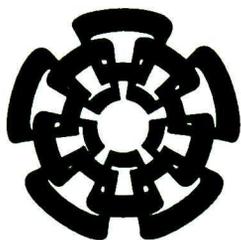


CT-854-SS1
POW. 7015



Centro de Investigación y de Estudios Avanzados
del Instituto Politécnico Nacional
Unidad Guadalajara

Análisis y simulación de un chip de visión neuromórfico para la detección de movimiento

Tesis que presenta:

César Rodolfo Acosta Méndez

para obtener el grado de:

Maestro en Ciencias

en la especialidad de:

Ingeniería Eléctrica

Director de Tesis

Dr. Juan Luis Del Valle Padilla

**CINVESTAV
IPN
ADQUISICION
LIBROS**

CLASIF.. CT00755
ADQUIS.. CT-854-581
FECHA: 01-07-2015
PROCED.. DDV17015
\$ _____

Análisis y simulación de un chip de visión neuromórfico para la detección de movimiento

**Tesis de Maestría en Ciencias
Ingeniería Eléctrica**

Por:

César Rodolfo Acosta Méndez
Ingeniero en Mecatrónica

Universidad Autónoma de Yucatán 2004-2009

Becario de CONACYT, expediente no. 263550

Director de Tesis
Dr. Juan Luis Del Valle Padilla

CINVESTAV del IPN Unidad Guadalajara, Octubre de 2014.

Resumen

Implementar aplicaciones de ingeniería neuromórfica es siempre un reto, hay que entender en primera instancia cómo la naturaleza ha resuelto un problema, y aplicar lo que podemos observar de sus soluciones en dispositivos electrónicos. Un proceso que se atiende en esta tesis es el de la visión y en especial la detección y atención al movimiento. En este trabajo de tesis se analiza y simula un sistema neuromórfico que emula las propiedades elementales de los sistemas de visión biológicos, como son las propiedades cognitivas de atención y decisión de un sistema elemental de visión, con la característica de integrar en un solo chip las funcionalidades del hardware y software, y el empleo de una energía mínima para estos propósitos.

El objetivo del sistema de visión es mantenerse enfocado en el seguimiento de una línea por un vehículo. El sistema está compuesto por cinco bloques en los cuales la información fluye en forma jerárquica, desde un sistema de fotorrecepción hasta un bloque que determina las características más relevantes de la escena, el seguimiento y decisión.

Como una etapa previa a un proceso de fabricación de un sistema similar, hemos creído conveniente simularlo en Spice utilizando las herramientas de Tanner EDA, en una tecnología $0.5 \mu\text{m}$ de On Semiconductor, empleando los parámetros actualizados de los modelos de los transistores Mosfets Bsim3.v3, disponibles en MOSIS. Esta versión Bsim3.v3, es conveniente para la simulación en el régimen de débil inversión de los transistores, régimen que es la opción más favorable para la emulación de procesos biológicos por los transistores.

Aunque la información fluye en forma jerárquica, de manera continúa, la información es generada y procesada en paralelo por el intermedio de un arreglo unidimensional de fotorreceptores y de los bloques de procesamiento subsecuentes.

En este trabajo se presenta un análisis detallado y una simulación en Spice de los diferentes bloques del sistema, validados con las respuestas

experimentales de éstos, publicadas en la literatura, haciendo énfasis entre sus variables de entrada y de salida.

ABSTRACT

Build applications for neuromorphic engineering is always a challenge, you have to understand at first, how nature has solved a specific problem, and apply the solutions that you observe to electronics devices, in this thesis our efforts are centered to the vision and in special way to the detection and focussing of movement. In this thesis work, it is analyzed and simulated a neuromorphic system, that emulate the cognitive properties of attention and detection of movement of biological vision systems, as are the cognitive properties of attention and decision of an elemental vision system, with the characteristics of the integration on a chip all the functionalities of the hardware and software and the employment of a minimum energy for this purpose.

The objective of the vision system is to stay focused on tracking a line by a vehicle. The system is composed of five blocks in which information flows in a hierarchical manner, from a photo reception system to a block that determines the most relevant characteristics of the scene, monitoring and decision.

As a previous step of the manufacturing process, we thought it would be convenient to simulate the system in SPICE, using Tanner EDA tools, on a 0.5um technology of On Semiconductor, using the updated parameters of the MOS transistors Bsim3.v3 available in MOSIS. The Bsim3.v3 version is suitable for simulation in the regime of weak inversion of transistors, being this regimen is the most favorable one, for the emulation of biological processes by transistors.

Although the information flows in a hierarchical fashion, continuously, information is generated and processed in parallel by means of a one-dimensional array of photoreceptors and subsequent processing blocks.

In this work is presented a detailed analysis and simulation using the Spice simulation of the different blocks of the system, validated with experimental responses of these blocks, found in the literature on the subject, emphasizing the relationship among input and output variables.

Agradecimientos

Le agradezco a mi padre, Cesar Renan, a mi madre, Francisca Esther, por darme todo el soporte y los recursos para lograr mis objetivos, a mi hermana por su apoyo.

A mi asesor Dr. Juan Luis del Valle por su tiempo y apoyo para lograr terminar el programa de maestría, a mis profesores dentro del programa de maestría por haberme enseñado tanto.

A mis amigos y compañeros por su ayuda en los momentos difíciles, al hacer de este tiempo más llevadero.

Y finalmente a CONACYT por el apoyo económico que me otorgó con la beca de manutención, la cual me sirvió para sobrevivir lejos de casa.

Contenido

Agradecimientos.....	IX
Resumen.....	III
ABSTRACT	V
Índice de Figuras.....	XV
Índice de Tablas	XIX
Capítulo 1 Introducción.....	1
1.1 Antecedentes en ingeniería neuromórfica.....	1
1.2 Los chips neuromórficos de visión.....	4
1.3 Propuesta de tesis.	5
1.4 Objetivos de la tesis.	7
1.5 Organización de la tesis.....	8
Capítulo 2 Descripción del chip de visión neuromórfico.	9
2.1 El chip de visión.	9
2.2 Algoritmo de percepción del movimiento.	12
2.2.1 Un modelo basado en mapas de relevancia (<i>Saliency map</i>).....	13
2.2.2 El movimiento.	15
2.2.3 Esquemas basados en intensidad.	16
2.2.4 Detección de bordes en una imagen estática.	18
2.3 Propiedades en subumbral de un transistor.	21
2.4 Herramientas e infraestructura de simulación.	24
2.5 Cálculo de la irradiancia en el plano focal del sistema óptico.	24
2.6 Algoritmo para la emulación de la excitación luminosa a la entrada del fotorreceptor.....	28
2.7 Comparación del chip neuromórfico versus las funcionalidades de la visión humana.	31
2.8 Conclusiones.	34
Capítulo 3 Fotorreceptor.....	37
3.1 Principios y análisis de fotorreceptores neuromórficos.	37
3.1.1 Receptores logarítmicos simples.	39

3.1.2	Fotorreceptor logarítmico adaptativo.....	41
3.2	Propiedades del circuito.....	45
3.3	Simulaciones en SPICE.....	53
3.4	Conclusiones.....	58
Capítulo 4 Derivada espacial y selector.....		59
4.1	Introducción.....	59
4.2	Amplificador de transconductancia.....	60
4.2.2	Simulación de la respuesta del amplificador de transconductancia en respuesta a señales del fotorreceptor.....	65
4.3	Fundamentos de los circuitos en modo corriente en débil inversión.....	67
4.4	Descripción del circuito de detección de polaridad.....	71
4.4.1	Simulación de la respuesta del circuito de detección de polaridad en función de la respuesta del amplificador de transconductancia.....	72
4.5	Conclusiones.....	73
Capítulo 5 WTA.....		75
5.1	Principios del WTA.....	75
5.2	Aplicación en un circuito analógico.....	78
5.2.1	Comportamiento en estado estable de la red WTA básica.....	79
5.3	Simulaciones de un WTA.....	84
5.4	WTA histerético.....	88
5.5	Simulaciones de un WTA histerético.....	90
5.6	Redes resistivas.....	92
5.6.1	Redes excitadoras o de excitación lateral.....	93
5.6.2	Redes inhibitoras.....	99
5.7	WTA histerético completo.....	103
5.8	Simulaciones de un Chip de visión con atención.....	106
5.9	Descripción sobre el circuito capaz de determinar la posición de manera analógica con los datos del WTA.....	111
5.10	Conclusiones.....	115
Capítulo 6 Conclusiones.....		117
6.1	Introducción.....	117
6.2	Descubrimientos empíricos.....	118
6.3	Trabajo futuro.....	121
6.4	Conclusión.....	121
Apéndice A. Publicaciones.....		123

Bibliografia	125
---------------------------	------------

Índice de Figuras

Fig. 1.1: Computadora BlueGene (izquierda) y una abeja (derecha), diferentes tamaños similares capacidades.....	2
Fig. 1.2: Comparación entre propiedades de un transistor y una neurona.....	3
Fig. 2.1: Diagrama que ejemplifica el funcionamiento del vehículo acoplado con un sensor neuromórfico.....	10
Fig. 2.2: Diagrama de bloques de un sistema de seguimiento en un solo chip.....	11
Fig. 2.3: <i>Layout</i> del circuito fotorreceptor diseñado por Indiveri.....	12
Fig. 2.4: Diagrama esquemático de un modelo de atención selectiva basado en Mapas de relevancia.....	14
Fig. 2.5: Una imagen para la detección de movimiento en una dimensión.....	17
Fig. 2.6: El operador G'' , la segunda derivada de la distribución gaussiana.....	19
Fig. 2.7: Superficie que muestra el cambio de intensidades en un área.....	20
Fig. 2.8: Dobles derivadas de la superficie de intensidades en tres diferentes orientaciones.....	21
Fig. 2.9: Fotorreceptor seguidor de fuente.....	25
Fig. 2.10: Características I-V de una unión p-n bajo iluminación.....	26
Fig. 2.11: Descripción del plano focal.....	27
Fig. 2.12: Diagrama de los fotorreceptores y una barra negra.....	29
Fig. 2.13: Perfil de intensidad de una barra oscura pasando sobre un fondo blanco.....	30
Fig. 2.14: Procesamiento en el sistema visual.....	31
Fig. 2.15: Sistema de visión simplificado.....	32
Fig. 2.16: Emulación de las funciones de la retina y el núcleo lateral Geniculado por los bloques de circuitos neuromórficos equivalentes.....	33
Fig. 2.17: Emulación del proceso de detección de bordes, seguimiento, decisión y acción por la red WTA y el circuito de posición.....	34
Fig. 3.1: Una unión simple, el fotorreceptor logarítmico más simple. V_o se mantiene debajo el voltaje de sustrato, decreciendo logarítmicamente con la intensidad.....	40
Fig. 3.2: Un fotorreceptor logarítmico en seguidor de fuente. V_o decrece logarítmicamente con la intensidad de la luz empezando con una intensidad de cero en aproximadamente V_b	41
Fig. 3.3: El circuito propuesto por Shih-Chii Liu.....	42
Fig. 3.4: Circuito propuesto por Delbrück.....	43
Fig. 3.5: Características en DC.....	44
Fig. 3.6: Modelo de señal pequeña para el circuito de la Fig. 3.3.....	46

Fig. 3.7: Comparación entre los resultados del modelo usando parámetros constantes, dependientes de la frecuencia y su simulación en SPICE.	49
Fig. 3.8: Respuesta en frecuencia del fotorreceptor con bajas condiciones de iluminación, en función de las dimensiones del transistor M5.	50
Fig. 3.9: Dependencia del voltaje de polarización de compuerta V_{sg} , del transistor M5 en función de la variación de cinco décadas en el nivel de iluminación.....	50
Fig. 3.10: Adaptación de las características de filtrado del fotorreceptor en respuesta a la variación de la intensidad de la iluminación.....	51
Fig. 3.11: Efectos de la polarización V_m en el decaimiento del pulso de salida.	51
Fig. 3.12: Variación del voltaje de la fuente a la compuerta del transistor de adaptación como función del voltaje de polarización V_m . El voltaje de salida, V_{out} , depende únicamente de la intensidad de iluminación.	52
Fig. 3.13: Efecto de la relación de capacitancias C_d y C_i . Los valores de C_d fueron 4 pF y 10 pF respectivamente, C_i fue mantenido a un valor de 1 pF.	53
Fig. 3.14: Efecto de la relación de capacitancias C_d y C_i a bajas intensidades de iluminación.	53
Fig. 3.15: Ganancia transitoria Vs velocidad como función de la iluminación de la escena.	54
Fig. 3.16: Diagrama del fotorreceptor.	55
Fig. 3.17: Resultados a comparar.	56
Fig. 3.18: Cinco bloques fotorreceptores.	56
Fig. 3.19: Entrada de los 5 pixeles ante un borde de entrada de una barra oscura.....	57
Fig. 3.20: Entrada de 5 pixeles ante el borde de salida de una barra.	57
Fig. 3.21: Respuesta del fotorreceptor 3 ante ambos bordes.	58
Fig. 4.1: Par diferencial básico.	61
Fig. 4.2: Respuesta en corriente del par diferencial.	62
Fig. 4.3: Par diferencial con espejo de corriente también conocido como amplificador de transconductancia.	63
Fig. 4.4: Salida de un amplificador de transconductancia.	64
Fig. 4.5: Amplificador de transconductancia con salida desacoplada-.....	64
Fig. 4.6: Esquema de simulación del OTA con dos unidades de fotorecepción.....	65
Fig. 4.7: Perfil de voltajes de entrada para los OTA.....	66
Fig. 4.8: Respuesta de los OTA ante el perfil de entrada de la Fig 4.7.....	66
Fig. 4.9: <i>Current conveyer</i> simple usando solo dos transistores.	67
Fig. 4.10: Respuesta del <i>current conveyer</i>	68
Fig. 4.11: <i>Current Conveyer</i> clase II.	69
Fig. 4.12: Simulación del circuito <i>current conveyer</i> corrientes características.	70
Fig. 4.13: La diferencia entre las corrientes del <i>current conveyer</i> , son mínimas.	70
Fig. 4.14: Amplificador operacional suministrado por el <i>software Tanner</i>	71

Fig. 4.15: Circuito para la detección de bordes.	72
Fig. 4.16: Salida seleccionando la parte negativa de una señal.....	72
Fig. 4.17: Salida obviando la parte positiva de la señal.	73
Fig. 5.1: Red de N neuronas excitadoras (blancas) proyectándose a la neurona común inhibidora (Negra).....	76
Fig. 5.2: Circuito WTA clásico.....	79
Fig. 5.3: Circuito WTA con degeneración en fuente.	84
Fig. 5.4: Diagrama de simulación WTA clásico.	85
Fig. 5.5: Cambio de voltajes ante una diferencia de un 10% en las dos entradas.	85
Fig. 5.6: Respuesta en corriente de una red WTA.	86
Fig. 5.7: Comportamiento del circuito WTA ante diferentes órdenes de magnitud en la entrada.	86
Fig. 5.8: Respuesta con el largo del Transistor $M_1 L = 13.5\mu\text{m}$	87
Fig. 5.9: Respuesta WTA con transistor de degeneración en fuente	87
Fig. 5.10: Diagrama de simulación, WTA con degeneración en fuente.	88
Fig. 5.11: Circuito WTA con retroalimentación histerética.	89
Fig. 5.12: Diagrama de simulación WTA con retroalimentación histerética.....	90
Fig. 5.13: Salida de la celda WTA histerética ante cambios en la entrada de la celda 2.	91
Fig. 5.14: Respuesta de un WTA histerético ante una entrada ganadora en la celda 5 y una de menor cantidad de corriente en la celda 1.....	92
Fig. 5.15: Respuesta de un WTA clásico ante una entrada ganadora en la celda 5 y una de menor cantidad de corriente en la celda 1.....	92
Fig. 5.16: WTA con histéresis.....	94
Fig. 5.17: Respuesta de una red WTA con red excitadora.	97
Fig. 5.18: WTA sin red de excitación lateral.....	98
Fig. 5.19: Respuesta ante diferentes voltajes de excitación.....	98
Fig. 5.20: WTA con red inhibidora.	99
Fig. 5.21: WTA con red de inhibición.	100
Fig. 5.22: Diagrama de simulación de una red de WTA con inhibición.	102
Fig. 5.23: Resultado ganador de una red WTA con inhibición.	102
Fig. 5.24: Resultado ganador de una red WTA inhibidor en diferentes voltajes en los transistores de inhibición.....	103
Fig. 5.25: Respuesta ganadora de un WTA histerético con red de excitación e inhibición.	104
Fig. 5.26: Comparación entre WTA clásico y WTAh, características inhibitoras y excitadoras desactivadas.	104
Fig. 5.27: Entradas para ambos tipos de WTA.....	105

Fig. 5.28: Comparación entre WTA clásico y WTAh, características inhibitoras y excitadoras activadas.....	105
Fig. 5.29: Comparación entre WTA clásico y WTAh con entradas espaciadas.	106
Fig. 5.30: Entradas espaciadas para la comparación entre WTA clásico y WTAh.....	106
Fig. 5.31: Diagrama de simulación para un chip completo de visión.	107
Fig. 5.32: Entradas de la etapa WTAh del circuito de visión.	108
Fig. 5.33: Salidas de cada elemento WTAh.	108
Fig. 5.34: Representación de la ubicación de la celda ganadora.....	109
Fig. 5.35: Entrada de corriente que representa una escena visual inusual.	109
Fig. 5.36: Entradas de corriente para la etapa WTAh, con entrada mostrada en la Fig. 5.32.	110
Fig. 5.37: Salida de la etapa WTA.	110
Fig. 5.38: Representación de la ubicación de la celda ganadora, con entradas mostradas en la Fig. 5.32.	111
Fig. 5.39: Red de agregación.	112
Fig. 5.40: Red de resistores como divisor de voltaje.	113
Fig. 5.41: Circuito neurona, generador de pulsos.	114
Fig. 5.42: Circuito generador de pulsos de doble riel con el circuito de agregación.....	115

Índice de Tablas

Tabla 2.1: Valores de iluminación e irradiancia bajo condiciones típicas de iluminación. .25	25
Tabla 2.2: Características fotovoltaicas de módulos solares comerciales típicos.26	26
Tabla 2.3: Numero f contra el diámetro de apertura del diafragma.....27	27
Tabla 2.4: Estimación de la fotocorriente en el plano focal de la imagen.28	28
Tabla 2.5: Estimación de la fotocorriente en diferentes condiciones de iluminación.28	28
Tabla 2.6: Definición de longitudes importantes.29	29
Tabla 3.1: Dimensiones de transistores.....48	48

Capítulo 1

Introducción.

En este capítulo se presentan antecedentes de la ingeniería neuromórfica, un aspecto que motiva esta tesis, trabajos que se han hecho sobre el mismo tema y el propósito del tema de tesis, así como su organización.

1.1 Antecedentes en ingeniería neuromórfica.

Cualquier cerebro, desde el humano hasta el del más pequeño insecto, supera a cualquier computadora en tareas relacionadas con el mundo real, en donde entradas ambiguas se tienen que interpretar para iniciar acciones. En términos de consumo la eficiencia del cerebro está en 3×10^{14} operaciones por joule, mientras que un procesador está alrededor de 6.25×10^6 operaciones por joule. La súper computadora BlueGene diseñada por IBM y premiada como la mejor supercomputadora en el 2006 puede lograr un pico teórico de 3.67×10^{15} operaciones de punto flotante por segundo con una potencia consumida de 1×10^6 W, correspondiendo a 1×10^9 operaciones por joule, aún con muchos ordenes de magnitud menor que la

eficiencia del cerebro [1]. En comparación una abeja con solo 1 millón de neuronas tiene una eficiencia en energía de 7 órdenes de magnitud que el circuito más eficiente creado actualmente, realizando 10,000 GFLOPS que son 1,000 veces más que las computadoras más potentes. En resumen una computadora compuesta de muchos elementos precisos falla en resolver tareas que parecen triviales para el más pequeño cerebro [2].



Fig. 1.1: Computadora BlueGene (izquierda) y una abeja (derecha), diferentes tamaños similares capacidades.

Uno de los objetivos de la ingeniería neuromórfica es entender estas razones e implementar dispositivos más eficientes para explotar las estrategias desarrolladas por el cerebro y la física del sustrato de silicio.

Carver Mead [3], padre de la ingeniería neuromórfica, atribuye que las razones de la superioridad del cerebro sobre las computadoras digitales se deban al uso de fenómenos físicos elementales como primitivas computacionales. La ingeniería neuromórfica aplica los principios computacionales descubiertos en los organismos biológicos a esas tareas que los sistemas biológicos realizan fácilmente, pero que se han probado difíciles de hacer usando técnicas tradicionales. Los sistemas neuromórficos, en vez de implementar redes neurales remotamente relacionadas con estos tipos de sistemas, son dispositivos de hardware conteniendo circuitos analógicos que tratan de emular en detalle las propiedades y los procesos físicos embebidos bajo la computación neuronal. El medio más cercano ampliamente accesible para la comunidad de investigadores que permita implementar modelos detallados de sistemas neurológicos es el Silicio. La tecnología VLSI, suponiendo que se use en débil inversión o en el subumbral, comparte las mismas primitivas que las neuronas que no son explotadas por la tecnología digital [1] [4].

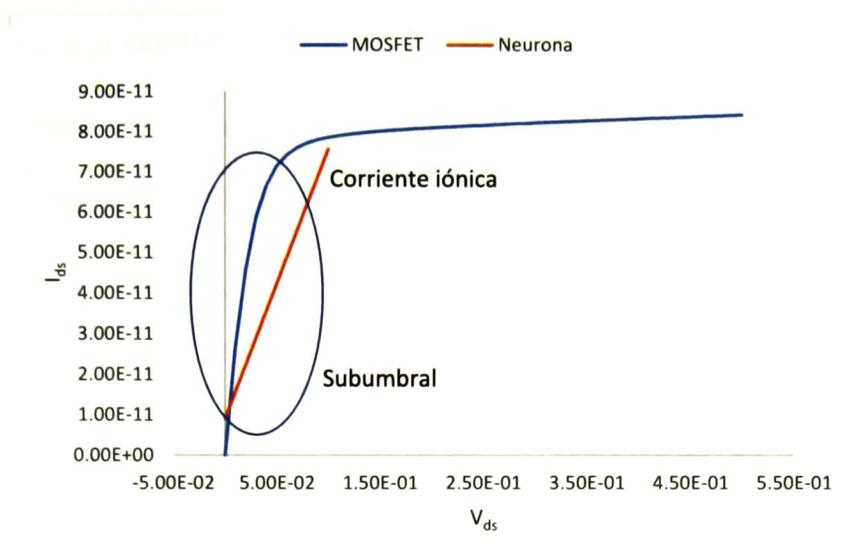


Fig. 1.2: Comparación entre propiedades de un transistor y una neurona.

En los circuitos neuromórficos, con transistores CMOS en subumbral, las primitivas de computación están basadas en la física de los dispositivos, y emulan el comportamiento de las neuronas [5] [6] [7] [8]. Los circuitos neuromórficos son parte de un concepto desarrollado por Carver Mead [3], en el que describe el uso de sistemas VLSI (del inglés *Very Large Scale Integration*) para imitar las arquitecturas neurobiológicas presentes en los sistemas nerviosos. Debido a que los circuitos neuromórficos comparten las mismas arquitecturas que sus contrapartes biológicas, estos pueden ser usados para implementar modelos biológicos para el procesamiento sensorial.

Los circuitos integrados basados en circuitos neuromórficos, para propósitos de detección del movimiento, han sido un intento por demostrar, en pequeña escala, las capacidades de las primitivas de computación de bloques elementales para realizar funciones complejas en tiempo real, que emulen algunas funciones de los mecanismos de visión de los seres vivos, incluyendo su bajo consumo de potencia. Su principal característica es que el único software que se requiere es la topología de los circuitos y sus interconexiones.

1.2 Los chips neuromórficos de visión

Un sensor de visión implementado típicamente en dispositivos analógicos VLSI, son dispositivos que emulan modelos de sistemas visuales biológicos y que pueden ser usados para realizar tareas de visión [5]. Este tipo de dispositivos y sistemas ofrecen una alternativa de bajo costo para una máquina de visión, ya que pueden ser usados para reducir la carga computacional en un sistema digital, o de manera ideal, realizar toda la computación sin necesidad de hardware adicional.

Los sistemas de visión neuromórficos basan su comportamiento en lo que se conoce actualmente como retinas de silicio, sensores que emulan con éxito a la retina biológica de los seres vivos, en sus propiedades de preprocesamiento, de extracción, codificación y transmisión de la información visual, así como en la optimización de su consumo energético. Carver Mead, propuso el primer modelo de una retina sintética empleando circuitos neuromórficos [6] capaz de emular el comportamiento de las células bipolares en la retina de los animales vertebrados, posteriormente C. Mead y M. Mahowald [7] publicaron un importante artículo que discute la funcionalidad de la retina desde una perspectiva de ingeniería. En 1996 T. Delbruck y C. Mead publican los detalles de una nueva retina de tiempo continuo, adaptativa, logarítmica basada en fotodiodos y circuitos innovativos de retroalimentación [8]. K. Boahen [9] propone una aproximación usando foto transductores en modo corriente, amplificación adaptativa, filtrado y cuantización, que permiten la transmisión de la información a un bloque jerárquico superior para su posterior procesamiento. En el 2008, T. Delbruck, P. Lichtsteiner y C. Posch desarrollaron una cámara de visión asíncrona de contraste temporal con 120 dB de ganancia y 15 μ s de latencia que puede comunicar su información basada en eventos a una computadora o a otro chip neuromórfico por medio de una interface llamada AER (Address Events Representation). Esta cámara está ya en su etapa comercial.

En lo que se refiere a las aplicaciones de estas retinas neuromórficas para la detección del movimiento, pueden mencionarse los artículos basados en métodos de igualación explícita en el dominio del tiempo [10], en los procesos basados en el gradiente [5] [11] y en los métodos de correlación

[12]. Estas propuestas se basaron en arreglos unidimensionales de sensores neuromórficos de la época.

Actualmente los arreglos de sensores son bidimensionales y se han agregado sistemas que a partir de la información de preprocesamiento producida por los sensores, encausan la información hacia sistemas de mayor jerarquía [13] [14] [15] [16], tal como lo hace el córtex del cerebro humano para tomar decisiones cognitivas, inteligentes. Esta habilidad, conocida como “atención visual” en los sistemas biológicos, es capaz de concentrar los recursos de computación limitados, en situaciones de sobrevivencia (bottom-up) o de cognición, evolutivas (top-down), y es representada en los circuitos neuromórficos como redes WTA, “el ganador toma todo”, lo más importante prevalece.

El propósito de este trabajo es profundizar en el análisis y simulación de distintos bloques de visión neuromórficos publicados en la literatura.

1.3 Propuesta de tesis.

Los chips neuromórficos de visión analizados generalmente están compuestos de los siguientes bloques:

1. Un bloque fotorreceptor adaptativo de contraste espacio-temporal de amplio rango dinámico.
2. Un bloque convertidor diferencial de modo voltaje a modo corriente, un amplificador de transconductancia, que determina la derivada espacial de la señal.
3. Un bloque seleccionador en modo corriente que determina el signo del cambio en la derivada espacial.
4. Una red en modo corriente que puede tomar decisiones sobre la señal más importante sobre las demás, “Winner Take All, WTA” y seguirla en forma dinámica, así como eliminar el ruido de las derivadas y definir los campos de visión del fotorreceptor (contraste por extremos).

5. Un bloque que determina la posición de la señal anterior con propósitos de control de la trayectoria de un vehículo en el cual está montado el chip de visión.

En esta Tesis se analizó y simuló un chip de visión propuesto por Giacomo Indiveri [5] del Instituto de Neuroinformática de la Universidad de Zurich. El Chip utiliza un fotorreceptor unidimensional formado por 60 píxeles, que forman parte de 60 columnas con los bloques 2 y 3. Los bloques 4 y 5 son bloques integradores. El algoritmo de percepción del movimiento utilizado es el “algoritmo de intensidad constante” o el “método del gradiente” [11]. La simulación fue realizada empleando los parámetros Spice actualizados de Bsim3.v3 de una tecnología 0.5 μm de ON Semiconductor, proporcionados por MOSIS, en una plataforma de simulación de Tanner EDA y validados, en circuitos individuales, usando AIMSpice. La metodología utilizada fue la de reproducir, en simulación, los resultados experimentales de las fuentes originales de los bloques y de circuitos analizados más trascendentes en la literatura. La principal señal de excitación fue la de una barra negra de longitud L desplazándose en un fondo blanco con una velocidad determinada en una dimensión.

Dos grandes bloques componen el chip de visión.

Un primer bloque que integra los fotosensores con una capacidad de contraste temporal y espacial que pudiera producir un mapa de características ópticas de movimiento sobresalientes (“salicency map”).

Un segundo bloque capaz de seleccionar la más fuerte, de una pluralidad de señales de entrada y tomar decisiones acerca de la posición de esta señal y comandar una señal de respuesta conveniente para la navegación segura.

El segundo bloque utiliza tecnología CMOS con procesos de fabricación estándar. En el caso de los fotorreceptores se requiere la fabricación de fotodiodos. Los fotodiodos son transistores canal n , en los cuales, en la región de la fuente se deja una apertura para que pase la luz al diodo formado por la región $n+$ y el substrato tipo p y genere una señal opto electrónica, el transistor funcionara en una configuración seguidor de fuente.

El segundo bloque comprendería el circuito WTA y los circuitos actuadores, definiendo bien las especificaciones y la funcionalidad necesaria para resolver este problema. En este caso la naturaleza y rangos de las variables de entrada que simulen la dinámica del sistema.

1.4 Objetivos de la tesis.

Una de las principales motivaciones de esta tesis es iniciar una etapa de aprendizaje acerca del diseño de circuitos y sistemas basados en la ingeniería neuromórfica para la emulación de funcionalidades propias de los sistemas biológicos. Un chip de visión como los ilustrados en la literatura, es un buen ejemplo para este propósito.

En esta tesis implementaremos la demostración de un sistema neuromórfico simple en una aplicación de navegación, en la que se ejemplifican las primitivas de computación propias de los circuitos neuromórficos con funciones específicas: sensado, procesamiento, atención y actuación. Definiendo apropiadamente las especificaciones y la funcionalidad necesaria para resolver este problema.

1.4.1.1 Objetivo General.

Analizar y simular los bloques de un chip de visión.

1.4.1.2 Objetivos Particulares.

Analizar de forma detallada los diferentes bloques del sistema usando herramientas de simulación.

1. Desarrollo para la infraestructura de simulación de sensores opto electrónicos:

Irradiancia al nivel del plano de la imagen. Magnitud de fuentes de corriente en Spice.

Algoritmos para las señales de entrada a los bloques del chip de visión en función de la velocidad de una barra y sus dimensiones.

2. Desarrollo de los métodos de análisis y simulación en Spice de los diferentes bloques de construcción del chip de visión.

Implementación en simulación usando la herramienta “Schematic Editor” de Tanner EDA Circuit Simulator.

3. Validación de los métodos en función de los resultados experimentales de la literatura conducente.

1.5 Organización de la tesis.

Capítulo 2. Se presenta la arquitectura de un sistema de visión, se describe una metodología para detectar el movimiento de la escena visual basada en un modelo llamado mapa de relevancia “Saliency Map”, en el cual la escena visual se descompone en información elemental para el procesamiento del movimiento.

Capítulo 3 se describe el bloque de fotorreceptor, parámetros importantes y características de este dispositivo.

Capítulo 4 se describe el bloque de derivada espacial y de selección. Este bloque del sistema es parte del procesamiento del movimiento y tiene como principal objetivo realizar una derivada espacial de la escena visual que se proporciona por los fotorreceptores.

Capítulo 5 se describe el Bloque WTA, su concepto en las ciencias de la computación, simulaciones y funcionamiento en bloque del sistema completo.

Capítulo 6 se establecen las conclusiones finales.

Capítulo 2

Descripción del chip de visión neuromórfico.

Este capítulo presenta la descripción general de un chip de visión, su arquitectura, los algoritmos de percepción del movimiento, la infraestructura para su simulación y la comparación de las propiedades de emulación del chip neuromórfico con las funcionalidades de la visión.

2.1 El chip de visión.

Estos sistemas típicamente contienen elementos fotosensibles y elementos de procesamiento, aplican la secuencia de selección competitiva a un estímulo visual sentido y evaluado en el mismo plano de procesamiento y se realizan operaciones de seguimiento visual.

En este capítulo se describe una arquitectura de seguimiento que reduce el costo computacional de las diferentes etapas de procesamiento

interconectadas a ésta, llevando a cabo una extensa cantidad de cómputo, y transmitir solo el resultado de este cálculo, en lugar de grandes cantidades de datos que representan la imagen de entrada en bruto. Esta arquitectura selecciona bordes de alto contraste independientes de la iluminación absoluta de la escena y usa un WTA histerético con retroalimentación positiva y acoplamiento lateral para anclar y suavemente seguir un objetivo seleccionado.

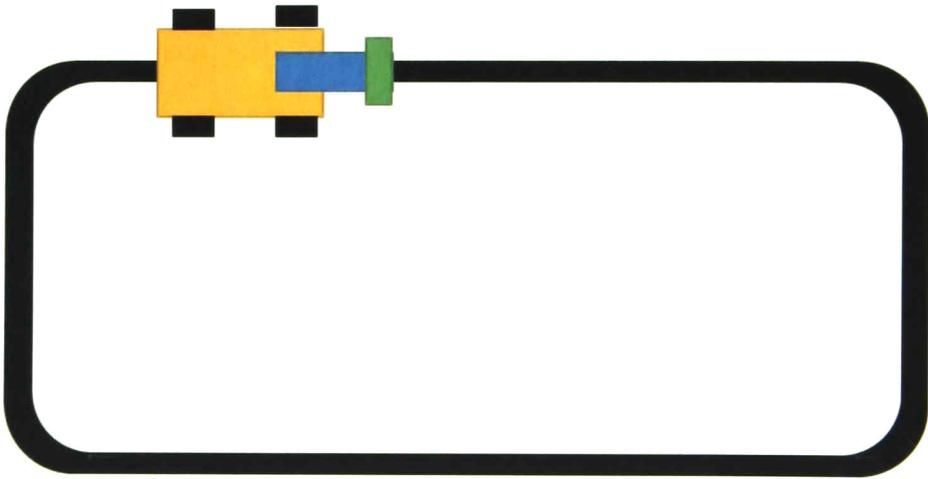


Fig. 2.1: Diagrama que ejemplifica el funcionamiento del vehículo acoplado con un sensor neuromórfico.

El chip de visión tiene como tarea principal seguir un camino indicado por una línea negra en un fondo blanco, el sensor de visión es puesto en el frente del vehículo motor, dependiendo del movimiento de la línea en el campo visual, el sensor controlará los motores de tal forma que éste pueda girar y mantener la línea en una posición central al campo de visión, en la **Fig. 2.1** se aprecia un diagrama donde se muestra el vehículo acoplado con el sensor neuromórfico en una pista negra sobre un fondo blanco.

Los sensores de visión neuromórficos, típicamente son dispositivos analógicos VLSI que implementan modelos de sistemas biológicos y que pueden ser usados para tareas de visión por computadora [5]. Giacomo Indiveri propuso una arquitectura para hacer la tarea de atención visual usando componentes neuromórficos, este trabajo es, en especial, interesante debido la división de las tareas en bloques independientes, donde se puede identificar la tarea de cada elemento.

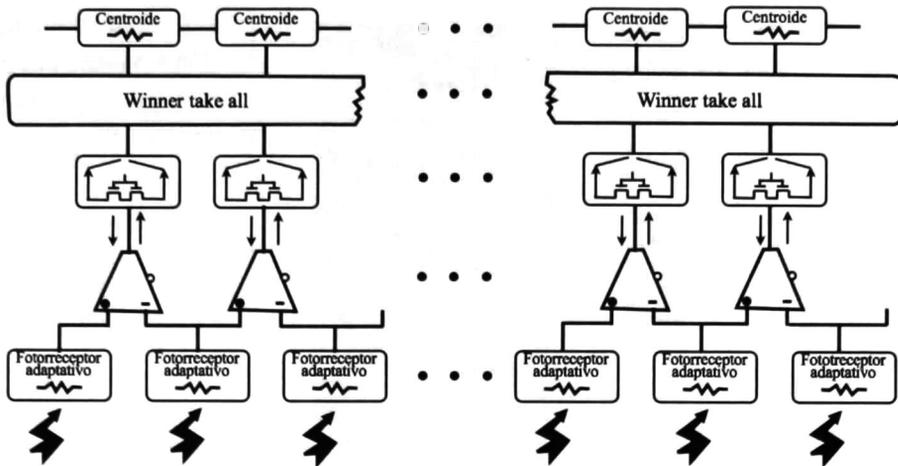


Fig. 2.2: Diagrama de bloques de un sistema de seguimiento en un solo chip.

En esta arquitectura la iluminación es procesada en paralelo por cinco bloques principales en una estructura jerárquica y puede ser implementado en un dispositivo simple.

Indiveri propone una arquitectura unidimensional, que se diseña como columnas de manera delgada y alargada, como una forma para optimizar espacio e incrementar el número de píxeles, **Fig. 2.3**.

Indiveri [5] construyó dos chips de un tamaño aproximado de $2\text{ mm} \times 2\text{ mm}$ usando tecnologías CMOS de $2\text{ }\mu\text{m}$ y $1.2\text{ }\mu\text{m}$. Las columnas de procesamiento son de un tamaño de 60λ donde λ es el parámetro de escalamiento de la tecnología correspondientes a $1\text{ }\mu\text{m}$ en la tecnología de $2\text{ }\mu\text{m}$ y $0.6\text{ }\mu\text{m}$ en la tecnología de $1.2\text{ }\mu\text{m}$; como algunos elementos del circuito no son escalables con λ (como los capacitores) los chips son algo diferentes. El chip de $2\text{ }\mu\text{m}$ tiene un *pitch* de $60\text{ }\mu\text{m}$ y contiene 25 columnas de procesamiento, mientras que el chip de $1.2\text{ }\mu\text{m}$ tiene un *pitch* de $36\text{ }\mu\text{m}$ y 40 columnas de procesamiento.

El primer nivel de la arquitectura consta de un arreglo de fotorreceptores que convierte la intensidad de la imagen en un voltaje logarítmico. La respuesta de este circuito es invariante a la intensidad absoluta de la escena, el fotorreceptor adaptativo exhibe características de un filtro pasa banda temporal, con frecuencias de corte, altas y bajas ajustables.

