNOS EEPROM Products", IEEE JSSC, Vol. 24, NO. 6, PP: 1714-1722.

Lee B. W., Sheu B. J. and Yang H. (Jun. 1991), "Analog Floating Gate Synapses for General Purpose VLSI Neural Computation", Transactions On Circuit and Systems, Vol. 38, NO. 6: PP. 654-658.

Lazzaro J., Ryckebusch S., Mahowald M. A. and Mead C. A. (1989), "Winner-Take-All of O(N) Complexity", Advances in Neural information processing systems, I. D. S. Touretzky, Ed. San Mateo, CA: Morgan Kaufmann, PP:703-711.

Lai Fang-shi and Hwang W. (Apr. 1997), "Design and Implementation of Differential Cascode Voltage Switch with Pass-Gate Logic for High-Performance Digital System", IEEE JSSCS, Vol. 32, No. 4, PP: 563-573.

Lenzlinger M. and Snow E. H. (Jan. 1969), "Fowler-Nordheim Tunneling into Thermally Grown SiO₂", J. Appl. Phys., Vol.40, PP: 278-283.

Lippmann R. P. (Apr. 1987), "An Introduction to Computing with Neural Nets", IEEE ASSP Magazine, PP: 4-22.

Meta-Software, "Hspice User's Manual Version H-92 Analysis and Methods", Vol.3.

Parpia Z., Salama C. A. and Hadaway A. (Nov. 1987), "Modeling and Characterization of CMOS-Compatible High-Voltage Device Structures", IEEE Transactions On Electron Devices, Vol. ED-34, NO. 11, PP: 2335-2343.

Pasternak J. H., Shubat A. S. and Salama C. A. (Apr. 1987), "CMOS Differential Pass-Transistor Logic Design", IEEE JSSC, Vol. SC-22, NO. 2, PP: 216-222.

Parpia Z., Salama C. A. and Hadaway R. A. (Oct. 1988), "A CMOS-Compatible High-Voltage IC Process", IEEE Transactions On Electron Devices, Vol. 35, NO. 10, PP: 1687-1694.

Parpia Z., Mena J. G. and Salama C. A. (Dec. 1986), "A Novel CMOS-Compatible High-Voltage Transistor Structure", IEEE Transactions On Electron Devices, Vol. ED-33, NO. 12, PP: 1948-1952.

Perfetti R. and Dreng (Oct. 1990), "Winner-Take-All Circuit for Neurocomputing applications", IEE proceedings, Vol. 137, NO. 5, PP: 353-359.

Soo, D. C. and Meyer, R.G. (Dec. 1982), "A Four-Quadrant NMOS Analog Multiplier",

- IEEE JSSC, vol.SC-17, NO. 6: PP 1174-1178.
- Schwartz, D. B., Howard, R. E. and Hubbards, W. E. (Apr. 1989), "A Programmable Analog Neural Network Chip", IEEE JSSC, vol.24, NO. 2: PP. 313-319.
- Tsividis Y. (Aug. 1983), "Principles of Operation and Analysis of Switched-Capacitor Circuits", Proceedings of IEEE, Vol. 71, PP: 926-940.
- Tsividis Y., Satyanarayana S. (Nov. 1987), "Analogue Circuits for Variable Synapse Electronic Neural Networks", Electronics Letters, Vol. 23, NO. 24, PP: 1313-1314.
- Tsividis, "Operation And Modeling Of The MOS Transistor", McGraw-Hill Book Company.
- Thomsen A. and Brooke A. (Mar. 1991), "A Floating Gate MOSFET with Tunneling Injector Fabricated Using a Standard Double-Polysilicon CMOS Process", IEEE Electron Device Letters, Vol. 12, NO. 3, PP: 111-113.
- Ugur Cilingiroglu (Jan. 1993), " A Charge Based Neural Hamming Classifier", IEEE JSSC, vol. 28, NO. 1, PP: 59-67.
- Ugur Cilingiroglu (Feb. 1991), "A purely Capacitive Synaptic Matrix for Fixed Weight Neural Networks", IEEE Transactions On Circuits And Systems, Vol. 38, NO. 2, PP: 210-217.
- Ugur Cilingiroglu, He Y. and Sanchez-Sinencio E. (Mar. 1993), "A High Density and Low Power Charge Based Hamming Network", IEEE Transactions On VLSI Systems, Vol. 1, No. 1, PP: 56-62.
- Yuping He, Ugur Cilingiroglu (May 1993), "A Charge-Based On-Chip Adaptation Kohonen Neural Network", IEEE Transactions On Neural Networks, Vol. 4, NO. 3, PP: 462-469.
- Watanabe, K. and Ogawa, S. (Sep. 1988), "Clock-Feedthrough Compensated Sample/Hold Circuits", Electronics Letters, vol.24, No.19: PP. 1110-1114.
- Wang S. T. (Jan. 1980), "Charge Retention of Floating Gate Transistors Under Applied Bias Conditions", IEEE Transactions On Electron Devices, Vol. ED-27, NO.1, PP: 297-299.
- Wolf S. and Tauber R. N., "Silicon Processing for the VLSI Era", lattice Press.
- Yong Q. L., Salama A. T., Seufert M., Schvan P. and King M. (Feb. 1997), "Design and Characterization of Submicron BICMOS Compatible High-Voltage NMOS and P-

MOS Devices", IEEE Transactions On Electron Devices, Vol. 44, NO. 2, PP: 331-338.

