

CMOS (K)WINNERS-TAKE-ALL CIRCUITS
OF O(N) COMPLEXITY

100798



Ph.D. Thesis by
Barbaros ŞEKERKIRAN, M.Sc.

(492D032)

100798

Date of submission : 31 August 1999

Date of defence examination: 29 December 1999

Supervisor (Chairman): Prof.Dr. Uğur ÇİLİNGİROĞLU

Members of the Examining Committee Prof.Dr. Duran LEBLEBİCİ (İ.T.Ü.)

Prof.Dr. Hakan KUNTMAN (İ.T.Ü.)

Prof.Dr. Atilla ATAMAN (Y.T.Ü.)

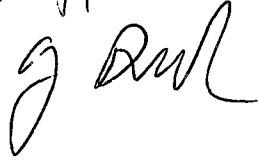
Assoc. Prof.Dr. Günhan DÜNDAR (B.Ü.)

DECEMBER 1999

**O(N) KARMAŞIKLIĞINDA
CMOS (K-)-KAZANAN-HEPSİNİ-ALIR DEVRELERİ**

**DOKTORA TEZİ
Y. Müh. Barbaros ŞEKERKIRAN
(492D032)**

**Tezin Enstitüye Verildiği Tarih : 31 Ağustos 1999
Tezin Savunulduğu Tarih : 29 Aralık 1999**

Tez Danışmanı : Prof.Dr. Uğur ÇİLİNGİROĞLU 
Diğer Jüri Üyeleri Prof.Dr. Duran LEBLEBİCİ (İ.T.Ü.) 
Prof.Dr. Hakan KUNTMAN (İ.T.Ü.) 
Prof.Dr. Atilla ATAMAN (Y.T.Ü.) 
Doç.Dr. Günhan DÜNDAR (B.Ü.) 

ARALIK 1999

PREFACE

I would like to express my gratitude for my supervisor Dr. U. Çilingiroğlu for his vision and support. I have been fortunate to work at The Scientific and Technical Research Council of Turkey where I found the opportunity to take part in realization of the circuits I designed during my Ph.D. work. That experience gave me a broad perspective stretching from system level design to testing passing through transistor level design layout drawing and CMOS processing. Without help from Dr. D. Leblebici this would never be possible. I am also indebted to Alcatel-Teletaş for providing the financial support to participate in and present my work at international conferences. I would like to express my thanks to Dr. İ. Çilesiz for her help in redaction. Without the support of my family my decades long education would not have been possible at all.

December, 1999

Barbaros Şekerkıran

CONTENTS

ABBREVIATIONS	v
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF SYMBOLS	ix
SUMMARY	x
ÖZET	xv
1. INTRODUCTION	1
1.1. Statement of Thesis	1
1.2. Motivation	1
1.3. Objectives	1
1.4. Overview	2
2. (K-)WINNER-TAKE-ALL CIRCUITS	3
2.1. Introduction	3
2.2. Distinction Between Selection Mechanisms	6
2.2.1. Activation-based WTA mechanism	7
2.2.2. Temporal WTA process	8
2.3. Synchronicity of the (K)WTA Process	8
2.3.1. Synchronous WTA process	8
2.3.2. Synchronous KWTA process	12
2.4. (K)WTA Circuit Implementation	13
2.4.1. Lazzaro WTA circuit	14
2.4.2. Urahama-Nagao implementation	15
2.4.2.1 WTA circuit	15
2.4.2.2 KWTA circuit	17
2.4.3. Choi-Sheu implementation	20
2.4.4. Çilingiroğlu implementation	21
2.4.5. Serrano-Barranco implementation	22
3. IMPROVEMENT ON CURRENT-MODE WTA CIRCUITS	24
3.1. Lazzaro WTA Circuit	24
3.1.1. Stability analysis	25
3.1.2. Limitations	28
3.1.2.1 Channel length modulation	28

3.1.2.2 Device mismatch	30
3.1.2.3 Gate area	31
3.1.2.4 Aspect ratio	31
3.1.2.5 Design curves	31
3.2. Proposed Modifications	33
3.2.1. Simple cascoding	34
3.2.2. Gain-boosted regulated-cascoding	36
3.2.3. Simulation results	38
4. THE PROPOSED WTA CIRCUIT	41
4.1. Circuit Description and Operation	41
4.2. Resolution	47
4.3. Dynamic Analysis	48
4.4. Implementation and Test Results	51
4.4.1. Offset voltages	53
4.4.2. Transient response	53
5. THE PROPOSED KWTA CIRCUIT	56
5.1. Circuit Description and Operation	56
5.2. Low Input Voltage Region	63
5.3. Dynamic Analysis	65
5.4. Implementation and Test Results	68
5.4.1. Offset voltages	68
5.4.2. Number of winners	70
5.4.3. Transient response	70
6. CONCLUSIONS	74
REFERENCES	76
APPENDIX	80
BIOGRAPHY	88

ABBREVIATIONS

AWTA	: Activation Based Winner-Take-All
CMOS	: Complementary Metal Oxide Semiconductor
DC	: Direct Current
DRAM	: Dynamic Random Access Memory
GBW	: Gain-Bandwidth Product
KWTA	: K-Winners-Take-All
MAXNET	: Maximum Selection Neural Network
MOS	: Metal Oxide Semiconductor
NMOS	: N-Channel Metal Oxide Semiconductor
PACNET	: Pair-Compared Neural Network
PMOS	: P-Channel Metal Oxide Semiconductor
TWTA	: Temporal Winner-Take-All
VLSI	: Very Large Scale Integration
WTA	: Winner-Take-All
YITAL	: Yarıiletken Teknolojileri Araştırma Laboratuvarı

LIST OF TABLES

	<u>Page No</u>
Table 3.1. The technology related parameters for a 0.6 μm technology and circuit design parameters.....	32
Table 4.1. Transistor dimensions	43
Table 4.2. Offset voltage distribution among cells, given in mV.....	53
Table 5.1. Transistor dimensions.....	58
Table 5.2. Offset voltage distribution among cells, given in mV	70



LIST OF FIGURES

	<u>Page No</u>
Figure 2.1 : The Lazzaro WTA circuit.....	15
Figure 2.2 : Circuit implementing Eq. (2.27) and (2.28).....	17
Figure 2.3 : Circuit implementing Eq. (2.48).....	19
Figure 2.4 : KWTa circuit with voltage inputs.....	19
Figure 2.5 : Schematic of Choi-Sheu WTA circuit.....	20
Figure 2.6 : Circuit diagram of the Çilingiroğlu WTA circuit.....	21
Figure 2.7 : Circuit diagram of the Serrano-Barranco WTA circuit.....	22
Figure 3.1 : Architecture of the Lazzaro WTA circuit.....	24
Figure 3.2 : a) Small-signal equivalent of the Lazzaro circuit, b) its block diagram, dashed line indicating the feedback connection	25
Figure 3.3 : Open-loop configuration of a cell, horizontal dashed line indicating the unity-gain feedback loop.....	29
Figure 3.4 : Design curves for M2.....	32
Figure 3.5 : Cascode voltage dependent current sources; a) simple (nonregulated), b) gain-boosted regulated.....	33
Figure 3.6 : Simple cascode WTA circuit.....	35
Figure 3.7 : Circuit diagram of gain-boosted regulated-cascode WTA circuit.....	37
Figure 3.8 : Simulated DC transfer curves of the 10-cell original Lazzaro WTA circuit. $IIN_i = 10 \mu A - i \times 50 nA$, $i = 2, 3, \dots, 10$	39
Figure 3.9 : Simulated DC transfer curves of the 10-cell simple-cascode, WTA circuit. $IIN_i = 10 \mu A - i \times 5 nA$, $i = 2, 3, \dots, 10$	39
Figure 3.10 : Simulated DC transfer curves of the 10-cell gain-boosted regulated-cascode, WTA circuit. $IIN_i = 10 \mu A - i \times 50 pA$, $i = 2, 3, \dots, 10$	40
Figure 4.1 : Architecture of the proposed WTA circuit.....	42
Figure 4.2 : The HSPICE simulated evaluation node voltages of 10-cell WTA circuit for initial evaluation node voltages a) $V_i = 5V + i \times 40mV$, $1 \leq i \leq 10$; b) $V_i = 2V + i \times 40mV$, $1 \leq i \leq 10$; a) $V_i = 0V + i \times 40mV$, $1 \leq i \leq 10$	44
Figure 4.3 : General description of the proposed circuit configuration.....	46
Figure 4.4 : Simulated evaluation node voltages of 10-cell WTA circuit for initial evaluation node voltages a) $V_i = 2V + i \times 10mV$, $1 \leq i \leq 10$; $C_{EVAL} = 400fF$; b) $C_{CSN} = 800fF$, $V_i = 2V + i \times 10mV$, $1 \leq i \leq 10$; $C_{EVAL} = 100fF$, $C_{CSN} = 800fF$. Oscillatory behavior of the waveforms verifies the importance of the ratio between C_{EVAL} and C_{CSN} as explained in Section 4.3.....	47
Figure 4.5 : Simplified circuit diagram of a cell.....	49
Figure 4.6 : Evaluation node voltages during charging of C_{CSN}	51

Figure 4.7	: Photomicrograph of the chip.....	52
Figure 4.8	: Experimental waveforms. The first and second traces belong to control signals ϕ_1 , ϕ_2 , respectively, the third and the forth traces show the outputs of the winning cell and its successor. The input voltage of the winning cell and its successor are 1.010V and 1V respectively. Other inputs are grounded.....	54
Figure 4.9	: Measured effect of input difference voltage on response time..	55
Figure 5.1	: Architecture of the proposed KWTa circuit.....	57
Figure 5.2	: General description of the cell configuration.....	58
Figure 5.3	: The HSPICE simulated evaluation node voltages of 10-cell KWTa circuit for initial evaluation node voltages a) $V_i = 5V + i \times 40mV$, $1 \leq i \leq 10$; b) $V_i = 2V + i \times 40mV$, $1 \leq i \leq 10$; a) $V_i = 0V + i \times 40mV$, $1 \leq i \leq 10$; $I_{TAIL}/I_{M3} = 2.5$	60
Figure 5.4	: The HSPICE simulated evaluation node voltages of 10-cell KWTa circuit for initial evaluation node voltages $V_i = 5V + i \times 40mV$, $1 \leq i \leq 10$; a) $I_{TAIL}/I_{M3} = 2.5$; b) $I_{TAIL}/I_{M3} = 4.5$; c) $I_{TAIL}/I_{M3} = 2.5$ $I_{TAIL}/I_{M3} = 8.5$	61
Figure 5.5	: General description of the proposed circuit configuration.....	62
Figure 5.6	: Simulated evaluation node voltages of 10-cell WTA circuit for initial node voltages a) $V_i = 2V + i \times 10mV$, $1 \leq i \leq 10$; $I_{TAIL}/I_{M3} = 2.5$, $C_{EVAL} = 400fF$, $C_{CSN} = 800fF$; b) $V_i = 2V + i \times 10mV$, $1 \leq i \leq 10$; $I_{TAIL}/I_{M3} = 2.5$, $C_{EVAL} = 100fF$, $C_{CSN} = 800fF$. Oscillatory behavior of the waveforms verifies the importance of the ratio between C_{EVAL} and C_{CSN} , as explained in Section 5.3.....	64
Figure 5.7	: Stability chart of 10-cell circuit as a function of k for three different I_{M5}/I_{M3} values (stable operating areas are indicated by arrows).. a) rising edge for $I_{M5}/I_{M3} = 0.25$; b) falling edge for $I_{M5}/I_{M3} = 0.25$ c) rising edge for $I_{M5}/I_{M3} = 0.5$; d) falling edge for $I_{M5}/I_{M3} = 0.5$ rising edge for $I_{M5}/I_{M3} = 0.75$; f) falling edge for $I_{M5}/I_{M3} = 0.75$	67
Figure 5.8	: Photomicrograph of the circuit	69
Figure 5.9	: Measured dependence of number of winners, k, on the ratio of the tail current to I_{M3}	71
Figure 5.10	: Experimental waveforms. a) The first, second and third traces belong to ϕ_1 , ϕ_2 and ϕ_3 , respectively; the fourth trace shows the buffered output voltage of cell-9. b) The configuration is same as (a) except for ϕ_3 coming before the end of the competition. $I_{TAIL} = 30\mu A$, $I_{M3} = 20\mu A$, $V_{IN(9)} = 1.3V$, $V_{IN(5)} = 1.250V$, all other inputs = 0V.....	72
Figure 5.11	: Measured effect of input difference voltage on response time for k=1.....	73
Figure 5.12	: Measured effect of tail current on response time for k=1.....	73

LIST OF SYMBOLS

A_{β}^2	: transconductance standard deviation constant
A_{VTH}^2	: threshold voltage standard deviation constant
β	: device transconductance parameter
C	: capacitance
GND	: ground
gm	: source transconductance
gmb	: bulk transconductance
gds	: drain-to-source conductance
I	: current
Id_{SAT}	: drain saturation current
k	: number of winners in a KWTa circuit
k_p	: process transconductance parameter
L	: gate length of a MOS transistor
L_{eff}	: effective geometrical gate length
L_{sr}	: slew rate limit
n	: total cell number in a (K)WTA circuit
O	: complexity
P	: probability of occurrence
p	: pole frequency
R	: resistance
Rds	: drain-to-source resistance
T	: connection weight between cells
t	: time
τ	: time constant
V	: voltage
V_c	: common voltage
Vd_{SAT}	: drain saturation voltage
VDD	: supply voltage
VG	: gate voltage
V_{TH}	: threshold voltage
W	: gate width of a MOS transistor
x, u	: activation

CMOS (K-)WINNER(S)-TAKE-ALL CIRCUITS OF $O(N)$ COMPLEXITY

SUMMARY

The demand for high performance winner-take-all (WTA) circuits has become more pronounced because of their promise for analog applications. These circuits are used to select the one with the highest magnitude among elements of an input vector. K-winners-take-all circuit (KWTA), on the other hand, selects the most activated k cells, among n total cells and can be considered as a generalized version of WTA circuits. (K)WTA circuits are among key components of most neural network architectures which are increasingly being used for solving complex engineering classification, control and optimization problems. A variety of signal processing circuits using WTA function, especially in filtering sections have also been reported. WTA circuits of $O(N^2)$ complexity requires an individual interaction between each cell pair. As n (input count) increases implementation becomes more difficult due to large number of interconnections. In these situations, architectures of $O(N)$ complexity become advantageous. The most widely used WTA circuit is the Lazzaro circuit shown in Fig. 1.

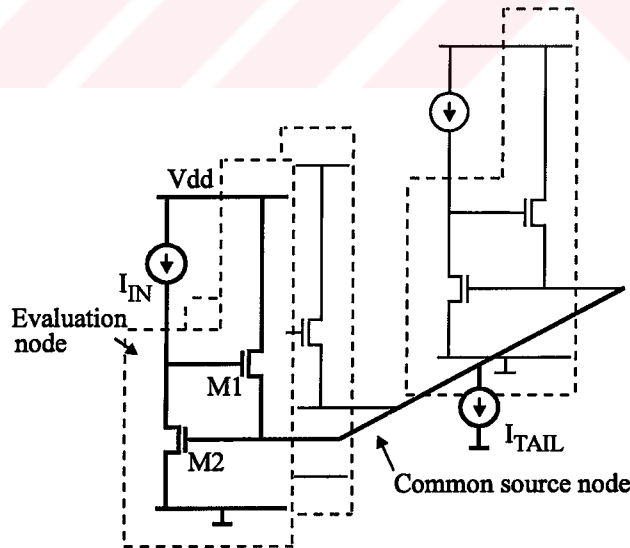


Figure 1. Lazzaro WTA circuit.

With either of the two proposed modifications in which the transistor M2 is replaced by a cascode current or a gain-booster-regulated-cascode current source, the

resolution of this circuit can significantly be increased. The modified circuits are shown in Fig. 2 and Fig. 3, respectively.

The resolution enhancement is related to the small-signal output resistance increase and is $\frac{gm_{M3}}{gds_{M3}}$ fold for the simple cascode circuit while $\frac{gm_{M3}}{gds_{M3}} \frac{gm_{M4}}{(gds_{M4} + g_{oBIAS})}$ fold for gain-booster-regulated-cascode current source case.

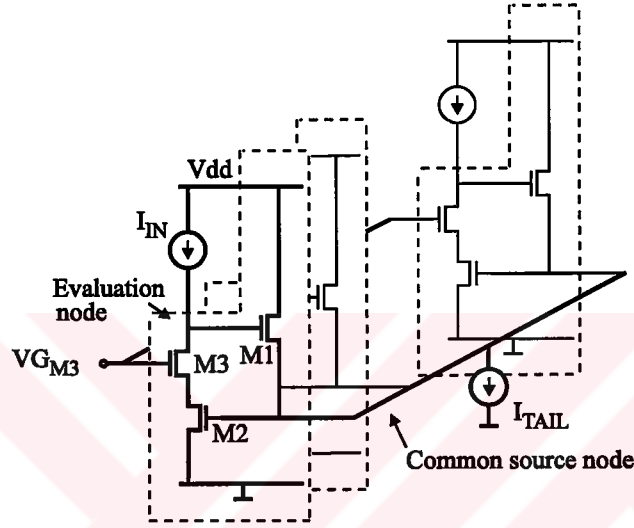


Figure 2. Simple cascode Lazzaro WTA circuit

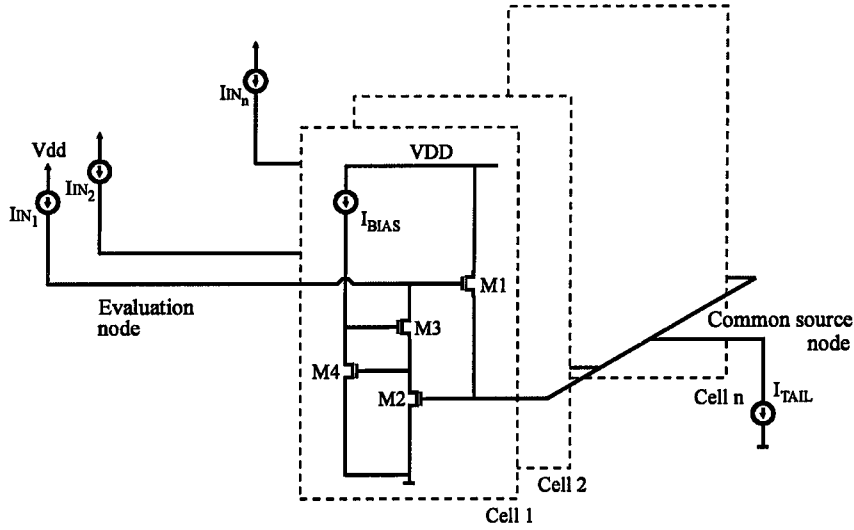


Figure 3. Gain-booster regulated cascode WTA circuit.

Positive feedback is an effective tool to increase discrimination capability of WTA and KWTa circuits. However, it causes the circuit to become stuck in a stable state after the competition phase. Before a new selection, the circuit should be initialized by suppressing the positive feedback in order to prevent the cells corresponding to the winners of the previous phase to become privileged. In this dissertation, two new circuit topologies of $O(N)$ complexity have been proposed. Both are based on positive-feedback approach. The first one is a WTA circuit shown in Fig. 4.

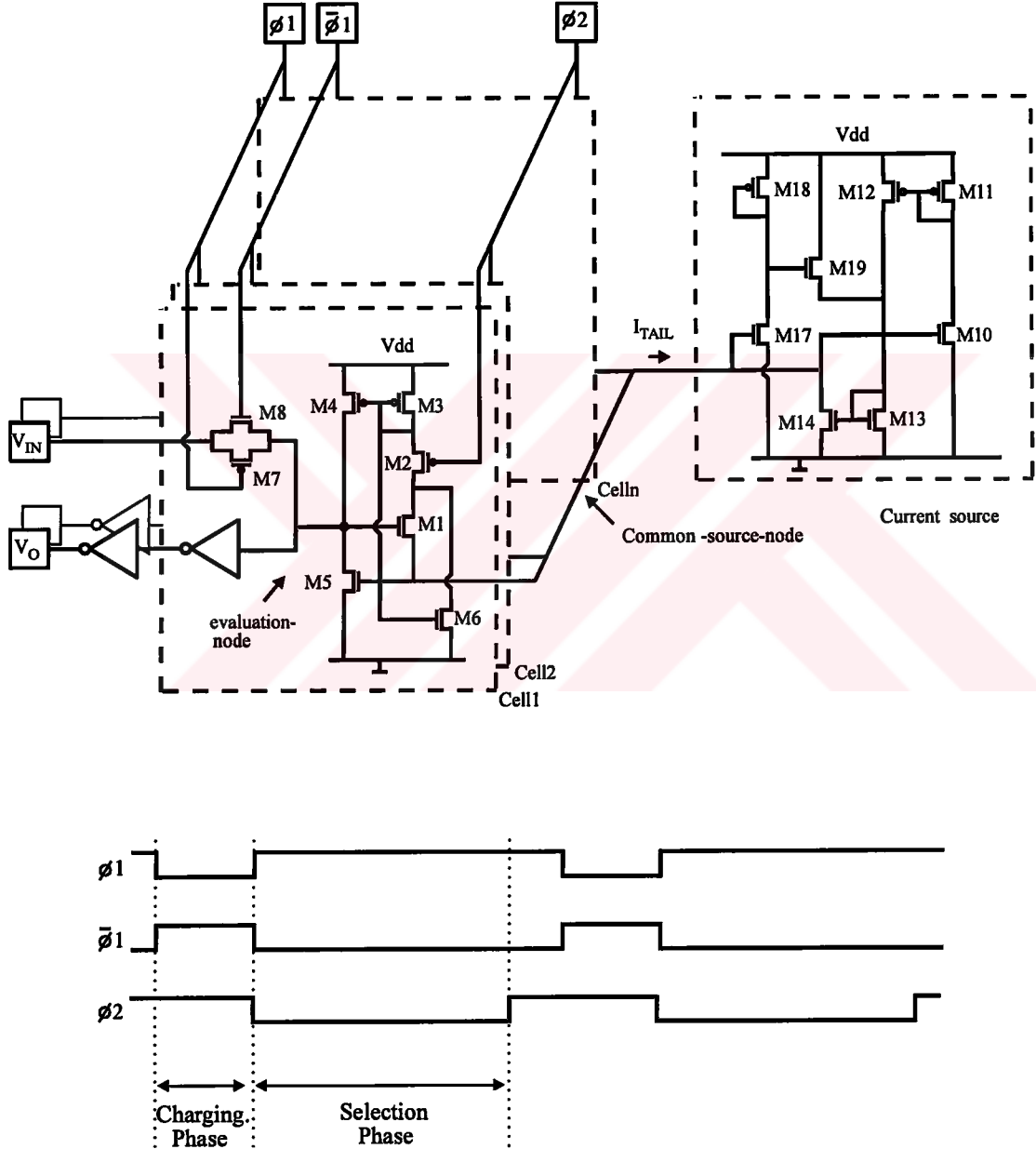


Figure 4. The proposed WTA circuit of $O(N)$ complexity.

The proposed circuit automatically limits the amount of positive feedback applied, keeping the ratio, r , of the tail current to the discharging current, constant. If this ratio is less than 1, the circuit cannot select any winner, on the other hand if it is greater than 2, the tail current becomes high enough to supply excessive charging current to the cells following the winner, leading the circuit to select more than one winner. Then, 1.5 seems to be the most appropriate value for this ratio.

The second proposed topology is for a KWTa circuit shown in Fig. 5.

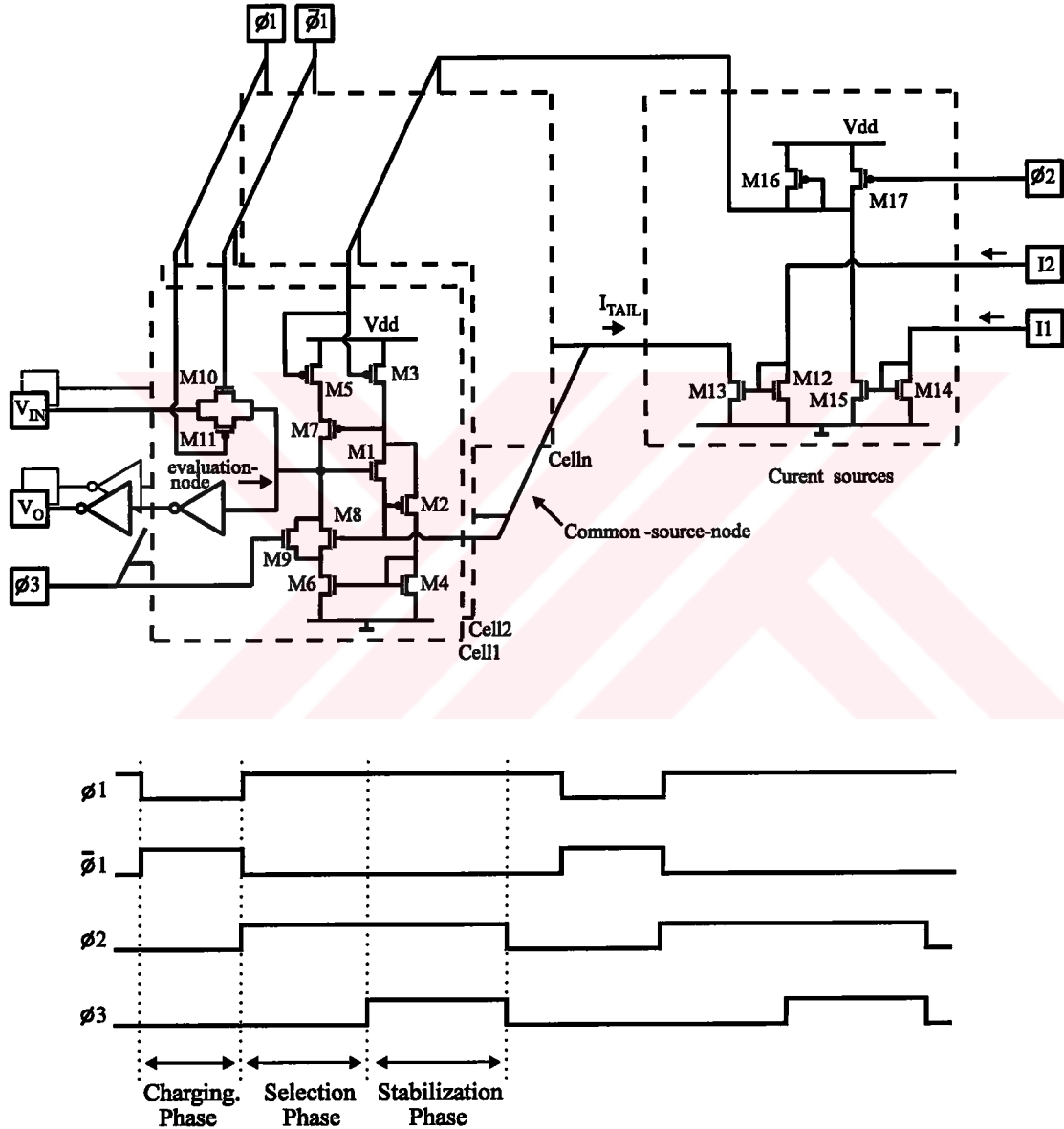


Figure 5. The proposed KWTa circuit of $O(N)$ complexity.

When the drain current of M13 is set to $I_{M13} = (k + r) I_{M3}$ (where k is an integer and r is a real number between zero and one) the proposed circuit ensures

selection of the highly activated k cells (k evaluation nodes having highest initial voltages).

Both circuits have been implemented in a $3\mu\text{m}$, n-well, single-polysilicon, single-metal, CMOS technology. The stability issue is examined and the bounds for a stable operation are determined. The functionality of the circuits has been verified based on measurement results. Test results also confirmed that the circuits achieve a high speed and a very good resolution along with a simple structure which make them highly suitable for neural network and analog VLSI signal processing implementations.

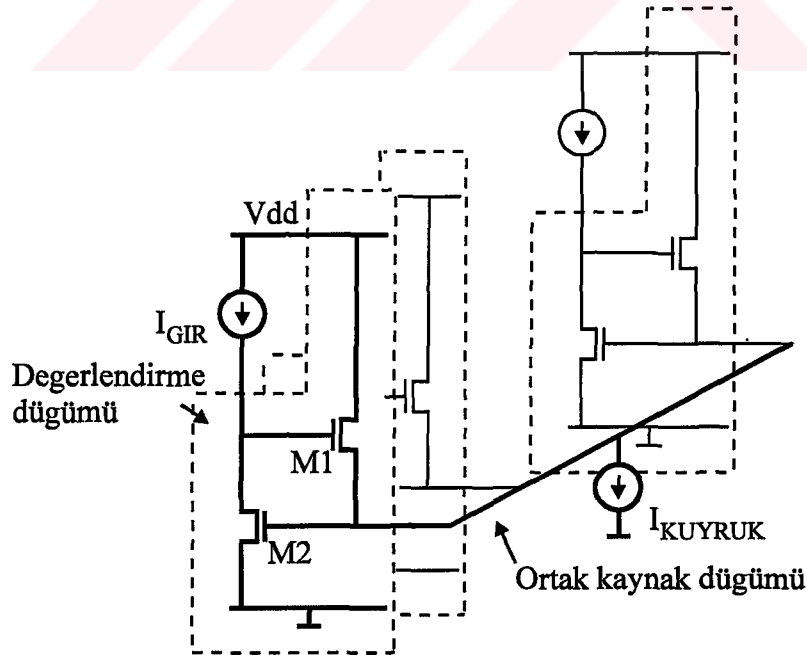


O(N) KARMAŞIKLIĞINDA

CMOS (K-) KAZANAN-HEPSİNİ-ALIR DEVRELERİ

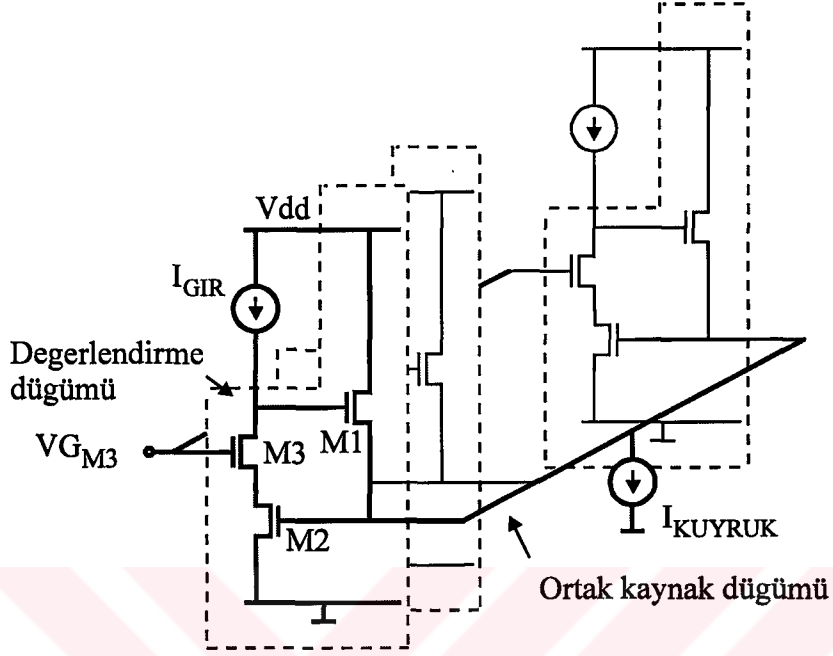
ÖZET

Yüksek performanslı Kazanan-Hepsini-Alır (KHA) devrelerine duyulan gereksinim, analog uygulamalar açısından sundukları avantaj dolayısıyla önem kazanmıştır. Bu bloklar, uygulanan giriş vektörünün en yüksek genlikli elemanını seçerler. Diğer yandan K-Kazanan-Hepsini-Alır (KKHA) devreleri toplam n adet hücreden en yüksek genlikli işaret uygulanan k adet hücreyi seçer ve aslında KHA devrelerinin daha genel hali olarak görülebilirler. (K)KHA devreleri, karmaşık sınıflama, kontrol ve optimizasyon problemlerini çözmekte gittikçe daha fazla kullanılan yapay sinir ağları mimarilerinin vazgeçilmez bileşenlerindendir. Özellikle filtre bölümlerinde KHA işlemini kullanan birçok işaret işleme devresi tasarlanmıştır. $O(N^2)$ karmaşıklığındaki devrelerde her hücre karşılıklı etkileşim içinde bulunmalıdır. Giriş sayısı arttıkça, bağlantı sayısının artması nedeniyle bu mimarilerin gerçekleştirilmelerinde zorluklar ortaya çıkar. Bu durumlarda $O(N)$ karmaşıklığındaki devreler daha avantajlı olur. KHA devrelerinden en sık kullanılanı Şek. 1’de görülen Lazzaro devresidir.