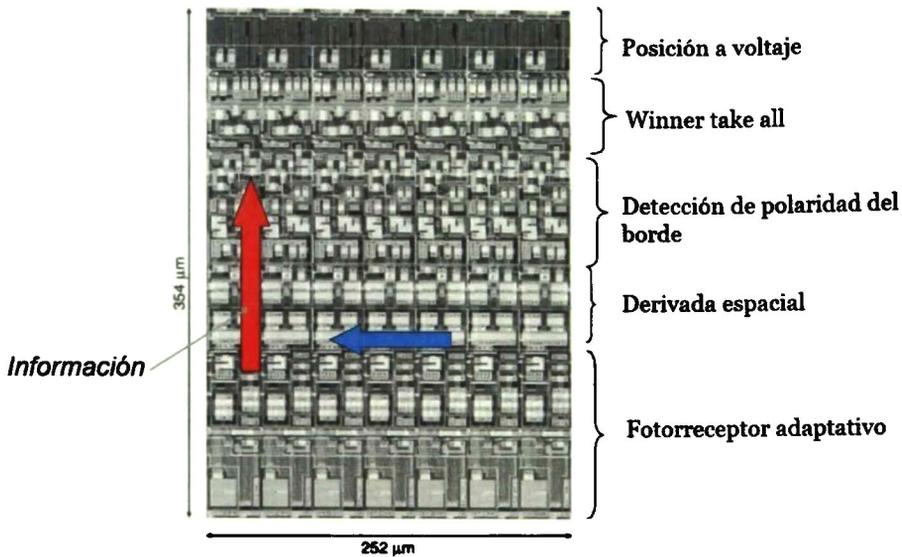


Fig. 2.3: *Layout* del circuito fotorreceptor diseñado por Indiveri.

El segundo nivel está compuesto por un arreglo de amplificadores de transconductancia simples operados en el régimen del subumbral, estos reciben las salidas de voltaje de sus fotorreceptores vecinos, la amplitud de su salida de corriente codifican la intensidad del contraste de los bordes y el signo de su polaridad.

En una tercera etapa la polaridad de cada borde se computa para que el sensor responda selectivamente ante bordes que incrementan el contraste, ON, o que disminuyan a éste, OFF, o ambos. La cuarta etapa se compone de una red hWTA que selecciona y bloquea una característica con el mayor contraste espacial moviéndose a una velocidad que mejor capte el fotorreceptor. Finalmente la última etapa consiste en un circuito de posición a voltaje que codifica la posición espacial del WTA en una salida analógica simple.

2.2 Algoritmo de percepción del movimiento.

El circuito diseñado por Indiveri fue concebido para implementar modelos de atención en hardware [1]. La atención es una de las más poderosas estrategias usadas en los sistemas biológicos donde toda la computación artificial puede tener una ventaja. En un sensor biológico la atención visual

actúa como un filtro que selecciona la característica más sobresaliente, subsecuentemente centrando recursos computacionales para analizar los detalles del objetivo. Parte de la investigación en el contexto de la atención está siendo encaminado a descubrir los mecanismos para la selección del objetivo a considerar [4].

En nuestra época la atención visual es básicamente definida de la siguiente forma: “Es una habilidad en la que a una criatura se le permite dirigir su mirada rápidamente hacia un objeto de interés en el campo visual” Este enunciado tiene dos conceptos importantes: el primero “rápidamente”, lo que significa es que se da en tiempo real, y el segundo, que tenemos es “objeto de interés”, que se refiere a la habilidad de identificar esos objetos o regiones en el campo visual que contengan información importante. El problema es identificar qué es importante en este momento, eso es extremadamente complejo. La atención visual es una tarea muy intrincada la cual es realizada muy bien por las criaturas vivientes y como resulta ser es una tarea difícil de imitar por los sistemas artificiales, demanda enormes capacidades de procesamiento [17].

En el trabajo propuesto por Treisman and Gelade se estudia la atención visual en el contexto de una búsqueda visual, distinguiendo dos diferentes tipos de estímulos: aquellos estímulos que “saltan” que son inmediatamente reconocidos por el observador sin importar el número de distractores de la escena, y están los estímulos por los que se está realizando una búsqueda explícita [18]. De estas observaciones se distinguen dos modelos de observación: uno es rápido, un proceso de reflejo, la cual dirige la mirada a un estímulo inesperado, llamado bottom-up lo que sugiere un origen en las áreas bajas del cerebro y se proyecta hacia las áreas corticales donde se aplica el estímulo, el segundo es lento, un método intencional el que se origina en las zonas altas del cerebro y viaja hasta los ojos, este método es llamado top-down.

2.2.1 Un modelo basado en mapas de relevancia (*Saliency map*).

Los mecanismos de atención visual son requeridos para resolver el problema de saturación por información, de una limitada capacidad de procesamiento del sistema. Estos se presentan en los sistemas sensoriales biológicos y

pueden ser una herramienta de ingeniería útil para sistemas visuales artificiales.

Entre los modelos de atención selectiva [1] basados en el concepto de mapas de relevancia, se encuentra el propuesto por Koch y Ullman [19], este modelo ha llevado a muchas implementaciones en software aplicadas a tareas de visión por computadora y robótica. Estos son especialmente llamativos porque se acoplan bastante bien a implementaciones de hardware.

Un diagrama que describe el proceso principal de los modelos basados en relevancia se muestra en la **Fig. 2.4**. Un conjunto de características es extraído de la escena visual. A todas estas características se les denominan *feature maps*, éstos son normalizados y combinados en un mapa de relevancia en donde topológicamente se guarda la información más sobresaliente de la escena visual, en donde un circuito WTA selecciona la característica más relevante y enfoca la atención.

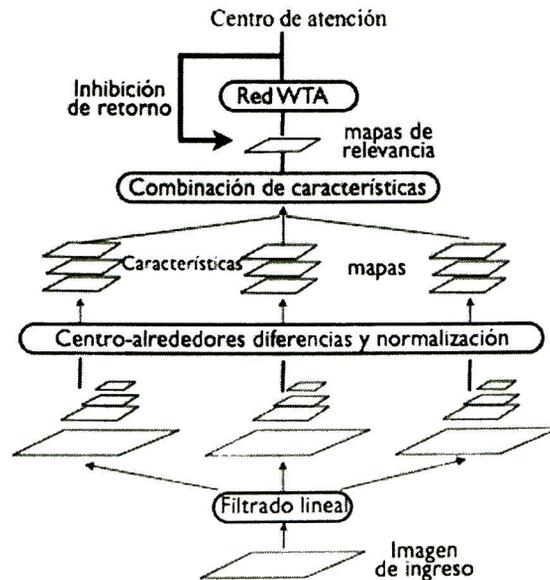


Fig. 2.4: Diagrama esquemático de un modelo de atención selectiva basado en Mapas de relevancia.

Estos modelos de relevancia son bastante modulares por lo que un sistema neuromórfico que lo implementa puede ser escalado a sistemas de atención visual muy complejos.

Muchos de los modelos que trabajan con mecanismos bottom-up usan los mismos principios. Primero el valor resaltante final está compuesto por varios valores sobresalientes cada uno pertenece a diferentes características de bajo nivel (éstos son llamados características dependientes de valores sobresalientes *feature dependent saliency values*). Segundo, en un punto el valor relevante esencialmente depende del contexto de los alrededores, esto es, el valor no es igual al máximo en un valor absoluto, pero es proporcional al contraste que compone con sus alrededores cercanos. Tercero, el mapa de relevancia está basado en valores de relevancia dependientes de las características, con diferentes pesos; los pesos dependen de una modulación top-down y puede influenciarse a través de un entrenamiento. Cuarto, el entendimiento de la escena y el reconocimiento de objetos están fuertemente relacionado con la dirección de la mirada.

2.2.2 El movimiento.

El movimiento de elementos y regiones en una imagen no está dado directamente, por lo que tiene que ser calculado de mediciones más elementales. El primer registro de la luz por el ojo o por un dispositivo electrónico puede ser descrito como un arreglo de valores de intensidad de la luz dependientes en el tiempo, $I(x, y, t)$. El movimiento en la imagen se puede describir en términos de un vector de movimiento $V(x, y, t)$ que determina la velocidad en un punto de la imagen en las coordenadas (x, y) en un tiempo t . El primer problema para analizar el movimiento es el cálculo de $V(x, y, t)$ desde $I(x, y, t)$.

En algunos casos puede ser suficiente con tan solo detectar ciertas propiedades del vector de movimiento, en vez de medirlo por completo. Por ejemplo al identificar el movimiento de un objeto rápidamente, el movimiento se debe detectar, pero no necesariamente se debe medir.

Para detectar el movimiento se puede realizar en diferentes niveles en el procesamiento de una imagen, utilizando diferentes primitivas de movimiento. Es útil hacer una distinción entre dos esquemas principales. En un nivel bajo, las mediciones de movimiento pueden ser basados directamente en cambios locales de valores en la intensidad de luz; estos son llamados esquemas basados en intensidad. Alternamente es posible primero identificar características como bordes, esquinas o regiones, y luego medir el movimiento al juntar estas características con el tiempo, y detectando su

cambio de posición, estos tipos de esquemas son llamados *token-matching*. Estos modelos de detección y medición de movimiento proponen diferentes problemas de cálculo, y consecuentemente diferentes tipos de proceso tanto en los sistemas de visión por computadora como biológicos [20].

2.2.3 Esquemas basados en intensidad.

Un método basado en la intensidad es el método del gradiente, en este esquema la medición de movimiento se deriva de la comparación entre gradientes de intensidad y cambios de intensidad temporales. Un ejemplo unidimensional que ilustre el principio básico es mostrado en la **Fig. 2.5**. Considere el perfil de intensidad (intensidad I como función de x) mostrado por la curva sólida de la **Fig. 2.5**. En el punto x_0 se tiene una pendiente positiva, si hay movimiento a la derecha, como se muestra por la curva punteada, la intensidad $I(x_0)$ decrecerá, al moverse al lado derecho ésta incrementará, por lo que indica la dirección del movimiento, y de donde los cambios de magnitud de las intensidades espaciales y temporales determinan la velocidad del movimiento. En principio, se pueden obtener mediciones del movimiento donde sea que el gradiente de intensidad no sea cero, aunque la medición es más confiable en los bordes donde el salto de intensidad es inducido.

La velocidad del movimiento y la pendiente se determinan por observaciones locales. De estos dos valores el observador puede inferir la velocidad de la curva de intensidad que pasa. El observador puede implementarse como un arreglo de sensores, cada sensor produce un voltaje relacionado a la intensidad de la luz que incide en él. En el punto x_0 la derivada espacial puede determinarse al tomar la diferencia entre las intensidades de los sensores vecinos y dividiéndose por el espacio entre estos. Esta aproximación es bastante buena si los sensores están cercanos.

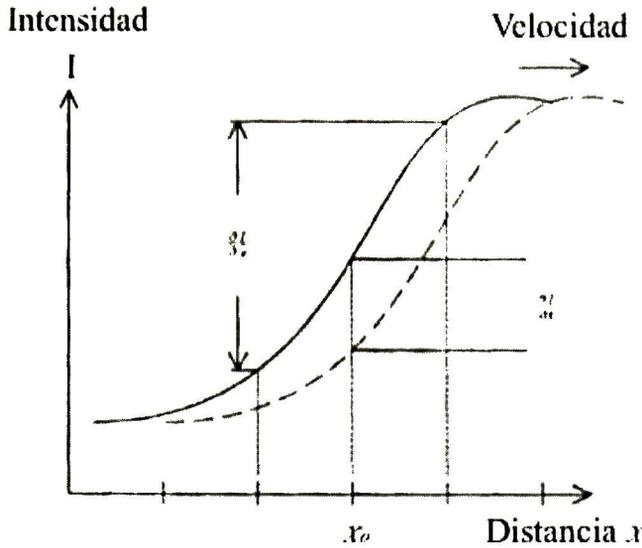


Fig. 2.5: Una imagen para la detección de movimiento en una dimensión.

También se puede determinar el cambio de la intensidad local en el tiempo al tomar la derivada en el tiempo de la señal de intensidad. Al conocer el gradiente de intensidad (pendiente) y al conocer qué tan rápido cambia la intensidad se puede calcular la velocidad. La ecuación de la línea tangente de la curva de intensidad en el punto x_0 es:

$$I(x, t) = m[(x - x_0) - vt] + I_0 \quad (2.1)$$

donde v es la velocidad de la imagen y m es la pendiente de la línea.

Las derivadas del tiempo y espacio son:

$$\begin{aligned} \frac{\partial I}{\partial x} &= m \\ \frac{\partial I}{\partial t} &= -mv = -\frac{\partial I}{\partial x} v \end{aligned} \quad (2.2)$$

Por lo tanto

$$v = -\frac{\frac{\partial I}{\partial t}}{\frac{\partial I}{\partial x}} \quad (2.3)$$

Una gran dificultad en las imágenes naturales es que en el mundo hay una gran variedad de motivos que causan cambios en la intensidad de una

imagen, estos son: cambios de iluminación, éstos incluyen las sombras, fuentes de luz visibles, y gradientes de iluminación; cambios en la orientación o distancia del observador a las superficies visibles; cambios en la reflectancia de la superficie de objetos en la imagen. Un punto crítico aquí es considerar que estos motivos de cambios de intensidad se pueden localizar espacialmente [21].

Cuando ocurra un cambio de intensidad habrá un pico correspondiente a la primera derivada espacial. Un cambio de intensidad se puede definir de esta manera, así la tarea de detectar esos cambios puede reducirse a encontrar los picos de la derivada espacial de intensidades [21]. Desde un punto de vista computacional, restringiendo la detección del movimiento a la localización de los máximos en el gradiente de intensidades se obtienen ventajas, al detectar los lugares donde el gradiente de intensidad es máximo se puede obtener una medición de la velocidad más confiable; los cambios máximos de intensidad están más ligados a características físicas. Por ejemplo, los bordes de un objeto, quiere decir que si un máximo de intensidades se mueve es posiblemente a consecuencia del movimiento de un objeto [20].

2.2.4 Detección de bordes en una imagen estática.

Para la detección de bordes en una imagen D. Marr [21] hace notar la importancia de filtrar la imagen para una debida localización. Hay dos consideraciones físicas para determinar el tipo de filtro: uno sobre el dominio del espacio y otro sobre el dominio de la frecuencia. Estos requerimientos en el dominio de la frecuencia y el espacio son conflictivos, sin embargo, la única distribución que optimiza esta relación es la Gaussiana.

$$f(x) = G''(x) * I(x) \quad (2.4)$$

donde $I(x)$ es la imagen y $G''(x)$ es la segunda derivada del gaussiano, este operador en una dimensión se puede escribir como:

$$G''(x) = \left(-\frac{1}{\sigma^3 \sqrt{2\pi}} \right) \left(1 - \frac{x^2}{\sigma^2} \right) e^{-\frac{x^2}{2\sigma^2}} \quad (2.5)$$

La segunda derivada del gaussiano se asemeja al operador “sombrero Mexicano” y comparte similitud con el operador de la diferencia de gaussianos, el cual tiene la propiedad de aproximarse a un filtro pasa-banda.

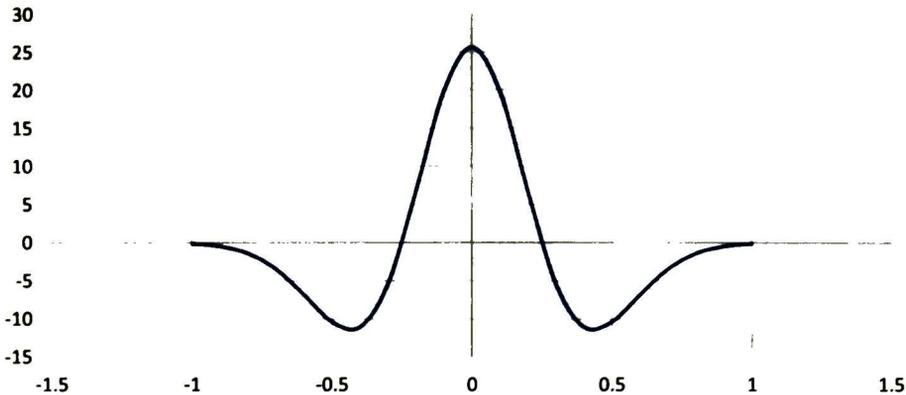


Fig. 2.6: El operador G'' , la segunda derivada de la distribución gaussiana.

Así la tarea de detectar los cambios de intensidad en una imagen se reduce en encontrar el cruce por cero de la segunda derivada de la imagen filtrada por las gaussianas. Este tipo de filtros son ampliamente usados en el procesamiento de imágenes debido a que imita al proceso que hace la retina para detectar bordes.

Un problema que surge al usar este tipo de filtros es identificar la orientación asociada a la doble derivada, para escoger qué derivada direccional usar. Este es un problema particular para nosotros debido al tipo de sensor que se plantea, éste es un sensor unidimensional donde la operación se aplica sobre una dirección particular.

Para determinar la orientación hay que tener en cuenta que la motivación de detectar los cambios en la intensidad, y que estos cambios corresponden a propiedades físicas de un objeto a detectar, deben ser consideradas cuidadosamente. Estas propiedades son espacialmente continuas y casi todas pueden ser asociadas con una dirección que se proyecta en una orientación de la imagen. La orientación que escogamos es tal que la dirección del objeto coincida con sus cruces por cero. De igual manera esta orientación es donde los cruces por cero tienen su pendiente máxima, medido perpendicularmente desde el eje de orientación donde se encuentran los cruces por cero.

Para ejemplificar en donde se encuentra la orientación de una imagen, vea la **Fig. 2.7**, en esta figura la derivada direccional no es cero exceptuando a lo largo del eje Y.

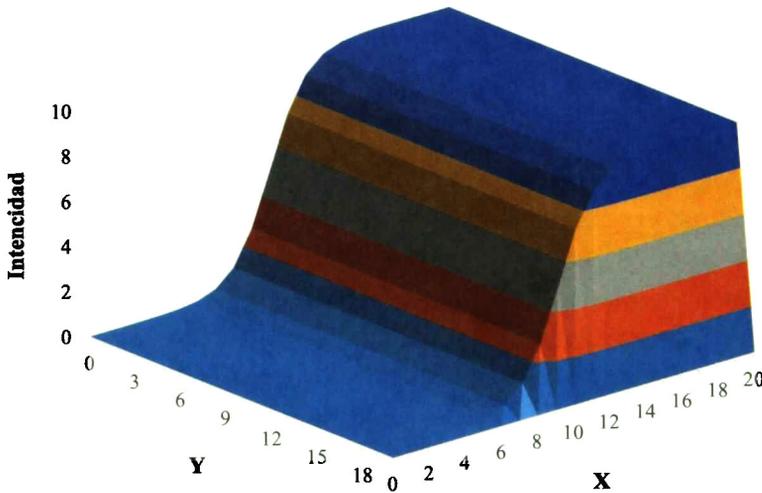


Fig. 2.7: Superficie que muestra el cambio de intensidades en un área.

En la **Fig. 2.7**, la orientación está en el eje y, debido a que la segunda derivada de este perfil en el eje y es cero sobre todo este eje, por lo dicho anteriormente la derivada direccional que se debe escoger es la perpendicular al eje de orientación, es decir $\partial^2 I / \partial x^2$, además la pendiente de los cruces por cero tomados en este eje es la máxima como se muestra en la **Fig 2.8**, donde se comparan las dobles derivadas del perfil de intensidades en diferentes orientaciones.

Al aplicar este concepto al caso que se trata en esta tesis, podemos decir que la orientación de la imagen a procesar se encuentra en la dirección de movimiento del vehículo, por lo tanto la operación para determinar los bordes de la imagen será realizada en la perpendicular a la dirección, es decir, una tira de sensores unidimensionales será colocada de manera perpendicular a lo largo de la línea.

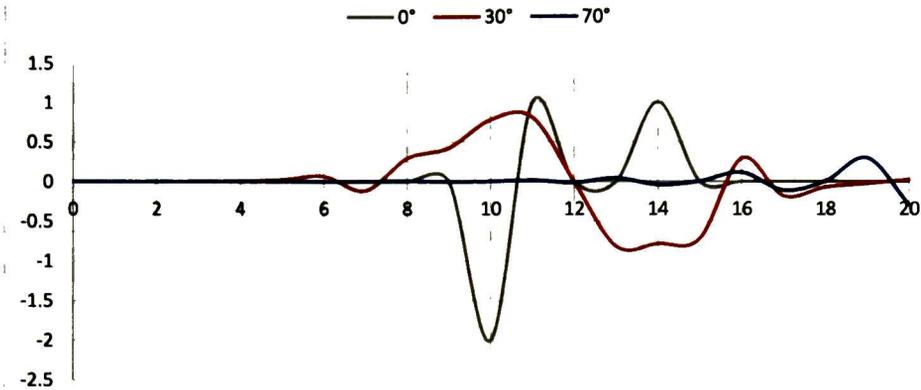


Fig. 2.8: Dobles derivadas de la superficie de intensidades en tres diferentes orientaciones.

2.3 Propiedades en subumbral de un transistor.

Como se mencionó antes, en la electrónica es posible implementar los mismos procesos físicos que suceden en las neuronas al usar transistores MOS operados en la región de subumbral.

Los transistores MOS están en la región de subumbral cuando el voltaje entre la compuerta y la fuente (V_{gs}) está debajo del voltaje de umbral (V_{th}). En el subumbral la corriente de drenaje tiene una relación exponencial, la cual está dada por [22]:

$$I_{ds} = \frac{W}{L} I_0 e^{(1-\kappa)\frac{V_{BS}}{U_T}} e^{\left(\kappa\frac{V_{GS}}{U_T}\right)} \cdot \left(1 - e^{\left(\frac{-V_{ds}}{U_T}\right)} + \frac{V_{DS}}{V_E}\right) \quad (2.6)$$

donde W y L son el ancho y la longitud del transistor, I_0 es la corriente de polarización, κ es el coeficiente de subumbral, U_T es el voltaje térmico, V_E es el voltaje de *early* y V_{GS} , V_{DS} y V_{BS} son los voltajes compuerta-fuente, drenaje-fuente, sustrato-fuente, respectivamente.

En el subumbral si el voltaje V_{DS} es menor a $4U_T$ el arrastre de los portadores mayoritarios en el transistor NMOS es trivial. Como consecuencia, I_{ds} es principalmente producido por los portadores minoritarios (electrones) que se difunden en la dirección del gradiente de concentración, que es controlado por el campo eléctrico horizontal creado por V_{ds} (los electrones, portadores minoritarios, fluirán desde la fuente hacia

el drenador si $V_{ds} > 0$, o desde el drenador a la fuente si $V_{ds} < 0$.) En otras palabras, en el subumbral la corriente de empobrecimiento (generada por difusión) sobrepasa la corriente de inversión (generada por arrastre).

Entendido lo anterior, podemos decir que un transistor MOS en subumbral se comporta como un transistor bipolar npn, donde el drenador, el substrato y la fuente actúan como el colector, la base y el emisor, respectivamente. (Cuando $V_{ds} < 0$, el drenador actúa como emisor y la fuente como el colector) [22] [23].

Así, empleando las formulas físicas clásicas apropiada para describir un transistor bipolar, encontramos la descripción general de I_d en un transistor NMOS en el subumbral, como sigue [22] [23]:

$$I_d = I_0 e^{\left(\frac{\kappa V_g - V_s}{U_T}\right)} \left[1 - e^{-\frac{V_{ds}}{U_T}} \right] \quad (2.7)$$

I_0 se conoce como la corriente escalada cuyo valor depende principalmente de la geometría física de los transistores [22]. El símbolo κ representa la derivada del potencial de superficie Ψ_s con respecto al voltaje de compuerta V_g , que describe un divisor de voltaje capacitivo consistente de una capacitancia de compuerta u óxido, C_{ox} , y una capacitancia de empobrecimiento, C_d (κ es una constante que toma un valor entre 0.5 y 0.9) [23]).

Como puede verse, como en un transistor bipolar npn, la corriente que fluye en un transistor NMOS en el subumbral varía exponencialmente con respecto a la diferencias de voltaje entre sus terminales. Más aún, cuando $V_{ds} > 4V_T$, el último término en la anterior ecuación es trivial (esto es, I_d es casi independiente de V_{ds}) y el transistor en el subumbral se dice que opera en saturación. En tal caso, dado un V_{gs} fijo, la corriente del dispositivo en una gráfica I_d vs V_{ds} es casi plana, implicando una pequeña dependencia mostrada por I_d en V_{ds} cuyo peso es controlado por el factor de modulación de la longitud del canal (λ) que es el recíproco del voltaje *early*, V_e [22] [23] [24]. Esta particular característica es similar a la de un transistor que opera arriba del umbral.

Para describir un transistor NMOS saturado en el subumbral, podemos reescribir la ecuación (2.7) como sigue:

$$I_d = I_0 e^{\frac{\kappa V_g - V_s}{U_T}} \quad (2.8)$$

dado un potencial de la fuente dado (p. ej. V_s es una constante), la transconductancia del transistor NMOS puede expresarse como:

$$g_m = \frac{\partial I_d}{\partial V_{gs}} = I_0 e^{\frac{\kappa V_g - V_s}{U_T}} \frac{\kappa}{U_T} = I_d \frac{\kappa}{U_T} \quad (2.9)$$

que es similar a la expresión para la transconductancia de un transistor bipolar.

Con lo anterior en mente, se puede ver que un transistor MOS en el subumbral típicamente tiene una mayor transconductancia por unidad de corriente (g_m/I_d) comparada con la de un transistor arriba del umbral (suponiendo que ambos tienen una geometría idéntica) [22] [23] [24]. Como resultado, un transistor MOS en el subumbral es capaz de alcanzar un nivel de sensibilidad y una mayor ganancia de voltaje, que su contraparte arriba del umbral, mientras que ambos consumen la misma cantidad de potencia activa. Esta propiedad es quizás una de las principales razones por la cual se usan más los transistores MOS en el subumbral en las aplicaciones analógicas de baja potencia como son los circuitos WTA en tecnología CMOS.

Cuando el voltaje $V_{ds} > V_{gs} - V_{th}$, y $V_{DS} > 4V_T = 4kT/q$ se dice que el transistor está en saturación y la relación I_d vs. V_{ds} se describe por (2.10), ésta es ampliamente conocida como la relación I - V de ley cuadrática.

$$I_d = \mu_n C_{OX} \frac{W}{L} \frac{(V_{gs} - V_{th})^2}{2} = \mu_n C_{OX} \frac{W}{L} \frac{V_{ds}^2}{2} \quad (2.10)$$

donde μ_n es la movilidad electrónica cerca de la superficie del silicio, C_{ox} es la capacitancia de la compuerta por unidad de área. Si se toma en cuenta el fenómeno de la modulación de la longitud del canal (p. ej., el efecto *early*) [22] [23] [24], entonces la mencionada ecuación debe reescribirse como:

$$I_d = \mu_n C_{OX} \frac{W}{L} \frac{(V_{gs} - V_{th})^2}{2} \cdot \left(1 + \frac{V_{ds}}{V_c}\right) \quad (2.11)$$

V_c es el voltaje de *early*, y tiene una expresión general dada por:

$$V_c = L \frac{\partial V_d}{\partial L} \quad (2.12)$$

Adicionalmente el voltaje de *early* se emplea usualmente para caracterizar la resistencia de salida de pequeña señal de un transistor NMOS, esto es $r_{ds} \approx V_c/I_d$.

2.4 Herramientas e infraestructura de simulación.

El simulador elegido para realizar las simulaciones es Tanner SPICE, además de AIM SPICE para comparar los diferentes métodos de simulación, ambos simulan los transistores usando las librerías tomadas de Mosis usando una tecnología 0.5 μ m.

Una gran ventaja de la librería de Tanner es el simulador gráfico que simplifica la tarea de simulación, lo que permite una gran facilidad a la hora de realizar grandes simulaciones con múltiples complementos y conexiones.

Parte del problema de la simulación de un fotorreceptor es una forma de emular la reacción de la luz en los dispositivos; en la sección 2.5 se explica cómo calcular la irradiancia de un fotorreceptor, esta irradiancia es simulada con una fuente de corriente en paralelo del diodo receptor.

Gracias a este simulador la tarea de realizar las conexiones entre circuitos y los bloques numerosos de fotorreceptores no es complicado, con un entendimiento de la herramienta el trabajo se hace de manera eficiente; otros bloques simulados son los amplificadores, OTA y WTA, en los capítulos siguientes se explica con más atención cómo se simula cada uno y sus propiedades en la tarea de la detección y atención del movimiento.

2.5 Cálculo de la irradiancia en el plano focal del sistema óptico.

Los fotorreceptores modernos, basados en eventos en tiempo real poseen un rango dinámico de operación hasta seis décadas de iluminación, desde la iluminación producida en una noche con luna llena hasta un día soleado. La **Tabla 2.1** muestra estos rangos de operación.

Tabla 2.1: Valores de iluminación e irradiancia bajo condiciones típicas de iluminación.

Fuente de luz	Iluminancia	Irradiancia
Luna llena	1	0.1
Calle iluminada	10	1
Luz oficina	100-1000	10-100
Pleno sol	100000	1000

La Fig. 2.9 muestra un arreglo elemental básico de un circuito detector formado por un fotodiodo polarizado en inversa y un transistor NMOS en configuración de seguidor de emisor.

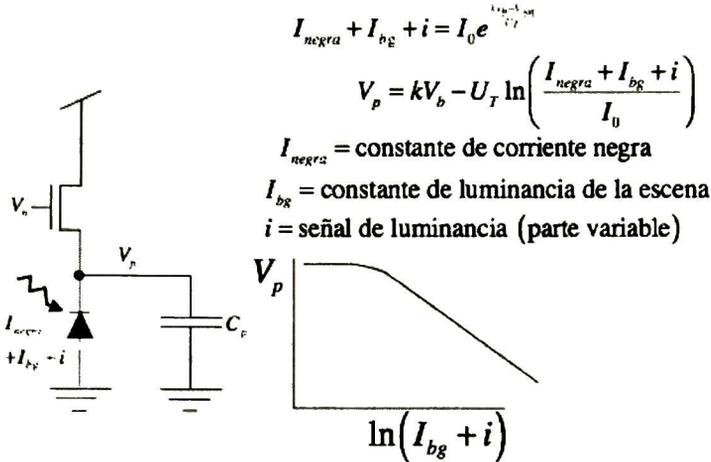


Fig. 2.9: Fotorreceptor seguidor de fuente.

La fotocorriente será proporcional a la irradiancia recibida y al área del fotodiodo. Aquí se utilizará para los cálculos un fotodiodo n-p, de un pixel desarrollado en el laboratorio de la Dra. Serrano en la Universidad de Sevilla [1] en el lado de la fuente del Transistor, de 40 μm de perímetro y 100 μm^2 de superficie.

La Fig. 2.10 muestra las características I-V de un elemento fotodetector de unión p-n. Pueden observarse dos distintas regiones de operación. La del cuarto cuadrante corresponde a su operación como celda fotovoltaica, en tanto que en el tercer cuadrante se muestra su operación como fotodiodo con polarización negativa. Una buena aproximación consiste en considerar que fotocorriente generada en el diodo polarizado en inversa es sensiblemente igual a la corriente de corto circuito de la celda fotovoltaica.

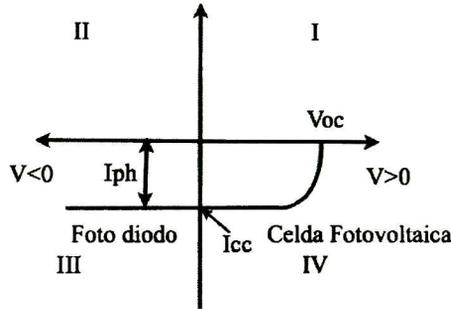


Fig. 2.10: Características I-V de una unión p-n bajo iluminación.

Las características de módulos solares comerciales se muestran en la **Tabla 2.2**. Los módulos están formados por arreglos de celdas solares conectados en serie, éstas son unidades cuadradas de silicio monocristalino de 12.5 cm de lado, con un área $A = 156.25 \text{ cm}^2$, capaces de generar una corriente de corto circuito de 5.72 A con una eficiencia global de conversión del 15% bajo una irradiancia AM1.5 de 1000 W/cm^2 . Luego, la densidad de corriente para esta irradiancia es, $J_{cc0} = 36.6 \text{ mA/cm}^2 = 36.6\text{E-}3 \text{ A/cm}^2$.

Tabla 2.2: Características fotovoltaicas de módulos solares comerciales típicos.

Tipo	205-72M	200-72M	195-72M	190-72M	185-72M
Potencia máxima (P_{mpp})	205	200	195	190	185
Voltaje en circuito abierto (V_{oc})	45.6	45.3	45.1	45.0	44.8
Corriente de cortocircuito (I_{sc})	5.82	5.72	5.63	5.56	5.48
Eficiencia del módulo	16.06%	15.67%	15.27%	14.88%	14.49%
Celdas	6x12 piezas en serie de celdas solares monocristalinas de 125mm×125mm (5 pulgadas)				

Si el fotodiodo de un pixel fuera comparable ópticamente a esta celda solar de alta eficiencia, produciría una fotocorriente de $I_{ph} = J_{cc} \cdot A_{diodo} = 3.66\text{E-}8 \text{ A}$, bajo las mismas condiciones de irradiancia.

Dado que la aplicación del circuito fotorreceptor implica el uso de una lente de distancia focal determinada, y un diafragma mecánico de pasos para ajustar el nivel de luminancia y formar una imagen en el plano focal **Fig. 2.11**, la irradiancia en el plano de la imagen, donde se sitúa el arreglo de pixeles, resulta inferior a la irradiancia de la escena o del objeto. Esta relación entre la luminancia en el plano del chip y la de la correspondiente luminancia de la escena, se describe en la referencia [8], mediante la fórmula:

$$I_{chip} = \frac{I_{scene}R}{8f^2} \quad (2.13)$$

donde I_{chip} , I_{scene} , es la irradiancia o luminancia del chip y de la escena, respectivamente, R es un coeficiente que toma en cuenta la luz reflejada, con un valor típico del 18% y f es el f -number o la “apertura relativa”, y es la relación entre la distancia focal de la lente al diámetro del diafragma. Este diafragma emula el papel de la pupila en el ojo humano.

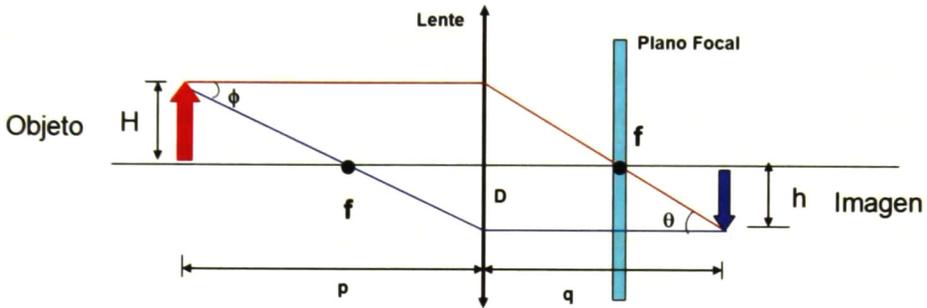


Fig. 2.11: Descripción del plano focal.

La **Tabla 2.3** muestra diferentes valores del f -number [3] en función del diámetro del diafragma en posiciones fijas (*stops*) para una lente de distancia focal de 6 mm, la distancia focal usual en aplicaciones de fotorreceptores basados en eventos.

Tabla 2.3: Numero f contra el diámetro de apertura del diafragma.

Stop	f (mm)	Apertura (mm)	A mm ²	f number
1	6	6	28.3	1.0
2	6	4.2	14.1	1.4
3	6	3.0	7.1	2.0
4	6	2.1	3.5	2.8
5	6	1.5	1.8	4.0
6	6	1.1	0.9	5.7
7	6	0.8	0.4	8.0

Los diámetros del diafragma (apertura) han sido escogidos para reducir el área del diafragma al 50%, para cada *stop* consecutivo.

Las densidades de corriente son directamente proporcionales a las irradiancias y se calculan usando la fórmula:

$$\frac{I_{rr0}}{I_{rr1}} = \frac{J_{cc0}}{J_{cc1}} \rightarrow J_{cc1} = J_{cc0} \frac{I_{rr1}}{I_{rr0}} \quad (2.14)$$

La fotocorriente generada por el fotodiodo se calcula usando la relación:

$$I_{ph} = J_{cc1} A_{diodo} \quad (2.15)$$

Tabla 2.4: Estimación de la fotocorriente en el plano focal de la imagen.

R	0.18
Numero f	1.4
Área de diodo (cm ²)	1.0E-6
Jcco (A/cm ²)	0.0366

Estimando la fotocorriente generada por un fotodiodo de un pixel, fabricado en la región de la fuente de un transistor MOS con pozo n, en función de la irradiancia de la escena. El rango estimado va de 1E-15 a 1E-9 A, para condiciones de iluminación de una noche con luna llena a una iluminación a pleno sol. Estos valores pueden ser usados para valores de fuentes de corriente en paralelo con el diodo para propósitos de simulación de fotorreceptores basados en eventos.

Tabla 2.5: Estimación de la fotocorriente en diferentes condiciones de iluminación.

Condición	I_{rre} (W/m ²)	I_{rrch} (W/m ²)	Jcce(A/cm2)	Jccch(A/cm2)	Iph pixel
Pleno sol	1000	11.480	3.66E-2	4.2E-4	4.2E-10
Luz interior	100	1.148	3.66E-3	4.2E-5	4.2E-11
Luz interior tenue	10	0.115	3.66E-4	4.2E-6	4.2E-12
Alumbrado Publico	1	0.011	3.66E-5	4.2E-7	4.2E-13
Luna llena	0.1	1.15E-4	3.66E-7	4.2E-9	4.2E-15

2.6 Algoritmo para la emulación de la excitación luminosa a la entrada del fotorreceptor.

Para calcular la forma de onda de entrada de un fotorreceptor neuromórfico de un arreglo de pixeles unidimensional, formado por n pixeles de dimensión igual a la separación entre centros de fotodiodos (*pitch*), una longitud D de los fotodiodos, una barra negra de longitud igual a un número entero de la dimensión del *pitch*, que se mueve de izquierda a derecha a una velocidad V_b determinada, en la dirección del arreglo de pixeles, en un fondo blanco que genera una corriente I_{ph} en el fotodiodo, como variables independientes

En la Fig. 2.12 se muestra un diagrama físico (*layout*) del arreglo de pixeles y la barra negra.