Yong - Yoong Chai and Johnson L. G. (June, 1996), "A 2x2 Analog Memory Implemented With a Special Layout Injector", IEEE Journal Of Solid-State Circuits, Vol. 31, No. 6, PP:856-859.

Yong - Yoong Chai. (December, 1994), "A 2x2 Analog Memory Implemented With a Special Layout Injector", Dissertation, Oklahoma State University.

Yong Q. L., Salama A. T., Seufert M., Schvan P. and King M. (Feb. 1997), "Design and Characterization of Submicron BICMOS Compatible High-Voltage NMOS and P-MOS Devices", IEEE Transactions On Electron Devices, Vol. 44, NO. 2, PP: 331-338.

APPENDIX A

Define

$V_{dd}=5$	Volts	
$V_T=.98$	Volts	Reference voltage pattern voltage
$V_{ref}=3$	Volts	N-channel threshold voltage
$V_{in}=3.5$	Volts	Input voltage pattern voltage
$W=4 \cdot 10^{-6}$	μm	Width of all of the n-channel transistors
$L=2 \cdot 10^{-6}$	μm	Length of all of the n-channel transistors
$k=30 \cdot 10^{-6}$	amps/Volt ²	
$\beta = k \cdot \frac{W}{L}$	amps/Volt ²	
$\alpha=1$		
$\gamma=2$		
$h=8$		
$C_L=40 \cdot 10^{-15}$	Farad	Load at the precharging nodes (C_{load})
$C_B=C_L$	Farad	
$C_A=C_B$	Farad	
$G_{B1}=3 \cdot \beta$	1/Ohm	Conductance of the transistor M_{B1}
$G_{A1}=3 \cdot \beta$	1/Ohm	Conductance of the transistor M_{A1}

$$G_{\text{eff}_B} = 4 \cdot \beta \quad 1/\text{Ohm} \quad \text{Conductance of the transistor } M_B$$

$$G_{\text{eff}_A} = 4 \cdot \beta \quad 1/\text{Ohm} \quad \text{Conductance of the transistor } M_A$$

$$G_{B2} = (V_{\text{in}} - V_T) \cdot \frac{\beta}{h} \quad 1/\text{Ohm} \quad \text{Conductance of the transistor } M_{B2}$$

$$G_{A2} = (V_{\text{ref}} - V_T) \cdot \frac{\beta}{h} \quad 1/\text{Ohm} \quad \text{Conductance of the transistor } M_{A2}$$

$$R_B = \frac{1}{G_{B1}} + \frac{1}{G_{B2}} \quad \text{Ohm} \quad \text{Eq.1}$$

$$R_A = \frac{1}{G_{A1}} + \frac{1}{G_{A2}} \quad \text{Ohm} \quad \text{Eq.2}$$

$$G_{A1} = G_{B1} = G_1 \quad 1/\text{Ohm}$$

$$I_{DA} = G_{\text{eff}_A} \cdot (V_B - V_{A1} - V_T) \quad \text{Eq.3}$$

$$I_{DB} = G_{\text{eff}_B} \cdot (V_A - V_{B1} - V_T) \quad \text{Eq.4}$$

$$V_{A1} = I_{DA} \cdot R_A \quad \text{Eq.5}$$

$$V_{B1} = I_{DB} \cdot R_B \quad \text{Eq.6}$$

$$I_{DA} = \frac{V_{A1}}{R_A} \quad \text{Eq.7}$$

$$I_{DB} = \frac{V_{B1}}{R_B} \quad \text{Eq.8}$$

$$I_{DA} = \frac{G_{eff_A}}{(1 + G_{eff_A} \cdot R_A)} \cdot V_B - G_{eff_A} \cdot \frac{V_T}{(1 + G_{eff_A} \cdot R_A)} \quad \text{Eq.9}$$

$$C_L \cdot \frac{d}{dt} V_A + \frac{G_{eff_A}}{(1 + G_{eff_A} \cdot R_A)} \cdot V_B = G_{eff_A} \cdot \frac{V_T}{(1 + G_{eff_A} \cdot R_A)} \quad \text{Eq.10}$$

$$I_{DB} = \frac{-(-G_{eff_B} \cdot V_A + G_{eff_B} \cdot V_T)}{(1 + G_{eff_B} \cdot R_B)} \quad \text{Eq.11}$$

$$C_L \cdot \frac{d}{dt} V_B + \frac{G_{eff_B}}{(1 + G_{eff_B} \cdot R_B)} \cdot V_A = G_{eff_B} \cdot \frac{V_T}{(1 + G_{eff_B} \cdot R_B)} \quad \text{Eq.12}$$

