CAPITULO 3

NEUROCOMPUTACIÓN

3.1 INTRODUCCIÓN

Los modelos neuronales artificiales están basados tanto en el estudio de las redes neuronales biológicas como en consideraciones teórico-prácticas convenientes para la representación del mundo físico. Si bien el uso de estos modelos se ha extendido a muy diversas ramas de la ciencia, el estudio del presente capítulo y sus conclusiones, se refieren a la aplicación de sistemas neuronales en ingeniería; aún cuando estas propiedades pueden ser extensibles a otras áreas.

Los avances teóricos en redes neuronales han sido significativos, particularmente en los últimos diez años. Si bien, la investigación en el tema ha sido continua desde las épocas de McCulloch-Pitts y Hebb, no cabe duda que ha recibido un notable impulso luego de la aparición del modelo de Hopfield y el algoritmo de aprendizaje basado en retropropagación del error, para redes neuronales estáticas multicapa. Las aplicaciones de la teoría en ingeniería comprenden tareas de clasificación, procesamiento de señal, modelado de procesos complejos, optimización, control, visión, etc. Aquí, la teoría de redes neuronales ha dado lugar a un nuevo enfoque o punto de vista sobre la manera de resolver estos problemas. Cabe aclarar, que no es obvio que los modelos neuronales sean más eficientes que otras teorías en la resolución de un problema en particular. Son pocos los casos en que se puede demostrar formalmente que un modelo neuronal es mejor para la resolución de una clase específica de problemas. Muchas veces debemos conformarnos con acumular evidencia comparativa entre diversas técnicas aplicadas a casos particulares de problemas de un mismo tipo o clase. De todas maneras esta situación es común a otras áreas de la teoría en problemas de ingeniería. Sin embargo, existe una propiedad de los modelos neuronales comparativamente superior, la capacidad de computo paralelo que redunda en un elevado rendimiento.

Ya fue dicho durante los capítulos 1 y 2 que amen de la diversidad de modelos neuronales existentes, existen dos propiedades unificantes: el paralelismo masivo de unidades computacionales simples y la adaptatividad de los parámetros del modelo en función de las características de los datos de entrada. Una dada arquitectura neuronal, se caracteriza por: la cantidad de funciones que puede representar, su algoritmo de aprendizaje y la forma en que esta red neuronal ha sido construida. En el capítulo 2 se analizó, la capacidad representativa de una clase de modelos neuronales y sus algoritmos de adaptatividad. En este capítulo se realiza un análisis sobre la implementación paralela de redes neuronales artificiales.

Este es, sin duda, un tema fundamental en la investigación sobre modelos neuronales, ya que la construcción de los mismos en hardware posibilita el pleno aprovechamiento de la capacidad de cómputo paralela de la red. Actualmente, más del 90% de las redes neuronales existentes son simuladas en computadoras convencionales. Estas simulaciones han servido para desarrollar la teoría de los diferentes modelos, descubrir nuevos enfoques y algoritmos para la resolución de un problema o una clase de problemas y dar solución a aplicaciones específicas de la ingeniería. Sin embargo, las redes neuronales son particularmente atractivas cuando se aplican a problemas complejos, para los cuales los métodos tradicionales sólo dan soluciones pobres, y en los que se puede hacer valer las características adaptivas de dicha red. Pero justamente, estos problemas complejos resueltos con métodos neuronales necesitan una gran potencia de cómputo. Simular los modelos de la red neuronal en una computadora convencional sólo tiene el beneficio de la flexibilidad del sistema de cómputo pero se torna rápidamente ineficiente en función de la complejidad del problema. En las secciones subsiguientes se analizan diferentes formas de generar dispositivos y sistemas de cómputo neuronal.

3.2 DISPOSITIVOS Y SISTEMAS NEURONALES

Implementar un sistema neuronal que sea capaz de representar un dado modelo neuronal, implica generar una arquitectura paralela de las unidades computacionales de dicha red. Cada unidad computacional se convierte en un procesador neuronal. En la simulación de modelos neuronales en computadoras los procesadores neuronales son simulados por uno o más microprocesadores convencionales. En una computadora neuronal debe existir un arreglo paralelo de procesadores especialmente orientados para la computación de un dado algoritmo. Aquí surgen varios interrogantes en el diseño de los procesadores neuronales en función de: la tecnología digital o analógica a utilizar, la flexibilidad que debe tener el sistema para adaptarse a distintos modelos neuronales y a distintos tamaños de una red; la integración que debe tener el sistema con señales sensoriales externas o con usuarios.

Las distintas elecciones hechas en función de estos interrogantes afectan en forma directa propiedades del sistema tales como:

- precisión de la función de cómputo,
- ancho de banda de procesamiento,
- forma de almacenamiento de los parámetros adaptativos,
- interconexión y comunicación entre unidades neuronales,
- consumo de potencia,
- tipo de arquitectura,
- capacidad,
- costo.

Básicamente, puede pensarse que un procesador neuronal elemental es una unidad RISC (Reduced Instruccion Set Controller) digital o analógica. Si no se requiere mayor flexibilidad, cada neurona computa un algoritmo fijo, consistente en multiplicaciones, sumas y la aplicación de una función no-lineal. Sin lugar a dudas, la implementación de redes neuronales no sería posible si el gran avance que ha tenido la tecnología de integración en gran escala (VLSI). Actualmente, los procesos tecnológicos más avanzados permiten integrar algunas decenas de millones de transistores en un solo dispositivo. Se

espera que con los materiales y procesos actuales el límite de la tecnología pueda ser mejorado en al menos dos órdenes de magnitud en la próxima década. Lo cual constituye un pronóstico promisorio para alcanzar un número considerable de neuronas por cada dispositivo. Esto es fundamentalmente importante debido a que, en general la interconexión entre neuronas es mucho mayor que con las señales sensoriales externas.

Los dispositivos VLSI permiten explotar el paralelismo intrínseco existente en una red neuronal generando múltiples unidades idénticas, con una topología regular y simplificando el diseño del mismo. Además, proveen un mecanismo eficiente y robusto para la implementación de resultados teóricos sobre modelos neuronales e investigación sobre alta densidad de componentes y transistores de canal corto.

3.3 ANALÓGICO vs. DIGITAL

Un punto fundamental en la generación de sistemas neuronales es la unidad de procesamiento neuronal. Estas unidades tienen una importancia mucho mayor que en una computadora secuencial, debido a que las neurocomputadoras aumentan su potencialidad aumentando el número de unidades. Además, la mayoría de la unidades neuronales incluyen la memoria de la neurocomputadora. Como se dijo, una unidad de computo neuronal es un procesador digital o analógico que realiza un conjunto reducido de operaciones simples. La física de dispositivos electrónicos permite representar excelentemente los modelos matemáticos neuronales.

Dentro de las tecnologías electrónicas disponibles, sin dudas, la CMOS (*Complementary Metal Oxide Semiconductor*) se ha adueñado del tema en forma casi excluyente. Los motivos fundamentales de esta preferencia son, entre otros: la gran versatilidad que posee la tecnología CMOS, posibilitando el diseño de múltiples funciones; conversión corriente a tensión (y viceversa) con un solo transistor; bajo consumo de potencia; dos regiones de funcionamiento en modo activo, como fuente de corriente en la zona de saturación y como

resistencia variable en la región de funcionamiento ohmico o lineal. Conmutación de alta velocidad que permite trabajar en circuitos digitales con frecuencias de reloj superior a 100 MHz.

El apéndice B describe al transistor MOS en tres regiones de funcionamiento: deplexión, inversión débil e inversión fuerte. La curva de corriente de drenador Id en función de la tensión compuerta-substrato Vgs sigue inicialmente una ley exponencial. Este modo de funcionamiento se denomina sub-umbral y se extiende hasta valores Vgs ~ 0.7v. Para un rango de valores de Vgs más elevados Id disminuye su respuesta exponencial frente a la entrada Vgs hasta que para Vgs > 2.Vt ley que las relaciona es aproximadamente cuadrática.