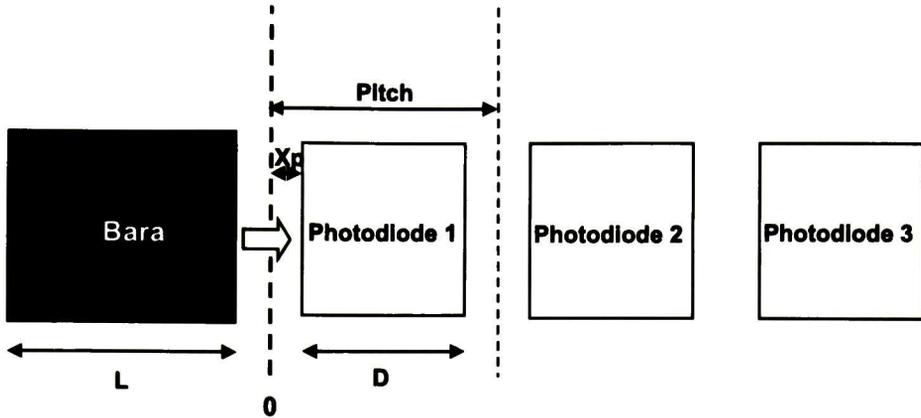


Fig. 2.12: Diagrama de los fotorreceptores y una barra negra

Por conveniencia se definen las siguientes longitudes en la **Tabla 2.6** con dimensiones en μm . Por razones de simplificación considérese que la longitud de la barra es igual a la longitud del *pitch* del arreglo (longitud de un pixel).

Tabla 2.6: Definición de longitudes importantes.

$$\begin{aligned}
 X_0 &= 0 \\
 X_1 &= X_p = \frac{(P - D)}{2} \\
 X_2 &= D \\
 X_3 &= (L - P) + 2X_1 \\
 X_4 &= X_2 \\
 X_5 &= L \\
 L &= n \cdot P
 \end{aligned}$$

X_0 representa el origen, $X_0 = 0$, $X_1 = X_p$ define la diferencia entre el pitch del arreglo y la longitud del fotodiodo, $X_2 = D$ la longitud del fotodiodo, $X_3 = 2X_p$ es la longitud que separa dos fotodiodos consecutivos, $X_4 = X_2 = D$, la longitud del fotodiodo y $X_5 = L$ es la longitud de la barra.

La imagen de la barra posee una velocidad V_b a lo largo del arreglo. Cuando el extremo derecho de la barra está en la posición $X_0 = 0$, empieza el conteo del tiempo. El primer pixel estará iluminado y genera una fotocorriente de $1\text{E-}10$ A (una estimación para una escena de fondo blanco). Esto se mantendrá hasta que la barra alcance el extremo izquierdo del fotodiodo. El tiempo necesario para que esto suceda será inversamente proporcional a la velocidad de la barra y directamente a la longitud X_1 . A partir de este tiempo el fotodiodo se oscurece gradualmente por el paso de

la barra, hasta estar completamente cubierto cuando la barra se desplace una distancia $X_2 = D$. En esta situación la corriente fotogenerada se ha estimado en $1E-13$ A. Esta situación se mantiene hasta que la barra haya avanzado una distancia igual a $X_3=2X_p$, para una barra de longitud igual a la distancia de un pixel= P . En este punto el extremo izquierdo de la barra empieza a descubrir el fotodiodo, y se inicia una transición creciente del nivel de obscuridad al de plena iluminación. La transición se completa cuando la barra avanza una distancia $X_4 = D$. En el caso de una sucesión de barra oscuras de longitud L , separadas entre ellas por una distancia igual a su longitud, la forma de onda se repetirá periódicamente en el primer pixel. Esto será útil para estudiar las propiedades dinámicas de un fotorreceptor determinado. Para el caso de un arreglo de n pixeles, la iluminación se mantendrá alta para el primer pixel indefinidamente después de este último periodo de transición. Para el segundo pixel la forma de onda se repetirá a partir de un retraso de tiempo necesario para que la barra viaje una distancia igual a un pixel, y así sucesivamente para el resto de los pixeles.

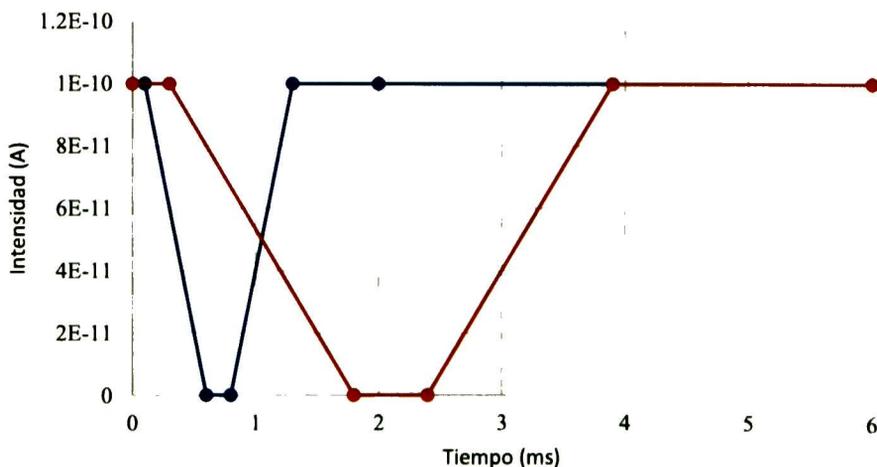


Fig. 2.13: Perfil de intensidad de una barra oscura pasando sobre un fondo blanco.

Con la finalidad de determinar los parámetros de una fuente de corriente independiente, del tipo de segmentos lineales (*Piece-wise Linear*, PWL) para su simulación en *Spice* o su equivalente en la representación de Pulsos, es necesario mapear en el tiempo, las sumas acumuladas de las distancias críticas. Llamando S a las sumas acumuladas, el algoritmo utilizado fue el siguiente:

$$S_{i+1} = S_i + X_i \quad (2.16)$$

La ecuación de mapeo en el tiempo en:

$$V_b = \frac{S}{t} \rightarrow t = \frac{S}{V_b} \quad (2.17)$$

La Fig. 2.13 muestra la forma de onda resultante.

2.7 Comparación del chip neuromórfico versus las funcionalidades de la visión humana.

La percepción es por mucho el proceso sensorial más importante, por el que obtenemos información del ambiente. La visión empieza con la luz reflejada en la superficie de los objetos, luego la luz es transformada en energía eléctrica-química por medio de los fotorreceptores. Este proceso es altamente complejo, la intrincada estructura de la retina que consiste en un arreglo de células organizadas en capas de neuronas especializadas, interconectadas entre sí, así como, la biología de la visión se describen de una manera comprensible en la referencia [26].

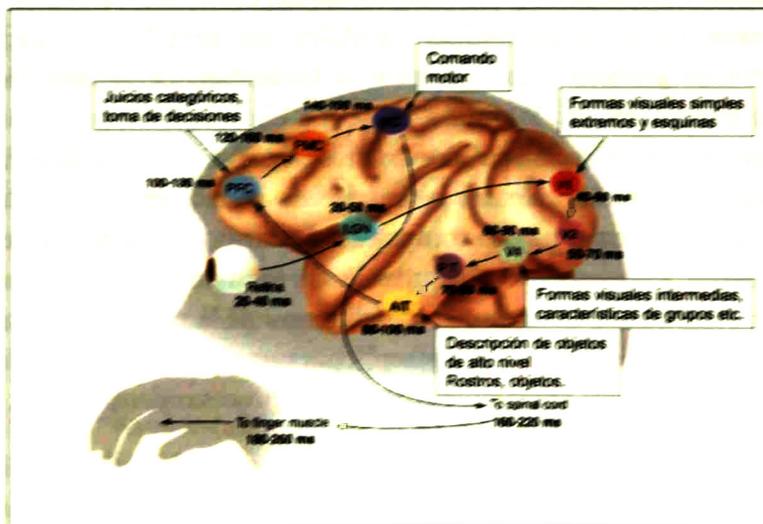


Fig. 2.14: Procesamiento en el sistema visual.

Un diagrama esquemático sobre el flujo del procesamiento de la información visual, se muestra en la Fig. 2.14. Un diagrama más reducido

de las primeras etapas de visión se muestra en la **Fig. 2.15**, que corresponderían a las funciones emuladas por el chip de visión propuesto.

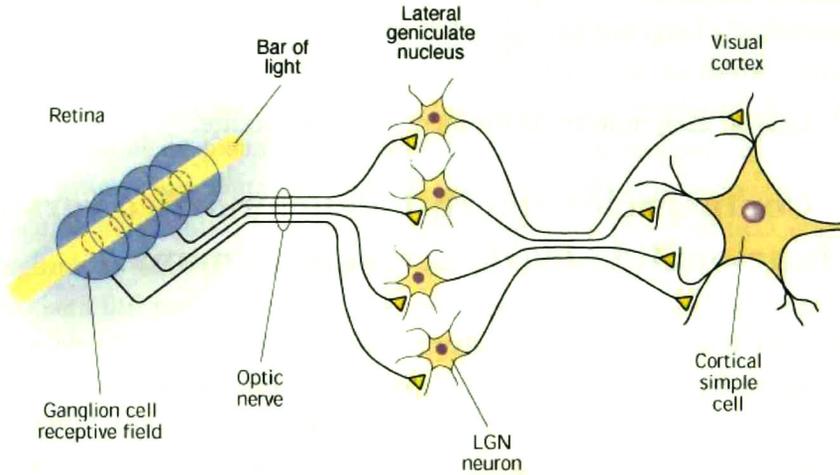


Fig. 2.15: Sistema de visión simplificado.

Las características visuales de la escena, que sirven como datos de entrada a la retina, son múltiples: tamaño, posición, orientación, movimiento, iluminación, sombras, contraste, color, etc. La retina humana preprocesa estas características dándoles un sentido, un significado, mediante un proceso consistente en: la traducción de señales visuales a señales neuronales, de fotones a impulsos electroquímicos; la extracción eficiente de la información en los extremos o bordes de los objetos, mediante un proceso que se conoce como la creación de campos receptores ("receptor fields") que definen zonas de contraste del centro y sus alrededores, y que incrementan grandemente el contraste de los bordes y esquinas de los objetos. Finalmente las neuronas ganglionares de la retina crean un código mediante impulsos eléctricos ("spikes"). para su comunicación eficiente con otras neuronas en las áreas subsiguientes del procesamiento visual.

Los axones de las neuronas ganglionares de la retina conducen la información codificada, nervio óptico, a otro conjunto de neuronas que se encuentran en una área del cerebro conocida como Núcleo lateral Geniculado, donde se forma un mapa topográfico de eventos en dos dimensiones que representa las posiciones de los campos receptivos y otras características relevantes de la escena visual, como el color, tal como se originan en la retina.

La información representada en el Núcleo Lateral Geniculado es retransmitida a la Corteza Visual Primaria, V1, donde es procesada en formas visuales simples: contornos y esquinas. Esto prosigue en formas cada vez más complejas del procesamiento de la información.

Estas funciones “simples” de la retina biológica y el núcleo lateral Geniculado son emuladas por los bloques neuromórficos del fotorreceptor adaptativo, el amplificador diferencial de transconductancia y el bloque de detección de polaridad y su selección, con la excepción del proceso de generación de los campos receptivos, la Fig. 2.16 ilustra esta comparación.

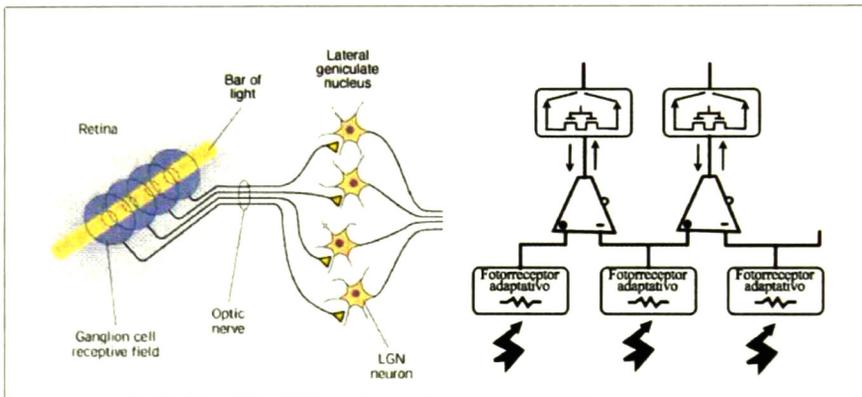


Fig. 2.16: Emulación de las funciones de la retina y el núcleo lateral Geniculado por los bloques de circuitos neuromórficos equivalentes.

Mientras que esto último parecería ser una limitación en la emulación, el autor del diseño del chip neuromórfico [5] pospuso esta función para implementarla en el bloque “winner take-all”, mediante una red resistiva de retroalimentación positiva y negativa, tomando en consideración el ruido inherente generado por las no idealidades del proceso de fabricación del chip (“mismatch”) y el de las operaciones de derivación.

La red “Winner-take-all” tiene como entrada el mapa de relevancia producido por los bloques anteriores, selecciona la salida más relevante, a costa de las demás, implementa una operación de convolución (determina su campo de recepción) y sigue esta salida relevante. El bloque de determinación de la posición transforma esta señal en un comando para la acción de cambio de dirección del vehículo en el cual está montado el chip de visión (ver Fig. 2.17).

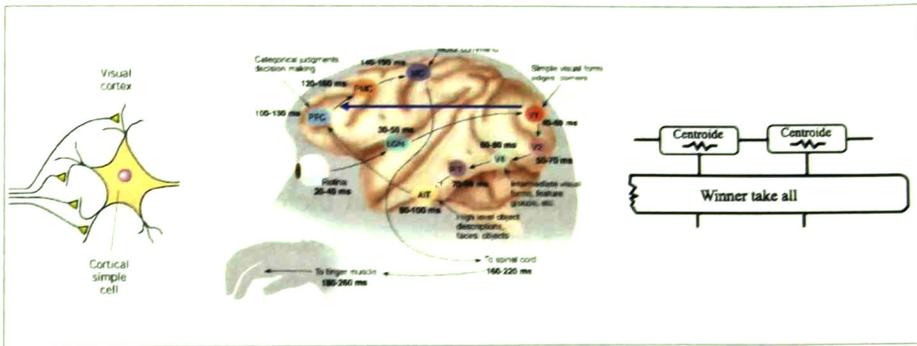


Fig. 2.17: Emulación del proceso de detección de bordes, seguimiento, decisión y acción por la red WTA y el circuito de posición.

El circuito neuromórfico es una síntesis para la detección del movimiento. Este procesamiento no es una emulación en cómo se haría en un sistema biológico, debido a que las neuronas biológicas se comunican con pulsos de corriente, para investigar estos pulsos se necesitan redes que emulen estos pulsos. Sin embargo la tarea de la atención visual es muy bien realizada, gracias a la simplificación de un algoritmo para la detección del movimiento y al circuito WTA que elegantemente logra capturar muchas de las características vistas en las redes de competencia biológicas.

Algunas limitaciones del sensor comparadas con la biología, serían que el circuito no realiza una tarea completa de atención basada en Saliency Maps, esto debido a que el circuito es un arreglo unidimensional y en un saliency map se usan arreglos bidimensionales. El sistema fotorreceptor aún es sensible ante defectos de fabricación, los errores en el proceso de fabricación causan cambios notables en el funcionamiento del circuito.

2.8 Conclusiones.

Un chip de visión inspirado por la naturaleza, diseñado para un propósito específico, (seguir una línea), es una aplicación que parece simple, sin embargo difícil de lograr. Es importante antes de aventurarse a resolver este problema, es entenderlo, comprender qué es el movimiento, qué eventos genera, cómo detectarlo, cómo lo hacen los organismos vivos.

En este capítulo se mostró cómo obtener de una imagen espacial al filtrar información de ésta, un problema que existe al aplicarlo en un sistema de seguimientos es la orientación de éste, esta orientación se puede elegir al

calcular cuál es la dirección en la que la operación para detectar los elementos más sobresalientes sea más significativa.

Otro problema relacionado con la simulación es que los programas usados para simular circuitos es que generalmente no tienen información de fotosensores, esta información se estima con información que se tiene sobre éstos, y se puede simular al aplicar una fuente de corriente en paralelo con un diodo. Estos son problemas operacionales inherentes al circuito, aparte de estos el chip es un circuito neuromórfico modular, que emula operaciones que se encuentran en la vía de la visión encontrada en seres vivos, en los próximos capítulos se describen estos módulos.

Capítulo 3

Fotorreceptor

Este capítulo se enfoca en la descripción de un fotorreceptor logarítmico adaptativo basado en un modelo de pequeña señal y los resultados de su simulación del circuito en SPICE. La simulación de este circuito se realiza al emplear una fuente de corriente en paralelo con un fotodiodo para emular la corriente generada por la absorción de fotones en el dispositivo, para generar un rango de fotocorrientes con un sentido físico se debe tener una referencia conocida de esta fotocorriente a un nivel de irradiancia determinada, para esto escogemos la densidad en corto circuito de una celda solar moderna de alta eficiencia.

3.1 Principios y análisis de fotorreceptores neuromórficos.

Una variedad de tareas que necesita el procesamiento de imágenes, pueden simplificarse considerablemente en escenas dinámicas, si la información básica del movimiento está preprocesada previamente. En tareas de movimiento visual, en donde se utiliza

información extraída para guiar el movimiento del sistema, el procesamiento se debe de hacer dentro del tiempo en que la escena no cambia de manera significativa, en tiempo real [10]. El procesamiento de imágenes en tiempo real usando electrónica digital típicamente requiere un ancho de banda bastante alto. Mucha de la carga computacional cae en el preprocesamiento de la imagen, antes de que algún algoritmo de visión global sea aplicado. Maquinas especializadas en procesadores paralelos digitales como la maquina PIPE se han usado para hacer procesamiento de imágenes en tiempo real [27].

Mucho del trabajo previo en retinas de silicio se ha orientado en los aspectos espacio-temporales del preprocesamiento de imagen, en estos modelos el tiempo es tratado como una variable independiente.

Los sistemas de visión móviles a menudo representan severas restricciones en tamaño, peso y consumo de energía. Sin embargo el grado de integración es limitado por el gran tamaño y la disipación de energía de una arquitectura PIPE. Un sistema analógico bien diseñado es mucho más económico en área y en consumo de energía que los sistemas tradicionales digitales con un procesamiento comparable [10].

Durante mucho tiempo se han estado construyendo varios chips de visión neuromórficos que hacen cálculos en el dominio del tiempo además de hacer procesos espacio temporales continuos en el tiempo, que toman lugar antes de algún tipo de muestreo o de comunicaciones de larga distancia [8]. Como por ejemplo procesamiento de movimiento, detección de cambios, procesamiento en retinas neuromórficas, y síntesis de imágenes con información auditiva para escenas visuales.

Estos procesos requieren circuitos fotorreceptores que traduzcan la luz que incide en el chip en una señal eléctrica. Los fotorreceptores producen una salida analógica continua que puede ser directamente conectada a circuitos analógicos adyacentes, como por ejemplo circuitos que procesan el movimiento de la imagen; ésta es una característica que contrasta con la vasta mayoría de dispositivos usados comercialmente como por ejemplo CCD, su disponibilidad y operación confiable han propiciado su uso en aplicaciones de visión, sin embargo su uso ha impedido la investigación de algoritmos y arquitecturas de visión que usen el tiempo de una manera natural y eficiente, esto debido a que es difícil acoplar información de tiempo

en un flujo serial de las salidas de los circuitos analógicos de imagen. Si queremos hacer procesos de visión analógico en el dominio del tiempo es sensato construir circuitos fotorreceptores analógicos continuos en el tiempo y acoplar sus salidas localmente a los circuitos que hacen la operación.

En este capítulo presentamos un circuito foto-receptor originalmente diseñado por Tobi Delbrück [5] y mejorado por Shih-Chii Liu [27]. El objetivo principal del fotorreceptor es procesar el contraste independiente de la iluminación absoluta y adaptar su respuesta en el tiempo en función de las condiciones de iluminación.

3.1.1 Receptores logarítmicos simples.

Un receptor con una respuesta logarítmica a la iluminación tiene el tipo correcto de respuesta porque el cambio de intensidad está dado por:

$$d \log I = \frac{dI}{I} = \frac{i}{I_{bg}} \quad (3.1)$$

Un receptor simple logarítmico puede fabricarse con una unión p-n, éste sería una unión simple en el sustrato ligeramente dopado y una región altamente dopada (difusión de contacto del drenaje), cuando la luz incide en el silicio se generan pares electrón-hueco.

Cuando los electrones son generados en el sustrato p-, son barridos por el campo eléctrico hacia la región n++. Lo mismo pasa con los huecos creados en la región n++. El resultado de esto es una fotocorriente circulando de n++ a p-, por esto la región n++ es negativamente cargada con respecto al sustrato. Este voltaje negativo produce una corriente negativa en la unión que compensa la fotocorriente. Debido a que la fotocorriente es exponencial en el voltaje de la unión, el voltaje en la región n++ es logarítmica en la intensidad.

Un receptor logarítmico que produce una salida útil consiste en un simple transistor MOS, en donde la fuente del transistor forma el fotodiodo mostrado en la **Fig. 3.1** El canal del transistor MOS forma la barrera que resulta en una respuesta logarítmica a la intensidad. El voltaje en la fuente decrece mientras la intensidad se incrementa. Este circuito sería un muy buen fotorreceptor logarítmico en un mundo ideal, la ganancia de 60

mV/década resulta en un rango típico de salida de voltajes alrededor de 20mV en escenas naturales, lo que afectaría a los demás circuitos con variaciones de corriente en factores de dos lo que es suficiente para el procesamiento.

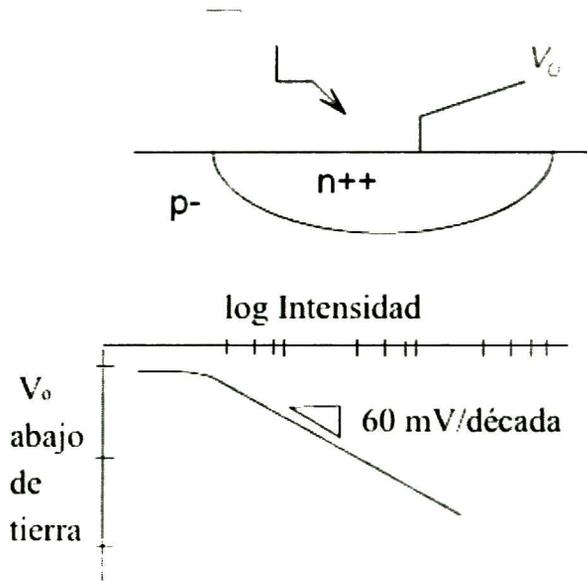


Fig. 3.1: Una unión simple, el fotorreceptor logarítmico más simple. V_o se mantiene debajo del voltaje de sustrato, decreciendo logarítmicamente con la intensidad.

Hay dos problemas importantes, el *mismatch* y la respuesta lenta. Las diferencias entre las supuestas salidas idénticas del fotorreceptor son tan grandes como las variaciones típicas de la señal producidas en escenas reales, haciendo que un sistema de fotorreceptores no sea utilizable más que para condiciones demostrativas. El otro problema se debe a que ante iluminaciones bajas la respuesta es lenta. Esto es debido a la capacitancia por unidad de área de la unión, hacer el fotodiodo más grande no ayuda porque la capacitancia se escala de manera lineal con el área. Y empeora cuando el tamaño se reduce, porque la densidad del sustrato se reduce y la capacitancia de la unión será incrementada.

Por esto la necesidad de la adaptación para tratar el *mismatch* del circuito y una retroalimentación activa para tratar el problema de la respuesta lenta.

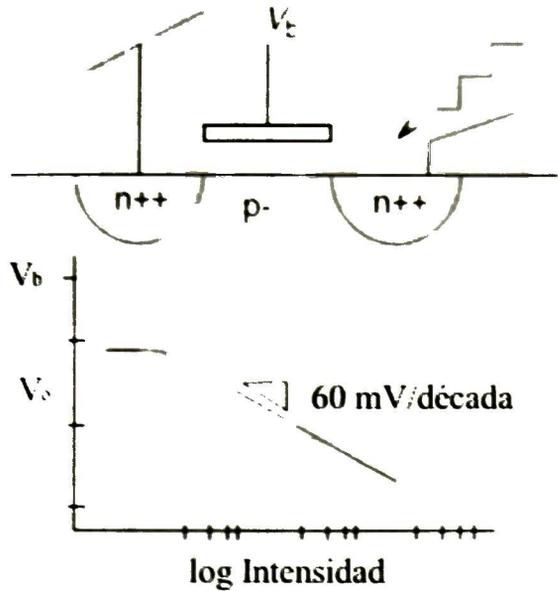


Fig. 3.2: Un fotorreceptor logarítmico en seguidor de fuente. V_0 decrece logarítmicamente con la intensidad de la luz empezando con una intensidad de cero en aproximadamente V_b .

3.1.2 Fotorreceptor logarítmico adaptativo.

En el circuito que Liu propone [27] el elemento de adaptación es un resistor no lineal formado por un transistor tipo P y la inclusión de dos transistores. Los transistores incluidos permiten que el tiempo de adaptación sea configurado por una polarización exterior.

3.1.2.1 Descripción del circuito.

El circuito el cual tiene la forma básica del receptor de Delbrück, se ha reemplazado el elemento adaptativo de su circuito receptor por un resistor no lineal consistente de un transistor PMOS, Q_1 , la implementación de un resistor flotante controlado por voltaje ha sido descrito por Banu y Tsvividis, la polarización de Q_1 , V_b es generada por Q_3 y Q_4 la conductancia de Q_1 está determinada por la salida en voltaje de V_1 y la polarización externa V_m .

El circuito receptor adaptativo está formado por una retroalimentación negativa a la compuerta del transistor de retroalimentación en la configuración de seguidor de fuente. Conceptualmente el circuito usa un modelo interno para hacer una predicción sobre la señal de entrada al

apoyarse de la salida que viene de una comparación entre la entrada y la predicción, el lazo de retroalimentación es completado al usar aprendizaje para refinar el modelo y así las predicciones sean más acertadas.

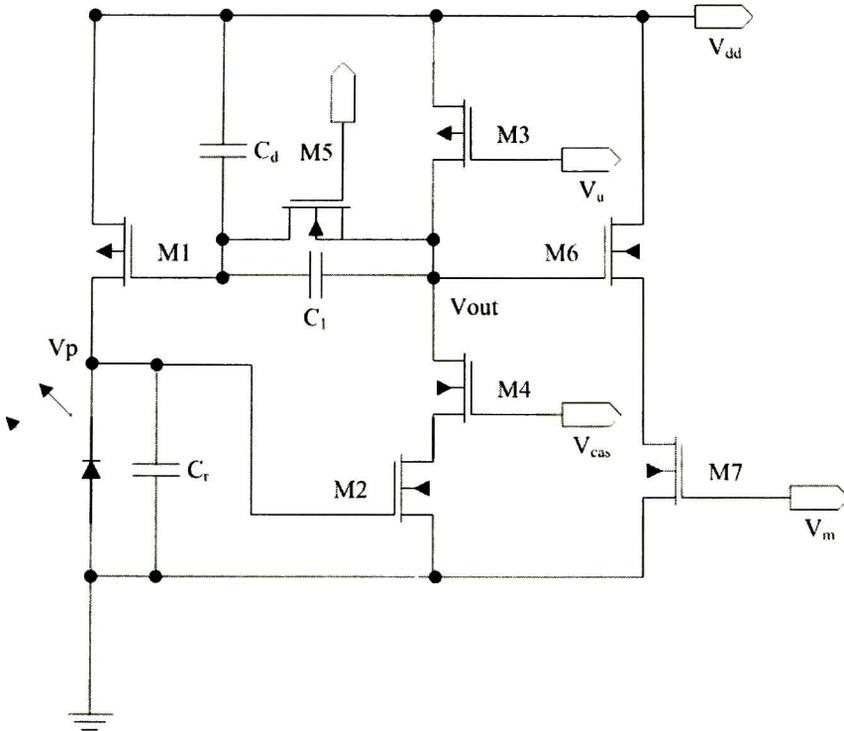


Fig. 3.3: El circuito propuesto por Shih-Chii Liu.

El circuito está formado únicamente por siete transistores y tres capacitores.

Todos los transistores operan en condiciones de débil inversión, donde la corriente del drenador varía con el voltaje de compuerta en forma exponencial. El fotorreceptor se adapta automáticamente a condiciones variables de la intensidad de iluminación en varias décadas.

En condiciones de CD, si la fotocorriente crece, la corriente del transistor M1, configurado como seguidor de fuente, debe crecer también. Esto es asegurado mediante las ramas de amplificación y retroalimentación de los transistores M2 y M5.

El resultado neto es que el voltaje en el fotodiodo se mantiene constante. La corriente de polarización provista por el transistor M3 es controlada mediante el voltaje de polarización V_u . En condiciones de CD el voltaje de salida V_{out} es igual al voltaje de compuerta del transistor M1 V_{gate} , por ser este último nodo, uno de alta impedancia. El circuito “aprende” comparando un modelo de corriente con la fotocorriente de excitación I_{bg} y se adapta a ésta por medio de la rama de retroalimentación.

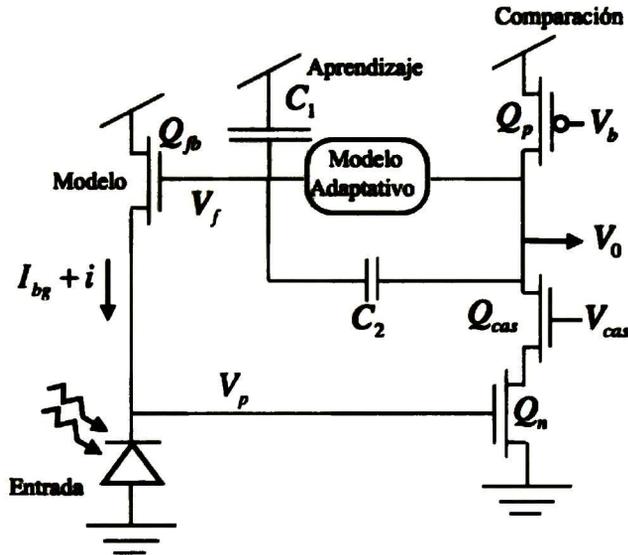


Fig. 3.4: Circuito propuesto por Delbrück.

La Fig. 3.5 muestra los voltajes de CD en distintos nodos en el circuito para $V_u = 2.0V$.

Los voltajes en el drenador y fuente del transistor amplificador M2 se mantienen constantes independientemente del nivel de iluminación del fondo de la escena, I_{bg} , lo que implica que la conductancia y transconductancia del transistor no dependerá del nivel de iluminación.

La etapa de entrada y el modelo de predicción.

Consiste en una fotorreceptor en seguidor de voltaje, el transistor de retroalimentación Q_{fb} opera en el subumbral así el voltaje V_p es logarítmico en la fotocorriente, V_p está bajo V_f a cualquier voltaje que se necesite para

encender Q_b para generar la fotocorriente. Conceptualmente el voltaje en C_i actúa como modelo para la intensidad de entrada.

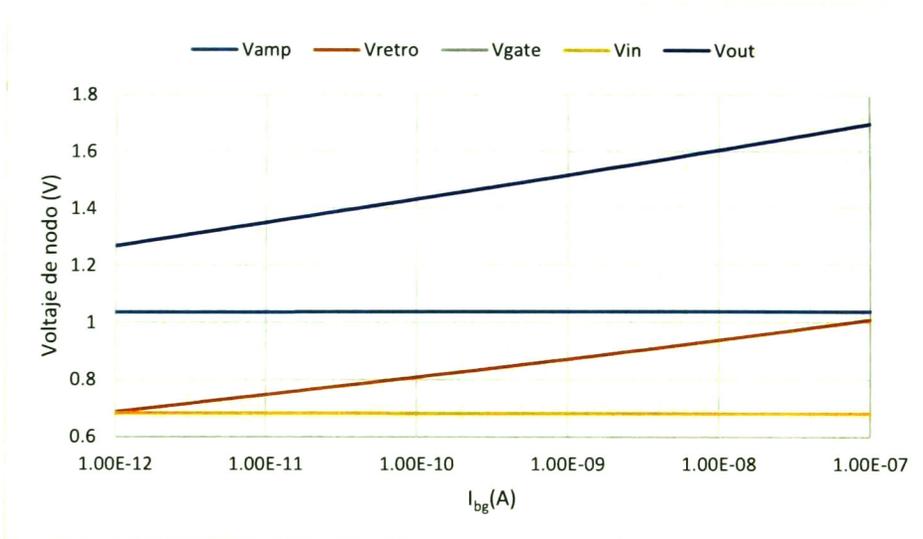


Fig. 3.5: Características en DC.

3.1.2.2 La etapa de comparación.

La comparación entre la entrada y el modelo está realizada por un amplificador inversor que consiste en Q_n y Q_p en conjunto a un transistor adicional Q_{cas} . El voltaje de entrada V_p controla la corriente que drena Q_n , la corriente suministrada por Q_p está fija a valores de polarización de V_b . El voltaje de la ganancia de este amplificador está determinado por la relación entre la transconductancia de Q_n , y la conductancia de la salida del amplificador está determinado por el tamaño de los transistores en el amplificador. El voltaje V_b determina la frecuencia de corte del receptor. El lazo de retroalimentación está completo cuando la salida V_o está conectada de vuelta a V_f por medio del elemento adaptativo y sobre el divisor de capacitancias formado por C_1 y C_2 .

3.1.2.3 El lazo de retro alimentación.

Un incremento pequeño en i intenta bajar el voltaje V_p en una razón de $(i/I_{bg})U_T$. En respuesta V_o sube a una razón de A_{amp} la salida está acoplada a la compuerta de Q_b por medio de un divisor capacitivo con una ganancia de quizá 0.1. Subiendo la compuerta de Q_b subimos la fuente de Q_b . El amplificador de retroalimentación y la señal de entrada compiten para

controlar la fuente de voltaje de Q_b pero el amplificador gana gracias a su alta ganancia. El voltaje V_p se mueve lo suficiente para que el voltaje V_o se mueva lo suficiente para que V_f se mueva lo suficiente y así V_p se mantenga casi fijo. Mientras más grande sea C_1 en comparación con C_2 mayor será la ganancia.

Transistor de *Cascode* Q_{cas} .

Tiene dos efectos: protege el drenaje de Q_n de los grandes cambios de voltaje de V_o porque la conductancia de fuente de Q_{cas} es mayor que la conductancia de drenador por un factor aproximado de A_{amp} , el drenaje de Q_n se mueve tanto como V_p . Sin Q_{cas} los grandes voltajes que pasan sobre la capacitancia compuerta-drenaje de Q_n reducen la carga en el nodo de entrada. Hacen que la capacitancia compuerta drenaje, de fF, aparezca en el nodo de entrada como si estuviera en la escala de pF, un fenómeno llamado el efecto Miller.

Q_{cas} también multiplica la resistencia de drenaje de Q_n por un factor aproximado de A_{amp} , esto incrementa la ganancia del amplificador por un factor de 2.

Ambos, el incremento de la capacitancia de entrada y el incremento de la ganancia, se traducen como un incremento en la velocidad. Este incremento hace que sea operable en situaciones de bajas intensidades de luz. Incrementando su rango dinámico por una década.

3.2 Propiedades del circuito.

Las respuestas temporales y las propiedades de adaptación de este circuito son expuestas en las siguientes secciones:

3.2.1.1 Función de transferencia.

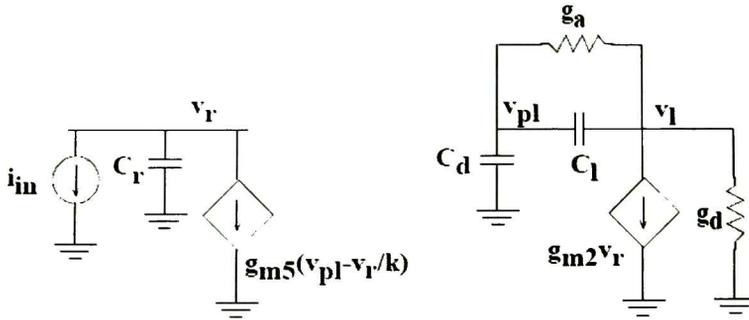


Fig. 3.6: Modelo de señal pequeña para el circuito de la Fig. 3.3.

Podemos resolver la función de transferencia del circuito al escribir las ecuaciones KCL del modelo de pequeña señal mostrado en la **Fig. 3.6**, la función de transferencia está dada por:

$$\frac{V_0}{i_{in}} = \frac{1}{g_{m5}} \left[\frac{\frac{s(\tau_{ld} + \tau_l) + g_a/g_{m2}}{s\tau_l + g_a/g_{m2}}}{(s\tau_r + 1/\kappa) \left(1/A_{amp} + s\tau_{ld} + \frac{s\tau_{ld}}{A_{amp}(s\tau_{ld} + g_a/g_{m2})} \right)} \right] \quad (3.2)$$

donde g_a es la conductancia de salida de Q_1 y C_r es la capacitancia parasita en el nodo V_r , las constantes de tiempo τ_d , τ_l , τ_r , se definen como:

$$\tau_l = \frac{C_l}{g_{m2}} \quad ; \quad \tau_r = \frac{C_d}{g_{m2}} \quad ; \quad \tau_{ld} = \frac{C_r}{g_{m2}} \quad (3.3)$$

Y la ganancia del amplificador está dada como:

$$A_{amp} = \frac{g_{m2}}{g_d} \quad (3.4)$$

en donde g_m es la transconductancia, y g_d es la conductancia de salida de un transistor.

Las unidades de la función de transferencia son Ohms. La función puede expresarse como una función sin unidades, para su expresión en decibels, normalizando el voltaje de salida por unidad de voltaje térmico, U_T , y la señal de entrada de corriente por unidad de irradiancia de la escena I_{bg} . De esta manera la Ecuación (3.2) se expresa de la siguiente manera.

$$\frac{V_{in} I_{bg}}{i_{in} U_T} = \frac{I_{bg}}{U_T g_{m5}} \left[\frac{s(\tau_{ld} + \tau_l) + a}{(s\tau_l + a)(s\tau_r + b) \left(s\tau_{ld} + d + \frac{sd\tau_{ld}}{(s\tau_l + a)} \right) + 1} \right] \quad (3.5)$$

donde a , b y d son expresiones que tienen la forma siguiente:

$$\begin{aligned} a &= g_a / g_{m2} \\ b &= 1/\kappa \\ d &= 1/A_{amp} \end{aligned} \quad (3.6)$$

La función de transferencia normalizada puede expresarse en forma estándar como sigue:

$$\frac{V_{in} I_{bg}}{i_{in} U_T} = B \left[\frac{e \left(\frac{s}{z_1} + 1 \right)}{\left(\frac{s}{p_2} + 1 \right) \left(d \left(\frac{s}{p_3} + 1 \right) \left(\frac{s}{p_1} + 1 \right) + fs \right) + 1} \right] \quad (3.7)$$

donde:

$$\begin{aligned} e &= 1/b = \kappa \\ f &= d \tau_{ld} / a = \frac{g_d \tau_{ld}}{g_a} \\ B &= \frac{I_{bg}}{U_T g_{m1}} \end{aligned} \quad (3.8)$$

Los polos p_1 , p_2 , y p_3 son los polos de las expresiones de primer orden del denominador, y no son los polos de un polinomio de tercer orden del mismo. Se utilizan aquí para visualizar sus efectos en términos de los parámetros del circuito. Sus valores se expresan en términos de estos parámetros como sigue:

$$\begin{aligned} z_1 &= g_a / (C_1 + C_d) \\ p &= g_a / C_1 \\ p &= g_{m2} / C_r \\ p &= g_d / C_d \end{aligned} \quad (3.9)$$

Para calcular la función de transferencia normalizada se utilizan las condiciones y parámetros del circuito que se muestran en la **Tabla 3.1**. Las dimensiones de los transistores $M1$ y $M2$ fueron establecidas extrapolando los valores de diseño de un fotorreceptor logarítmico de contraste temporal de última generación en tecnología $0.3 \mu\text{m}$. Las dimensiones del transistor $M5$, fueron ajustados a los valores reportados.