Şekil 1. Lazzaro KHA devresi.

Bu devrenin ayırma kabiliyeti M2 transistörünün yerine basit veya regüle edilmiş kaskod akım kaynağı yerleştirilerek artırılmıştır. Bu şekilde değiştirilmiş devreler Şek. 2 ve Şek. 3’de görülmektedir.



Şekil 2. Basit kaskod Lazzaro devresi

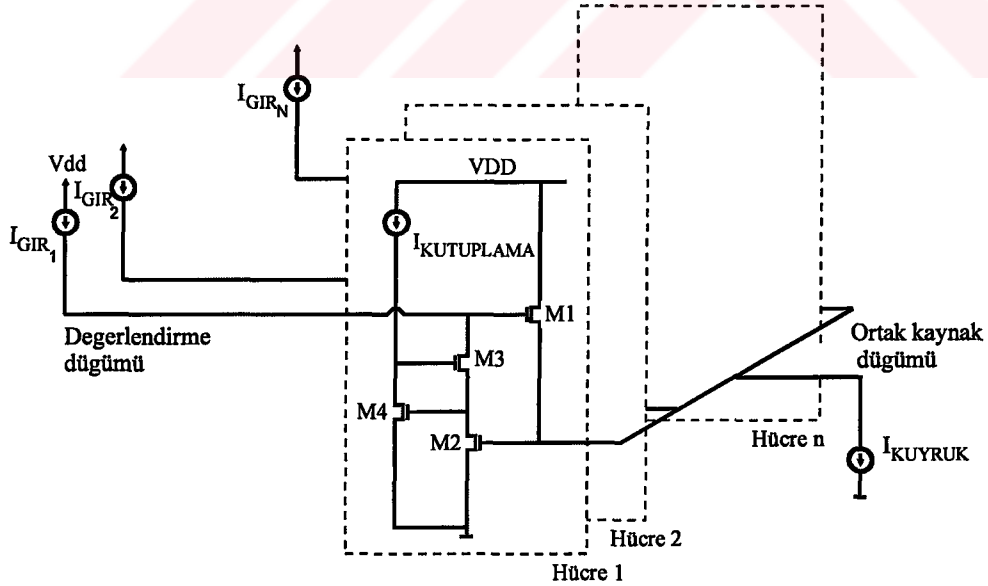
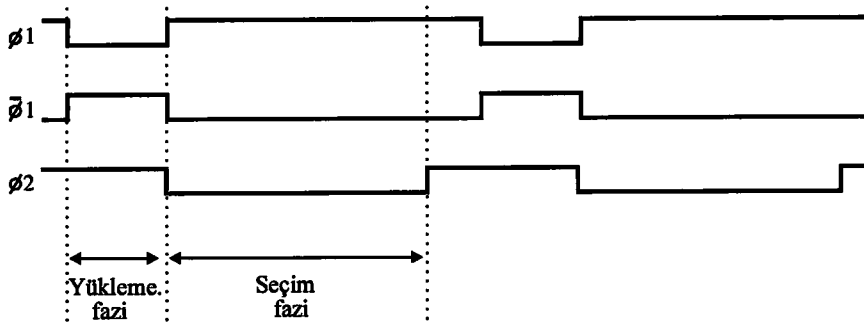
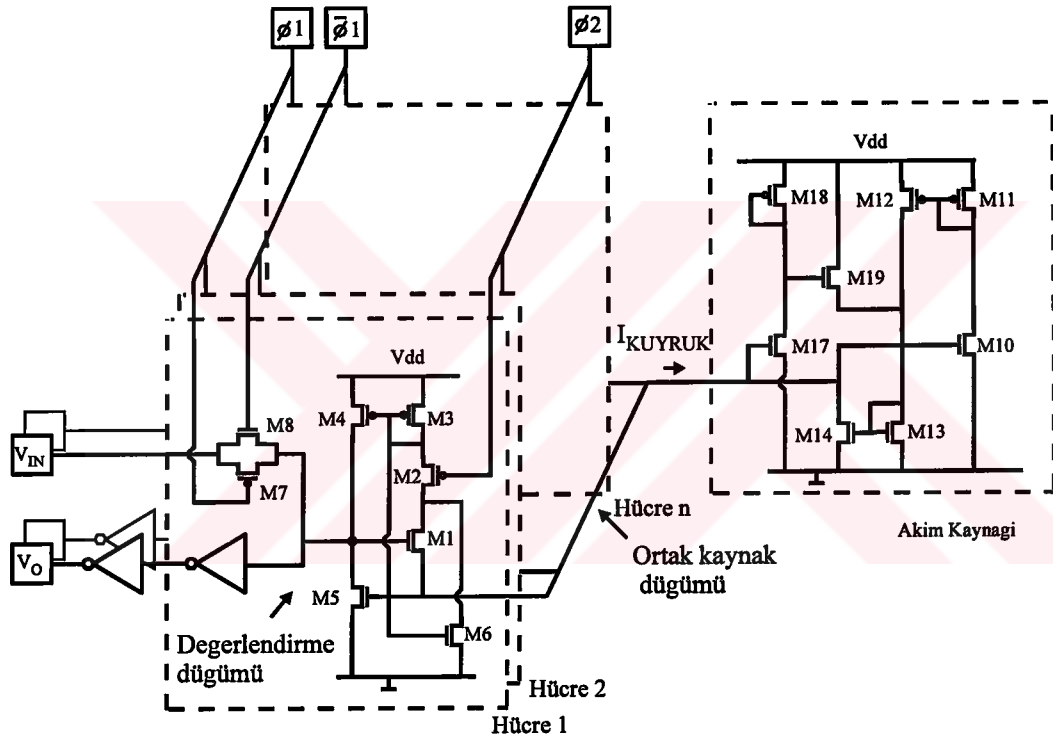


Figure 3. Regüle edilmiş kaskod Lazzaro KHA devresi.

Devrenin ayırma kapasitesindeki artış akım kaynağının küçük işaret çıkış direncindeki artmaya bağlıdır ve ilk devre için $\frac{g_{mM3}}{g_{dsM3}}$ kat, ikinci devre için ise

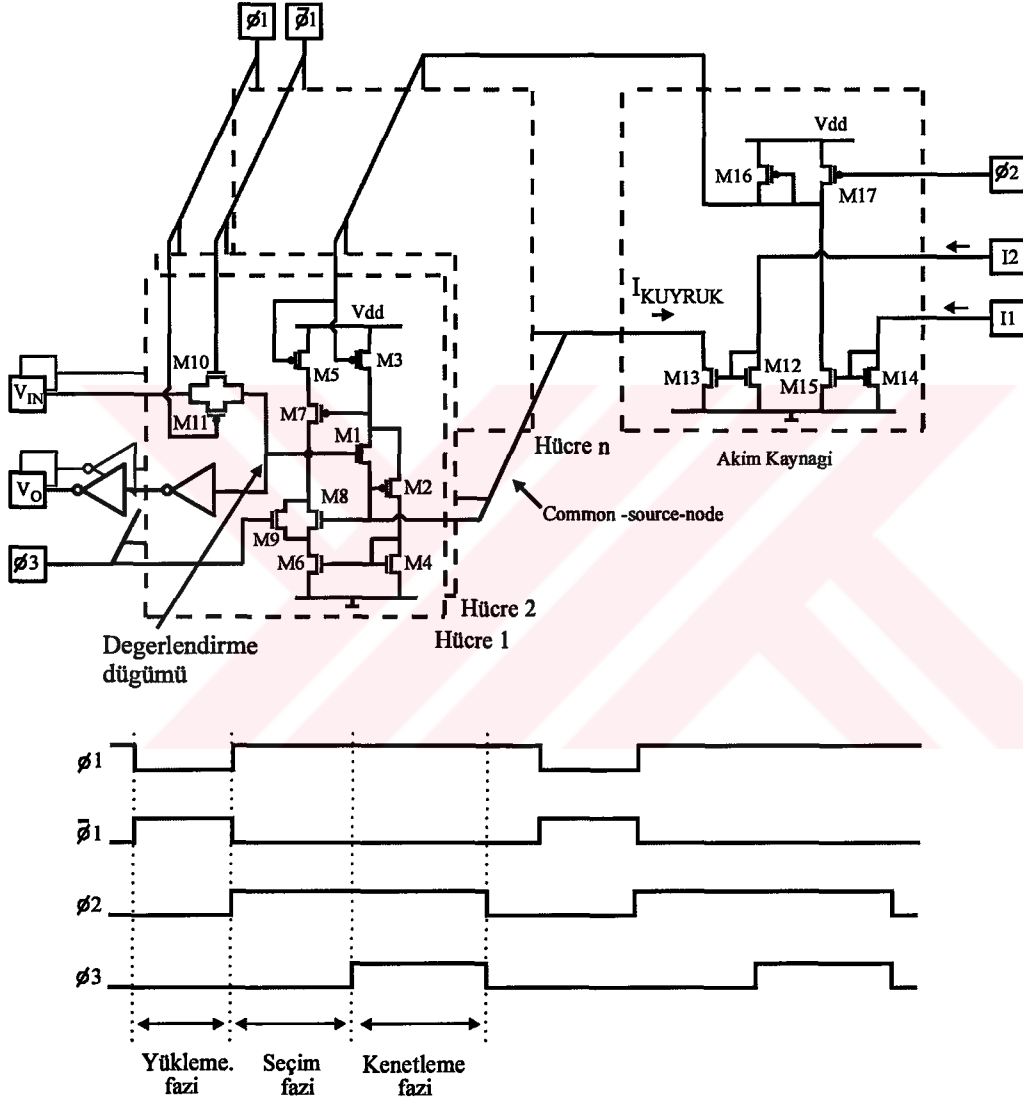
$$\frac{g_{mM3}}{g_{dsM3}} \frac{g_{mM4}}{(g_{dsM4} + g_{oBIAS})} \text{ kattır.}$$

Pozitif geribeleme (K)KHA devrelerinin ayırma kabiliyetini arttırmak için kullanılabilecek etkili bir yöntemdir. Fakat devre her seçme işleminden sonra bir durumda takılı kalır. Yeni bir seçme işleminden önce, bir önceki kazananın öncelikli olmaması için bir eşitleme işlemi yapılmalıdır. Bu tezde, O(N) karmaşıklığında iki adet pozitif geribeslemeli devre önerilmektedir. Bunlardan ilki Şek. 4’de görülen KHA devresidir.



Şekil 4. Önerilen KHA devresi.

Önerilen devre, uygulanan pozitif beslemenin miktarını kuyruk akımı ile boşaltma akımı arasındaki oranı sabit tutarak otomatik olarak ayarlamaktadır. Bu oran birden küçükse devre kazanan hücreyi seçemezken, bu oranın ikiden büyük olması halinde kazanan hücre sayısı birden çok olabilir. Bunun nedeni, kuyruk akımının kazanan hücreyi takip eden hücrelere kazanan olmaları için yetecek doldurma akımını sağlayacak kadar yüksek olmasıdır. Bu nedenle, söz konusu oran için en uygun değer 1.5 dir. Önerilen ikinci pozitif geribeslemeli yapı ise Şek. 5’de görülen KKHA devresidir.



Şekil 5. Önerilen KKHA devresi.

M13 transistörünün savak akımı $I_{M13} = (k + r) I_{M3}$ eşitliğini sağlayacak şekilde ayarlandığında (burada k bir tamsayı, r ise birden küçük sıfırdan büyük bir sayıdır), önerilen devre en yüksek giriş gerilimine sahip k adet hücreyi seçer.

Her iki devre de n-kuyulu , tek metal ve tek polisilisyum katmanlı 3µm CMOS teknolojisinde gereklenmiřtir. Devrelerin kararlılıęı incelenmiř ve devrelerin kararlı alıřabilmesi iin tasarımıñ hangi sınırlar iinde kalınarak yapılması gerektięi belirtilmiřtir. Devrelerin iřlevsellięi lüm sonularına dayanılarak gsterilmiřtir. lüm sonuları, nerilen yapıların yksek hız ve yksek ayırma kabiliyeti gereken uygulamalarda kullanılabileceęini gstermiřtir. Bu sonuca basit devre yapılarıyla eriřilmiř olması devreleri yapay sinir aęları ve iřaret iřleme uygulamaları iin daha da uygun kılmaktadır.



1. INTRODUCTION

1.1 Statement of Thesis

The intention of this dissertation is to present new implementations of Winner-Take-All (WTA) and K-Winners-Take-All (KWTa) processes and to propose modifications on Lazzaro WTA circuit in order to enhance its performance.

1.2 Motivation

A large volume of work has been published on theoretical analysis of WTA process. Surprisingly however few realized WTA circuits have been reported. In theoretical works, WTA circuit is always defined based on idealistic mathematical formulas and the impact of microelectronic element characteristics such as the MOS transistor was always disregarded. As a consequence, an increasingly enlarging gap appeared between theoretical and practical works. So, realization of new WTA circuits, is of crucial importance. It is very hard for a circuit designer to select the topology of the circuit and to size the components based on existing neural network models. The other important issue is the requirement for the architectures capable of processing large number of inputs simultaneously. Individual interaction between each cell pair is required in WTA circuits of $O(N^2)$ complexity . As n (input count) increases, implementation becomes more difficult due to large number of interconnections. Hence within the scope of the dissertation only architectures of $O(N)$ complexity are discussed.

1.3 Objectives

The objectives of this work are

- ☐ to find out new more effective WTA circuit topologies,
- ☐ to analyze their stability,

- to determine the design parameters,
- to present reliable models for the investigated circuits.

1.4 Overview

Following this introduction, this dissertation is divided into five chapters. They are not presented in the order in which the research was performed, but in the order of functional and operational complexity.

The definition of WTA and KWTa processes, classification of WTA circuit types, brief presentation of mostly used previous circuits form the contents of the first chapter. In the second chapter, Lazzaro WTA circuit is analyzed and two modifications to enhance its resolution are presented. In the succeeding two chapters, a WTA and a KWTa circuit with $O(N)$ complexity are presented, respectively. Concluding remarks are drawn in the last chapter.

2. (K-) WINNER-TAKE-ALL CIRCUITS

2.1 Introduction

The need for selection arises in many electronic systems. A class of circuits called comparators has traditionally and extensively been used for determining the higher between two voltages or currents. With the advent of more complex electronic systems, the need for networks capable of making comparison among larger number of inputs became more pronounced. Hence a new class of circuit called Winner-Take-All (WTA) has been introduced. K-Winners-Take-All (KWTa) circuit, on the other hand, selects the most activated k cells among n total cells and can be considered as a generalized version of WTA circuits. Variety of tasks such as, pattern recognition and classification, speech recognition and processing, reasoning and learning (in neural networks) require such a subsystem.

The WTA function has been used in various signal-processing applications, including image feature extraction [1], Hamming network [2] and silicon binaural hearing [3]. In [4], one neuron with the maximum excitation is selected for each pixel by the WTA circuit in the array of velocity-sensitive neurons for image processing applications. In [5], the WTA circuits selects the winner which has the minimum distance between the input vector and the stored codebook vectors for video compression.

In many VLSI neural network chips, an efficient learning mechanism is necessary. A proper choice of the learning scheme is of crucial importance. There are two basic types of learning methods [6,7]: the supervised and the unsupervised learning. Supervised learning such as the backpropagation learning has been proven to be effective in many applications. However, in the case where the desired outputs are unknown in advance, the unsupervised learning method is more appropriate.

There are several types of unsupervised learning schemes [8]: the Hebbian learning, the competitive learning and the Boltzman learning. Among architectures suggested for artificial neural network implementations, the self-organizing network using competitive learning has the desirable property of effectively producing spatially organized representation of various features of the input signals [9-13]. In

competitive learning as the name implies, the output neurons of a neural network compete among themselves to become active. Whereas in neural networks based on Hebbian learning several output neurons may be active simultaneously, in competitive learning only a single output neuron is active at any given time. It is this feature that makes competitive learning highly suited to discover statistically salient features that may be used to classify a set of input patterns. In the competitive layer of a self-organized neural network the WTA operation is executed to select the winner.

Among previously proposed CMOS WTA circuits [14-24], the most widely used one is the Lazzaro circuit [14]. It is a current-mode circuit containing only inhibitory feedback loops. Circuits without positive feedback suffer from systematic resolution limitations due to the finite open loop gain of their cells stages. They are unable to select the winner if the differences between inputs of their most excited cells are small. Although performance can be enhanced by cascading identical stages [15], excessive real estate consumption and slower response remain as outstanding problems. A modified version of Lazzaro's basic circuit containing both excitatory and inhibitory feedback loops has been presented [16]. In general, positive feedback increases the resolution, but an excessive amount may cause oscillation or multiple winner selection. Recently, further improvement on this circuit has been reported, where by replacing the independent current source by a voltage dependent one, a constant ratio is maintained between the positive and the negative feedback [17]. This configuration ensures identification of the cell having maximum initial excitation even if differential input voltages are infinitesimal. Thus only random resolution limitation, which stems from device mismatch and noise remain as error sources which can be minimized by proper selection of circuit parameters.

Besides conventional areas where the WTA process is used frequently, KWTa process offers exciting implementation possibilities in median filtering and genetic algorithms applications. In basic median filtering, each point is replaced by the median data of its neighboring points. This is an effective way of achieving noise reduction while preserving edges [25], [26]. Selection of the median data can be performed with a KWTa circuit by determining $n/2$ winner out of n input vector then selecting $n/2+1$ winner and identifying the vector among the first loser group which becomes winner in the second step. The other potential application domain of k -winners-take-all circuits is implementation of genetic algorithms. A genetic algorithm alters a population whose members are generated initially to be solutions to an optimization or control problem so that the characteristics of members that define fitter traits to the solution are retained while those that reduce the fitness are

rejected [27]. To do this, it selects a group of members exhibiting high fitness to be parents of new members that form a new population. Since the offspring are outcome of breeding processes that take place among selected individuals, the overall fitness of the new population is expected to be higher. The selection and breeding processes are then repeated until a desired level of fitness is achieved. If a set of input vectors representing the fitness level of members is applied, KWTA circuit can perform the parent selection process.

A WTA network should stay invariant with respect to all permutation of indexes. In other words, there should be no privileged cells. Hence the following lemma holds.

Lemma: Order Preserving Dynamics: If in a WTA the activation of any two cells c and d is,

$$x_c(t_0) > x_d(t_0) \quad (2.1)$$

then for all finite times $t_2 > t_0$

$$x_c(t_2) > x_d(t_2) \quad (2.2)$$

Thus, the ordering of the cells by the magnitude of their respective states remains unchanged throughout the selection process.

Proof: Assume that $x_c(t_0) > x_d(t_0)$ and $x_c(t_2) < x_d(t_2)$, then by the continuity of states there exists $t_0 < t_1 < t_2$ such that $x_c(t_1) \equiv x_d(t_1)$ which contradicts uniqueness of solutions and interchangeability of cells. As this consideration applies to all pair of cells, the ordering of the cells at $t=t_0$ is preserved for $t > t_0$.

More precisely, the purpose of conventional WTA network is to select a unit with the highest activation strength from a set of contending units. Many WTA mechanisms have been suggested for the purpose of selection. Most of these mechanisms use lateral inhibition, where each unit inhibits every other unit. The amount of inhibition applied by the unit is proportional to its activation level. In this situation, the unit with the highest initial activation will inhibit all the other units more strongly than it is inhibited by any other unit. As a result, it is inhibited less than the others are. The way the inhibition from different units is combined to affect a given unit varies significantly from system to system, though addition of the individual inhibitions is usually part of the scheme. This process will eventually

result in the unit with the highest initial activation obtaining an activation value, which is significantly larger than the rest. This unit is then declared a winner and is selected. Yet, if differences between most activated units are small, then the activation level of the winner may be low and close to the activation level of the losers. To overcome this problem two approaches can be used. The first and the simplest one is to increase the activation of all cells through a source which is identical for each cell. The other solution is to relate the activation, which is constant in the previous approach, to the activation level of the corresponding cell. This is called positive feedback. In general, positive feedback increases the resolution but an excessive amount may cause oscillation or multiple winner selection. It also causes the circuit to become stuck in a stable state after the competition phase. Before a new selection, the circuit should be initialized by suppressing the positive feedback in order to prevent the cells corresponding to the winner of the previous phase to become privileged.

2.2 Distinction Between Selection Mechanisms

Now let's us make a clear distinction between different selection mechanisms and selection criteria and comment on the relationship between mechanism and criteria. Various important criteria for selection include the following:

Highest Activation Strength: This is the criterion used in traditional WTA networks as described in the previous section. Typically, the contending units represent alternative decisions and the strength of activation of a unit encodes the degree of confidence with which the decision is asserted.

Arbitrary choice: Here the system should arbitrarily choose a unit. One application of this criterion is in unit recruitment, where the system has to choose a unit from a pool of free units, and it does not matter which is picked up. The need for unit recruitment arises in any connectionist system in which the operative part of the network is required to grow by mechanisms implemented in the network itself. It is used in some knowledge intensive learning systems.

Earliness: In situations where several neural modules are competing to produce a decision, it may be appropriate to choose the decision arriving first.

2.2.1 Activation-based WTA mechanism

Activation based WTA (AWTA) circuits can be classified as current, voltage and charge mode circuits depending on the type of the physical quantity applied to the input. In microelectronics, making comparison between currents is easier than comparing voltages or charges. The most straightforward way to carry out this process is to source and sink the currents to be compared through a simple node. If the magnitude of the sourced current is higher, then the voltage of the node goes towards the positive rail, else this voltage decreases down to the ground level. Hence, in many voltage or charge based circuits the input physical quantity is converted to current and the comparison process is executed.

Existing AWTA mechanisms are of course most naturally suited to the highest activation strength criterion. However, they could in principle also be applied to the other two. For arbitrary choice, the units in the choice set can be made to generate random activation levels. This is exactly the way unit recruitment is performed in [28].

The differences among the arrival times can be converted into activation differences, again allowing an AWTA mechanism to be applied. The conversion could be done, for instance, by dumping equal initial activation into each unit and then decreasing these activations of each cell gradually until the termination signal pertaining to this unit arrives. The easiest way to do this is to discharge the capacitors through a constant current source during a time span determined by the signal arrival time. The earliest the signal arrives the less the charge sunk to the ground and consequently the highest the voltage level. This conversion could also be done by feeding each module's output line into a unit whose output rises gradually to the level of its input. One simple type of unit with this property obeys

$$\tau \frac{dy}{dt} + y = x \quad (2.3)$$

where x is the input value, y is the output, and τ is a time constant. Such units are used, for instance, in a network described in [29]. Some extra circuitry is needed to sample these unit's outputs at some time sufficiently soon after the start of the earliest termination signal. The sampled values will then differ in a way that reflects the differences between the times of onset of the corresponding termination signal, and be fed into a AWTA network.

2.2.2 Temporal WTA process

The temporal WTA mechanism is time based. The selection is done by an arbiter subnetwork, connected to a fixed set of possibly contending units [30]. The arbiter subnetwork makes the choice based on differences in the times of arrival of ‘ready’ signals sent to it by the units that are actually contending on the current occasion. These differences are variations both across units on a given occasion of contention and across different occasions for a given unit.

In outline, the arbiter works as follows. It tries to identify the unit sending the earliest arriving ready signal, so that it can send an ‘individual winner’ signal to just that unit. It performs this by using each ready signal to block the arrival of all others, so that only the earliest one survives. However, if there is a sufficiently close tie for earliest arrival, nonearliest signals may scrape through before the blocking has had time to take effect. Therefore the arbiter causes the contenders, that sent the closely tying ready signals, to enter a new bound to contention. If a unique selection again cannot be made, a further round of contention is initiated, and so on. It is crucial to bear in mind in what follows that the set of actually contending units is typically reduced from one round to the next.

2.3 Synchronicity of the (K)WTA Process

(K)WTA networks can also be categorized based on the initialization of the competition phase. The first type is attained by continuously active networks, whose competitive processes are asynchronously converged until the dynamic states are stabilized. The second category, which is performed by iterative procedures, is treated as an iterative updating (K)WTA process controlled by a synchronous system clock. The input physical quantity is sampled at predetermined times and the selection process is executed based on these levels. So we will call this type of circuits as synchronous (K)WTA circuits.

2.3.1 Synchronous WTA process

The synchronous WTA competition process can be modeled based on the behavior defined as [31,32];

$$\frac{dx_i(t)}{dt} = x_i(t) + a_{ij}y_j(t) + \kappa_i \quad (2.4a)$$

$$y_i(t) = f(x_i(t)) \quad (2.4b)$$

where the feedback coefficients between i'th and j'th cell, a_{ij} , and the cell bias values κ_i are constant real parameters, and i and j indicates the cell number (1,2,...n) in the n-cell system. The cell output function f is defined by,

$$f(x_i) = \begin{cases} -1, & \text{if } x_i \leq -1 \\ x_i, & \text{if } |x_i| < 1 \\ 1, & \text{if } x_i \geq 1 \end{cases} \quad (2.5)$$

On the other hand, we assume the following parameters,

$$a_{ij} = \begin{cases} \alpha, & \text{if } i \neq j \\ \beta, & \text{if } i = j \end{cases} \quad (2.6a)$$

$$\kappa_i = \kappa \quad (2.6b)$$

These are obviously invariant with respect to all permutations of cell indexes. Also, WTA networks are characterized by positive self-feedback and negative inter-cell coupling. It can therefore be expected that $\alpha < 0$, $\beta > 0$. A WTA network with the above feedback operator, is fully interconnected, the number of interconnections being the square of the number of cells ($O(N^2)$ complexity). It is possible to simplify the network to $O(N)$ complexity. For this purpose, it will be convenient to decompose

$$\beta = \alpha + \delta + 1 \quad (2.7)$$

which together with (2.4a) and (2.6a) gives a cell equation that can be rewritten as.

$$\frac{dx_i(t)}{dt} = -x_i(t) + \alpha \sum_{j=1}^n y_j(t) + (\delta + 1)y_i + \kappa \quad (2.8)$$

The very same sum occurs in the dynamic equation of each cell. This simplification can be made explicit in the way the system is defined. Here, the feedback operator has the form

$$a_{ij} = \begin{cases} \alpha, & \text{if } i \neq j \\ \alpha + \delta + 1, & \text{if } i = j \end{cases} \quad (2.9)$$

A single global adder is used to compute the sum

$$\sigma = \sum_{j=1}^n y_j \quad (2.10)$$

The first step in the design process is the classification of output patterns into stable and unstable ones [33]. Due to symmetry between cells, all patterns with the same number of winners are equivalent. In other words, the ordering of the cells by the magnitude of their respective states remains unchanged throughout the selection process. In n -cell network $n+1$ patterns need to be considered. P_k represents all patterns with k winning cells. So the possible values for k are $0, 1, 2, \dots, n$ ($n+1$ patterns). Only P_1 is considered to be stable while all other patterns must be unstable.

$$P_k = \begin{cases} \text{unstable} & \text{if } k \neq 1 \\ \text{stable} & \text{if } k = 1 \end{cases} \quad (2.11)$$

Now let's examine the design inequalities associated with different cases. The pattern P_0 is completely symmetric. Hence there is effectively no choice for the cell on which the instability condition is centered. The only possible instability requirement for P_0 (where $\sigma = -n$) is

$$\frac{dx_i(t)}{dt} = +1 + \alpha(-n) - (\delta + 1) + \kappa > 0 \quad (2.12)$$

which is equivalent to

$$\kappa - \delta > n\alpha \quad (2.13)$$

The design inequalities for the stable pattern P_1 (where $\sigma = 2-n$) must be derived for every cell. Because of symmetry, only two different cases exist, corresponding to whether the inequality is derived from an eventually winning or losing cell.

$$\frac{dx_i(t)}{dt} = +1 + \alpha(-n+2) - (\delta + 1) + \kappa > 0 \quad (2.14)$$

and

$$\frac{dx_i(t)}{dt} = -1 + \alpha(-n+2) - (\delta+1) + \kappa > 0 \quad (2.15)$$

The simplified forms of these two inequalities are

$$\kappa - \delta < (n-2)\alpha \quad (2.16)$$

$$\kappa + \delta > (n-2)\alpha \quad (2.17)$$

Finally, all patterns P_j (where $\sigma=2j-n$) with $j>1$ must be unstable. Although in principle there is still a choice of centering the instability condition on a winning or losing cell, WTA behavior here requires the number of winning cells to be reduced (as eventually only one is allowed) and therefore only the choice of a winning cell makes sense. Thus the instability condition as derived from any winning cell (all of which are again equivalent because of symmetry) is

$$\frac{dx_i(t)}{dt} = -1 + \alpha(-n+2j) - (\delta+1) + \kappa < 0 \quad 2 \leq j \leq n \quad (2.18)$$

and equivalently,

$$\kappa + \delta < (n-2j)\alpha \quad 2 \leq j \leq n \quad (2.19)$$

In particular, for $j=2$ this reduces to

$$\kappa + \delta < (n-4)\alpha \quad (2.20)$$

Of all inequalities only the last one is adequate. In proof, first add (2.13) to (2.16) multiplied by -1 . From the resulting inequality $0 > 2\alpha$ it is seen that α should be negative. Now, if $j_1 < j_2$ then

$$(n-2j_1)\alpha > (n-2j_2)\alpha \quad (2.21)$$

and if (2.19) holds for $j=2$, it holds for every larger j as well. Therefore, (2.20) is sufficient to describe the stability of the system.