$$C_L \cdot \frac{d}{dt} V_A + \frac{G_{eff_A}}{(1 + G_{eff_A} \cdot R_A)} \cdot V_B = G_{eff_A} \cdot \frac{V_T}{(1 + G_{eff_A} \cdot R_A)} \quad \text{Eq.13}$$

$$C_L \cdot \frac{d}{dt} V_B + \frac{G_{eff_B}}{(1 + G_{eff_B} \cdot R_B)} \cdot V_A = G_{eff_B} \cdot \frac{V_T}{(1 + G_{eff_B} \cdot R_B)} \quad \text{Eq.14}$$

$$\begin{pmatrix} C_L & 0 \\ 0 & C_L \end{pmatrix} \cdot \begin{bmatrix} \frac{d}{dt} V_A \\ \frac{d}{dt} V_B \end{bmatrix} + \begin{bmatrix} 0 & \frac{G_{eff_A}}{(1 + G_{eff_A} \cdot R_A)} \\ \frac{G_{eff_B}}{1 + G_{eff_B} \cdot R_B} & 0 \end{bmatrix} \cdot \begin{pmatrix} V_A \\ V_B \end{pmatrix}$$

$$\mathbf{i} = \begin{bmatrix} \text{Geff}_A \cdot \frac{V_T}{(1 + \text{Geff}_A \cdot R_A)} \\ \text{Geff}_B \cdot \frac{V_T}{(1 + \text{Geff}_B \cdot R_B)} \end{bmatrix}$$

Eq.15

$$\mathbf{C} = \begin{pmatrix} C_L & 0 \\ 0 & C_L \end{pmatrix}$$

Let

$$G_A = \frac{\text{Geff}_A}{(1 + \text{Geff}_A \cdot R_A)}$$

Eq.16

$$G_B = \frac{\text{Geff}_B}{1 + \text{Geff}_B \cdot R_B}$$

Eq.17

$$\mathbf{G} = \begin{pmatrix} 0 & G_A \\ G_B & 0 \end{pmatrix}$$

$$\mathbf{K} = \begin{pmatrix} G_A \cdot V_T \\ G_B \cdot V_T \end{pmatrix}$$

$$\mathbf{C}^{-1} = \begin{bmatrix} \frac{1}{C_L} & 0 \\ 0 & \frac{1}{C_L} \end{bmatrix}$$

$$S = C^{-1} \cdot G$$

Eq.18

$$S = \begin{bmatrix} 0 & \frac{1}{C_L} \cdot G_A \\ \frac{1}{C_L} \cdot G_B & 0 \end{bmatrix}$$

Finding the eigen values

$$\begin{bmatrix} 0 & \frac{1}{C_L} \cdot G_A \\ \frac{1}{C_L} \cdot G_B & 0 \end{bmatrix} - \lambda \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 0$$

Eq.19

$$\begin{bmatrix} -\lambda & \frac{1}{C_L} \cdot G_A \\ \frac{1}{C_L} \cdot G_B & -\lambda \end{bmatrix} = 0$$

$$\lambda_0 = \frac{1}{C_L} \cdot \sqrt{G_A} \cdot \sqrt{G_B}$$

$$\lambda_1 = -\frac{1}{C_L} \cdot \sqrt{G_A} \cdot \sqrt{G_B}$$

Or

$$\lambda_0 = -\lambda_1$$

Now finding the first eigen vector

$$\begin{bmatrix} -\lambda_0 & \frac{1}{C_L} \cdot G_A \\ \frac{1}{C_L} \cdot G_B & -\lambda_0 \end{bmatrix} \cdot \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = 0$$

Eq.20

$$X_1 = \sqrt{G_A} \cdot \frac{X_2}{\sqrt{G_B}}$$

$$v_0 = \begin{bmatrix} \sqrt{G_A} \cdot \frac{1}{\sqrt{G_B}} \\ 1 \end{bmatrix}$$

Normalized eigen vector

$$v_0 = \frac{1}{\sqrt{G_A + G_B}} \cdot \begin{pmatrix} \sqrt{G_A} \\ \sqrt{G_B} \end{pmatrix}$$

Eq.21

Now finding the second eigen vector

$$\begin{bmatrix} -\lambda & \frac{1}{C_L} \cdot G_A \\ \frac{1}{C_L} \cdot G_B & -\lambda \end{bmatrix} \cdot \begin{pmatrix} X_3 \\ X_4 \end{pmatrix} = 0$$

Eq.22

$$X_3 = -\sqrt{G_A} \cdot \frac{X_4}{\sqrt{G_B}}$$

$$v_1 = \begin{bmatrix} -\sqrt{G_A} \cdot \frac{1}{\sqrt{G_B}} \\ 1 \end{bmatrix}$$

Normalized eigen vector

$$v_1 = \frac{1}{\sqrt{G_A + G_B}} \cdot \begin{pmatrix} -\sqrt{G_A} \\ \sqrt{G_B} \end{pmatrix}$$

Eq.23

The general solution is

$$V_{(t)} = G^{-1} \cdot K + U \cdot e^{-SD \cdot t} \cdot U^{-1} \cdot [V_{(0)} - G^{-1} \cdot K]$$