Una unidad de procesamiento neuronal puede ser realizada utilizando los modelos de la electrónica analógica, digital o una combinación de ambas. Cada una de ellas tiene ventajas y desventajas comparativas que se detallarán a continuación.

3.3.1 NEURONAS ANALOGICAS

Los modelos analógicos son esencialmente compactos cuando representan las funciones aritméticas neuronales tales como: multiplicación, suma, y ciertas funciones no lineales (3.1) y (3.2).

$$u_k = \sum_{j=1}^{p} w_{kj} x_j \qquad (3.1)$$

$$y_k = \varphi(u_k - \theta_k) \qquad (3.2)$$

En el procesamiento de señales analógicas pueden usarse tensiones o corrientes. Si bien, salvo en el caso de conductancias o resistencias de valor infinito, toda tensión lleva una corriente asociada, la señalización puede ser más fácilmente manipulada en función de una de ellas. Como se verá expresamente en el capítulo 4, la operación suma puede ser representada,

siguiendo la ley de Kirchoff, simplemente por un nodo al cual llegan fuentes de corriente.

Un producto de términos puede ser modelado por medio de un amplificador de transconductancia con ganancia variable. En este caso existe una conversión de la señal de una tensión de entrada a una corriente de salida.

El modelo electrónico analógico de la función no-lineal φ depende, obviamente, del tipo de función requerida. Si la función es de tipo sigmoidea o umbral con discontinuidad (función signo), puede ser representada por un amplificador diferencial. La pendiente de la función sigmoidea puede ser ajustada variando la ganancia de tensión de dicho amplificador. En el caso límite de la función umbral esta ganancia puede tomar valores muy grandes usando técnicas de conmutación como en las compuertas lógicas.

La electrónica analógica es muy compacta comparada con la digital, las funciones mencionadas pueden ser realizadas con un número mínimo de transistores en un rango de 1 a pocas decenas. Tienen un ancho de banda de procesamiento mucho mayor que el de la electrónica digital, especialmente cuando se los utiliza en el modo de corriente.

3.3.1.1. ARQUITECTURA DE NEURONAS ANALOGICAS

Según la ecuación (.1) cada salida y es función del producto interno de los vectores **x** y **w**. Si extendemos el espacio de salida a un vector m-dimensional, el vector **y** es el resultado del producto de una matriz de pesos w por un vector de entradas **x** (3.3). Esta ecuación puede ser representada por una matriz de sinapses como indica la figura 3.1.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} w_{11} \cdots\cdots\cdots w_{1n} \\ \vdots \quad \cdots \quad \vdots \\ w_{m1} \cdots\cdots\cdots w_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad (3.3)$$

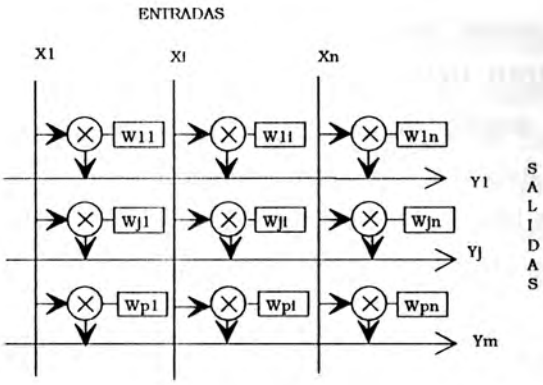

Figura 3.1 Matriz sináptica

Una red neuronal monocapa cuya dimensionalidad de entrada es **n** y de salida **m**, debe realizar n.m multiplicaciones (y sumas) sinápticas, y sólo m transformaciones no-lineales. Es decir, es crítico el tamaño y complejidad del circuito multiplicador, ya que consume un área importante del VLSI. La figura 3.2 representa un multiplicador propuesto por Bult y Wallinga (Burt, 1987). Este multiplicador provee una corriente de salida:

$$I_{SAL} = \frac{2 I_a I_b}{\left[k(V_2 - 2V_t)^2\right]} = \frac{I_a I_b}{2 I_0} \qquad (3.4)$$

donde Ia e Ib representan fuentes externas de corriente en forma de modo común y modo diferencial. En este circuito los transistores trabajan en zona de saturación post-umbral (ley cuadrática), en la ecuación (3.4) k representa la transconductancia y Vt la tensión de umbral del MOS.

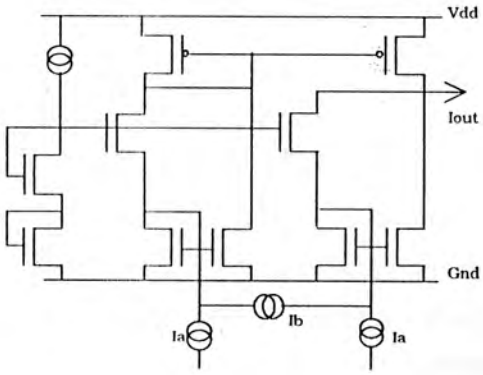

Figura 3.2 multiplicador de Bult-Wallinga

Este circuito posee buen rango dinámico lineal y gran ancho de banda. Los autores aseguran tiempos de subida menores a 50 nseg, *slew-rate* de .28μ A/nseg y un ancho de banda superior a 5 MHz. Sin embargo el circuito necesita dos fuentes externas, idénticas, de corriente y una referencia I0 común a todas las celdas multiplicadoras. En el próximo capítulo se analizaran en detalle otros circuitos multiplicadores analógicos.

Otro punto clave en el diseño de una red neuronal en VLSI es el conexionado entre neuronas. Existen tantas conexiones como productos en el circuito, es decir m.n para una red de una capa plenamente conectada. Este conexionado metálico consume gran cantidad de espacio del VLSI y complica en gran medida el diseño, ya que las conexiónes se entrecruzan entre todas las entradas de una capa con todas las neuronas de dicha capa. La redes neuronales digitales pueden solucionar en parte este problema generando buses compartidos en forma de redes de datos y utilizando algún mecanismo de multiplexado.

3.3.2 NEURONAS DIGITALES

La tecnología CMOS digital aporta una mayor flexibilidad para el diseño de unidades de procesamiento neuronal. Estas unidades pueden diseñarse, verdaderamente, como un procesador RISC elemental, el cual posea una unidad aritmética y un multiplicador. Las funciones no-lineales pueden ser implementadas dentro del dispositivo o con la ayuda de tablas en memoria.

La figura 3.3 representa un esquema simplificado de la arquitectura de una unidad de procesamiento neuronal digital. Cada unidad de procesamiento es conectada a dos buses para entrada y salida de datos. Estos datos deben ser almacenados en registros internos. Además la lógica de comunicaciones debe poseer un par de registros para la identificación de la neurona (número de capa y posición dentro de ella). Dicha lógica debe encargarse de recibir señales de entrada y enviar los resultados del procesamiento a las neuronas conectadas.

La unidad de procesamiento debe contener al menos una unidad aritmético-lógica (ALU), un registro acumulador (Ax) y un registro auxiliar para la multiplicación (Mx). En este caso la multiplicación se hace por algoritmo de suma y corrimiento (*add-shift*). Implementar un multiplicador por hardware representa un costo adicional de área de integración, especialmente, sabiendo que, cada unidad de procesamiento necesita su propio multiplicador. Las operaciones son controladas por el procesador consistente en un puntero de instrucciones, un registro de datos (MDR), un registro de instrucciones (ICR) y la memoria de datos/programa.

Un procesador neuronal de este tipo puede incluir mas de una unidad de procesamiento en un mismo dispositivo VLSI. Inclusive, manteniendo una estructura de comunicaciones fija entre las unidades de un mismo dispositivo (Fig. 3.4). De esta manera no se aumenta la cantidad de pines. La figura 3.4 muestra un arreglo bidimensional de procesadores neuronales comunicados por conexiones punto a punto en una de las direcciones del arreglo (horizontal) y por medio de un nodo especial encargado de hacer un algoritmo de enrutamiento para la otra dirección. Este nodo debe posibilitar que los mensajes lleguen a todas las unidades neuronales que poseen conexión sináptica con cada una de ellas.