2.3.2 Synchronous KWTa process

The continuous Hopfield model is fully characterized by a system of first order differential equations which governs the evaluation of the state of the network ($i=1,2,...N$) [34].

$$C \frac{dx_i}{dt} = -\lambda_i x_i + \sum_{j=1}^N a_{ij} g(x_j) - \kappa_i \quad (2.22)$$

where $C>0$ and

$$\lambda_i = \sum_j |a_{ij}| \quad (2.23)$$

The sigmoid function $g(x)$ is defined by: $g(x) = f(G, x)$, where $G > 0$ is the gain of the sigmoid, and $f(x)$ is defined by

$$\forall x, 0 < f'(x) \leq f'(0) = 1 \quad (2.24)$$

and by

$$\lim_{x \rightarrow +\infty} f(x) = 1 \text{ and } \lim_{x \rightarrow -\infty} f(x) = -1 \quad (2.25)$$

The KWTa Network is characterized by mutually inhibitory interconnections $a_{ij} = -1$ for $i \neq j$, an excitatory connection $a_{ii} = a$, ($|a| < 1$) and an external input (identical to every node) which identifies the number of winners, K , desired and the size of the network N : $\kappa_i = 2K - N$. The differential equations for the KWTa network are therefore:

$$\text{for all } i, C \frac{dx_i}{dt} = -\lambda_i x_i + (a+1)g(x_i) - \left(\sum_{j=1}^N g(x_j) - \kappa \right) \quad (2.26)$$

where $\lambda = N-1+|a|$, $-1 < a < +1$, and $\kappa = 2K - N$.

To the author's knowledge, there is only one synchronous KWTa implementation [35].

2.4 (K)WTA Circuit Implementation

In designing WTA circuits the following design issues should be addressed,

- ☐ high speed
- ☐ compactness
- ☐ high accuracy
- ☐ low power consumption
- ☐ robustness
- ☐ large input activation range
- ☐ small real-estate consumption

Some of these issues such as compactness and small real-estate consumption are correlated. Tradeoffs also exist between these requirements. For high speed operation large gate overdrive voltage for MOS transistors is a must but it is well known that offset voltage between input transistors increases as the gate overdrive voltage increases, deteriorating accuracy of the circuit. This also would cause large power consumption. Fabrication tolerances imply large gate area devices for robust circuits, but such devices cause larger parasitic capacitances. In order to charge these capacitance in a given time span the current flowing through the circuit should be large, resulting in high power consumption. Note that, all of these correlations and tradeoffs are also valid for other analog circuits. A good compromise between these issues can only be achieved taking.

- ☐ process characteristics (element parameters)
- ☐ statistical parameter deviation
- ☐ demanded circuit performance

into account. Circuit techniques which are known to be efficient in performance enhancement of other analog blocks (i. e., comparators-opamps) are also applicable to WTA circuits, cascoding being a good example. Guidelines for layout preparation are given in Appendix A.

Implementation of (K)WTA circuits can be realized in one or several layers. Selection of number of layers is a key design issue determining the performance parameters stated above [36].

The pair-compared neural network (PACNET) is a multi-layer feed-forward structure which is composed of pair comparison subnets used to pick a maximum in

a pair-by-pair hierarchy. Each subnet picks the maximum of two inputs and forwards it to the next layer for further comparison. Since an n -input PACNET is constructed in a pair by pair hierarchy, the number of layers becomes the nearest integer greater than $\log_2 n$ ($\lceil \log_2 n \rceil$). It should be noted that the convergence will be reached after $\lceil \log_2 n \rceil$ cycles. Thus, for n inputs, the PACNET needs $\lceil \log_2 n \rceil$ cycles to complete a WTA process which is independent of the distribution of the original activations. The convergence of the PACNET is obviously assured. However, $\lceil \log_2 n \rceil$ layers are complicated for hardware realization, especially where n is large. In those cases, larger real estate and power consumption is inescapable. Besides these drawbacks, the most advantageous aspect of PACNET appears to be its inherent stability. In other words, local feedback is applied to individual comparison cells (where necessary) having relatively simpler structures which makes it easier to stabilize the networks.

The maximum selection network (MAXNET) is first initialized by the inputs, then the competition process is enabled. This process continues until one of the cells becomes winner. The simple configuration, i. e., one layer structure, is the favorable advantage of the MAXNET. However, the convergence speed slows down as the input count enlarges. It also largely depends on the distribution of activations described in Appendix B.

Now lets take a look at some previous well-known WTA circuit implementations.

2.4.1 Lazzaro WTA circuit

The Lazzaro WTA circuit [14] shown in Fig.2.1 is the simplest CMOS implementation of current-mode asynchronous WTA circuits. It is composed of n interacting cells and a tail current source. The input currents compete with the current sunk by M2's. The voltage appearing at the input nodes propagate through source follower configured M1's. M1 having the highest gate voltage flows the larger part of the tail current and also determines the gate voltage of M2's. Owing to this negative feedback, the circuit equalizes the current sunk by M2's to the maximum of the input currents. Hence lower magnitude input currents remain below this level and the corresponding evaluation node voltages die out. More detailed explanation of this circuit will be given in the following chapter. The discrimination capability of the original circuit has been increased significantly applying simple cascoding [37] and gain-boosted regulated cascoding techniques [38]. These

modifications confer the circuit with a high resolution enabling it to compete with more recent current-based circuits.

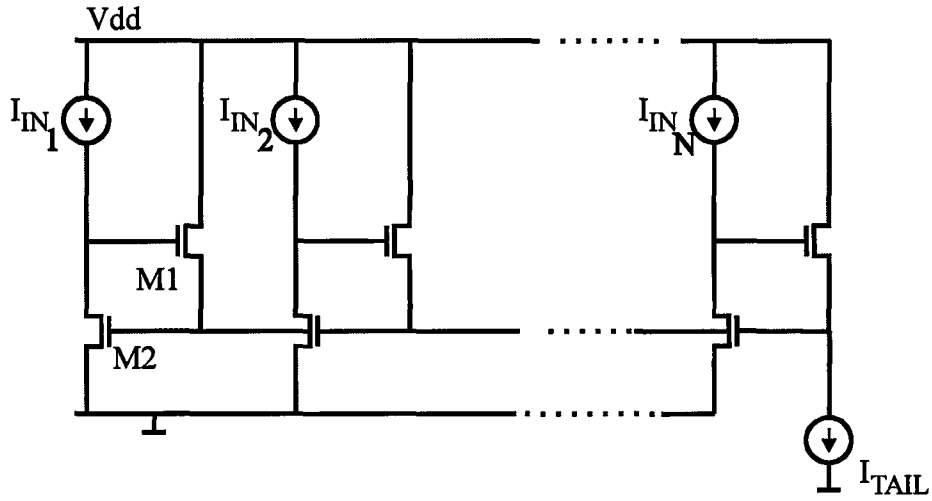


Figure 2.1 The Lazzaro WTA circuit.

2.4.2 Urahama-Nagao implementation

2.4.2.1 WTA circuit

Selection of the largest number among n real ones, $a_1, a_2, \dots, a_n$ can be formulated by the following zero-one integer programming problem [39].

$$\min \quad - \sum_{i=1}^n a_i x_i \quad (2.27)$$

$$\text{subj. to} \quad \sum_{i=1}^n x_i = 1 \quad (2.28)$$

In the optimum solution of this problem, one x_i corresponding to the largest a_i takes the value of 1 and all the remaining x_i 's become 0. Through the extension of the domain of x_i from binary points $\{0,1\}$ to the continuous interval $[0,1]$, this problem can be formulated by nonlinear programming

$$\min \quad - \sum_{i=1}^n a_i x_i \quad (2.29)$$

$$\text{subj. to } \sum_{i=1}^n x_i = 1 \quad (2.30)$$

$$\sum_{i=1}^n x_i \ln x_i = 0 \quad (2.31)$$

The mathematical programming theory states that the solution of (2.29) is a saddle point of the following Lagrange function, i.e., the minimum with respect to x and maximum with respect to λ and μ .

$$L = -\sum_{i=1}^n x_i a_i + \lambda \left(\sum_{i=1}^n x_i - 1 \right) + \mu \sum_{i=1}^n x_i \ln x_i \quad (2.32)$$

Let us fix μ to a sufficiently small positive value. The optimum solution of (2.29) then satisfies the following two equations

$$\frac{\partial L}{\partial x_i} = -a_i + \lambda + \mu(\ln x_i + 1) = 0 \quad (2.33)$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^n x_i - 1 = 0 \quad (2.34)$$

We get

$$x_i = e^{\frac{a_i - \lambda - \mu}{\mu}} \quad (2.35)$$

from (2.33) and its insertion into (2.34) yields the expression

$$e^{\frac{\lambda + \mu}{\mu}} = \sum_{i=1}^n e^{\frac{a_i}{\mu}} \quad (2.36)$$

whose insertion into (2.35) reads

$$x_i = \frac{e^{\frac{a_i}{\mu}}}{\sum_{j=1}^n e^{\frac{a_j}{\mu}}} \quad (2.37)$$

which approaches the solution of (2.29) when μ becomes sufficiently small. Equation (2.37) can be implemented with the circuit illustrated in Fig. 2.2 where MOS transistors are biased in their subthreshold range where their saturated drain current varies exponentially with their gate to source voltage as

$$y_i = J e^{\alpha(a_i - v)} \quad (2.38)$$

with positive J and α . The currents y_i 's come together at the common wire connecting the source of transistors, and their sum is forced to be equal to the current of the independent current source, I ; by Kirchoff's Law, as shown in Fig. 2.2.

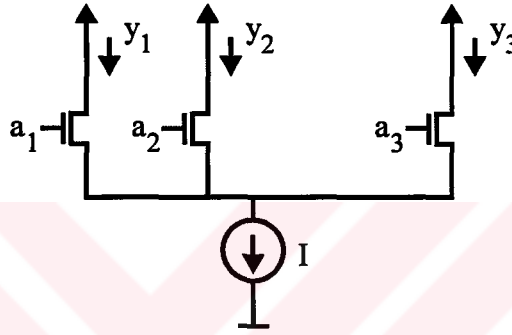


Figure 2.2 Circuit implementing Eq. (2.27) and Eq. (2.28).

$$\sum_{i=1}^n y_i = I \quad (2.39)$$

If we relate the parameters and variables I, J, α, v and y_i with λ, μ , and x_i as $\alpha = 1/\mu$, $v = \mu \ln(J/I) + \lambda + \mu$ and $y_i = I x_i$ the Equation (2.35) coincides with (2.38) and (2.39) reduces to (2.34). The circuit devised by Lazzaro et al. [14], whose inputs are currents, has a configuration modified from that of Fig. 2.2 whose inputs are voltages.

2.4.2.2 KWTA circuit

The derivation of the WTA networks in the previous section can be extended to KWTA circuits. The selection of the k largest numbers from n real ones $a_1, a_2, \dots, a_n$ can be formulated by the following zero-one integer programming;

$$\min \quad - \sum_{i=1}^n a_i x_i \quad (2.40)$$

$$\text{subj to } \sum_{i=1}^n x_i = k \quad x_i \in \{0,1\} \quad (2.41)$$

In the optimum solution of this problem, x_i 's corresponding to the largest a_i 's take the value 1 and all the remaining x_i 's become 0. The problem can be reformulated as the following nonlinear programming.

$$\min \quad - \sum_{i=1}^n a_i x_i \quad (2.42)$$

$$\text{subj to } \sum_{i=1}^n x_i = k \quad (2.43)$$

$$\sum_{i=1}^n [x_i \ln x_i + (1 - x_i) \ln(1 - x_i)] = 0 \quad (2.44)$$

The expression in (2.42) is adopted to make the Lagrangian result in a form that fits a suitable circuit. The solution of (2.42) is a saddle point of the following Lagrange function, i.e., minimum with respect to x and maximum with respect to λ and μ .

$$L = - \sum_{i=1}^n a_i x_i + \lambda \left(\sum_{i=1}^n x_i - k \right) + \mu \sum_{i=1}^n [x_i \ln x_i + (1 - x_i) \ln(1 - x_i)] \quad (2.45)$$

The Lagrange multiplier μ is fixed to a sufficiently small positive value. The optimum of (2.42) is then a solution of two equations.

$$\frac{\partial L}{\partial x_i} = -a_i + \lambda + \mu \ln \frac{x_i}{1 - x_i} = 0 \quad (2.46)$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^n x_i - k = 0 \quad (2.47)$$

We get,

$$x_i = \frac{e^{a_i/\mu}}{e^{a_i/\mu} + e^{\lambda/\mu}} \quad (2.48)$$

from (2.46). Since expression (2.48) is a special case of (2.37) with n being 2, this form can be implemented with the circuit shown in Fig. 2.3.

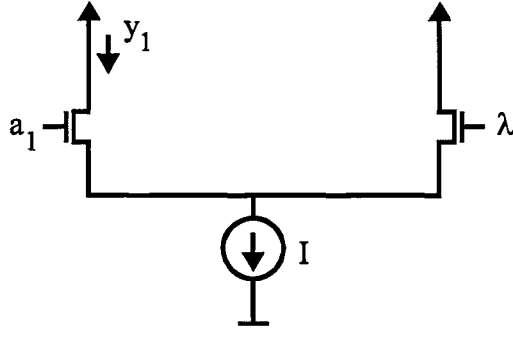


Figure 2.3 Circuit implementing Eq. (2.48).

The current y_i in Fig. 2.3, is proportional to x_i in (2.48) with the factor I as $y_i = Ix_i$. thus (2.47) becomes

$$\sum_{i=1}^n y_i - kI = 0 \quad (2.49)$$

Copying y_i in Fig. 2.3 using current mirrors and adding a common wire ensuring (2.49) we obtain the whole configuration of the KWTa circuit shown in Fig. 2.4.

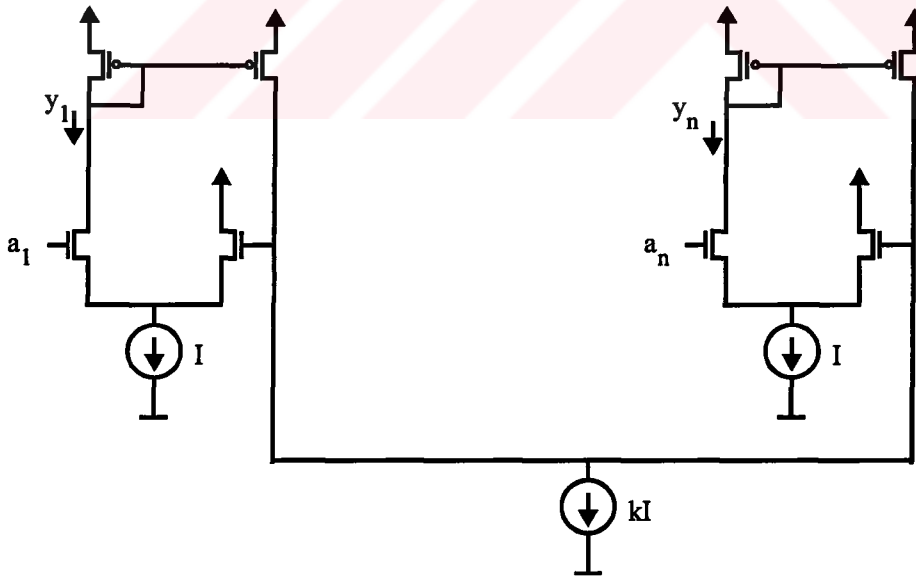


Figure 2.4 KWTa circuit with voltage inputs.

It is an asynchronous voltage-mode circuit of $O(N)$ complexity [15]. Transistor-level schematic diagram for the WTA circuit is shown in Fig. 2.5. All transistors operate in the saturation region. Each WTA cell consists of two branches. The first branch ($M1$, $M2$ and $IM5$) converts an input voltage into cell current. These currents are compared and redistributed along the common signal node. All source terminals of the input transistors $M1$'s are tied to this common signal node. In the second branch ($M3$ and $IM4$), the current in each cell is converted into the output voltage. Since the source terminals of transistor $M1$ in each cell are at the same potential for all the cells, the current flowing through each cell is related to the square of the individual input voltage. Thus, the largest input voltage can secure the largest amount of current from the total bias current. This largest current is converted and amplified to produce the output voltage of the winning cell. If the input voltage differences are sufficiently large, the output voltage of the winning cell is saturated at the positive supply value, while the output voltages of the other cells are grounded. Through the use of a common signal node, the total bias current is provided by the constant current source $IM5$ in every cell. As the number of inputs increases, the layout of the WTA circuit can easily be extended by abutting the common signal node of additional cells to the existing circuit.



The topology of the circuit is too complex for such a medium resolution circuit. To obtain a high-resolution circuit, cascading of the identical stages is suggested which makes the circuit topology even more complex. On the other hand the stability issue has not been examined.

2.4.4 Çilingiroğlu implementation

This synchronous circuit of $O(N^2)$ complexity is charge-based [18]. The circuit topology is depicted in Fig. 2.6. The elements of the input vector are charges dumped into capacitances connected to the input nodes. The NMOS transistors turn on in saturation and start discharging the rows. The rate of discharge is the lowest for the row of the highest voltage because the lower voltages of other rows minimize its pull-down current. Therefore, the voltages of all losing cells eventually fall below V_{TH} while that of the most activated cell remains above V_{TH} . The moment this happens, the winning row returns to a high-impedance-state and thereafter retains its above V_{TH} voltage while other rows continue to be discharged until all are grounded. This completes the first phase of discrimination. When row pull-up current sources are turned on, all row voltages start an upward swing but only the voltage of the row having an initial advantage of at least one V_{TH} and the highest rate of pull-up,

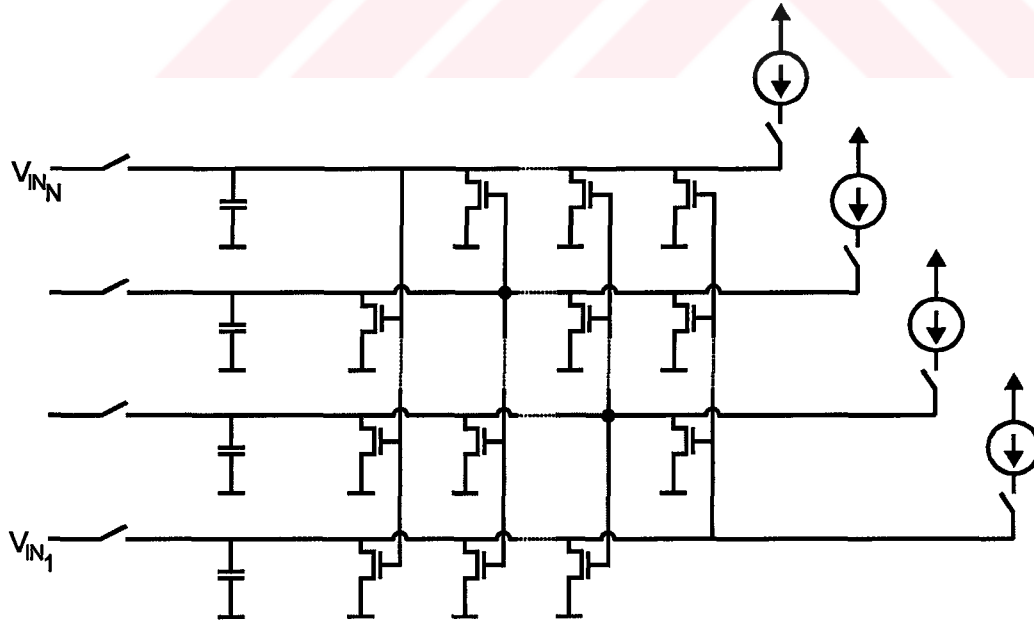


Figure 2.6 Circuit diagram of the Çilingiroğlu WTA circuit.

reaches VDD. Other row voltages are kept below V_{TH} by the pull-down transistor. Output buffers digitize these signals to VDD or to ground and transfer them to the output.

This circuit is advantageous in terms of operational simplicity. The most important drawback stems from its complexity of $O(N^2)$. For large input-count the number of elements composing the transistor matrix becomes large.

2.4.5 Serrano-Barranco implementation

This is an asynchronous current-mode circuit of $O(N)$ complexity [21]. The circuit schematic is shown in Fig. 2.7. Each cell consists of a two-output current mirror (M2, M3, M4), a CMOS inverter, and a MOS transistor (M1). Each cell receives two input currents and delivers one output current. The inverter acts as a current comparator. If the inverter output is low, the NMOS M1 is off. If the inverter output is high, M1 is on, and its drain current becomes equal to the cell input current. The current flowing through all M1's are summed across the diode connected PMOS transistor MP. This current is mirrored into each cell through M5's to be compared with the corresponding input current.

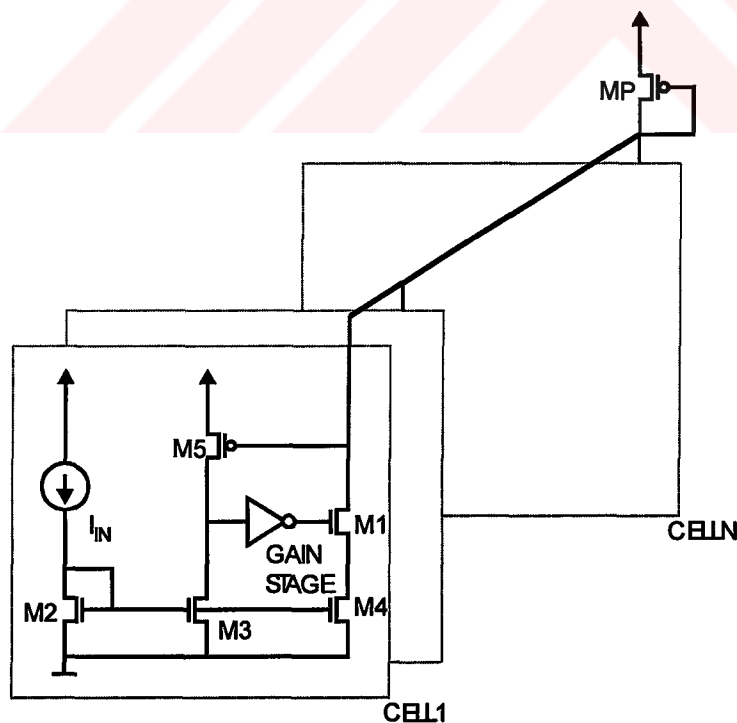


Figure 2.7 Circuit diagram of the Serrano-Barranco WTA circuit.

This circuit has several drawbacks. Firstly, the signal passes through two current mirrors (M2, M3, M4) and (MP and M5's) and deviations in the transfer ratio of these mirrors are multiplied, causing larger current discrepancies. Secondly, due to large gain of the inverter stage inserted within the negative feedback loop, this circuit can only be stabilized using large compensation capacitances connected across the input and the output of each inverter. The last drawback comes from power dissipation issue. If the successors are very close to the winner, all of the corresponding inverters may be biased close to their toggle point. This causes large current to flow through the PMOS and NMOS transistors composing the inverter.



3. IMPROVEMENTS ON CURRENT-MODE WTA CIRCUITS

3.1 Lazzaro WTA Circuit

The Lazzaro WTA circuit is shown in Fig.3.1. It consists of n interacting cells and a tail current source. The unidirectional input currents strive against the current sunk by M2's to pull the gate voltage of M1's up. Since all M2's share the same gate-to-source voltage, they also try to sink the same inhibitory current. A comparison process takes place at the evaluation-node of cells between their input currents and the inhibitory current. For a moment, suppose that the maximum input current prevails over the inhibitory current. Then, the evaluation-node voltage (gate voltage of M1) of the corresponding cell increases. Since the source voltage of M1 follows its gate voltage, the common-source-node voltage (gate voltage of M2's) also increases. This upward swing continues until the inhibitory currents become equal to the maximum input current. In all cells except for the winning cell, where a balance is sustained, inhibitory current dominates over the input current.

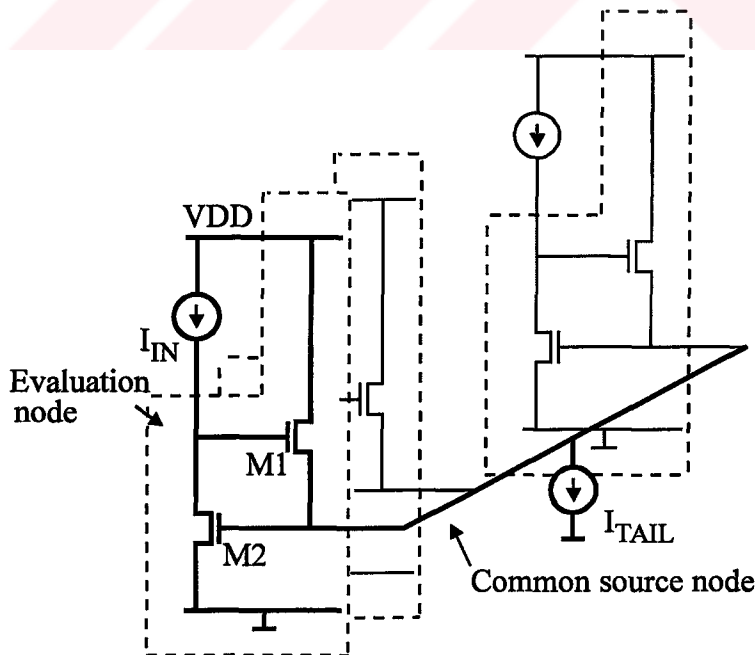


Figure 3.1 Architecture of the Lazzaro WTA circuit.

This unbalance implies a decrement of evaluation-node voltage of losing cells and force their M2's to operate in linear region, decreasing their drain current down to their input current. Hence, distinct voltage levels appear for the winning and losing cells at their evaluation-node. The entire tail current is steered towards M1 of the winning cell.

3.1.1 Stability analysis

The small signal equivalent of the Lazzaro WTA circuit is shown in Fig. 3.2.

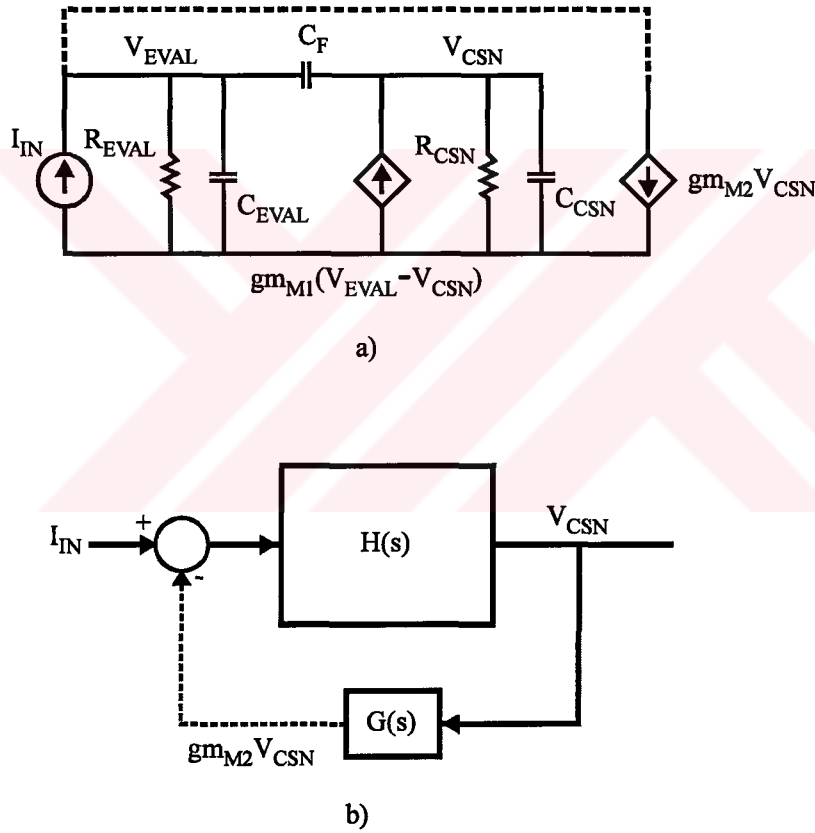


Figure 3.2. a) Small-signal equivalent of the Lazzaro circuit b) its block diagram, dashed line indicating the feedback connection.

Here, V_{EVAL} , C_{EVAL} , R_{EVAL} , V_{CSN} , C_{CSN} , R_{CSN} are voltage, capacitance and resistance associated with evaluation-node and common-source-node, gm_{M1} and gm_{M2} are

transconductance of transistors M1 and M2 respectively. The gate-to-source capacitance of M1 is presented by C_F , connected between the evaluation and common-source nodes.

This is a 2-pole 1-zero system. Kirchoff's Law implies the equations,

$$I_{IN} = \frac{V_{EVAL}}{R_{EVAL}} + V_{EVAL}sC_{EVAL} + (V_{EVAL} - V_{CSN})sC_F \quad (3.1)$$

$$(V_{EVAL} - V_{CSN})sC_F + gm_{M1}(V_{EVAL} - V_{CSN}) = \frac{V_{CSN}}{R_{CSN}} + V_{CSN}sC_{CSN} \quad (3.2)$$

leading to the following open-loop transfer function

$$\frac{I_{IN}}{V_{CSN}} = \frac{s(C_{CSN} + C_F) + gm_{M1} + \frac{1}{R_{CSN}}}{sC_F + gm_{M1}} \left[\frac{1}{R_{EVAL}} + s(C_{EVAL} + C_F) \right] - sC_F \quad (3.3)$$

Arranging the terms we get,

$$\frac{V_{CSN}}{I_{IN}} = H(s) = \frac{sC_F + gm_{M1}}{s^2(C_{EVAL}C_{CSN} + C_{EVAL}C_F + C_{CSN}C_F) +} \quad (3.4)$$

$$\frac{1}{s \left(C_{EVAL}gm_{M1} + \frac{C_{EVAL} + C_F}{R_{CSN}} + \frac{C_{CSN} + C_F}{R_{EVAL}} \right) + \frac{gm_{M1}}{R_{EVAL}} + \frac{1}{R_{EVAL}R_{CSN}}}$$

Since closed-loop transfer function takes the form,

$$\frac{V_{CSN}}{I_{IN}} = T = \frac{H}{1 + GH} \quad (3.5)$$

Substitution of (3.4) for H and gm_{M2} for G in (3.5) yields the transfer function of the closed-loop system as,

$$T = \frac{sC_F + gm_{M1}}{s^2(C_{CSN}C_{EVAL} + C_{CSN}C_F + C_{EVAL}C_F) +}$$

$$\frac{1}{s \left(C_{EVAL}gm_{M1} + C_Fgm_{M2} + \frac{C_{EVAL} + C_F}{R_{CSN}} + \frac{C_{CSN} + C_F}{R_{EVAL}} \right) + \frac{gm_{M1}}{R_{EVAL}} + \frac{1}{R_{EVAL}R_{CSN}}}$$

$$\frac{\frac{gm_{M1}}{R_{EVAL}} + \frac{1}{R_{EVAL}R_{CSN}} + gm_{M1}gm_{M2}}{(3.6)}$$

Since R_{CSN} and R_{EVAL} are large, (3.6) simplifies to,

$$T = \frac{sC_F + gm_{M1}}{s^2(C_{CSN}C_{EVAL} + C_{CSN}C_F + C_{EVAL}C_F) + gm_{M1}gm_{M2}} \quad (3.7)$$

Now let us examine the system for $C_F = 0$. The transfer function takes the form

$$T = \frac{\frac{1}{gm_{M2}}}{s^2 \frac{C_{CSN}C_{EVAL}}{gm_{M1}gm_{M2}} + s \frac{C_{EVAL}}{gm_{M2}} + 1} \quad (3.8)$$

Assuming the general form for second-order low-pass system [40]

$$T = \frac{1}{\frac{s^2}{\omega_n^2} + \frac{2\zeta}{\omega_n}s + 1} \quad (3.9)$$

we obtain the ringing frequency and the damping ratio as

$$\omega_n = \sqrt{\frac{gm_{M1}gm_{M2}}{C_{CSN}C_{EVAL}}} \quad (3.10)$$

$$\zeta = \frac{1}{2} \sqrt{\frac{gm_{M1}C_{EVAL}}{gm_{M2}C_{CSN}}} \quad (3.11)$$

The frequency response exhibits damped oscillations (ringing), if ξ is less than $1/\sqrt{2}$. Hence the structural limit for ringing free settlement can be expressed as

$$\frac{gm_{M1}C_{EVAL}}{gm_{M2}C_{CSN}} > 2 \quad (3.12)$$

In general the M1 of each cell with non-zero drain current contributes to the conductance seen looking into the common-source-node. Hence (3.12) takes the form,

$$\frac{C_{CSN}}{C_{EVAL}} < \frac{\sum_{i=1}^n gm_{M1_i}}{2 gm_{M2}} \quad (3.13)$$

The sum of transconductance of M1's is minimized and therefore the bound becomes more pronounced when the whole tail current is directed towards the winning cell. M2 of the winning cell sinks the maximum of the applied input currents I_{IN-MAX} . Thus (3.13) can be expressed in terms of applied currents and device parameters as

$$\frac{C_{CSN}}{C_{EVAL}} < \frac{1}{2} \sqrt{\frac{I_{TAIL}}{I_{IN-MAX}}} \sqrt{\frac{(W/L)_{M1}}{(W/L)_{M2}}} \quad (3.14)$$

The greater I_{IN-MAX} , the smaller the upper bound expressed on the right-hand side of (3.14). Therefore, for the worst case analysis, I_{IN-MAX} should be replaced by the maximum permissible input current, $I_{IN-LIMIT}$. On the other hand, the greater the C_F , the more stable the circuit. Hence, satisfaction of (3.14) ensures the stability of the system.