Eq.24

Where matrix U is a 2x2 matrix where each column is the normalized eigen vectors and

$$e^{-SD \cdot t} = \begin{pmatrix} e^{-\lambda_0 \cdot t} & 0 \\ 0 & e^{\lambda_0 \cdot t} \end{pmatrix}$$

Setting initial conditions

$$V_{A(0)} = V_{dd}$$

$$V_{B(0)} = V_{dd}$$

$$G^{-1} = \begin{bmatrix} 0 & \frac{1}{G_B} \\ \frac{1}{G_A} & 0 \end{bmatrix}$$

$$K = \begin{pmatrix} G_A \cdot V_T \\ G_B \cdot V_T \end{pmatrix}$$

$$G^{-1} \cdot K = \begin{pmatrix} V_T \\ V_T \end{pmatrix}$$

Constructing the U matrix

$$U = \frac{1}{\sqrt{G_A + G_B}} \cdot \begin{pmatrix} \sqrt{G_A} & -\sqrt{G_A} \\ \sqrt{G_B} & \sqrt{G_B} \end{pmatrix}$$

$$U^{-1} = \frac{1}{2} \cdot \sqrt{G_A + G_B} \cdot \begin{bmatrix} \frac{1}{\sqrt{G_A}} & \frac{1}{\sqrt{G_B}} \\ \sqrt{G_A} & \sqrt{G_B} \\ -1 & 1 \\ \sqrt{G_A} & \sqrt{G_B} \end{bmatrix}$$

$$U \cdot e^{-SD \cdot t} \cdot U^{-1} = \begin{bmatrix} \frac{1}{2} \cdot \exp(-\lambda_0 \cdot t) + \frac{1}{2} \cdot \exp(\lambda_0 \cdot t) & \frac{1}{2} \cdot \sqrt{G_A} \cdot \frac{(\exp(-\lambda_0 \cdot t) - \exp(\lambda_0 \cdot t))}{\sqrt{G_B}} \\ \frac{1}{2} \cdot \sqrt{G_B} \cdot \frac{(\exp(-\lambda_0 \cdot t) - \exp(\lambda_0 \cdot t))}{\sqrt{G_A}} & \frac{1}{2} \cdot \exp(-\lambda_0 \cdot t) + \frac{1}{2} \cdot \exp(\lambda_0 \cdot t) \end{bmatrix}$$

The solution is

$$V_{(t)} = G^{-1} \cdot K + U \cdot e^{-SD \cdot t} \cdot U^{-1} \cdot [V_{(0)} - G^{-1} \cdot K]$$

$$V_{A(t)} = a \cdot \exp(-\lambda_0 \cdot t) + b \cdot \exp(\lambda_0 \cdot t) + V_T \quad \text{Eq.25}$$

$$V_{B(t)} = c \cdot \exp(-\lambda_0 \cdot t) + d \cdot \exp(\lambda_0 \cdot t) + V_T \quad \text{Eq.26}$$

$$a = \frac{-1}{2} \cdot \frac{(-\sqrt{G_B} \cdot V_{A(0)} + V_T \cdot \sqrt{G_B} - \sqrt{G_A} \cdot V_{B(0)} + V_T \cdot \sqrt{G_A})}{\sqrt{G_B}} \quad \text{Eq.27}$$

$$b = \frac{-1}{2} \cdot \frac{(-\sqrt{G_B} \cdot V_{A(0)} + V_T \cdot \sqrt{G_B} + \sqrt{G_A} \cdot V_{B(0)} - V_T \cdot \sqrt{G_A})}{\sqrt{G_B}} \quad \text{Eq.28}$$

$$c = \frac{-1}{2} \cdot \frac{(-\sqrt{G_B} \cdot V_{A(0)} + V_T \cdot \sqrt{G_B} - \sqrt{G_A} \cdot V_{B(0)} + V_T \cdot \sqrt{G_A})}{\sqrt{G_A}} \quad \text{Eq.29}$$

$$d = \left[\frac{1}{2} \frac{(\sqrt{G_B} \cdot V_{A(0)} - V_T \cdot \sqrt{G_B} - \sqrt{G_A} \cdot V_{B(0)} + V_T \cdot \sqrt{G_A})}{\sqrt{G_A}} \right] \quad \text{Eq.30}$$

$$V_{A(t)} = a \cdot \exp(-\lambda_0 \cdot t) + b \cdot \exp(\lambda_0 \cdot t) + V_T \quad \text{Eq.31}$$

$$V_{B(t)} = c \cdot \exp(-\lambda_0 \cdot t) + d \cdot \exp(\lambda_0 \cdot t) + V_T \quad \text{Eq.32}$$

Final voltage has reached when $dV_A/dt=0$

$$\frac{d}{dt} (a \cdot \exp(-\lambda_0 \cdot t) + b \cdot \exp(\lambda_0 \cdot t) + V_T) = 0 \quad \text{Eq.33}$$

$$-a \cdot \exp(-\lambda_0 \cdot t) + b \cdot \exp(\lambda_0 \cdot t) = 0 \quad \text{Eq.34}$$