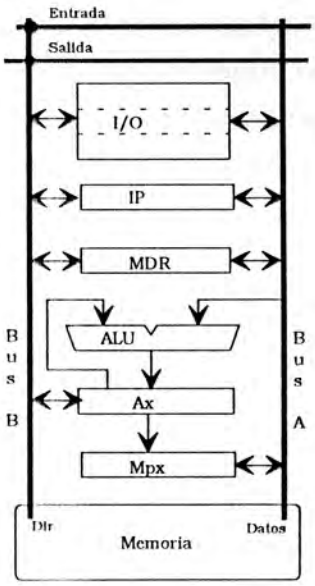

Figura 3.3 Unidad de procesamiento neuronal digital

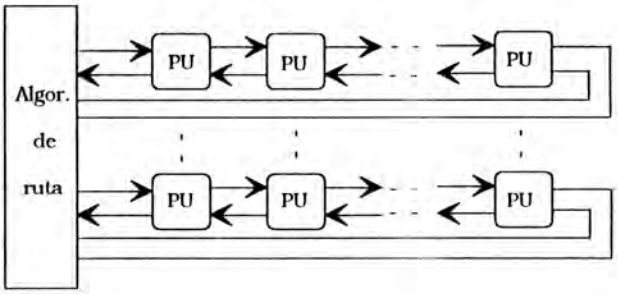

Figura 3.4 Sistema neuronal digital

3.3.3 NEURONAS MIXTAS ANALOGICO-DIGITALES

De las técnicas híbridas analógico-digiteles, sin duda, la que mayor interés ha suscitado es la aritmética de pulsos. Probablemente, el interés se deba a la similitud que puede encontrarse con las neuronas biológicas en lo que se refiere al pulso como mecanismo neurotransmisor. Esta técnica ha sido desarrollada por Murray (Murray, 1987a, 1987b, 1989) y trata de tomar ciertos beneficios de ambas técnicas. Básicamente, existe una codificación de la señal en forma digital y analógica conviviendo en distintas partes del circuito neuronal. La señal analógica se representa por una tensión $\mathbf{v_i}$ como una función continua del tiempo. La señal digital es codificada en pulsos digitales a una frecuencia de portadora mayor que el ancho de banda de la señal continua. La figura 3.5a muestra el esquema de una neurona la cual recibe una señal de entrada analógica $\mathbf{v_i}$ y proporciona una salida $\mathbf{y_j}$ codificada en pulsos, cuya frecuencia es proporcional a la amplitud de la entrada. Este comportamiento puede ser generado por medio de un oscilador controlado por tensión (VCO). La función sináptica en este modelo es la de regular (modular) la cantidad de pulsos que conforman la señal neurotransmisora (Fig. 3.5b). Las salidas neuronales son sumadas en tensión o en corriente utilizando funciones de tipo OR.

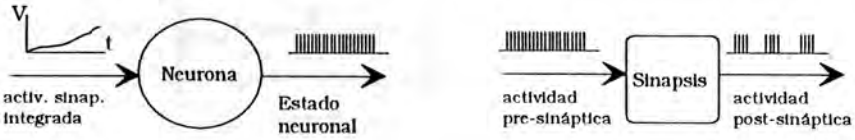

Figura 3.5 Codificación de pulsos: a) Neurona, b) Sinapsis

La figura 3.6 muestra un esquema en el cual se suman los aportes sinápticos excitatorios e inhibitorios de los pulsos y su conversión a una tensión en un capacitor. Esta tensión resultante controla un VCO entre límites de frecuencia ϕ_{min} y ϕ_{max}.

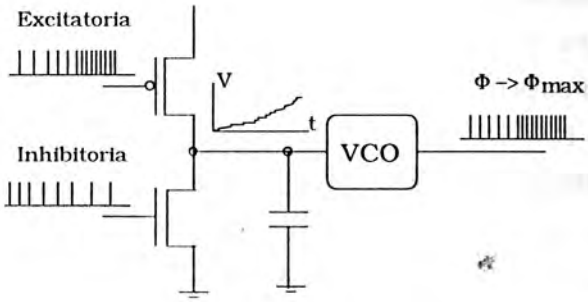

Figura 3.6 Sumador

La función sináptica se realiza utilizando pulsos cuya relación de trabajo decrece desde 1/2 hasta un cierto número $1/2^N$. Este número N indica el número de bits de precisión del circuito. El coeficiente de peso sináptico es almacenado en memoria en forma digital y modula la habilitación de los pulsos mencionados anteriormente. Es decir cada bit **i** de memoria en 1 habilita el paso de pulsos de la portadora hacia la salida por un intervalo $1/2^i$. Los bits de la palabra de peso en 0 inhabilitan que dichos pulsos pasen a la salida sináptica. La función sinaptica puede ser representada por la ecuación:

$$o_j = S_j\left(W_{j1}c_1 + W_{j2}c_2 + \ldots + W_{jN}c_N\right) \qquad (3.5)$$

La figura 3.7 muestra la manera en que estos pulsos son combinados mediante una función OR e inyectados a la columna exitatoria o inhibitoria que conducen al cuerpo de la neurona.

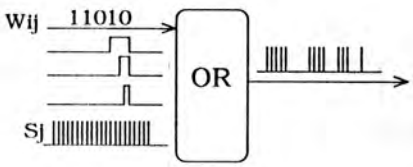

Figura 3.7 Modulador de pesos sinápticos

3.3.3.1 PESOS ANALOGICOS EN EL SISTEMA DE CODIFICACION DE PULSOS

La memoria digital de almacenamiento del peso en la función sináptica puede ser reemplazada por una versión plenamente analógica (Murray,1989). En este caso el par inversor T1-T2 de la figura 3.8 está asociado a la función sináptica. La misma muestra que el coeficiente de peso es almacenado dinámicamente como una tensión sobre el capacitor cm. El pulso de portadora S_j maneja el par inversor. Cuando S_j=0 el capacitor de salida se carga a la tensión almacenada en la memoria dinámica. Cuando S_j=1 se habilita la descarga del capacitor a través de T2 con una constante de tiempo mucho mayor que la de carga (Fig. 3.9). Como puede verse en la figura 3.9 la memoria dinámica controla el valor de continua de salida, y por lo tanto el ancho del pulso que se conforma a la salida del segundo inversor.

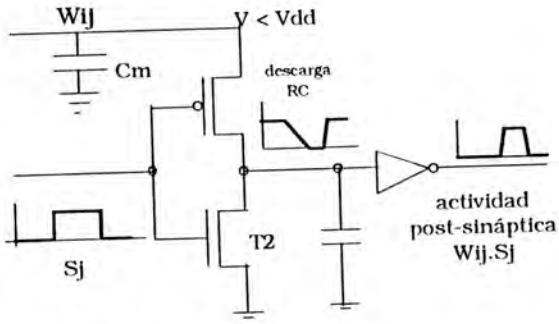

Figura 3.8 Memoria analógica en el sistema de codificación de pulsos

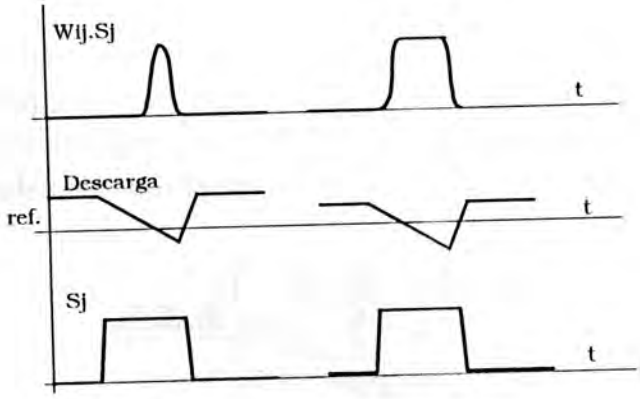

Figura 3.9 Generación del pulso en función del valor almacenado en memoria

Este método de almacenamiento analógico reduce considerablemente el área destinada a la matriz de coeficientes de pesos respecto del modelo digital. Además, el uso de pseudo-relojes del modelo digital en complicado de implementar y produce ruido y picos de consumo de corriente dentro del VLSI.