3.1.2 Limitations

Now, let us give the resolution of this circuit a glance. The input current sources will be assumed to be ideal. Besides noise factor, the resolution of this circuit is subject to two distinct limitations. One of them is the random resolution limitation which stems from device mismatch. This limitation depends on physical and geometrical parameter deviation. The other is the systematic resolution limitation. This factor can be ascribed to the drain-to-source resistance of M2's.

3.1.2.1 Channel length modulation

Fig. 3.3. depicts the open-loop configuration of a cell. The open-loop cell gain assumes the form

$$\frac{\delta V_{OUT}}{\delta V_{IN}} = \frac{gm_{M1}}{gm_{M1} + gmb_{M1}} gm_{M2} R_{EVAL} \quad (3.15)$$

where gm_{M1} and gm_{M2} denote small-signal gate transconductance of M1 and M2, gmb_{M1} is small-signal substrate transconductance of M1, R_{EVAL} describes the parallel combination of the output resistances associated with the input current source and M2. The first ratio on the right-hand side of (3.15) describes gain of source follower connected M1 and can be assumed to be unity if the body effect factor is not noticeable.

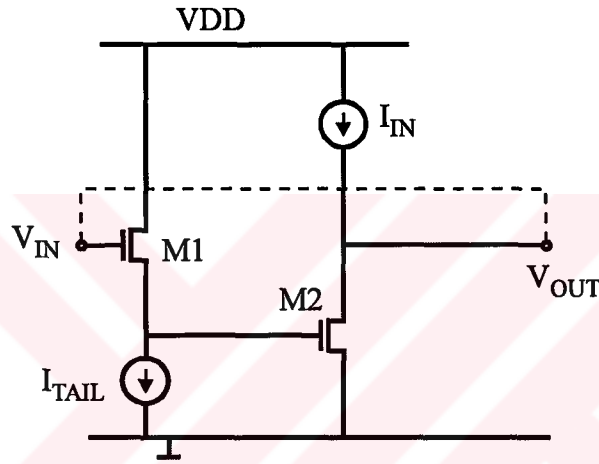


Figure 3.3. Open-loop configuration of a cell, horizontal dashed line indicating the unity-gain feedback loop.

The differential voltage between evaluation-nodes of the winning cell and its successor can be supposed to be equal to the voltage drop across R_{EVAL} generated by their differential input current. The linear dependence of the gain on both gm_{M2} and R_{EVAL} implies their enlargement for higher gain. Since the current through M2 is identified by the input current and is not a design parameter, the transconductance of M2 can only be increased by selecting its aspect ratio as high as possible, taking into account the adverse impact of the reduction of channel length on output resistance.

Consequently, the channel length modulation exhibited by M2 is the most crucial source of systematic precision limitation. The effect of the channel length modulation is included in the drain current expression as,

$$I_d = \frac{I_{d_{SAT}}}{1 - \frac{\Delta L}{L_{eff}}} \quad (3.16)$$

where $I_{d_{SAT}}$ is the drain current at the limit of the pinch-off ($V_d = V_{d_{SAT}}$), ΔL is the length of the pinch-off region, L_{eff} is the effective geometrical gate length. ΔL is function of the drain voltage and takes different values ΔL_W and ΔL_L for M2 of the winning cell and its successor respectively, if both are in saturation. The deviation in the inhibitory current can be obtained as,

$$\frac{\Delta I_{d_{M2}}}{I_{d_{M2}}} \approx \frac{\Delta L_W - \Delta L_L}{L_{eff}} \quad (3.17)$$

Consequently, the higher L_{eff} , the better the current discrimination capability of the circuit.

3.1.2.2 Device mismatch

The standard deviation in drain current mismatch between two identical transistors can be modeled as [41],

$$\frac{\sigma^2(I_d)}{I_d^2} = \frac{4}{(V_{GS} - V_{TH})^2} \frac{A_{VTH}^2}{WL} + \frac{A_\beta^2}{WL} \quad (3.18)$$

where β is the transconductance parameter, A_{VTH}^2 and A_β^2 are threshold voltage and transconductance standard deviation constants. The first term on the right hand side represents the effect of threshold voltage discrepancy and the second term displays the effect of deviation in transconductance. Replacing $(V_{GS} - V_{TH})^2$ with $I_d L / k_p W$, (3.18) can be expressed in terms of drain current as

$$\frac{\sigma^2(I_d)}{I_d^2} = \frac{4A_{VTH}^2}{L^2} \frac{k_p}{I_d} + \frac{A_\beta^2}{WL} \quad (3.19)$$

where $k_p = \mu_n C_{ox}$, is the process transconductance parameter. It is clear that the deviation in inhibitory currents can be kept low only by selecting both the gate length and the gate area of M2 large.

3.1.2.3 Gate area

Increasing the gate area causes higher real-estate consumption and lower speed due to the enhanced gate capacitance contribution of M2 to the common-source-node capacitance. For high speed applications the slewing capability of the common-source-node, C_{CSN} , should not exceed some limit, namely L_{SR}

$$C_{CSN} \leq \frac{I_{TAIL}}{L_{SR}} \quad (3.20)$$

Hence, the gate area of M2 is limited by one of this two constraints whichever is lower.

$$(WL)_{M2} \leq \min (area_{REAL-ESTATE} , area_{SPEED}) \quad (3.21)$$

Note also that the circuit has two poles and is unconditionally stable. Yet, if a ringing free settling behavior is desirable, the pole pertaining to the common-source-node should occur at high frequencies.

3.1.2.4 Aspect ratio

The voltage levels appearing at the evaluation-nodes for the losing and winning cells can be expressed as,

$$V_{LOW} = \sqrt{\frac{2I_{IN} \max \left(\frac{L}{W} \right)}{kp}}_{M2} \quad (3.22)$$

$$V_{HIGH} \approx V_{LOW} + V_{TH} + \sqrt{\frac{2I_{TAIL} \left(\frac{L}{W} \right)}{kp}}_{M1} + V_{TH} (V_{LOW} + V_{TH}) \quad (3.23)$$

where the second and the last terms stand for the threshold voltage of M2 and M1, respectively. If the aspect ratio of M2 is low, both V_{LOW} and V_{HIGH} shift towards VDD and may cause the PMOS transistor providing the input current to enter nonsaturation.

3.1.2.5 Design curves

The limitations discussed so far are illustrated in Fig 3.4. The technology related parameters are typical for a 0.6 μm CMOS technology and are listed in Table 3.1.

The curves are drawn for a precision of 1 %. The geometry of M2 should be chosen to fall inside the shaded region of acceptability.

The aspect ratio of M1 has no effect on the precision of the circuit, but it should be selected as large as possible to push the pole associated with the common-source-node to higher frequencies and to decrease V_{HIGH} at the same time.

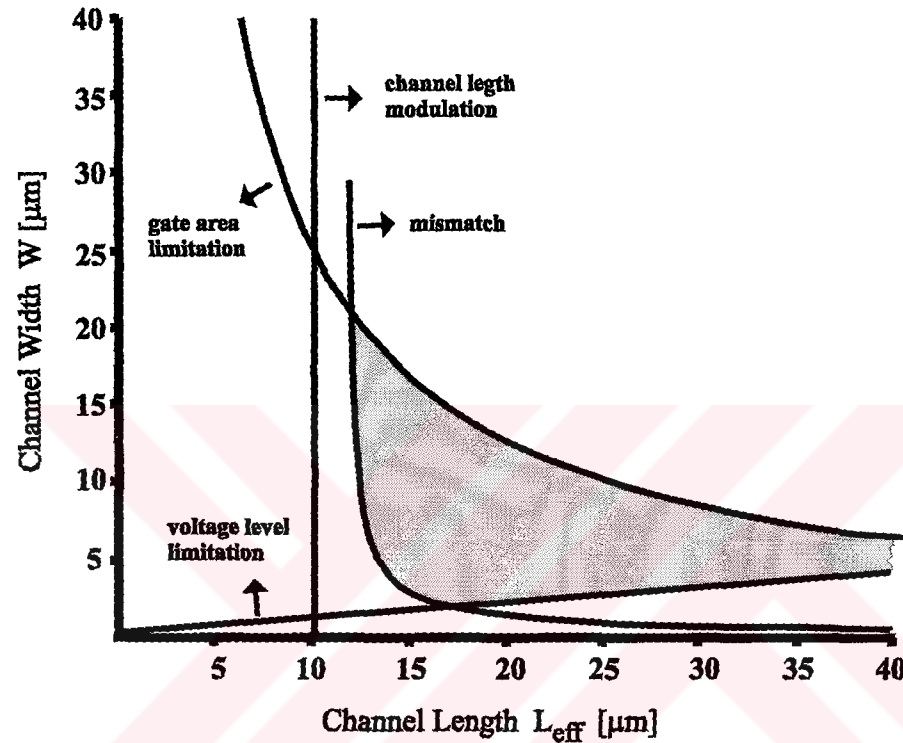


Fig. 3.4 Design curves for M2.

Table 3.1. The technology related parameters for a 0.6 μm CMOS technology and circuit design parameters.

V_{DD}	3.3 V	V_{HIGH}	2 V
A_{VTO}	15 mV μm	I_{IN}	5 μA
A_{β}	4 % μm	I_{TAIL}	10 μA
L_{SR}	10 V/ μs	N	100
V_{TH}	0.5 V	K_p	100 $\mu\text{A}/\text{V}^2$

3.2 Proposed Modifications

A WTA circuit designed to have a good systematic resolution makes a clear separation between the winner and the losers. If it is not as good as expected in random resolution, it is probable that the selected cell is not the one having the highest input current. For some applications, a good random resolution may not be achievable due to real-estate or speed restrictions, nevertheless a very high systematic resolution may still be desirable. The reason is that the selection of an element among the most excited elements may be much more profitable than the situation where no winner is selected (total indeterminism). Cascoding techniques can be used to obtain WTA circuits with high resolution.

Schematic diagrams of both simple (nonregulated) cascode and gain-boosted regulated-cascode voltage dependent current sources are shown in Fig. 3.5. M2 converts the control voltage V_{CSN} into current which also flows through M3. In order to lessen the dependence of this current on the output voltage, the drain voltage of M2 should be kept as stable as possible. In the simple (nonregulated) cascode circuit, this is done by combining M3 in a cascode connection with M2. Hence, the drain of M1 is loaded with the low source resistance of M3. In the regulated cascode current source, the cascoding effect of M3 is increased by adding a gain stage consisting of M4 and the constant current source, I_{BIAS} . The negative feedback forces M3 to flow a drain current equal to I_{BIAS} . This implies a fixed gate-to-source voltage for M3 [42,43].

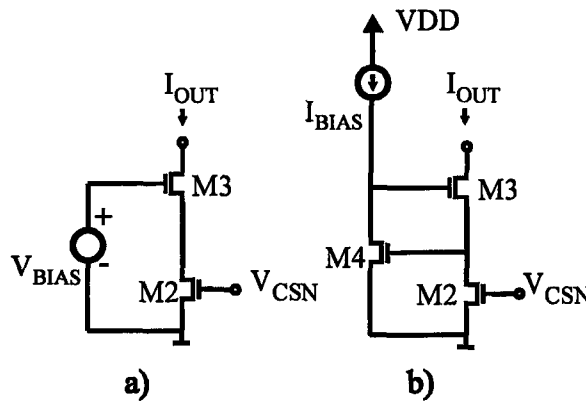


Figure 3.5. Cascode voltage dependent current sources; a) simple (nonregulated) , b) gain-boosted regulated.

The small-signal output resistance for low frequencies of simple (nonregulated) cascode and gain-booster regulated cascode circuits can be approximated by

$$r_{oSIM-CAS} = \frac{1}{g_{dsM2}} \frac{g_{mM3}}{g_{dsM3}} \quad (3.24)$$

and,

$$r_{oREG-CAS} = \frac{1}{g_{dsM2}} \frac{g_{mM3}}{g_{dsM3}} \frac{g_{mM4}}{(g_{dsM4} + g_{oBIAS})} \quad (3.25)$$

respectively, where g_{mM_v} is the gate transconductance, g_{dsM_v} is the output conductance of the respective transistor M_v ($v=1,2,3$) and g_{oBIAS} is the output conductance of the current source I_{BIAS} . It is seen from (3.24) that the output resistance of M2 is increased g_{mM3}/g_{dsM3} fold for simple cascoding. This resistance can be increased by an additional factor of $g_{mM4} / (g_{dsM4} + g_{oBIAS})$ using the gain-booster regulated-cascode circuit as seen from (3.25).

3.2.1 Simple cascoding

Referring to Fig 3.1, an efficacious arrangement can be made to improve output resistance of M2 by using the configuration shown in Fig. 3.6 which combines M2 in a cascode connection with M3. The effect of cascode connection is an increase in the output resistance of the part sinking the inhibitory current which is equal to R_{dsM2} in the basic circuit, by a factor of g_{mM3}/g_{dsM3} , resulting in an improvement in the resolution by the same amount. The output resistance defining the resolution is the parallel combination of the output resistance of this improved current sinking part and that of input current source. In order for this configuration to be profitable, output resistance of input current source should also be high and be formed using cascode configuration. The output voltage level of the winner is equal to the sum of gate-to-source voltage of M1 conducting the entire tail current and M2 sinking the maximum of the applied input currents. On the other hand, the output voltage level for the loser is equal to $V_{G_{M3}} - V_{TH}$.

The maximum separation between winner and loser output levels is obtained by decreasing $V_{G_{M3}}$ hence drain voltage of M2 down to the linear region boundary of M2. As this boundary depends on the maximum input current, the simplest

solution is to bias the gate of M3's at a constant level enabling M2's to operate in saturation even in the worst case where the maximum permissible input current is applied to one of the cells. Should the maximum permissible input current be described by $I_{IN-LIMIT}$, then the maximum saturation drain voltage of M2's becomes

$$V_{dsat_{M2}} = \sqrt{\frac{2I_{IN-LIMIT}}{\beta_{M2}}} \quad (3.26)$$

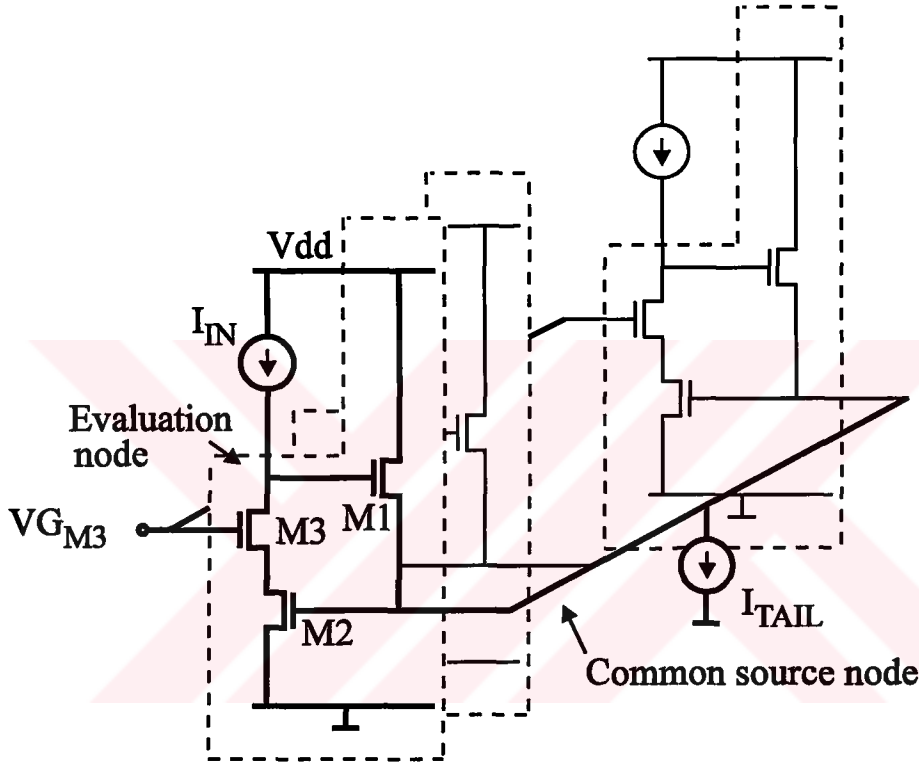


Figure 3.6 Simple cascode WTA circuit.

where β_{M2} is the device transconductance parameter of M2. The gate voltage to be applied to M3's is given by

$$V_{GM3} = V_{dsat_{M2}} + V_{TH}(V_{dsat_{M2}}) + \sqrt{\frac{2I_{IN-LIMIT}}{\beta_{M3}}} + \Delta V \quad (3.27)$$

where β_{M3} is the device transconductance parameter of M3. The sum of the second and the third terms express the gate-to-source voltage of M3 including the body effect. Taking discrepancies coming from device mismatches into account, the last

term is added to keep M2 away from linear region. Although dependent on the technology used, ΔV should be selected in the order of 100 mV.

Insertion of M3 in the improved circuit creates an additional pole at its source node, which occurs at

$$p_S = \frac{gm_{M3}}{2\Pi (C_{s_{M3}} + C_{d_{M2}})} \quad (3.28)$$

where $C_{s_{M3}}$ and $C_{d_{M2}}$ are the source capacitance of M3 and the drain capacitance of M2, respectively. The stability analysis done for the original Lazzaro circuit and consequently (3.14) are still valid for the improved circuit provided that p_s occurs at frequencies higher than ω_n given by (3.10). Else, p_s becomes the first non-dominant pole. Substituting gm_{M1} by gm_{M3} and C_{CSN} by $C_{s_{M3}} + C_{d_{M2}}$ in (3.12), the limitation for ringing free settling can be obtained as,

$$\frac{C_{s_{M3}} + C_{d_{M2}}}{C_{EVAL}} < \frac{1}{2} \sqrt{\frac{(W/L)_{M3}}{(W/L)_{M2}}} \quad (3.29)$$

which can be satisfied by proper selection of transistor sizes, regardless of tail and input current levels.

As stated above, the Lazzaro WTA circuit is unconditionally stable. Even if the bound (3.14) is not satisfied, the circuit accomplishes the discrimination process properly, but the evaluation can only be assumed to be valid when the ringing fades away and the circuit enters the steady state. Yet the modified version has three poles and can sustain oscillation if the phase-margin becomes negative.

3.2.2 Gain-boosted regulated-cascoding

The detailed diagram of the WTA circuit obtained replacing M2 by the gain-boosted regulated-cascode current source depicted in Fig. 3.5 b) is shown in Fig.3.7. It consists of n interacting cells and a current source, I_{TAIL} . The unit cell is shown within the dotted outline.

The operation principle of the circuit is similar to that of the original circuit and is not given here to avoid repetition.

When the evaluation-node voltage of losing cells drops below the voltage level, V_{OLOS} , given by

$$V_{OLOS} = V_{TH} + \sqrt{\frac{2I_{BIAS}}{\beta_{M4}}} + \sqrt{\frac{2I_{IN}}{\beta_{M3}}} \quad (3.30)$$

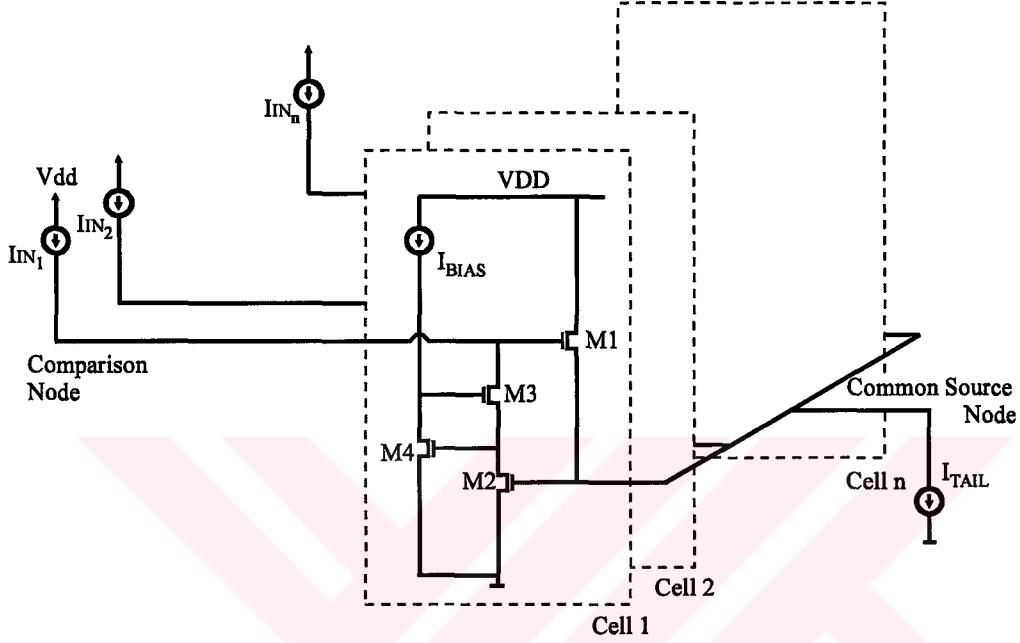


Figure 3.7 Circuit diagram of gain-booster regulated-cascode WTA circuit.

M3 leaves the saturated region. β_{Mv} 's are device transconductance parameters of transistor M_v ($v=1,2,3,4$). The transconductance and the drain-to-source resistance of a MOS transistor exhibit a noticeable degradation in the linear region. This results in lowering of the output resistance exhibited by the gain-booster regulated-cascode structure, given by (3.25). The voltage level appearing on the evaluation-node of the winning cell, is given as

$$V_{OWIN} = V_{TH} + \sqrt{\frac{2I_{IN-MAX}}{\beta_{M2}}} + V_{TH}(V_{gsM2}) + \sqrt{\frac{2I_{TAIL}}{\beta_{M1}}} \quad (3.31)$$

where I_{IN-MAX} is the maximum input current. The sum of first and the second terms on the right-hand side of (3.31) represents the gate-to-source voltage of M2. The third term is the threshold voltage of M1, including the body effect. The sum of this term with the last one gives the gate-to-source voltage of M1 of the winning cell.

Note that the entire tail current is steered towards M1 of the winning cell since all other M1's are off. The separation between winning and losing cells is given by $V_{DIFF} = V_{WIN} - V_{LOS}$. Using (3.30) and (3.31) this can be expressed as

$$V_{DIFF} = V_{TH}(V_{GS_{M2}}) + \sqrt{2I_{IN-MAX}} \left[\frac{1}{\sqrt{\beta_{M2}}} - \frac{1}{\sqrt{\beta_{M3}}} \right] + \sqrt{\frac{2I_{TAIL}}{\beta_{M1}}} - \sqrt{\frac{2I_{BIAS}}{\beta_{M4}}} \quad (3.32)$$

For a robust operation, the dependence of V_{DIFF} on input currents should be reduced. This implies the enlargement of the aspect ratio of M2 in order to keep $V_{GS_{M2}}$ low. Hence, the impact of the body effect represented by the first term on right-hand side of (3.32) decreases. The second term can be cancelled by selecting the aspect ratios of M2 and M3 equal. To achieve a larger difference between the evaluation-node voltage of the winning and losing cells other design parameter should be addressed. V_{DIFF} enlarges with increasing $(W/L)_{M4}$ and I_{TAIL} and with decreasing $(W/L)_{M1}$ and I_{BIAS} . The frequency stability considerations are principally the same as explained in Section 3.1.1. If the poles associated with the new (added) nodes occurs at high frequencies compared to ω_n given in (3.10) then the analysis given in the Section 3.1.1 can still be considered valid for this circuit.

3.2.3 Simulation results

Discrimination performance of the original, simple cascoded and gain-boosted regulated-cascoded circuit was compared based on DC simulations. All three circuits have 10 cells composed only of NMOS transistors ($W=10\mu m$, $L=0.8\mu m$). Level-47 parameter set provided by a well-known European foundry for its $0.6\mu m$ CMOS technology was used in all simulations (See Appendix C). The reason to keep transistor dimensions and technology the same for all three circuits is to illustrate advantages offered by modifications.

The DC transfer curve of original circuit is shown in Fig. 3.8. Referring to Fig 3.1, the tail current was set to $50\mu A$. Input current applied to cells, Cell-2 to Cell-10, were kept constant. The ordering of the magnitude of input currents were inversely dependent on the cell number, input current of Cell-2 being higher than that of Cell-3 which is higher than that of Cell-4, so on. The difference between input currents was constant and equal to $100nA$. Input current of Cell-2 was set to $9.9\mu A$. Evaluation node voltages were plotted as function of input current applied to Cell-1, which was changed from $9\mu A$ to $11\mu A$. It is shown from Fig. 3.8 that the determination of winner occurs only when, input current of cell-1 exceeds $10.5\mu A$.

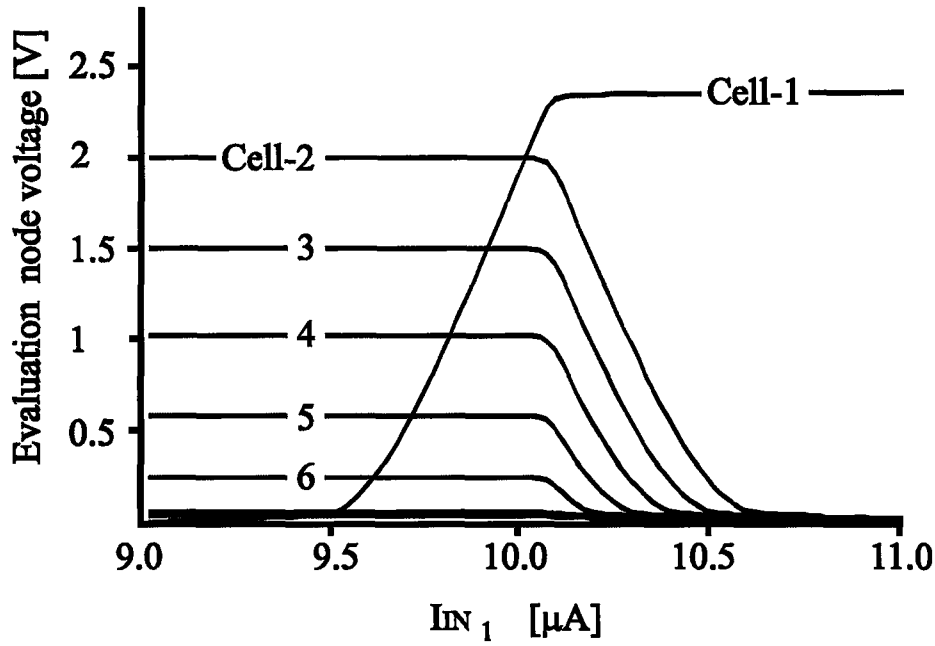


Figure 3.8 Simulated DC transfer curves of the 10-cell original Lazzaro WTA circuit.
 $I_{IN_i} = 10 \mu A - i \times 100 \text{ nA}$, $i = 2, 3, \dots, 10$.

Hence for a clear discrimination between the winner and loser (between Cell1 and Cell-2) a minimum input current difference of 600 nA is necessary.

The DC transfer curve of the simple cascode circuit is shown in Fig. 3.9. Referring to Fig 3.6, the tail current was set to 50 μA and the biasing voltage ($V_{G_{M3}}$)

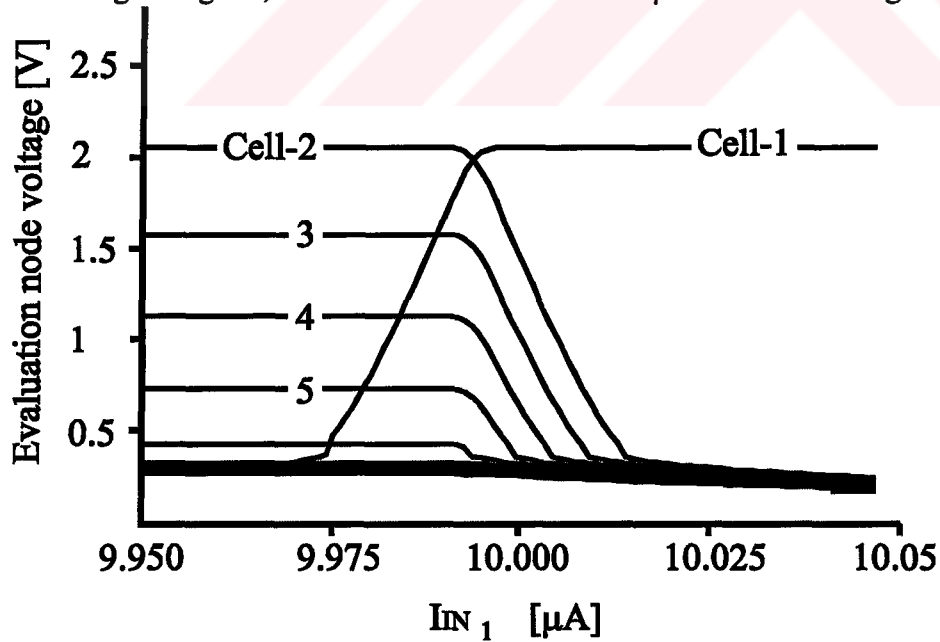


Figure 3.9 Simulated DC transfer curves of the 10-cell simple-cascode, WTA circuit.
 $I_{IN_i} = 10 \mu A - i \times 5 \text{ nA}$, $i = 2, 3, \dots, 10$.

was kept at 1.2V. The stimuli used for the simulation was similar to the one applied to the original circuit described above only input current levels being different. Input current level of cell-2 was set to 9.995 μA and the difference between equally spaced inputs was selected to be 5 nA. It is shown that the circuit makes a clear differentiation between the winner and losers when input current of Cell-1 is higher than 10.015 μA . This implies an input difference current of 20 nA between the first and the second cell. Hence, the discrimination capability is enhanced $600 \text{ nA}/20 \text{ nA} = 30$ times compared to the original circuit. The difference between voltage levels produced for the winner and losers decreased from 2.1 V to 1.8V.