Multiplying both sides by $\exp(\lambda_0 \cdot t)$

Solve for the time $dV_A/dt=0$

$$\exp(\lambda_0 \cdot t) = \sqrt{\frac{a}{b}} \quad \text{Eq.35}$$

And

$$\exp(-\lambda_0 \cdot t) = \sqrt{\frac{b}{a}} \quad \text{Eq.36}$$

Plug above expressions in to the $V_{A(t)}$ to get the steady state Voltage

$$V_{Af} = 2 \cdot \sqrt{a \cdot b} + V_T \quad \text{Eq.37}$$

$$V_{Af} = \sqrt{\left(V_{A0} - V_T\right)^2 - \frac{G_A}{G_B} \cdot \left(V_{B0} - V_T\right)^2} + V_T \quad \text{Eq.38}$$

Same procedure is followed to find the steady state voltage of the V_B

$$V_{Bf} = 2 \cdot \sqrt{c \cdot d} + V_T \quad \text{Eq.39}$$

$$V_{Bf} = \sqrt{\left(V_{B0} - V_T\right)^2 - \frac{G_B}{G_A} \cdot \left(V_{A0} - V_T\right)^2} + V_T \quad \text{Eq.40}$$

APPENDIX B

Define

$V_{Tf} = 1.4$	Volts	Threshold of any n_channel transistor with body effect
$W_f = 4$	μm	Width of the transistor M_f
$L_f = 8$	μm	Length of the transistor M_f
$W_f = 20$	μm	Width of the transistor M_f
$L_f = 2$	μm	Length of the transistor M_f
$W_n = 6$	μm	Width of all of the pull down transistors
$L_n = 2$	μm	Length of all of the pull down transistors
$W_p = 20$	μm	Width of the p-channel current source transistor
$L_p = 2$	μm	Length of the p-channel current source transistor
$V_{Tp} = 1$	Volt	Threshold of the p-channel transistor with no body effect
$V_{Tn} = 0.9$	Volts	Threshold of any n-channel transistor with no body effect
$V_{dd} = 5$	Volts	
$V_{Tfo} = 1.4$	Volts	Threshold of any n_channel transistor with body effect
$C_{ox} = 50$	Farad	Oxide capacitance per unit area
μ_{0n}	$\mu m/V.nS$	N-channel mobility
$\mu_{op} = \frac{\mu_{0n}}{2.5}$	$\mu m/V.nS$	P-channel mobility

$v_{\max} = 20$ $\mu \text{ m/nS}$ **Maximum Saturation velocity**

$V = v_{\max} \cdot \frac{L}{\mu_0}$ **V** **Carrier saturation voltage**

$K_p = \mu_0 \cdot C_{ox}$ **amps/Volt²**

$K_n = \mu_0 \cdot C_{ox}$ **amps/Volt²**

$\beta_n = K_n \cdot \frac{W_n}{L_n}$ **amps/Volt²**

$\beta_p = K_p \cdot \frac{W_p}{L_p}$ **amps/Volt²**

$\beta_I = K_I \cdot \frac{W_I}{L_I}$ **amps/Volt²**

$V_{cn} = v_{\max} \cdot \frac{L_n}{\mu_{0n}}$ **Volts**

$V_{cp} = v_{\max} \cdot \frac{L_p}{\mu_{0p}}$ **Volts**

$V_{cf} = v_{\max} \cdot \frac{L_f}{\mu_{0n}}$ **Volts**

$V_{Dsatf} = V_{cf}$ **Volts**

$V_{Dsatn} = V_{cn}$ **Volts**

$V_{bias} = 3.5$ **Volts**

V_{cam} is the output of the mismatched CAM in the winner match line

V_{caml} is the output of the mismatched CAM in the loser match line

The following equation are used

$$I_D = \beta \cdot (V_{GS} - V_T) \cdot V_{DS} \quad V_{DS} < V_{DSat}$$

$$I_D = \beta \cdot V_c \cdot (V_{GS} - V_T) \quad V_{DS} \geq V_{DSat}$$

Summing the current in and out of the match line

$$0 = \beta_p \cdot (V_{dd} - V_{bias} - V_{Tp}) \cdot (V_{dd} - V_{mw}) - \sum_{i=1}^n \beta_n \cdot V_{cn} \cdot [(V_{cam})_i - (V_{nw})_i - V_{Tf}]$$

$$0 = G_a \cdot (V_{dd} - V_{mw}) - \sum_{i=1}^n \beta_n \cdot V_{cn} \cdot [(V_{cam})_i - (V_{nw})_i - V_{Tf}] \quad \text{Eq.1}$$

$$G_a = \beta_p \cdot (V_{dd} - V_{bias} - V_{Tp})$$

$$0 = \beta_n \cdot V_{cn} \cdot (V_{cam} - V_{nw} - V_{Tf}) - 2 \cdot \beta_n \cdot (V_f - V_{Tn}) \cdot V_{nw} \quad \text{Eq.2}$$

$$0 = \beta_f \cdot V_{cf} \cdot (V_{mw} - V_f - V_{Tfo}) - \beta_l \cdot (V_{dd} - V_{Tn}) \cdot V_f$$