3.3.4 COMPARACION ENTRE NEURONAS ANALOGICAS, DIGITALES E HIBRIDAS: ANALISIS Y DISCUSION

La importancia de llevar los modelos neuronales teóricos al campo de las aplicaciones es fundamental desde el punto de vista del rendimiento de la red neuronal en la solución de sistemas complejos. El elemento constitutivo básico de las neurocomputadoras es la unidad de procesamiento. Estas unidades computacionales deben ser dispuestas de manera de generar una arquitectura de cómputo en paralelo en una organización regular de una o más capas.

Los dispositivos neuronales en VLSI pueden concentrar un número importante de unidades de cómputo, dependiendo del límite de la tecnología empleada y de la complejidad requerida por la unidad neuronal. A este respecto puede decirse que los modelos analógicos poseen una amplia ventaja sobre los digitales e híbridos. Un multiplicador sináptico puede ser desarrollado con una decena de transistores, al igual que los módulos que implementan las funciones no-lineales de salida. En particular, si se utiliza la tecnología MOS trabajando en la zona previa al umbral (*subthreshold*) la característica de transferencia entre corriente y tensión sigue una ley exponencial, lo cual permite implementar, con pocos transistores la respuesta sigmoidea típica en los modelos tipo Perceptrón. En consecuencia, podemos hacer un cálculo aproximado de la cantidad de transistores necesaria en una neurona analógica a través de la ecuación:

$$N_T = \sum_{j=1}^{n} \left(n_{mul}^{j} \cdot conex^{j-1} + n_f^{j} \right) conex^{j+1} \quad (3.6)$$

en la cual N_T es el número total de transistores utilizados, n_{mul} y n_f, el número de transistores del multiplicador y de la función no lineal respectivamente; y **conex** representa el número de conexiones o *fan-in* de cada neurona. En todos los casos el supra-índice representa el número de capa de la red neuronal.

Las neuronas digitales, por el contrario no son compactas. La forma más económica de hacer un multiplicador de punto fijo es utilizando una unidad aritmético-lógica (ALU) y un registro de desplazamiento. En este caso la multiplicación se realiza por sumas y desplazamientos hacia la izquierda del multiplicador. Aún cuando los transistores de la tecnología digital son más pequeños que los de la analógica, tanto la ALU como un registro de desplazamiento, consumen un área mucho mayor que en el caso anterior. Esta área aumenta con el número de bits de precisión del multiplicador. Además, aumenta el número de iteraciones en la cuenta y en consecuencia, se reduce el ancho de banda de la red neuronal. Si bién existen multiplicadores que realizan la operación en un solo paso, el área que consumen es prohibitiva si el circuito debe repetirse para cada unidad de procesamiento.

En la tecnología híbrida pueden darse ambas situaciones, pero si se opta por la memoria analógica dinámica, el área consumida es pequeña y comparable a la del multiplicador analógico.

Otro punto importante es el referente a la potencia consumida por cada neurona. Si se desea alcanzar un nivel de integración elevado, el consumo de potencia por neurona debe ser extremadamente bajo. Por ejemplo, una neurona biológica disipa del orden de 10^{-12}watts, un transistor trabajando en conmutación, alrededor de 1 millon de veces más. Aún cuando parece imposible mejorar el consumo en esta cifra, pueden realizarse algunas mejoras respecto de los circuitos tradicionales usados tanto en tecnología analógica como digital. Una es bajar el nivel de tensión de la fuente a valores cercanos a 1 o 2 volts. La otra, es utilizar al MOS en la zona *subthreshold*, donde la corriente de drenador es del orden de los nanoamperes o inclusive picoamperes. Esta última facilita el hecho de poder disminuir la fuente ya que la zona *subthreshold*, se encuentra para valores de tensión compuerta-fuente menores a 0,7 volts.

Podemos definir al ancho de banda de un circuito neuronal como el número de patrones procesados por segundo. Aquí las ventajas no son tan claras. El tiempo de procesamiento típico medido por el tiempo de respuesta (rise-time) en una neurona analógica es del orden 100 nseg. Una neurona digital con un reloj elevado necesita del orden de 200 a 300 nseg. De todas maneras estas diferencias no son grandes y las ventajas podrian revertirse si se utiliza tecnologías submicrón con relojes mayores de 100 MHz.

Los diseños digitales aventajan a los analógicos, fundamentalmente, en dos aspectos, exactitud y precisión de los cómputos. La exactitud está relacionada a la mejor relación señal-ruido de las señales digitales frente a las analógicas. Este aspecto puede ser importante, sobre todo, cuando la red neuronal tiene varias capas de procesamiento o cuando aumenta la distancia de las conexiones sinápticas. En contrapartida, aunque las neuronas analógicas sean más suceptibles al ruido, es lógico que en un VLSI digital haya mas ruido generado por la misma electrónica digital. Si, además, esta electrónica es asincrona, las conmutaciones de los relojes asincrónicos pueden ocasionar cambios de estado en lineas vecinas (*cross-talk*). Es importante notar que, aún cuando los ruidos pueden ser muy perjudiciales, en una red neuronal la gran cantidad de conexiones sinápticas y neuronas en paralelo hace que un error simple tenga una influencia moderada en el cómputo final. Además, como se vio en el capítulo 2, las funciones no-lineales de salida saturan para un nivel elevado de la entrada, lo cual favorece la afirmación anterior.

Respecto de la precisión, no cabe duda que los sistemas digitales son ideales para fijar con exactitud la precisión de un circuito, medida en bits. Por la última afirmación del párrafo anterior, las redes neuronales no requieren una alta precisión. Las redes de Hopfield y todas aquellas que tienen una función no-lineal tipo umbral sólo necesitan dos estados de salida, es decir, un bit de precisión. En general, para las neuronas con salida continua tampoco se necesitan muchos bits de precisión, alcanza con 3 a 8 bits, dependiendo de la aplicación.

Los circuitos analógicos son en general compactos, por ende no se utilizan técnicas de compartición o multiplexado de los elementos de cómputo o canales de comunicación. En cambio, estas técnicas pueden ser muy útiles en los sistemas digitales. Como se desprende de lo analizado, los circuitos neuronales digitales son más costosos en área de VLSI, por consiguiente, puede ser ventajoso diseñar circuitos con tecnologías avanzadas y veloces, con alta capacidad de procesamiento. Estas unidades de procesamiento real (UP) pueden emular múltiples neuronas virtuales (EP). Esto reduce el número de UP que deben integrarse en un mismo dispositivo y pueden compartirse los canales internos de comunicación de datos entre neuronas. Por otro lado, se complica la arquitectura de diseño debido a que debe implementarse una manera de distinguir las diferentes EP virtuales dentro de una misma UP. Este es el caso de una arquitectura como la descripta en el ejemplo del procesador RISC de la sección 3.3.2. La figura 3.3 muestra como múltiples unidades UP pueden ser conectadas a buses comunes de datos y direcciones.

En las neuronas digitales los buses compartidos son esenciales debido a que es necesario el acceso a zonas de memoria global. Un caso típico es el de la generación de la función no-lineal de salida. Las funciones exponenciales, trigonométricas, hiperbólicas o gaussianas deben ser almacenadas en forma de una tabla en una zona de memoria de acceso global. El problema es justamente la saturación del canal de comunicaciones entre las neuronas y la zona de memoria global. Además, las funciones llevadas a tablas consumen mucha memoria, aún cuando la precisión sea baja.