Tabla 3.1: Dimensiones de transistores.

Transistor	W (μm)	L (μm)
M1	3.6	3.6
M2	3.6	2.2
M5	1.2	4.2

Los parámetros de conductancia y transconductancia fueron estimados de forma cuasi-estática a partir de las características I_{ds} vs V_{ds} o I_{ds} vs V_g de los transistores individuales en los puntos de polarización y de simulaciones a baja frecuencia. Las variables C_l , C_d y C_r fueron establecidas como valores típicos obtenidos de la literatura en el tema. Las polarizaciones externas se mantuvieron constantes: $V_{dd}=3\text{V}$, $V_u = 2.0\text{V}$, $V_{cas} = 2.0\text{V}$ y $V_m = 0.3\text{V}$. La iluminación de la escena fue de 10 pA, equivalente a una iluminación nocturna con luna llena.

Tabla 3.2 Valores numéricos de los parámetros.

g_{m2}	3.50E-07 S
g_{d2}	8.50E-08 S
g_d	3.00E-10 S
C_d	4.00E-12 F
C_l	1.00E-12 F
C_r	1.00E-12 F
g_a	1.00E-12 S
κ	7.00E-01
g_{m1}	2.80E-10 S
i_{bg}	1.00E-11 A
U_T	2.60E-02 V
A_{amp}	1.17E+03
t_l	2.86E-06
t_{ld}	1.14E-05
t_r	2.86E-06

3.2.1.2 Respuesta temporal.

Se calculó la respuesta del modelo usando la ecuación (3.9) por medio de Excel. La **Fig. 3.7** muestra una comparación entre los resultados del modelo usando parámetros constantes, independientes de la frecuencia y los resultados de la simulación en SPICE y los parámetros correspondientes a BSIM v1.3 de la tecnología 0.5 μm de ON Semiconductor, tomados de corridas reportadas por MOSIS. Debido a que usando parámetros constantes se muestran resultados pobres se hace un esfuerzo de expresar los resultados de simulaciones de los parámetros de interés con respecto a la

frecuencia, con modelos matemáticos de ajuste de curvas empleando la técnica de mínimos cuadrados.

Habiendo demostrado que el modelo de la simulación sigue al modelo de señales pequeñas seguimos analizando algunas de las propiedades del circuito empleando sus resultados de simulación.

La adaptación del fotorreceptor a bajas frecuencias, en condiciones de baja iluminación, es una condición necesaria para que el circuito pueda emular la respuesta de una retina biológica. El transistor PMOS M5, dentro de la rama de retroalimentación, es clave para garantizar esta función y su conductancia el parámetro más importante.

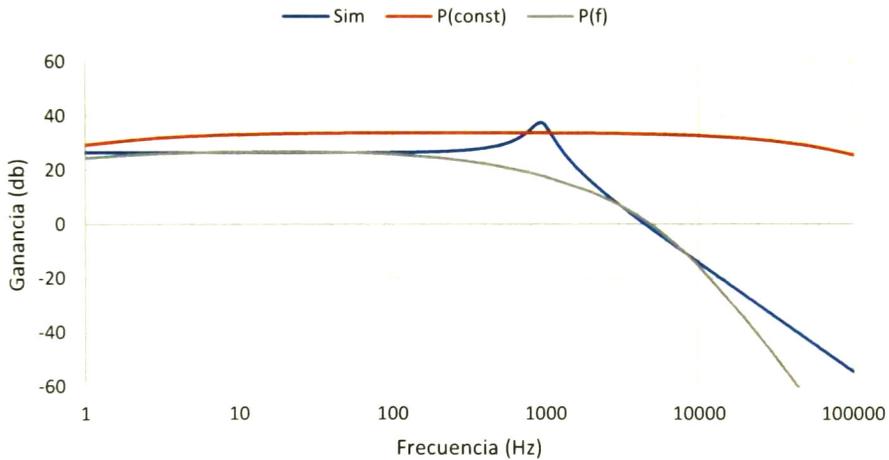


Fig. 3.7: Comparación entre los resultados del modelo usando parámetros constantes, dependientes de la frecuencia y su simulación en SPICE.

La conductancia del transistor de adaptación depende de las dimensiones de éste y más importante, para dimensiones fijas, dependerá automáticamente del nivel de iluminación y podrá controlarse mediante la polarización V_m .

La **Fig. 3.8** muestra la respuesta en frecuencia para bajas intensidades de iluminación en función de las dimensiones del ancho W , y largo L , del transistor en micrómetros. El transistor de menores dimensiones $W = 1.2 \mu\text{m}$, $L = 1.2 \mu\text{m}$ es el de menor conductancia.

Para dimensiones físicas del transistor fijas, la conductancia del transistor dependerá del voltaje de la compuerta del transistor de adaptación, V_{gs} , que a vez dependerá del nivel de iluminación, como se muestra en la Fig. 3.9. Un cambio de 100 mV en el voltaje de compuerta puede generar un cambio de corriente del orden de dos décadas de magnitud en el transistor, debido a su operación en la región de débil nivel de inversión.

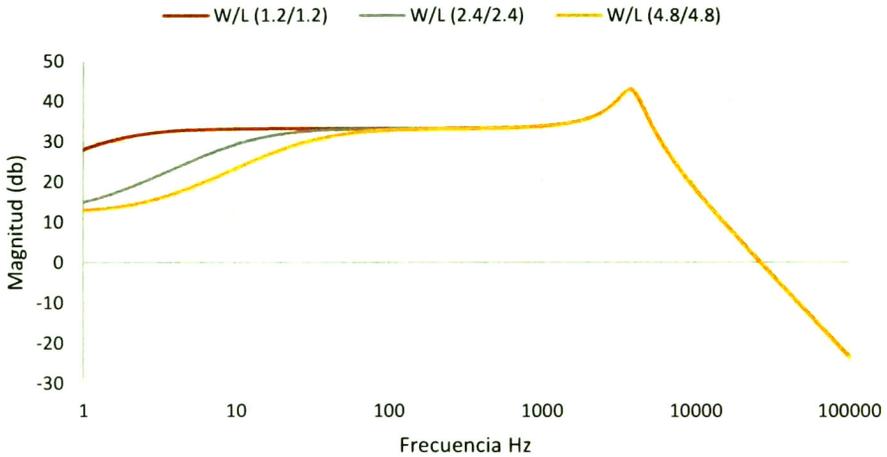


Fig. 3.8: Respuesta en frecuencia del fotorreceptor con bajas condiciones de iluminación, en función de las dimensiones del transistor M5.

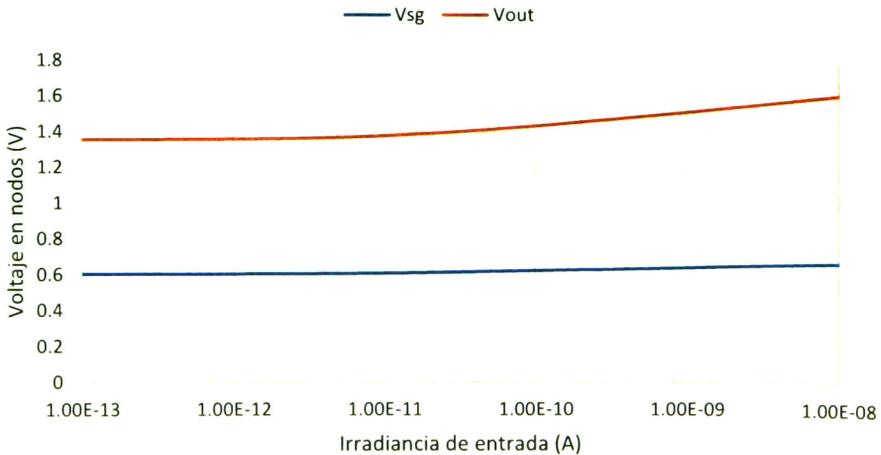


Fig. 3.9: Dependencia del voltaje de polarización de compuerta V_{sg} , del transistor M5 en función de la variación de cinco décadas en el nivel de iluminación.

La Fig. 3.10 es una gráfica de frecuencias del circuito RL sobre 5 décadas de intensidad en el fondo. Esta grafica muestra que en el rango de 10 a 1 kHz, el circuito es un filtro pasa banda en altos niveles de luz, y se reduce a un filtro pasa bajos en bajos niveles de luz.

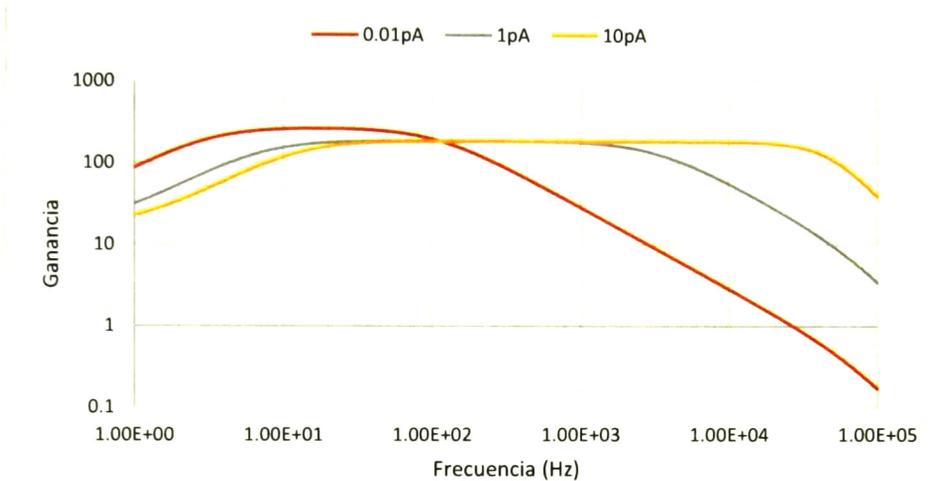


Fig. 3.10: Adaptación de las características de filtrado del fotorreceptor en respuesta a la variación de la intensidad de la iluminación.

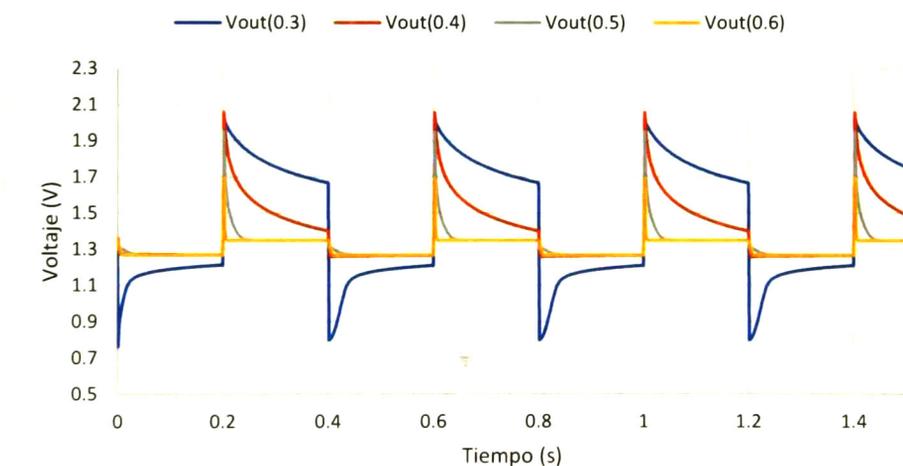


Fig. 3.11: Efectos de la polarización V_m en el decaimiento del pulso de salida.

Las respuestas a los transitorios pueden controlarse empleando la polarización V_m del transistor, como se ilustra en la Fig. 3.11, la razón de este comportamiento puede encontrarse en el incremento de la

conductancia del transistor de adaptación M5. La polarización de su compuerta aumenta a medida que la polarización V_m se incrementa, como se ilustra en la **Fig. 3.12**.

Luego el tiempo de respuesta del fotorreceptor puede controlarse externamente, hasta hacerlo tan pequeño que el circuito se vuelva insensible a los transitorios, función que resulta útil para inhibir la respuesta de una arreglo de fotodetectores para crear una ventana o región de atención.

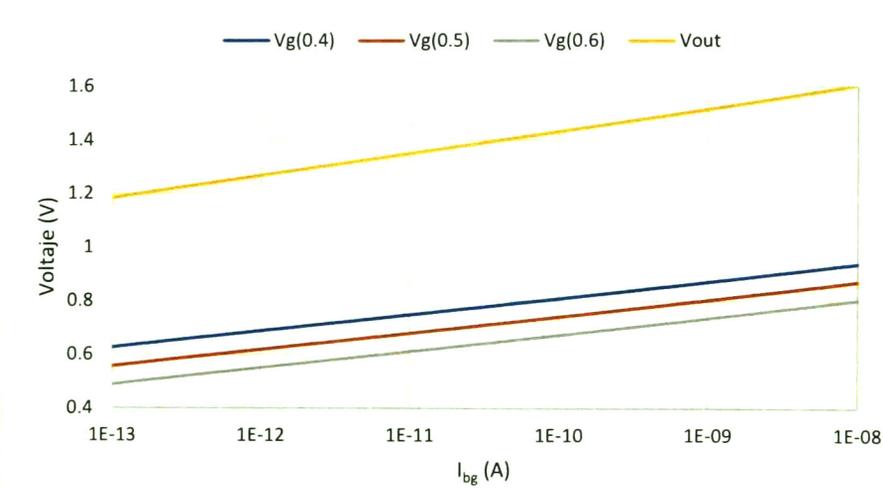


Fig. 3.12: Variación del voltaje de la fuente a la compuerta del transistor de adaptación como función del voltaje de polarización V_m . El voltaje de salida, V_{out} , depende únicamente de la intensidad de iluminación.

Vale la pena aquí hacer un comentario sobre el efecto de las capacitancias C_d y C_1 ; éstas forman parte de la rama de retroalimentación y determinan la relación entre los voltajes de salida V_{out} y el voltaje de compuerta del transistor M_1 , V_{fb} , por sus propiedades como divisor de voltaje capacitivo. La siguiente ecuación cuantifica este efecto:

$$\frac{V_{out}}{V_{fb}} = \frac{C_d + C_1}{C_1} \quad (3.10)$$

Esta relación no solo determina el coeficiente de retroalimentación a altas frecuencias sino que también determina la ganancia del filtro pasa bandas a intensidades elevadas, como lo muestra la **Fig. 3.13**.

A bajas frecuencias el efecto neto es decrecer la frecuencia de corte.

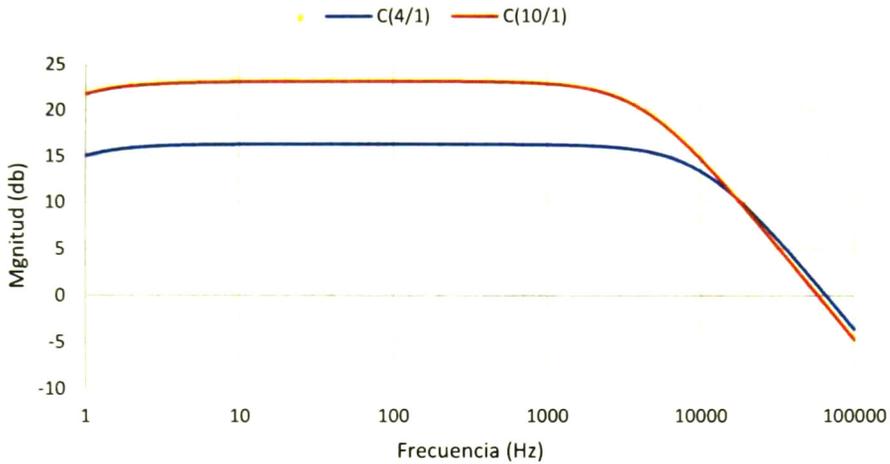


Fig. 3.13: Efecto de la relación de capacitancias C_d y C_i . Los valores de C_d fueron 4 pF y 10 pF respectivamente, C_i fue mantenido a un valor de 1 pF.

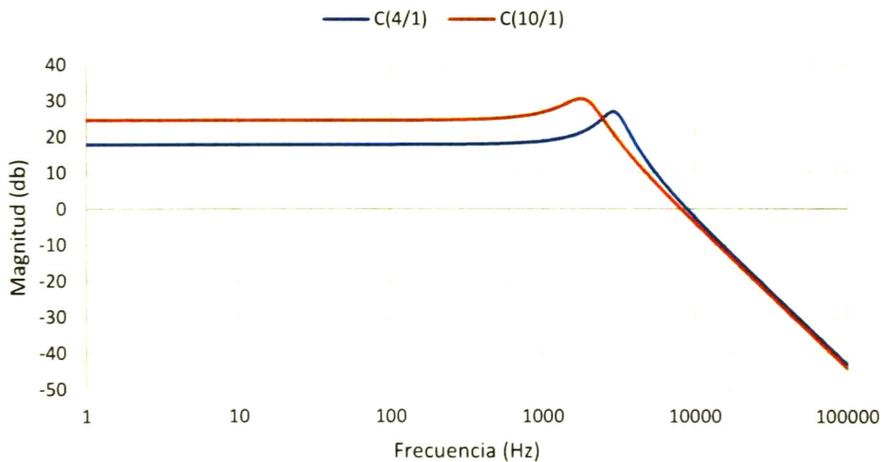


Fig. 3.14: Efecto de la relación de capacitancias C_d y C_i a bajas intensidades de iluminación.

3.3 Simulaciones en SPICE.

A continuación se presentan los resultados de la simulación de una barra oscura viajando a distintas velocidades sobre un fotorreceptor, además se

varía la iluminación de la escena en cada velocidad para revisar el comportamiento en diferentes ambientes visuales.

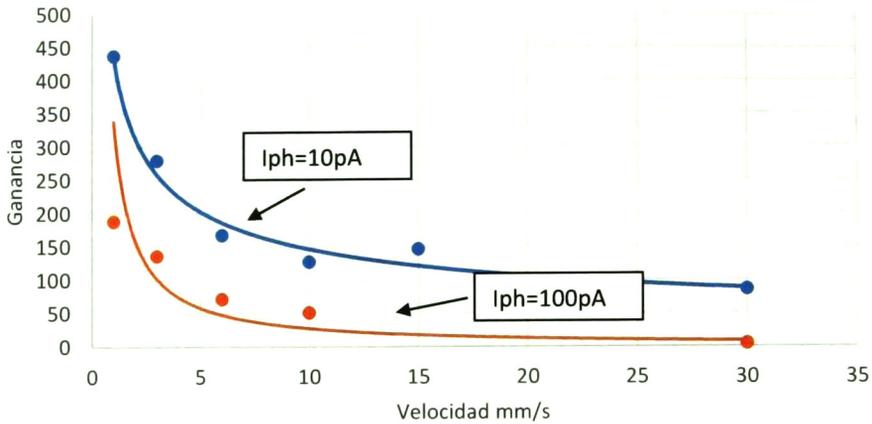


Fig. 3.15: Ganancia transitoria Vs velocidad como función de la iluminación de la escena.

El fotorreceptor fue simulado en Tanner SPICE S-Edit la **Fig. 3.16** muestra el esquemático del simulador, este bloque básico es el de cada fotorreceptor que se repite tantos pixeles hay en el arreglo.

La simulación de la barra nos muestra el comportamiento de la ganancia del circuito dependiendo de la velocidad y de la intensidad de la iluminación.

Las simulaciones se comparan con los resultados que obtuvo Indiveri [5] en su trabajo con los fotorreceptores neuromórficos como se muestra en **Fig. 3.17**, en esta figura se ilustra la respuesta de todos los bloques importantes del sistema de visión, ya explicado en el capítulo anterior, la excitación entrante es causada por una barra negra pasando en un fondo blanco.

Siguiendo con las simulaciones, a continuación se presenta la respuesta de un arreglo de 5px donde el esquemático de la simulación está presentado en la **Fig. 3.18**, aquí se considera al arreglo como un fotorreceptor grande por lo tanto la entrada de iluminación representa una barra negra entrando en el campo de visión del arreglo de pixeles está expresado en la **Fig. 3.19**, y en la siguiente se muestra cuando la barra va permitiendo pasar luz.

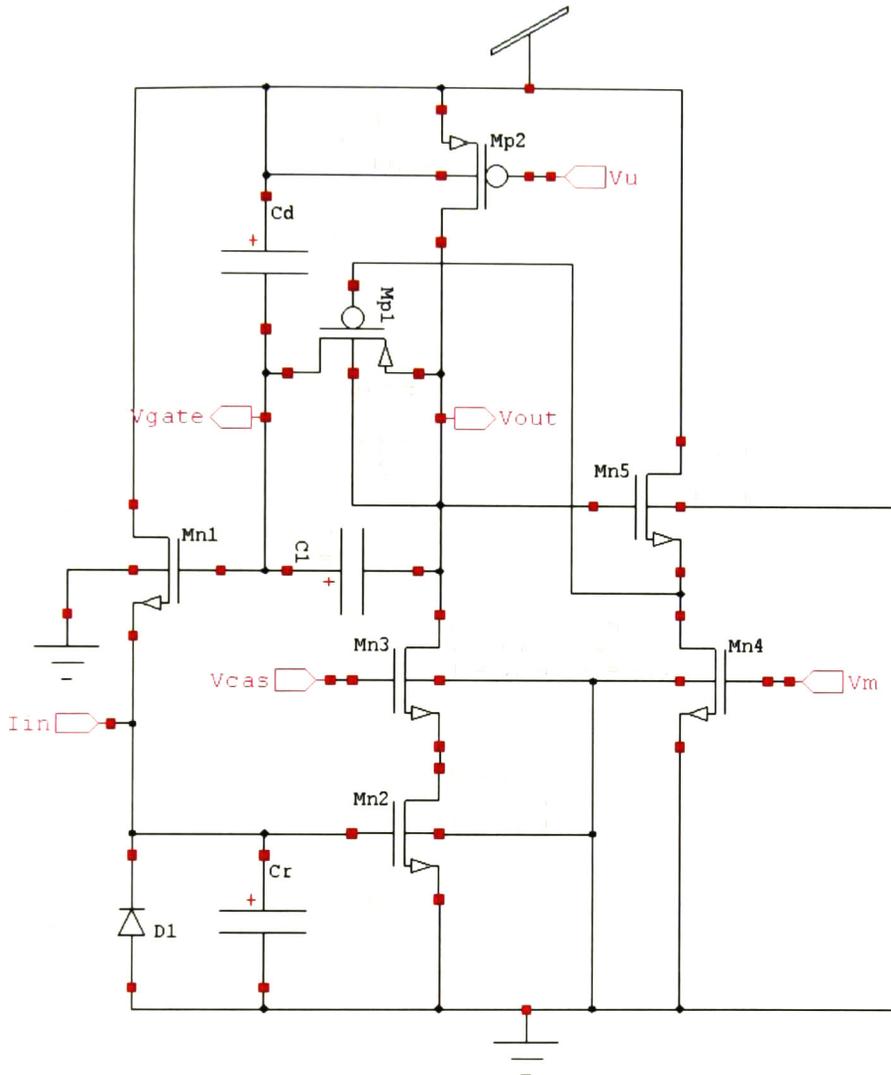


Fig. 3.16: Diagrama del fotorreceptor.

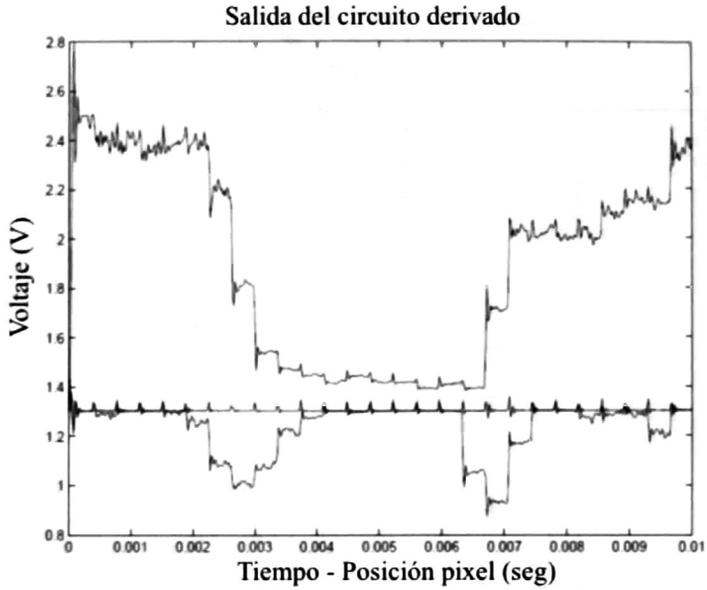


Fig. 3.17: Resultados a comparar.

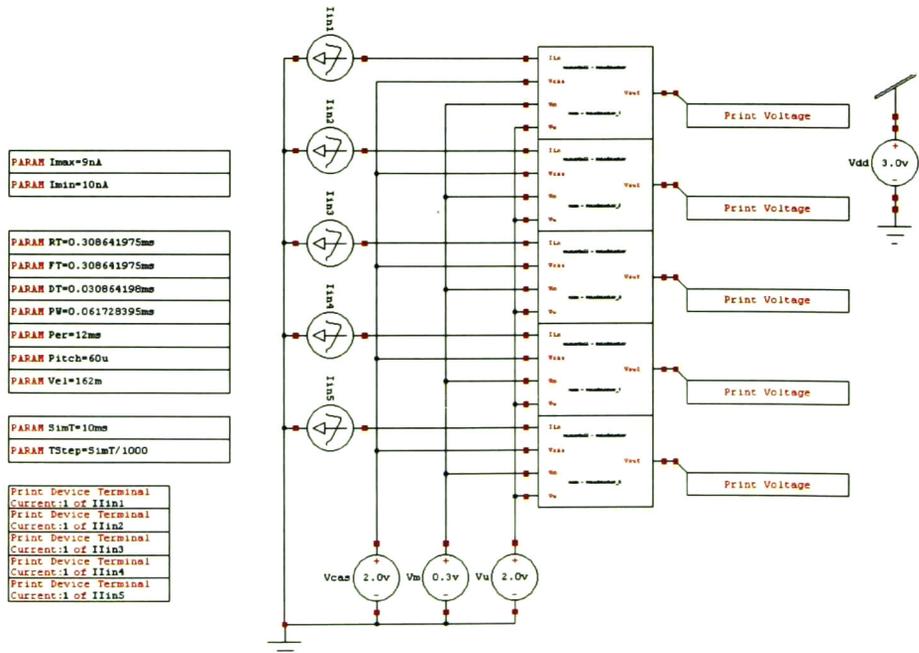


Fig. 3.18: Cinco bloques fotorreceptores

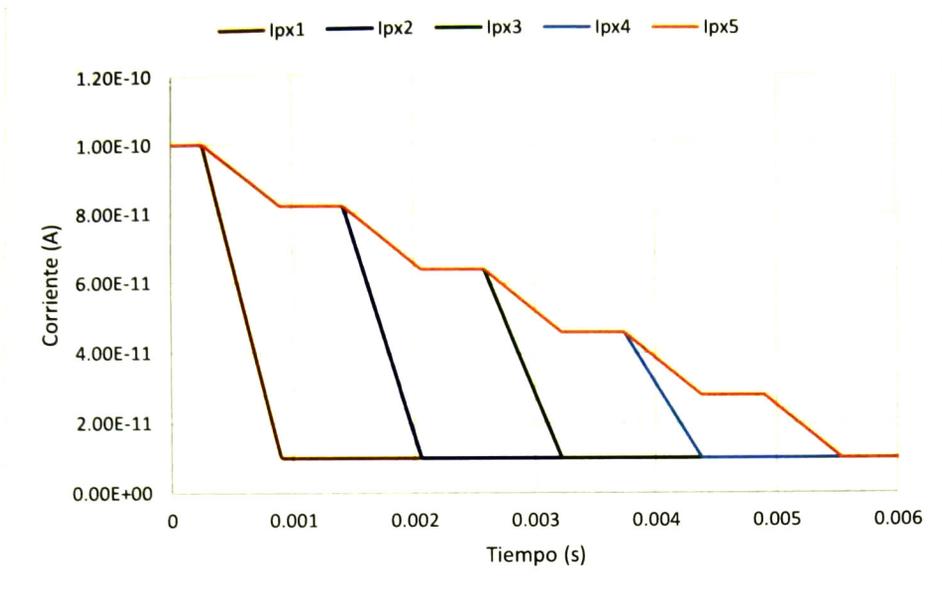


Fig. 3.19: Entrada de los 5 píxeles ante un borde de entrada de una barra oscura.

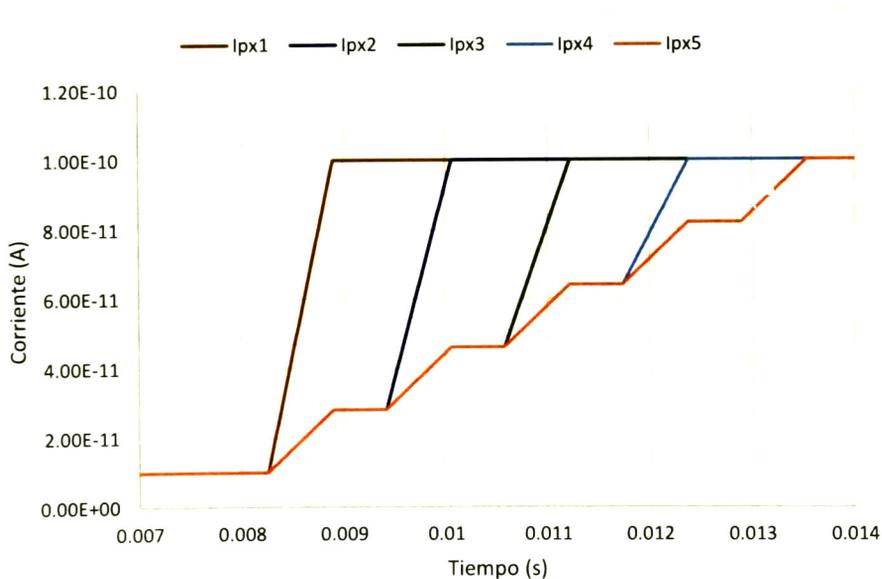


Fig. 3.20: Entrada de 5 píxeles ante el borde de salida de una barra.

Esta es la salida del fotorreceptor 3, al comparar esta salida con la respuesta obtenida por Indiveri, podemos notar cierto parecido. La cual es una buena representación de los resultados publicados por Giacomo.

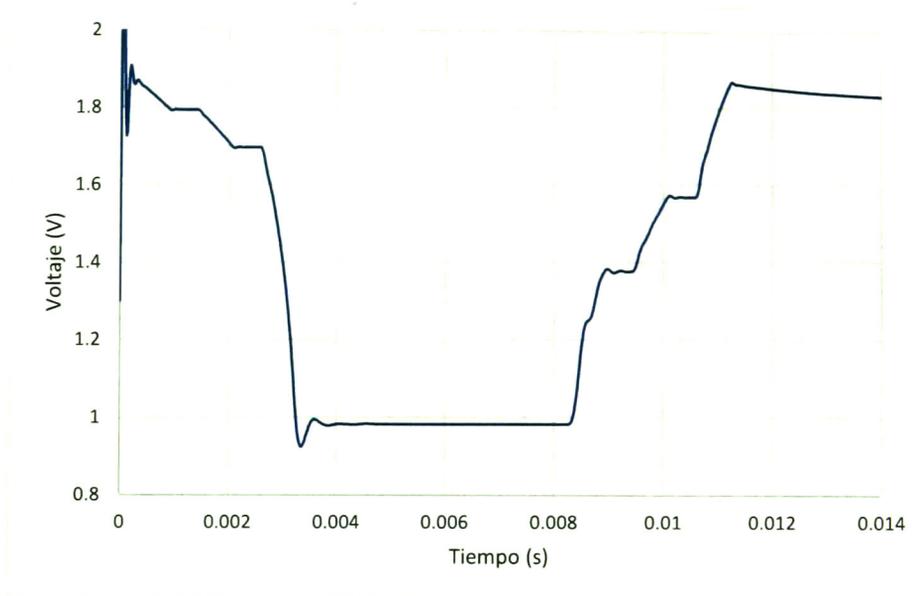


Fig. 3.21: Respuesta del fotorreceptor 3 ante ambos bordes.

3.4 Conclusiones.

En este capítulo se habló sobre los fotorreceptores, cómo se construyen, qué se espera de su funcionamiento y los elementos que lo conforman, algo interesante de este fotorreceptor es la habilidad de adaptarse ante la iluminación constante y así reaccionar ante solo los cambios de iluminación, esta propiedad hace del fotorreceptor de un elemento indispensable para la detección del movimiento ya que aísla los elementos móviles de la escena, este es el primer paso para la detección del movimiento, con las simulaciones comprobamos el funcionamiento, así como estimar la ganancia del dispositivo ante cambios en diferentes iluminaciones y velocidades, se aplicó un perfil de entrada que corresponde a un escenario visual comparable con las pruebas que realizó Giacomo Indiveri en su dispositivo.

Capítulo 4

Derivada espacial y selector.

En este apartado se presentan circuitos básicos que son muy usados para la implementación de circuitos neuromórficos en tecnologías VLSI, en especial se presentara el circuito diferenciador OTA y un circuito selector.

4.1 Introducción.

El elemento computacional más elemental en un sistema neuronal es la membrana de una neurona. Es una estructura estable que se comporta como un aislante perfecto, la corriente que fluye por la membrana puede controlarse por canales iónicos que se comportan de manera similar a los dispositivos MOSFET operados en la región subumbral, característica que propicia a un consumo menor de energía. Otra característica atractiva que favorece al uso de elementos analógicos para el procesamiento de información, es que se provee un procesamiento paralelo con una velocidad limitada solo por el retraso de las señales de la red neuromórfica.

En la ciencia de la inteligencia artificial se utilizan neuronas artificiales que están compuestas de varios elementos computacionales como lo son un multiplicador, un sumador y una función sigmoïdal, estas operaciones pueden ser realizadas por elementos analógicos en tecnología VLSI, en especial la operación sigmoïdal.

Una de las configuraciones mostradas en este capítulo es el par diferencial, esta configuración tiene la particularidad de transformar la información de voltaje a corriente, con una función de transferencia que es similar a la función canónica elemental en el campo de la mecánica estadística: la función de Fermi.

La función de Fermi es una función sigmoïdal, éstas son funciones diferenciables acotadas por límites reales que tienen una derivada positiva en donde sea. Esta función es comúnmente usada en redes neurales con retroalimentación positiva debido a su no linealidad y la simplicidad de su derivada.

4.2 Amplificador de transconductancia.

Uno de los recursos más comunes usados en la biología y dispositivos creados para el cálculo de mediciones insensibles a los valores absolutos de referencia y robustos al ruido, es usar un diferenciador de señales. El amplificador de transconductancia genera en su salida una corriente que es una función de la diferencia entre las dos entradas de voltaje. Una conductancia ordinaria transforma el voltaje en corriente.

4.2.1.1 Par diferencial.

Muchos de los circuitos que representan una diferencia entre voltajes utilizan alguna variante del par diferencial mostrado en el la **Fig. 4.1**. El par diferencial es un circuito compacto compuesto de solo tres transistores que son muy usados en muchos sistemas neuromórficos. Tiene la propiedad deseable de aceptar una diferencia de voltaje como entrada proveyendo una salida de corriente diferencial con características muy útiles.

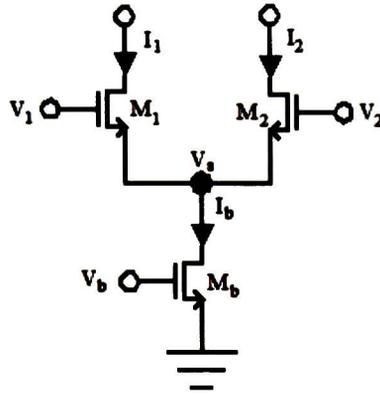


Fig. 4.1: Par diferencial básico.

El par diferencial tiene la misma estructura que un seguidor de fuente exceptuando la corriente de polarización I_b que está compartiéndose por dos transistores MOSFET M_1 y M_2 cuyas fuentes están conectadas al drenaje M_b como se muestra en la figura, la corriente se comparte entre los transistores M_1 y M_2 manteniendo una dependencia en sus respectivos voltajes de compuerta V_1 y V_2 . Si todos los transistores operan bajo el umbral y en saturación, asumiendo que M_1 y M_2 tienen el mismo factor κ_n . La corriente de saturación de un transistor se define como:

$$I_{sat} = I_0 e^{\frac{\kappa V_g - V_s}{U_T}} \quad (4.1)$$

Al aplicar esta ecuación a los transistores M_1 y M_2 obtenemos

$$I_1 = I_0 e^{\frac{\kappa V_1 - V_s}{U_T}} \quad \text{y} \quad I_2 = I_0 e^{\frac{\kappa V_2 - V_s}{U_T}} \quad (4.2)$$

La suma de las dos corrientes debe de ser igual a la corriente I_b

$$I_b = I_1 + I_2 = I_0 e^{-\frac{V_s}{U_T}} \left(e^{\frac{\kappa V_1}{U_T}} + e^{\frac{\kappa V_2}{U_T}} \right) \quad (4.3)$$

Resolviendo la ecuación para el voltaje V

$$e^{-\frac{V_s}{U_T}} = \frac{I_b}{I_0} \frac{1}{e^{\frac{\kappa V_1}{U_T}} + e^{\frac{\kappa V_2}{U_T}}} \quad (4.4)$$

Sustituyendo la ecuación (4.4) en (4.2)

$$I_1 = I_b \frac{e^{\kappa V_1 / U_T}}{e^{\kappa V_1 / U_T} + e^{\kappa V_2 / U_T}} \quad (4.5)$$

$$I_2 = I_b \frac{e^{\kappa V_2 / U_T}}{e^{\kappa V_1 / U_T} + e^{\kappa V_2 / U_T}} \quad (4.6)$$

Si V_1 es mucho más positivo que V_2 , el transistor M_2 se apaga así que esencialmente toda la corriente pasa por M_1 , I_1 es aproximadamente igual a I_b e I_2 es aproximadamente igual a 0. La dependencia de estas dos corrientes en la entrada diferencial de voltaje se muestra en la **Fig. 4.2**. Son casi lineales para pequeñas diferencias de voltaje y se saturan en I_b para altas diferencias. Estas funciones son muy útiles para la implementación de diferentes funciones, en especial en redes neurales.

