

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

DEPARTAMENTO DE INGENIERÍA ELÉCTRICA
SECCIÓN DE ELECTRÓNICA DEL ESTADO SÓLIDO

TÍTULO DE LA TESIS:

**PROCESAMIENTO Y CLASIFICACIÓN DE SEÑALES
UTILIZANDO LÓGICA DIFUSA**

QUE PARA OBTENER EL GRADO DE
DOCTOR EN CIENCIAS CON ESPECIALIDAD EN
INGENIERÍA ELÉCTRICA PRESENTA:

M. C. OLIVERIO ARELLANO CÁRDENAS *

ASESORES:

DR. JOSÉ ANTONIO MORENO CADENAS
DR. FELIPE GÓMEZ CASTAÑEDA

* Becario del CONACYT

Octubre del 2000

PROCESAMIENTO Y CLASIFICACION DE SEÑALES UTILIZANDO LÓGICA DIFUSA

OBJETIVO

Diseño y realización de un circuito integrado basado en la metodología de la lógica difusa para el procesamiento y clasificación de señales.

OBJETIVO ESPECIFICO

Diseño y realización de un circuito integrado, en tecnología CMOS, que realice las funciones correspondientes a las capas 1 y 2 de la arquitectura denominada: Sistema de Inferencias Difusas basado en Redes Adaptables (ANFIS), para el procesamiento y clasificación de señales.

AGRADECIMIENTOS

Es difícil saber con certeza cual es la contribución que de alguna forma tuvieron todas aquellas personas e instituciones quienes, en su momento, me dieron la instrucción, el apoyo y la motivación para poder llegar hasta este punto de mi trayectoria profesional y de mi vida en general. Tratando de ser justo con todos ellos, quiero agradecer:

- ✧ A mi madre, por todo el cariño, apoyo, comprensión y enseñanzas que siempre me ha dado.
- ✧ A mi hermana, por haber compartido conmigo grandes momentos y por su apoyo incondicional.
- ✧ A mis profesores y profesoras, quienes en las diferentes etapas de mi vida académica se esforzaron por transmitirme los conocimientos necesarios para enfrentar los problemas que se me pudieran presentar.
- ✧ A mis dos asesores, por las acertadas contribuciones, consejos e ideas que me proporcionaron, no sólo para la elaboración de este trabajo, sino en toda mi estancia de posgrado; y a quienes agradezco su sincera amistad.
- ✧ A los miembros de mi jurado, por sus valiosas observaciones, las cuales me permitieron realizar un mejor trabajo.
- ✧ A todos mis amigos y amigas en general, compañeros y compañeras (algunos de ellos, antiguos profesores), que de forma directa o indirecta me motivaron a seguir adelante, para concluir satisfactoriamente esta etapa de mi vida.
- ✧ A las instituciones (IPN, CONACYT, CINVESTAV-IPN, etc.) que proporcionaron los recursos materiales para poder realizar mi educación.
- ✧ A todos aquellos que, por omisión, no hubiera mencionado en los párrafos anteriores, y que también hayan contribuido de alguna forma a este logro.
- ✧ A la vida misma.

PREFACIO

En los últimos años (toda la década de los 90's), tanto en la comunidad científica como en el sector industrial, se ha acrecentado un enorme interés en la utilización de la combinación de los paradigmas de la *Lógica Difusa* y las *Redes Neuronales* en lo que se denomina *Sistemas Neurodifusos*. Así, este último concepto ha permitido aplicar las características de representación lingüística y bases del conocimiento, propiedades de los sistemas difusos, conjuntamente con las características de aprendizaje, almacenamiento y generalización de las redes neuronales, pudiéndose contar en la actualidad con sistemas poderosos de procesamiento de información para la resolución de problemas complejos de identificación y adaptabilidad.

La aplicación de técnicas neurodifusas a problemas de control en tiempo real, requiere de circuitos que lleven a cabo funciones de procesamiento en forma eficiente, lo cual resulta complicado de lograr con circuitos digitales. En la actualidad gran parte del procesamiento de información se realiza digitalmente; sin embargo, cuando se interactúa con el mundo real, se deben manejar entes analógicos, por lo cual normalmente se requieren interfaces A/D y D/A; esto hace que surja un interés creciente por desarrollar circuitos integrados analógicos, capaces de procesar en tiempo real dichas señales.

Dentro de este contexto, el objetivo principal de esta tesis es la construcción, caracterización y prueba de un sistema que opere en base a la metodología de los sistemas neurodifusos, utilizando circuitos analógicos que se puedan integrar en una estructura CMOS. La arquitectura que se propone para este fin se fundamenta en un *sistema de inferencias difuso basado en redes adaptables* (ANFIS) [1], originalmente desarrollado para ser programado y funcionar en una computadora (*software*); en otras palabras, se trata de un sistemas difuso que emplea técnicas de optimización propias de las redes neuronales para el ajuste de sus parámetros. Esta es una arquitectura que consta de cinco capas, cada una de las cuales realiza una función específica.

La presente tesis es parte de un proyecto, el cual abarca el desarrollo de un sistema funcional que incluye todas las capas de la arquitectura ANFIS para su aplicación en el procesamiento y clasificación de señales. La parte del proyecto correspondiente a este trabajo, contempla el diseño y construcción de los circuitos necesarios para conformar las dos primeras capas de esta arquitectura (capas 1 y 2), que corresponden a una etapa *fusificadora* y una de *evaluación de reglas difusas*.

Para el problema que se plantea en el título de la tesis (procesamiento y clasificación de señales), las redes neuronales se han utilizado ampliamente como una solución, la cual se basa en el aprendizaje a partir de ejemplos. Uno de los modelos más conocidos de redes neuronales es el de *retropropagación* (back-propagation) [2]. Entre las principales desventajas de este modelo se puede mencionar que él no asegura un aprendizaje exitoso ya que puede quedar atrapado en un mínimo local y, además, puede requerir de un gran tiempo de entrenamiento.

Otra metodología que utiliza redes neuronales probabilísticas (PNN) [3] tiene varias ventajas sobre la de retropropagación, ya que utiliza un algoritmo que corrige la mayoría de los problemas que presenta esa metodología. Sin embargo, el modelo de las PNN puede presentar algunos problemas cuando el conjunto de datos de entrenamiento es muy grande. Y, más aún, este modelo puede resultar inadecuado cuando hay incertidumbre en la información.

El sistema ANFIS constituye una propuesta de sistema neurodifuso que se basa en un sistema difuso de tipo Sugeno de 1er orden [4] (el cual tiene ventajas con respecto a otros sistemas difusos), y cuyo algoritmo de aprendizaje resuelve de manera muy eficiente los

problemas que presentan las redes arriba mencionadas, puesto que este sistema es el resultado de una comparación exhaustiva, realizada por el investigador que la desarrolló inicialmente [5], de distintos métodos de optimización. La implantación del modelo correspondiente a este sistema fue desarrollada, en sus inicios, en forma algorítmica, alcanzando tal importancia que, incluso, fue considerado para formar parte de las herramientas (*toolboxes*) del programa comercial Matlab, el cual es ampliamente reconocido en el ámbito científico.

Ya que la tendencia generalizada es la de contar con sistemas de procesamiento en tiempo real, se consideró, como una cuestión natural, el trasladar el sistema ANFIS a una versión de circuito integrado. A la fecha del término de esta tesis, el desarrollo de una arquitectura semejante ha sido realizado y publicado por otros grupos de investigación [6], [7], aunque ellos presentan una arquitectura empleando la aproximación Takagi-Sugeno de orden cero y, en el caso de [6], utilizan funciones de membresía triangulares y trapezoidales; en el caso de esta tesis, se ha desarrollado una aproximación Takagi-Sugeno de orden 1, la cual utiliza polinomios de salida y funciones de membresía tipo campana, permitiendo características de una mejor aproximación de funciones que las que permite un sistema de orden cero.

Dada la complejidad de realizar en *hardware* el sistema ANFIS y al mismo tiempo incluir características de adaptabilidad en línea, en este trabajo se procedió únicamente a la implantación de la arquitectura, lo cual representó en sí un reto considerable, realizando el entrenamiento de la red fuera de línea. Sin embargo, como se menciona en las conclusiones, en trabajos futuros, el incluir circuitos que permitan el aprendizaje en línea constituye también un problema abierto de investigación a mediano plazo.

El sistema ANFIS forma parte del *Toolbox* de Lógica Difusa del programa Matlab, el cual se ejecuta desde una computadora personal estándar (PC). Allí se efectúa el entrenamiento de ANFIS para obtener los valores optimizados de todos sus parámetros modificables, de acuerdo a los datos de un problema específico. Estos valores se toman como base para realizar los ajustes pertinentes en las variables del circuito (en forma de voltajes y corrientes) que definen la arquitectura. Las características de ANFIS lo hacen adecuado para usarse en problemas de predicción de valores (regresión no lineal), identificación de sistemas no lineales, clasificación de datos y procesamiento adaptable de señales.

Como ejemplo de aplicación del circuito diseñado, se propone su utilización para la detección de arritmias cardíacas, a partir de la forma que presenta el registro electrocardiográfico (ECG) de pacientes tomados de la base de datos MIT-BIH.

La composición de este escrito se divide en 7 capítulos, los cuales se explican brevemente a continuación:

El Capítulo 1 aborda los aspectos teóricos de lógica difusa que fundamentan la conformación de la arquitectura ANFIS, así como la descripción de la regla de aprendizaje de retropropagación, utilizada en las redes adaptables, que es la base del algoritmo de aprendizaje híbrido con el cual se realiza el entrenamiento de ANFIS. Debido a la gran extensión de la teoría de la lógica difusa, los tópicos aquí tratados se limitan a los conceptos y definiciones necesarios para entender el funcionamiento de un sistema de inferencias difuso del tipo Sugeno de 1er orden, en el cual se basa el sistema ANFIS desarrollado en este trabajo. La elección del modelo de Sugeno es debido a que ofrece un buen compromiso entre sencillez de los circuitos para construirlo y eficiencia de control.

La arquitectura ANFIS se describe en el Capítulo 2, en donde se detallan las funciones que se realizan en los nodos de cada capa de la arquitectura; se presenta una estructura de dos entradas, tres funciones de membresía por entrada y nueve reglas difusas, que es la que se

desarrolla en este trabajo, y que divide el espacio difuso en nueve regiones difusas; asimismo, se considera el desarrollo de la regla de aprendizaje híbrida, con la que se lleva a cabo la optimización de los parámetros de la arquitectura al realizar el entrenamiento.

Considerando las funciones que se deben efectuar en cada capa de la arquitectura, en particular las capas 1 y 2 que se desarrollan en este trabajo, en el Capítulo 3 se hace el diseño de los circuitos que llevan a cabo dichas funciones, todos ellos en tecnología CMOS. Para verificar su funcionamiento, tanto en forma individual como en conjunto, se realizan simulaciones eléctricas con el programa PSpice.

El Capítulo 4 presenta los diseños geométricos (layout) de los circuitos propuestos. Estos diseños se elaboran con ayuda del programa L-Edit, considerando las reglas de diseño de la compañía ORBIT para su tecnología de $2\mu\text{m}$, pozo N, doble metal y doble polisilicio. La estructura que contiene las dos capas de la arquitectura ANFIS propuesta, se construye en un área aproximada de silicio de $2\text{mm} \times 2\text{mm}$, en un encapsulado DIP de 40 terminales.

Una aplicación de los circuitos integrados fabricados se describe en el Capítulo 5. Esto se hace en conjunto con los circuitos construidos para la parte del proyecto que contempla las capas 3, 4 y 5 de ANFIS, y se propone su uso para clasificar arritmias cardíacas. Puesto que el universo de arritmias que pueden presentarse es muy amplio, el ejemplo se limita a identificar entre la condición normal y latidos ectópicos, producidos por una contracción ventricular prematura (PVC). Se emplea un método que consiste en obtener dos coeficientes de predicción lineal (LPC) y el valor cuadrático medio de cada complejo QRS de la señal ECG, y tomarlos como datos de entrenamiento y prueba para el sistema ANFIS.

En el capítulo 6 se describen las pruebas y mediciones hechas a los circuitos integrados fabricados, para verificar sus límites de operación en condiciones reales de funcionamiento. Se muestran comparaciones entre resultados medidos y simulados, para valorar el desempeño del sistema propuesto. Esto incluye simulaciones tanto a nivel eléctrico (hechas con Pspice) como a nivel funcional (hechas en Matlab).

Finalmente, en el Capítulo 7 se presentan las conclusiones de la tesis.

CONTENIDO

PREFACIO	i
CAPITULO I	
Introducción a los sistemas neurodifusos.	
1.1. Introducción	1
1.2. Sistemas difusos	2
1.2.1. Conceptos básicos	2
1.2.2. Funciones de membresía	7
1.2.3. Sistemas de inferencia difusos	10
1.2.4. Modelo difuso tipo Sugeno	13
1.3. Redes adaptables	16
1.3.1. Arquitectura	16
1.3.2. Regla de aprendizaje de retropropagación	18
1.4. Sistemas neurodifusos	22
1.5. Resumen	24
CAPITULO II	
La Arquitectura ANFIS.	
2.1. Introducción	25
2.2. Sistema de inferencias difuso basado en redes adaptables: ANFIS ..	25
2.2.1. La arquitectura ANFIS	25
2.2.2. Regla de aprendizaje híbrida	29
2.2.3. ANFIS y la regla de aprendizaje híbrida	32
2.3. Resumen	33
CAPITULO III	
Diseño eléctrico de celdas CMOS para la arquitectura ANFIS.	
3.1. Introducción	35
3.2. Diseño de celdas básicas	35
3.2.1. Circuito generador de campanas	36
3.2.2. Circuito para realizar la operación MIN	39
3.3. Simulación de celdas	41
3.3.1. Celdas básicas	41
3.3.2. Variación de temperatura	44
3.3.3. Capas 1 y 2 de ANFIS	45
3.4. Resumen	47
CAPITULO IV	
Diseño geométrico de los circuitos CMOS.	
4.1. Introducción	49
4.2. Diseño geométrico	50
4.2.1. Celdas básicas	51
4.2.2. Construcción de la arquitectura (capas 1 y 2)	52
4.2.3. Estructuras auxiliares	53

4.2.4. Circuito integrado completo	55
4.3. Resumen	57
 CAPITULO V	
Aplicación propuesta.	
5.1. Introducción	59
5.2. Principios de electrocardiografía	59
5.2.1. Antecedentes	59
5.2.2. Derivaciones	61
5.2.3. El electrocardiograma normal	62
5.2.4. Variaciones del electrocardiograma normal	65
5.3. Obtención de los datos de entrenamiento para el sistema ANFIS	69
5.3.1. Preprocesamiento de la señal ECG	69
5.3.2. Conjunto de datos de entrenamiento	72
5.3.3. Entrenamiento de ANFIS con los datos obtenidos	74
5.4. Resumen	77
 CAPITULO VI	
Resultados experimentales.	
6.1. Introducción	79
6.2. Pruebas y mediciones de celdas individuales	79
6.3. Pruebas y mediciones del sistema ANFIS completo	82
6.4. Mediciones y resultados de la aplicación propuesta	86
6.5. Resumen	90
 CAPITULO VII	
Conclusiones	91
Trabajo futuro	93
TPublicaciones en congresos	95
 APENDICES	
Apéndice A: Método descendiente abrupto	97
Apéndice B: Método de mínimos cuadrados	100
Apéndice C: Parámetros tecnológicos de Spice proporcionados por Mosis	105
 Referencias	 108
Bibliografía	109

CAPITULO 1

INTRODUCCION A LOS SISTEMAS NEURODIFUSOS

“ ... Tan sólo los objetos de la naturaleza están subordinados al tiempo, en tanto que son finitos; lo verdadero, la idea, el espíritu, es, por el contrario, *eterno*.”