$$0 = \beta_f \cdot V_{cf} \cdot (V_{mw} - V_f - V_{Tfo}) - G_b \cdot V_f \quad \text{Eq.3}$$

$$G_b = \beta_l \cdot (V_{dd} - V_{Tn})$$

Solving for V_{nw}

$$0 = \beta_n \cdot V_{cn} \cdot (V_{cam} - V_{nw} - V_{Tf}) - 2 \cdot \beta_n \cdot (V_f - V_{Tn}) \cdot V_{nw}$$

$$V_{nw} = -V_{cn} \cdot \frac{(V_{cam} - V_{Tf})}{(-V_{cn} - 2 \cdot V_f + 2 \cdot V_{Tn})} \quad \text{Eq.4}$$

$$0 = G_a \cdot (V_{dd} - V_{mw}) - 2 \cdot \beta_n \cdot (V_f - V_{Tn}) \cdot V_{nw}$$

$$V_{nw} = \frac{1}{2} \cdot G_a \cdot \frac{(-V_{dd} + V_{mw})}{[\beta_n \cdot (-V_f + V_{Tn})]} \quad \text{Eq.5}$$

Substituting Eq.4 in Eq.5

Solving for V_f in terms of V_{mw}

$$V_f = V_{Tn} - \frac{1}{2} \cdot V_{cn} \cdot G_a \cdot \frac{(-V_{dd} + V_{mw})}{(\beta_n \cdot V_{cn} \cdot V_{cam} - \beta_n \cdot V_{cn} \cdot V_{Tf} - G_a \cdot V_{dd} + G_a \cdot V_{mw})} \quad \text{Eq.6}$$

Now solving for V_{mw} in terms of V_f Using Eq.3

$$0 = \beta_f \cdot V_{cf} \cdot (V_{mw} - V_f - V_{Tfo}) - G_b \cdot V_f$$

$$V_{mw} = \left[1 + \frac{G_b}{(\beta_f \cdot V_{cf})} \right] \cdot V_f + V_{Tfo} \quad \text{Eq.7}$$

Substituting Eq.7 in Eq.6 for V_{mw} and solving for V_f

$$V_f = V_{Tn} - \frac{1}{2} \cdot G_a \cdot V_{cn} \cdot \frac{\left[-V_{dd} + \left[1 + \frac{G_b}{(\beta_f \cdot V_{cf})} \right] \cdot V_f + V_{Tfo} \right]}{\left[\beta_n \cdot V_{cn} \cdot V_{cam} - \beta_n \cdot V_{cn} \cdot V_{Tf} - G_a \cdot V_{dd} + G_a \cdot \left[\left[1 + \frac{G_b}{(\beta_f \cdot V_{cf})} \right] \cdot V_f + V_{Tfo} \right] \right]} \quad \text{Eq.8}$$

$$A \cdot (V_f)^2 + B \cdot V_f + C = 0 \quad \text{Eq.9}$$

$$A = (-2 \cdot G_b - 2 \cdot \beta_f \cdot V_{cf}) \cdot G_a$$

$$B(V_{cam}) = \left[(2 \cdot \beta_n \cdot V_{Tf} - G_a) \cdot V_{cn} + 2 \cdot G_a \cdot V_{dd} - 2 \cdot \beta_n \cdot V_{cn} \cdot V_{cam} - 2 \cdot G_a \cdot V_{Tfo} + 2 \cdot V_{Tn} \cdot G_a \right] \cdot V_{cf} \cdot \beta_f \dots \\ + (2 \cdot V_{Tn} \cdot G_a \cdot G_b - G_a \cdot G_b \cdot V_{cn})$$

$$C(V_{cam}) = \left[(2 \cdot V_{cf} \cdot V_{cam} - 2 \cdot V_{Tf} \cdot V_{cf}) \cdot \beta_n \cdot \beta_f \cdot V_{Tn} + (V_{cf} \cdot V_{dd} - V_{Tfo} \cdot V_{cf}) \cdot \beta_f \cdot G_a \right] \cdot V_{cn} \dots \\ + (-2 \cdot V_{cf} \cdot V_{dd} + 2 \cdot V_{Tfo} \cdot V_{cf}) \cdot \beta_f \cdot G_a \cdot V_{Tn}$$

$$V_f(V_{cam}) = \frac{1}{(2 \cdot A)} \cdot \left(-B(V_{cam}) - \sqrt{B(V_{cam})^2 - 4 \cdot A \cdot C(V_{cam})} \right) \quad \text{Eq.10}$$

Substitute for V_f to find V_{nw} and V_{mw} in Eqs. 5&7

$$V_{mw}(V_{cam}) = \left[1 + \frac{G_b}{(\beta_f \cdot V_{cf})} \right] \cdot V_f(V_{cam}) + V_{Tfo} \quad \text{Eq.11}$$

$$V_{nw}(V_{cam}) = \frac{1}{2} \cdot G_a \cdot \frac{(-V_{dd} + V_{mw}(V_{cam}))}{\left[\beta_n \cdot (-V_f(V_{cam}) + V_{Tn}) \right]} \quad \text{Eq.12}$$