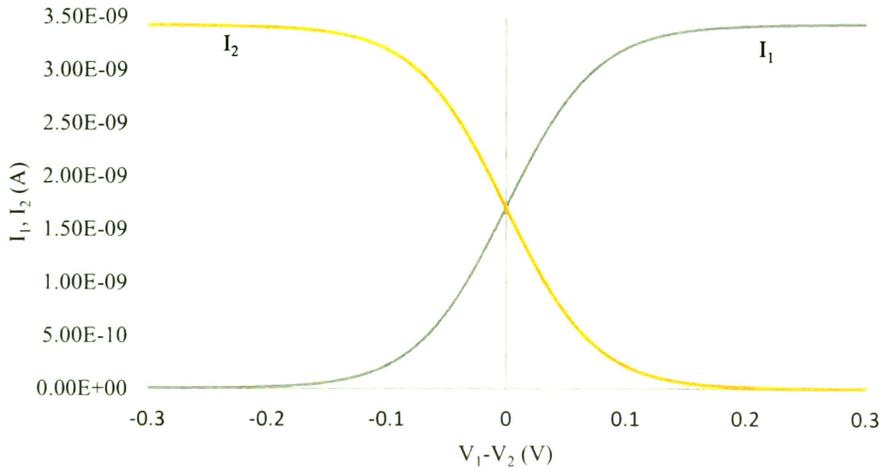


Fig. 4.2: Respuesta en corriente del par diferencial.

En la simulación del par diferencial se utilizaron transistores del mismo tamaño, un voltaje de alimentación V_{aa} de 3V y un voltaje de polarización V_a de 0.6 V, generando 3.5×10^{-9} A de corriente de polarización.

Para proveer una salida del término diferencial $I_1 - I_2$ usando una simple terminal, uno necesita simplemente conectar un espejo de corriente de tipo complementario para las salidas diferenciales. El circuito entonces sería un amplificador de transconductancia.

Al conectar un espejo de corriente en un par diferencial como se muestra en la **Fig. 4.3**. Este circuito es conocido como la versión más simple de un amplificador de transconductancia, mientras todos los transistores se mantengan en saturación y el par este operado en subumbral la salida de corriente está dada por la ecuación (4.7).

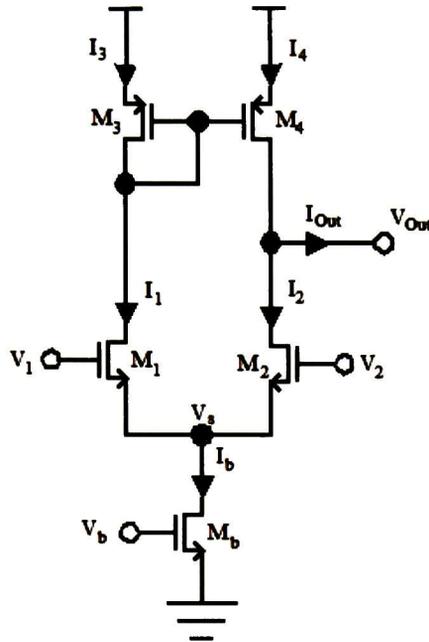


Fig. 4.3: Par diferencial con espejo de corriente también conocido como amplificador de transconductancia.

Los espejos de corriente se usan para generar una salida en corriente que es proporcional a la diferencia entre las corrientes de drenaje, esta diferencia está dada como:

$$I_1 - I_2 = I_b \frac{e^{\kappa V_1 / U_T} - e^{\kappa V_2 / U_T}}{e^{\kappa V_1 / U_T} + e^{\kappa V_2 / U_T}} \quad (4.7)$$

Multiplicando ambos, el numerador y denominador de la ecuación (4.7) por $e^{-(V_1+V_2)/2U_T}$ podemos expresar cada componente exponencial en términos de diferencias de voltaje, el resultado es:

$$I_1 - I_2 = \frac{e^{\kappa(V_1-V_2)/2U_T} - e^{-\kappa(V_1-V_2)/2U_T}}{e^{\kappa(V_1-V_2)/2U_T} + e^{-\kappa(V_1-V_2)/2U_T}} = I_b \tanh \frac{\kappa(V_1 - V_2)}{2U_T} \quad (4.8)$$

La ventaja de esta función de transferencia está en las propiedades de la tangente hiperbólica: pasa por el origen con una pendiente unitaria, se comporta de forma lineal con pequeñas entradas diferenciales, y se satura suavemente para grandes entradas diferenciales.

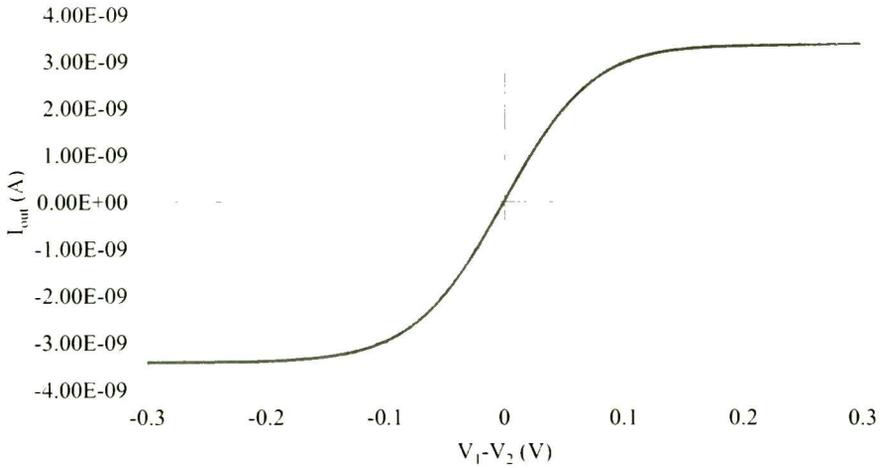


Fig. 4.4: Salida de un amplificador de transconductancia.

Si una aplicación requiere un rango extendido en la salida, el circuito puede ser expandido como se muestra en la **Fig. 4.5**, la corriente es reflejada dos veces, de tal manera que la salida sea simétrica y desacoplada de la etapa de entrada. La desventaja de esta extensión es que se utiliza el doble de transistores e incrementa el efecto de *missmatch* debido a la fabricación.

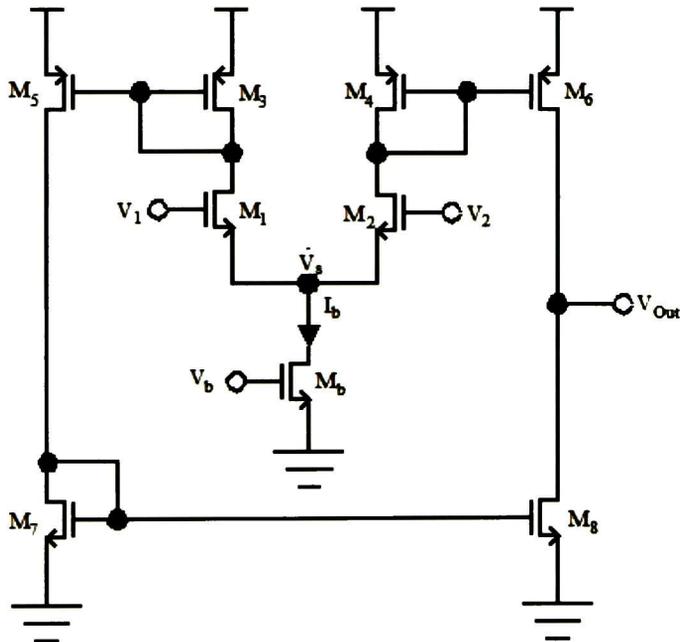


Fig. 4.5: Amplificador de transconductancia con salida desacoplada-

4.2.2 Simulación de la respuesta del amplificador de transconductancia en respuesta a señales del fotorreceptor.

Para verificar el funcionamiento del amplificador de transconductancia (OTA por sus siglas en inglés) se simula con entradas que se obtienen de la salida de los fotorreceptores, el circuito a simular se compone por lo mostrado en la Fig. 4.6, donde las entradas del circuito OTA está dado por lo mostrado en la Fig. 4.7, esta entrada es en respuesta a lo mostrado en el capítulo 3.

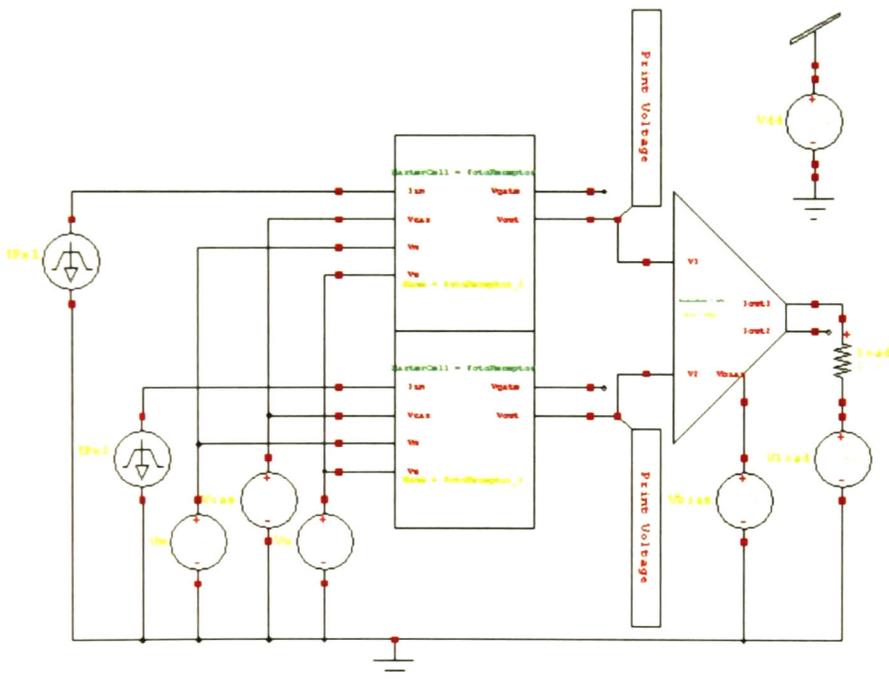


Fig. 4.6: Esquema de simulación del OTA con dos unidades de fotorecepción

La simulación consta de 5 fotorreceptores y 4 OTA conectados entre sí, con un voltaje de polarización de 0.6V y un voltaje de alimentación de 3.0V

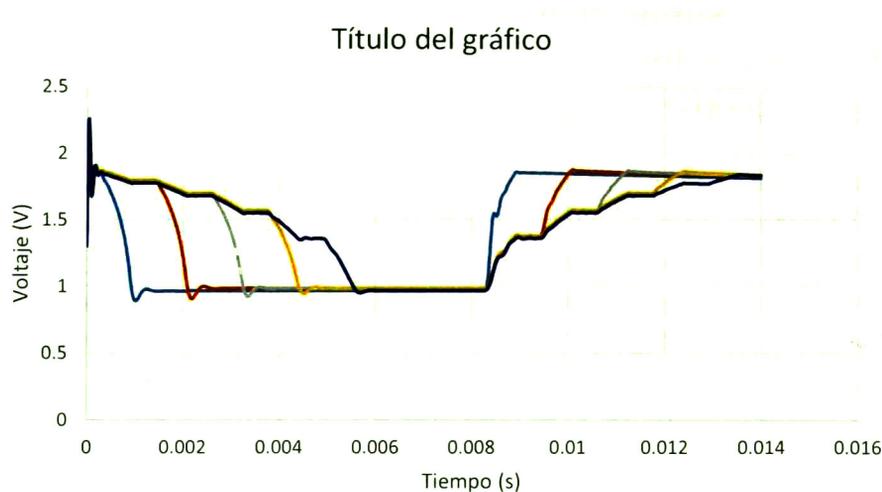


Fig. 4.7: Perfil de voltajes de entrada para los OTA

La respuesta del circuito OTA está mostrado en la **Fig. 4.8**

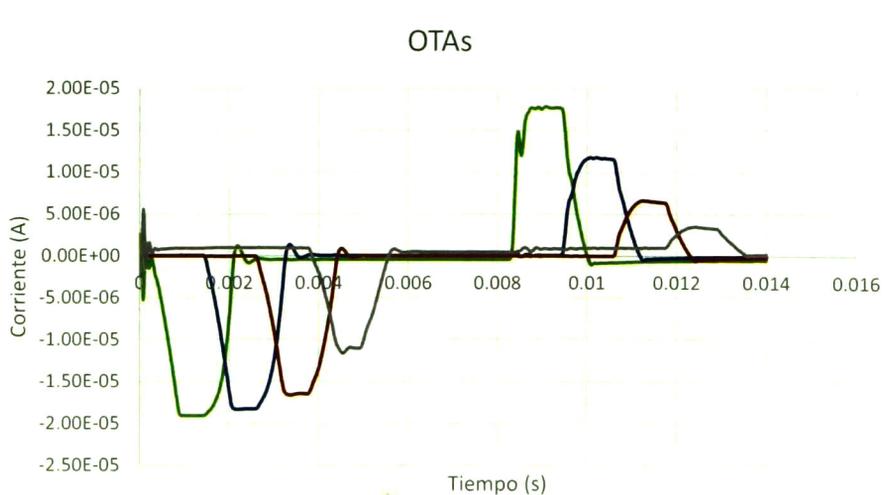


Fig. 4.8: Respuesta de los OTA ante el perfil de entrada de la Fig 4.7

Como se nota en la figura la mayor salida es de $20 \mu\text{A}$ que es la corriente de polarización del OTA. El elemento de mayor diferencia se detecta al entrar la barra oscura al campo visual, mientras va entrando hasta cubrir todo el campo visual no se generan cambios, este circuito realiza la derivada espacial del mapa de intensidades en el campo visual.

4.3 Fundamentos de los circuitos en modo corriente en débil inversión.

En los circuitos en modo voltaje el bloque principal usado para hacer operaciones básicas es el amplificador operacional. En circuitos en modo corriente este bloque sería el *current conveyor*.

El circuito tiene las siguientes propiedades:

- El potencial en la entrada de la terminal X es igual al voltaje aplicado en la terminal Y.
- Una corriente que es forzada en el nodo X resulta en una corriente igual a la que fluye en el nodo Y.
- La entrada de corriente fluyendo en el nodo X es transportada al nodo Z, la cual tiene la característica de ser una fuente de corriente con una salida de alta impedancia.

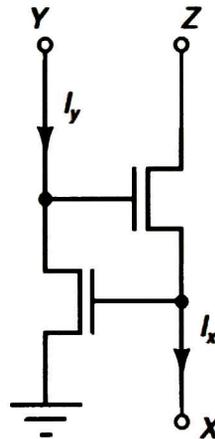


Fig. 4.9: *Current conveyor* simple usando solo dos transistores

La más simple implementación de un *current conveyor* es un transistor MOS, una forma más elaborada se muestra en la **Fig. 4.9**, esta forma básica con dos transistores es usada en muchos circuitos neuromórficos y es el elemento clave para el WTA que se analiza en el capítulo 5. Como se mencionó antes tiene la característica que I_z es igual a I_x y el voltaje de V_x tiene una relación logarítmica de I_y .

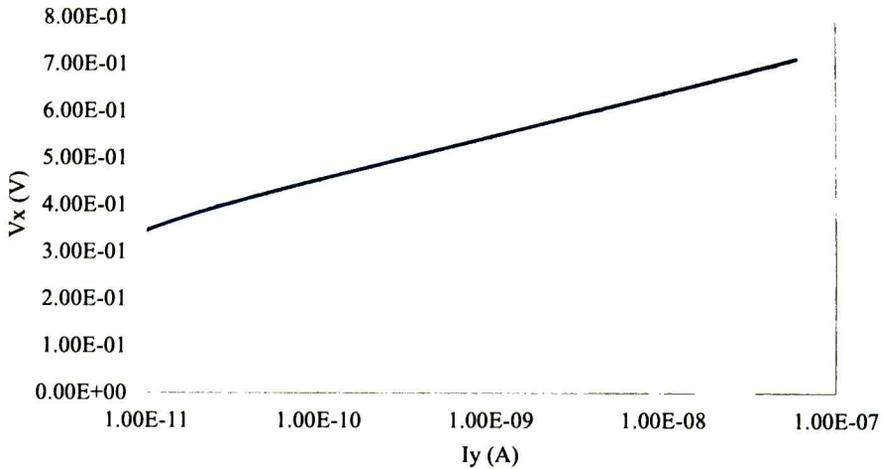


Fig. 4.10: Respuesta del current conveyour

En la **Fig. 4.10**, por otro lado, se muestra la respuesta logarítmica del voltaje V_x ante las variaciones de I_y . Esta respuesta resulta de simular el circuito mostrado en la **Fig. 4.9**, donde cada transistor guarda una relación $W/L = 3.0\mu\text{m}/3.0\mu\text{m}$ y las entrada de corriente en el nodo X es una constante de 6 nA mientras que en la entrada Y se encuentra una rampa de corriente de 0 a 12 nA.

Sedra y Smith reformularon la definición del current conveyour describiendo un circuito que combine ambas características de procesamiento, voltaje y corriente. Este nuevo tipo es conocido como current conveyour II (CCII).

Las características de entrada y salida se definen como:

$$\begin{bmatrix} V_x \\ I_y \\ I_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \pm 1 \end{bmatrix} \begin{bmatrix} V_y \\ V_z \\ I_x \end{bmatrix} \quad (4.9)$$

Las entradas de voltaje V_x y V_y están unidas por una ganancia unitaria, la terminal Y tiene una impedancia infinita y la corriente que es forzada en el nodo X es transportada a la salida en el nodo Z con una ganancia ± 1 . Un ejemplo de un circuito que ejemplifica el CCII es mostrado en la **Fig. 4.11**.

En las Fig. 4.12 se ejemplifica la corriente inyectada en el nodo Y es transportada e invertida en el nodo Z, al realizar la suma de las señales la diferencia es mínima Fig. 4.13.

Para simular el amplificador operacional, se utilizó uno que se encuentra en una librería de Tanner.

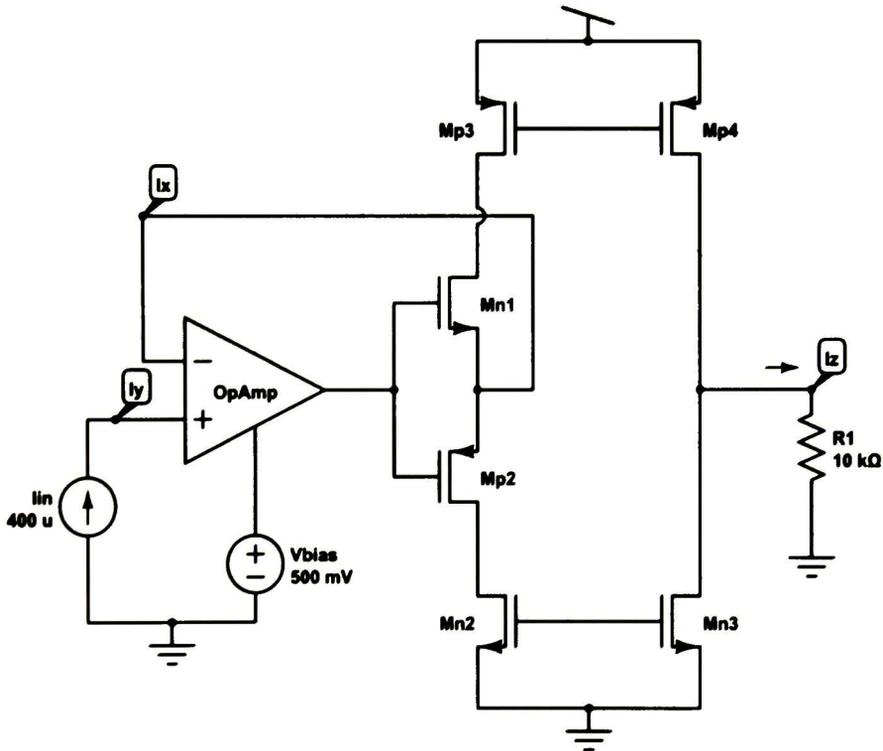


Fig. 4.11: Current Conveyor class II.

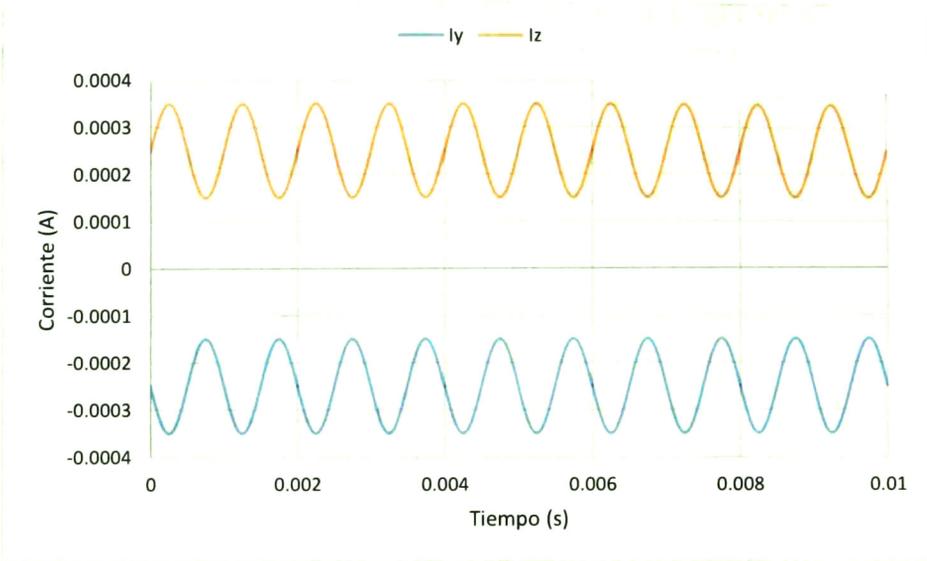


Fig. 4.12: Simulación del circuito *current conveyour* corrientes características.

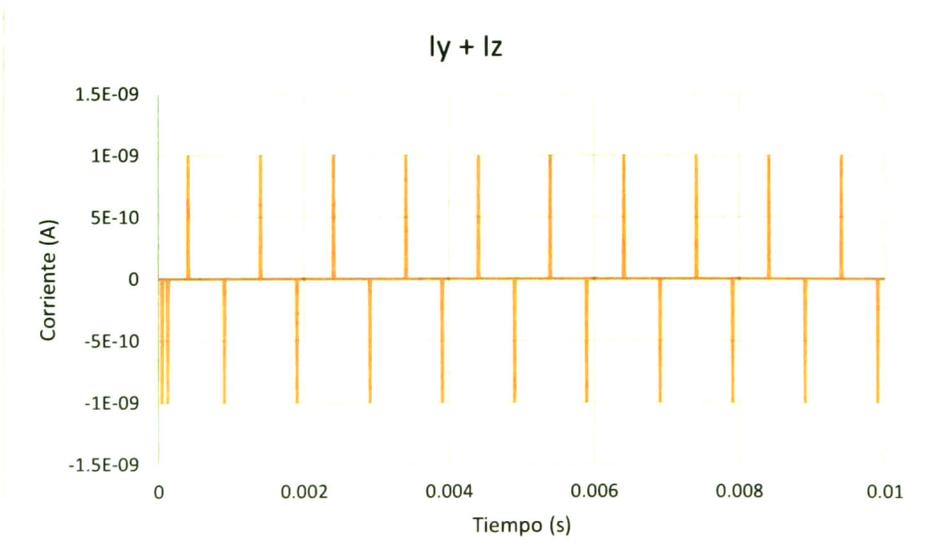


Fig. 4.13: La diferencia entre las corrientes del *current conveyour*, son mínimas.

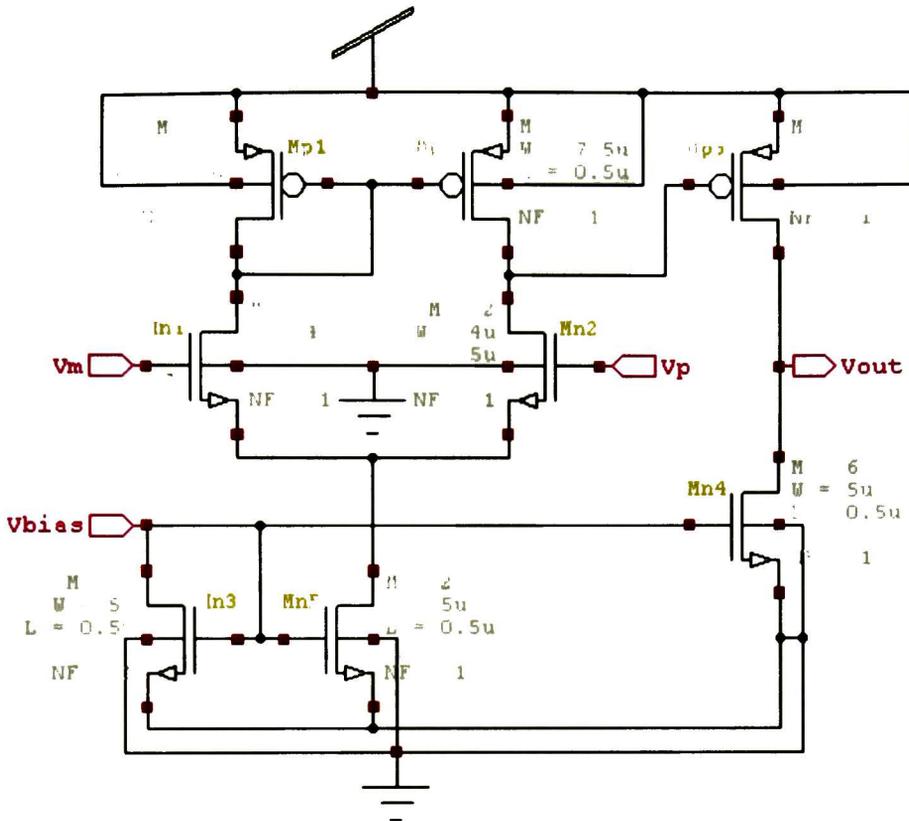


Fig. 4.14: Amplificador operacional suministrado por el *software* Tanner.

4.4 Descripción del circuito de detección de polaridad.

La polaridad de los bordes en la escena visual está codificada por la señal en corriente que nos dan los amplificadores de transconductancia. Cada una de estas corrientes es la entrada del circuito mostrado en la Fig. 4.15.

El circuito de detección de polaridad y selección de corrientes positivas o negativas, compara la señal de contraste (brillante-oscuro u oscuro-brillante) en un extremo (derivada) o cuando el contraste está creciendo o decreciendo en ese extremo.

Los transistores M1-M6 implementan un transportador de corriente II. Dependiendo de los voltajes en REF y CTRL se seleccionan las componentes positivas, negativas o ambas de la corriente de entrada

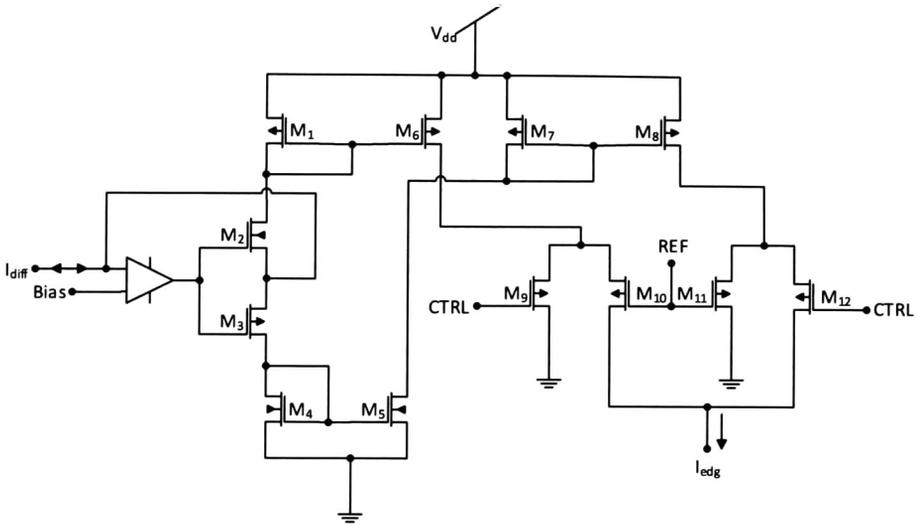


Fig. 4.15: Circuito para la detección de bordes.

4.4.1 Simulación de la respuesta del circuito de detección de polaridad en función de la respuesta del amplificador de transconductancia.

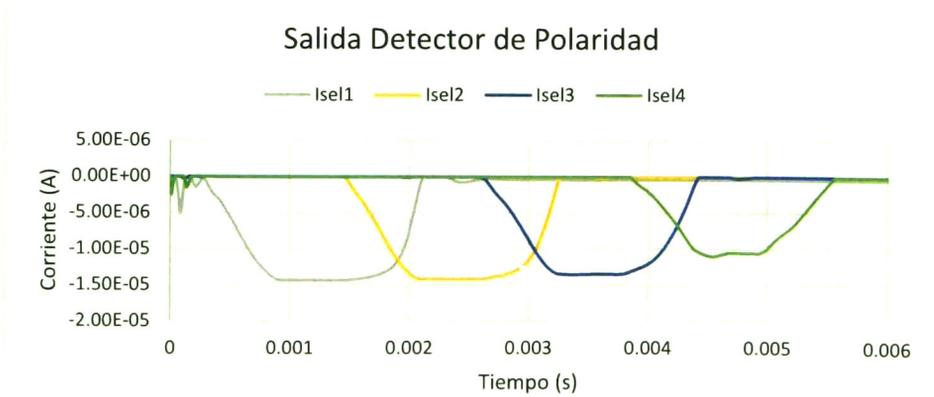


Fig. 4.16: Salida seleccionando la parte negativa de una señal.

Con las salidas generadas por el amplificador de transconductancia que dependen del perfil de entrada visto en capítulo 3 se obtienen estos resultados. La primera figura muestra la salida para un perfil de entrada y la segunda figura muestra la salida de un perfil de salida. El selector está

configurado para seleccionar los perfiles de apagado que se puede considerar como el perfil de entrada de la barra negra pasando sobre un fondo blanco.

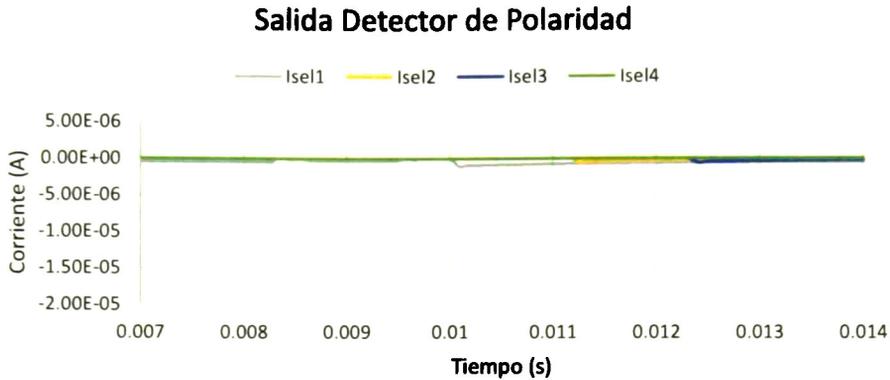


Fig. 4.17: Salida obviando la parte positiva de la señal.

4.5 Conclusiones.

En la ciencia de la inteligencia artificial se describen modelos de redes neuronales usando un conjunto de operaciones básicas, de las cuales en este capítulo se mostró la que más se utiliza para el procesamiento de imágenes, así también como circuitos que nos ayudan a conservar la integridad de la información para un mejor computo.

En especial, esta parte es del procesamiento de imágenes es el más crítico ya que aquí es donde se define el contraste de la imagen para después determinar el elemento más prominente y realizar una decisión sobre lo que ocurre en la escena visual.

Capítulo 5

WTA

En este capítulo se presenta el circuito WTA con diferentes configuraciones para poder realizar operaciones útiles en el procesamiento de imágenes, en especial para detectar de una mejor manera los bordes y que pueda mantener una selección de un objeto de importancia en la escena visual. Estas configuraciones se basan en una retroalimentación positiva que dota de características histeréticas a la red, configuraciones con redes resistivas excitadoras e inhibitoras, suavizan la entrada y desacoplan la respuesta deseada permitiendo una mejor selección del ganador.

5.1 Principios del WTA.

El sistema de procesamientos de la información visual es extraordinariamente complejo, pero quizás la más importante parte del sistema es el circuito WTA (*Winner take all*), que identifica las señales de más alta intensidad de múltiples entradas.

El sistema humano de procesamiento visual está construido de numerosas y complejas capas neuronales que se comunican unas con otras por medio de conexiones neuronales en retroalimentación positiva (*feedforward*) y negativa (*feedback*). Por el intermedio de estas conexiones, cada neurona hace frecuentemente señales a las otras en localizaciones dentro de las capas (*intro-layers*) o entre capas (*inter-layers*) emitiendo pulsos de corrientes eléctricas. Cada vez que una neurona genera un pulso, la información de la dirección es sentida por una unión neuronal llamada sinapsis (*synapse*), que está temporalmente conectada a una línea central sensorial (también conocida como bus), donde muchas otras neuronas están compitiendo simultáneamente por el derecho de propagarse a fin de viajar más. En tal competencia, la regla general es: La neurona recipiente en el extremo de la línea del bus únicamente escuchará a las neuronas que son activas cuando están activadas (p. ej., las ganadores son aquellas que tienen la señal más fuerte y consistente) e ignoran el resto.

La red WTA es la celda de computación aplicada a modelos de redes neuronales en la cual, las neuronas compiten entre sí para la activación, en principio solo la neurona de mayor valor se mantiene activa. En sentido estricto el circuito realiza una función máxima. En la **Fig. 5.1** se representa un modelo de red WTA simple (pero poderoso) que describe la población de N unidades excitadoras homogéneas que estimulan a una sola unidad inhibidora global y ésta se retroalimenta para inhibir todas las unidades estimuladoras.

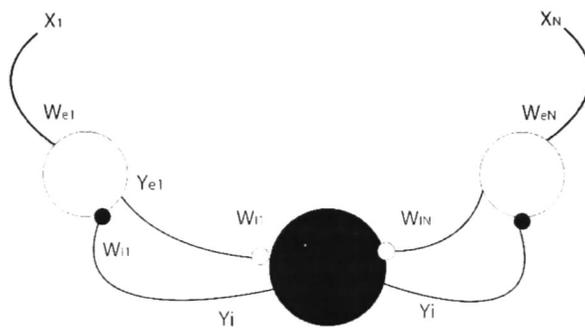


Fig. 5.1: Red de N neuronas excitadoras (blancas) proyectándose a la neurona común inhibidora (Negra).

Considere la red mostrada en la **Fig. 5.1**, donde la entrada externa de la j th neurona estimuladora es x_j , y la respuesta de la j th neurona excitadora es y_{ej} , la respuesta de la celda inhibidora será expresada por y_i , en la cual los

pesos de las sinapsis de las entradas externas a las neuronas excitadoras (de las neuronas excitadoras a las inhibidoras y viceversa), se representan por w_{ej} , w_{ij} y w_{lj} respectivamente. Con esta definición puede describirse esta red como:

$$y_{ei} = f(w_{ej}x_j - w_{ij}y_i) \quad (5.1)$$

$$y = f\left(\sum_{j=1}^N w_{lj}y_{ej}\right) \quad (5.2)$$

donde $f()$ denota la función de transferencia tanto para las neuronas inhibidoras como para las excitadoras; en la práctica, este sistema describe las interacciones recurrentes entre las neuronas excitadoras e inhibidoras. Considerando el caso en el que las neuronas tengan un comportamiento de umbral lineal y una entrada externa sea mucho mayor a las demás.

Se asume que el peso para una unidad excitadora externa de entrada, j_0 , excede la entrada inhibidora de la misma unidad y los demás pesos de las unidades restantes no lo logran, bajo estas condiciones:

$$\begin{aligned} y_{ej_0} &= (w_{ej_0}x_{j_0} - w_{ij_0}y_i) \\ y_{ej} &= 0 \quad \forall j \neq j_0 \\ y_i &= w_{lj_0}(w_{ej_0}x_{j_0} - w_{ij_0}y_i) \end{aligned} \quad (5.3)$$

de donde se puede simplificar obteniendo:

$$\begin{aligned} y &= \frac{w_{ej_0}x_{j_0}}{1 + w_{lj_0}w_{ij_0}} \\ y_{ej} &= 0 \quad \forall j \neq j_0 \\ y_i &= \frac{w_{ej_0}w_{lj_0}x_{j_0}}{1 + w_{lj_0}w_{ij_0}} \end{aligned} \quad (5.4)$$

La función de umbral lineal se expresa por la función de rectificación de media onda $f(x) = \text{Max}(0, x)$ que es una representación más biológica y realista de una neurona.

Esta solución (5.4) solo es satisfecha mientras la entrada x_{j_0} sea la entrada con el mayor valor.

5.2 Aplicación en un circuito analógico.

Un circuito WTA, que identifica la más alta intensidad de la señal entre señales de entrada múltiples, es uno de los bloques más importante de construcción en varias redes neuronales, sistemas de control probabilísticos (*fuzzi*), y más frecuentemente, en sensores de imagen integrados en chips neuromórficos de visión con el objetivo de emular la capa óptica extremadamente sensitiva de la parte posterior del ojo humano que recibe la imagen producida por los lentes, la retina. Una vez que la neurona (también referida como la celda) con la señal más alta de entrada es exitosamente seleccionada por el circuito WTA, un cierto valor será asignado a esta celda ganadora por medio de una corriente o un voltaje, mientras que los valores nominales de las otras celdas serán puestos en nulo (p. ej., ellas pierden).

De hecho, se ha encontrado que los circuitos WTA diseñados en tecnología CMOS son apropiados para la implementación de chips neuromórficos de baja potencia y alta densidad.

El circuito mostrado en la **Fig. 5.2** es un circuito WTA analógico, continuo en el tiempo, propuesto por Lazzaro [29] y es muy usado en una amplia variedad de aplicaciones. El circuito es compacto y sencillo. Entre sus características destacan el procesamiento paralelo de las señales de entradas realizado solamente con dos transistores por celda, y en el que un solo transistor es el común de todas las celdas. El cálculo colectivo y la conectividad global se obtienen usando un nodo común a todas las celdas. En el arreglo mostrado en la **Fig. 5.2** cada celda consta de un *current conveyor*, esta celda es el bloque primitivo del WTA, tiene la propiedad de que I_z es igual a I_x y el voltaje de V_x tiene una relación logarítmica con I_y .