(G. W. F. Hegel)

“No tiene término el Universo en parte alguna...
Ni bordes tiene, ni límite, ni fin.
Y no importa en qué parte del mundo te halles:
Estés donde estés, desde el sitio que ocupas,
infinito siempre será en todas sus direcciones.”
(Tito Lucrecio Caro, *Sobre la Naturaleza de las Cosas*)

1.1. INTRODUCCIÓN

En la actualidad, una de las diferencias básicas que existen entre computadoras y seres humanos tiene que ver con la utilidad práctica del material memorizado. Esta diferencia consiste en que, mientras una computadora almacena información que únicamente puede ser llamada de manera explícita por un programa, el ser humano “aprende”. En el hombre, la adquisición de nueva información modifica su comportamiento (o su forma de pensar), en la medida que le aporta nuevos criterios para la toma de decisiones, tanto a escala consciente, como inconsciente. El hombre debe su gran capacidad de adaptación, también llamada *plasticidad*, a su habilidad para aprender.

Para poder construir máquinas que tengan ciertas características de inteligencia es necesario el diseño de arquitecturas radical y conceptualmente diferentes a las que utilizan métodos de cálculo tradicionales. Esta es la filosofía presupuesta en el estudio de las llamadas *redes neuronales artificiales*, los *sistemas difusos* y, más recientemente, los denominados *sistemas neurodifusos*. Estas metodologías constituyen un concepto totalmente nuevo y diferente de la computación, el cual se basa en un modelo simplificado de la manera en que las neuronas se organizan en redes dentro del cerebro, así como la forma de razonamiento y toma de decisiones de los seres humanos.

Las redes neuronales artificiales se encuentran aún en sus primeras fases de desarrollo, sin embargo, se ha encontrado que tienen capacidades autónomas de aprendizaje, y que la información que almacenan es recuperable con base en su contenido. También se ha observado que estos sistemas pueden “reconocer” un patrón dado a partir de información incompleta o con mucho ruido. Todas éstas son características análogas a las de la mente humana.

Por otra parte, mucho del razonamiento humano involucra el uso de variables cuyos valores son conjuntos difusos. Esta observación es la base para el concepto de *variable lingüística*, esto es, una variable cuyos valores son definidos por palabras, en vez de números. El uso de estas variables representa un cambio significativo en el análisis de sistemas. Tanto las redes neuronales como los sistemas difusos, rompen con la tradición histórica del pensamiento que establece que el mundo puede ser caracterizado en forma rigurosa y sin ambigüedades, dividirlo en categorías y, entonces, manipular estas descripciones de acuerdo a reglas formales y precisas.

Recientemente ha habido un gran auge de la investigación en estas áreas y como consecuencia se han logrado grandes avances de carácter fundamental. Estos avances comienzan a verse reflejados en el plano tecnológico, ya que se empiezan a desarrollar computadoras que funcionan con base en estos principios. En estos momentos ya existen en el mercado algunos programas que operan en computadoras convencionales y que funcionan “imitando” estas nuevas arquitecturas. Estos programas son ejemplos modestos de lo que se tendrá en los años por venir.

1.2. SISTEMAS DIFUSOS

1.2.1. Conceptos básicos.

Uno de los primeros conceptos que se debe conocer para entender la lógica difusa es el de **conjunto difuso**. Básicamente, un conjunto difuso es un conjunto que no tiene límites claramente definidos o precisos. A diferencia de los conjuntos clásicos, en los conjuntos difusos la

transición de la pertenencia o no-pertenencia de un elemento a un cierto conjunto, es gradual, y esta transición está caracterizada por *funciones de membresía*, las cuales les dan a los conjuntos difusos flexibilidad para modelar expresiones lingüísticas empleadas cotidianamente.

Sea X una colección de objetos denotados genéricamente por x ; entonces, un conjunto difuso A en X se define como el siguiente conjunto de pares ordenados:

$$A = \{(x, \mu_A(x)) \mid x \in X\} \tag{1.1}$$

donde $\mu_A(x)$ se conoce como la **función de membresía** (abreviado FM) de A . La FM transforma cada elemento de X hacia un grado de membresía (o valor de membresía) entre 0 y 1.

Generalmente, X es llamado el **universo de discurso** o, simplemente, el universo, el cual puede consistir de objetos discretos (ordenados o no ordenados) o ser un espacio continuo.

Por ejemplo, el universo ordenado discreto definido por el número de hijos que una pareja puede desear tener, se expresa como: $X = \{0, 1, 2, 3, 4, 5, 6\}$. De esta forma, el conjunto $A =$ “número ideal de hijos en una familia” se puede describir como:

$$A = \{(0, 0.1), (1, 0.3), (2, 0.7), (3, 1), (4, 0.6), (5, 0.2), (6, 0.1)\}$$

donde los grados de membresía asignados, son medidas subjetivas.

Un ejemplo de universo continuo es el de los números reales, el cual se expresa como: $X = \mathbf{R}$. El conjunto difuso $B =$ “números cercanos a 5” se puede expresar de la siguiente manera:

$$B = \{(x, \mu_B(x)) \mid x \in X\},$$

donde $\mu_B(x)$ se puede definir como:

$$\mu_B(x) = \frac{1}{1 + (x - 5)^2}$$

En la figura 1.1 se muestran las FM correspondientes a los conjuntos difusos A y B arriba descritos.

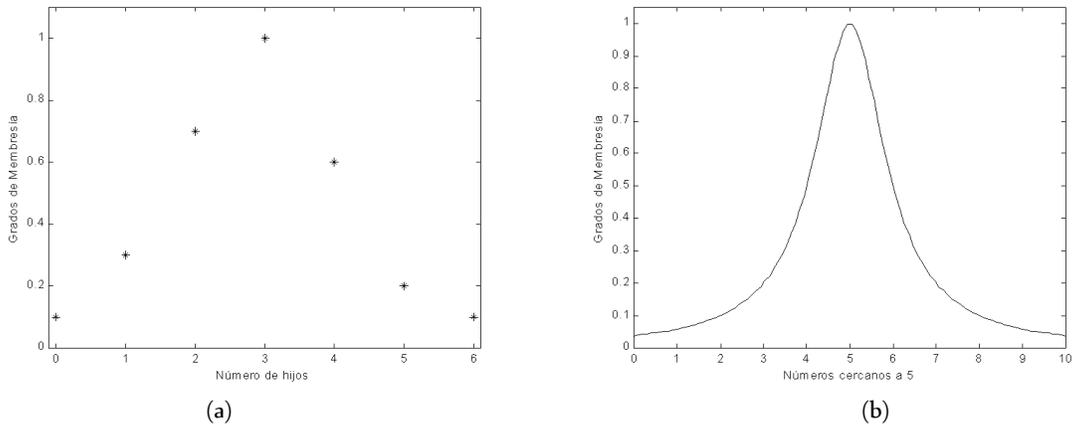


Fig. 1.1 Funciones de membresía para un universo discreto (a), y para un universo continuo (b).

En la práctica, cuando el universo de discurso X es un espacio continuo, generalmente se divide en varios conjuntos difusos cuyas FM cubren a X de una manera más o menos uniforme. Estos conjuntos se denominan *valores lingüísticos* o *etiquetas lingüísticas* y, normalmente, se les asignan nombres de adjetivos utilizados en el lenguaje común, tales como “grande”, “frío”, “alto”, etc.

Una **variable lingüística** se define mediante el quinteto de elementos siguiente: $(x, \mathcal{T}(x), X, G, M)$, donde x es el nombre de la variable; $\mathcal{T}(x)$, el conjunto de términos de x , esto es, el conjunto de sus valores lingüísticos o términos lingüísticos; X es el universo de discurso; G , una regla sintáctica que genera los términos en $\mathcal{T}(x)$; y, M , una regla semántica, la cual le asocia a cada valor lingüístico A un valor correspondiente $M(A)$, que denota un conjunto difuso en X .

Por ejemplo, si *temperatura* se interpreta como una variable lingüística, entonces su conjunto de términos $\mathcal{T}(\text{temperatura})$ podría ser:

$$\mathcal{T}(\text{temperatura}) = \{ \text{frío, no frío, muy frío, no muy frío, ...}, \\ \text{tibio, medio tibio, algo tibio, ...}, \\ \text{caliente, no caliente, muy caliente, no muy caliente, ...} \}$$

donde cada término en $\mathcal{T}(\text{temperatura})$ se caracteriza por un conjunto difuso de un universo de discurso que puede ser: $X=[0, 100]$. Normalmente se dice “la temperatura es fría” para denotar la asignación del valor lingüístico “frío” a la variable lingüística *temperatura*. Por otro lado, cuando la variable *temperatura* se interpreta como una variable numérica, se emplea la expresión “temperatura=30°C” para asignarle el valor numérico “30” a la variable numérica *temperatura*. La regla sintáctica se refiere a la forma en que se generan los valores lingüísticos en el conjunto de términos $\mathcal{T}(\text{temperatura})$. La regla semántica define la función de membresía de cada valor lingüístico del conjunto de términos.

La figura 1.2 muestra una distribución de algunas de las funciones de membresía propuestas para la variable *temperatura*.

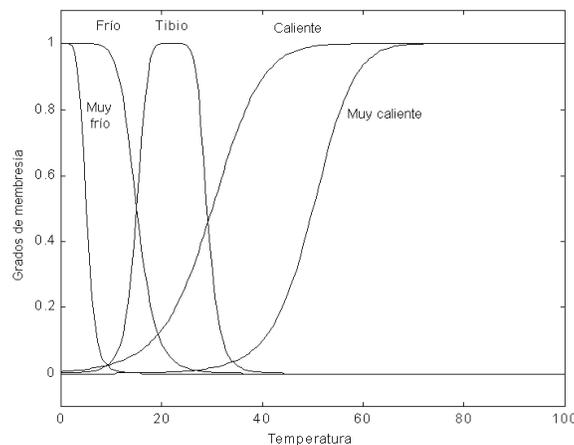


Fig. 1.2 Funciones de membresía típicas para el conjunto de términos $\mathcal{T}(\text{temperatura})$.

Del ejemplo previo se observa que el conjunto de términos se compone de varios *términos primarios* (frío, tibio, caliente), modificados por la *negación* (“no”) y/o algunos *modificadores* (muy, algo, un poco, demasiado, etc.), que se pueden enlazar con *conectivas* tales como *y, o, ni ... ni ...*. Las conectivas, los modificadores y la negación se consideran como operadores que cambian el significado de sus operandos en una forma específica, independiente del contexto.

Otros conceptos que comúnmente se emplean al hablar de conjuntos difusos se dan a continuación:

Soporte: El soporte de un conjunto difuso A es el conjunto de todos los puntos x en X , tal que $\mu_A(x) > 0$, esto es:

$$\text{soporte}(A) = \{x \mid \mu_A(x) > 0\} \quad (1.2)$$

Núcleo: El núcleo de un conjunto difuso A es el conjunto de todos los puntos x en X , tal que $\mu_A(x) = 1$:

$$\text{núcleo}(A) = \{x \mid \mu_A(x) = 1\} \quad (1.3)$$

Normalidad: Un conjunto difuso A es normal si su núcleo no está vacío, esto es, si se puede encontrar un punto $x \in X$, tal que $\mu_A(x) = 1$.

Puntos de cruce: El punto de cruce de un conjunto A es un punto $x \in X$, en el cual $\mu_A(x) = 0.5$

$$\text{cruce}(A) = \{x \mid \mu_A(x) = 0.5\} \quad (1.4)$$

Impulso difuso: Se refiere a un conjunto difuso cuyo soporte es un único punto en X con $\mu_A(x) = 1$.

Corte α : El corte α o *conjunto de nivel α* de un conjunto difuso A , es un conjunto certero, definido por:

$$A_\alpha = \{x \mid \mu_A(x) \geq \alpha\} \quad (1.5)$$

De igual forma, un *corte α fuerte* o *conjunto de nivel α fuerte* de un conjunto difuso A , es un conjunto certero, definido por:

$$A'_\alpha = \{x \mid \mu_A(x) > \alpha\} \quad (1.6)$$

Usando esta notación, el soporte y el núcleo de un conjunto difuso se pueden expresar como: $\text{soporte}(A) = A'_0$ y $\text{núcleo}(A) = A_1$, respectivamente.

Convexidad: Un conjunto difuso A es convexo si y sólo si para cualquier $x_1, x_2 \in X$ y cualquier $\lambda \in [0, 1]$, se cumple que:

$$\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \min\{\mu_A(x_1), \mu_A(x_2)\} \quad (1.7)$$

Asimismo, el conjunto A es convexo si todos sus cortes α son convexos.

Números difusos: Un número difuso A es un conjunto difuso que satisface las condiciones de normalidad y convexidad. Estos son los tipos más básicos de conjuntos difusos.