Now solving for V_{cam} voltage at which the feed back transistor leaves the ohmic region and enters the saturation region. This is the value of the V_{cam} at which the $V_{nw}(V_{cam})$ in Eq.12 is equal to the V_{Dsat} of the n-channel transistor. Furthermore, lets define the feed back voltage at which V_{nw} is equal to the V_{Dsat} to be called V_{fc} (the critical feedback voltage).

$$V_{nw} = \frac{1}{2} \cdot G_a \cdot \frac{(-V_{dd} + V_{mw}(V_{cam}))}{\left[\beta_n \cdot (-V_{fc} + V_{Tn}) \right]} \quad \text{Eq.12a}$$

Solving Eq.12a for V_{fc}

$$V_{fc} = -\beta_f \cdot V_{cf} \cdot \frac{\left[(-V_{dd} + V_{Tfo}) \cdot G_a - 2 \cdot V_{cn} \cdot \beta_n \cdot V_{Tn} \right]}{\left[(G_a + 2 \cdot V_{cn} \cdot \beta_n) \cdot \beta_f \cdot V_{cf} + G_a \cdot G_b \right]} \quad \text{Eq.12b}$$

Eq. 12b is the feed back voltage at which the feed back transistor leaves the ohmic region and enters the saturation region.

Rearrange the parameters A, B(V_{cam}) and C(V_{cam}) in Eq. 9 for simplicity

$$a = -2 \cdot V_{cn} \cdot V_{cf} \cdot \beta_n \cdot \beta_f$$

$$b = (2 \cdot V_{Tn} - V_{cn}) \cdot G_a \cdot G_b + (2 \cdot V_{dd} - 2 \cdot V_{Tfo} + 2 \cdot V_{Tn} - V_{cn}) \cdot \beta_f \cdot V_{cf} \cdot G_a + 2 \cdot \beta_f \cdot V_{cf} \cdot V_{cn} \cdot \beta_n \cdot V_{Tf}$$

$$B(V_{cam}) = a \cdot V_{cam} + b$$

$$c = 2 \cdot V_{cn} \cdot V_{cf} \cdot \beta_n \cdot V_{Tn} \cdot \beta_f$$

$$d = \left[\left[(-2 \cdot V_{Tn} + V_{cn}) \cdot V_{dd} + 2 \cdot V_{Tfo} \cdot V_{Tn} - V_{cn} \cdot V_{Tfo} \right] \cdot \beta_f \cdot G_a - 2 \cdot V_{cn} \cdot \beta_n \cdot V_{Tf} \cdot V_{Tn} \cdot \beta_f \right] \cdot V_{cf}$$

$$C(V_{cam}) = c \cdot V_{cam} + d$$

$$V_{fc} = \frac{1}{(2 \cdot A)} \cdot \left[- (a \cdot V_{cam} + b) - \sqrt{(a \cdot V_{cam} + b)^2 - 4 \cdot A \cdot (c \cdot V_{cam} + d)} \right] \quad \text{Eq.12c}$$

Now solving the CAM's voltage in terms of V_{fc}

$$V_{camc} = \frac{[-(V_{fc})^2 \cdot A - d - V_{fc} \cdot b]}{(c + V_{fc} \cdot a)} \quad \text{Eq. 12d}$$

This is the value of the V_{cam} which causes the feed back transistors to leave the ohmic region and enter the saturation region.

The Only transistor out of linear is the M_n transistor

Summing the current in and out of the match line

$$0 = \beta_p \cdot (V_{dd} - V_{bias} - V_{Tp}) \cdot (V_{dd} - V_{mw}) - \sum_{i=1}^n \beta_n \cdot V_{cn} \cdot [(V_{cam})_i - (V_{nw})_i - V_{Tf}]$$

$$0 = G_a \cdot (V_{dd} - V_{mw}) - \sum_{i=1}^n \beta_n \cdot V_{cn} \cdot [(V_{cam})_i - (V_{nw})_i - V_{Tf}]$$

$$G_a = \beta_p \cdot (V_{dd} - V_{bias} - V_{Tp}) \quad \text{Eq.13}$$

$$0 = \beta_n \cdot V_{cn} \cdot (V_{cam} - V_{nw} - V_{Tf}) - 2 \cdot \beta_n \cdot V_{cn} \cdot (V_f - V_{Tn}) \quad \text{Eq.14}$$

$$0 = \beta_f \cdot V_{cf} \cdot (V_{mw} - V_f - V_{Tfo}) - \beta_l \cdot (V_{dd} - V_{Tn}) \cdot V_f$$

$$0 = \beta_f \cdot V_{cf} \cdot (V_{mw} - V_f - V_{Tfo}) - G_b \cdot V_f \quad \text{Eq.15}$$

$$G_b = \beta_l \cdot (V_{dd} - V_{Tn})$$

Solving for V_{nw} using Eq.14

$$0 = \beta_n \cdot V_{cn} \cdot (V_{cam} - V_{nw} - V_{Tf}) - 2 \cdot \beta_n \cdot V_{cn} \cdot (V_f - V_{Tn}) \quad \text{Eq.16}$$

$$V_{nw} = V_{cam} - V_{Tf} - 2 \cdot V_f + 2 \cdot V_{Tn}$$

Solving for V_{mw} using Eq.15

$$0 = \beta_f \cdot V_{cf} \cdot (V_{mw} - V_f - V_{Tfo}) - G_b \cdot V_f$$

$$V_{mw} = \left[1 + \frac{G_b}{(\beta_f \cdot V_{cf})} \right] \cdot V_f + V_{Tfo} \quad \text{Eq.17}$$