La red WTA es modular y puede ser extendida a N celdas. Al conectar las celdas adicionales al nodo V_c , las nuevas corrientes de entrada pueden ser representadas esquemáticamente por fuentes de corriente. En una versión de circuitos integrados, las fuentes de corriente se pueden implementarse mediante un transistor CMOS de tipo p funcionando en saturación en el subumbral y su valor controlado por medio de su voltaje de compuerta.

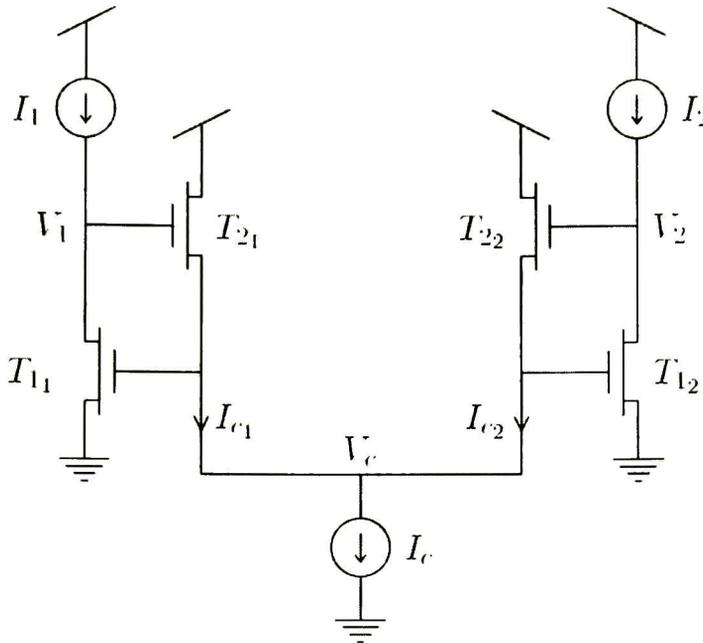


Fig. 5.2: Circuito WTA clásico.

El funcionamiento de este circuito se describe a continuación: Las señales de salida están codificadas por las corrientes I_{c1} e I_{c2} y los voltajes V_1 y V_2 , que son los voltajes de drenaje de los transistores T_{11} y T_{12} . La corriente de saturación está impuesta por I_c . Los transistores T_{11} y T_{12} descargan el nodo V_1 y V_2 , por tanto implementan una retroalimentación inhibitoria, por otra parte, los transistores T_{21} y T_{22} implementan la retroalimentación excitadora al cargar el nodo V_c . En general el circuito mostrado en la **Fig. 5.2**, toma como entradas a I_1 e I_2 y como salidas I_{c1} e I_{c2} . Todos los transistores son de canal n y de idéntica relación W/L ; en este circuito las corrientes I_1 e I_2 compiten entre sí para activar su salida I_{c1} o I_{c2} dependiendo del ganador, éste se decide en la retroalimentación de los transistores T_{11} y T_{12} ya que el voltaje que habrá entre éstos (y en su compuerta) decidirá cuál salida se mantiene activa.

5.2.1 Comportamiento en estado estable de la red WTA básica.

Se analiza el circuito en estado estable al proporcionar entradas de corriente constantes y midiendo la respuesta bajo tres casos particulares: ambas

entradas idénticas en todo sentido, una entrada mucho mayor que la otra, es decir que exista una diferencia considerable entre las entradas, y cuando las entradas sean diferidas en una pequeña cantidad.

5.2.1.1 Ambas entradas iguales.

Empezaremos por analizar este circuito considerando un escenario en el que sus dos corrientes de entrada son idénticas ($I_{in1} = I_{in2} = I$). Los transistores $T1_1$ y $T1_2$ actúan como sumideros de corriente que drenan las correspondientes corrientes I dentro de tierra. Los transistores $T2_1$ y $T2_2$ drenan corrientes I_{c1} e I_{c2} respectivamente. Debido a que $T1_1$ y $T1_2$ tienen idénticos potenciales de compuerta y de fuente, mientras que inyectan la misma corriente a tierra, sus potenciales de drenador son idénticos ($V_1 = V_2 = V_{out}$). Por esto, $T2_1$ y $T2_2$ tienen idénticos potenciales de compuerta, de fuente y de drenaje, resultando en un par de corrientes de drenaje. Considerando que ambas celdas tienen idénticos espejos de corriente de salida, tendremos: $I_{c1} = I_{c2} = I_{out}$

Los transistores de cada celda WTA en la **Fig. 5.2** operan todos en el subumbral, como se mencionó previamente, ofreciendo la posibilidad de disipación de potencia ultra baja, que es crucial para un sistema neuronal integrado.

Describiendo $T1_1$ y $T1_2$ empleando la ecuación (2.8):

$$I_d = I_0 e^{\frac{\kappa V_c}{U_T}} \left(1 - e^{-\frac{V_{out}}{U_T}} \right) \quad (5.5)$$

$T2_1$ y $T2_2$ pueden describirse de manera similar.

$$I_d = I_0 e^{\frac{\kappa V_{out} - V_c}{U_T}} \left(1 - e^{-\frac{V_{dib} - V_c}{U_T}} \right) \quad (5.6)$$

V_{dib} es el voltaje de drenador del transistor $T2_i$. Resolviendo para V_{out} como una función de I e I_c , encontramos que V_{out} codifica logarítmicamente la corriente de entrada I (con una compensación determinado por la magnitud de I_c) [23].

Si ambos transistores están en saturación, las corrientes de salida deben ser idénticas, por la ley de Kirchoff de corriente en el nodo común V_c .

$$I_{Out,1} = I_{Out,2} = \frac{I_b}{2} \quad (5.7)$$

Este comportamiento implica que cada celda WTA puede manejar un más que amplio rango dinámico, dado que V_{out} varía linealmente entre dos límites: apagando y saturando T_{11} y T_{12} de acuerdo con un rango de corriente de corrientes de entrada I_{in} en varios órdenes de magnitud. Adicionalmente, cada celda implementa esencialmente un convertidor corriente a voltaje de características de transferencia, entrada/salida, logarítmica.

5.2.1.2 Una entrada mucho mayor que la otra.

Prosiguiendo, consideremos la situación en la que $I_{in1} \gg I_{in2}$. Basados en la ecuación (2.8), encontramos que la diferencia entre I_{in1} e I_{in2} se refleja por el término entre paréntesis cuadrados, esto es:

$$\left[1 - e^{-\frac{V_{out}}{U_T}} \right] \quad (5.8)$$

dado que T_{11} y T_{12} comparten el mismo voltaje de compuerta, V_c .

Por esto, $I_{in1} > I_{in2}$ se traslada a $V_{out1} \gg V_{out2}$. Suponiendo que V_{out2} es suficientemente pequeño para apagar T_{22} . Encontramos que I_{c2} se reduce a cero. Como resultado la corriente de polarización debe fluir a través de T_{21} , significando que $I_{c1} = I_c$. De esta manera, decimos que la celda 1 ha ganado sobre la celda 2 y toma todas las corrientes de polarización, también puede mostrarse que el potencial ganador V_{out1} codifica I_{in1} logarítmicamente.

Recordando que la corriente de un transistor está compuesta por dos componentes, la componente directa (*forward*) I_f (5.9) y la componente inversa (*reverse*) I_r (5.10). Cuando el voltaje de fuente V_s del transistor es aproximadamente igual al voltaje de drenaje V_d , la componente *reverse* es comparable con la componente *forward*.

$$I_f = I_0 e^{\frac{\kappa V_g - V_s}{U_T}} \quad (5.9)$$

$$I_r = I_0 e^{\frac{\kappa V_g - V_d}{U_T}} \quad (5.10)$$

En el caso donde el voltaje de drenaje de T_{11} sea mayor que el voltaje de drenaje de T_{12} , si el transistor T_{11} está en saturación el componente dominante será la corriente de *forward* y el voltaje de compuerta V_c se incrementará de tal manera que:

$$I_{d1} = I_{f1} = I_0 e^{\frac{\kappa V_c}{U_T}} = I_{in1} \quad (5.11)$$

Aunque las corrientes en los puertos de entrada sean diferentes, la componente *forward* de T_{11} y T_{12} es igual, porque los dos transistores comparten el mismo voltaje de compuerta V_c y ambos tienen sus fuentes conectadas a tierra.

Para este caso en particular, la corriente de drenaje I_{d2} del transistor T_{12} puede solo ser igual a la corriente de entrada I_{in2} bajo la siguiente condición:

$$I_{r2} = I_{in1} - I_{in2} \gg 0 \quad (5.12)$$

La componente reverse de I_{d2} se hace significativa solo cuando V_{d2} decrezca lo suficiente para operar en la región óhmica. En este caso la salida del transistor T_{22} esté apagado de manera efectiva e $I_{out2} = 0$ consecuentemente T_{21} conduce toda la corriente de polarización I_b tal que

$$I_b = I_0 e^{\kappa V_{d1} - V_c} \quad (5.13)$$

El anterior análisis de la celda ganadora se base sobre una consideración de señal grande en la que I_{in1} e I_{in2} son suficientemente diferentes. Cuando I_{in1} e I_{in2} son bastante similares, la anterior consideración no se mantiene y consecuentemente debemos hacer un análisis de pequeñas señales, donde el voltaje *early* (V_e) será importante.

5.2.1.3 Dos entradas que difieran por una pequeña cantidad.

Ahora, consideremos que $I_{in1} = I + \Delta I$, $I_{in2} = I$, y que todos los transistores T_{11} , T_{12} , T_{21} y T_{22} están operando en el subumbral y en saturación. Para entender el cambio en V_{out1} debido a ΔI , veamos a T_{11} como un resistor r_{ds} , cuya resistencia está dada por V_e/I_d como se mencionó, por lo que tenemos

$$\Delta V_{out1} = \Delta I \cdot r_{ds} \approx \frac{V_e}{I + \Delta I} \quad (5.14)$$

Enseguida, se investiga cómo M_3 ayudará a la Celda 1 a ganar sobre la Celda 2 a la luz de ΔI . Asumiremos primero que el nodo V_{c1} permanece al mismo potencial a pesar de ΔI (Esta consideración es muy cercana a la realidad especialmente cuando ΔI es pequeña, ya que cualquier decrecimiento en V_{c2} será compensado por una cantidad equivalente de

incremento en V_{C1}). Por medio del empleo de (2.8), encontramos que la transconductancia de $T2$, será

$$g_{m3} = I_{bias} \left(\frac{\kappa}{U_T} \right) \quad (5.15)$$

De las ecuaciones anteriores, vemos que el cambio en I_{C1} debido a ΔI está dado por,

$$\Delta I_{C1} = g_{m3} \cdot \Delta V_{out1} = \left(\frac{I_{bias}\kappa}{U_T} \right) \cdot \Delta I \cdot \frac{V_B}{I_{int}} \quad (5.16)$$

$$\Delta I_{C1} = \frac{\kappa V_E}{U_T} \cdot \frac{\Delta I}{I_{int}} \cdot I_{bias} \quad (5.17)$$

El término entre paréntesis en la ecuación (5.16) representa el factor de ganancia que describe cuan sensitiva es la Celda con respecto a la diferencia de entrada. Asignando valores típicos para estos parámetros ($\kappa = 0.5$, $V_e = 50V$, y $U_T = 26mV$) se obtiene una ganancia de alrededor 962 o equivalentemente 60 dB.

Debido a la gran magnitud del factor de ganancia, un cambio muy pequeño ΔI será suficiente para causar un incremento significativo en I_{C1} (y este incremento será proporcionado por la fuente de corriente de la celda perdedora).

De la ecuación de voltaje de *early*, vemos que conforme el canal del transistor $T1$, es más largo, más grande será el voltaje *early* y a su vez el circuito WTA será más sensitivo (en respuesta a la diferencia de entrada), esto es, el circuito WTA demuestra una respuesta abrupta.

$$V_e = L \frac{\partial V_d}{\partial L} \quad (5.18)$$

El mismo efecto se puede alcanzar al agregar un transistor en degeneración de fuente como se muestra en la **Fig. 5.3**, debido a que se incrementa la transconductancia de entrada, un efecto similar al que ocasiona incrementar la longitud de canal de ese transistor.

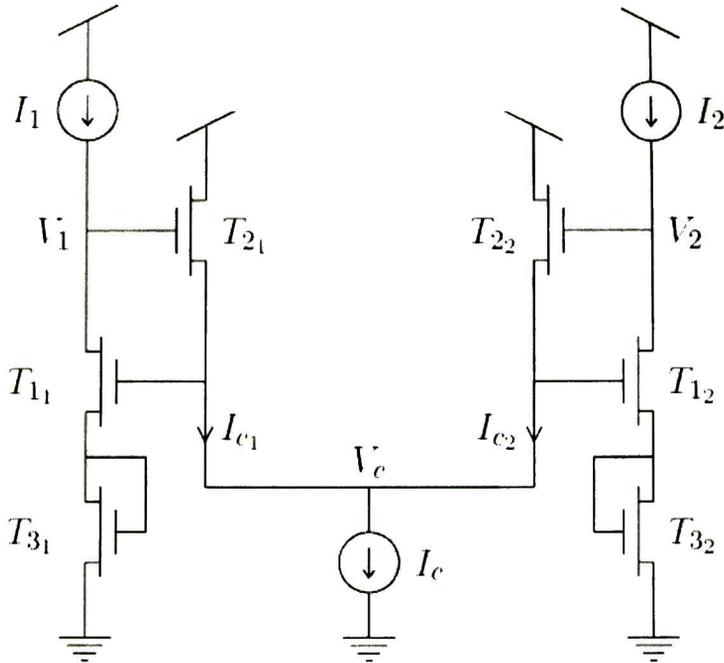


Fig. 5.3: Circuito WTA con degeneración en fuente.

5.3 Simulaciones de un WTA.

Se simula un WTA simple, **Fig. 5.4**, que consta de dos bloques principales CCI y una corriente de polarización de 10 nA con una entrada variando desde 7.5 nA hasta 4.5 nA la otra entrada se encuentra constante en un valor de 6 nA.

La **Fig. 5.5** muestra la respuesta en voltaje de dos celdas WTA ante diferencias de corriente en sus entradas, esta figura representa la competencia de cada célula por ser quien emita la señal de ganador.

En la **Fig. 5.6** se muestra la respuesta en corriente de una red WTA clásica, se puede apreciar los estados mencionados en la sección anterior, cuando las corrientes son iguales, se puede notar que la corriente en las salidas son idénticas expresado por la ecuación (5.7), cuando hay una diferencia mínima existe una relación expresada por (5.17) hasta que la diferencia empieza hacerse mayor, esto es alrededor de una diferencia de

alrededor de un 0.25%, en este momento la corriente de salida de la célula ganadora es la corriente de polarización.

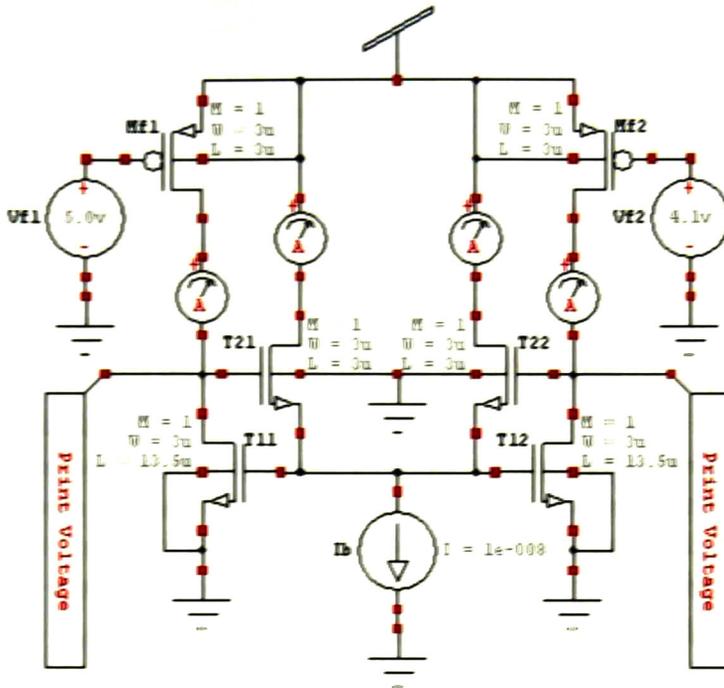


Fig. 5-4: Diagrama de simulación WTA clásico.

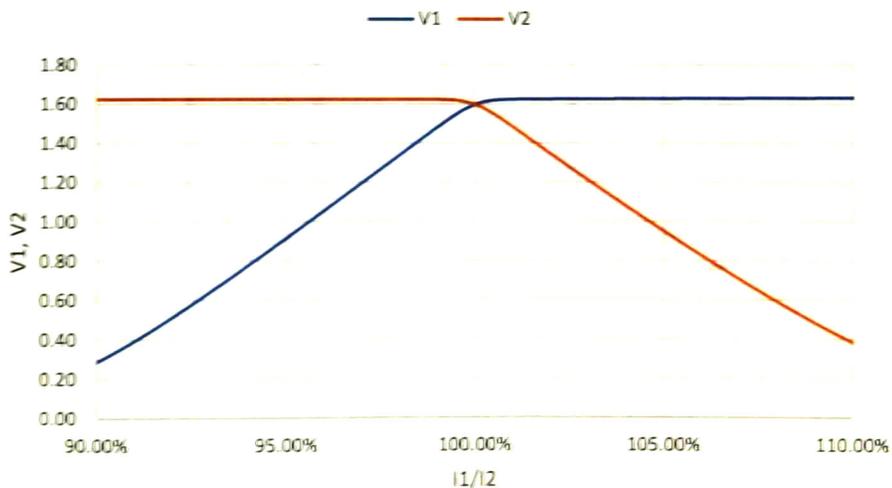


Fig. 5-5: Cambio de voltajes ante una diferencia de un 10% en las dos entradas.

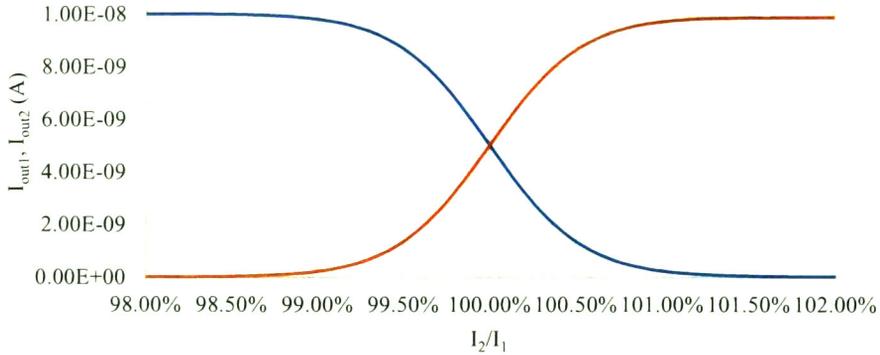


Fig. 5.6: Respuesta en corriente de una red WTA.

La siguiente figura muestra el amplio rango dinámico y las propiedades logarítmicas del circuito, se repitió el mismo experimento que en la figura anterior con corrientes de diferentes órdenes de magnitud, la línea continua que pasa sobre todos los órdenes de magnitud representa el nivel de voltaje cuando las corrientes son iguales, en 5 órdenes de magnitud diferentes se aprecia la funcionalidad del circuito.

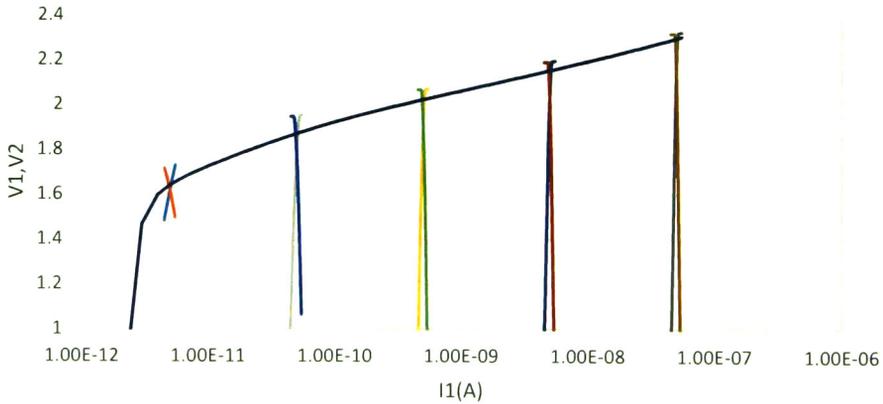


Fig. 5.7: Comportamiento del circuito WTA ante diferentes órdenes de magnitud en la entrada.

El comportamiento del circuito WTA cambia al incrementar el tamaño de los transistores $T1$, esto debido a que su conductancia determina la respuesta de pérdida del circuito. El voltaje de *early* es una medida de la conductancia de un transistor en saturación, hay una dependencia entre el

largo del canal y la respuesta del circuito, al cambiar la longitud del canal la pendiente se incrementa, esto hace que el circuito WTA se comporte como un WTA ideal, en la Fig. 5.8 y Fig. 5.9 se muestran una comparación usando L de $3\mu\text{m}$ y $13.5\mu\text{m}$.

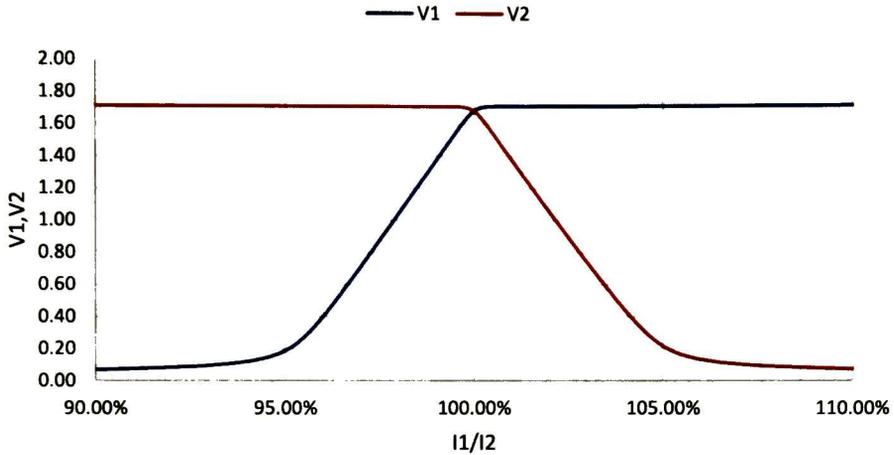


Fig. 5.8: Respuesta con el largo del Transistor M1 L = $13.5\mu\text{m}$

Como se mencionó en la sección anterior se puede obtener el mismo efecto al usar un transistor en degeneración de fuente, los resultados al colocar el transistor se muestran en la figura siguiente.

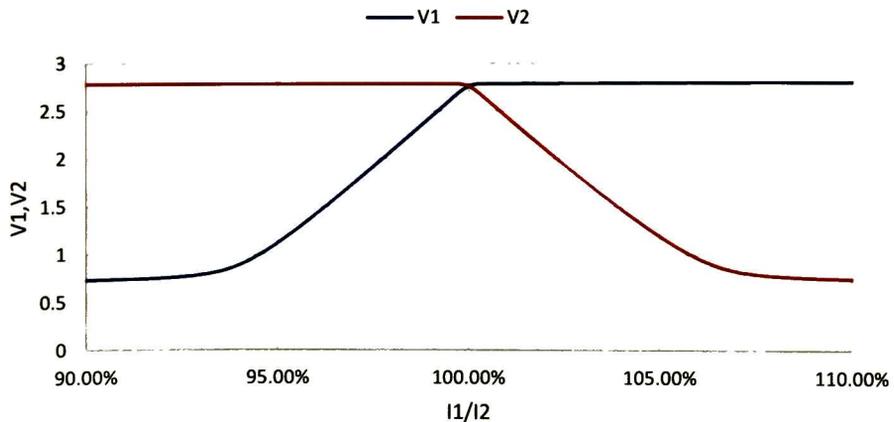


Fig. 5.9: Respuesta WTA con transistor de degeneración en fuente

La simulación consta del mismo circuito con el transistor extra del mismo tamaño, el efecto es similar al logrado con el aumento del largo del transistor.

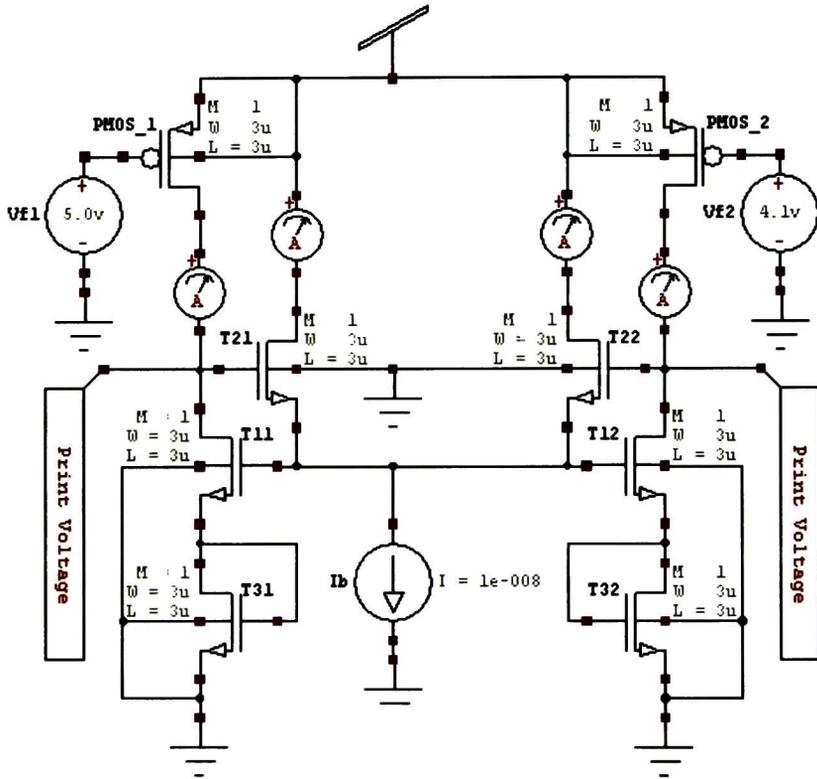


Fig. 5.10: Diagrama de simulación, WTA con degeneración en fuente.

5.4 WTA histerético.

El circuito WTA básico elige un ganador de un grupo de señales de entrada. Al usar el circuito para la selección y seguimiento de objetos, puede ocurrir un problema cuando el objeto de interés desaparece de la escena visual o cuando hay mucha actividad en la escena. Esto se debe a que el circuito WTA comparara las señales de entrada que se representan como niveles de intensidad del entorno de la imagen en tiempo real y ésta no tiene forma de mantener la atención en una en particular. De aquí la necesidad de usar filtros espaciales para el procesamiento de la señal.

El circuito WTA tiene una ganancia no lineal alta, esto presenta la posibilidad de que aparezcan cambios ilegítimos en la salida al procesar elementos con valores similares. Para resolver este problema, en [30] se propone agregar una retroalimentación local al WTA para lograr lo que se conoce como histéresis, ésta otorga la propiedad de mantener un estado en la ausencia del estímulo que lo ha generado, este tipo de propiedades dotan al circuito un mecanismo de atención.

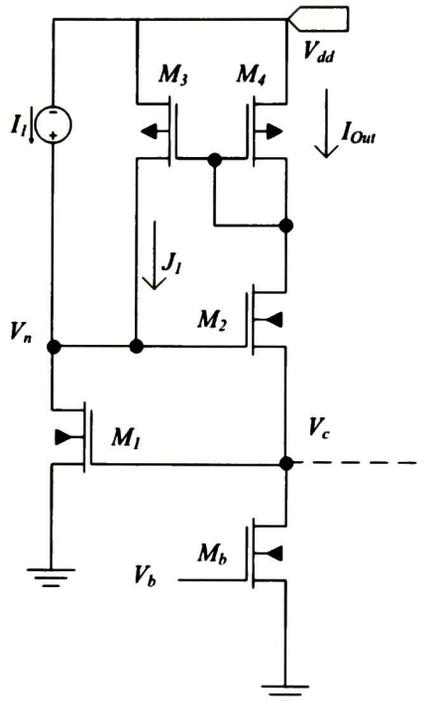


Fig. 5.11: Circuito WTA con retroalimentación histérica.

La **Fig. 5.11**, muestra el circuito de histéresis local, en éste cuando el circuito selecciona al ganador una corriente proporcional a la corriente de polarización de la red es retroalimentada a la entrada por medio de un espejo de corriente formado por los transistores M_3 y M_4 creando una corriente, que se denominará como histérica, es sumada con la corriente de entrada. La corriente histérica es aproximadamente igual a la corriente de polarización del nodo ganador y esencialmente cero para los demás, esto debido a la operación no lineal del WTA. Al sumar la corriente de entrada con la corriente histérica se es capaz de incrementar la habilidad para mantener el estado ganador de la celda ganadora.

Esta habilidad es determinada por el valor de la corriente histerética, la cual es la cantidad de corriente adicional que cada celda competitiva tiene que sobrepasar para lograr un estado ganador. Si la corriente de polarización está en el subumbral el factor de retroalimentación estará dado por la diferencia entre el voltaje de alimentación de M_3 y M_4 , esto a causa de las propiedades de un espejo de corriente.

Las propiedades estables del circuito WTA histerético son las mismas a las del circuito clásico. De manera similar la respuesta dinámica del WTA histerético es la misma a la del WTA clásico y depende principalmente del valor que tenga la corriente de polarización y del total de corrientes que entren a los puertos de entrada de las células WTA.

5.5 Simulaciones de un WTA histerético.

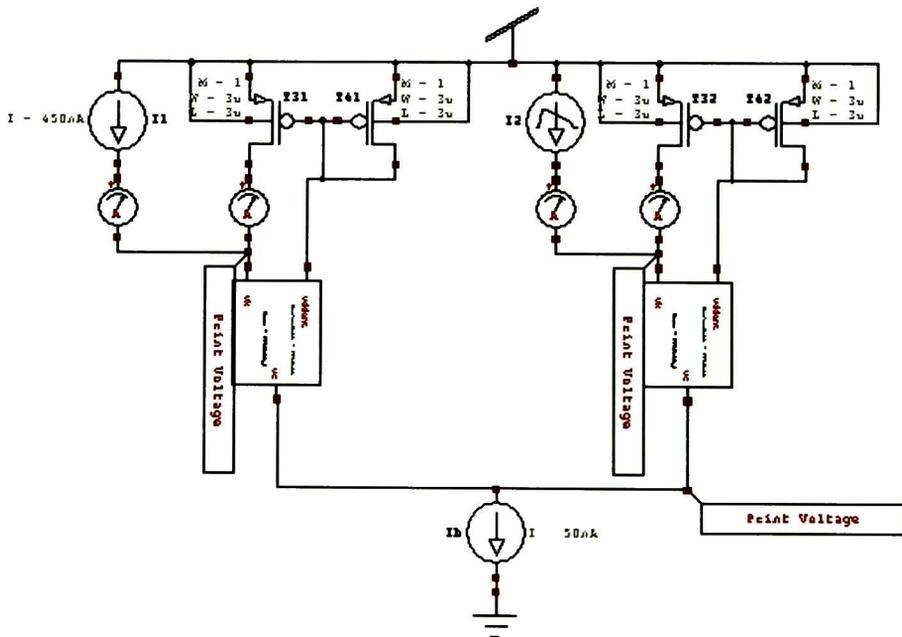


Fig. 5.12: Diagrama de simulación WTA con retroalimentación histerética.

En este caso para la simulación de un WTA histerético se usó, dos celdas básicas de WTA con retroalimentación, en la celda 1 se mantuvo una alimentación constante de corriente de 450 nA, mientras que la segunda celda se hizo un barrido de 400 nA a 700 nA, esta configuración se mantuvo para cuatro corrientes de polarización, 50 nA, 100nA, 150 nA, 200 nA.

El objetivo de esta simulación es mostrar la dependencia que tiene el efecto histerético ante cambios de la corriente de polarización, esto se logra al aumentar esta corriente, en la Fig. 5.13 se observa la necesidad de cada vez más corriente en la entrada para lograr romper el estado ganador de la celda ganadora.

Al aplicar la configuración que otorga de histéresis a una red de 8 WTA y compararlo contra una red clásica, se puede observar un comportamiento interesante los resultados son mostrados en la Fig. 5.14 y Fig. 5.15.

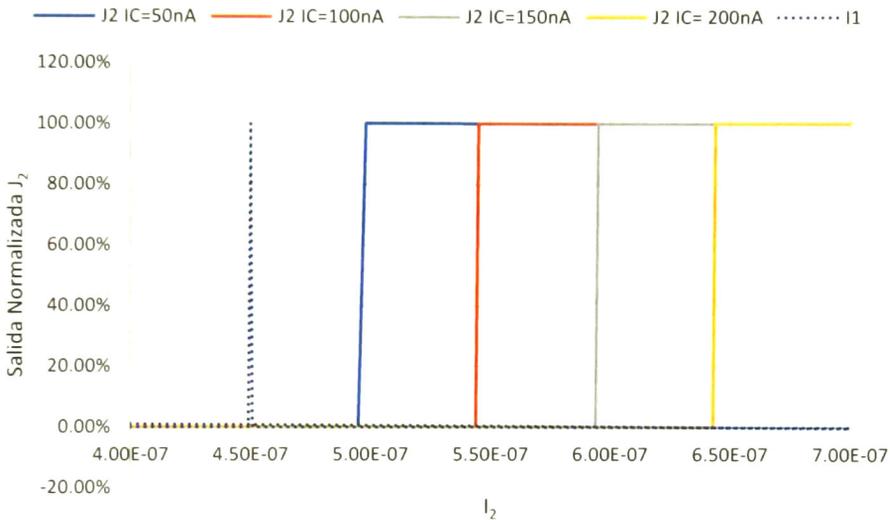


Fig. 5.13: Salida de la celda WTA histerética ante cambios en la entrada de la celda 2.

A ambas redes se les aplica la misma entrada, esta consiste en un barrido de corriente en la celda 1 desde 1000nA hasta 5nA mientras que la celda 2 mantiene un valor de 450nA y la celda 5 mantiene un valor de 800nA, en ambos casos se tomaron los datos cuando el barrido de la entrada en la celda 1 alcanzó un valor de 780, se escogió este momento en el barrido ya que se puede observar el momento de cambio de ganador, el resultado es diferente debido a que la corriente histerética está evitando que la red WTA seleccione otra salida como ganadora sin antes la entrada de esta célula sobrepase la corriente histerética, en contra parte el WTA clásico cambia la célula ganadora en el momento en que sobrepasa a sus hermanas, es de

observar que la corriente constante en la celda 2 no afecta a sus vecinos, es decir el efecto histerético está limitado a un espacio único.

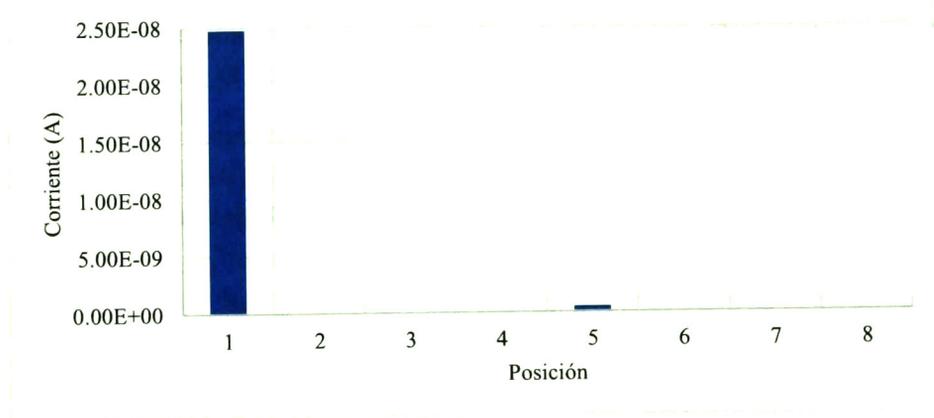


Fig. 5.14: Respuesta de un WTA histerético ante una entrada ganadora en la celda 5 y una de menor cantidad de corriente en la celda 1.

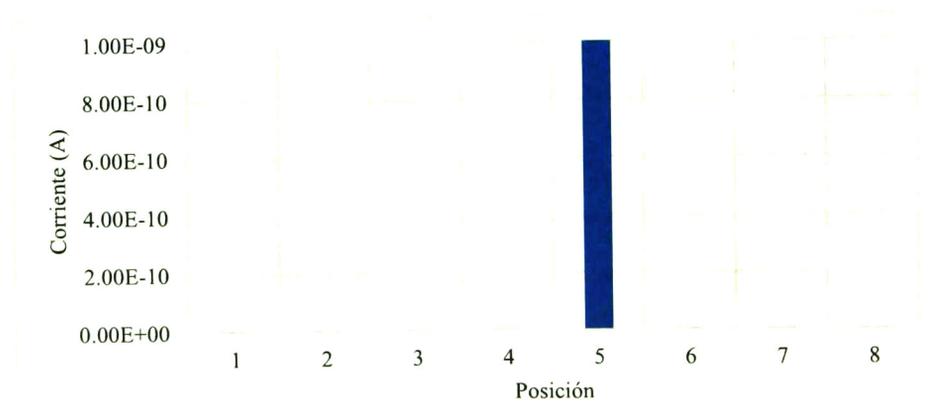


Fig. 5.15: Respuesta de un WTA clásico ante una entrada ganadora en la celda 5 y una de menor cantidad de corriente en la celda 1.

5.6 Redes resistivas.

En el capítulo 2 se presentaron diferentes formas para la detección de los elementos importantes en la escena visual, así como también de la aplicación de filtros a la escena. Tiene una particular importancia el operador G debido a la semejanza que tiene con la operación que realizan las redes neuronales biológicas para identificar bordes.

Muy a menudo este tipo de filtros es aplicado usando una red de resistencias [31] [32] [33], ya sean pasivas o también se puede implementar usando transistores MOS cuyas terminales de fuente y drenaje sean simétricas y cuyos voltajes de compuerta y sustrato se pueda ajustar para proveer control de la difusión de corriente. Estas redes de difusión son usadas extensamente en varios tipos de circuitos neuromórficos, ya sean retinas o cócleas.

Para nuestro caso se usan estas redes difusivas en diferentes puntos del circuito WTA, éstas se conocen como redes de inhibición y de excitación, la función de las redes de inhibición es desacoplar de forma espacial la célula WTA ganadora y asegurar una mejor selección de ésta, en cuanto a las redes excitadoras, éstas tienen la función de suavizar la señal de entrada, esto tiene una función especial cuando el objeto de interés está cambiando de posición de manera continua.