Ancho de banda: Para un conjunto normal y convexo, el ancho de banda o su anchura, se define como la distancia entre sus dos únicos puntos de cruce:

$$\text{anchura}(A) = |x_2 - x_1| \quad (1.8)$$

donde $\mu_A(x_1) = \mu_A(x_2) = 0.5$

Simetría: Un conjunto difuso A es simétrico si su FM es simétrica alrededor de un cierto punto $x=c$, de tal forma que:

$$\mu_A(c - x) = \mu_A(c + x) \text{ para toda } x \in X \quad (1.9)$$

Abierto por la derecha, abierto por la izquierda, cerrado: Un conjunto difuso A es abierto por la derecha si $\lim_{x \rightarrow +\infty} \mu_A(x) = 1$ y $\lim_{x \rightarrow -\infty} \mu_A(x) = 0$; abierto por la izquierda, si $\lim_{x \rightarrow +\infty} \mu_A(x) = 0$ y $\lim_{x \rightarrow -\infty} \mu_A(x) = 1$; y, cerrado, si $\lim_{x \rightarrow \pm\infty} \mu_A(x) = 0$.

Contenido o subconjunto: Un conjunto difuso A está *contenido* en un conjunto difuso B (o, equivalentemente, A es un subconjunto de B , o A es menor o igual que B) si y sólo si $\mu_A(x) \leq \mu_B(x)$ para toda x . Simbólicamente:

$$A \subseteq B \Leftrightarrow \mu_A(x) \leq \mu_B(x) \quad (1.10)$$

Con los conjuntos difusos se pueden efectuar operaciones similares a las que se realizan con conjuntos clásicos. Las tres operaciones básicas son la *unión*, la *intersección* y el *complemento*.

Unión (disyunción): La unión de dos conjuntos difusos A y B es un conjunto difuso C (escrito como $C = A \cup B$ o como $C = A \vee B$), cuya FM se relaciona con las de A y B por:

$$\mu_C(x) = \max(\mu_A(x), \mu_B(x)) = \mu_A(x) \vee \mu_B(x) \quad (1.11)$$

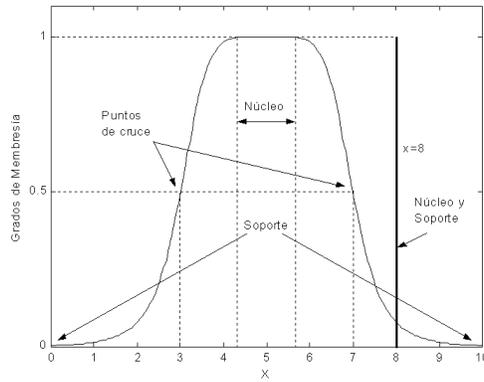
Intersección (conjunción): La intersección de dos conjuntos difusos A y B es un conjunto difuso C (escrito como $C = A \cap B$ o como $C = A \wedge B$), cuya FM se relaciona con las de A y B por:

$$\mu_C(x) = \min(\mu_A(x), \mu_B(x)) = \mu_A(x) \wedge \mu_B(x) \quad (1.12)$$

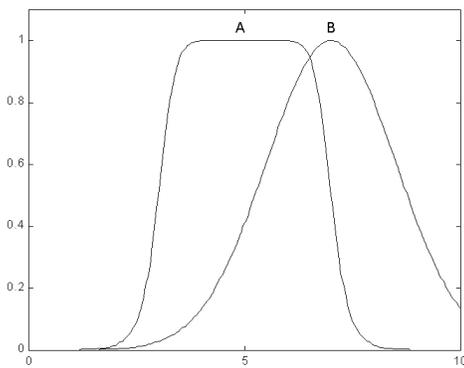
Complemento: El complemento de un conjunto difuso A , denotado por \bar{A} ($\neg A$, No A), se define como sigue:

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad (1.13)$$

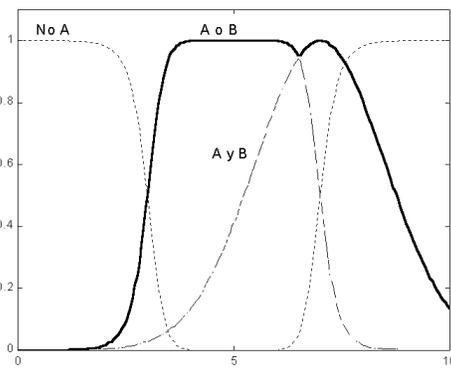
En la figura 1.3, se muestran gráficamente algunas de las definiciones anteriores.



(a)



(b)



(c)

Fig. 1.3 a) Núcleo, soporte y puntos de cruce de un conjunto difuso y de un impulso difuso en $x=8$; b) Dos conjuntos difusos A y B ; c) Los conjuntos difusos \bar{A} , $A \cup B$ y $A \cap B$.

1.2.2. Funciones de membresía.

Un conjunto difuso se caracteriza completamente por su FM. Una forma concisa y conveniente de definir una FM es expresándola como una fórmula matemática. A continuación se describirán algunas clases de funciones parametrizadas empleadas comúnmente para definir FM en una dimensión, esto es, que tienen un sola entrada. Para FM de un orden mayor se puede usar un razonamiento análogo.

Función de membresía triangular.

Una FM triangular se especifica mediante tres parámetros $\{a, b, c\}$, de la siguiente forma:

$$\text{triángulo}(x; a, b, c) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases} \quad (1.14)$$

Utilizando operadores min y max, la expresión anterior se puede expresar como sigue:

$$\text{triángulo}(x; a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \quad (1.15)$$

Los parámetros $\{a, b, c\}$ (con $a < b < c$) determinan las coordenadas en x de los tres vértices de la FM triangular en cuestión.

Función de membresía trapezoidal.

Una FM trapezoidal se determina con cuatro parámetros $\{a, b, c, d\}$, como sigue:

$$\text{trapecio}(x; a, b, c, d) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases} \quad (1.16)$$

También se puede definir esta FM en una forma más concisa, empleando los operadores min y max, como sigue:

$$\text{trapecio}(x; a, b, c, d) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{d-x}{d-c}\right), 0\right) \quad (1.17)$$

Los parámetros $\{a, b, c, d\}$ (con $a < b \leq c < d$) determinan las coordenadas en x de los cuatro vértices de la FM trapezoidal definida.

Debido a la sencillez de sus fórmulas, así como a su eficiencia computacional, las FM triangulares y trapezoidales han sido ampliamente utilizadas, especialmente en aplicaciones de tiempo real. Sin embargo, puesto que ambas funciones se componen de segmentos de líneas rectas, no son suaves en los puntos de los vértices especificados por los parámetros. Por ello, también se emplean FM definidas por funciones suaves y no lineales.

Función de membresía gaussiana.

Una FM gaussiana se especifica con dos parámetros $\{c, \sigma\}$, como se muestra a continuación:

$$\text{gauss}(x; c, \sigma) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \quad (1.18)$$

Una FM gaussiana se determina completamente con c y σ ; c representa el centro de la FM y σ determina su anchura.

Función de membresía tipo campana generalizada.

Una FM del tipo campana generalizada (o FM tipo campana) se caracteriza mediante tres parámetros $\{a, b, c\}$, como sigue:

$$\text{campana}(x; a, b, c) = \frac{1}{1 + \left| \frac{x - c}{a} \right|^{2b}} \quad (1.19)$$

donde c y a definen el centro y el ancho de la FM, respectivamente, mientras que el parámetro b controla las pendientes en los puntos de cruce. Este parámetro b normalmente es positivo; b negativo generaría una FM con forma de campana invertida. Esta FM es una generalización directa de la función de distribución de Cauchy definida en la teoría de probabilidad, por ello también es referida como FM de Cauchy.

Debido a su suavidad y notación concisa, las FM gaussiana y tipo campana, han llegado a ser cada vez más empleadas para especificar conjuntos difusos. Las funciones gaussianas son bien conocidas en probabilidad y estadística, y tienen propiedades útiles como aquella en la que no cambia su forma al multiplicarse por otra (el producto de dos gaussianas es una gaussiana con un factor de escalamiento) ni con una transformada de Fourier (la transformada de Fourier de una gaussiana es una gaussiana). La FM tipo campana tiene un parámetro más que la gaussiana, por lo cual tiene un grado más de libertad para ajustar sus pendientes en los puntos de cruce.

Aunque las FM gaussiana y de tipo campana logran una mayor suavidad, no permiten especificar FM asimétricas, las cuales se emplean en ciertas aplicaciones. Una FM sigmoideal puede ser abierta por la izquierda o por la derecha, y el producto o la diferencia absoluta de dos de ellas permite definir FM asimétricas. Una FM sigmoideal se define por:

$$\text{sig}(x; a, c) = \frac{1}{1 + \exp[-a(x - c)]} \quad (1.20)$$

donde a controla la pendiente en el punto de cruce c .

Otras FM se pueden crear para aplicaciones específicas, si ello se requiere. En general, cualquier tipo de función de distribución de probabilidad estadística puede emplearse como FM. Por ejemplo, si se requiere una FM con pendientes distintas en ambos de sus lados, se puede construir como sigue:

$$f(x; a, b, c) = \begin{cases} F_I\left(\frac{c-x}{a}\right), & x \leq c \\ F_D\left(\frac{x-c}{b}\right), & x \geq c \end{cases} \quad (1.21)$$

donde $F_I(x)$ y $F_D(x)$ son funciones monótonas decrecientes, definidas en $[0, \infty)$, tal que $F_I(0)=F_D(0)=1$ y $\lim_{x \rightarrow \infty} F_I(x)=\lim_{x \rightarrow \infty} F_D(x)=0$.

En la figura 1.4 se muestran las formas de las FM triangular, trapezoidal, gausiana y de Cauchy. Las FM que se muestran en la figura corresponden a las definidas por los siguientes valores: triángulo(x; 6, 7, 9), trapecio(x; 5, 6, 8, 10), gauss(x; 0.7, 3) y campana(1.5, 5, 3).

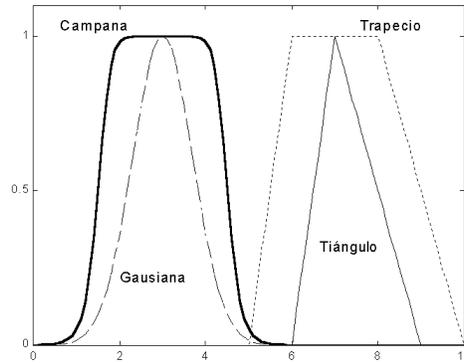


Fig. 1.4 Ejemplos de cuatro tipos de FM parametrizadas.

1.2.3. Sistemas de inferencias difusos.

Un sistema de inferencias difuso es una entidad de cómputo, basada en los conceptos de la teoría de conjuntos difusos, reglas difusas del tipo *si...entonces* y el razonamiento difuso [2]. Para entender esta idea, se definirán inicialmente algunos conceptos.

Operador norma-T.

Un operador norma-T es una función que se aplica a una pareja de valores, de la forma $T(*, *)$, y satisface las siguientes condiciones:

$$\begin{aligned}
 &T(0, 0) = 0, \quad T(a, 1) = T(1, a) = a \quad (\text{límites}) \\
 &T(a, b) \leq T(c, d) \quad \text{si } a \leq c \text{ y } b \leq d \quad (\text{monotonicidad}) \\
 &T(a,b) = T(b, a) \quad (\text{conmutatividad}) \\
 &T(a, T(b, c)) = T(T(a, b), c) \quad (\text{asociatividad})
 \end{aligned}
 \tag{1.22}$$

La primera condición establece la generalización para los conjuntos certeros; la segunda, implica que un decremento en los valores de membresía de A o B , no producirá un incremento en el valor de membresía de $A \cap B$. La tercera condición indica que el orden es indiferente al orden de los conjuntos difusos a ser combinados; la cuarta, permite tomar la intersección de cualquier cantidad de conjuntos, en cualquier orden, agrupándolos por parejas. Algunos de los operadores norma-T más comunes son:

$$\begin{aligned}
 \text{Minimum (MIN):} & \quad T_{\min}(a,b) = \min(a,b) = a \wedge b \\
 \text{Producto Algebraico :} & \quad T_{pa}(a,b) = ab \\
 \text{Producto Limitado :} & \quad T_{pl}(a,b) = 0 \vee (a+b-1) \\
 \text{Producto Drástico:} & \quad T_{pd}(a,b) = \begin{cases} a, & \text{si } b = 1 \\ b, & \text{si } a = 1 \\ 0, & \text{si } a, b < 1 \end{cases}
 \end{aligned} \tag{1.23}$$

Operador conorma-T (norma-S).

Un operador conorma-T (o norma-S) es una función que se aplica a un par de valores, de la forma $S(*, *)$, y satisface lo siguiente:

$$\begin{aligned}
 S(1, 1) = 1, S(a, 0) = S(0, a) = a & \quad (\text{límites}) \\
 S(a, b) \leq S(c, d) \text{ si } a \leq c \text{ y } b \leq d & \quad (\text{monotonicidad}) \\
 S(a, b) = S(b, a) & \quad (\text{conmutatividad}) \\
 S(a, S(b, c)) = S(S(a, b), c) & \quad (\text{asociatividad})
 \end{aligned} \tag{1.24}$$

La justificación de estos requerimientos es similar a la de los correspondientes para los operadores norma-T. En forma similar, se presentan a continuación cuatro operadores norma-S comunes:

$$\begin{aligned}
 \text{Maximum (MAX):} & \quad S_{\max}(a,b) = \max(a,b) = a \vee b \\
 \text{Suma Algebraica :} & \quad S_{sa}(a,b) = a + b - ab \\
 \text{Suma Limitada :} & \quad S_{sl}(a,b) = 1 \wedge (a + b) \\
 \text{Suma Drástica :} & \quad S_{sd}(a,b) = \begin{cases} a, & \text{si } b = 0 \\ b, & \text{si } a = 0 \\ 1, & \text{si } a, b > 0 \end{cases}
 \end{aligned} \tag{1.25}$$

Reglas difusas si ... entonces.

Una **regla difusa** del tipo **si ... entonces** (también llamada simplemente regla difusa, implicación difusa o declaración condicional difusa) asume la siguiente forma:

$$\text{si } x \text{ es } A \text{ entonces } y \text{ es } B \tag{1.26}$$

donde A y B son valores lingüísticos, definidos por conjuntos difusos en los universos de discurso X y Y , respectivamente. Frecuentemente a “ x es A ” se le llama el **antecedente** o **premisa**, mientras que a “ y es B ” se le llama el **consecuente** o **conclusión**. En el lenguaje común se tienen muchos ejemplos de reglas difusas de este tipo: Si la presión es alta, entonces el volumen es pequeño, si la velocidad es alta, entonces se aplican los frenos ligeramente, si la manzana es roja, entonces está madura, etc.