3.3.4.1 FLEXIBILIDAD: DISPOSITIVOS DE PROPOSITOS GENERALES Y DISPOSITIVOS ORIENTADOS A APLICACIONES ESPECIFICAS

Sin duda, uno de los aspectos más importantes en el diseño de un VLSI neuronal está referido a su flexibilidad para implementar diferentes modelos. En otras palabras, si el diseño del circuito neuronal está orientado a un único modelo neuronal o si la estructura interna permite ejecutar otros algoritmos. Al

respecto, podemos hacer una categorización simple dividiendo los mismos en circuitos de:

- propósito general
- propósito especial u orientados hacia una sola aplicación

Esta clasificación está bastante correlacionada con la posibilidad de utilizar una tecnología analógica o digital. La flexibilidad de los diseños analógicos es mucho menor. Cambiar un modelo neuronal significa cambiar una parte sustancial de la función sináptica o neuronal, lo cual en un circuito analógico se traduce en una organización totalmente diferente de los transistores que lo integran. Además, dos funciones de complejidad matemática similar pueden traducirse en complejidades de diseño electrónico muy distintas. Por ejemplo: función tgh(.) y función gaussiana. En consecuencia, los circuitos neuronales analógicos deben ser orientados hacia un modelo en particular. Más aún, muchas veces, hacia una aplicación en particular.

Hoy en día, esta falta de flexibilidad se compensa, parcialmente, por el gran avance de los procesos de integración y fabricación; comúnmente llamados *custom*. Es muy probable que se incrementen mucho más las facilidades para la fabricación "a medida" de circuitos electrónicos orientados a una aplicación determinada.

De todas maneras, la tecnología digital proporciona una flexibilidad mayor para el desarrollo de modelos neuronales, especialmente cuando se trata de prototipos de laboratorio. Los procesadores neuronales, como el descripto en la sección 3.3.2, pueden ejecutar varios algoritmos de cómputo correspondiente a distintos modelos o diferentes arquitecturas de interconexión y número de capas dentro de un mismo modelo. Si las unidades de computo digitales ejecutan un programa en memoria, cambiar de modelo sólo significa modificar el programa. Sin embargo, esta flexibilidad no es totalmente gratis, ya que tener una memoria de programa, buses compartidos y otras características propias de la arquitectura lleva a un costo extra de área del VLSI que disminuye enormemente el número de unidades de procesamiento que pueden integrarse en un mismo dispositivo.

3.3.4.2 EL PROBLEMA DEL ALMACENAMIENTO DE LOS COEFICIENTES DE PESO

El problema de almacenar eficientemente los parámetros de la red neuronal es central en el diseño de la misma. Los parámetros de una red neuronal son: los pesos, Wij, los umbrales θj, y algunos coeficientes propios del modelo, tal como la temperatura de la sigmoide en los Perceptrones y redes de Boltzman.

Estos coeficientes son modificados durante la ejecución del algoritmo de aprendizaje y se mantienen constantes en la faz de procesamiento paralelo. En su mayoría, los algoritmos de entrenamiento son iterativos, por lo cual cada parámetro es modificado un gran número de veces, generalmente, en pequeñas cantidades. Esto impone la necesidad de tener una memoria de lectura escritura, ya sea digital o analógica.

El almacenamiento en estas memorias puede ser temporario en celdas dinámicas o estáticas, o semi-permanente en memorias eléctricamente reprogramables. Las memorias digitales pueden tener cualquiera de estos tres esquemas. Las analógicas pueden usar celdas dinámicas o eléctricamente reprogramables. El almacenamiento en memoria de los pesos en forma digital es idéntico al utilizado en las memorias digitales comunes. En cambio las memorias analógicas constituyen un fenómeno novedoso en el cual aún resta mucho por investigar. Un ejemplo de memorias analógicas es el utilizado en el modelo analógico-digital de la sección 3.3.3.1.

Las memorias analógicas pueden ser dinámicas o semi-permanentes. La celda de la memoria analógica dinámica es similar a su correspondiente digital. El valor del dato almacenado es función de la tensión en el capacitor de entrada de un transistor MOS. El mecanismo de control de esta tensión puede hacerse usando una llave electrónica como en los circuitos de capacitores conmutados. Las llaves son MOS con una corriente de fuga muy pequeña, no obstante, estas memorias requieren un sistema de refresco. Vale aclarar que el refresco aquí es mucho más crítico que en el caso de las memorias digitales, debido a que éstas sólo tienen dos estados, en cambio las memorias analógicas deben mantener la precisión de bit seleccionada en el diseño. Una ventaja de las memorias

dinámicas es que son compactas y el ancho de banda de programación es elevado. Sin embargo, tienen el problema que necesitan sistemas de refresco asincrónicos y complicados, debido a que los pesos de las sinapses son modificados, también, en forma asíncrona.

Una técnica que ha despertado mucho interés es la de escritura semi-permanente o borrable eléctricamente EEPROM. Estas memorias se diseñan utilizando un transistor MOS con dos compuertas. Una de ellas es flotante, es decir, sin acceso desde el exterior. El propósito de esta compuerta es la de controlar la característica tensión-corriente del MOS a través de la carga almacenada (Fig. 3.10). La tensión umbral, Vt, se modifica en función de la carga en la compuerta flotante.

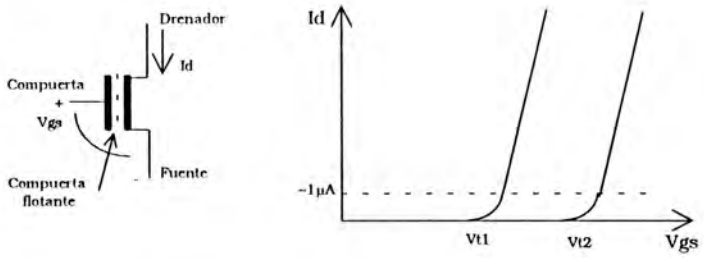

Figura 3.10 variación de Vt en función de la carga en la compuerta flotante

Dado que esta compuerta no está accesible desde el exterior (Fig. 3.10), se utiliza el fenómeno túnel o Fowler-Norldheim para llevar cargas, que quedan almacenadas en forma estática. El apéndice A está dedicado al análisis del efecto Fowler-Norldheim en memorias de compuerta flotante. La modificación de carga en la compuerta requiere de potenciales de programación elevados, entre 14 y 24 volts; que generan un campo eléctrico suficientemente alto como para producir el efecto túnel. Este proceso puede modelarse a través de la ecuación:

$$\Delta V_t = V_{TOt} + V_{PP2} - V_{TOwt} - 1,375.10^{-6} B/\ln\left[e^{(B/E_0)} + 32420.B.T_P\right] \qquad (3.7)$$

donde VPP2 y TP representan la tensión y el ancho del pulso de programación y ΔVt la diferencia entre el valor de tensión de compuerta actual y el deseado. B, Eo y VTO son constantes propias del MOS.

La gran ventaja de estas memorias analógicas es el poder retener la carga almacenada sin necesidad de una fuente de alimentación. Para evacuar la carga de la compuerta flotante se requiere un campo eléctrico similar aplicado en sentido inverso. Si bien existen pequeñas fugas de carga, puede asegurarse una variación menor al 1% en varios meses.

En las celdas estandar se necesitan pulsos de programación elevados, por lo que existe el riesgo de alcanzar tensiones de ruptura del transistor. Para evitar ello, se utilizan capas de OSi delgadas entre las compuertas, o superficies texturadas que aumentan el campo eléctrico a bajas tensiones de programación (Apéndice A)(Ellis, 1982)(Carley, 1989).

3.4 DISPOSITIVOS NEURONALES: ESTADO DEL ARTE

En los últimos cinco años han surgido algunas decenas de proyectos orientados a la generación de dispositivos de computación neuronal. Los mismos varían en un amplio rango de complejidad, versatilidad y área de aplicaciones. La revista IEEE Transaction on Neural Networks ha dedicado tres números completos a la generación de redes neuronales en hardware (IEEE, 1991, 1992, 1993). Otras actividades sobresalientes son las realizadas por el grupo del Dr. Carver Mead del Instituto Tecnológico de California, Caltech, quien desarrolló sistemas neuronales especiales orientados a substituir los órganos que realizan la primera etapa de la función sensorial humana; tales como una Retina y una Coclea. Además ha implementado un sensor de movimiento óptico y un dispositivo llamado *seehear* destinado a traducir imágenes ópticas en señales auditivas capaces de ser percibidas por personas ciegas o disminuidos visuales. Los trabajos de Mead utilizan tecnología CMOS en la zona de funcionamiento anterior al umbral (*subthreshold*). Las propiedades y resultados han sido volcados en el libro *Analog VLSI and Neural Sistems*, (Mead, 1989).