The DC transfer curve of the gain-boosted regulated-cascode circuit is shown in Fig. 3.10. Referring to Fig 3.7, the tail current was set to 50 μA . The stimuli used for the simulation was similar to the previous ones, only input current levels being different. Input current level of cell-2 was set to 9.9999 μA and the difference between equally spaced inputs was selected to be 100 pA. A clear differentiation between the winner and losers was secured when input current of Cell-1 is higher than 10.00030 μA . This implies an input difference current of 400 pA between the first and the second cell. Hence the discrimination capability is enhanced $20 \text{ nA}/400 \text{ pA} = 50$ times compared to the simple regulated circuit Voltage levels produced for the winner and losers decreased further down to 1.3V.

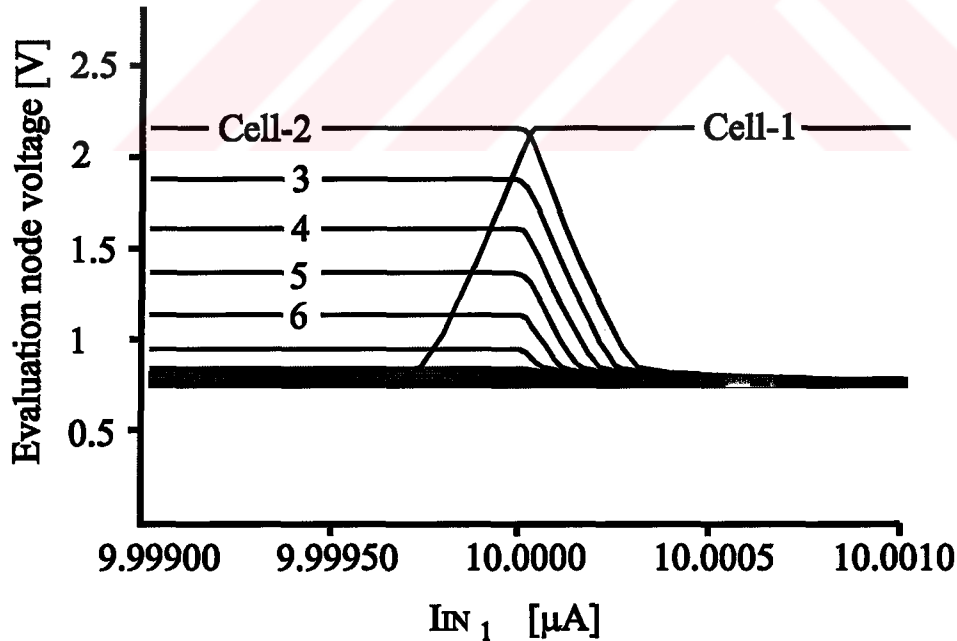


Figure 3.10 Simulated DC transfer curves of the 10-cell gain-boosted regulated-cascode, WTA circuit. $I_{INi} = 10 \mu\text{A} - i \times 100 \text{ pA}$, $i = 2, 3, \dots, 10$.

4 THE PROPOSED WTA CIRCUIT

4.1 Circuit Description and Operation

The architecture of the proposed WTA circuit and aspect ratios of transistors are shown in Fig. 4.1 and Table 4.1 respectively. It consists of n identical cells together with a voltage dependent current source common to all the cells. Each classification cycle consists of two consecutive phases (charging and selection) determined by the voltage waveforms ϕ_1 , ϕ_2 . During the charging phase, the initial input voltages are dumped onto the circuit. The circuit selects the winning cells during the second step called the selection phase. At the end, distinct voltage levels appears at the evaluation nodes for the winner and losers

During the charging phase, the gate voltage of M2 is pulled high, disabling the drain currents of M1, M3 and M4 and forcing the evaluation-nodes into high impedance state. Then the transmission gates consisting of M7 and M18 are turned on. The initial input voltages are dumped onto the equal evaluation-node capacitances and those voltages serve as the initial gate voltages of the NMOS transistors, M1's. This phase is terminated by M2 being turned on and the transmission gates being turned off.

At this point, the selection phase starts. The sum of currents flowing through all M1 transistors is equal to the current sunk by the current source, I_{TAIL} . At the beginning of the competition phase, M1 of the cell having the greatest input voltage (say the i 'th cell of n total cells), has the greatest drain current. The PMOS current mirror of each cell, formed by M3 and M4, copies the drain current of its M1 into its input capacitance. Since the source nodes of all M1 transistors are tied together, the difference between the current through $M1_i$ and the current through all other M1 transistors is enlarged. This positive feedback eventually causes all M1's other than $M1_i$ to turn off, because the entire tail current I_{TAIL} is steered to the latter. Since all the M5's share a common gate-to-source voltage, they should also sink the same amount of discharging current, providing an inhibitory negative feedback loop. However, if the differences between input voltages of the i 'th cell and its closest successors are small, then the tail current is divided almost equally among M1's of these highest-input cells. The negative feedback loop shifts all input voltage levels to

make the drain currents of their M4 and M5 equal. This leads the circuit to a metastable period, where it stays until the positive feedback loop stated above becomes dominant and the input voltage of the winning cell starts an upward swing while others decrease. When all input voltages are below the threshold voltage of NMOS transistors, none of M1's is on, and there is no charging current forcing the circuit to escape from this low voltage region. In this case, currents through M6's transferred by M3's and M4's to the input capacitances increase the input voltages linearly until the greatest becomes high enough to turn its M1 on.

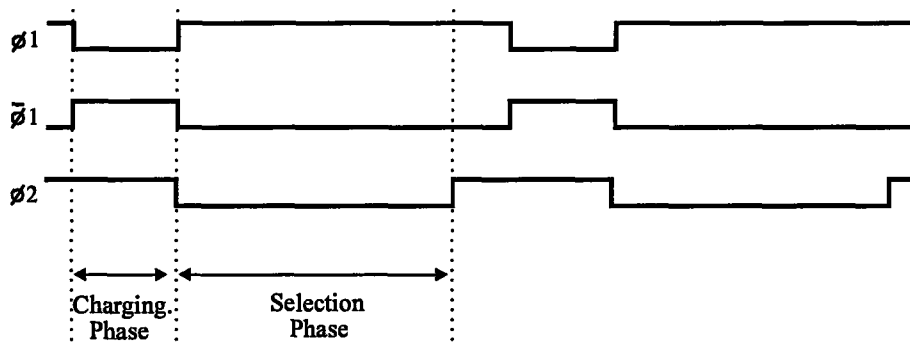
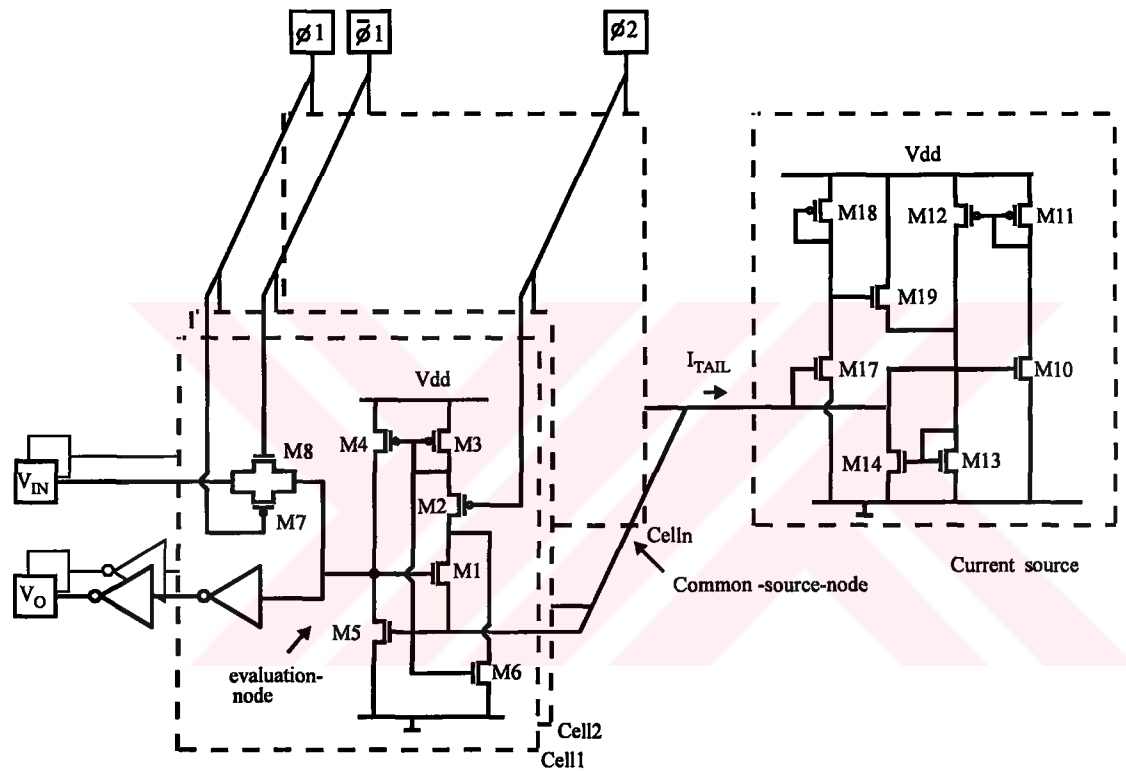


Figure 4.1. Architecture of the proposed WTA circuit.

Table 4.1: Transistor dimensions.

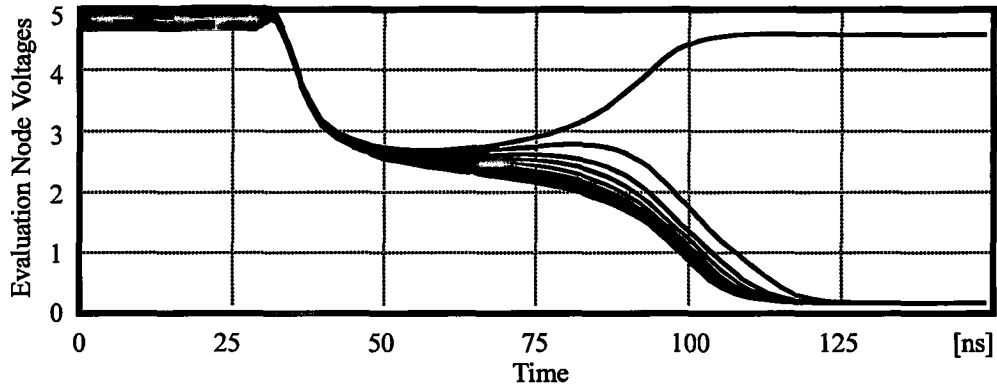
Transistors	W [μm]	L [μm]
M1, M11	13	3
M2, M3, M4	27	3
M5, M10, M13, M14, M17	7	3
M6	4	20
M12	19	3
M18	4	7
M19	4	9

At the end of the competition, the entire tail current is steered to $M3_i$. The drain currents of $M4_i$ and $M5_i$ are equal. The drain voltage of $M5_i$ should be as high as possible since it is output voltage of the winning cell. This requires that $M4_i$ be a wide transistor and operate in nonsaturated mode. Since $M3_i$ and $M4_i$ share the same gate-to-source voltage, source-to-drain voltage of $M4_i$ can be expressed as,

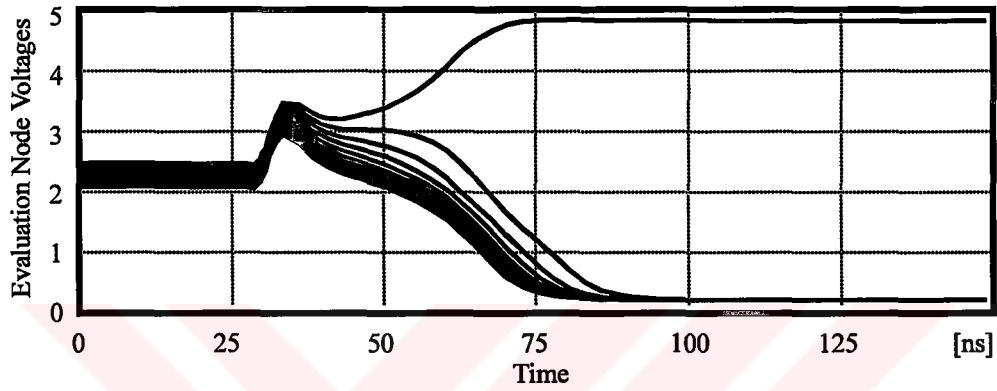
$$V_{ds_{M4}} = \sqrt{\frac{2I_{TAIL}}{\beta_{M3}}} \left[1 - \sqrt{1 - \frac{\beta_{M3}}{\beta_{M4}} \frac{I_{M5}}{I_{TAIL}}} \right] \quad (4.1)$$

where β is termed the device transconductance parameter. In order for the circuit to produce high output voltage at the input node of the winning cell $M4$, must be much larger than $M3$ and I_{TAIL} / I_{M5} ratio must be high. Yet, at the same time, they are the most important parameters for stable operation and are discussed in the following section. Referring to the circuit diagram of the current source, the current through $M10$ is mirrored first by $M11$ and $M12$ and then transferred to the common-source-node by $M13$ and $M14$. If the voltage of the common-source-node is low, the current through $M10$ is small. In this case, the inverter consisting of $M17$ and $M18$ turns $M19$ on, and increases the tail current, to supply more charging current to the cells and to improve the transient performance.

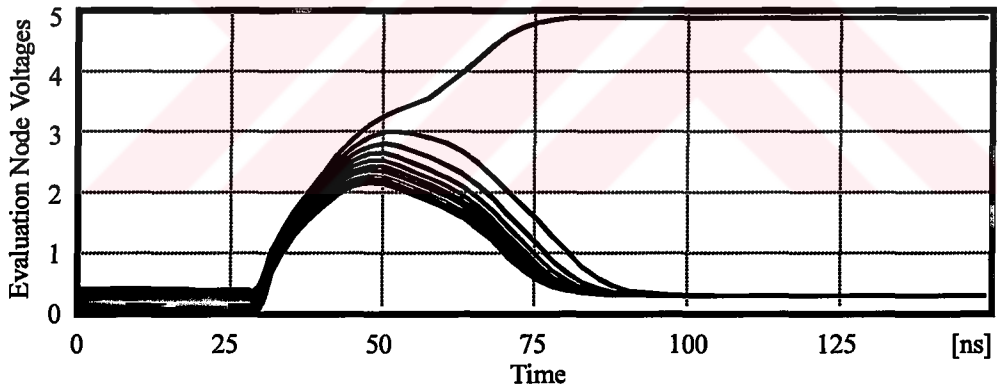
Post-layout HSPICE simulations were carried out to verify the performance of the proposed circuit. Transistor parameters provided by YITAL for its $3\mu\text{m}$ nwell technology were used in all simulations (see Appendix C). The circuit was simulated for various sets of input vectors and for different common voltage levels. Typical simulation results are shown in Fig 4.2.



a)



b)



c)

Figure 4.2 The HSPICE simulated evaluation node voltages of 10-cell WTA circuit for initial evaluation node voltages a) $V_i = 5V + i \times 40mV$, $1 \leq i \leq 10$; b) $V_i = 2V + i \times 40mV$, $1 \leq i \leq 10$; c) $V_i = 0V + i \times 40mV$, $1 \leq i \leq 10$.

It was observed that the circuit selects one and only one winner even if the differences between input voltages are small. In all transient simulations the selection phase is initiated at the 20'th ns. Referring to Fig 4.1, the rising edge of ϕ_1 and ϕ_2 arrives at the above specified time. The circuit produces nearly ideal binary values to

the winning and losing cells and can directly drive intra-chip logic gates, provided that each cell is loaded with the same capacitance.

The operation of the circuit relies on the positive feedback which strives against the inhibitory current, I_{M5} . Its role is to increase the share of the tail current allotted to M1 of a cell by increasing its evaluation-node voltage. At the end of the selection phase, only one cell can be winner since there is no sufficient amount of tail current for more cells. The general description of the circuit configuration is shown in Fig. 4.3. n initial activations $u_1(0), \dots, u_n(0)$, are fed into the circuit at time zero. Activation of the cells, standing for voltages at evaluation-nodes, for $t > 0$ are given by,

$$u_i(t) = u_i(t-1) + \frac{f_1 \frac{(W/L)_{M4}}{(W/L)_{M3}} y_i(t) - f_2 r I_{TAIL}(t)}{C_{EVAL}/dt} \quad (4.2)$$

where,

$$f_1 = \begin{cases} 0 & \text{if } z(t-1) > V_{dd} - V_{TH} \text{ or } u_i(t-1) > V_{DD} \\ 1 & \text{else} \end{cases} \quad (4.3)$$

$$f_2 = \begin{cases} 0 & \text{if } u_i(t-1) < 0 \\ 1 & \text{else} \end{cases} \quad (4.4)$$

$$y_i(t) = \begin{cases} 0 & \text{if } u_i(t-1) - z(t-1) < V_{TH} \\ \frac{\beta_{M1}}{2} (u_i(t-1) - z(t-1) - V_{TH})^2 & \text{if } u_i(t-1) - z(t-1) > V_{TH} \end{cases} \quad (4.5)$$

$$I_{TAIL}(t) = \begin{cases} I_{T-MIN} & \text{if } z(t-1) < V_{TH} \\ r \frac{\beta_{M5}}{2} (z(t-1) - V_{TH})^2 & \text{if } z(t-1) > V_{TH} \end{cases} \quad (4.6)$$

$$z(t) = z(t-1) + \frac{\sum_{i=1}^n y_i(t) - I_{TAIL}(t)}{C_{CSN}/dt} \quad (4.7)$$

C_{EVAL} and C_{CSN} represent the capacitances associated with the evaluation-nodes and common-source-node, respectively. I_{T-MIN} is the tail current provided by the sub-circuit M17-M19, r is a constant to be defined in (4.8).

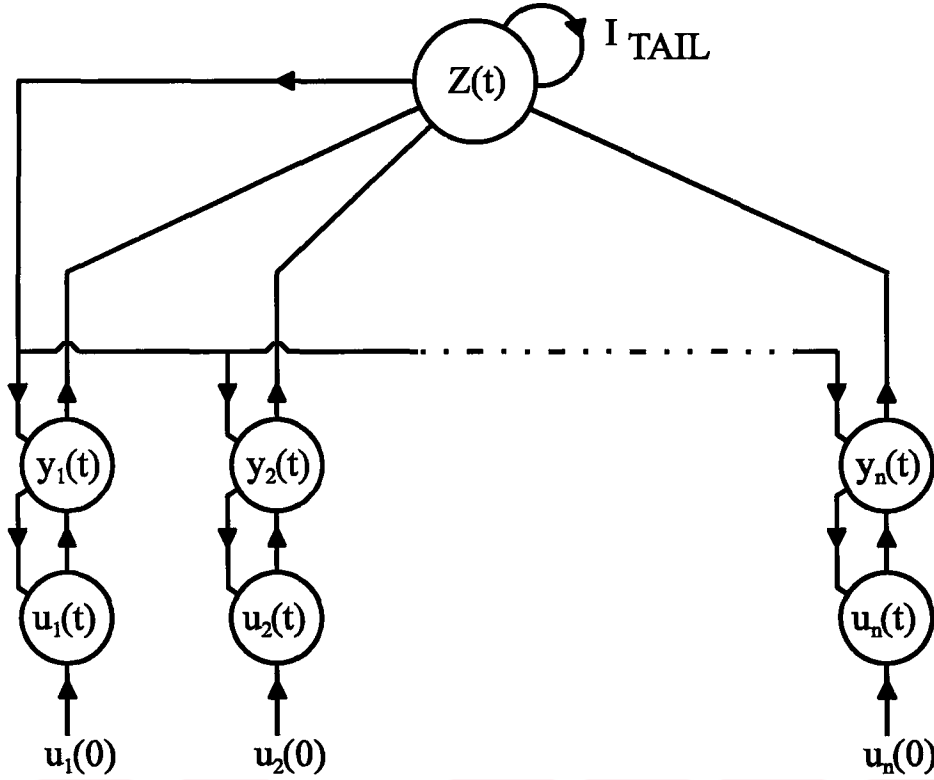
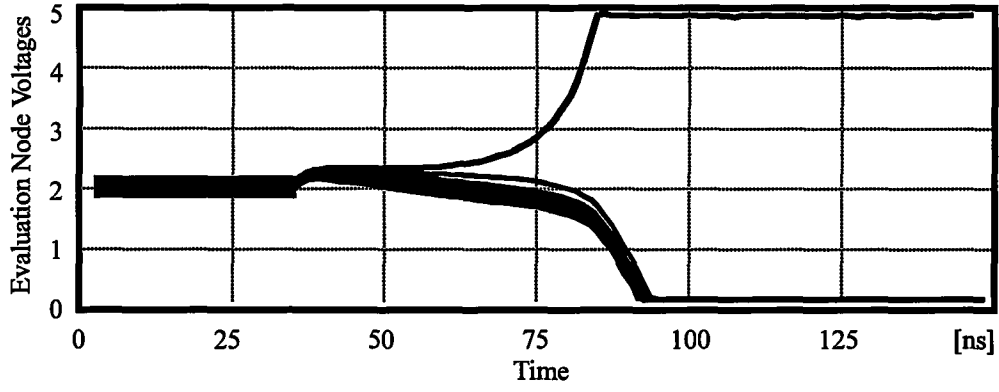
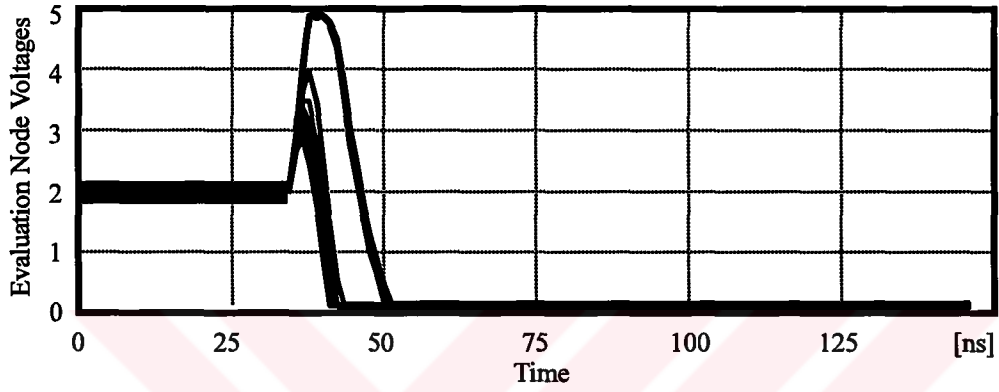


Fig. 4.3. General description of the proposed circuit configuration.

The current forwarded to the common-source-node is equal to the difference between the sum of all currents through M1's and I_{TAIL} . The currents conducted by M1's are expressed in (4.5), where β_{M1} is the transconductance parameter of M1. At the end of each step, the differential voltage resulting from charging or discharging of C_{CSN} adds to the previous level of the common-source-node voltage represented by (4.7). Notice that $z(0)=0$. Many simulations based on this algorithm (see Appendix D) were carried out for different values of circuit parameters and initial input activations, as shown in Fig 4.4. Agreement between the results of those simulations and the HSPICE simulated transient responses verified the validity of the model as although many secondary effects were disregarded.



a)



b)

Figure 4.4 Simulated evaluation node voltages of 10-cell WTA circuit for initial evaluation node voltages a) $V_i = 2V + i \times 10mV$, $1 \leq i \leq 10$; $C_{EVAL} = 400fF$; b) $C_{CSN} = 800fF$. $V_i = 2V + i \times 10mV$, $1 \leq i \leq 10$; $C_{EVAL} = 100fF$, $C_{CSN} = 800fF$.

Oscillatory behavior of the waveforms verifies the importance of the ratio between C_{EVAL} and C_{CSN} as explained in Section 4.3.

4.2 Resolution

As discussed earlier, positive feedback is an effective tool to enhance the discrimination capability of WTA circuits. Yet, an excessive amount of positive feedback may cause oscillations or multiple winner selection. The proposed circuit automatically limits the amount of positive feedback applied, keeping the ratio, r , of the tail current to the discharging current constant. If this ratio is less than 1, the circuit cannot select any winner, on the other hand, if it is greater than 2, the tail current becomes high enough to supply excessive charging current to the cells following the winner, leading the circuit to select more than one winner. Then 1.5 seems to be the most appropriate value for this ratio. If M5 of the cell and M10 of the current source are identical, then ignoring the current through M6 and the channel length modulation effect, r is given by,

$$r = \frac{\left(\frac{W}{L}\right)_{M12} \left(\frac{W}{L}\right)_{M14} \left(\frac{W}{L}\right)_{M4}}{\left(\frac{W}{L}\right)_{M11} \left(\frac{W}{L}\right)_{M13} \left(\frac{W}{L}\right)_{M3}} \quad (4.8)$$

The aspect ratio (W/L) of M6 has to be selected small enough to ensure that its current has a negligible effect on the total charging current.

4.3 Dynamic Analysis

Referring to the simplified schematic of a cell shown in Fig. 4.5, the dynamic characteristics of the circuit can be explained. C_{EVAL} is the evaluation-node capacitance and C_{CSN} is the total common-source-node capacitance. During the charging period M2's and consequently, M1's are off. The tail current discharges the common-source-node capacitance down to the ground potential. The initial evaluation voltages dumped into the circuit may have values up to the supply voltage. When M2's are turned on, the large voltage difference appearing between the gate and the source node of M1's of highly activated cells may cause large currents through these transistors. Suppose that input voltages of highly activated m cells are set at the vicinity of a common voltage level V_a . At the beginning of the selection phase the gate overdrive voltage of M1's become $V_a - V_{TH}$. If V_a is high, M1's begin to flow a large drain current charging C_{CSN} . After the charging of C_{CSN} , the current through each individual M1 decreases down to I_{TAIL}/m . The voltage difference between the gate and the source node of M1's can be assumed to be equal to $V_{TH}(V_{CSN})$.

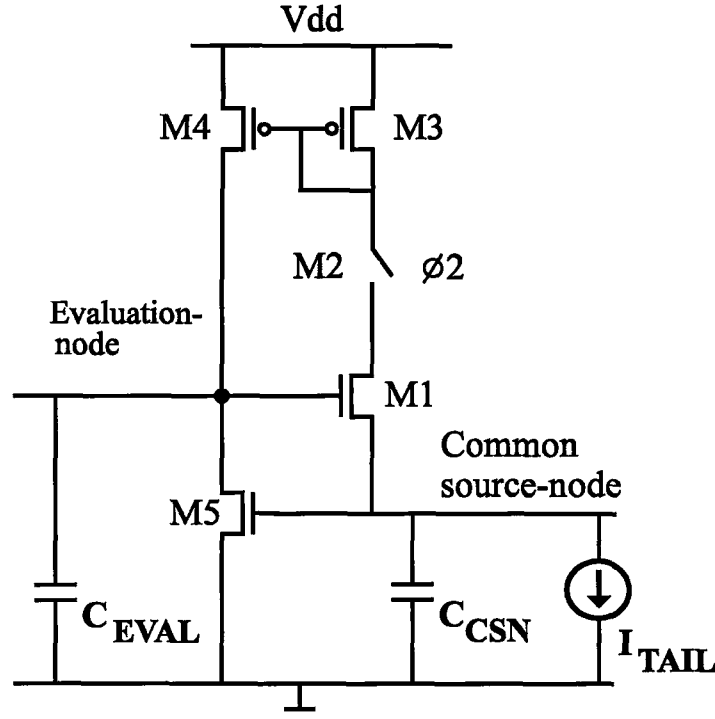


Figure 4.5 Simplified circuit diagram of a cell.

However, during this process the voltage of the evaluation-nodes increases from V_a to V_b , due to the charging current supplied to C_{EVAL} through the current mirror M3-M4. Beyond this point, the selection process enters its steady state (M1's act as source follower). In order to preserve initial voltage differences between evaluation-node of different cells, M4's should always stay in saturation. The amount of charge injected to C_{CSN} become approximately equal to $(V_b - V_{TH}) C_{CSN}$, while this injection causes each C_{EVAL} (of the activated cells) to be dumped with a charge of $\frac{(W/L)_{M4}}{m(W/L)_{M3}} C_{CSN} (V_b - V_{TH})$. The gate to source capacitance of M1, C_{gsM1} ,

can be seen as a feed-forward capacitance. The initial and final voltage values appearing across this capacitance are $(V_a - 0)$ and $[V_b - (V_b - V_{TH})]$. The charge transferred to the evaluation-node becomes $(V_a - V_{TH}) C_{gsM1}$. Based on the charge conservation theory V_b can easily be determined as

$$V_b = \frac{(C_{EVAL} + C_{gsM1})V_a - \left(\frac{(W/L)_{M4}}{(W/L)_{M3}} \frac{C_{CSN}}{m} + C_{gsM1} \right) V_{TH}}{C_{EVAL} - \frac{(W/L)_{M4}}{(W/L)_{M3}} \frac{C_{CSN}}{m}} \quad (4.9)$$

Note that this equation is valid only if $C_{EVAL} > \frac{(W/L)_{M4}}{(W/L)_{M3}} \frac{C_{CSN}}{m}$ due to the fact that during the analysis V_b is assumed to have a finite value. Yet, it is clear that if this case is not valid the positive feedback would make impossible to limit V_b . (4.9) is more critical for large values of V_a and consequently, V_b . Both C_{EVAL} and V_a are greater than $\frac{(W/L)_{M4}}{(W/L)_{M3}} \frac{C_{CSN}}{m}$ and V_{TH} , respectively. Hence the second term of the dividend in (4.9) can be neglected. Arranging (4.9) we get,

$$V_b = \frac{1 + \frac{C_{gsM1}}{C_{EVAL}}}{1 - \frac{(W/L)_{M4}}{(W/L)_{M3}} \frac{C_{CSN}}{mC_{EVAL}}} V_a \quad (4.10)$$

For a stable system the denominator should be positive. Hence the stability criteria can be set as

$$\frac{C_{EVAL}}{C_{CSN}} \frac{(W/L)_{M3}}{(W/L)_{M4}} > \frac{1}{m} \quad (4.11)$$

The smaller m , the harder the satisfaction of (4.11). The smallest possible value for m is 1. Yet, this also implies that only one cell is considerably activated and selection of one winner is ensured. Hence, the most critical value of m becomes 2.

$$\frac{C_{EVAL}}{C_{CSN}} > \frac{(W/L)_{M4}}{2(W/L)_{M3}} \quad (4.12)$$

For a given maximum common input voltage level, the design parameters, $(W/L)_{M4}$, $(W/L)_{M3}$, C_{CSN} and C_{EVAL} , should be selected to keep V_b below the limit, above which $M4$ enters the nonsaturated region. If this condition is not satisfied, input difference voltages decrease during the transient effects stemming from C_{CSN} and the discrimination process begins with lower difference voltages. As a consequence the precision of the selection process becomes much more prone to device mismatch and the response of the circuit is delayed. Note that, the larger the input difference voltages, the faster the circuit responds. Fig 4.6 depicts the change in the evaluation-node voltages during this transient phase. In order to increase the common mode input range, the aspect ratio of $M4$ should be selected large. Consequently, the saturation limit of its drain-to-source voltage, V_{dsSAT} , is decreased.