La expresión 1.26 se puede abreviar como $A \rightarrow B$. Esencialmente, esta expresión describe una relación entre dos variables (x, y) , lo cual sugiere que una regla difusa de este tipo se puede

definir como una relación binaria difusa R en el espacio difuso $X \times Y$. La expresión $A \rightarrow B$ puede interpretarse básicamente de dos formas: A está acoplada con B , o A implica B .

En el primer caso (A está acoplada con B) la relación se define como:

$$R = A \rightarrow B = A \times B = \int_{X \times Y} \mu_A(x) \tilde{*} \mu_B(y) / (x, y) \quad (1.27)$$

donde $\tilde{*}$ es un operador norma-T y $A \rightarrow B$ representa la relación difusa R . En el segundo caso (A implica B), se tienen los siguientes casos:

Implicación material: $R = A \rightarrow B = \neg A \cup B \quad (1.28)$

Cálculo propositivo: $R = A \rightarrow B = \neg A \cup (A \cap B) \quad (1.29)$

Cálculo propositivo

extendido: $R = A \rightarrow B = (\neg A \cap \neg B) \cup B \quad (1.30)$

Modus ponens

generalizado: $\mu_R(x, y) = \sup\{c \mid \mu_A(x) \tilde{*} c \leq \mu_B(y) \text{ y } 0 \leq c \leq 1\} \quad (1.31)$

donde $R = A \rightarrow B$ y $\tilde{*}$ es un operador norma-T.

Aunque todas estas fórmulas son diferentes en apariencia, todas ellas se reducen a la identidad $R = A \rightarrow B \equiv \neg A \cup B$ cuando A y B son proposiciones en el sentido de lógica bipolar (de dos valores). La figura 1.5 siguiente ilustra las dos interpretaciones de una regla difusa $A \rightarrow B$.

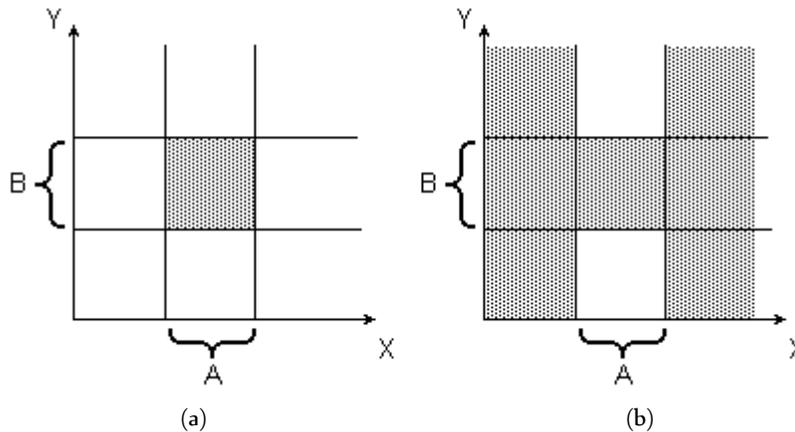


Fig. 1.5 Dos interpretaciones de la regla difusa $A \rightarrow B$: a) A acoplada con B ; b) A implica B .

Se pueden formular diferentes métodos calificados para calcular la relación difusa $R = A \rightarrow B$, basándose en las dos interpretaciones y considerando diferentes operadores norma-T y conorma-T. La relación R puede verse como un conjunto difuso con una FM bidimensional: $\mu_R(x, y) = f(\mu_A(x), \mu_B(y)) = f(a, b)$, con $a = \mu_A(x)$ y $b = \mu_B(y)$; la función f realiza la labor de transformar los grados de membresía de x en A y de y en B hacia los de (x, y) en $A \rightarrow B$.

Razonamiento difuso.

El razonamiento difuso es un procedimiento de inferencia que deriva conclusiones a partir de un conjunto de reglas difusas *si ... entonces* y hechos conocidos [8]. La regla básica de inferencia en lógica tradicional es el *modus ponens*, de acuerdo con el cual es posible inferir la

veracidad de una proposición B a partir de la veracidad de A y la implicación $A \rightarrow B$. Por ejemplo, si A se identifica con el hecho “la manzana es roja”, y B con “la manzana está madura”, entonces, si el hecho “la manzana es roja” es cierto, entonces el hecho “la manzana está madura” también será cierto.

Sean A , A' y B conjuntos difusos de X , X y Y , respectivamente. Si se asume que la implicación difusa $A \rightarrow B$ se expresa como la relación difusa R en $X \times Y$, entonces el conjunto difuso B inducido por “ x es A ” y la regla difusa “si x es A , entonces y es B ” queda definido por:

$$\begin{aligned}\mu_{B'}(y) &= \max_x \min[\mu_A(x), \mu_R(x, y)] \\ &= \vee_x [\mu_A(x) \wedge \mu_R(x, y)]\end{aligned}\tag{1.32}$$

O, equivalentemente: $B' = A' \circ R = A' \circ (A \rightarrow B)$

De esta forma, se puede utilizar el procedimiento de inferencia o razonamiento difuso para derivar conclusiones si la implicación difusa $A \rightarrow B$ se define como una relación binaria difusa apropiada.

Sistema de inferencias difuso.

En la literatura, estos sistemas han recibido diferentes nombres: Sistema basado en reglas difusas, sistema experto difuso, modelo difuso, memoria asociativa difusa, controlador lógico difuso o, simplemente, sistema difuso [9], [10].

Su estructura básica consiste de tres componentes: Una **base de reglas**, la cual contiene una selección de reglas difusas; una **base de datos** (o diccionario), el cual define las funciones de membresía empleadas con las reglas difusas; y, un mecanismo de razonamiento, el cual realiza el procedimiento de inferencia sobre las reglas y los hechos proporcionados, para producir una salida razonable o conclusión.

El sistema de inferencias difuso básico puede tener tanto entradas difusas como certeras (las cuales son vistas como impulsos difusos), pero las salidas que produce son casi siempre conjuntos difusos. En algunas situaciones se necesita una salida certera, por ejemplo, cuando el sistema se emplea como un controlador. Entonces, surge la necesidad de un método de **defusificación** para extraer el valor certero que represente mejor a un conjunto difuso.

Un sistema difuso con entradas y salidas certeras efectúa una transformación no lineal de su espacio de entrada hacia su espacio de salida. Esta transformación se lleva a cabo mediante un número de reglas difusas *si ... entonces*, cada una de las cuales describe el comportamiento local de la transformación. En particular, el antecedente de una regla define una región difusa en el espacio de entrada, mientras que el consecuente especifica la salida en la región difusa.

Existen diferentes tipos de sistemas de inferencia difusos, los cuales se han utilizado en diversas aplicaciones. Las diferencias básicas pueden estar en los consecuentes de sus reglas difusas, así como en los métodos de defusificación que emplean. Los nombres que se les han dado normalmente se toman de las personas que primero los propusieron; así, por ejemplo, se tiene el sistema de inferencias difuso tipo Mamdani [11], el de tipo Takagi-Sugeno-Kang (TSK) [4], [12] o el de tipo Tsukamoto [13].

El caso particular del que se ocupa el presente trabajo, se basa en un sistema de inferencias difuso tipo TSK (también llamado simplemente de tipo Sugeno). Por ello, a continuación se describe brevemente en qué consiste este sistema difuso, que será retomado nuevamente en el capítulo siguiente, al hacer la descripción de la arquitectura ANFIS.

1.2.4. Modelo difuso tipo Sugeno.

Este modelo fue propuesto inicialmente por Takagi, Sugeno y Kang [4], [12], en un esfuerzo por desarrollar un método sistemático para generar reglas difusas a partir de un conjunto dado de datos de entrada/salida. Una regla difusa típica en un modelo difuso tipo Sugeno tiene la forma: Si x es A y y es B , entonces $z = f(x, y)$, donde A y B son conjuntos difusos en la parte de antecedentes y $z = f(x, y)$ es una función certera en la parte de consecuentes. Normalmente $f(x, y)$ es un polinomio dependiente de las variables de entrada x y y , pero, en general, puede ser cualquier función que pueda describir apropiadamente la salida del modelo dentro de la región difusa especificada por el antecedente de la regla. Cuando $f(x, y)$ es un polinomio de primer orden, el sistema de inferencias difuso resultante es llamado un *modelo difuso de Sugeno de primer orden*; si f es una constante, entonces se tiene un *modelo difuso de Sugeno de orden cero*. Cabe mencionar que, aunque también se ha desarrollado un *modelo difuso de Sugeno de segundo orden*, esto es, $f(x, y)$ es un polinomio de segundo grado, el presente trabajo se basa en un sistema que emplea el modelo de 1er orden, aprovechando la facilidad de aplicar métodos de optimización a funciones lineales.

La salida de un modelo de Sugeno de orden cero es una función suave de sus variables de entrada, a condición de que las vecindades de sus FM en el antecedente tengan suficiente traslape. En otras palabras, el traslape de las FM en el antecedente determina la suavidad del comportamiento de entrada/salida resultante. El traslape de las FM en el consecuente no tiene un efecto decisivo en la suavidad.

En la figura 1.6 se muestra el procedimiento de razonamiento difuso para un modelo difuso de Sugeno de 1er orden.

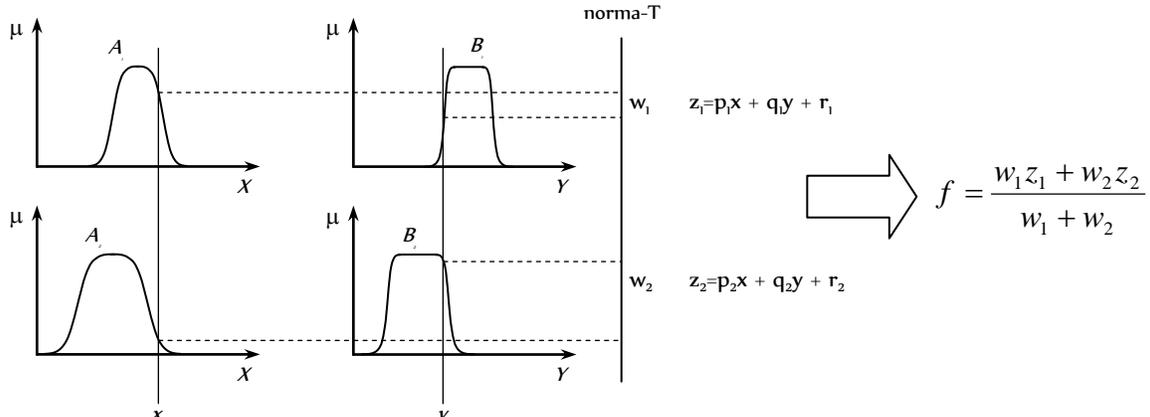


Fig. 1.6 El modelo difuso de Sugeno.

Puesto que cada regla tiene una salida certera, la salida global se obtiene mediante un *promedio ponderado*. En la práctica, la operación de promedio ponderado se sustituye con una *suma ponderada* (esto es, $f = w_1 z_1 + w_2 z_2$ en la figura anterior), a fin de reducir los cálculos, especialmente durante el entrenamiento de un sistema de inferencias difuso. Sin embargo, esta simplificación puede llevar a la pérdida de significado lingüístico de las FM, a menos que la suma de las intensidades de disparo (esto es, $\sum_i w_i$) sea cercana a la unidad.

Un ejemplo de un modelo difuso de Sugeno con dos entradas, una salida y cuatro reglas se expresa como:

Si X es pequeño y Y es pequeño, entonces $z = -x + y + 1$

Si X es pequeño y Y es grande, entonces $z = -y + 3$

Si X es grande y Y es pequeño, entonces $z = -x - 3$

Si X es grande y Y es grande, entonces $z = x + y + 2$

En un sistema de inferencias difuso, el antecedente de una regla define una región difusa local, mientras que el consecuente describe el comportamiento dentro de esa región mediante varios componentes. En el caso de un modelo de Sugeno, tales componentes pueden ser una constante (modelo de orden cero) o una ecuación lineal (modelo de 1er orden). El espacio difuso de entrada se puede particionar con diferentes métodos para formar los antecedentes de las reglas difusas. Los tres métodos básicos son: Partición de rejilla, partición de árbol y partición dispersa.

La partición de rejilla sólo requiere un pequeño número de FM para cada entrada; sin embargo, tiene algunos problemas cuando se tiene un número moderadamente grande de entradas. Por ejemplo, un modelo difuso con 10 entradas y 2 FM por cada una de ellas, podría resultar en $2^{10} = 1024$ reglas difusas, lo cual es prohibitivamente grande.

En la partición de árbol cada región se puede especificar en forma única a través de un árbol de decisiones correspondiente. Esta partición resuelve el problema de un incremento exponencial del número de reglas, pero se requieren más FM en cada entrada para definir tales regiones difusas, y esas FM generalmente no tienen un significado lingüístico claro tal como “pequeño”, “grande”, etc.

Con la partición dispersa se puede limitar el número de reglas a una cantidad razonable; sin embargo, esta partición normalmente se define por los pares de datos de entrada/salida deseados, debido a lo cual no se mantiene ortogonalidad en X , Y o en $X \times Y$. Esto hace difícil estimar la función de transformación global directamente de los consecuentes de cada salida de las reglas.

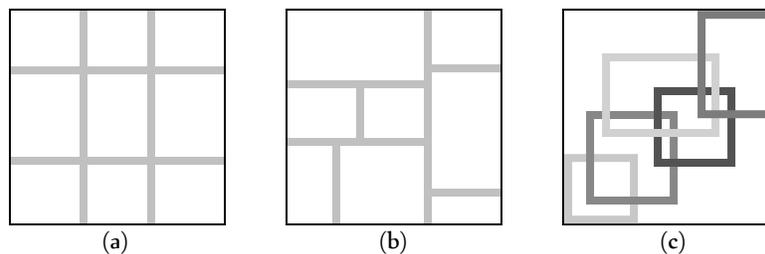


Fig. 1.7 Diferentes métodos de particionar el espacio de entrada: a) Partición de rejilla; b) partición de árbol; c) partición dispersa.