El propósito de esta sección es presentar brevemente algunas de las realizaciones principales con el motivo de hacer un análisis comparativo del

estado del arte de los dispositivos de computación neuronal. Este análisis ha servido para elaborar un criterio sobre la conveniencia de utilizar una determinada tecnología, arquitectura, y otros propósitos, de diseño, ya enumerados, en las realizaciones hechas para esta tesis y que son volcadas en los capítulos siguientes.

3.4.1 ANNA (Analog Neural Network Arithmetic and logic unit)

Este dispositivo es híbrido y está orientado a la arquitectura de Perceptrón multicapa. Utiliza un esquema combinado analógico-digital, por el cual las primeras capas de procesamiento son analógicas, compactas y de baja precisión, la última capa es digital y aporta mayor precisión. Este dispositivo permite construir Perceptrones con un número variable de capas. La comunicación de entrada y salida es digital. El vector de entrada es cargado en un registro de desplazamiento especial de 64 bits (*barrel shifter*) y afecta a la unidad de computación neuronal, donde se realizan las operaciones sinápticas. Este dispositivo posee una memoria interna de 4096 pesos analógicos que pueden ser flexiblemente agrupados según las necesidades de cada capa.

Las características principales del ANNA son: la flexibilidad de su arquitectura, el elevado número (teórico) de computaciones sinápticas, 10.000 Mc/s, y su sistema de precisión controlada. Por el contrario, posee un método de control de la memoria complejo, con funciones de almacenamiento dinámico y refresco; y necesita de un procesador adicional (DSP) que genere una secuencia de palabras de microcódigo correspondientes a la topología implementada.

3.4.2 CNAPS (Connected Network of adaptive Processors)

Este dispositivo es un procesador digital SIMD, es decir, trabaja por una secuencia de una instrucción y múltiples datos. Su arquitectura interna es un arreglo lineal de unidades de procesamiento conectadas a un único módulo

encargado de generar las secuencias de instrucciones. Este último está encargado también de manejar la entrada y salida de datos.

La representación de los datos es en punto fijo. Cada procesador posee un multiplicador de 9 x 16 bits, un sumador de 32 bits y 32 registros de almacenamiento local. CNAPS posee, internamente, 2 M de pesos de 1 bit y 256 K pesos de 8 bits. Con un reloj de 25 MHz es capas de realizar 1.600 Mc/s

Este dispositivo, en configuración de Perceptrón multicapa, ha sido utilizado eficazmente para implementar una versión del sistema NETtalk (Senowski, 1987) y demostró una rendimiento equivalente a 2000 veces el de una estación SUN-Sparc.

3.4.3 ETANN (Electrically Trainable Analog Neural Network)

El dispositivo neuronal Etann es, hasta la fecha, la implementación más avanzada realizada en VLSI orientada a redes neuronales. Este dispositivo, fabricado con tecnología CMOS de 1µ, integra 128 neuronas analógicas en dos capas de procesamiento en cascada. En el modo de procesamiento paralelo (PDP) cada neurona en cada capa computa en forma analógica la función

$$O_i = sigmoide\left(\sum_{j=1}^{64}\left(W_{i,j}.I_j\right)+\theta_i\right) \qquad (3.8)$$

donde $\mathbf{O_i}$ es una de las salidas analógicas de la red, $\mathbf{I_j}$ son las entradas analógicas, $\mathbf{W_{ij}}$ son los coeficientes de peso internos, y $\boldsymbol{\theta_i}$ los valores de umbral. La dimensión máxima del vector de entrada es 128 y la del vector de salida 64. Los 8192 coeficientes $\mathbf{W_{ij}}$, distribuidos en dos matrices de 64x64, y los 2048 umbrales $\boldsymbol{\theta_i}$ son reprogramables eléctricamente (Apéndice A).

La capacidad máxima de procesamiento del Etann, para un vector de entrada de dimensión 128 y de salida 64 es de unos 8K multiplicaciones y sumas en 1µ seg; es decir, más de 8 exp(9) multiplicaciones y sumas por segundo, con una precisión equivalente a 7 bits. La capacidad de procesamiento del Etann es

equivalente a unos 800 procesadores digitales trabajando a 10 MIPS, considerando que estos procesadores pueden ser acoplados en forma perfecta.

El Etann posee una función de activación continua de tipo sigmoidea que puede ser modificada a una de tipo umbral usando un modo de alta ganancia en el amplificador de salida. Este modo permite construir redes con salida binaria, por ejemplo, tipo Hopfield, con la ayuda de capacitores externos.

La figura 3.11 representa el diagrama en bloques del Etann. El dispositivo es fundamentalmente analógico, pero posee electrónica de control digital para la selección individual de sinapses, umbrales, y neuronas; y para el control de los diferentes modos de funcionamiento. Las señales lógicas externas para decodificación son compatibles TTL

En la figura se destacan dos arreglos matriciales de 64x64 que representan los coeficientes reprogramables **Wij** de cada una de las dos capas que son configurables en cascada para construir una red con capa oculta; y otros dos bloques correspondientes a los umbrales **θi** de 16x64 asociados a las matrices antes mencionadas. En la figura 3.1 puede verse en detalle la disposición del arreglo matricial respecto de las entradas. Las salidas de los multiplicadores son sumadas y procesadas para conformar el nodo de salida. Las señales de entrada, **Io-I63**, admiten un rango dinámico de 0 a 5volts. El rango dinámico de las salidas, **Oo-O63**, es de 0volt al doble del valor de una referencia externa.

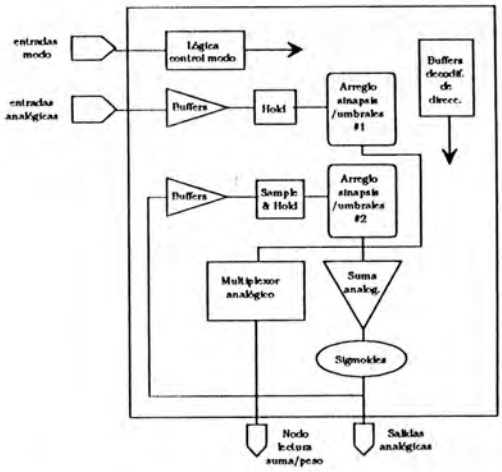

Figura 3.11 Diagrama en bloques

Puede verse en la figura 3.11 que existe una realimentación de las salidas hacia el segundo bloque matricula de pesos. Esto permite implementar una segunda capa de procesamiento dentro del mismo dispositivo. Todas las señales de entrada y realimentación pasan a través de *buffers* separadores y son almacenadas temporariamente (*holds*).

3.5 NEUROCOMPUTADORAS

Una computadora neuronal es, esencialmente, un arreglo paralelo de procesadores interconectados que operan en forma concurrente. Los procesadores neuronales son, en general, de la misma naturaleza, (analógicos, digitales o híbridos) y poseen la misma arquitectura interna. Las neurocomputadoras pueden estar destinadas a un propósito general u orientadas a una aplicación especifica. Las primeras permiten la computación de diversos algoritmos neuronales, redes de distinta configuración y número de capas; y de distintos algoritmos de entrenamiento. Las segundas son mucho menos flexibles, poseen, en general, una arquitectura fija o con pocas variantes, representan un solo modelo neuronal y poseen una interfaz sensorial adaptada para el problema especifico.

La complejidad de las neurocomputadoras es más variable aún que la de los dispositivos neuronales. El mínimo del rango de complejidad lo constituyen las computadoras secuenciales convencionales, simulando algoritmos de modelos neuronales. El máximo esta dado por supercomputadoras que disponen distintas formas de arreglos paralelos con una arquitectura orientada a la computación neuronal. La figura 3.12 representa el número de nodos de una neurocomputadora en función de la complejidad de los mismos, para el espectro existente.