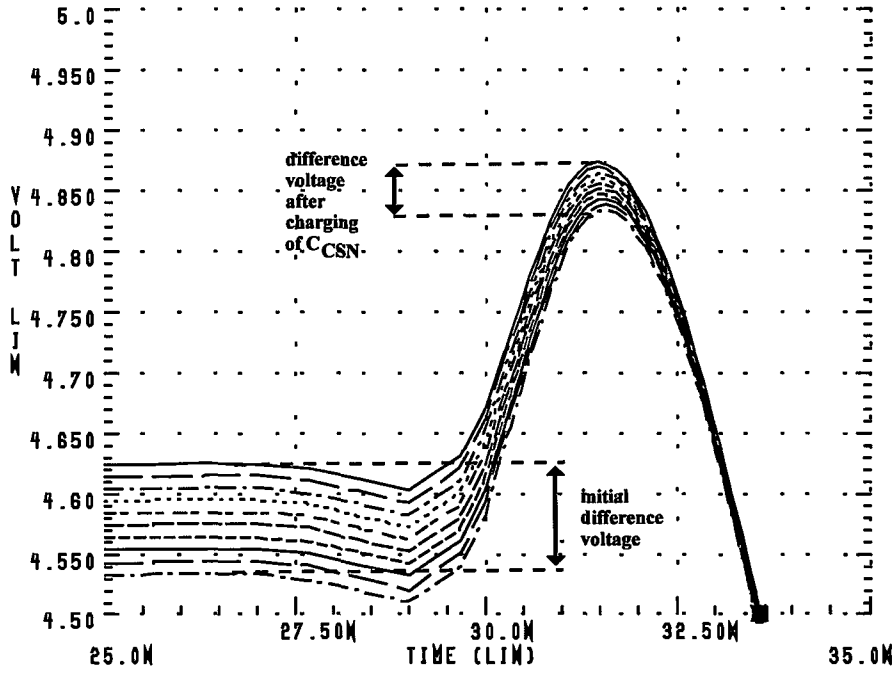


Figure 4.6 Evaluation-node voltages during the charging of C_{CSN} .

This also causes M3 and M4 to operate with low gate overdrive voltage, $(V_{GS} - V_{TH})$. On the other hand, the current mismatch modeled using the formula

$$\frac{\sigma^2(I_d)}{I_d^2} = \frac{4}{(V_{GS} - V_{TH})^2} \frac{A_{V_{TH}}^2}{WL} + \frac{A_{\beta}^2}{WL} \quad (4.13)$$

implies large current discrepancies among M4's of different cells having low gate overdrive voltages. This trade-off can be overcome by selecting $(V_{GS} - V_{TH})$ small, but the gate area, WL, large. This countermeasure would not only impair the high speed characteristics, but would also cause large real estate consumption.

4.4 Implementation and Test Results

An experimental circuit having 10 cells was designed and fabricated in YITAL $3\mu\text{m}$, 5V, n-well, single-poly, single-metal, CMOS technology. Fig. 4.7 shows photomicrograph of the chip.

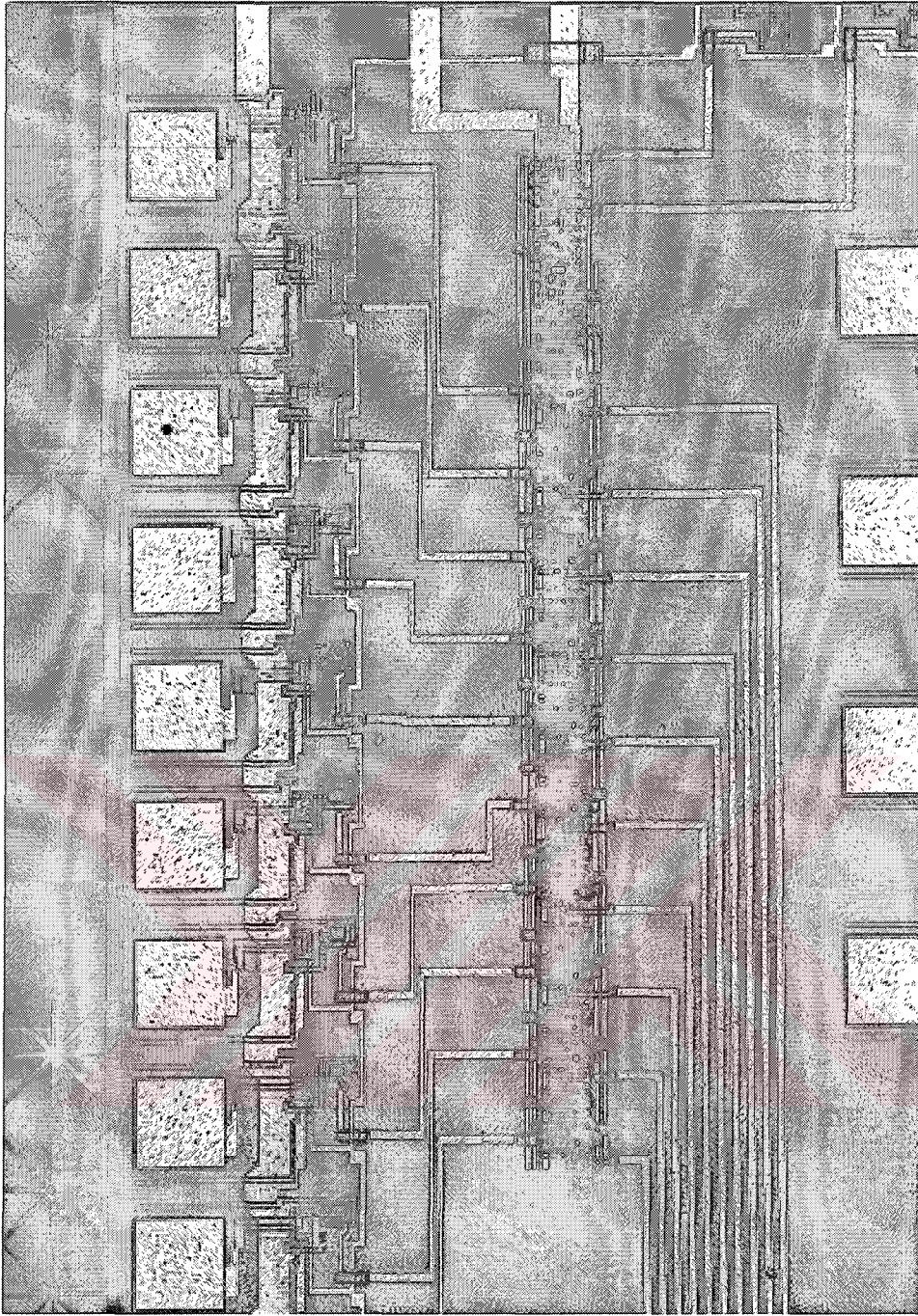


Figure 4.7 Photomicrograph of the chip.

Each cell occupies $120\mu\text{m} \times 120\mu\text{m}$. Buffers composed of a cascade of 2 inverters are used to drive output pads. First stage inverters are embedded in cells in order to minimize and equalize input capacitances. Therefore, only digitized output values are measurable. Although power consumption of the circuit is dependent on input activation distribution, an average level of 7 mW can be given for an operation at 5MHz.

4.4.1 Offset voltages

Table 4.2 shows the results of the offset voltage measurements. The whole input voltage range extending from ground to the positive supply rail was examined for all cells. All input voltages were set to a common voltage level. Then the input voltage of Cell-5 was increased by 200 mV to make it the winning cell. Other input voltages were increased sequentially until the corresponding cell became the winner. At this stage, the difference between input voltages was recorded as the offset voltage of the winning cell with respect to Cell-5. The reason of high offset voltages at small and large input voltage region had been explained in previous sections.

Table 4.2 Offset voltage distribution among cells, values given in mV.

Cell	1	2	3	4	5	6	7	8	9	10
5V→	56	-25	-53	35	0	72	25	-30	-19	37
4V→	12	-2	-12	22	0	35	18	-21	-13	22
3V→	11	3	25	15	0	8	10	-12	3	16
2V→	9	21	24	12	0	4	-3	8	-4	11
1V→	30	-13	17	20	0	17	-23	11	-19	8
0V→	55	-32	60	50	0	82	-41	65	-35	47

4.4.2 Transient response

Input voltage of 8 cells were set to -10 mV and only two cells were activated. In this case, the circuit acts as a two-port differential charge-sensing amplifier that is conventionally used in DRAM's. Transient responses were observed for 10 possible combination of randomly selected cell pairs. A typical transient response of the circuit is shown in Fig.4.8. Dependence of response time on the input difference voltage is illustrated in Fig. 4.9. In order to measure the response of the circuit for various differential voltages, all input voltages were set to 2V. Then, the input voltage of Cell-4 was increased by 250 mV to make it a clear winner. Afterwards, the input voltage of Cell-6 was increased slowly until it became the new winner. It is assumed that the differential voltage between the inputs of Cell-4 and Cell-6 was zero where the transition (winner Cell-4 → winner Cell-6) occurred. Beyond this point, the input voltage of Cell-6 was increased by 40 mV steps and the delays

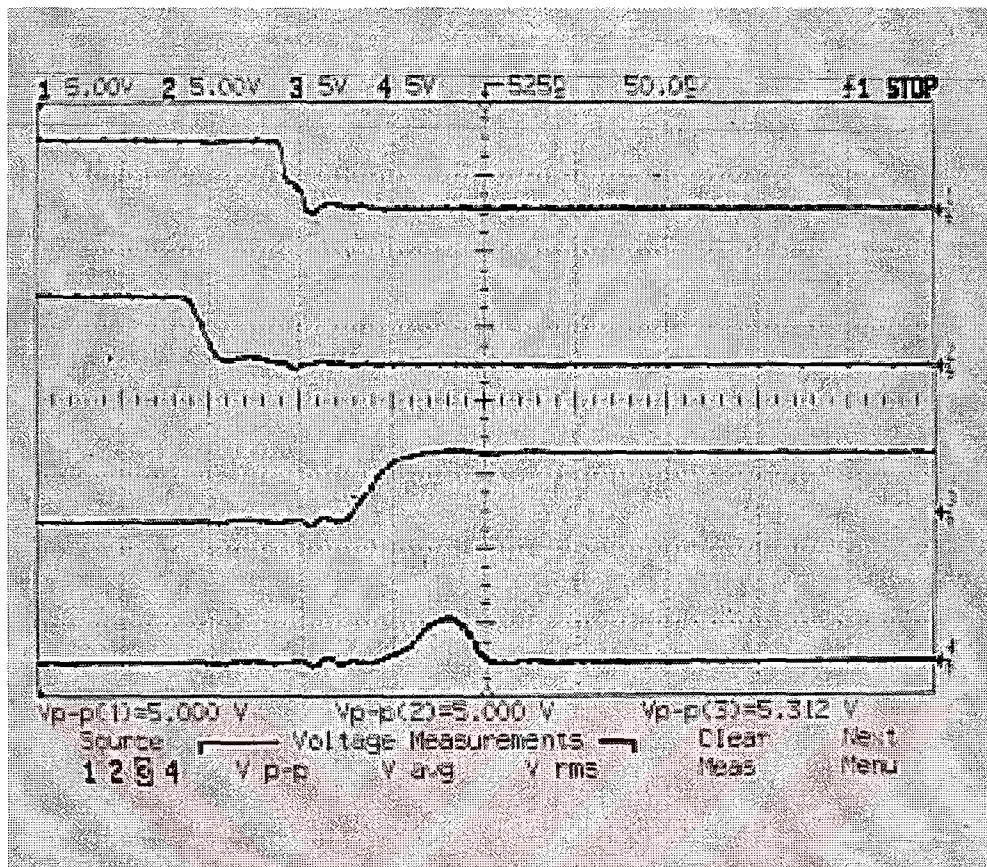


Figure 4.8 Experimental waveforms. The first and second traces belong to control signals ϕ_1 , ϕ_2 , respectively, the third and the forth traces show the outputs of the winning cell and its successor. The input voltage of the winning cell and its successor are 1.010V and 1V respectively. All other inputs are grounded.

associated with difference voltages (40mV, 80mV, . . .) were recorded. The circuit selected only one winner for each input vector set applied during the tests and showed no evidence of instability.

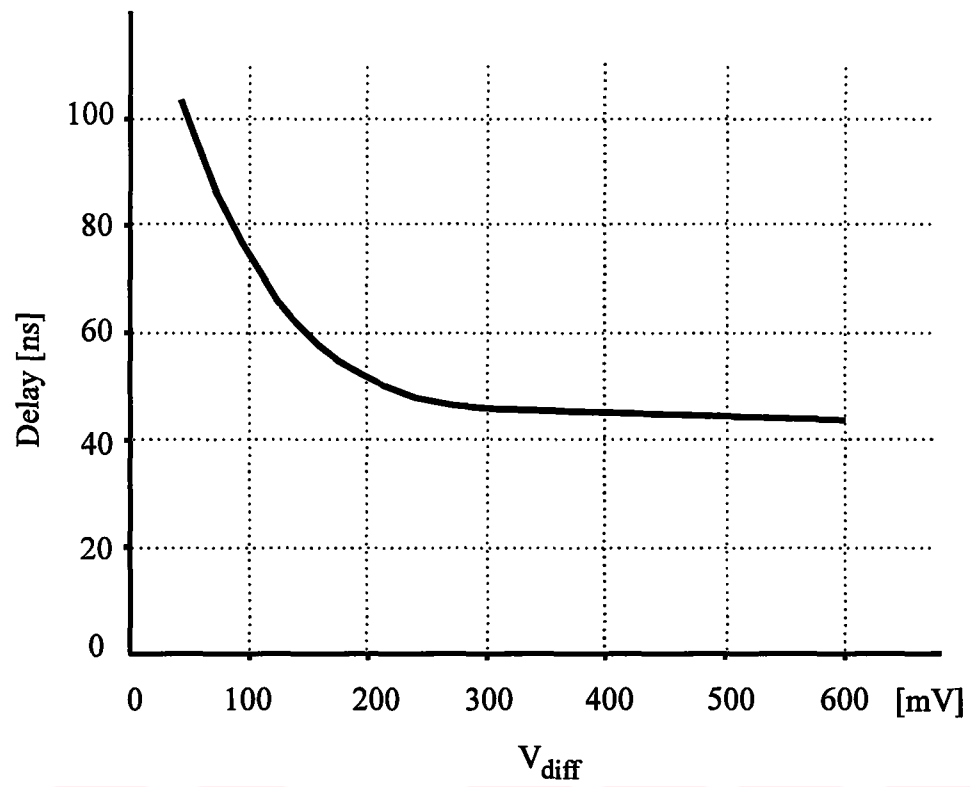


Figure 4.9 Measured effect of input difference voltage on response time.

5. THE PROPOSED KWTa CIRCUIT

5.1 Circuit Description and Operation

The architecture of the proposed KWTa circuit and aspect ratios of transistors are shown in Fig. 5.1 and Table 5.1, respectively. It consists of n identical cells together with two independent externally adjustable current sources common to all the cells. Fig. 5.2 depicts the simplified circuit diagram of a cell. Each classification cycle consists of three consecutive phases (charging, selection and stabilization) determined by the voltage waveforms ϕ_1 , ϕ_2 , ϕ_3 . During the charging phase, the initial input voltages are dumped onto the circuit. The circuit selects the winning cells during the second step called the selection phase. The last step is the stabilization phase where the competition among cells is suppressed and the circuit is made stuck in a stable state.

During the charging phase, the gate voltage of M17 is pulled down, disabling the drain currents of PMOS transistors M3's and M5's and forcing the evaluation-nodes into high impedance state. Then the transmission gates consisting of M10 and M11 are turned on. The initial input voltages are dumped onto the equal evaluation-node capacitances and those voltages serve as the initial gate voltages of the NMOS transistors, M1's. This phase is terminated by both M17 and the transmission gates being turned off. At this point, M3's and M5's begin to source their drain currents, I_{M3} and I_{M5} , to the cells and the selection phase starts. The source terminals of all M1's are tied together and the sum of their drain currents is equal to the tail current, I_{TAIL} , provided by M13. The higher the evaluation-node voltage (gate voltage of M1) of a cell, the larger the portion of the tail current steered toward its M1. Drain current of M1 is provided by M3, hence, M1 sinks a part of I_{M3} . The residual current is conducted through M2 then mirrored by M4 and M6 and transferred by M8 to the evaluation-node as discharging current. Consequently, the cells having higher input voltages have smaller discharging currents. Their evaluation-node voltages are pulled up further by I_{M5} . On the other hand, only a small part of I_{M3} is absorbed by M1 in the cells having low evaluation-node voltages. The residual part transferred through M2, M4, M6 and M8 strives against I_{M5} to pull the evaluation-node voltages further towards ground. The available discharging current $I_{M3} (W/L)_{M6} / (W/L)_{M4}$ should be

higher than the total charging current I_{M5} . This requires that the constant α defined by

$$\alpha = \frac{(W/L)_{M5}(W/L)_{M4}}{(W/L)_{M3}(W/L)_{M6}} \quad (5.1)$$

be lower than unity. During the selection phase, M13 and M3's should operate in the saturation region to supply predetermined charging and discharging currents. Otherwise, instead of staying in the metastable region, the circuit tends to decrease I_{TAIL} or I_{M3} , until it sticks to a permanently stable state. M7's and M8's are implemented to force the circuit to stay away from this uncontrollable stance.

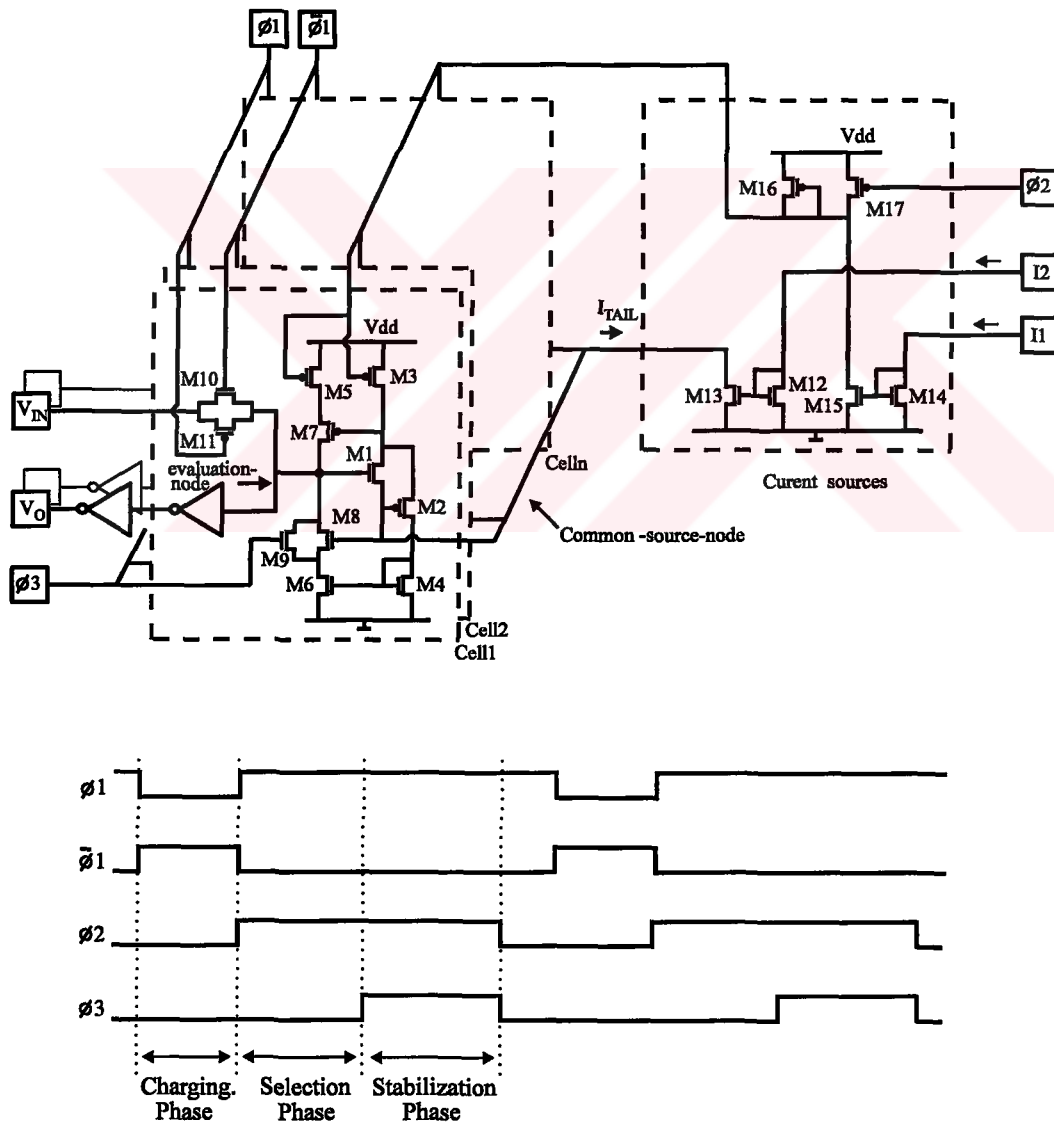


Figure 5.1 Architecture of the proposed KWTA circuit.

Table 5.1 Transistor dimensions.

Transistors	W [μm]	L [μm]
M1	9	5
M2	17	4
M3, M16	21	4
M4, M6	11	5
M5, M7	7	4
M8, M9	13	5
M10, M11	4	5
M12 - M15, M17	31	5

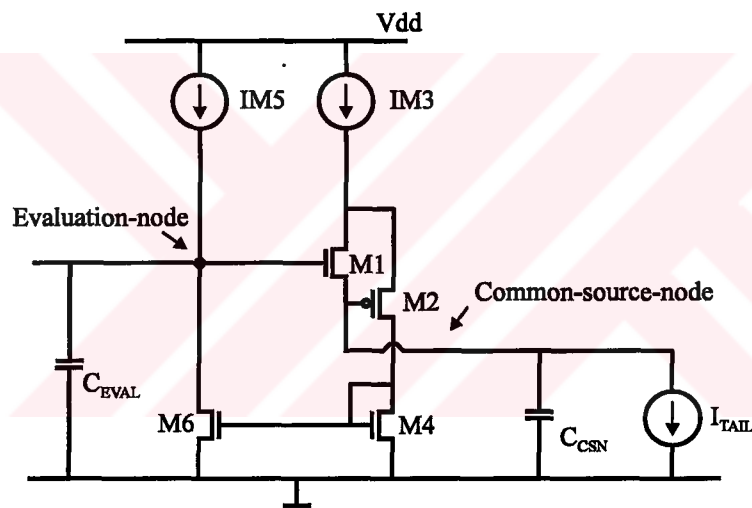


Figure 5.2 General description of the cell configuration.

If the sum of currents through all M1 transistors is higher than I_{TAIL} , the common-source-node voltage increases. If this upward swing continues, M5's enter the nonsaturated region before M3's, due to the voltage shifts towards VDD caused by the gate-to-source voltages of M7's. Then I_{M5} 's decrease while I_{M3} 's remain unchanged. The evaluation-node voltages go towards ground pulling the common-source-node voltage down. On the other hand, if the sum of the currents flowing through M1's is less than I_{TAIL} , the common-source-node voltage decreases. Since M9's are off during the whole selection phase, M8's keep the drain voltages of M6's

lower than the drain voltage of M13 by their gate-to-source voltages. Therefore, M6's enter the nonsaturated region before M13. The discharging currents decrease, the evaluation-node voltages increase thus the common-source-node is pulled high. This negative feedback loop forces the sum of the currents flowing through all M1's to be equal to I_{TAIL} . At the end of the selection period, the evaluation-node voltages of the winning cells go towards VDD and their M1's transfer the entire drain current of their M3's to the common-source-node. If the tail current is set to

$$I_{TAIL} = (k + r) I_{M3} \quad (5.2)$$

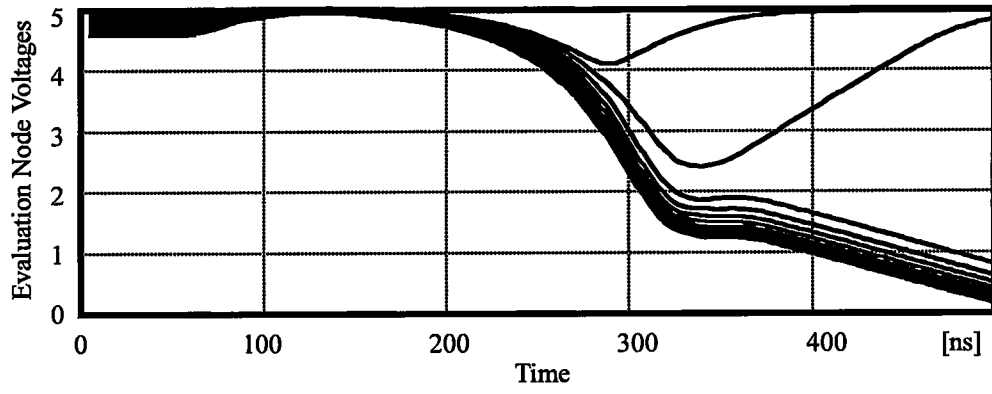
where k is an integer and $0 \leq r < 1$, the process stated above ensures selection of k cells having higher initial evaluation-node voltages. The residual current, $I_{RES} = r I_{M3}$, is steered towards M1's of the most activated successors of the k cells selected before. Another winner selection process takes place among losing cells, after which M1 of the cell having $(k+1)$ 'th highest evaluation-node voltage conducts all of the residual current. Yet, its evaluation-node voltage depends on the residual tail current which is difficult to determine, especially for high k . Therefore, this secondary selection process must be suppressed.

In order to force the circuit to stay in a stable state after the selection of k winners, M9's clocked to ϕ_3 are turned on and the stabilization phase starts. This shorts the sources and drains of M8's, enabling the common-source-node voltage to decrease until M13 enters the nonsaturated region and its drain current decreases to $I_{TAIL} = k I_{M3}$. In this case, the entire I_{M3} is transferred by M2, M4, M6 and M8 to the evaluation-node in the losing cells since their M1's are off. Their evaluation-node voltages are pulled down to a voltage level where M6's operate in the linear region and their nonsaturated drain currents become equal to I_{M5} . Equating nonsaturated drain current expression of M6 to I_{M5} and rearranging the terms, output voltage level of losing cells appearing at the evaluation-nodes can easily be obtained as

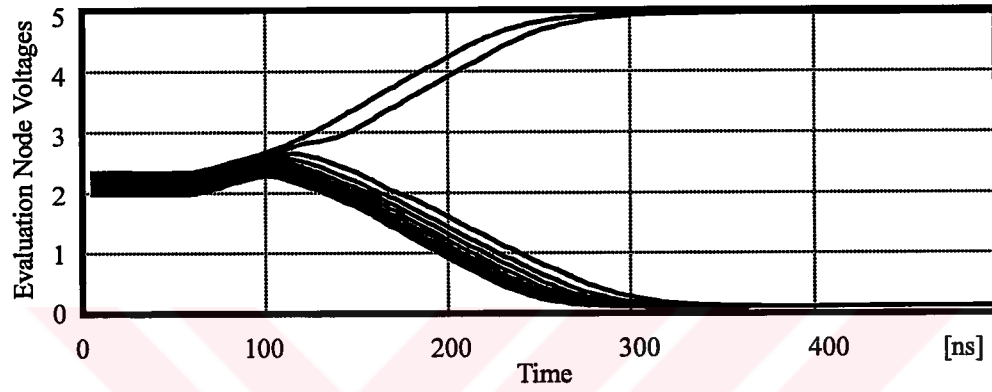
$$V_{LOW} = \sqrt{\frac{2I_{M3}}{\beta_{M6}}} \left(1 - \sqrt{1 - \alpha} \right) \quad (5.3)$$

where β_{M6} represents the transconductance parameter of M6 and α is the constant defined by (5.1).

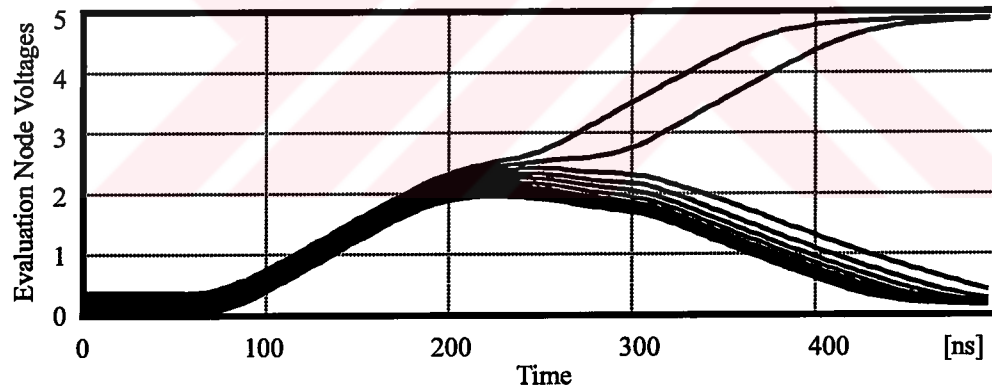
Post-layout HSPICE simulations were carried out to verify the performance of the proposed circuit. Typical simulation results are shown in Fig. 5.3 and 5.4.



a)



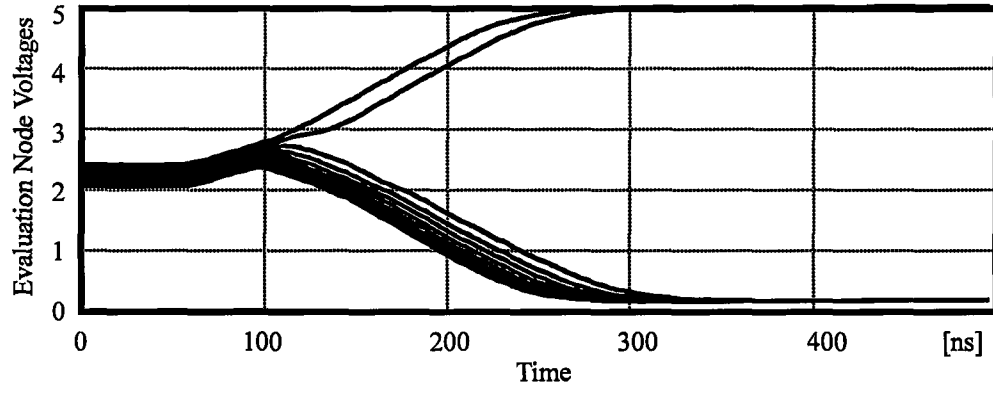
b)



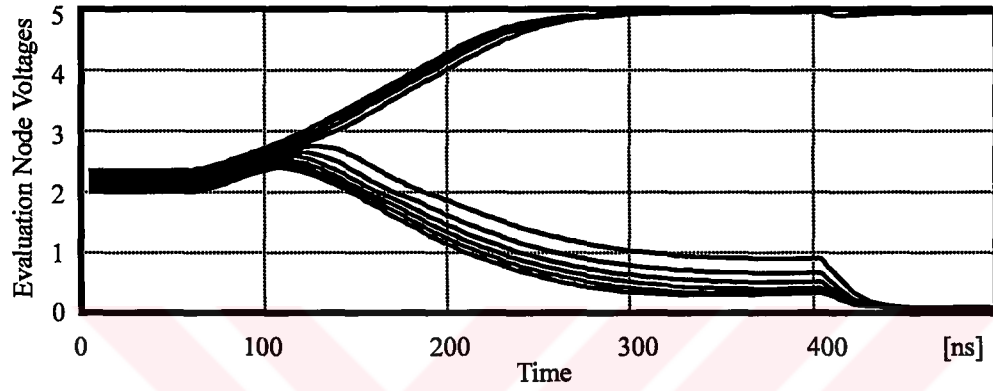
c)

Figure 5.3 The HSPICE simulated evaluation node voltages of 10-cell KWTa circuit for initial evaluation node voltages a) $V_i = 5V + i \times 40mV$, $1 \leq i \leq 10$; b) $V_i = 2V + i \times 40mV$, $1 \leq i \leq 10$; c) $V_i = 0V + i \times 40mV$, $1 \leq i \leq 10$; $I_{TAIL}/I_{M3} = 2.5$

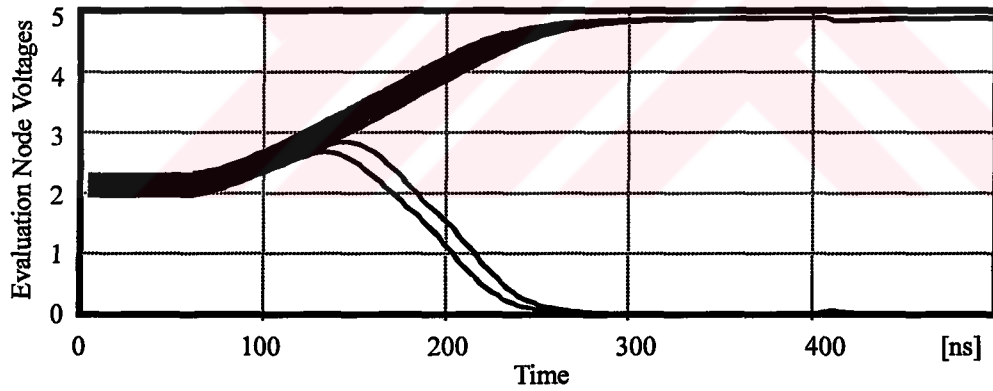
Transistor parameters provided by YITAL for its 3 μm nwell technology were used in all simulations (see Appendix C). The circuit was simulated for various sets of input vectors and for different common voltage levels.



a)



b)



c)

Figure 5.4 The HSPICE simulated evaluation node voltages of 10-cell KWTa circuit for initial evaluation node voltages $V_i = 5V + i \times 40mV$, $1 \leq i \leq 10$;
a) $I_{TAIL}/I_{M3}=2.5$; b) $I_{TAIL}/I_{M3}=4.5$; c) $I_{TAIL}/I_{M3}=2.5$ $I_{TAIL}/I_{M3}=8.5$

The operation of the circuit relies on the positive feedback which strive against the inhibitory current, I_{M3} . Its role is to increase the share of the tail current allotted to M1 of a cell up to $I_{M3} = I_{TAIL} / k$, by increasing its evaluation-node voltage. At the end of the selection phase, only k cells can be winners since there is no sufficient amount of tail current for more cells. The general description of the circuit

configuration is shown in Fig. 5.5. n initial activations $u_1(0), \dots, u_n(0)$, are fed into the circuit at time zero.