Con todos los elementos hasta aquí establecidos, se puede definir el proceso mediante el cual se construye un sistema de inferencias difuso para una aplicación específica. A este proceso normalmente se le llama **modelado difuso** y, conceptualmente, puede dividirse en dos etapas que no son totalmente independientes. La primera etapa es la identificación de la *estructura de la superficie*, la cual incluye las siguientes tareas:

- ❖ Seleccionar las variables de entrada y salida relevantes.
- ❖ Elegir un tipo específico de sistema de inferencias difuso.
- ❖ Determinar el número de términos lingüísticos asociados con cada variable de entrada y salida. (En un modelo de Sugeno, se debe determinar el orden de las ecuaciones del consecuente).

- ◇ Definir un conjunto de reglas difusas del tipo *si ... entonces*.

Con la primera etapa del modelado difuso, se obtiene una base de reglas que, por medio de términos lingüísticos, describe en forma aproximada el comportamiento del sistema que se quiere reproducir. El significado de estos términos lingüísticos se determina en la segunda etapa, *identificación de estructura lingüística*, la cual determina las FM de cada término lingüístico (y los coeficientes de cada polinomio de salida de regla si el modelo difuso empleado es tipo Sugeno). Esta etapa incluye las siguientes tareas:

- ◇ Elegir una familia adecuada de FM parametrizadas.
- ◇ Consultar humanos expertos del sistema que se desea reproducir, para determinar los parámetros de las FM usadas en la base de reglas.
- ◇ Refinar los parámetros de las FM usando regresión y técnicas de optimización.

Esta última tarea asume la disponibilidad de un conjunto de datos de entrada/salida deseado, mientras que las dos primeras parten de la disponibilidad de un humano experto. Las técnicas de optimización y adaptación expanden las aplicaciones de los sistemas de inferencia difusos a campos tales como el control adaptable, procesamiento de señal adaptable, regresión no lineal y el reconocimiento de patrones.

1.3. REDES ADAPTABLES

1.3.1. Arquitectura.

Como su nombre lo indica, una **red adaptable** es una estructura cuyo comportamiento global de entrada/salida queda determinado por una colección de parámetros modificables. Específicamente, la configuración de una red adaptable se compone de un conjunto de nodos conectados por enlaces direccionados, donde cada nodo realiza un tipo de *función de nodo estática* (por ejemplo, una función sigmoide) en las señales que entran a él, para generar una sola *salida de nodo*, y cada enlace especifica la dirección del flujo de señales de un nodo a otro. Generalmente, una función de nodo es una función parametrizada con parámetros modificables; cambiando dichos parámetros, se modifica la función de nodo y, en consecuencia, el comportamiento global de la red adaptable. La regla de aprendizaje especifica la forma en que estos parámetros deben ajustarse para minimizar una medida de error dada, la cual es una expresión matemática que mide la desviación entre la salida real de la red y una salida deseada.

Aquí se debe mencionar que si la regla de aprendizaje se aplica empleando pares de datos de entrada/salida, esto es, que para cada dato de entrada se le especifica a la red la respuesta que debe tener en su salida, se habla de una **regla de aprendizaje supervisado**, y se le denomina una red de aprendizaje supervisado (o, simplemente, red supervisada); por otro lado, si únicamente se utilizan vectores de entrada, se tendrá un método de aprendizaje sin supervisión, y se hablará de una red de aprendizaje no supervisado, la cual “evoluciona” para extraer características de regularidades en los patrones que se le presentan, sin que se le especifique qué salidas o clases asociadas con los patrones de entrada se desean, esto es, el sistema detecta o asigna categorías a características persistentes sin ninguna retroalimentación del medio. Puesto que el sistema empleado en esta tesis utiliza aprendizaje supervisado, el enfoque será en esa dirección.

La regla básica de aprendizaje de una red adaptable es el **método descendiente abrupto** (steepest descent method), en el cual el *vector gradiente* se deriva mediante invocaciones sucesivas de la regla de la cadena en el cálculo de derivadas de una función continua. Este es el método que se usó originalmente para hallar el gradiente en una red neuronal multicapa, bajo el nombre de *regla de aprendizaje de retropropagación* (backpropagation learning rule), nombre que actualmente resulta muy conocido en los tópicos de redes neuronales.

Los parámetros de una red adaptable se distribuyen hacia sus nodos, de tal forma que cada nodo tiene un conjunto local de parámetros. La unión de estos conjuntos locales de parámetros genera el conjunto global de parámetros de la red. Si el conjunto de parámetros de un nodo tiene elementos ajustables, entonces su función de nodo depende de los valores de los parámetros; en este trabajo se usará un cuadrado para representar este tipo de *nodo adaptable*. Por otra parte, si un nodo tiene un conjunto de parámetros sin elementos ajustables, entonces su función es fija y se representa con un círculo (*nodo fijo*). Los enlaces de una red adaptable se utilizan meramente para señalar la dirección de propagación de las salidas de los nodos; generalmente no hay pesos o parámetros asociados con ellos. La figura 1.8 ilustra lo dicho arriba para una arquitectura de dos entradas, tres nodos en una capa oculta y dos nodos en la capa de salida (abreviadamente: 2-3-2).

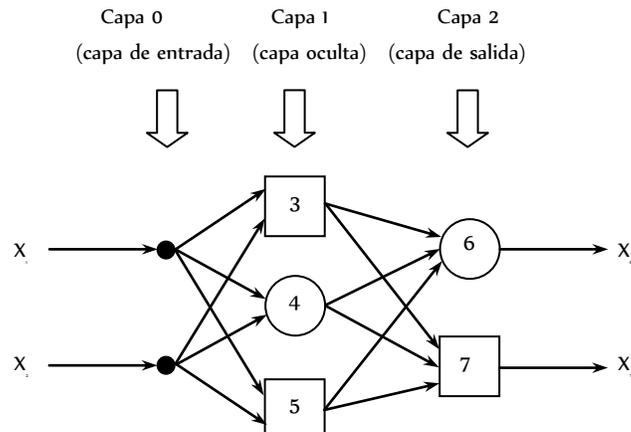


Fig. 1.8 Ejemplo de red adaptable 2-3-2.

Las redes adaptables generalmente se clasifican en dos categorías de acuerdo al tipo de enlaces que tienen: Redes de avance (feedforward) y redes recurrentes (recurrent). La figura 1.8 es un ejemplo de red de avance, puesto que la salida de cada nodo se propaga del lado de entrada (izquierda) hacia el lado de salida (derecha). Si existe algún enlace de retroalimentación que forme una trayectoria circular en una red, entonces la red es recurrente.

Conceptualmente, una red adaptable de avance realiza una transformación estática entre sus espacios de entrada y de salida. Esta transformación puede ser una relación lineal simple o una altamente no lineal, dependiendo de la estructura de la red y la función de cada nodo. El objetivo es lograr construir una red para realizar una transformación no lineal que esté regulada por un conjunto de datos de pares deseados de entrada/salida de un sistema que se quiera modelar. A este conjunto de datos normalmente se le llama el **conjunto de datos de entrenamiento**, y a los procedimientos que se siguen para ajustar los parámetros para optimizar el desempeño de la red, normalmente se les llama **reglas de aprendizaje** o **algoritmos de adaptación**. El desempeño de una red se determina midiendo el error entre la salida deseada y la salida de la red bajo las mismas condiciones de entrada. A esta variable se le llama la **medida del error** y puede tomar diferentes formas para diferentes aplicaciones. En general, una regla de

aprendizaje se deriva al aplicar una técnica de optimización específica a una medida de error dada.

1.3.2. Regla de aprendizaje de retropropagación.

En los apéndices A y B se explican en detalle dos métodos de optimización comunes para redes adaptables, los cuales forman la base de la **regla de aprendizaje híbrida**, que es el método de optimización utilizado para ajustar los parámetros de la arquitectura ANFIS (Adaptive Network-based Fuzzy Inference System: Sistema de inferencias difuso basado en redes adaptables) [1], la cual es la parte central de este trabajo de tesis. Dichos métodos son el descendiente abrupto y el **estimador de mínimos cuadrados** (Least Squares Estimator o LSE) [14]. A continuación se dará una breve semblanza del uso de la regla de aprendizaje descendiente abrupto en redes adaptables.

La parte central de esta regla de aprendizaje consiste en la obtención recursiva de un vector gradiente en el cual cada elemento se define como la derivada de una medida de error con respecto a un parámetro. Esto se hace por medio de la regla de la cadena, una fórmula básica para derivar funciones compuestas. El procedimiento para encontrar un vector gradiente en una estructura de red, generalmente se conoce como **retropropagación** (backpropagation), puesto que el vector gradiente se calcula en dirección opuesta al flujo de la salida de cada nodo. Una vez que se obtiene el vector gradiente, se utilizan técnicas de optimización y regresión basadas en derivadas para ajustar los parámetros.

Si se considera una red adaptable de avance que tiene L capas y la capa l (l=0, 1, ..., L; l=0 representa la capa de entrada) tiene N(l) nodos, entonces la salida y la función de nodo i (i=1, ..., N(l)) en la capa l se pueden representar como x_{li} y f_{li} , respectivamente, como se muestra en la figura 1.9. Sin perder generalidad, se asume que no hay saltos de enlaces (esto es, enlaces que conecten capas no consecutivas). Puesto que la salida de un nodo depende de las señales que entran a él, y de su conjunto de parámetros, se tiene la siguiente expresión general para la función de nodo f_{li} :

$$x_{li} = f_{li}(x_{l-1,p} \dots x_{l-1,N(l-1)}, \alpha, \beta, \gamma, \dots) \tag{1.33}$$

donde α, β, γ , etc. son los parámetros de este nodo.

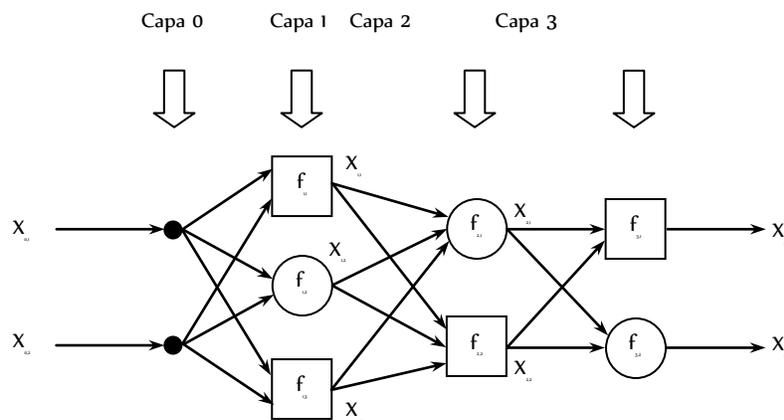


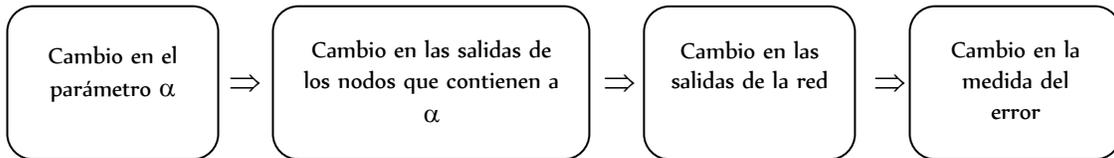
Fig. 1.9 Red de avance mostrando sus funciones de nodo.

Asumiendo que el conjunto de datos de entrenamiento dado tiene P entradas, se puede definir una medida del error para la p-ésima ($1 \leq p \leq P$) entrada de los datos de entrenamiento, como la suma de los errores cuadráticos:

$$E_p = \sum_{k=1}^{N(L)} (d_k - x_{L,k})^2 \quad (1.34)$$

donde d_k es el k-ésimo componente del p-ésimo vector de salida deseado y $x_{L,k}$ es el k-ésimo componente del vector de salida real, producido al presentar el p-ésimo vector de entrada a la red. Obviamente, cuando $E_p = 0$, la red es capaz de reproducir exactamente el vector de salida deseado en el p-ésimo par de datos de entrenamiento. Así, la tarea es minimizar una medida de error global, la cual se define como $E = \sum_{p=1}^P E_p$. En situaciones o aplicaciones específicas, es posible utilizar otras definiciones de E_p .

Si se desea emplear el método descendiente abrupto en la minimización de la medida del error, primero debe hallarse el vector gradiente. Para entender esto, se observa la siguiente relación causal:



donde la flecha \Rightarrow , indica una relación causal. En otras palabras, un cambio pequeño en un parámetro α afectará la salida del nodo que contiene al propio α ; a su vez, éste afectará la salida de la capa final y, por consiguiente, a la medida del error. Por lo tanto, el concepto básico al calcular el vector gradiente es propagar un tipo de información derivada, iniciando desde la capa de salida y retrocediendo capa por capa, hasta alcanzar la capa de entrada.

Si se define la *señal de error* $\epsilon_{l,i}$ como la derivada de la medida del error E_p con respecto a la salida del nodo i en la capa l , tomando en consideración tanto las trayectorias directas como las indirectas, se puede expresar como sigue:

$$\epsilon_{l,i} = \frac{\partial^+ E_p}{\partial x_{l,i}} \quad (1.35)$$

La expresión anterior se conoce como la **derivada ordenada** [15]. La diferencia de ésta con una derivada parcial ordinaria recae en la forma en que se considera la función a ser diferenciada. Para un nodo de salida interno $x_{l,i}$ (donde $l \neq L$), la derivada parcial $\partial E_p / \partial x_{l,i}$ es igual a cero, ya que E_p no depende directamente de $x_{l,i}$. Sin embargo, es evidente que E_p sí depende de $x_{l,i}$ (indirectamente), puesto que un cambio en $x_{l,i}$ se propagará, a través de alguna trayectoria indirecta hacia la capa de salida produciendo un cambio correspondiente en el valor de E_p . Por lo tanto, se puede ver a $\epsilon_{l,i}$ como la razón de esos dos cambios, cuando ambos se hacen infinitesimales.

Como ejemplo para demostrar la diferencia entre una derivada ordinaria y una ordenada, se considerará la red mostrada en la figura 1.10, en donde z es una función de x y y , a su vez, y es una función de x :

$$\begin{aligned} z &= g(x, y) \\ y &= f(x) \end{aligned}$$

Para la derivada parcial ordinaria $\partial z / \partial x$, se asume que todas las demás variables de entrada son constantes (en este caso, y):

$$\frac{\partial z}{\partial x} = \frac{\partial g(x, y)}{\partial x}$$

Esto significa que las entradas x y y a la función g se consideran independientes, sin considerar que, de hecho, y es una función de x .