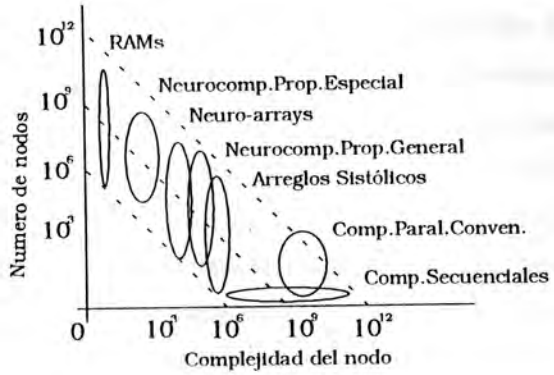

Figura 3.12 Diagrama de cantidad de nodos en función de la complejidad

Las computadoras neuronales son afectadas en forma directa por las propiedades de los modelos analógicos, digitales o híbridos descriptos anteriormente. En los últimos años ciertas lineas de investigación sobre computación paralela se han volcado al estudio de neurocomputadoras. Ambas ciencias comparten reglas comunes tales como múltiples unidades de cómputo, arquitectura de interconexión de estas unidades, compartición de la memoria, estructura de comunicación entre unidades de cómputo, sincronización, etc.

En especial, las neurocomputadoras de propósito general son bastante similares a las computadoras paralelas o supercomputadoras. Inclusive, se han desarrollado algunas neurocomputadoras utilizando procesadores convencionales tipo DSP (Morgan, 1992) o RISC (Treleaven, 1989). En general, cualquiera de estos procesadores posee una complejidad mayor que la requerida para ejecutar algoritmos neuronales. Una posibilidad es utilizar un subconjunto de la arquitectura de estos procesadores para disminuir la complejidad del sistema asociado, interconexión, compartición de recursos, etc.

Otra opción es la generación de arreglos sistólicos especialmente diseñados para modelos neuronales (Lehmann, 1993). Ambos casos aceptan distintos modelos neuronales y algoritmos de entrenamiento.

La neurocomputación de propósitos generales con procesadores digitales comparte también muchos de los inconvenientes de las computadoras paralelas, fundamentalmente, debido a que los procesadores no son lo suficientemente simples y compactos como para utilizar uno por sinapsis o por

neurona del modelo elegido. Cada procesador simula la operación de más de una unidad de procesamiento virtual, con lo cual, surgen problemas como el de asignación de tareas y cargas a los procesadores, granularización de los programas, etc.

También los procesadores analógicos pueden ser utilizados para el diseño de computadoras neuronales de propósitos generales. Si bien, se dijo, estas son menos flexibles en función de los modelos que pueden representar y del número de capas, neuronas y sinapsis que conforman una arquitectura dada. Si bien, aún la mayoría de la investigación en redes analógicas se encuentra a nivel del dispositivo, se consignan algunos desarrollos de neurocomputadora analógicas utilizando circuitos convencionales y/o *custom* (Fisher, 1991).

Aún no han aparecido en la literatura científica computadoras que utilizan dispositivos híbridos con la técnica de codificación de pulsos. Murray y Del Corso, se hallan abocados a un desarrollo de este tipo.

Dentro de los sistemas neuronales de propósito especial se destacan claramente los de Carver Mead del Instituto tecnológico de California (Caltech). Este grupo ha realizado varios desarrollos en el área de la bioingeniería, tales como una retina electrónica, una coclea, un sistema VLSI binaural, y un sensor de movimiento óptico (Mead, 1989).

En la sección siguiente, se presentan algunos ejemplos que reflejan el estado del arte de las neurocomputadoras, tanto de propósito general como aquellas orientadas a una aplicación en particular.

3.5.1 GENES: Arreglo sistólico bidimensional

El proyecto GENES utiliza la tecnologia CMOS digital para generar un arreglo sistólico bidimensional orientado a la computación de varios algoritmos neuronales. Cada dispositivo que integra el sistema de cómputo, implementa un arreglo bidimensional de 4x4 celdas como indica la figura 3.13a. La

arquitectura de la celda está orientada al cómputo de la función sináptica y también a la ecuación que determina la actualización del coeficiente de peso del algoritmo de aprendizaje. Esta computación, salvo para las redes que utilizan el algoritmo de retropropagación, es completamente local. La arquitectura de la celda se detalla en la figura 3.13b. Debido al mayor número de sinapses respecto del de neuronas, las celdas han sido orientadas al cómputo sináptico, dejando la función neuronal para ser almacenadas en tablas.

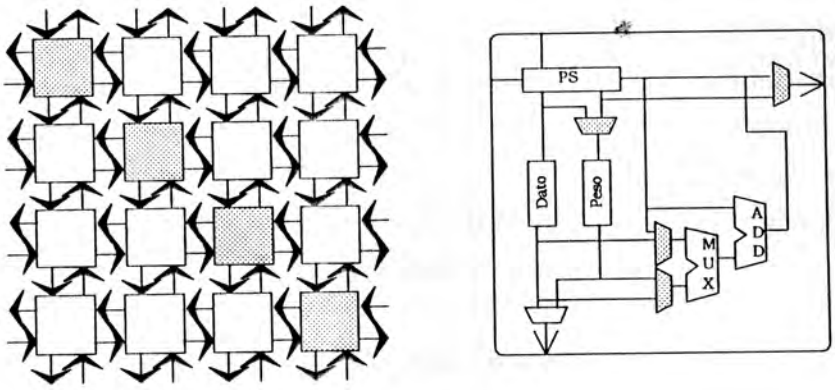

Figura 3.13 Arreglo sistólico a) arquitectura bidimensional b) esquema interno

La ecuación (3.9) representa el algoritmo de aprendizaje de Hebb como ejemplo de funcionamiento del arreglo sistólico. La figura 3.14a muestra como se introducen los vectores X e Y en el arreglo y circulan por el mismo en direcciones de norte a sur y de este a oeste, respectivamente. En régimen permanente pueden haber más de un patrón circulando por el arreglo a un dado tiempo, lo cual permite explotar toda la capacidad paralela del mismo. El algoritmo de ruta y cómputo de los patrones puede ser modificado para mejorar su rendimiento (Fig. 3.14b). El extremo derecho del arreglo realiza la función de ingreso del dato y de la computación de la función no lineal $\phi(.)$.

$$w_{ij}(n+1) = w_{ij}(n) + \eta x_j y_i \qquad \forall i, j \qquad (3.9)$$

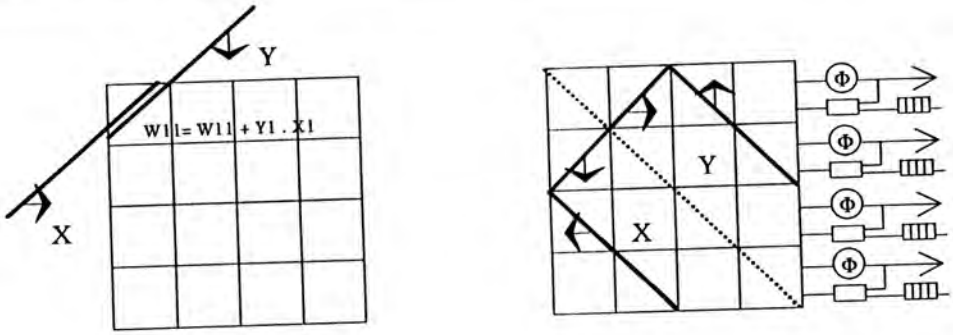

Figura 3.14 Algoritmo de ruta para la actualización de los pesos

El dato computado debe ser transmitido a los cuatro vecinos de cada celda. Estos datos deben ser almacenados dentro de cada celda, por lo cual se disponen cinco registros dedicados especialmente a cada uno de los parámetros almacenados. El corazón de la celda lo constituye un multiplicador-acumulador tipo MAC. La precisión de las operaciones es de 8 bits en punto fijo, para disminuir el consumo de área.

El rendimiento de este sistema ha sido probado para 16 neuronas, con una frecuencia de reloj de 10 MHz, obteniendo un promedio de 50 Mc/s.