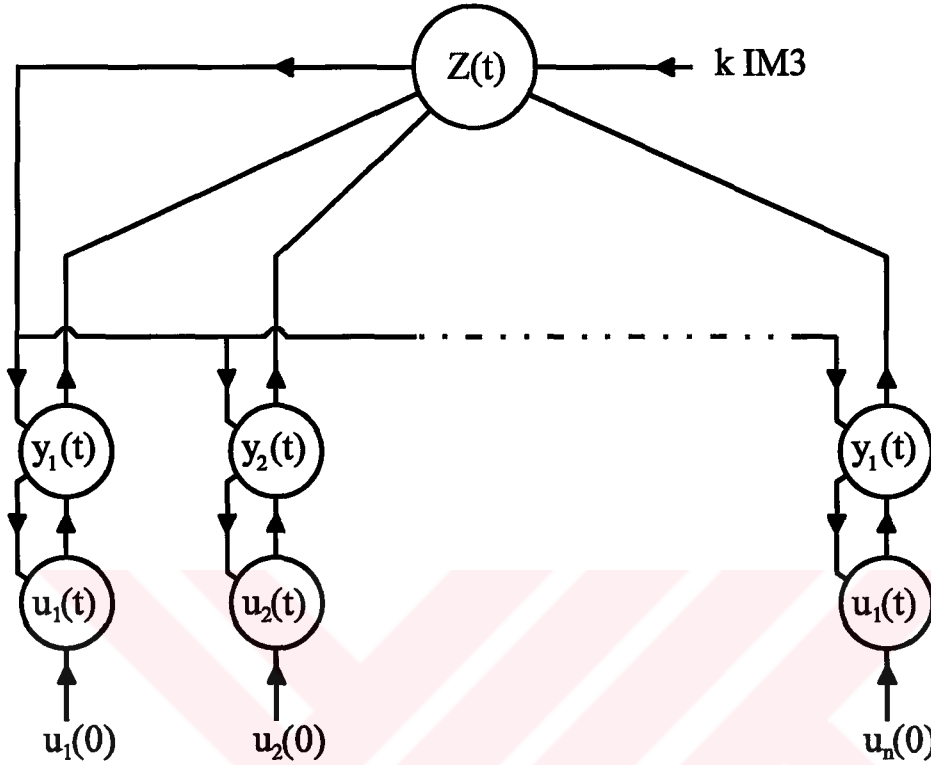


Figure 5.5. General description of the proposed circuit configuration.

Activation of the cells, standing for voltages at evaluation-nodes, for $t > 0$ are given by

$$u_i(t) = u_i(t-1) + \frac{f_1 I_{M5} - f_2 (I_{M3} - I_i(t))}{C_{EVAL}/dt} \quad (5.4)$$

where,

$$f_1 = \begin{cases} 0 & \text{if } z(t-1) > V_{DD} - V_{TH-M7} \text{ or } u_i(t-1) > V_{DD} \\ 1 & \text{else} \end{cases} \quad (5.5)$$

$$f_2 = \begin{cases} 0 & \text{if } z(t-1) < V_{TH-M8} \text{ or } u_i(t-1) < 0 \\ 1 & \text{else} \end{cases} \quad (5.6)$$

$$I_i(t) = \begin{cases} I_{M3} & \text{if } \sqrt{2I_{M3}/\beta_{M1}} < u_i(t-1) - z(t-1) \\ \frac{\beta_{M1}}{2}(u_i(t-1) - z(t-1))^2 & \text{if } 0 < u_i(t-1) - z(t-1) < \sqrt{2I_{M3}/\beta_{M1}} \\ 0 & \text{if } u_i(t-1) - z(t-1) < 0 \end{cases} \quad (5.7)$$

$$z(t) = z(t-1) + \frac{\sum_{i=1}^n I_i(t) - I_{TAIL}}{C_{CSN}/dt} \quad (5.8)$$

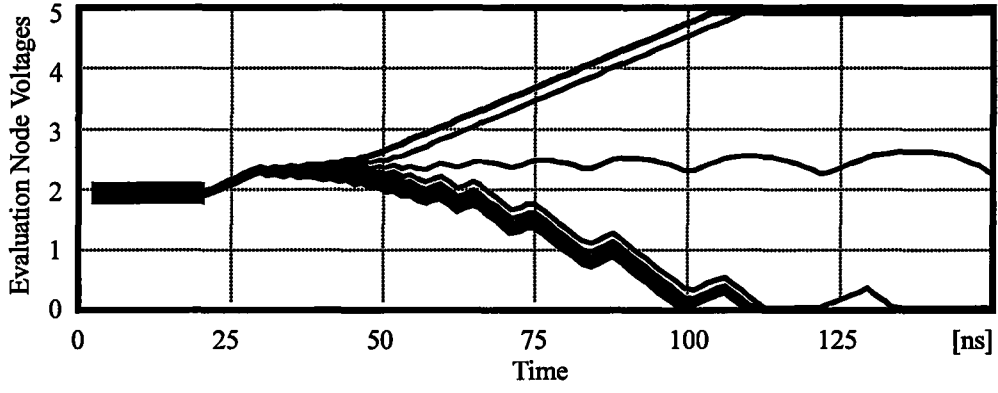
C_{EVAL} and C_{CSN} represent the capacitances associated with the evaluation-nodes and common-source-node respectively. V_{TH-M7} and V_{TH-M8} are the threshold voltages of transistors M7 and M8. First bounds in if blocks indicate the behavior of M7 and M8 in (5.5) and (5.6). Second bounds simulate the decrement of the pull-up and pull-down currents as the drain-to-source voltages of M5's and M6's decrease. The current forwarded to the common-source-node is equal to the difference between the sum of all currents through M1's and I_{TAIL} . The currents conducted by M1's are expressed in (5.7), where β_{M1} is the transconductance parameter of M1. At the end of each step, the differential voltage resulting from charging or discharging of C_{CSN} adds to the previous level of the common-source-node voltage represented by (5.8). Notice that $z(0)=0$. Many simulations based on this algorithm (see Appendix E) were carried out for different values of circuit parameters and initial input activations, as shown in Fig. 5.6. Agreement between the results of those simulations and the HSPICE simulated transient responses verified the validity of the model although many secondary effects were disregarded.

5.2 Low Input Voltage Region

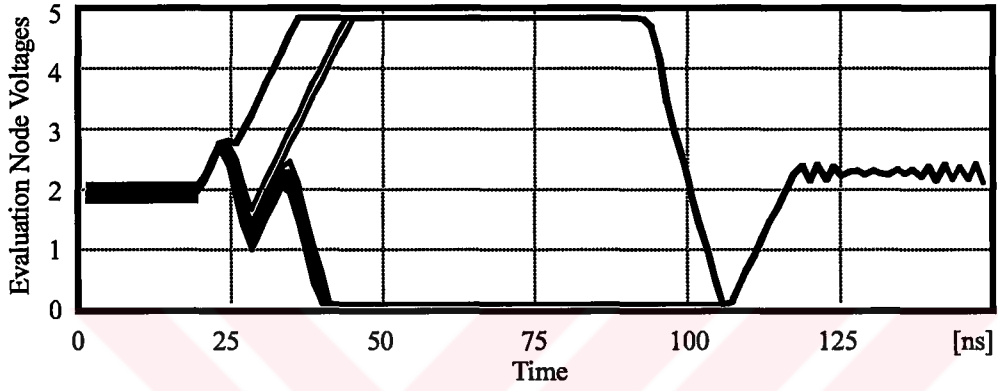
In the proposed circuit, the positive feedback loops are active if M1's and M8's are conductive. Yet, this is not always the case. When all the evaluation-node voltages are below the threshold voltage V_a which can be defined as

$$V_a = V_{TH-M8} + V_{TH-M1} \quad (5.9)$$

M8's are off and constant currents flowing through M5's and M7's charge the evaluation-node capacitances linearly. Here, V_{TH-M8} and V_{TH-M1} denote the threshold voltages of M8 and M1. If the charging currents and capacitances were identical in



a)



b)

Figure 5.6 Simulated evaluation node voltages of 10-cell WTA circuit for initial node voltages a) $V_i = 2V + i \times 10mV$, $1 \leq i \leq 10$; $I_{TAIL}/I_{M3}=2.5$, $C_{EVAL}=400fF$, $C_{CSN}=800fF$; b) $V_i = 2V + i \times 10mV$, $1 \leq i \leq 10$; $I_{TAIL}/I_{M3}=2.5$, $C_{EVAL}=100fF$, $C_{CSN}=800fF$. Oscillatory behavior of the waveforms verifies the importance of the ratio between C_{EVAL} and C_{CSN} , as explained in Section 5.3.

all cells, the cell having the maximum initial input voltage would be the first surpassing V_a . However, device mismatches may impel some cells to dominate over the most activated cell. Now, suppose that all evaluation-nodes are set to V_c ($0 < V_c < V_a$) and all cells are identical, except for Cell-i, M5 of which provides p percent higher charging current than that of others. In this case, when the evaluation-node voltage of Cell-i reaches V_a , all other evaluation-node voltages become

$$V_d = V_a - (V_a - V_c) (p/100) \quad (5.10)$$

This means that in order to have all cells reach V_a at the same time, the initial voltage applied to Cell-i should be lower than others by

$$V_{off} = (V_a - V_c) (p/100) \quad (5.11)$$

The developed offset model implies a linear dependence of offset voltage on the difference between V_a and the input common voltage level. This implication agrees well with the measurement results between 0 and 2 V to be given in Section 5.4. In order to estimate the order of the magnitude of V_{off} , consider the case of $V_a=2V$, $V_c=0V$, $p=1\%$. According to (5.11), V_{off} is expected to be 20 mV. Other mismatches such as discrepancies in C_{EVAL} 's can also be represented by the factor p . For a given technology, mismatches and consequently offset voltages can be reduced by avoidance of using transistors having minimum length or width at the expense of more real estate consumption.

5.3 Dynamic Analysis

Referring to the simplified schematic of a cell shown in Fig. 5.2, the dynamic characteristics of the circuit can be explained. C_{EVAL} is the evaluation-node capacitance and C_{CSN} is the total common-source-node capacitance. Neglecting the capacitive contribution of the tail current source, C_{CSN} can be modeled as

$$C_{CSN} = n C_{CSN-CELL} \quad (5.12)$$

where n is the number of cells and $C_{CSN-CELL}$ denotes the capacitive contribution of a cell consisting of the gate capacitances of M2 and M8, source capacitance of M1 and the capacitance of interconnection associated with a single cell. During the charging phase, M3's are turned off preventing M1's from sourcing charging current to the common-source-node while M13 discharges C_{CSN} completely. When M3's and M5's are turned on, the common-source-node voltage starts an upward swing and the charging current of its capacitance adds to the tail current. The greater the current sunk from the common-source-node, the smaller the inhibitory currents applied to evaluation-nodes and the higher the number of winners the circuit attempts to select. Now suppose that the evaluation-node voltages of all cells are approximately equal. Small differences between the gate voltages of M1's imply that the current sunk from the common-source-node be equally shared between n cells. Then the drain current of each M1 can be expressed as

$$I_{M1} = \frac{1}{n} \left(I_{TAIL} + C_{CSN} \frac{d V_{csn}}{dt} \right) \quad (5.13)$$

If the slew rate is high enough to enable M1 to sink the entire I_{M3} , M2 stays in the cut-off region. In this case, all I_{M3} 's charge C_{CSN} and since M4's and M6's are not conducting, I_{M5} 's charge C_{EVAL} 's. Although conducting, M1's act as short circuit elements and their drain currents are not controlled by their gate voltages. The discrimination process is suppressed and the rising edge of an oscillatory phase is assured. Therefore, for a stable operation $I_{M1} < I_{M3}$ is required. Now, let us examine the stability of the circuit for two cases separately. If $\frac{I_{M5}}{C_{EVAL}} > \frac{nI_{M3}}{C_{CSN}}$, the common-source-node voltage is unable to follow the evaluation-node voltages although the sum of all I_{M3} 's charges C_{CSN} . For the reasons stated above, an oscillatory rising edge becomes unavoidable. On the other hand, if $\frac{I_{M5}}{C_{EVAL}} < \frac{nI_{M3}}{C_{CSN}}$, M1's act as source followers. Ignoring the body effect, they provide approximately unity voltage gain. Consequently, the slew rate of both the common-source-node and the evaluation-node voltages can be considered to be equal. Using I_{M3} for I_{M1} in (5.13) and replacing the slew rate term with I_{M5} / C_{EVAL} we obtain

$$\frac{C_{CSN}}{C_{EVAL}} < n \frac{I_{M3} - I_{TAIL} / n}{I_{M5}} \quad (5.14)$$

Substituting (5.2) for I_{TAIL} and (5.12) for the common-source-node capacitance and also ignoring r , we get,

$$\frac{C_{CSN-CELLL}}{C_{EVAL}} < \frac{I_{M3}}{I_{M5}} \left(1 - \frac{k}{n} \right) \quad (5.15)$$

If this bound is not satisfied, the evaluation-node voltages increase until M5's enter the nonsaturated region. At this point, the slew rate of the common-source-node voltage decreases and the inhibitory currents increase. This implies a decrement at the evaluation-node voltage levels. If C_{CSN} is so large to prevent the voltage of the common-source-node to follow this decremental slope, the gate-to-source voltages of M1's reduce and these transistors are disabled. In this case, the entire I_{M3} is copied into the evaluation-node as discharging current and C_{CSN} is discharged by the tail current. This forces the cells to decrease their evaluation-node voltages until M6's enter the nonsaturated region. Obviously, for a stable operation M1's should not be off. Assuming that the current transfer ratio of the current mirror formed by M4 and M6 is unity, the stability condition for the falling edge can be expressed as

$$\frac{I_{M3} - I_{M5}}{C_{EVAL}} < \frac{I_{TAIL}}{C_{CSN}} \quad (5.16)$$

As in the previous case, insertion of (5.2) for I_{TAIL} and (5.12) for the common-source-node capacitance in (5.16) yields the bound

$$\frac{C_{CSN-CELL}}{C_{EVAL}} < \frac{k}{n} \left(\frac{1}{1 - \frac{I_{M5}}{I_{M3}}} \right) \quad (5.17)$$

If both conditions expressed in (5.15) and (5.17) are not satisfied, then the circuit sustains oscillations suppressing the discrimination process. The higher of the upper limits set in these two conditions is effective for a sustainable oscillation. Given n (total number of cells) and I_{M5}/I_{M3} , the most critical value of k is reached where this effective upper limit has a minimum value. For a higher k , the upper limit set in (5.15) decreases while for a smaller k the upper limit set in (5.17) decreases as shown in Fig. 5.7.

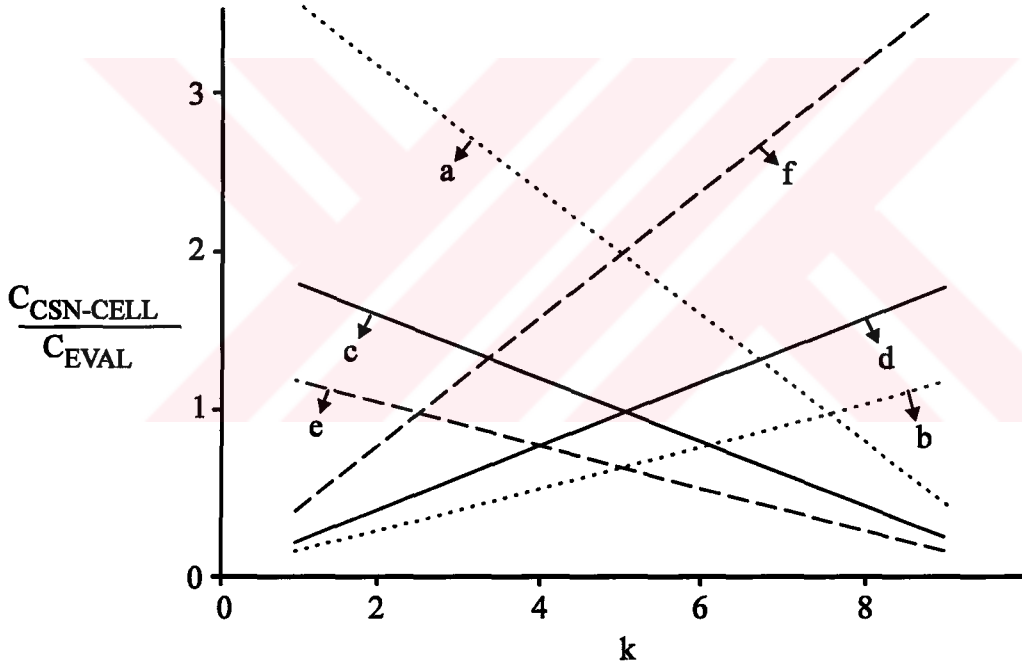


Figure 5.7. Stability chart of 10-cell circuit as a function of k for three different I_{M5}/I_{M3} values (stable operating areas are indicated by arrows).

- a) rising edge for $I_{M5}/I_{M3} = 0.25$; b) falling edge for $I_{M5}/I_{M3} = 0.25$
- c) rising edge for $I_{M5}/I_{M3} = 0.5$; d) falling edge for $I_{M5}/I_{M3} = 0.5$
- e) rising edge for $I_{M5}/I_{M3} = 0.75$; f) falling edge for $I_{M5}/I_{M3} = 0.75$

The minimum effective upper limit for $C_{CSN-CELL} / C_{EVAL}$ is reached for the value of k given by

$$k_{CRIT} = n \left(1 - \frac{I_{M5}}{I_{M3}} \right) \quad (5.18)$$

where the two bounds become equal. Inserting (5.18) in (5.15) or (5.17) we obtain

$$\frac{C_{CSN-CELL}}{C_{EVAL}} < 1 \quad (5.19)$$

which defines the bound ruling out the sustainable oscillation possibility. It does not depend on the I_{M5}/I_{M3} nor on the total cell number. On the other hand, if only one of the bounds expressed in (5.15) or (5.17) is satisfied, the circuit enters the metastable region with progressive diminution of free oscillation amplitude and then executes the discrimination process. This can be considered as a transition zone between the oscillatory and stable regions. If a double metal process is to be used, the source nodes of M1's should be connected with a narrow second metal line to minimize C_{CSN} . Another countermeasure is to increase C_{EVAL} by adding a double-polysilicon or MOS capacitor to each evaluation-node to alleviate the inevitable influence of charge injection from the analog switch consisting of M10 and M11.

5.4 Implementation and Test Results

A 10-cell experimental circuit was designed and fabricated in YITAL, 3 μ m, n-well, single-polysilicon, single-metal, CMOS technology. The photomicrograph of the KWTA chip is shown in Fig. 5.8. Each cell core occupies an area of 60 $\times$ 80 μ m². In order to isolate the evaluation-nodes from high and uneven output pad and connection capacitances, buffers consisting of two cascaded inverters are used.

The first inverter is embedded in the cell to keep the parasitic interconnection capacitance component small and equal for each cell. Therefore, only digitized output values are measurable. Although power consumption is dependent on input activation distribution, an average level of 9 mW can be given for an operation at 5MHz.

5.4.1 Offset voltages

Table 5.2 shows the results of the offset voltage measurements. The whole input voltage range extending from ground to the positive supply rail was examined for all cells. The circuit was adjusted to select only one winner and all input voltages were

set to a common voltage level. Then the input voltage of Cell-5 was increased by 200 mV to make it the winning cell. Other input voltages are increased sequentially until the corresponding cell becomes the winner. At this stage, the

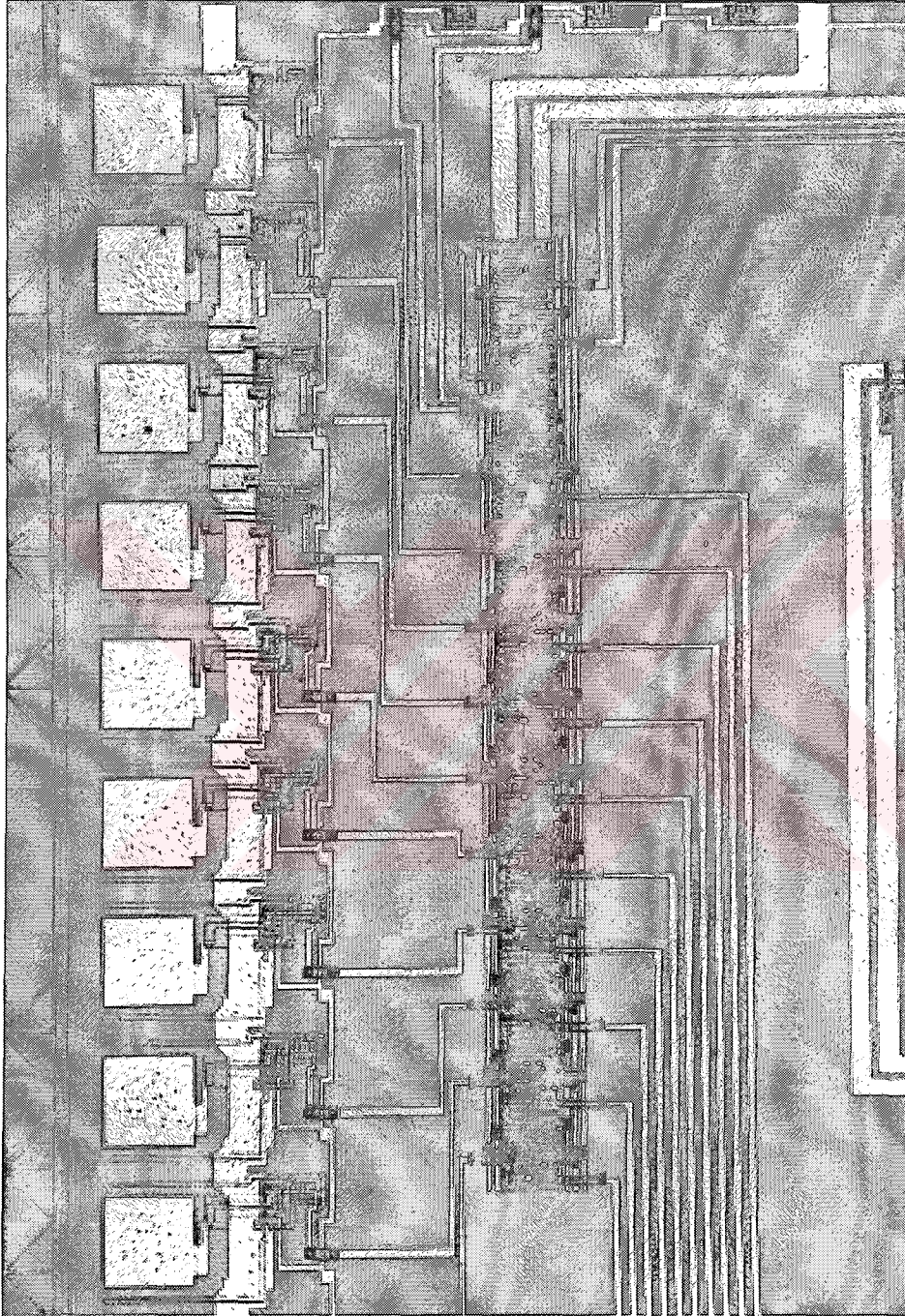


Figure 5.8 Photomicrograph of the chip.

difference between input voltages was recorded as the offset voltage of the winning cell with respect to Cell-5. The reason of high offset voltages at small input voltage region is explained in Section 5.2.

Table 5.2. Offset voltage distribution among cells, given in mV.

Cell→	1	2	3	4	5	6	7	8	9	10
5V→	12	-2	13	-15	0	-12	-4	-10	1	-21
4V→	10	-2	16	-16	0	-15	-8	-11	-1	-18
3V→	16	3	26	-13	0	-11	-10	-9	3	-18
2V→	39	31	24	-17	0	35	8	32	14	13
1V→	121	113	57	-52	0	97	63	141	68	87
0V→	240	206	107	5	0	173	141	293	140	165

5.4.2 Number of winners

The number of winners, k , can be determined before each selection phase by adjusting I_{TAIL} and I_{M3} externally. As function of their ratio, the number of winners identified by the circuit is shown in Fig. 5.9. A good agreement between the number of winners predicted by (5.2) and obtained from measurement, especially for small k , was observed.

5.4.3 Transient response

Typical transient response of the circuit programmed to select one winner is shown in Fig. 5.10(a). The impact of M9 on the winner selection is illustrated in Fig. 5.7(b) where M9 is turned on before a clear separation of the winner. In this case, M9 shorts M8 that prevents M13 to enter the nonsaturated region. The tail current fades out and the evaluation-node voltages are pulled further towards ground. Consequently, the selection phase ends without winner.

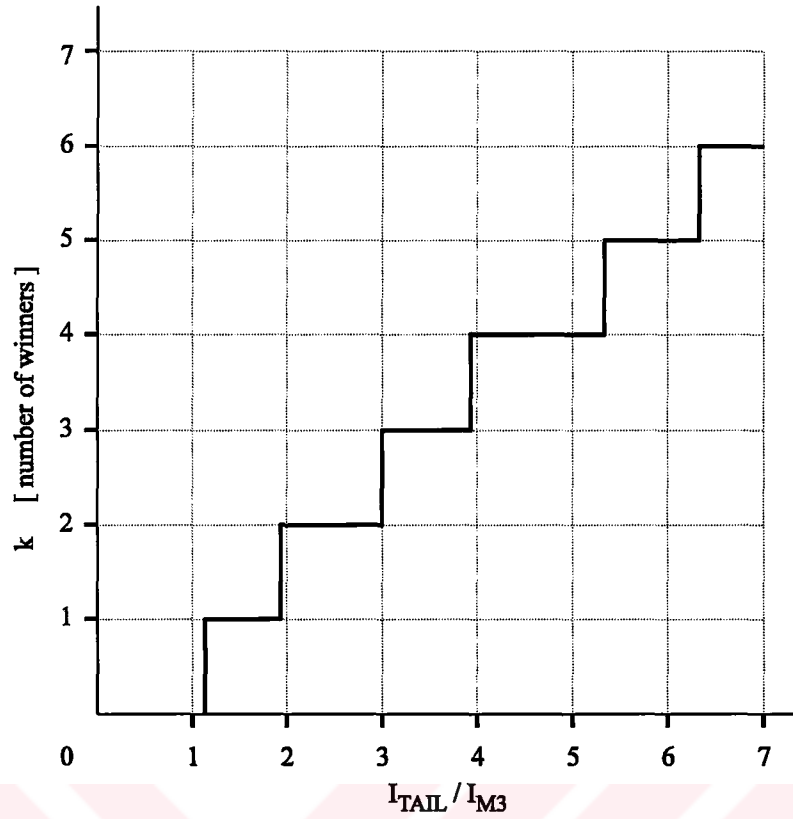
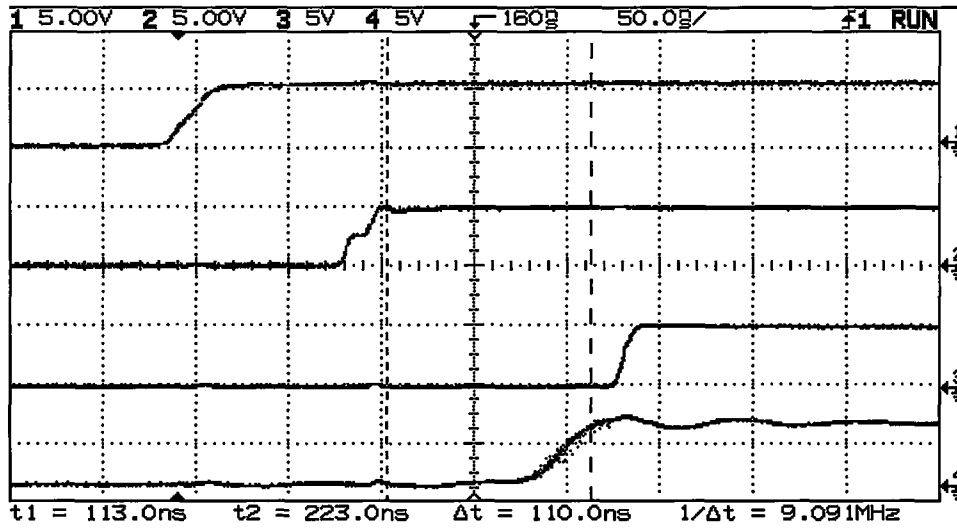
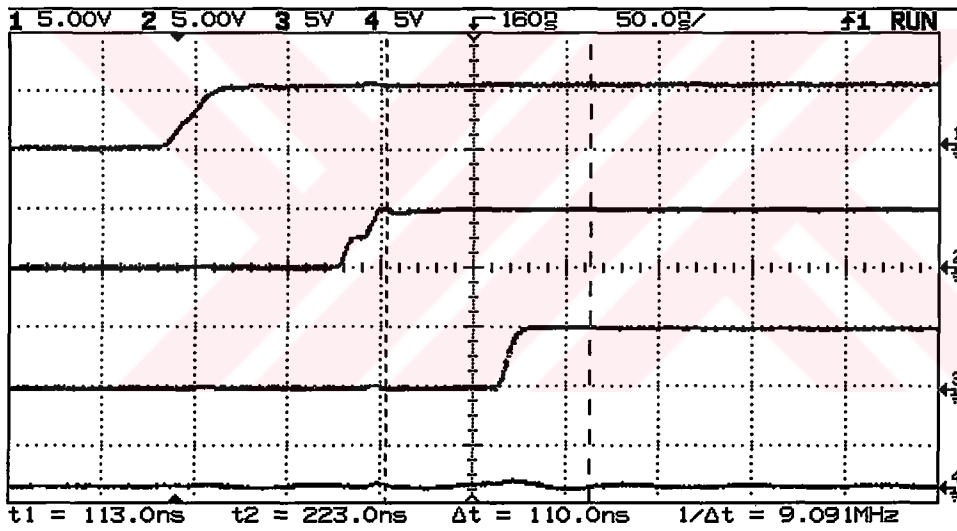


Figure 5.9. Measured dependence of number of winners ,k, on the ratio of the tail current to I_{M3} .