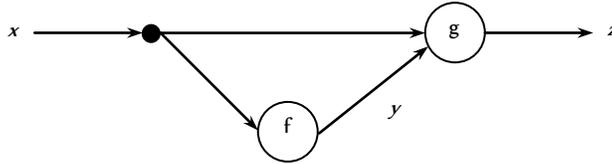


Fig. 1.10 Red para comparar las derivadas ordenadas y las derivadas parciales.

Para la derivada ordenada, se toma en cuenta esa relación causal, lo cual significa que se considera tanto la trayectoria directa como las indirectas de dicha relación, de tal forma que:

$$\begin{aligned} \frac{\partial^+ z}{\partial x} &= \frac{\partial g(x, f(x))}{\partial x} \\ &= \left. \frac{\partial g(x, y)}{\partial x} \right|_{y=f(x)} + \left. \frac{\partial g(x, y)}{\partial y} \right|_{y=f(x)} \frac{\partial f(x)}{\partial x} \end{aligned}$$

La señal de error para el i -ésimo nodo de salida (en la capa L) se puede calcular directamente, considerando la ec. 1.35 para el caso de $l=L$:

$$\epsilon_{L,i} \frac{\partial^+ E_p}{\partial x_{L,i}} = \frac{\partial E_p}{\partial x_{L,i}} \tag{1.36}$$

Lo anterior es igual a $\epsilon_{l,i} = -2(d_i - x_{L,i})$ si E_p es como en la ec. 1.34. Para el nodo interno en la i -ésima posición de la capa l , se puede derivar la señal de error por la regla de la cadena:

$$\epsilon_{l,i} = \frac{\partial^+ E_p}{\partial x_{l,i}} = \sum_{m=1}^{N(l+1)} \frac{\partial^+ E_p}{\partial x_{l+1,m}} \frac{\partial f_{l+1,m}}{\partial x_{l,i}} = \sum_{m=1}^{N(l+1)} \epsilon_{l+1,m} \frac{\partial f_{l+1,m}}{\partial x_{l,i}} \tag{1.37}$$

\uparrow
 Señal de error en la capa l

\uparrow
 Señal de error en la capa $l+1$

donde $0 \leq l \leq L-1$. Esto es, la señal de error de un nodo interno en la capa l se puede expresar como una combinación lineal de la señal de error presente en los nodos de la capa $l+1$. Por lo tanto, para cualquier l e i ($0 \leq l \leq L$ y $1 \leq i \leq N(l)$), es posible encontrar $\epsilon_{l,i} = \partial^+ E_p / \partial x_{l,i}$ aplicando inicialmente la ec. 1.36 una vez para obtener las señales de error en la capa de salida y, entonces, aplicando la ec. 1.37 iterativamente, hasta alcanzar la capa l deseada. Este proceso es el llamado de retropropagación, puesto que las señales de error se obtienen secuencialmente yendo de la capa de salida hacia la de entrada.

El vector gradiente se define como la derivada de la medida del error con respecto a cada parámetro, así que se requiere aplicar nuevamente la regla de la cadena para encontrar al vector gradiente. Si α es un parámetro del i -ésimo nodo en la capa l , se tiene que:

$$\frac{\partial^+ E_p}{\partial \alpha} = \frac{\partial^+ E_p}{\partial x_{l,i}} \frac{\partial f_{l,i}}{\partial \alpha} =_{\in_{l,i}} \frac{\partial f_{l,i}}{\partial \alpha} \quad (1.38)$$

Si diferentes nodos comparten el parámetro α , la ec. 1.38 se debe cambiar a una forma más general:

$$\frac{\partial^+ E_p}{\partial \alpha} = \sum_{x^* \in S} \frac{\partial^+ E_p}{\partial x^*} \frac{\partial f^*}{\partial \alpha} \quad (1.39)$$

donde S es el conjunto de nodos que contienen a α como un parámetro; x^* y f^* son, respectivamente, la salida y la función de un nodo genérico en S . La derivada de la medida del error global E con respecto a α es:

$$\frac{\partial^+ E}{\partial \alpha} = \sum_{p=1}^P \frac{\partial^+ E_p}{\partial \alpha} \quad (1.40)$$

De acuerdo con lo anterior, la fórmula de actualización del parámetro general α para el método descendente abrupto simple es:

$$\Delta \alpha = -\eta \frac{\partial^+ E}{\partial \alpha} \quad (1.41)$$

donde η es la **razón de aprendizaje**, la cual se puede expresar como [10]:

$$\eta = \frac{\kappa}{\sqrt{\sum_{\alpha} \left(\frac{\partial E}{\partial \alpha} \right)^2}} \quad (1.42)$$

en donde κ es el **tamaño del paso**, esto es, la dimensión de cada transición a lo largo de la dirección del gradiente en el espacio de parámetros. Generalmente el tamaño del paso se puede cambiar para variar la velocidad de convergencia [1].

Existen dos paradigmas de aprendizaje para satisfacer las necesidades de diferentes aplicaciones. Uno es el **aprendizaje fuera de línea** (o aprendizaje por lote), en el cual la fórmula de actualización de los parámetros α se basa en la ec. 1.40, y la acción de actualización tiene lugar solamente hasta que todo el conjunto de datos de entrenamiento se ha presentado a la red, esto es, después de cada **época** o **barrido**. Por otro lado, en el **aprendizaje en línea** (o aprendizaje patrón por patrón), inmediatamente después de que cada par de entrada/salida se presenta a la red, se actualizan los parámetros, y la fórmula de actualización se basa en la ec. 1.38. En la práctica, es posible combinar estos dos modos de aprendizaje y actualizar los parámetros después de que se han presentado k datos de entrenamiento, donde el valor de k está entre 1 y P y, algunas veces, es referido como el **tamaño de la época**. Estos dos paradigmas de aprendizaje forman parte de la *regla de aprendizaje híbrido*, que será explicada en el siguiente capítulo.

1.4. SISTEMAS NEURODIFUSOS

Hasta este punto ya se han descrito todos los elementos necesarios para construir un sistema neurodifuso propiamente dicho. En los sistemas puramente difusos, el problema de encontrar las FM apropiadas así como las reglas difusas para un problema específico, frecuentemente es un proceso tedioso de prueba y error. Por lo tanto, la idea de aplicar algoritmos de aprendizaje a estos sistemas surgió inicialmente con los llamados *controladores difusos auto-organizativos*. Sin embargo, otra posibilidad más aceptada actualmente para el ajuste de los parámetros de los sistemas difusos surge de los métodos de aprendizaje empleados en redes neuronales.

Los sistemas neurodifusos modernos normalmente se representan como una red neuronal multicapa de avance. Esto se hace simplemente para facilitar el análisis en el momento de aplicar el proceso de optimización e ilustrar el flujo de datos dentro del sistema, así como su naturaleza en paralelo. Los nodos en estas redes difieren de los utilizados en las redes neuronales en cuanto al tipo de función que realizan y la interpretación que se les da.

Aunque existen muchos métodos diferentes, el término de *sistemas neurodifusos* se puede restringir a aquellos sistemas que presentan las siguientes propiedades:

- ✧ Un sistema neurodifuso es un sistema difuso que se entrena por medio de un algoritmo de aprendizaje (comúnmente) derivado de la teoría de las redes neuronales. El procedimiento de aprendizaje opera en información local, y origina solo modificaciones locales en el sistema difuso de base. El proceso de aprendizaje no se basa en conocimiento, sino en un conjunto de datos dado.
- ✧ Un sistema neurodifuso puede verse como una red neuronal multicapa de avance. Las unidades en sus nodos emplean operaciones norma-T o conorma-T en lugar de las funciones de activación que normalmente usan las redes neuronales. La primera capa representa las variables de entrada; las FM y las reglas difusas se representan mediante enlaces y capas ocultas; y, la última capa, representa las variables de salida. Sin embargo, esta representación en forma de red neuronal no es un requisito indispensable para aplicar un procedimiento de aprendizaje, simplemente se hace por conveniencia.
- ✧ Un sistema neurodifuso siempre puede interpretarse como un sistema de reglas difusas, esto es, antes, durante y después del entrenamiento. Es posible crear el sistema a partir de datos de entrenamiento únicamente, o se pueden inicializar sus parámetros mediante un conocimiento previo en forma de reglas difusas.
- ✧ El procedimiento de entrenamiento de un sistema neurodifuso toma en consideración las propiedades semánticas del sistema difuso base. Esto resulta en restricciones en las posibles modificaciones de los parámetros del sistema.
- ✧ Un sistema neurodifuso aproxima una función de n dimensiones (desconocida) que se define parcialmente mediante los datos de entrenamiento. Las reglas difusas codificadas dentro del sistema representan muestras vagas, y se pueden interpretar como prototipos vagos de los

datos de entrenamiento. Un sistema neurodifuso no debe entenderse como un tipo de sistema (difuso) experto, y no se relaciona con la lógica difusa en un sentido estrecho.

Por lo tanto, un sistema neurodifuso se considera como una técnica para ajustar a un sistema difuso a partir de datos, o para enriquecerlo mediante aprendizaje, a partir de ejemplos.

En el capítulo siguiente se describirá en detalle el sistema ANFIS, el cual se basa en un sistema difuso de tipo Sugeno de 1er orden, y utiliza una regla de aprendizaje híbrida para el ajuste de sus parámetros. Esta regla, como ya se mencionó se basa en el método descendiente abrupto y el estimador de mínimos cuadrados.

1.5. RESUMEN

El presente capítulo da una introducción a las definiciones básicas, notaciones y operaciones de los conjuntos difusos, las diferentes representaciones de sus funciones de membresía, las operaciones básicas de la teoría de conjuntos (O, Y, NO) y algunos tipos de operadores avanzados de conjuntos difusos (norma-T y conorma-T).

La interpretación de las reglas difusas *si ... entonces* como relaciones difusas, deriva en varios esquemas de razonamiento difuso, lo cual es la base de los sistemas de inferencia difusos. Dentro de los sistemas de inferencia difusos, se describió el sistema de tipo Sugeno, el cual emplea una función dependiente de sus variables de entrada en la parte de los consecuentes. Asimismo, se explicaron diferentes formas de particionar el espacio difuso.

Finalmente, se describió la arquitectura y el proceso de aprendizaje de una red adaptable, como base para realizar la optimización de parámetros de un sistema difuso, a fin de poder construir un sistema neurodifuso. Puesto que todo esto es un campo muy amplio de estudio, el análisis se centró en aquellos elementos que se requieren para entender el funcionamiento de la arquitectura ANFIS.

CAPITULO 2

LA ARQUITECTURA ANFIS

“Una inteligencia que conociese todas las fuerzas que en un momento dado animan a la naturaleza, así como la posición relativa de todas sus partes integrantes, y si, además, fuese lo suficientemente amplia para someter esos datos al análisis, podría abarcar en una fórmula los movimientos de los cuerpos más grandes del Universo y los movimientos de los átomos más ligeros: Nada quedaría incierto para ella, y el futuro, lo mismo que el pasado, se extendería ante su vista. La perfección que la mente humana ha sabido proporcionar a la astronomía nos da una débil idea de lo que puede ser una inteligencia semejante.”

(P. S. Laplace, *Ensayo Filosófico sobre la Teoría de la Probabilidad*)

2.1. INTRODUCCION

Desde que Zadeh publicó en 1965 el primer artículo sobre una forma novedosa de caracterizar incertidumbres no probabilísticas [16], mediante los llamados *conjuntos difusos*, la lógica difusa y la teoría de conjuntos difusos ha evolucionado hacia un área que involucra varias disciplinas, tales como cálculo de reglas difusas *si ... entonces*, grafos difusos, interpolación difusa, topología difusa, razonamiento difuso, sistemas de inferencias difusos y modelado difuso. Las aplicaciones que, por naturaleza son multidisciplinarias, incluyen control automático, electrónica de consumo, procesamiento de señales, predicción de series de tiempo, recuperación de información, manejo de bases de datos, visión por computadora, clasificación de datos, toma de decisiones, etc.

Recientemente, el resurgimiento del interés en el campo de las redes neuronales artificiales ha dado una nueva perspectiva encaminada hacia la literatura difusa. La regla de aprendizaje de retropropagación, la cual tuvo poca atención hasta que se descubrió su aplicación en redes neuronales, es en realidad un paradigma universal de aprendizaje para cualquier modelo parametrizado "suave", incluyendo los sistemas de inferencia difusos. Como resultado, los sistemas difusos no solo pueden tomar información lingüística (reglas lingüísticas) de humanos expertos, sino que también se pueden adaptar a sí mismos, empleando datos numéricos (pares de datos de entrada/salida) para lograr un mejor desempeño.

Los sistemas neurodifusos surgen debido a la falta de métodos estándar para transformar el conocimiento humano o su experiencia hacia la base de reglas y datos de un sistema de inferencias difuso. Para lograr tal objetivo, se requiere de métodos efectivos para ajustar sus funciones de membresía, así como minimizar la medida del error de salida o maximizar su índice de desempeño.

Con esta idea en mente, surgió la arquitectura llamada ANFIS (Sistema de Inferencias Difuso basado en Redes Adaptables), la cual sirve como base para construir un conjunto de reglas difusas *si ... entonces* con funciones de membresía apropiadas para generar los pares de datos de entrada/salida óptimos. Esto se logra usando una regla de aprendizaje híbrida, basada en los métodos de optimización comunes en las redes adaptables (el descendiente abrupto y el estimador de mínimos cuadrados).

2.2. SISTEMA DE INFERENCIAS DIFUSO BASADO EN REDES ADAPTABLES: ANFIS

2.2.1. La arquitectura ANFIS.

La arquitectura que se propone para el desarrollo del presente trabajo, es un tipo de red adaptable, la cual, funcionalmente, es equivalente a un sistema de inferencias difuso. Esta arquitectura puede representar tanto modelos difusos de Sugeno [4] (de orden 1 y 0) como de Tsukamoto [13] y, bajo algunas restricciones, puede operar como una *red de función de base radial* (RBFN) [1]. Básicamente, ANFIS modela un sistema de inferencias difuso en el cual sus parámetros se ajustan mediante un algoritmo de retropropagación basándose en un conjunto de datos de entrada/salida (datos de entrenamiento), lo cual le permite al sistema aprender.