Manipulating Eqs.13&14

$$G_a \cdot (V_{dd} - V_{mw}) - 2 \cdot \beta_n \cdot V_{cn} \cdot (V_f - V_{Tn}) = 0$$

$$V_{mw} = \frac{[-2 \cdot V_{cn} \cdot (V_f - V_{Tn})]}{G_a} \cdot \beta_n + V_{dd} \quad \text{Eq.18}$$

Equating Eq. 17&18 and solving for V_f and renaming it to V_{fs} for the feed back voltage in the saturation region

$$V_{fs} = \left[(-V_{Tfo} + V_{dd}) \cdot G_a + 2 \cdot \beta_n \cdot V_{cn} \cdot V_{Tn} \right] \cdot \beta_f \cdot \frac{V_{cf}}{\left[(\beta_f \cdot V_{cf} + G_b) \cdot G_a + 2 \cdot \beta_n \cdot V_{cn} \cdot \beta_f \cdot V_{cf} \right]}$$

Eq.19

$$V_{nws}(V_{cam}) = V_{cam} - V_{Tf} - 2 \cdot V_{fs} + 2 \cdot V_{Tn}$$

Eq.20

$$V_{mws} = \left[1 + \frac{G_b}{(\beta_f \cdot V_{cf})} \right] \cdot V_{fs} + V_{Tfo}$$

Eq.21

Where V_{mws} and V_{nws} indicate the winner match line voltage and drain voltage of the feed back transistor when the feed back transistor is in saturation.

Summing the current in and out of the loser match line

$$0 = \beta_p \cdot (V_{dd} - V_{bias} - V_{Tp}) \cdot (V_{dd} - V_{mw}) - \sum_{i=1}^{n_l} \beta_n \cdot \left[(V_{caml})_i - (V_{nl})_i - V_{Tf} \right] \cdot \left[(V_{ml})_i - (V_{nl})_i \right]$$

$$0 = G_a - \sum_{i=1}^{n_l} \beta_n \cdot \left[(V_{caml})_i - (V_{nl})_i - V_{Tf} \right] \cdot \left[(V_{ml})_i - (V_{nl})_i \right]$$

Eq.22

$$G_a = \beta_p \cdot (V_{dd} - V_{bias} - V_{Tp})$$

$$0 = \beta_n \cdot (V_{caml} - V_{nl} - V_{Tf}) \cdot (V_{ml} - V_{nl}) - 2 \cdot \beta_n \cdot (V_f - V_{Tn}) \cdot V_{nl}$$

Eq.23

Assuming there is only one mismatch in the loser line

$$0 = G_a - \beta_n \cdot (V_{caml} - V_{nl} - V_{Tf}) \cdot (V_{ml} - V_{nl})$$

Eq.24

$$0 = \beta_n \cdot (V_{caml} - V_{nl} - V_{Tf}) \cdot (V_{ml} - V_{nl}) - 2 \cdot \beta_n \cdot (V_f - V_{Tn}) \cdot V_{nl}$$

Eq.25

Solving for V_{ml} and V_{nl}

$$0 = G_a - 2 \cdot \beta_n \cdot V_{nl} \cdot V_f + 2 \cdot \beta_n \cdot V_{nl} \cdot V_{Tn}$$

$$V_{nl}(V_{caml}) = \frac{G_a}{(2 \cdot \beta_n \cdot V_f(V_{caml}) - 2 \cdot \beta_n \cdot V_{Tn})}$$

Eq.27

The feed back (V_f) voltage is set by the winner match line (using eq. 1L)

$$V_{ml}(V_{caml}) = V_{nl}(V_{caml}) + \frac{G_a}{[(V_{caml} - V_{nl}(V_{caml}) - V_{Tf}) \cdot \beta_n]}$$

Eq.28

It is found that all of the pull down transistors in the losing match line are in the ohmic region over the entire operating region

VITA

Shahriar Rokhsaz

Candidate for the degree

Doctor of Philosophy

Thesis: AN ANALOG CONTENT ADDRESSABLE MEMORY

Major Field: Electrical Engineering

Biographical:

Education: Received Bachelor of Science degree in Electrical Engineering from Oklahoma State University, Stillwater, Oklahoma in December 1990. Received Master of Science in Electrical Engineering from Oklahoma State University, Stillwater, Oklahoma in December 1992. Completed the requirements for the Doctor of Philosophy degree with a major in Electrical Engineering at Oklahoma State University in December 1998.

Experience: Employed as a test engineer at Austin Semiconductor during the summer of 1993. Employed as an analog designer from 1995 to 1996 at North Shore Circuit Design. Employed as a senior analog designer from 1996 to present at Sigmatel.