Dependence of the response time on input voltage difference is shown in Fig. 5.11. In order to measure the response of the circuit for various differential voltages, all input voltages were set to 2V and the circuit was programmed to work in the WTA mode. Then, the input voltage of Cell-4 was increased by 250 mV to make it a clear winner. Afterwards, the input voltage of Cell-6 was increased slowly until it becomes the new winner. It is assumed that the differential voltage between the inputs of Cell-4 and Cell-6 was zero where the transition (winner Cell-4 $\rightarrow$ winner Cell-6) occurred. Beyond this point, the input voltage of Cell-6 was increased by 40 mV steps and the delays associated with difference voltages (40mV, 80mV, . . .) were recorded. A similar procedure was carried out to depict the effect of the tail current on the response time. In this case, all of the input voltages were set to 2V, except for Cell-4 and Cell-6 whose input voltages were 2.250 V and 2.300 V, respectively. I_{M3} was set to 20 μ A while the tail current was increased by 2 μ A steps, from 20 μ A to 40 μ A. The higher the tail current, the higher the current charging the evaluation-node capacitance of the winning cell and the sooner the circuit responds. Consequently, for a given input difference, linear inverse dependence of response time on the tail current is obvious as shown in Fig. 5.12.



(a)



(b)

Figure 5.10. Experimental waveforms. a) The first, second and third traces belong to ϕ_1 , ϕ_2 and ϕ_3 , respectively; the fourth trace shows the buffered output voltage of cell-9. b) The configuration is same as (a) except for ϕ_3 coming before the end of the competition. $I_{TAIL}=30\mu A$, $I_{M3}=20\mu A$, $V_{IN(9)}=1.3V$, $V_{IN(5)}=1.250V$, all other inputs = 1.2V.

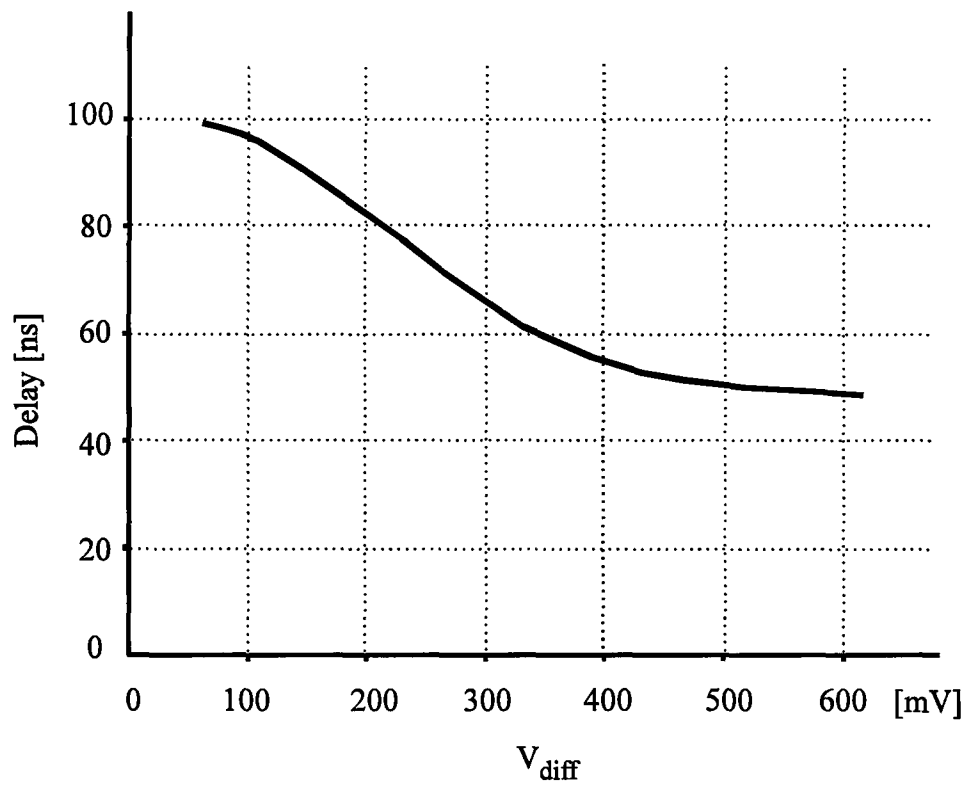


Figure 5.11. Measured effect of input difference voltage on response time for $k=1$.

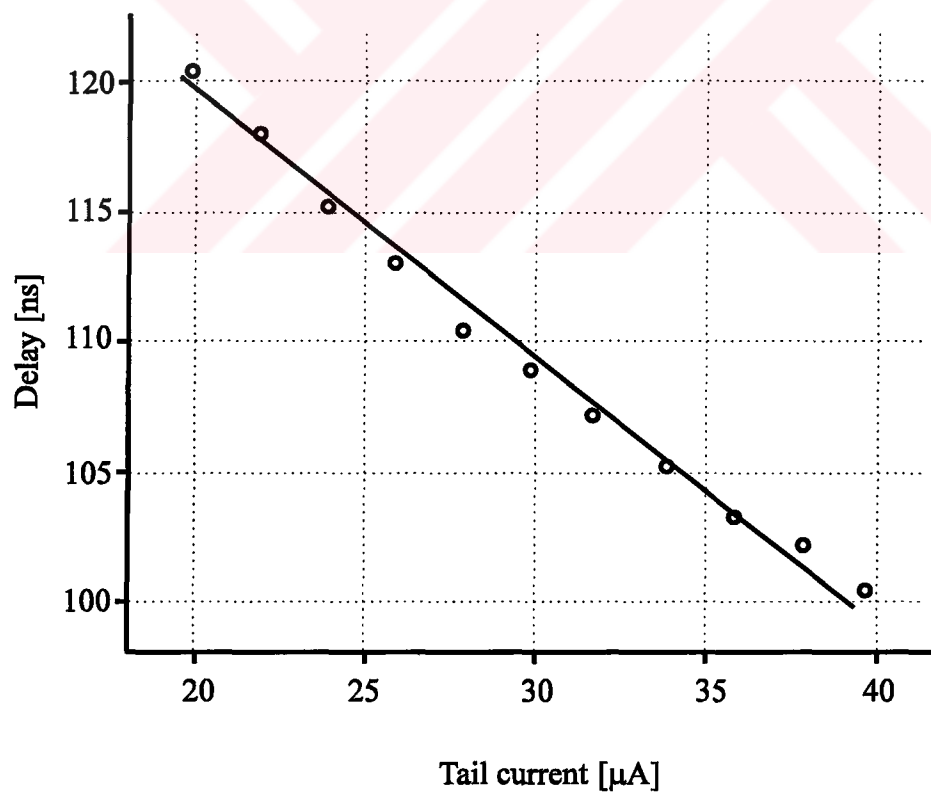


Figure 5.12 Measured effect of tail current on response time for $k=1$.

6. CONCLUSIONS

Two modifications on the Lazzaro WTA circuit and two novel circuit topologies constitute the backbone of this dissertation.

Modern MOS transistors having sub-micron channel length, exhibit pronounced channel-length modulation. This deficiency causes resolution degradation in WTA circuits. The above mentioned modifications are aimed to solve this problem for the Lazzaro WTA without impairing its advantages stemming from its simple architecture.

The modified circuits achieve a discrimination capability beyond that of all previously proposed architectures with still a simple architecture. Owing to this simplicity, their power and real estate consumption stays lower. They don't need large capacitors to stabilize the circuit as is the case in the Serrano-Barranco circuit [21] or positive feedback [16] in which an initialization phase becomes unavoidable. To the author's knowledge, the modified circuits are the most advantageous current-mode asynchronous circuits of the state-of-the art.

Beside these, two novel circuit topologies are proposed. Both are synchronous circuits of $O(N)$ complexity. The first one is a WTA circuit and is the first realized circuit of this kind. The second one is the world's first realized KWTA circuit. Please note that the information given above on the originality of the proposed circuits is limited to the best of author's knowledge. The measurement results not only proved the functionality of the circuits, but also showed good performance parameters attained with such simple architectures.

Most of previous work on (K)WTA circuits are focused on theoretical investigation of the WTA process. In order to analyze this process in details highly linearized models are adopted. Unfortunately, the extend to which these models present the circuits has not been questioned up to now. The circuits based on these linearized models can only be implemented by using large amount of local feedback. Yet, this would cause unacceptably slow response times and also high real estate consumption. As the number of realized circuits increases the above mentioned incoherence will be more and more pronounced. It is obvious that coping with

nonlinear behavior of active electronic elements, such as a MOS transistor, will lead us to less generalized solution mostly dependent on the circuit topology. Actually, transistor level simulation seems to be the only reliable method.

The comparison of the novel approach used to investigate the stability of the two proposed circuits (presented in Sections 4.3 and 5.3, respectively) with the method given in Section 2.3.1, sheds more light on this issue. The latter method is originally presented in [33]. It seems to be impossible to relate the restrictions given in this work with the parameters of the proposed WTA circuit although the method is aimed at analyzing the stability of the circuits of this kind. The method presented in this dissertation forms a seed for future work.

New topologies based on low power circuit techniques suitable for deep sub-micron technologies and high level models monitoring the trade-off between speed, resolution, precision, and power consumption, are the most important issues that need to be examined in the future.



REFERENCES

- [1] **Mahowald, M. A. and Delbruck, T.**, 1989, Cooperative stereo matching using static and dynamic image features, in *Analog VLSI implementation of neural systems*, 1, pp. 19-28, Ed. Mead, C and Ismail, M., Kluwer Academic
- [2] **Robinson, M. E., Yoneda, H. and Sanchez-Sinencio**, 1992, A modular CMOS design of a Hamming network, *IEEE Trans. Neural Networks.*, 3, 205-213.
- [3] **Mead, C. A., Arreguit, X. and Lazzaro, J.**, 1991, Analog VLSI model of binaural hearing, " *IEEE Trans on Neural Networks.*, 2 ,230-236.
- [4] **Lee, J. C. and Sheu, B. J.**, 1991, Analog VLSI neuroprocessors for early vision processing, in *VLSI signal processing*, 4, pp. 319-328, Ed. Moscovitz, H. S., Yao, K. and Jains, R., IEEE Press New York.
- [5] **Fang, W. C. and Sheu, B. J., Chen, O. T. C. and Choi, J.**, 1992, A VLSI neural processor for image data compression using self-organization neural network, *IEEE Trans. Neural Networks.*, 3, 506-518.
- [6] **Lippman, R. P.**, 1987, An introduction to computing with neural nets, *IEEE Acoustics, Speech , Signal Procesing*, 7, 4-22.
- [7] **Hech-Nielsen, R.**, 1990. *Neurocomputing*, Addison Wesley.
- [8] **Kosko, B.**, 1992. *Neural networks and fuzzy systems* , Prentice-Hall , New Jersey.
- [9] **Kohonen, T.**, 1990, The self-organizing map, *Proc. IEEE*, 7, 1464-1480.
- [10] **Kohonen, T.**, 1988. *Self-Organization and associative memory*, Springer-Verlag, New York.
- [11] **Grossberg, S.**, 1987, Competitive learning: From interactive activation to adaptive resonance, *Cognitive Sci.*, 11 , 23-63.

- [12] **Grossberg, S.**, 1976, Adaptive pattern classification and universal recording: I. Parallel development and coding of neural feature detectors, *Biolog. Cybern.*, **23** , 121-134.
- [13] **Grossberg, S.**, 1976, Adaptive pattern classification and universal recording: II. Feedback, expectation, illusions, *Biolog. Cybern.*, **23** , 187-202.
- [14] **Lazzaro, J., Reyckebusch, S., Mahowald M. A. and Mead, C.A.**, 1989, Winner-Take-All network of $O(N)$ complexity, in *Advances in neural information processing systems*, **1**, pp. 703-711, Ed. Touretzky, D. Morgan Kaufmann, San Mateo, CA.
- [15] **Choi, J. and Sheu, B. J.**, 1993, A high-precision VLSI winner-take-all circuit for self-organizing neural networks, *IEEE Journal of Solid-State Circuits*, **28**, 576-583.
- [16] **Starzyk, J. A. and Fang, X.**, 1993, CMOS current mode winner-take-all circuit with both excitatory and inhibitory feedback, *Electron. Lett.*, **29**, 908-910.
- [17] **Şekerkıran, B. and Çilingiroğlu, U.**, 1995, A high resolution CMOS winner-take-all circuit, *In Proceedings of the IEEE Int. Conf. Neural Networks*, Perth, Australia, December 1995, 2023-2026.
- [18] **Çilingiroğlu, U.**, 1993, A charge-based neural Hamming classifier, *IEEE Journal of. Solid-State Circuits.*, **28**, 59-67
- [19] **He, Y. and Sanchez-Sinencio, E.**, 1993, Min-net Winner-take-all CMOS implementation, *Electron. Let.*, **29**, 1237-1239.
- [20] **Erten, G. and Goodman, R. M.**, 1996, Analog VLSI implementation for stereo correspondence between 2-D images, *IEEE Trans. Neural Networks.*, **7**, 266-277.
- [21] **Serrano, T. and Linares-Barranco, B.**, 1996, A modular current-mode high-precision winner-take-all circuit, *IEEE Trans. Circuits Syst.*, **42**, 132-134.
- [22] **DeWeerth, S. P. and Morris, T. G.**, 1995, CMOS current mode winner-take-all circuit with distributed hysteresis, *Electron. Lett.*, **31**, 1051-1053.

- [23] **Pedroni, V. A.**, Neural n-port voltage comparator network, 1994, *Electron. Lett.*, **30**, 1774-1775.
- [24] **Johnson, L. G. and Jalaeddine S. M.**, 1991, MOS implementation of Winner-Take-All network with application to content-addressable memory, *Electron. Lett.*, **27**, 957-958.
- [25] **Oflazer, K.**, 1983, Design and implementation of a single chip 1-D Median filter, *IEEE Trans. Acousts., Speech, Signal Processing*, **31**, 1164-1168.
- [26] **Lin, H. M., and Wilson, A. N.**, 1988, Median filters with adaptive length, *IEEE Trans. Circuits Syst.*, **35**, 675-690.
- [27] **Fogel, D. B.**, 1994, An introduction to simulated evolutionary optimization, *IEEE Trans. Neural Networks.*, **5**, 3-14.
- [28] **Dierderich, J.**, 1991, Steps towards knowledge-intensive connectionist learning, in *Advances in Connectionist and Neural Computation Theory*, **1**, pp. 19-28, Ed. Barnden, J. A: and Pollock, J. B., Norwood, New Jersey
- [29] **Eliasbergh, V.**, 1988, Neuron layer with reciprocal inhibition as a mechanism for random choice, in *Proceedings of the IEEE Int. Conf. Neural Networks*, 17-25.
- [30] **Barnden, J. A. and Kankanahalli, S.**, 1993, Temporal winner-take-all networks: A time-based mechanism for fast selection in neural networks, *IEEE Trans. Neural Networks.*, **4**, 844-853.
- [31] **Nossek, J. A., Seiler, G., Roska, T. and Chua, L. O.**, 1992, Cellular neural networks: Theory and circuit design, *Int. J. Circuit Theory Appl*, **20**, 533-553.
- [32] **Seiler, G., Schuler A. J., and Nossek, J. A.**, 1993, Design of robust cellular neural networks, *IEEE Trans. Circ. Syst. I.*, **40**, 358-364.
- [33] **Seiler, G. and Nossek, J. A.**, 1993, Winner-Take-All Cellular Neural Networks., **4**, *IEEE Trans. Circ. Syst. II.*, **40**, 184-190.

- [34] **Majani, E., Erlanson, R. and Abu-Mostafa, Y.**, 1989, On the K-Winners-Take-All network, in *Advances in neural information processing systems*, 1, pp. 634-641, Ed. Touretzky, D. Morgan Kaufmann, San Mateo, CA.
- [35] **Şekerkıran, B. and Çilingiroğlu, U.**, 1999, A CMOS k-winner-take-all circuit with $O(N)$ complexity, *IEEE Trans. Circ. Syst. II.*, **46**, 1-5
- [36] **Yang, J. F., Chen, C. M., Wang, W. C. and Lee, J. Y.**, 1995, A general mean-based iterative winner-take-all neural, *IEEE Trans. Neural Networks.*, **6**, 14-24.
- [37] **Şekerkıran, B. and Çilingiroğlu, U.**, 1997, Improving the resolution of Lazzaro winner-take-all circuit in *Proceedings of the IEEE Int. Conf. Neural Networks*, Texas, USA, July 1997, 1005-1008.
- [38] **Şekerkıran, B. and Çilingiroğlu, U.**, 1998, Precision improvement in current-mode winner-take-all circuits using gain-boosted regulated-cascode CMOS stages in *Proceedings of the IEEE Int. Conf. Neural Networks*, Anchorage, USA, May 1998, 553-556.
- [39] **Urahama, K. and Nagao, T.**, 1995, K-Winners-Take-All Circuit with $O(N)$ complexity, *IEEE Trans. Neural Networks.*, **6**, 776-778.
- [40] **Lee, T. H.**, 1998. Formulas for second-order low-pass systems, in *The Design of CMOS Radio-Frequency Integrated Circuits*, pp 413-415, Cambridge University, Cambridge.
- [41] **Pelgrom, M. J. M., Duinmaijer, A. C. J. and Welbers, A. P. G.**, 1989, Matching properties of MOS transistors, *IEEE Journal of Solid-State Circuits*, **24**, 1433-1440.
- [42] **Hostica, B. J.**, 1979, Improvement of the gain of MOS amplifiers, *IEEE Journal of Solid-State Circuits*, **14**, 111-1114.
- [43] **Sackinger, and Gugenbühl, W.**, 1990, An active-feedback cascode current source, *IEEE Trans Circuits Syst. II*, **37**, 644-646.

APPENDIX A

Layout Considerations

It is important to remember that a simulation is only as good as its models. When the physical device being modeled will be manufactured, there will be manufacturing variations across the multiple copies which a single model will not reflect. Some of these discrepancies come from mask and patterning processes, while others stem from physical and electrical mismatches such as layer thickness and doping variations. These processes are more or less anisotropic. This means that the compatibility between devices oriented in the same direction is higher compared to the compatibility between rotated devices. So the critical devices should be laid on in the same direction.

Electromagnetic effects generated by the switching digital circuits can affect sensitive analog circuits via supply rail or substrate induced noise. The distance between these devices continues to reduce dramatically with tighter integration. It is possible to get rid of the supply rail coupling by using two separate supply rails and minimizing the capacitive and inductive coupling between them. This can be possible by keeping them away from each other and using a totally independent supplying scheme. Since p-substrate is common for the whole chip, it is impossible to suppress substrate coupling effects. A brief guide to be taken into account while drawing the layout is shown below.

- ☐ Use the same layout for all cells and place them in the same orientation.
- ☐ Form the common control line with a narrow second metal layer.
- ☐ Keep the stray capacitance of the common control line as low as possible.
- ☐ Keep the maximum distance between any cell pair short.
- ☐ Encircle the cells with guard rings.

APPENDIX B

Distribution of Initial Activations in WTA circuits.

Three frequently-occurring distributions, will be examined.

1. Uniform Distribution $[0,1]$: If the inputs, which are equally likely to take on any value within the given range, are uniformly distributed, the probability density function of the uniform distribution in $[0,1]$, for example, is given by,

$$p(x) = 1, \quad 0 < x < 1 \quad (\text{B.1})$$

To the WTA process, uniformly distributed inputs are the most difficult case among these three distributions in finding the maximum activated cell.

2. Normal distribution $(0,1)$: In the normal distribution of input exemplars, we assume a variance of one and a mean of zero. Thus, the probability density function is

$$p(x) = \sqrt{\frac{1}{2\pi}} e^{\left(\frac{-x^2}{2}\right)} \quad -\infty < x < \infty \quad (\text{B.2})$$

The normally distributed inputs are the most popular case for describing the activations, especially during random initializations.

3. Peak Uniform Distribution (α, β) : We define the peak uniform distribution as the inputs containing only one peak which is considerably larger than the rest. The maximum value is α , and the remaining $(n-1)$ inputs are uniformly distributed in the range $[0, \beta]$. The peak uniformly distributed inputs can be found in many classification problems in which the input exemplar well matches with the one and only one correspondence.

APPENDIX C

MOS transistor model parameters used in simulations.

C.1 0.6 μ m nwell process.

PMOS transistor model

```
.model modp bsim3 type=p capmod=bsim noisemod=2 subthmod=3.000e+00
satmod=2.000e+00 bulkmod=1.000e+00 k1=4.040e-01 k2=2.398e-02
k3=4.760e+00 k3b=-1.93000000E-01 npeak=3.681e+16 vth0=-9.49000000E-01
voff=-7.53000000E-02 dvt0=1.929e+00 dvt1=1.087e+00 dvt2=-4.51000000E-02
keta=1.527e-02 pscbe1=4.700e+08 pscbe2=1.000e-05 ua=0.000e+00
ub=8.473e-19 uc=-3.50000000E-02 uo=1.132e+02 dsub=5.000e-01
eta0=6.314e-02 etab=0.000e+00 nfactor=8.483e-01 vghigh=1.500e-01
vglow=-1.10000000E-01 eta=3.000e-01 pclm=3.019e+00 pdibl1=3.885e-01
pdibl2=0.000e+00 dROUT=5.000e-01 a0=5.953e-01 a1=0.000e+00 a2=1.000e+00
pvag=1.537e+00 vsat=7.490e+04 alpha=2.000e+00 ldd=2.000e-07 w0=0.000e+00
at=3.300e+04 ute=-1.42000000E+00 kt1=-4.89000000E-01 kt2=2.200e-02
kt1l=0.000e+00 ual=0.000e+00 ub1=0.000e+00 uc1=0.000e+00 cgdo=3.400e-10
cgso=3.400e-10 cgbo=1.300e-10 xpart=1.000e+00 rds0=1.000e+00
rdsw=4.516e+03 cdsc=8.221e-04 cdscb=0.000e+00 cit=9.958e-05
tox=1.264e-08 litl=1.067e-07 nlx=2.230e-07 xl=-5.00000000E-08 xw=0.000e+00
af=1.279e+00 kf=6.314e-29 ef=1.000e+00 rd=0.000e+00 rs=0.000e+00
rsh=5.700e+01 minr=1.000e-03 rdc=0.000e+00 rsc=0.000e+00 dl=7.134e-08
dw=3.122e-08 ldif=0.000e+00 hdif=8.000e-07 xj=3.000e-07 js=4.000e-05
n=1.000e+00 dskip= no tlev=0 tlevc=0 cj=4.400e-04 cjsw=3.100e-10 fc=0.000e+00
fcsw=0.000e+00 mj=5.600e-01 mjsw=3.300e-01 pb=9.100e-01 pbsw=9.100e-01
```

NMOS transistor model

```
.model modn bsim3 type=n capmod=bsim noisemod=2 subthmod=3.000e+00
satmod=2.000e+00 bulkmod=1.000e+00 k1=9.270e-01 k2=-8.02000000E-02
k3=4.864e+01 k3b=-4.04000000E-01 npeak=9.515e+16 vth0=7.968e-01
```

voff=-7.16000000E-02 dvt0=1.090e+00 dvt1=7.910e-01 dvt2=-3.93000000E-01
 keta=-2.79000000E-02 pscbe1=0.000e+00 pscbe2=1.000e-10 ua=4.400e-10
 ub=7.405e-19 uc=-1.27000000E-02 uo=4.426e+02 dsub=5.000e-01 eta0=3.664e-02
 etab=-2.44000000E-03 nfactor=7.659e-01 vghigh=1.500e-01
 vglow=-1.10000000E-01 eta=3.000e-01 pclm=1.433e+00 pdibl1=4.655e-02
 pdibl2=7.138e-03 drout=5.000e-01 a0=5.813e-01 a1=0.000e+00 a2=1.000e+00
 pvag=0.000e+00 vsat=8.979e+04 alpha=2.000e+00 ldd=2.000e-07 w0=2.611e-06
 at=3.300e+04 ute=-1.78000000E+00 kt1=-3.72000000E-01 kt2=2.200e-02
 kt1l=0.000e+00 ua1=0.000e+00 ub1=0.000e+00 uc1=0.000e+00 cgdo=3.400e-10
 cgso=3.400e-10 cgbo=1.300e-10 xpart=1.000e+00 rds0=1.000e+00 rdsw=1.703e+03
 cdsc=1.058e-03 cdscb=0.000e+00 cit=0.000e+00 tox=1.264e-08 litl=1.067e-07
 nlx=1.000e-10 xl=-5.00000000E-08 xw=0.000e+00 af=1.451e+00 kf=2.340e-26
 ef=1.000e+00 rd=0.000e+00 rs=0.000e+00 rsh=3.000e+01 minr=1.000e-03
 rdc=0.000e+00 rsc=0.000e+00 dl=1.545e-07 dw=-3.20000000E-09 ldif=0.000e+00
 hdif=8.000e-07 xj=3.000e-07 js=1.000e-05 n=1.000e+00 dskip= no tlev=0 tlevc=0
 cj=3.500e-04 cjsw=3.100e-10 fc=0.000e+00 fcsw=0.000e+00 mj=4.500e-01
 mjsw=4.900e-01 pb=9.400e-01 pbsw=9.400e-01

C.2 YITAL 3 μ m technology.

PMOS transistor model

.MODEL MP PMOS LEVEL=2 VT=-0.7 KP=15U GAMMA=0.8 PHI=0.7
 LAMBDA=0.03 UCRIT=6.15E4 UEXP=0.15 CJ=6E-4 CJSW=3.4E-10 MJ=0.5
 MJSW=0.33 CGDO=500E-12 CGSO=500E-12

NMOS transistor model

.MODEL MN NMOS LEVEL=2 VT=0.7 KP=45U GAMMA=0.11 PHI=0.5
 LAMBDA=0.03 UCRIT=8.23E4 UEXP=0.08 CJ=1E-4 CCJSW=3.9E-10 MJ=0.5
 MJSW=0.33 CGDO=500E-12 CGSO=500E-12

APPENDIX D

Algorithm modeling transient behavior of the proposed WTA circuit.

```
n = 10 total cell number

dt = 100e-12, simul_time = 200e-9 timestep and total simulation time

beta1 = 100e-6, beta5 = 60e-6 transconductance parameter of M1 and M5

ceval = 500e-15, ccsn = 800e-15 node capacitances

vt = 0.7, w = 1, r = 1.5 threshold voltage and circuit parameters

vdd = 5 supply voltage

for j = 1 to n input u(2,j) endfor initialize input initial evaluation node voltages

z(1)=0, z(2)=0 initialize common source node voltage

for j = 1 to dqeval(j) = 0 endfor initialize dqeval's

for i = 3 to simul_time/dt for each time point

    s1=z(i-1)-vt

    if s1 > 0 then itail(i)=beta5*s1*s1 compute tail current

    else itail(i)=0 endif

for j = 1 to n for each cell
```

```

s2=u(i-1,j)-z(i-1)-vt
if s2 > 0 then im1(j)=beta1*s2*s2           compute current through M1
else im1(j)=0 endif

f1 = 1, f2 = 1;

if z(i-1) > vdd-vt then f1=0 endif

if u(i-1,j) > vdd then f1=0 endif

if u(i-1,j) < 0 then f2=0 endif

u(i,j)=u(i-1,j)+(dt/ceval)*(f1*w*im1(i,j)-f2*itail(i)/r);

endfor

top = 0

for j=1 to n top = top + im1(j) endfor           add drain current of M1's

z(i) = z(i-1)+(dt/ccsn)*(top-itail(i-1)-10e-6);

end

```

APPENDIX E

Algorithm modeling transient behavior of the proposed KWTa circuit.

```
n = 10, k = 3.5                                total cell and winner number

dt = 1e-9, simul_time = 200e-9                timestep and total simulation time

beta = 50e-6                                   transconductance parameter of M3

im3 = 20e-6, im5 = 7e-6                       current through M3 and M5

ceval = 400e-15, ccsn = 2e-12                 internal node capacitances

for j = 1 to n  input u(1,j) endfor             initialize input evaluation node voltages

z(1)=0                                           initialize common source node voltage

s1=sqrt(im3/beta)

for i = 2 to simul_time/dt                     for each time point

  for j = 1 to n                               for each cell

    s2=u(i-1,j)-z(i-1);

    if s2 > s1 then g(i,j)=im3;

    elseif s2 > 0 then g(i,j)=beta*((u(i-1,j)-z(i-1))^2)

    else g(i,j)=0;

    f1 = 1, f2 = 1

    if z(i-1) > 4 then f1=0 endif
```

```

        if z(i-1) < 2 then f2=0; endif

        u(i,j)=u(i-1,j)+(dt/ ceval)*(f1*im5-f2*(im3-g(i,j)))

        if u(i,j) > 5 then u(i,j)=5 endif

        if u(i,j) < 0 then u(i,j)=0 endif

    endfor

    top = 0;

    for j=1 to n    top = top + g(i,j) endfor

    z(i) = z(i-1)+(dt/ccsn)*(top-(k*im3))    compute common source node voltage

endfor

```

BIOGRAPHY

Barbaros Şekerkıran was born in Istanbul in 1965. After completing his studies at Lycée Saint-Benoit in 1984, he entered Istanbul Technical University. He received the B.S. and M.S degrees in Electrical Engineering in 1989 and 1992, respectively. Since 1992 he has been pursuing his Ph.D. degree at the same university. From 1991 to 1997 he worked at the Scientific and Technical Research Council of Turkey (TÜBİTAK) in Kocaeli as CMOS process engineer. His research interests were ion implantation, CVD processes, transistor optimization and MOS transistor parameter extraction. From 1997 to 1999 he was employed as mixed-mode circuit designer at ETA Foundation in Istanbul. Since January 1999, he has been with Alcatel Microelectronics in Belgium, where he is mainly involved in high speed chip-to-chip communication projects, phase-aligners and clock recovery systems. His personal interests include prehistory of Asia Minor and the Balkans, anthropology, body-building, and gourmet cuisine.