3.5.2 PROCESADOR NEURONAL PROGRAMABLE:

El grupo de física aplicada de los laboratorios de investigaciones de la compañía Lockheed ha desarrollado un procesador neuronal analógico basado en el diseño *custom* de un resistor variable. Esta neurocomputadora posee 256 neuronas y 2048 sinapsis sobre 16 placas VME. El procesador permite computar algoritmos de modelos neuronales *feed-fordward* y Hopfield.

Cada neurona posee un amplificador operacional para la función no lineal y un dispositivo especial consistente en un conjunto de resistencias y llaves fabricado con tecnología *gate-array* analógico. La figura 3.15a muestra el esquema de las resistencias conmutadas por llaves bipolares programables. Ocho de estos grupos se integran en cada dispositivo custom. La figura 3.15b

esquematiza una neurona analógica que incluye uno de estos dispositivos junto al amplificador operacional.

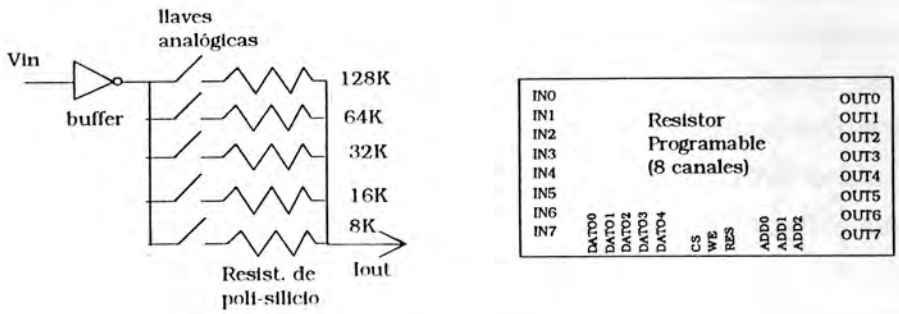

Figura 3.15 Sistema de resistencias conmutadas a) Esquema básico b) Dispositivo de 8 canales

El ancho de banda de esta neurocomputadora es de 90 KHz, y la resolución de 5 bits. Los pesos del sistema son controlados externamente por un procesador digital asociado. Si bien las entradas del sistema son puramente analógicas, se adicionó un conjunto de conversores digital a analógico con el fin de poder procesar patrones almacenados en la computadora.

Este procesador fue utilizado eficazmente en el control de lazo cerrado de un sistema de posicionamiento de espejos (Fisher, 1991).

3.5.3 RETINA ARTIFICIAL

La retina proyecta una imagen física sobre un arreglo de fotorreceptores generando la primera *imagen neural*. Esta imagen es posteriormente transmitida hacia el nervio óptico. C. Mead (1989), ha desarrollado una retina artificial. Este es un ejemplo de un sistema de neurocómputo especialmente dedicado a una aplicación. La retina electrónica consiste de: fotorreceptores, que computan el logaritmo de la señal de intensidad lumínica; celdas horizontales, conformadas por un arreglo bidimensional de resistores en formando hexágonos y que se encargan de distribuir y promediar la señal de entrada; las celdas bipolares, que producen una salida proporcional a la

diferencia entre la señal del fotorreceptor y la de las celdas horizontales. La figura 3.16 representa la disposición de las celdas horizontales. Se detalla además, el esquema de una celda bipolar con su fotorreceptor asociado.

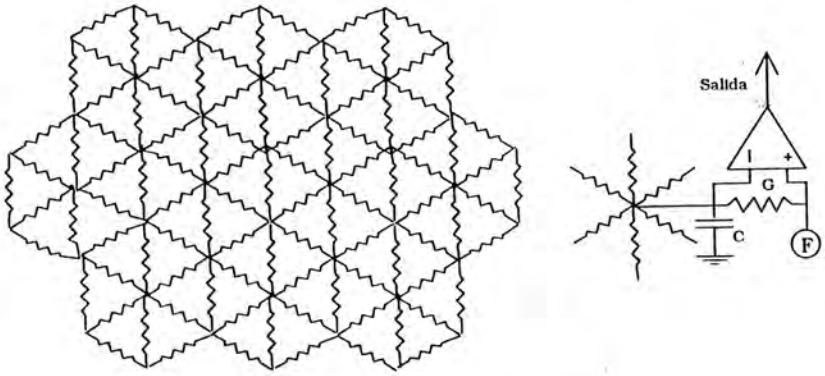

Figura 3.16 Retina artificial

Las señales de salida obtenidas por la retina artificial son muy similares a las reales, a velocidades electrónicas y con una alta tolerancia a imperfecciones del dispositivo, las cuales son características de los sistemas analógicos paralelos y masivos.

3.6 DISCUSION

En este capítulo se pone en evidencia las propiedades más importantes de los modelos analógicos, digitales e híbridos para la implementación de modelos neuronales, tanto a nivel de la celda neuronal mínima como a nivel del sistema neurocomputador. Se destaca en cada caso la influencia de los modelos analógicos y digitales en la precisión, ancho de banda, tipo de arquitectura, flexibilidad, almacenamiento de los parámetros, consumo de potencia, capacidad, comunicación e interconexión de las neuronas y el costo asociado al área del VLSI o a la integración del sistema.

A nivel del dispositivo, se han puesto de manifiesto los parámetros que redundan en las principales consideraciones que deben ser tenidas en cuenta en la implementación de modelos neuronales, tanto para ser utilizados como

parte constitutiva de neurocomputadoras de propósito general, como aquellas orientadas a una aplicación específica.

El análisis es fundamental para el desarrollo de esta tesis. Es obvio que no existe una única opción sobresaliente tanto para la generación de un VLSI neuronal, como para la de un sistema neuronal. De todas maneras, se ha optado por el modelo analógico tanto para el diseño de una celda neuronal como para el de una neurocomputadora. Las razones para esta decisión son:

Los dispositivos analógicos pueden integrar un número de neuronas mucho mayor en un mismo dispositivo.

Las celdas analógicas son compactas y tienen una relación uno a uno con las sinapsis y neuronas que computan.

Los modelos analógicos tienen una mayor proximidad con los modelos neuronales biológicos, en el modo en que realizan las funciones de procesamiento, en aspectos relacionados con el crecimiento potencial del sistema, consumo de potencia por neurona, crecimiento del interconexionado sináptico, etc.

Las celdas analógicas poseen un mayor ancho de banda, especialmente utilizando circuitos en el modo de corriente.

Aún cuando los modelos digitales poseen mayor resolución e inmunidad al ruido, estos aspectos quedan minimizados por el paralelismo masivo de unidades de cómputo de la red neuronal. Inclusive, un pequeño porcentaje de las neuronas y/o pesos sinápticos pueden estar fuera del rango de especificación, contaminados por ruido o poseer algún mal funcionamiento severo, sin afectar de forma determinante el funcionamiento global de la red. En este contexto, las neuronas analógicas saturan en forma natural a los valores de tensión de alimentación, mientras que en los dispositivos digitales debe controlarse estos límites así como el fenómeno de *overflow*.

Las neurocomputadoras analógicas pueden conectarse diréctamente a señales sensoriales y actuadoras externas.

Las lineas de investigación en computación neuronal analógica están menos exploradas y presentan un futuro más atractivo, especialmente para la aplicación de redes neuronales a problemas de tiempo real, tales como control, visión, reconocimiento de voz, y muchos otros.

En función de lo antedicho, este trabajo doctoral ha generado un modelo analógico a nivel neuronal. Este modelo será detalladamente descripto en el capítulo 4 de esta tesis y representa una neurona estática cuya función de activación es gaussiana. Se ha generado también una neurocomputadora analógica basada en el modelo perceptrón multicapa. Si bien, está orientada a una neurocomputadora de propósitos generales, es fácilmente adaptable a una aplicación específica y mantiene toda la potencialidad de cómputo intrínseca de dicha arquitectura. Los capítulos 5 y 6 de esta tesis reflejan lo hecho en este tema.