El enfoque de este trabajo es hacia ANFIS operando como un sistema de tipo Sugeno de ter orden, por lo que la siguiente descripción se centrará en ello, teniendo en cuenta que

fácilmente se puede hacer una generalización hacia los otros tipos de sistemas arriba mencionados.

Para explicar el funcionamiento de la arquitectura, se considera un sistema de inferencias con dos entradas (x, y) y una salida (f). Para un modelo difuso del tipo Sugeno de primer orden, un conjunto de reglas con dos reglas difusas *si ... entonces*, se define como:

Regla 1: Si x es A_1 y y es B_1 , entonces $z_1 = p_1 x + q_1 y + r_1$,

Regla 2: Si x es A_2 y y es B_2 , entonces $z_2 = p_2 x + q_2 y + r_2$.

La figura 2.1(a) ilustra el mecanismo de razonamiento para este modelo de Sugeno; la arquitectura ANFIS correspondiente se muestra en la figura 2.1(b), en la cual, los nodos en una misma capa realizan funciones similares. A continuación se da una descripción de la función que realiza cada capa en la arquitectura mostrada (en donde la salida del i -ésimo nodo en la capa l se denota como O_{li}).

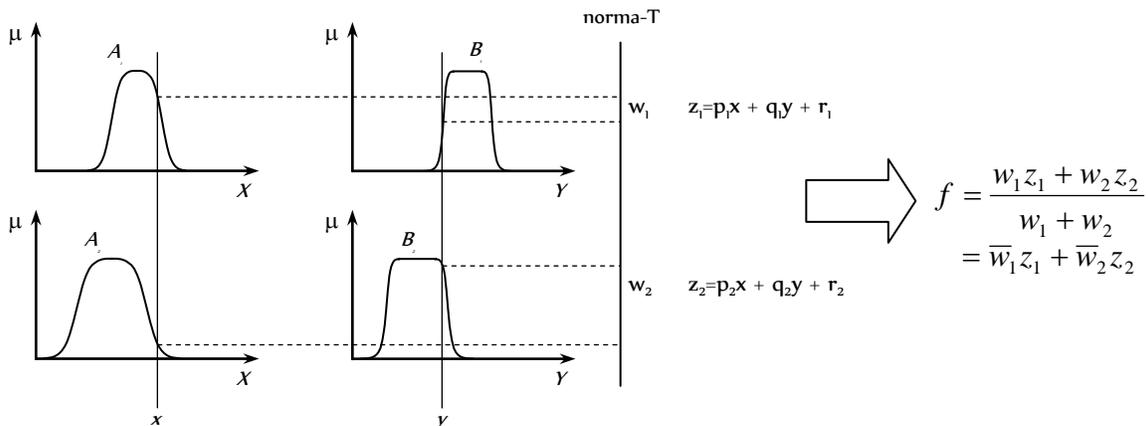
Capa 1. Cada nodo i en esta capa es un nodo adaptable con una función de nodo definida por:

$$\begin{aligned} O_{1,i} &= \mu_{A_i}(x), \quad \text{para } i=1, 2, \text{ ó} \\ O_{1,i} &= \mu_{B_{i-2}}(y), \quad \text{para } i=3, 4, \end{aligned} \tag{2.1}$$

donde x (o y) es la entrada al nodo i , y A_i (o B_{i-2}) es una etiqueta lingüística (tal como "pequeño" o "alto") asociada con este nodo. En otras palabras, $O_{1,i}$ es el grado de membresía de un conjunto difuso A ($= A_1, A_2$, o $B = B_1, B_2$), el cual especifica el grado en que una entrada (x o y) satisface al cuantificador A . La función de membresía de A , debe ser una función parametrizada que sea continua y derivable a tramos, tal como la función campana generalizada, cuya expresión matemática está dada por:

$$\mu_A(x) = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2b_i}}, \tag{2.2}$$

donde $\{a_i, b_i, c_i\}$ es el conjunto de parámetros que definen la posición y forma de la campana: a_i especifica el ancho de la campana; b_i , el valor de las pendientes; y c_i , la posición en el eje x de su eje de simetría. Esto se ilustra en la figura 2.2 siguiente. Los parámetros en esta capa son referidos como *parámetros de premisa*.



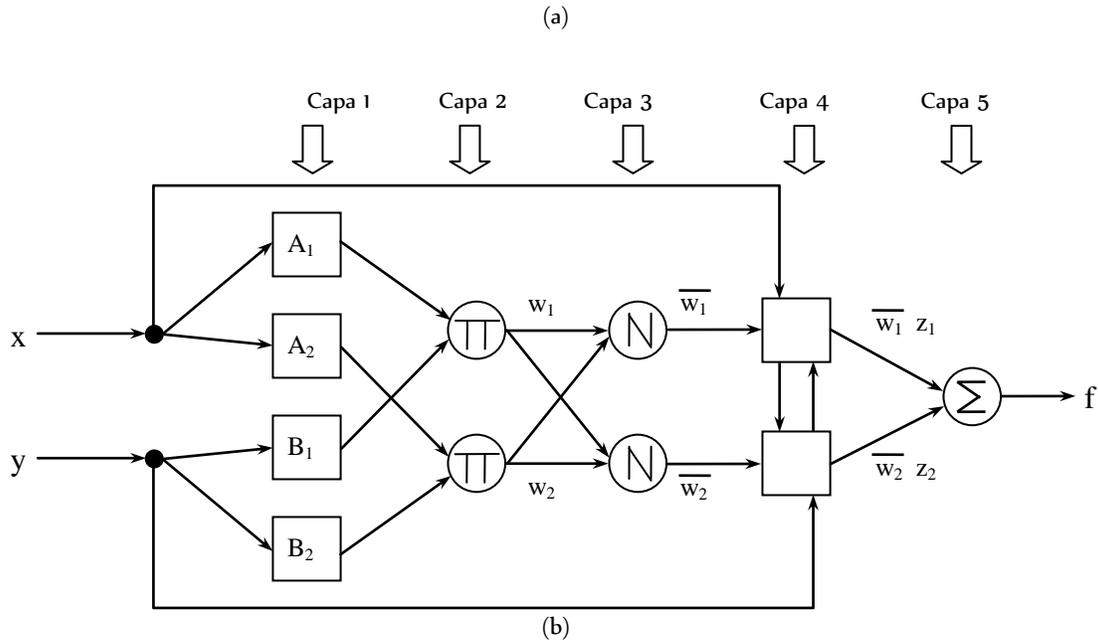


Fig. 2.1 (a) Modelo difuso de Sugeno de primer orden, con dos entradas y dos reglas difusas; (b) Arquitectura ANFIS equivalente.

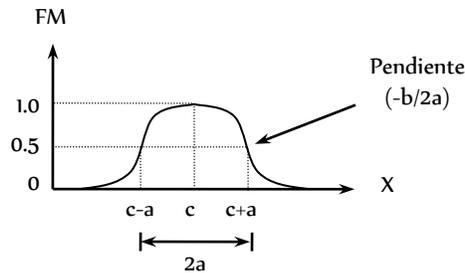


Fig. 2.2 Significado físico de los parámetros en la función de membresía tipo campana.

Capa 2. Los nodos de esta capa son nodos fijos, y se etiquetan con el símbolo Π ; la salida aquí es la operación norma-T de todas las señales que entran a él:

$$O_{2,i} = w_i = T(\mu_{A_i}(x), \mu_{B_i}(y)), \quad i=1,2. \quad (2.3)$$

Cada salida de estos nodos representa la intensidad de disparo de una regla. En general, se puede emplear cualquier operador norma-T como función de nodo en esta capa.

Capa 3. Esta capa se compone de nodos fijos etiquetados con N. El i -ésimo nodo calcula la relación de la intensidad de disparo de la i -ésima regla con respecto a la suma de las intensidades de disparo de todas las reglas:

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i=1,2. \quad (2.4)$$

A cada salida de esta capa se le llama *intensidad de disparo normalizada*.

Capa 4. Cada nodo i en esta capa es un nodo adaptable con una función definida por:

$$O_{4,i} = \bar{w}_i z_i = \bar{w}_i (p_i x + q_i y + r_i), \quad (2.5)$$

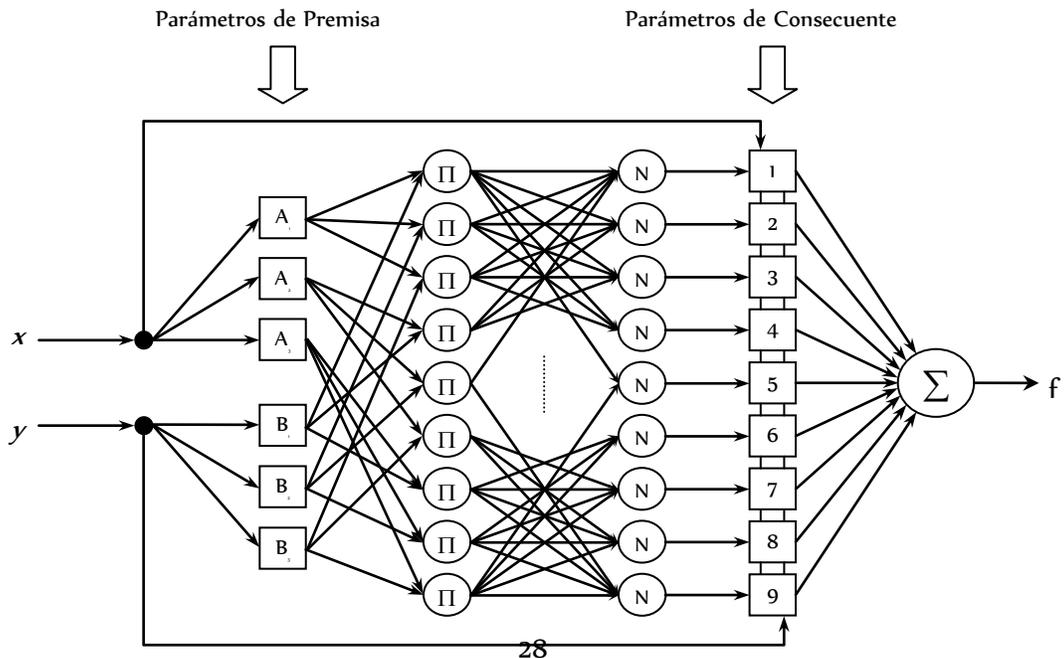
donde \bar{w}_i es una salida de intensidad normalizada de la capa 3, y $\{p_i, q_i, r_i\}$ es el conjunto de parámetros que define a los polinomios z_i en este nodo. Los parámetros en esta capa son referidos como *parámetros del consecuente*.

Capa 5. El único nodo presente en esta capa, es un nodo fijo denotado por Σ , el cual calcula la salida global como la suma de todas las señales que entran a él:

$$\text{Salida global} = O_{5,i} = f = \sum_i \bar{w}_i z_i = \frac{\sum_i w_i z_i}{\sum_i w_i} \quad (2.6)$$

De esta forma, lo que se obtiene es una red adaptable, la cual es, funcionalmente, equivalente a un modelo difuso de Sugeno. La estructura mostrada no es única, ya que se pueden hacer otras construcciones equivalentes, por ejemplo combinando las capas 3 y 4, con lo que la red tendría sólo cuatro capas.

El sistema completo que se tomó como base para este trabajo, es una arquitectura ANFIS equivalente a un modelo difuso tipo Sugeno de 1er orden con dos entradas y nueve reglas difusas, donde a cada entrada se le asocian tres funciones de membresía del tipo campana generalizada, como se muestra en la figura 2.3(a). La figura 2.3(b) ilustra la forma como se particiona el espacio de entrada bidimensional en nueve regiones difusas traslapadas, cada una de las cuales está gobernada por una regla difusa *si ... entonces*; en otras palabras, la parte de la premisa de una regla define una región difusa, mientras que la parte del consecuente especifica la salida dentro de esa región.



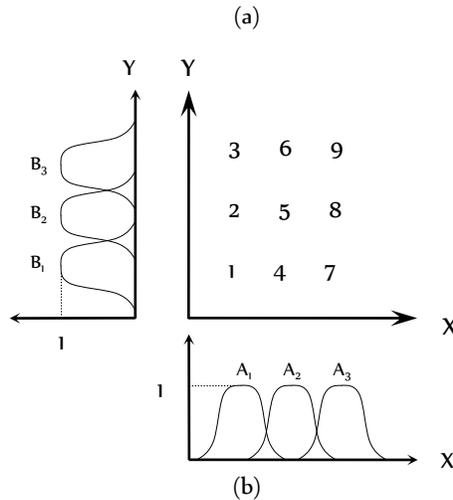


Fig. 2.3 (a) ANFIS de dos entradas y nueve reglas difusas
(b) Subespacios difusos correspondientes.

2.2.2. Regla de aprendizaje híbrida.

Aunque en una red adaptable se puede aplicar la regla de retropropagación o el descendiente abrupto para identificar sus parámetros, generalmente estos métodos son lentos para converger y tienden a quedar atrapados en mínimos locales. El método que aquí se emplea para determinar los parámetros de la red, es una regla de aprendizaje híbrida, la cual combina el método descendiente abrupto con el estimador de mínimos cuadrados (LSE), descritos en los apéndices A y B. Esto parte del hecho de que la salida de una red adaptable es lineal en algunos de sus parámetros, y estos se pueden identificar en forma rápida con el método LSE lineal.

Por sencillez, se asume que la red bajo consideración tiene sólo una salida:

$$o = F(\mathbf{i}, S) \tag{2.7}$$

donde \mathbf{i} es el vector de variables de entrada, S es el conjunto de parámetros y F es la función global que modela la red adaptable. Si existe una función H tal que la función composición $H \circ F$ sea lineal en algunos de los elementos de S , entonces esos elementos se pueden identificar con el método de mínimos cuadrados. En otras palabras, si el conjunto de parámetros S se puede descomponer en dos conjuntos:

$$S = S_1 \oplus S_2 \tag{2.8}$$

(donde \oplus representa una suma directa), tal que $H \circ F$ es lineal en los elementos de S_2 , entonces al aplicar H a la ec. 2.7 se obtiene:

$$H(o) = H \circ F(\mathbf{i}, S) \tag{2.9}$$

que es lineal en los elementos de S_2 . Si se tienen los valores de los elementos de S_2 , se pueden insertar P datos de entrenamiento en (2.9) y obtener una ecuación de matriz de la forma:

$$A\theta = \mathbf{y} \tag{2.10}$$

donde θ es un vector desconocido cuyos elementos son parámetros en S_2 . Sea $|S_2|=M$, entonces las dimensiones de A , θ y y son $P \times M$, $M \times 1$ y $P \times 1$, respectivamente. Puesto que P (número de pares de datos de entrenamiento) es normalmente mayor que M (número de parámetros lineales), se tiene un problema sobredeterminado y, generalmente, no hay una solución exacta para (2.10). Entonces, la mejor solución para θ , la cual minimiza el error cuadrático $\|A\theta - y\|^2$, es el estimador de mínimos cuadrados θ^* , definido por:

$$\theta^* = (A^T A)^{-1} A^T y \quad (2.11)$$

donde A^T es la transpuesta de A , y $(A^T A)^{-1} A^T$ es la pseudo inversa de A siempre que $A^T A$ no sea singular. La ec. 2.11 es concisa en notación, pero requiere mucho cálculo cuando trabaja con la matriz inversa, y queda pobremente definida si $A^T A$ es singular. Por ello, se utilizan fórmulas recurrentes para calcular LSE de θ . Este método recurrente de LSE es más eficiente, especialmente cuando M es pequeño.