5.6.1 Redes excitadoras o de excitación lateral.

La adición de la histéresis agrega resistencia al cambio de ganador, proveyendo un sistema de atención que resiste la selección de un nuevo ganador a menos que el estímulo sea de un valor mucho mayor que el ganador actual. Sin embargo si el estímulo ganador se está moviendo, la entrada ganadora puede que no se mueva a un elemento cercano. Para permitir los movimientos entre celdas adyacentes a la celda ganadora mientras se mantiene la atención en [34] y [35] se propone un circuito que provee histéresis no solo al ganador actual, sino también a sus vecinos.

En la **Fig. 5.16** se muestra la configuración de un circuito WTA con histéresis distribuida, se basa en un WTA simple con la retroalimentación para permitir histéresis, agregando la red resistiva que distribuye la corriente de retroalimentación a los elementos cercanos al nodo ganador. Esto se logra insertando transistores laterales que conectan las compuertas de los espejos de corriente, así se crea una red resistiva en modo corriente, que distribuye la corriente. El voltaje de compuerta de estos transistores sirve para determinar la constante espacial de la red. Mientras el voltaje de compuerta se incrementa, fluye más corriente del ganador a los vecinos, dando a estos elementos un estado preferencial en la contienda para determinar el elemento ganador. Esta operación, además de difundir la corriente histéretica, realiza un suavizado en la entrada, entonces el circuito

tendera a favorecer a las entradas que cuenten con una actividad promedio elevada, en vez de seleccionar la entrada singular con un máximo de actividad. Esto eliminará errores que provengan de ruido y de los errores al construir dispositivos VLSI.

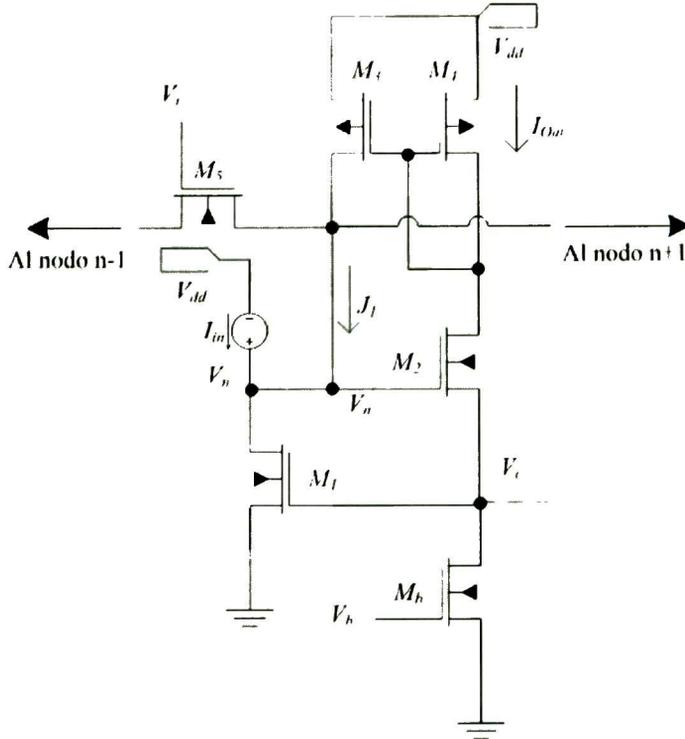


Fig. 5.16: WTA con histéresis.

Para estudiar analíticamente el principio de operación de la red de difusión excitadora olvidémonos por un momento del transistor M_5 , por simplicidad, y consideramos que la corriente de entrada está en subumbral y es además constante, aplicada solo al primer nodo de la red WTA histerética.

La corriente que fluye por los transistores de difusión puede dividirse en las corrientes de componente de *reverse* y *forward*, (5.9) y (5.10) de donde en este caso:

$$\begin{aligned} V_s &= V_i \\ V_d &= V_{i+1} \\ V_g &= V_{ex} \end{aligned} \tag{5.19}$$

De las ecuaciones de componentes de corriente e un transistor se sostiene la siguiente relación:

$$I_{f,i} = I_{r,i+1} \quad (5.20)$$

Al escribir las ecuaciones de corriente de Kirchhoff en el nodo i obtenemos:

$$I_{a,i} = (I_{f,i-1} - I_{r,i-1}) - (I_{f,i} - I_{r,i}) \quad (5.21)$$

Al usar las ecuaciones anteriores podemos expresar la ecuación anterior como:

$$I_{a,i} = 2I_{r,i} - I_{r,i-1} - I_{r,i+1} \quad (5.22)$$

pero si $I_{a,i}$ es una corriente de subumbral, la podemos escribir como

$$I_{a,i} = I_0 e^{\kappa \frac{V_c}{U_T}} \left(1 - e^{-\frac{V_i}{U_T}} \right) \quad (5.23)$$

Al expresar V_i en términos de $I_{r,i}$, podemos escribir:

$$I_{a,i} = I_0 e^{\kappa \frac{V_c}{U_T}} - e^{\kappa \left(\frac{V_c}{U_T} - \frac{V_{ex}}{U_T} \right)} I_{r,i} \quad (5.24)$$

Lo que se despeja en:

$$I_{r,i} = \lambda I_0 e^{\kappa \frac{V_c}{U_T}} - \lambda I_{a,i} \quad (5.25)$$

De donde $\lambda = e^{-\kappa \left(\frac{V_c}{U_T} - \frac{V_{ex}}{U_T} \right)}$, sustituyendo (5.25) en (5.22) tenemos la aproximación discreta de un laplaciano, usada muchas veces para el procesamiento de imágenes:

$$I_{a,i} = \lambda (I_{a,i-1} - 2I_{a,i} + I_{a,i+1}) \quad (5.26)$$

Reordenando obtenemos:

$$I_{a,i} = \frac{\lambda}{1 + 2\lambda} I_{a,i-1} + \frac{\lambda}{1 + 2\lambda} I_{a,i+1} \quad (5.27)$$

Al usar la ecuación de manera recursiva obtenemos:

$$I_{a,i} = \frac{\lambda}{1 + 2\lambda} I_{a,i-1} + \frac{\lambda^2}{(1 + 2\lambda)^2} (I_{a,i} + I_{a,i+2}) \quad (5.28)$$

Si $\lambda \ll 1$ la ecuación se reduce a:

$$I_{a,i} \approx \lambda I_{a,i-1} \quad (5.29)$$

Si queremos estimar la corriente que fluye por el transistor n de la red $I_{a,n}$, podemos usar la ecuación de manera recursiva hasta hallar la primera celda de la red.

$$I_{a,n} = I_{a,0} \lambda^n \quad (5.30)$$

Pero como $I_{a,0} = I_{in}$, si $\lambda \ll 1$ podemos escribir:

$$I_{a,n} = I_{in} e^{-n\kappa \left(\frac{V_c}{U_T} - \frac{V_{ex}}{U_T} \right)} \quad (5.31)$$

El término λ es definido como la constante espacial de la red, esta es modulada exponencialmente por el termino $-(V_c - V_{ex})$. Mientras V_{ex} es un parámetro del circuito directamente accesible, independiente de otros parámetros del circuito, el voltaje V_c depende logarítmicamente de la corriente de entrada.

$$I_{a,0} = I_0 e^{\kappa \frac{V_{ex}}{U_T}} \approx I_{in} \quad (5.32)$$

Con esta relación en mente, se puede reescribir λ como una función de V_{ex} y I_{in} así la ecuación se reduce a:

$$I_{a,n} = I_{in} \left(\frac{I_0 e^{\frac{\kappa V_{ex}}{U_T}}}{I_{in}} \right)^n \quad (5.33)$$

De acuerdo a esta ecuación, un incremento en V_{ex} incrementará exponencialmente la cantidad de esparcimiento de la corriente que pasa por el difusor. Consecutivamente un incremento en la amplitud de I_{in} disminuirá el esparcimiento de la corriente que pasa por la red de difusión.

Para casos realísticos esta estructura se comporta como un filtro no lineal que mejora la entrada con la mayor amplitud y suaviza la respuesta. La red de transistores implementada por los transistores M_1 y M_2 realiza la operación de suavizado en la entrada mientras el WTA agrega la corriente de histéresis retroalimentando la entrada con la máxima intensidad.

5.6.1.1 Simulación de redes excitadoras.

Para estimar el probable funcionamiento de una red WTA en conjunto con una red excitadora. Las principales características que se buscan corroborar son, la histéresis local, la cual permite el cambio de célula ganadora con preferencia en la vecindad. Otro punto importante es la dependencia de V_{ex} e I_{in} en la dispersión del efecto excitador de la red. Y por último corroborar la respuesta logarítmica ante estos cambios en V_{ex} e I_{in} .

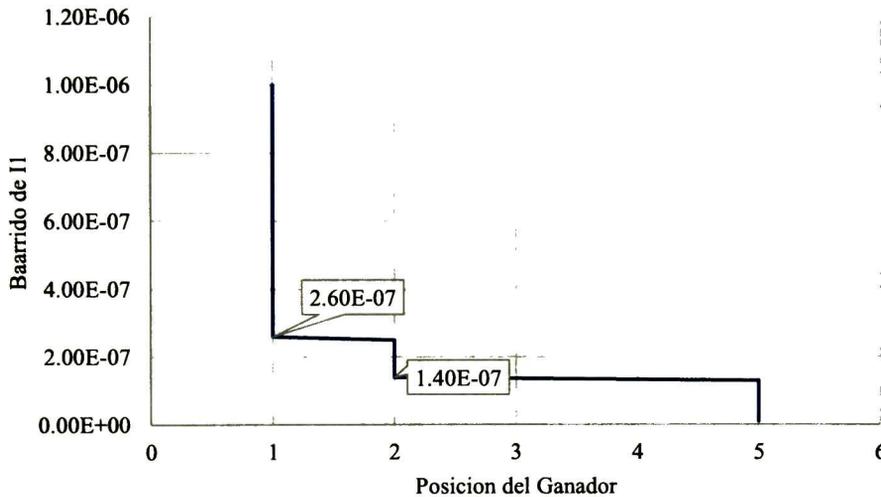


Fig. 5.17: Respuesta de una red WTA con red excitadora.

Para corroborar la acción histerética local, se hace un barrido de I_1 de 1000nA a 100nA en un arreglo de 8 celdas WTA, en donde las celdas 2 y 5 mantienen un valor constante de 450nA y 800nA respectivamente, la gráfica nos muestra la tendencia del ganador, al principio la celda 1 está ganando, este estado cambia cuando la corriente de la celda 1 es menor a 450nA, en 260 nA, la celda 2 se convierte en la ganadora y cuando la corriente decae a un valor de 140 nA la celda 5 gana, ésta es la habilidad histerética localizada en las cercanías de la celda 1.

En el caso de no estar activa la red de difusión excitadora el WTA con retroalimentación, con o sin retroalimentación histerética, dadas las mismas condiciones en todas las entradas y usando el mismo barrido de corriente en I_1 , se observa que al encontrarse con un valor de 756 nA el estado ganador cambia a la celda 5, obviando a la celda 2.

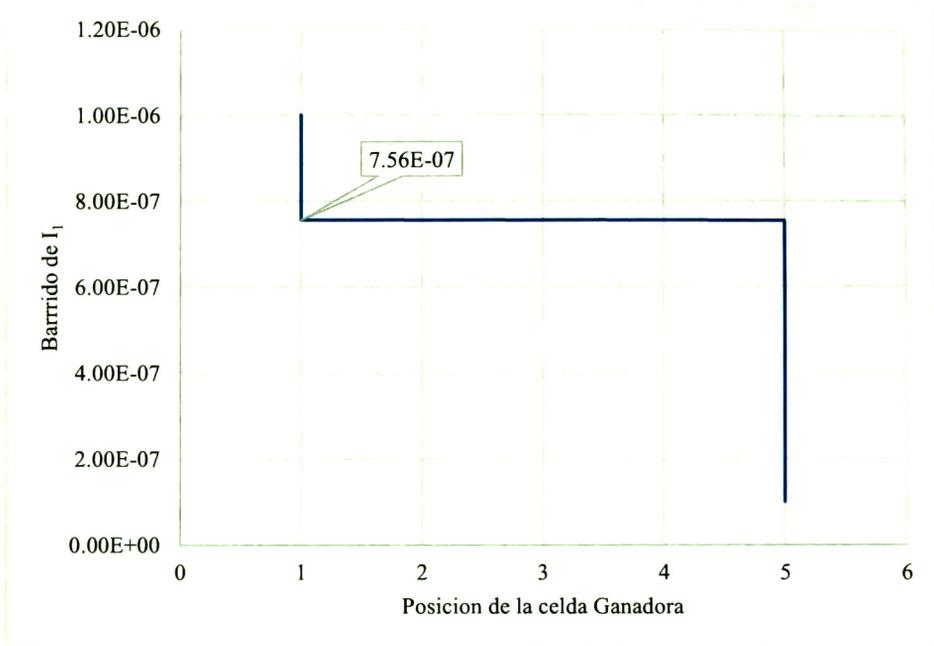


Fig. 5.18: WTA sin red de excitación lateral.

A variar el voltaje de V_{ex} la constante de excitación lateral se ve afectada, como se puede observar en la respuesta de la figura siguiente, en esta figura se puede observar que al aumentar el voltaje V_{ex} la diferencia entre la célula ganadora se va disminuyendo, las corrientes tomadas son las que deja pasar el transistor M_I sin tomar en cuenta la corriente de histéresis.

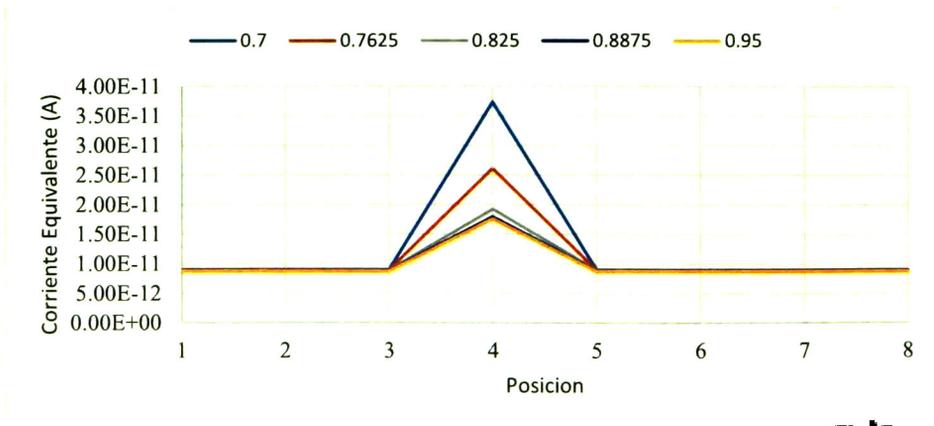


Fig. 5.19: Respuesta ante diferentes voltajes de excitación.

La diferencia entre tener y no tener la red de excitación es notable, este tipo de operaciones ayudan a aislar al objeto ganador en el espacio y elimina errores en la selección.

5.6.2 Redes inhibitorias.

La red de inhibición local, contiene la influencia de la celda ganadora en un rango espacial.

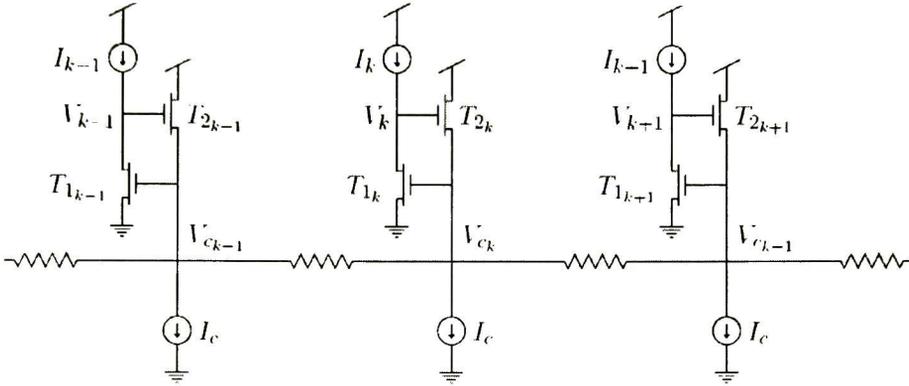


Fig. 5.20: WTA con red inhibitoria.

El circuito de la Fig. 5.20 realiza la operación de WTA con inhibición, es idéntico al WTA original excepto que cada neurona se conecta con su vecino más cercano con una resistencia no lineal. Cada resistor conduce una corriente I_r en respuesta a una diferencia de voltaje V , donde:

$$I_r = I_s \tanh(\Delta V / 2V_0) \quad (5.34)$$

I_s , la corriente de saturación del transistor, es un parámetro controlable. La corriente de polarización I_c es distribuida entre los resistores en el circuito WTA.

Para entender la operación, se considera la respuesta del circuito ante un impulso espacial, definido como $I_k \gg I_c$,

Cada resistor drena una corriente I_s cuando está saturado, el transistor T_{2k} entonces conduce $2I_s + I_c$.

Una versión ligeramente modificada de tal circuito se ilustra en la Fig. 5.21, donde los resistores mencionados son substituidos con

transistores NMOS que operan en saturación en el subumbral (mostrados como M1c y M2c), para alcanzar gran densidad de área y más baja disipación térmica. En este circuito los voltajes de compuerta de M1c y M2c están ambos controlados el voltaje de inhibición, V_{inh} . Empleando la ecuación (2.8), se puede escribir la expresión de cada corriente de drenador como sigue:

$$\begin{aligned} I_{d1c} &\cong I_0 e^{\frac{\kappa V_{inh} - V_{C1}}{U_T}} \\ I_{d2c} &\cong I_0 e^{\frac{\kappa V_{inh} - V_{C2}}{U_T}} \end{aligned} \quad (5.35)$$

La red inhibidora es equivalente en todos los aspectos a la red excitadora. Se puede mostrar que usando la misma metodología para analizar la red excitadora, la constante espacial de la red inhibidora depende de los voltajes aplicados en sus compuertas y de la corriente de entrada. Intuitivamente se puede ver cómo el voltaje de compuerta de los transistores que forman la red inhibidora permite modular el grado en el que las células del WTA compiten.

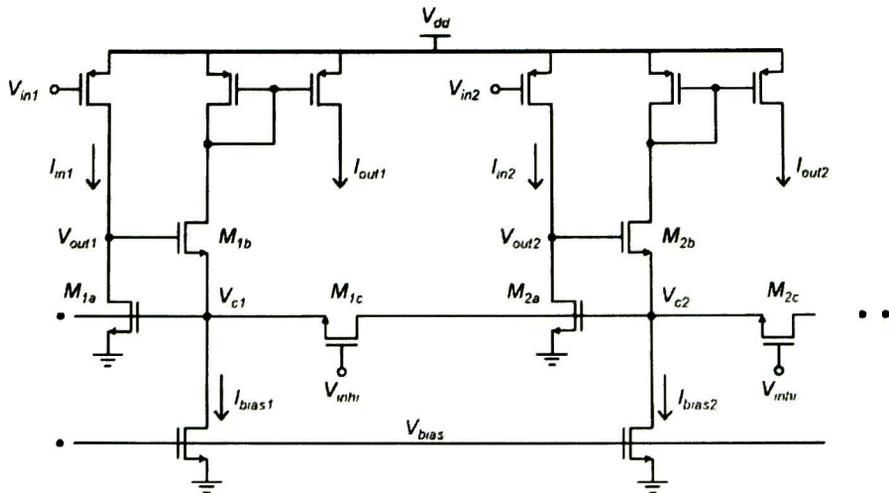


Fig. 5.21: WTA con red de inhibición.

Es evidente la dependencia que tiene la corriente que fluye a través de M1c (M2c), ésta depende exponencialmente de V_{inh} . Entre más elevado sea el voltaje de V_{inh} , más corriente fluirá hacia el nodo ganador VC a través de los transistores locales de desacoplamiento. Si V_{inh} alcanza el más alto potencial posible en el chip (V_{dd}), entonces el drenador y la fuente de M1c

estarán literalmente cortocircuitadas, resultando en un grado cero de desacoplamiento inhibitorio local. Por otro lado, si el voltaje Vinhi decrece suficientemente tal que M1c se apague efectivamente, entonces se alcanza el máximo grado de desacoplamiento y como resultado, cualquier celda puede ser seleccionado como la ganadora, independiente de sus vecinos. En una palabra, ajustando el valor de Vinhi, se puede controlar la cantidad de corriente Inhibidora que fluye lateralmente, que en su turno determina cuántas celdas vecinas se le permiten suprimir (inhibir) a una celda ganadora local.

La idea del circuito mencionado se originó de una configuración llamada, divisor de corriente de pseudos-conductancia (difusor) [36], que básicamente explota la similitud entre un transistor MOS y un transistor bipolar. La corriente que fluye en un transistor MOS en el subumbral, igual de la corriente en un transistor bipolar, puede ser definida entre componentes directas (*forward*) e inversas (*reverse*); esto es, $I_d = I_f - I_r$. Con esto en mente, revisaremos el transistor de desacoplamiento M1c, obteniendo lo siguiente:

$$I_{dic} = I_f - I_r = I_0 e^{\frac{\kappa V_{inh} - V_{C1}}{U_T}} \left(1 - e^{-\frac{V_{C1} - V_{C2}}{U_T}} \right) \quad (5.36)$$

Asumiendo que el valor de $(V_{C1} - V_{C2})$ es suficientemente pequeño comparado con V_T , se puede escribir la anterior expresión de I_{dic} como en la Ecuación (5.35).

Adicionalmente asumiendo que el transistor de la fuente de corriente de la celda 1 ganadora está operando en el subumbral, se encuentra que:

$$\frac{I_{dic}}{I_{bias}} \cong e^{\frac{\kappa V_{inh} - V_{bias}}{U_T}} \quad (5.37)$$

Entre mayor sea esta relación, el desacople será menos efectivo.

5.6.2.1 Simulación de redes inhibitoras.

Para la simulación de las redes inhibitoras se usó el diagrama de simulación mostrado en la Fig. 5.22, es una red consistente en 8 WTA en la Fig. 5.23 muestra el cálculo de la red inhibitor, en esta se aplica una señal ganadora de entrada en la cuarta célula suprimiendo la salida de sus vecinas,

las neuronas más alejadas tienen una respuesta significativa al codificar sus entradas.

En la Fig. 5.24 se muestran los resultados con diferentes relaciones entre las corrientes de polarización y la corriente de entrada.

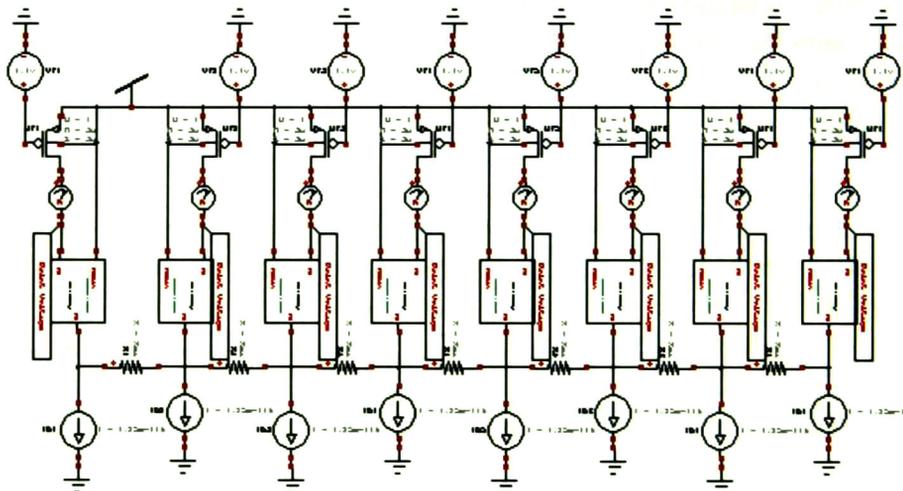


Fig. 5.22: Diagrama de simulación de una red de WTA con inhibición.

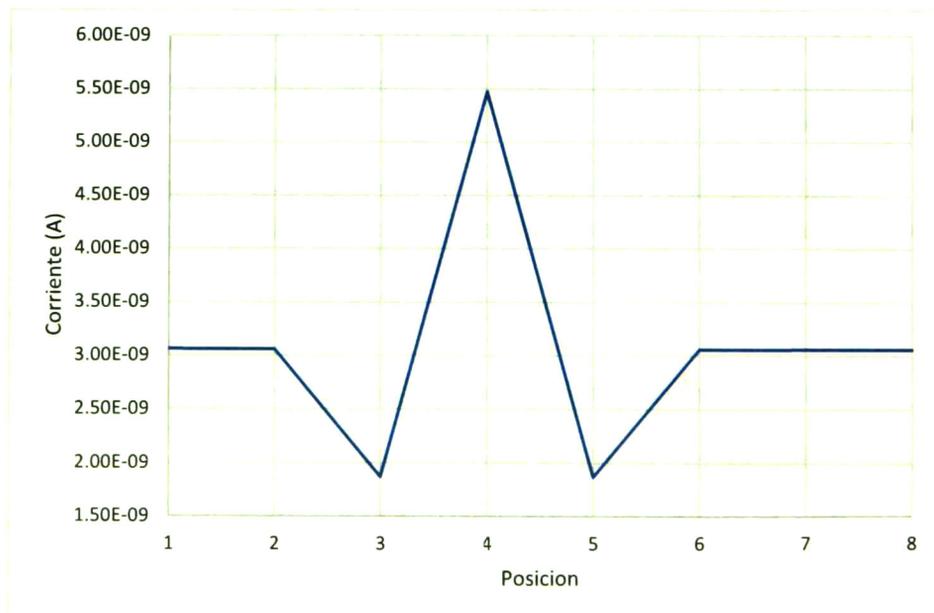


Fig. 5.23: Resultado ganador de una red WTA con inhibición.

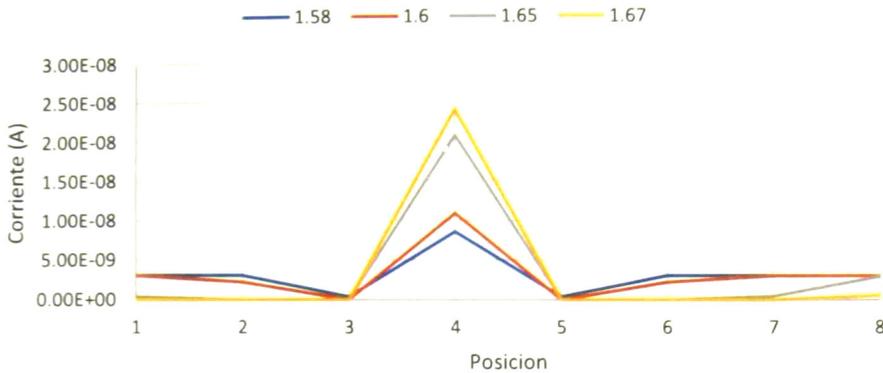


Fig. 5.24: Resultado ganador de una red WTA inhibidor en diferentes voltajes en los transistores de inhibición.

5.7 WTA histerético completo.

El circuito WTA que propuso Lázaro es uno muy elegante, es un circuito pensado en la optimización de espacio y en el bajo consumo de energía que procesa de forma paralela la información en tiempo real usando solo dos transistores por cada nodo de salida. En las secciones anteriores de este capítulo se han mostrado algunas extensiones al circuito básico que lo dotan de nuevas funciones, al combinar todas estas funciones, histéresis, redes de excitación e inhibición, el circuito WTA produce una respuesta espacial centralizada, propiedad que difiere de las obtenidas con otros tipos de redes de difusión.

Este tipo de respuesta ya se ha introducido en el capítulo 2, es una operación ideal para la detección de cambios de intensidad en sensores de visión, además de asemejarse al tipo de respuesta de muchas neuronas que se encuentran en el córtex visual de los mamíferos. Esto hace de este circuito particularmente útil para todo tipo de aplicaciones visuales.

La Fig. 5.25 muestra el resultado de una red WTA histerético con acoplamiento inhibitorio y excitador, la respuesta es muy semejante al operador G'' , esta característica es heredada de la red inhibitoria y gracias a las características de la red histerética y excitadora, el estado ganador es más resistente a errores en la visión.

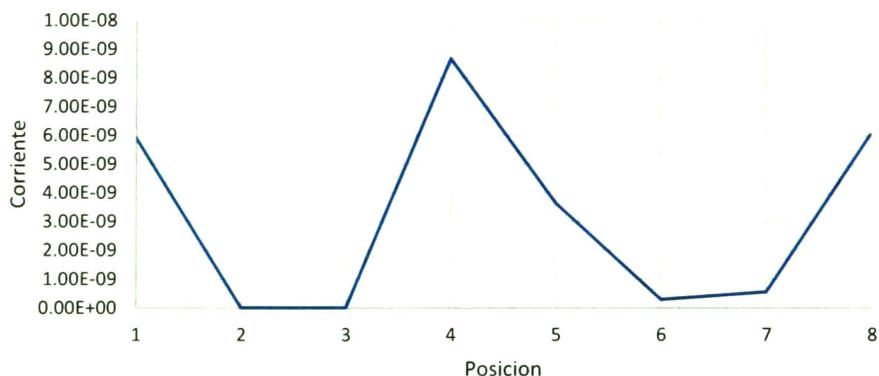


Fig. 5.25: Respuesta ganadora de un WTA histerético con red de excitación e inhibición.

En las siguientes figuras se meseta un comparativo con la red WTA clásica y diferentes configuraciones del WTA histerético.

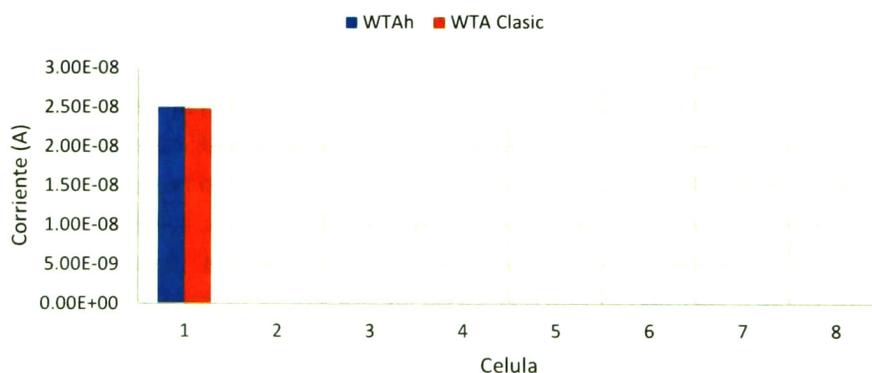


Fig. 5.26: Comparación entre WTA clásico y WTAh, características inhibitoras y excitadoras desactivadas.

En la **Fig. 5.26** se compara el funcionamiento del circuito WTA clásico e histerético, en el caso del circuito histerético los voltajes para las redes de inhibición y excitación son de V_{dd} , se aplican corrientes en las entradas 1, 5 y 6, siendo la mayor la entrada 1, en ambos casos los circuitos seleccionan la entrada 1 como la ganadora.

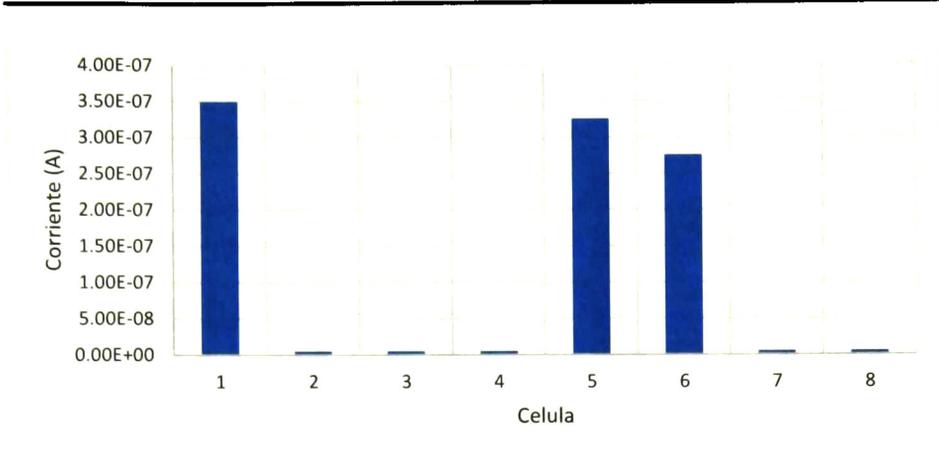


Fig. 5.27: Entradas para ambos tipos de WTA.

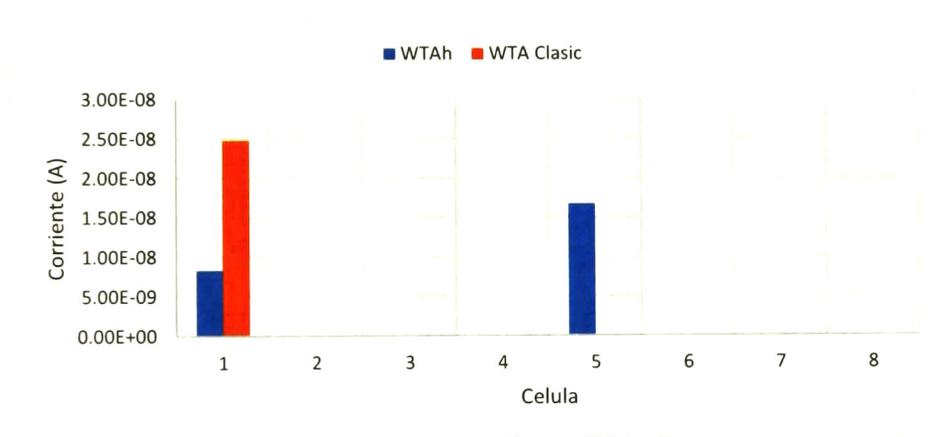


Fig. 5.28: Comparación entre WTA clásico y WTAh, características inhibitoras y excitadoras activadas.

En la **Fig. 5.28** se activa la excitación lateral, el WTA histerético cambia su selección a la entrada 5 ya que en la vecindad de esta entrada el efecto de activación es mayor que el generado por la corriente de la entrada 1.

En la **Fig. 5.29** se aplican corrientes en las entradas 2, 4 y 7, la inhibición es local en el circuito WTAh, al ser así el circuito permite la selección de varios elementos dependiendo de la separación entre estos. El circuito WTA clásico solo permite un ganador.

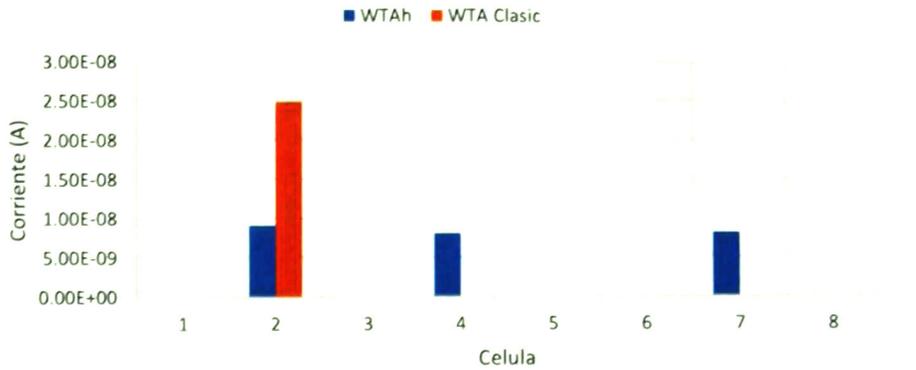


Fig. 5.29: Comparación entre WTA clásico y WTAh con entradas espaciadas.

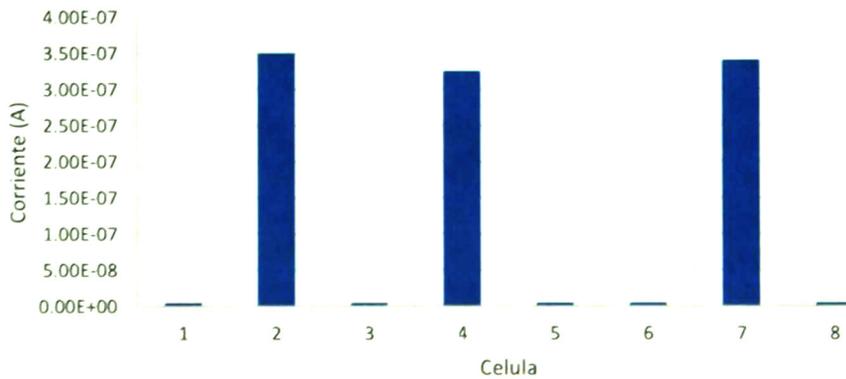


Fig. 5.30: Entradas espaciadas para la comparación entre WTA clásico y WTAh.

5.8 Simulaciones de un Chip de visión con atención.

A continuación se presenta la simulación de un chip de visión desde la etapa de captura de iluminación hasta la selección por parte del WTA.

Se utilizan los datos ya obtenidos en los capítulos anteriores para generar la base de la simulación, la cama de prueba es la misma que la vista en el capítulo 2 y 3, el modelo de prueba es más extendido.

En la Fig. 3.19 se muestra el pulso de entrada al arreglo de pixeles después de hacer todo el procesamiento de la información ya antes discutida en capítulos anteriores se obtienen los siguientes resultados.

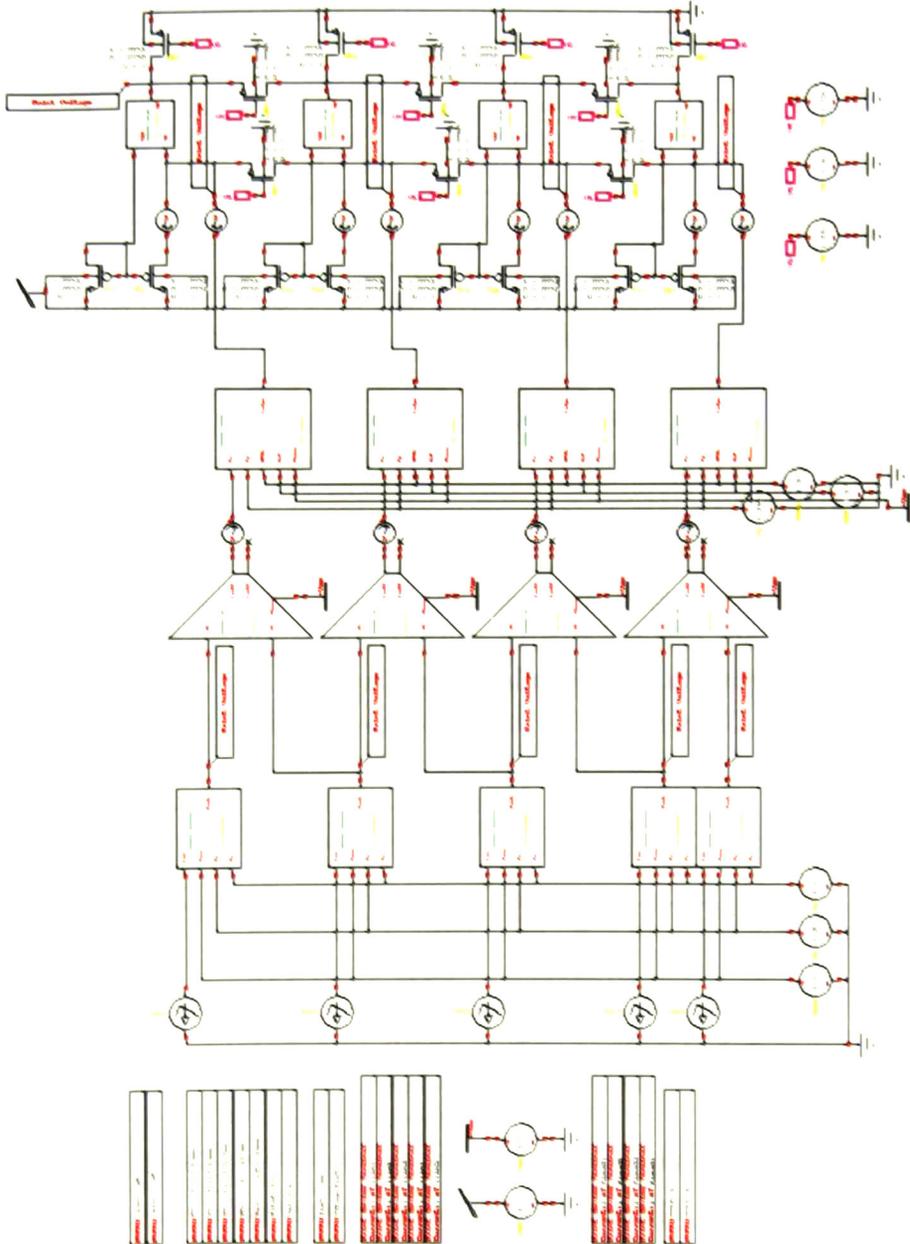


Fig. 5.31: Diagrama de simulación para un chip completo de visión.

Para la simulación se usaron un voltaje Vdd de 3V, los voltajes en los transistores de inhibición son de 1.90 V los voltajes de los transistores de excitación son de 0.965 V y el voltaje de los transistores de polarización son de 0.66 V y para el correcto funcionamiento del amplificador operacional se usa un voltaje de 2.0V.

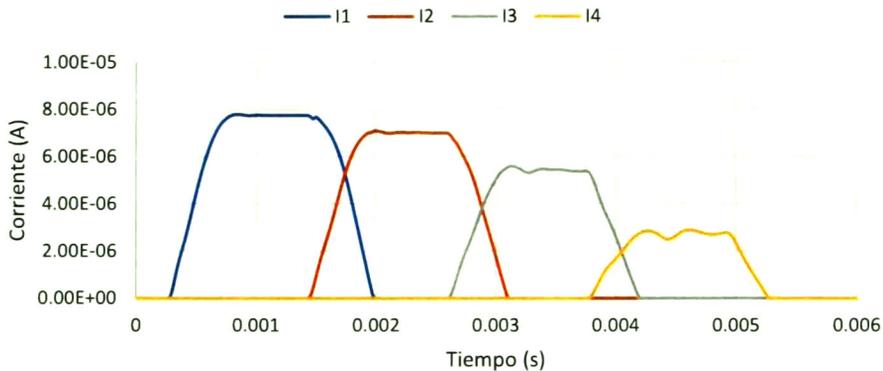


Fig. 5.32: Entradas de la etapa WTAh del circuito de visión.

La figura muestra la entrada de cada celda de la etapa de WTA en voltaje, esta nos dice cómo va cambiando el contraste en la escena, cada elemento representa una derivada de 2 píxeles.

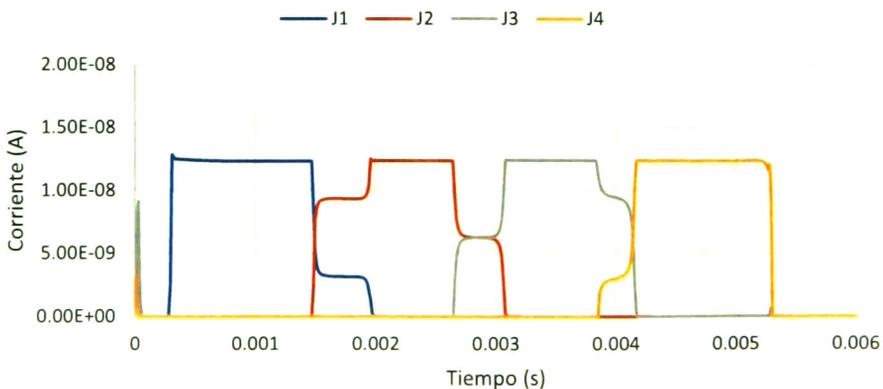


Fig. 5.33: Salidas de cada elemento WTAh.

La **Fig. 5.26** muestra la salida en corriente de cada célula de WTA, en ella podemos notar el movimiento del borde de la barra pasando por la escena, en la figura se aprecia la primera célula ganadora es la que primero

entra en contacto con el borde de la barra oscura, y esta se va moviendo de célula en célula, calculando donde se encuentra el objeto más prominente de la escena visual.

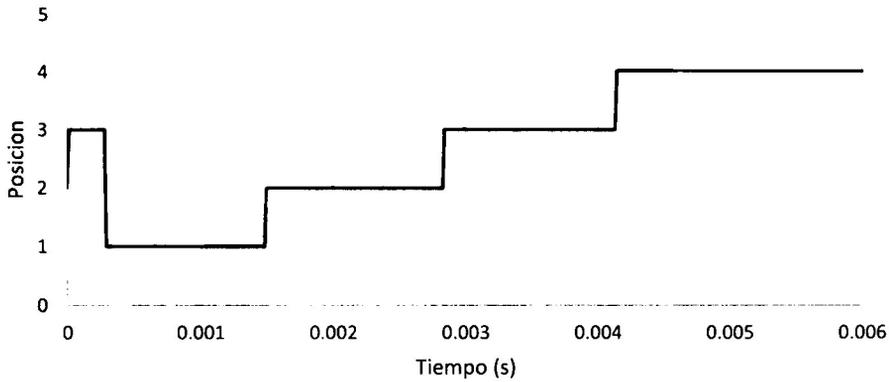


Fig. 5.34: Representación de la ubicación de la celda ganadora.

A continuación se aplica la misma señal más unos pulsos que simulan destellos en la escena visual, posibles elementos que pueden aparecer, en este caso son pequeños pulsos que pasan mientras una barra oscura pasa sobre el sensor, la **Fig. 5.35** muestra la entrada.

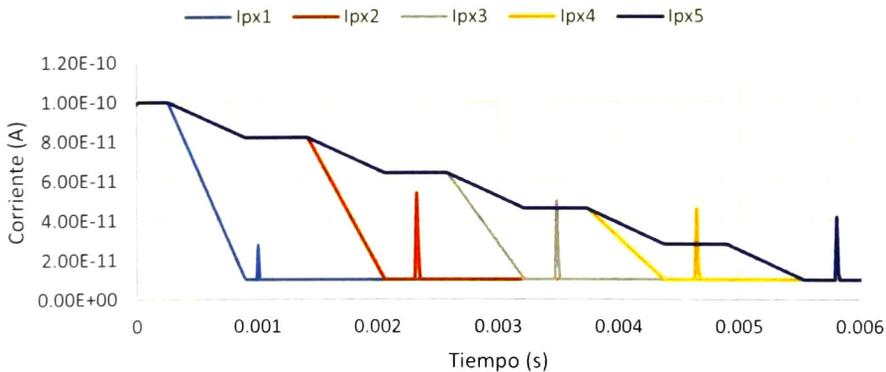


Fig. 5.35: Entrada de corriente que representa una escena visual inusual.

La entrada del WTA se muestra en la **Fig. 5.33** en ésta se puede observar que el efecto de los pulsos de luz se vuelven más influyentes conforme el promedio de la intensidad de la escena visual va disminuyendo,

esto porque en el lugar donde ocurre un pulso muy elevado significa un cambio de contraste grande.

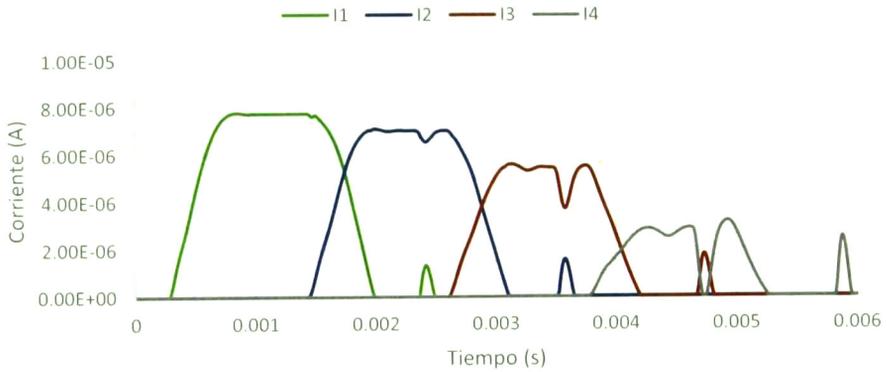


Fig. 5.36: Entradas de corriente para la etapa WTAh, con entrada mostrada en la Fig. 5.32.

La respuesta del WTA ante los pulsos se muestra en la Fig. 5.37, aquí se observan los cambios de corriente en la salida de cada célula WTA, en la Fig. 5.38 se resume en donde se encuentra la posición del ganador.

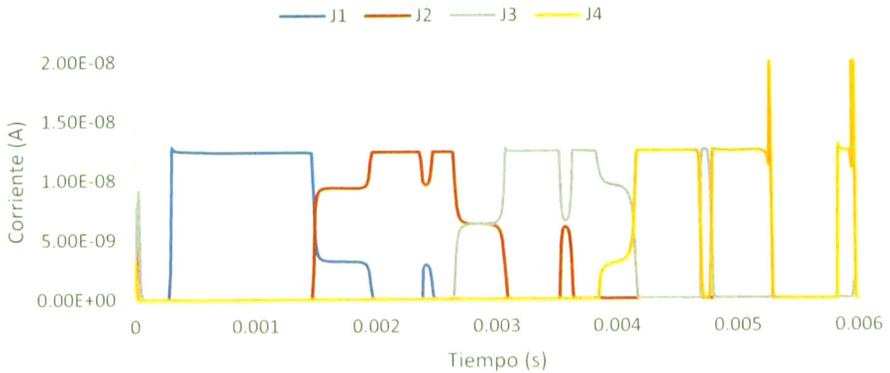


Fig. 5.37: Salida de la etapa WTA.

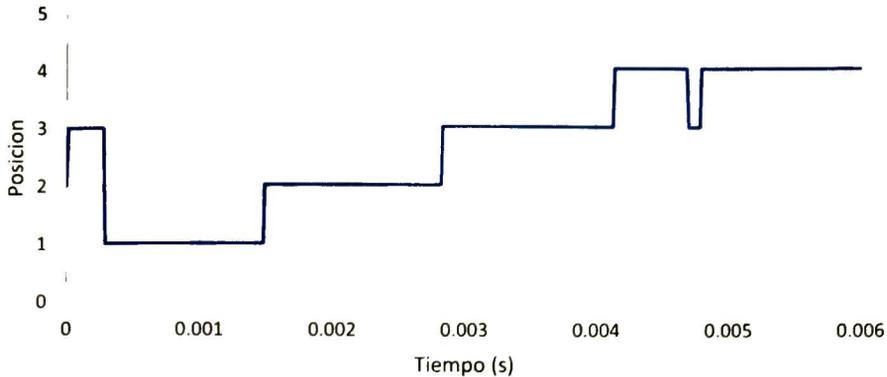


Fig. 5.38: Representación de la ubicación de la celda ganadora, con entradas mostradas en la Fig. 5.32.

5.9 Descripción sobre el circuito capaz de determinar la posición de manera analógica con los datos del WTA.

La tarea de este circuito está casi resuelta, lo que falta es convertir la señal codificada por el WTA en señales motoras que muevan el vehículo que sostiene al sensor, para que la trayectoria objetivo este centrada en el campo visual del sensor.

Para hacer esta tarea se necesita un circuito que sea capaz de transformar la información de un arreglo de entradas unidimensionales, en una salida que representa una combinación codificada de estas entradas. Un circuito que realiza esta tarea es el propuesto por DeWeert y P. Stephen, el circuito que proponen es capaz de convertir una imagen en una salida codificada en pulsos que pueden ser usados para el control motores bidireccionales que facilitan a la estructura moverse para centrar un estímulo en el campo sensorial de esta misma.

La tarea básica a resolver que se empezó a discutir en esta tesis es la de seguir un estímulo visual por medio de un sistema motor, esto es logrado al agregar un circuito de agregación.

El circuito de agregación computa una combinación de señales de entrada, resultando en un par de señales complementarias, reduciendo la representación sensorial en una representación escalar motora.

El circuito consta de un arreglo N+1 de circuitos de tipo par diferencial, configurados como se ve en la **Fig. 5.39**. Cada elemento tiene una entrada de corriente y una entrada diferencial de voltaje. Las salidas son sumadas por medio de un par de cables, resultando en un par de corrientes agregadas que son I_1 e I_2 . Donde:

$$I_1 - I_2 = \sum_n i_n \tanh\left(\frac{\Delta V_n}{2U_T}\right) \quad (5.38)$$

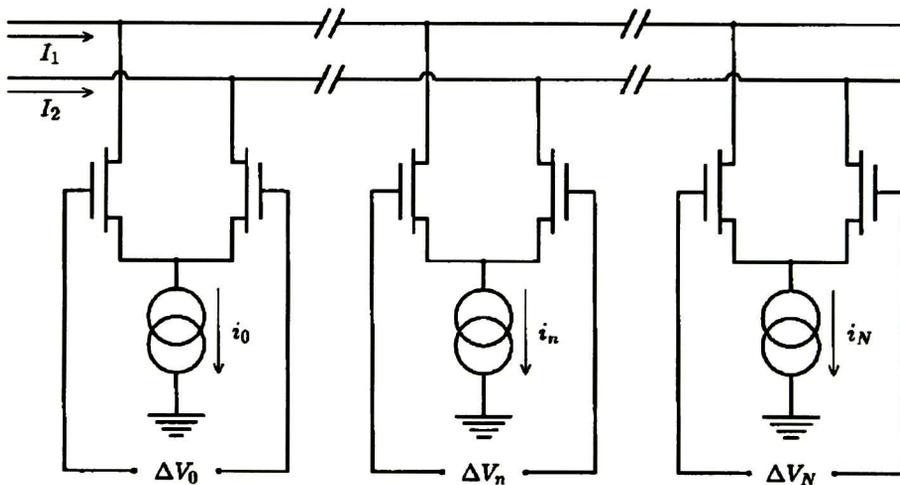


Fig. 5.39: Red de agregación.

En este circuito los voltajes diferenciales son independientes el uno del otro, sin embargo esto se puede usar para lograr codificar la posición de cada elemento de arreglo. Esto se logra mediante una serie de resistores N+1 usados como un divisor de voltaje, como se ve en la **Fig. 5.40**.

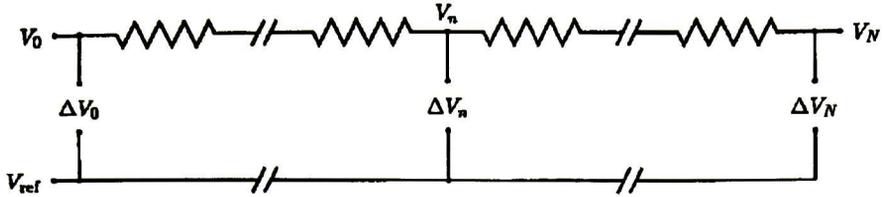


Fig. 5.40: Red de resistores como divisor de voltaje.

Al aplicar voltajes V_0 y V_N en los finales de la línea podemos crear un gradiente de voltaje por el arreglo de resistirse, donde el voltaje del nodo n está dado por:

$$V_n = \left(\frac{V_N - V_0}{N} \right) n + V_0 \quad (5.39)$$

La posición del elemento n está codificada por el voltaje diferencial $V_n - V_{ref}$, donde V_n está dado por el divisor y V_{ref} es una constante que pone el origen del arreglo, donde la corriente está distribuida igualmente entre las dos salidas de corriente.

Un modo de controlar el movimiento del sistema, es con motores DC modulado, por un tren de pulsos. Un tren de pulsos de doble riel se usa para impulsar un motor DC, una forma de generar un tren de pulsos es la propuesta por Mead, quien describe un circuito que convierte una corriente analógica en un tren de pulsos, llamado circuito neurona toma su nombre de las células biológicas que producen pulsos en el sistema nervioso de los animales.

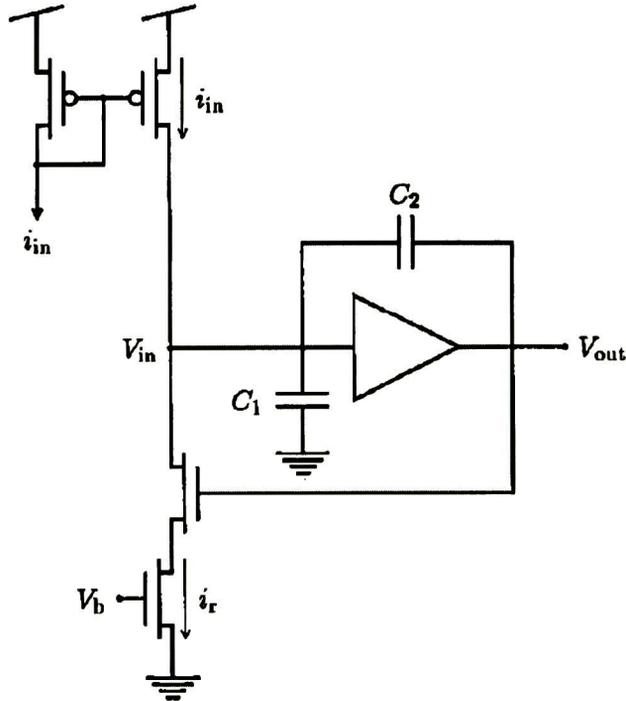


Fig. 5.41: Circuito neurona, generador de pulsos.

El circuito consiste de un espejo de corriente, un amplificador, dos capacitores y dos series de transistores. La corriente carga el capacitor de entrada, causa que el voltaje de entrada se eleve, cuando este voltaje llega el voltaje de activación del amplificador, la salida se activa causando una corriente de reset que descarga el nodo de entrada hasta que se desactive el amplificador, el capacitor de retroalimentación causa histéresis al acoplar el nodo de entrada con el de salida. Si las corriente de salida y entrada son constantes se puede mantener el un tren de pulsos en la salida.

Un rasgo importante para mover un motor por medio de un tren de pulsos, es el ciclo de trabajo, en el circuito esta característica es proporcional a la corriente de entrada, sin embargo se puede usar la corriente de reset como una entrada más, así el ciclo de trabajo es una proporción de estas dos corrientes.

Al desarrollar este circuito, que convierte una representación sensorial en una representación motora, éste debe permitir la adición de

todo número de señales de entrada y debe de producir un pulso de salida de doble riel. Ya se han presentado los circuitos que son de importancia en el momento de hacer esta conversión, resultando en un sistema completo construido con tecnología VLSI.

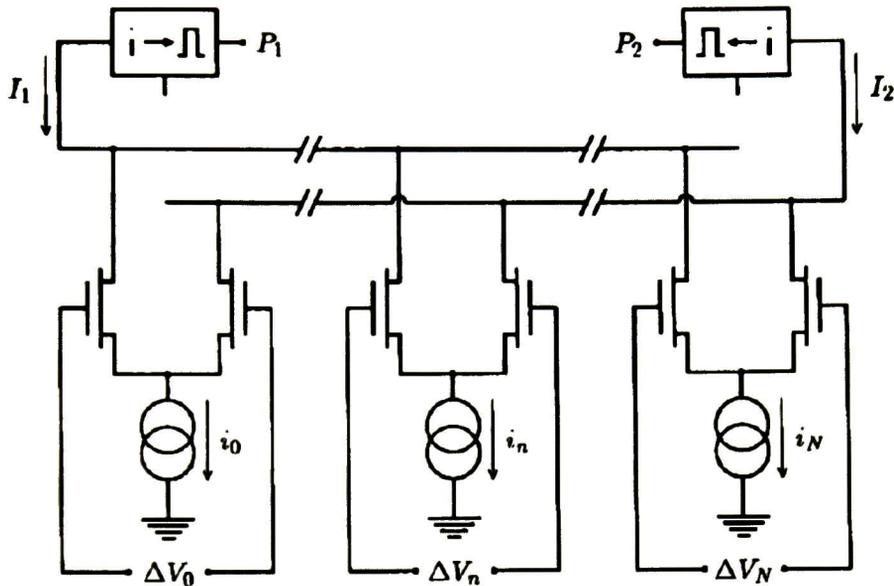


Fig. 5.42: Circuito generador de pulsos de doble riel con el circuito de agregación.

Las entradas de cada par diferencial son agregadas en un par de cables y estos son usados como entradas a los dos circuitos neuronales para la generación de pulsos que controlan a los motores, el ciclo de trabajo inherente a los pulsos es definido a razón de $(\sum i_n)/i_r$

5.10 Conclusiones.

En este capítulo se habló sobre cómo funciona un circuito WTA, desde el principio de operación, hasta su uso en aplicaciones de identificación de características importantes en imágenes y la similitud que se tiene entre la operación realizada por el circuito y la operación que se realiza en los sistemas biológicos. Esta comparación es posible gracias a las redes de excitación y de inhibición, configurados de diferentes formas se extiende el funcionamiento del WTA, se puede restringir el funcionamiento del circuito

en un rango espacial, dependiendo de la corriente de polarización se puede obviar ruido visual en la escena.

Como circuito de visión completo, ofrece un funcionamiento muy estable hasta que la iluminación total de la escena visual es bajo, en ese momento es esperado que brillos intensos en la escena visual causen que el circuito pierda por un momento el objetivo deseado.

El WTA es una herramienta poderosa con distintas aplicaciones sensoriales, esto es posible gracias a la operación G'' , además del seguimiento de objetos en el espacio visual, se puede aplicar a todo problema que necesite una operación de atención, en especial circuitos neuromórficos, como cóclea. Una aplicación que se puede explotar por sistemas de procesamiento digital es usar el WTA para concentrar el poder de cómputo en un área más pequeña así se podría optimizar tiempo de computo.

Capítulo 6

Conclusiones

En este capítulo se presenta una revisión de la tesis, en este capítulo se reafirma el objetivo de la tesis, se discuten los problemas que se encontraron y se establece un trabajo futuro.

6.1 Introducción.

La principal motivación de este trabajo de tesis: "Análisis y simulación de un chip de visión neuromórfico para la detección del movimiento" fue la de demostrar cómo un sistema neuromórfico específico, un chip de visión, puede ser integrado en una tecnología VLSI, para emular los principios elementales de un sistema de visión biológico, con tres atributos esenciales: el bajo consumo de potencia, la computación en paralelo de tipo analógico, y la ausencia de algoritmos de computación, otros que la topología de los bloques de circuitos analógicos utilizados en una configuración jerárquica.

Mientras que el análisis y simulación se refiere a una propuesta concreta. los conceptos, métodos y atributos, siguen siendo de actualidad, para entender los progresos, importancia y limitaciones de las propuestas modernas sobre la computación analógica, una guía de desarrollo sobre sistemas analógicos configurables y programables, FPAA, que predice nuevos métodos de computación basados en la computación analógica en el régimen de inversión débil, auxiliados por métodos computacionales digitales, basados en el empleo de elementos de memoria con dispositivos MOS de compuerta flotante.

El análisis de los diferentes bloques fue realizado basado en las referencias apropiadas consultadas, la simulación fue realizada usando la plataforma de herramientas de simulación en Spice, de Tanner, desarrollando la infraestructura necesaria para la simulación de dispositivos opto-electrónicos, para las señales de entrada.

6.2 Descubrimientos empíricos.

Durante la realización de esta tesis se fueron resolviendo problemas concretos de entendimiento y operación de los circuitos neuromórficos también de su aplicación ante tareas de visión.

Basados en el trabajo publicado por Indiveri et al. [5], donde se describe el funcionamiento de un chip de visión neuromórfico. Empezando con el flujo de información que se tiene que recorrer en orden de procesar una escena visual, hasta su aplicación en circuitos analógicos con el objetivo de generar señales para elementos motrices que sean capaces de realizar una operación simple de seguimiento de trayectoria, estos procesos son descritos de manera más detallada en cada capítulo.

Lo primero que se hizo fue definir una tarea específica a la cual un chip de visión pueda aplicarse, esta fue la simple tarea de proporcionar un sistema de guía a un vehículo que se mueve sobre una línea negra pintada sobre una superficie plana de color blanco.

El primer problema que se encontró fue entender cómo la naturaleza ataca esta tarea, la cual se resume en la aplicación de una operación computacional conocida como atención visual, una forma de solucionar la atención visual es mediante la aplicación de *saliency maps*, esta solución

tiene la ventaja de ser modular lo que hace que su transportación a un circuito VLSI sea fácil.

Seleccionar qué tipo de sensores usar y qué tipos de operaciones tienen que estar en los bloques es importante en esta etapa, para esto hay que entender qué es lo que se está viendo y cuál es la limitación física que se tiene. En primera instancia se tiene un objeto que se está moviendo siempre, por lo que un sensor que este enfocado a detectar cambios es más practico que uno dedicado a estar siempre detectando la escena, esto está apoyado por las células biológicas dedicadas al sensado de la luz [8]. En segunda se tiene una limitación física ya que se piensa en el uso de un sensor de una dimensión para la detección de la línea recta, para esta tarea es más que suficiente el uso de una dimensión el único inconveniente que se tiene es la orientación del sensor, este tiene que estar perpendicular al movimiento.

Debido a que en general varios programas d simulación no cuentan con una forma de aplicar entradas de iluminación, se tiene que investigar cómo los dispositivos semiconductores similares se comportan ante iluminación, proponer una respuesta que se acomode a la tarea demandada y simularla, la cual consiste en la aplicación de una fuente de corriente que genere pulsos similares a los esperados por los foto sensores.

Cada bloque del sistema tiene la tarea de aplicar operaciones de filtrado específicas para la detección de movimiento, éstas generan un mapa ya pre procesado de la escena visual, este pasa por un WTA quien selecciona la característica más relevante.

El bloque WTA, bloque importante en la operación de atención visual, este elemento es capaz de identificar y seguir en el espacio la característica más sobresaliente de sus entradas.

El circuito es uno que trabaja en modo corriente, trabaja con corrientes de subumbral, está construido con el propósito de ser eficiente en tamaño, por lo que éste es mínimo y la cantidad de corriente con la que trabajan está limitado por la tecnología implementada. Está hecho con un mínimo de 3 transistores considerando un transistor de polarización, es un circuito muy elegante, basado en un transportador de corriente, su respuesta es sensible al tamaño de los transistores, para un buen funcionamiento se recomienda usar transistores de mayor tamaño y si se desea aumentar la “sensibilidad” es decir qué tanto adquiere un estado de ganador ante

diferencias en la entrada, se recomienda agregar un transistor en degeneración en fuente.

El funcionamiento de este dispositivo está basado en la función máxima, donde solo el resultado es el elemento con mayor valor. Es usado en un sin número de aplicaciones y especialmente en circuitos neuromórficos y aplicaciones de atención, es un bloque principal en los mapas de relevancia. La función se resume en solo activar una celda donde se encuentre la mayor entrada.

Al agregar redes de distribución resistivas se dota al sistema de características similares a la operación sombrero de charro, dotando al sistema de una mejor identificación de objetos y, dependiendo de cómo se configuren las redes, se puede desacoplar especialmente un objeto de interés en el espacio o que se seleccionen varios ganadores distanciados en el espacio, dependiendo de la aplicación se puede cambiar la configuración del WTA.

Aplicando todo lo visto en un circuito completo se espera lo obvio, el funcionamiento es el esperado después de observar a detalle cada bloque. Para verificar esto se planea una cama de prueba donde se estima cómo son las señales de corriente generadas por los diodos receptores de luz ante el estímulo de una barra oscura pasando en un fondo brillante, este ejemplo se pensó así pues es el objetivo principal del sistema de visión a seguir, el sistema fotorreceptor adaptativo, amplifica y transforma la señal en información de voltaje, la que después es pasada a través del bloque selector de polaridad, discrimina la señal según su dirección y transforma la señal para que el WTA pueda procesarlo apropiadamente, este último bloque codifica la señal para seleccionar y ubicar la característica más sobresaliente.

Para transformar la información espacial en movimiento y centrar el objetivo, se necesita un circuito basado en una red de resistencias como divisor de voltaje y una red de agregación, los cuales modulan, mediante un circuito generador de pulsos, un motor de tipo DC el cual mueve la estructura del vehículo, sin embargo debido a que este circuito es muy complejo en su aplicación ya que se necesita conocimiento de control de motores y en general electrónica de potencia, en esta tesis solo se menciona este circuito, el cual se puede reemplazar por un dispositivo digital.

6.3 Trabajo futuro

Aunque los objetivos de esta tesis llegan sólo hasta la simulación de los bloques jerárquicos siempre hace falta un trabajo de construcción de un prototipo físico el cual verifique y proponga puntos de retroalimentación y de mejora para este sistema y resuelva de manera apropiada los problemas de layout.

6.4 Conclusión

La naturaleza es el mejor lugar para encontrar soluciones a problemas de todo tipo, aquí mostramos una forma de aplicar lo que la naturaleza propone para obtener un dispositivo que es asíncrono, eficiente en el consumo de energía, pequeño comparado con su contraparte digital, características en donde la naturaleza las ha resuelto con elegancia, quizá aún estamos lejos de igualarla pero ya estamos encaminados.

Apéndice A.

Publicaciones.

En esta sección se enlista la publicación realizada durante el año de tesis.

1. Cesar R. Acosta-Méndez, F. Sandoval-Ibarra, J. L. Del Valle-Padilla. Circuito Neuromórfico WTA: Fundamentos y Consideraciones de Diseño en Silicio. Reunión de Otoño de Potencia, Electrónica y Computación, IEEE, 2012.

Bibliografía

- [1] G. Indiveri, *Neuromorphic Winner-Take-All Circuits for Selective Attention Systems*, Genova: University of Genova, 1999.
- [2] L. D. Emiliano, *Sistemas Digitales y Arquitectura de Computadoras*, México, 1993.
- [3] C. Mead, «Neuromorphic Electronic Systems,» *Proceedings of the IEEE*, vol. 78, nº 10, pp. 1629-1636, October 1990.
- [4] C. Bartolozzi, *Selective Attention in Silicon: From the Design of an Analog VLSI Synapse to the Implementation of a Multi-Chip System*, Zurich, 2007.
- [5] G. Indiveri, «Neuromorphic Analog VLSI Sensor for Visual Tracking: Circuits and Applications Examples,» *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 46, nº 11, pp. 1337-1347, November 1999.
- [6] C. Mead y M. Mahowald, «A silicon model of early visual processing,» *Neural Networks*, vol. 1, nº 1, pp. 91-97, 1988.
- [7] M. Mahowald y C. Mead, «The Silicon Retina,» *Scientific American*, vol. 264, nº 5, pp. 76-82, May 1991.
- [8] T. Delbruck y C. Mead, «Analog VLSI Phototransduction by Continuous-time, Adaptive, Logarithmic Photoreceptor circuits,» *California Institute of Technology. Computation and Neural Systems Program*, vol. CNS Memo, nº 30, April 2, 1996.
- [9] K. Boahen, «The Retinomorph Approach: Pixel-Parallel Adaptive Amplification, Filtering and Quantization,» *Analog Integrated Circuits and Signal Processing*, vol. 13, nº 1-2, pp. 53-68, May-June 1997.
- [10] J. Kramer, «Compact Integrated Motion Sensor With Three-Pixel Interaction,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, nº 4, pp. 455-460, April 1996.
- [11] R. A. Deutschmann y C. Koch, «An Analog Velocity Sensor Using The Gradient Method,» *IEEE Intl. Symposium on Circuits and Systems*, pp. 649-652, 1998.
- [12] M. Ohtani, H. Y. T. Asai y N. Ohshima, «Analog Velocity Sensing Circuits Based on Bio-inspired Correlation Neural Networks,» *Microelectronics for Neural, Fuzzy and Bio-Inspired Systems*, pp. 366-373, Granda, Spain, 1999.

- [13] G. Indiveri, «A 2D Neuromorphic VLSI Architecture for Modeling Selective Attention,» *IEEE Computer society*, vol. 4, 2000.
- [14] G. Indiveri, R. M. y J. Kramer, «Active Vision Using an Analog VLSI Model of Selective Attention,» *IEEE Transactions on Circuits and Systems II*, vol. 48, 2001.
- [15] G. Indiveri, «Neuromorphic VLSI Models of Selective Attention: From Single Chip Vision Sensor to Multi-Chip Systems,» *In Sensors*, vol. 8, pp. 5352-5375, 2008.
- [16] D. Sonnleithner y G. Indiveri, «A Real-Time Event-Based Selective Attention System for Active Vision,» *Chapter in Advances in Autonomous Mini Robots*, 2012.
- [17] A. Lázár, *Modeling Visual Attention*, Budapest: Faculty of Information Technology Pázmány Péter Catholic University, 2008.
- [18] A. Treisman y G. Gelade, «A Feature-Integration Theory of Attention,» *Cognitive Psychology*, pp. 97-136, 1980.
- [19] C. Kooch y S. Ullman, «Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry,» *Human Neurobiol*, pp. 219-227, 1985.
- [20] C. Hildreth y S. Ullman, «The Measurment of Visual Motion,» *Massachusetts Institute of Technology Artificial Intelligence Laboratoory*, nº 699, 1982.
- [21] D. Marr y E. Hildreth, «Theory of Edge Detection,» *Proceedings of the Royal Society of London. Series B, Biological Sciences*, vol. 207, nº 1167, pp. 187-217, Feb 29 1980.
- [22] S.-C. Liu, J. Kramer, G. Indeveri, T. Delbrück y R. Douglas, *Analog VLSI Circuits and Principles*, Massachusetts: Massachusetts Institute of Technology, 2002.
- [23] S. M. Sze, *Physics of semiconductor devices*, New York: Wiley, 1981.
- [24] Y. P. Tsividis, *Operation and modeling of the MOS transistor*, New York: McGraw-Hill, 1987.
- [25] P. R. Gray, P. J. Hurst, S. H. Lewis y R. G. Meyer, *Analysis and design of analog integrated circuits*, New York: Wiley, 2001.
- [26] L. Simard, «Biological Vision,» *Fundamentals of Computer Vision*, pp. 308-558, 2012.
- [27] T. Delbruck y C. Mead, «Time-Derivative Adaptive Silicon Photoreceptor Array,» *Computation and Neural Systems Program*, 1991.
- [28] S.-C. Liu, «Silicon Retina With Adaptive Filtering Properties,» *Advances in Neural Information Processing Systems*, vol. 10, pp. 712-718, 1997.

-
- [29] J. Lazzaro, S. Ryckebusch, M. Mahowald y C. A. Mead, «Winner-Take-All Networks Of $O(N)$ Complexity,» de *Advances in Neural Signal Processing Systems 1*, San Francisco, CA, Morgan Kaufmann Publishers Inc., 1989, pp. 703-711.
- [30] S. P. DeWeerth y T. G. Morris, «Analog VLSI Circuits for Primitive Sensory Attention,» *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 6, pp. 507-510, Junio 1994.
- [31] W. Bair y C. Koch, «An Analog VLSI Chip for Finding Edges from Zero-crossings,» *Advances in Neural Information Processing Systems*, vol. 3, pp. 399-405, 1990.
- [32] E. A. Vittoz, «Pseudo-Resistive Networks and their Applications to Analog Collective Computation,» *Procesing MicroNewro*, pp. 163-173, 1997.
- [33] K. H. Wee, J.-j. Sit y R. Sarpeshkar, «Biasing techniques for subthreshold MOS resistive grids,» *IEEE*, pp. 2164-2167, 2005.
- [34] S. DeWeerth y T. Morris, «CMOS current mode winner-take-all circuit with distributed hysteresis,» *ELECTRONICS LETTERS*, vol. 31, nº 13, pp. 1051-1053, 1995.
- [35] G. Indiveri, «Winner-take-all networks with lateral excitation,» *Analog Integrated Circuits and Signal Processing*, vol. 13, nº 1-2, pp. 185-193, 1997.
- [36] A. G. Andreou, R. C. Meitzler, K. Strohbehn y K. A. Boahen, «Analog VLSI neuromorphic image acquisition and pre-processing systems,» *Neural Networks*, vol. 8, nº 7, pp. 1323-1347, 1995.
- [37] A. Licup, *A Prototype Imaging System With a Biologically Inspired Imager for an Extended Range Wlking Stick*, Trieste, Italy, September 2004.
- [38] T. Delbruck, «Frame-Free Dynamic Digital Vision,» *Proceedings of Intl. Symp. on Secure-Life Electronics, Advance Electronic for Quality Life on Society, Univ of Tokyo*, vol. 13, nº 1-2, pp. 21-26, Mar 6-7 2008.



CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL I.P.N. UNIDAD GUADALAJARA

El Jurado designado por la Unidad Guadalajara del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional aprobó la tesis

Análisis y simulación de un chip de visión neuromórfico para la
detección de movimiento

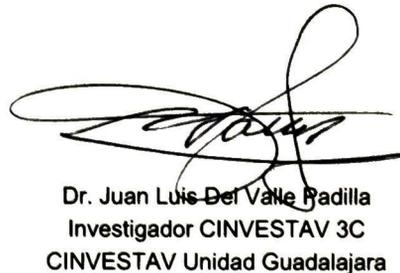
del (la) C.

César Rodolfo ACOSTA MÉNDEZ

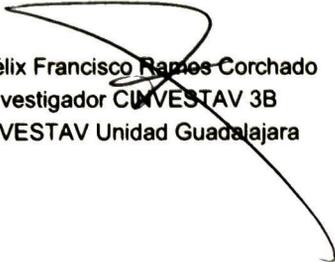
el día 03 de Octubre de 2014.



Dr. Federico Sandoval Ibarra
Investigador CINVESTAV 3C
CINVESTAV Unidad Guadalajara



Dr. Juan Luis Del Valle Padilla
Investigador CINVESTAV 3C
CINVESTAV Unidad Guadalajara



Dr. Félix Francisco Ramos Corchado
Investigador CINVESTAV 3B
CINVESTAV Unidad Guadalajara



Dra. Susana Ortega Cisneros
Investigador CINVESTAV 3A
CINVESTAV Unidad Guadalajara



CINVESTAV - IPN
Biblioteca Central



SSIT0012742