Sea a_i^T el i -ésimo vector fila de la matriz A definido en (2.10); y y_i^T , el i -ésimo elemento de y , entonces θ se puede calcular iterativamente utilizando las fórmulas siguientes:

$$\begin{aligned} \theta_{i+1} &= \theta_i + P_{i+1} a_{i+1} (y_{i+1}^T - a_{i+1}^T \theta_i) \\ P_{i+1} &= P_i - \frac{P_i a_{i+1} a_{i+1}^T P_i}{1 + a_{i+1}^T P_i a_{i+1}}, \quad i = 0, 1, \dots, P-1 \end{aligned} \quad (2.12)$$

donde P_i frecuentemente se llama la matriz de covarianza y el estimador de mínimos cuadrados θ^* es igual a θ_p . Las condiciones iniciales necesarias para determinar la ec. 2.12 son $\theta_0 = 0$ y $P_0 = \gamma I$, donde γ es un número positivo grande e I es la matriz identidad de dimensión $M \times M$. Cuando se trabaja con redes adaptables de múltiples salidas (esto es, la salida o en la ec. 2.7 es un vector columna), la ec. 2.12 se sigue aplicando, excepto que y_i^T es la i -ésima fila de la matriz y .

Ahora bien, se puede combinar el método del gradiente con el de LSE para actualizar los parámetros en una red adaptable. Para poder aplicar el aprendizaje híbrido en una forma iterativa, cada *época* de este procedimiento de aprendizaje está compuesta de un *paso en avance* y un *paso en retroceso*. En el paso en avance, se proporcionan los datos de entrada y las señales funcionales se propagan hacia adelante para calcular cada salida de nodo, capa por capa, hasta obtener las matrices A y y de la ec. 2.10, y los parámetros en S_2 se identifican mediante la ec. 2.11 o las fórmulas secuenciales de la ec. 2.12. Después de ello, se calcula la medida del error para cada par de datos de entrenamiento. En el paso en retroceso, las señales de error (la derivada de la medida del error con respecto a cada salida de nodo. Ver las ecs. 1.36 y 1.37) se propagan desde la salida global hacia la entrada de la red, y los parámetros en S_1 se actualizan mediante el método descendiente abrupto de la ec. 1.41.

Para valores fijos dados de los parámetros en S_1 , se puede garantizar que los parámetros así encontrados en S_2 están en el punto óptimo global en el espacio de parámetros de S_2 debido a la elección de la medida del error cuadrático. Esta regla de aprendizaje híbrido no sólo puede disminuir la dimensión del espacio de búsqueda en el método del gradiente, sino que, en general, reducirá substancialmente el tiempo de convergencia.

Si los parámetros se actualizan después de cada presentación de los datos, se tiene un esquema de aprendizaje *en línea* o *patrón por patrón*. Esta estrategia de aprendizaje es vital en la identificación de parámetros en línea para sistemas con características cambiantes. Para

modificar el proceso de la regla de aprendizaje, a fin de obtener una versión en línea, el descendiente abrupto debe basarse en E_p (ec. 1.38) en lugar de E .

Para que la fórmula recursiva de mínimos cuadrados considere las características variantes en el tiempo de los datos de entrada, los efectos de los pares de datos pasados deben decaer, conforme nuevos pares de datos se hacen presentes. Un método simple para ello es formular la medida del error cuadrático como una versión ponderada que asigne mayores factores de peso a los pares de datos más recientes. Esto equivale a agregar un factor de olvido λ a la fórmula recursiva original, como se muestra a continuación:

$$\begin{aligned}\boldsymbol{\theta}_{i+1} &= \boldsymbol{\theta}_i + \mathbf{P}_{i+1} \mathbf{a}_{i+1} (\mathbf{y}_{i+1}^T - \mathbf{a}_{i+1}^T \boldsymbol{\theta}_i) \\ \mathbf{P}_{i+1} &= \frac{1}{\lambda} \left[\mathbf{P}_i - \frac{\mathbf{P}_i \mathbf{a}_{i+1} \mathbf{a}_{i+1}^T \mathbf{P}_i}{\lambda + \mathbf{a}_{i+1}^T \mathbf{P}_i \mathbf{a}_{i+1}} \right]\end{aligned}\quad (2.13)$$

donde los valores típicos de λ están entre 0.9 y 1. Entre menor sea λ , se tendrá un efecto más rápido de decaimiento de los datos pasados. Sin embargo, una λ muy pequeña puede originar inestabilidad numérica, por lo que debe evitarse.

2.2.3. ANFIS y la regla de aprendizaje híbrida.

De la arquitectura ANFIS mostrada en la figura 2.1(b) se observa que, si los valores de los parámetros de la premisa son fijos, la salida global puede expresarse como una combinación lineal de los parámetros del consecuente. Esto es, la salida f en la figura se puede reescribir como:

$$\begin{aligned}f &= \frac{w_1}{w_1 + w_2} z_1 + \frac{w_2}{w_1 + w_2} z_2 = \bar{w}_1 z_1 + \bar{w}_2 z_2 \\ &= \bar{w}_1 (p_1 x + q_1 y + r_1) + \bar{w}_2 (p_2 x + q_2 y + r_2) \\ &= (\bar{w}_1 x) p_1 + (\bar{w}_1 y) q_1 + (\bar{w}_1) r_1 + (\bar{w}_2 x) p_2 + (\bar{w}_2 y) q_2 + (\bar{w}_2) r_2\end{aligned}\quad (2.14)$$

Como resultado de esto, se tiene lo siguiente, de la ec. 2.8:

S = Conjunto total de parámetros

S_1 = Conjunto de parámetros de la premisa (no lineales)

S_2 = Conjunto de parámetros del consecuente (lineales)

$H(\cdot)$ y $F(\cdot, \cdot)$ son la función identidad y la función del sistema de inferencia difusa, respectivamente. Por consiguiente, el algoritmo de aprendizaje híbrido descrito anteriormente se puede aplicar directamente: En el paso en avance, las salidas de los nodos van hacia adelante hasta la capa 4, y los parámetros del consecuente se identifican por el método de mínimos cuadrados. En el paso en retroceso, las señales de error se propagan hacia atrás y los parámetros de la premisa se actualizan mediante el descendiente del gradiente.

Sin embargo, debe notarse que la complejidad en el cálculo del método LSE es mayor que en el del descendiente del gradiente. De hecho, hay cuatro métodos para actualizar los parámetros, tal como se enumeran a continuación de acuerdo con la complejidad de cómputo asociada:

- 1) Solamente el descendiente del gradiente: Todos los parámetros se actualizan con el método del gradiente descendiente (en el paso en avance y en el de retroceso).
- 2) Descendiente del gradiente y un paso de LSE: El método de LSE se aplica sólo una vez al comienzo, para obtener los valores iniciales de los parámetros del consecuente y, entonces, el descendiente del gradiente se aplica para obtener todos los parámetros.
- 3) Descendiente del gradiente y LSE: Esta es la regla de aprendizaje híbrida propuesta.
- 4) Solamente LSE secuencial (aproximado): ANFIS se hace lineal con respecto a los parámetros de la premisa y se utiliza el algoritmo de Kalman extendido para actualizar a todos los parámetros. Esto se ha propuesto en la literatura de redes neuronales [3].

Como ya se ha mencionado, los parámetros del consecuente encontrados de esta forma son óptimos, bajo la condición de que los parámetros de la premisa sean fijos. Así, el método híbrido converge mucho más rápido, puesto que reduce las dimensiones del espacio de búsqueda del método original de retropropagación. La elección de alguno de los 4 métodos anteriores, dependerá de los recursos de cómputo disponibles y del nivel de desempeño requerido.

2.3. RESUMEN

En este capítulo se describieron todos los elementos que conforman la arquitectura ANFIS, los cuales se integran en cinco capas, dos de las cuales se construyen con nodos que tienen parámetros adaptables y, las otras tres, tienen funciones de nodo fijas. El tipo de sistema que se realiza con la arquitectura propuesta, es funcionalmente equivalente a un sistema de tipo Sugeno de 1er orden, ya que esto permite reconstruir una función altamente no lineal a partir de un grupo de funciones lineales.

Se describe una regla de aprendizaje híbrida, mediante la cual se pueden ajustar los parámetros de ANFIS, a partir de pares de datos de entrada/salida. Esta regla de aprendizaje combina dos métodos de optimización bien conocidos en la literatura de redes neuronales, los cuales son el descendiente abrupto y el estimador de mínimos cuadrados. Estos se emplean para minimizar la medida del error de la red, y se plantea su utilización para el caso en que se tenga aprendizaje en línea.

Finalmente, se mencionan cuatro métodos que se pueden emplear en el ajuste de parámetros, los cuales se originan como combinaciones distintas de los dos métodos que forman la regla híbrida, y cuya aplicación depende de los recursos con que se disponga y del nivel de desempeño que se desee obtener.

CAPITULO 3

DISEÑO ELECTRICO DE CELDAS CMOS PARA LA ARQUITECTURA ANFIS

“Lo conocido es finito, lo desconocido infinito; intelectualmente estamos en una isleta en medio de un ilimitado océano de lo inexplicable. Nuestra labor en cada generación consiste en reclamar un poco más de tierra.”
(T. H. Huxley)

“Si descubrimos una teoría completa, sus principios generales deben ser comprensibles, con el tiempo, para todo mundo, no sólo para unos cuantos científicos. Entonces todos, filósofos, científicos y gente común podremos participar en el análisis de por qué existimos nosotros y el universo. Si hallamos la respuesta, sería el máximo triunfo de la razón humana, y conoceríamos la mente de Dios.”
(Stephen Hawking, *Breve Historia del Tiempo*)

3.1. INTRODUCCION

La arquitectura ANFIS es el resultado directo de una metodología de cómputo que surgió en la década de los 90s denominada “Soft Computing”, la cual integra las habilidades de la mente humana para razonar y aprender, en un ambiente de incertidumbre e imprecisión. Esta metodología consiste de varios paradigmas computacionales que incluyen las redes neuronales, teoría de conjuntos difusos, razonamiento aproximado y métodos de optimización libres de derivadas, tales como los algoritmos genéticos y el recocido simulado (“simulated annealing”). Sin embargo, en su idea original, esta metodología depende de un sistema de cómputo, lo cual en ciertas circunstancias puede llegar a ser un inconveniente. Es por ello que surge la necesidad de construir esta arquitectura empleando dispositivos electrónicos, diseñados específicamente para realizar las funciones de cada una de sus capas y contenidos en uno o dos circuitos integrados, con lo cual se puede obtener un mejor desempeño, ya que de esta forma realmente se efectúa un procesamiento en paralelo (a diferencia de una computadora “clásica”, que realiza el procesamiento en forma secuencial) lo cual es una de las características que hacen poderosas a estos tipos de arquitecturas; y, al tener pocos componentes, se obtiene un mayor ahorro de energía y espacio, lo cual es indispensable si se quiere hablar de un sistema portátil.

Para construir el sistema ANFIS global, éste ha sido dividido en dos partes: Una que abarca las dos capas iniciales de la arquitectura (capas 1 y 2), y otra que incluye las otras tres (capas 3, 4 y 5), correspondiendo al presente trabajo de tesis el desarrollo de los circuitos necesarios para la primera parte.

La metodología que se siguió fue: Inicialmente diseñar, simular y enviar a fabricar celdas básicas que aproximen las operaciones requeridas en las dos primeras capas de la arquitectura para, posteriormente, poder construir la estructura completa que incluye las capas 1 y 2, interconectadas como se mostró en la figura 2.3(a), utilizando dichas celda básicas.

3.2. DISEÑO DE CELDAS BASICAS

El proyecto dentro del cual se desarrolla el presente trabajo, plantea algunas características que deben cumplir los circuitos aquí diseñados: Se deben emplear circuitos CMOS con voltaje de alimentación de 5V; las salidas del circuito deben ser corrientes, ya que el circuito defusificador con el cual se acoplará este diseño, utiliza corrientes en sus entradas; las variables de entrada deben ser voltajes, para mayor flexibilidad de interfaz con los posibles circuitos que conformen la etapa de entrada del diseño; las funciones de membresía deben ser ajustables, para modificarlas de acuerdo al entrenamiento realizado en Matlab; y, el espacio difuso debe quedar dividido en nueve regiones difusas, de acuerdo con la arquitectura que se plantea reproducir (figura 2.3). Sin embargo, no hay restricciones en cuanto al funcionamiento y características propias de los circuitos que se desarrollen.

3.2.1. Circuito Generador de Campanas.

Para construir la primera capa de la arquitectura ANFIS, se requiere de un circuito capaz de generar funciones del tipo campana, que es el tipo de función de membresía empleada en este trabajo. La expresión matemática de una campana generalizada es la ec. 2.2 mostrada anteriormente. Para aproximar esta función, se aprovecha la característica de transferencia de un amplificador diferencial, de tal forma que se obtiene una corriente controlada por voltaje.

La figura 3.1(a) muestra el circuito de un amplificador diferencial CMOS que usa transistores de canal N (M_1 y M_2) en el par diferencial, así como su curva característica normalizada (figura 3.1(b)). La característica de *gran señal* se puede obtener considerando que M_1 y M_2 siempre están en saturación. Esta condición es razonable en la mayoría de los casos, e ilustra el comportamiento del circuito aun cuando tal consideración no sea válida.

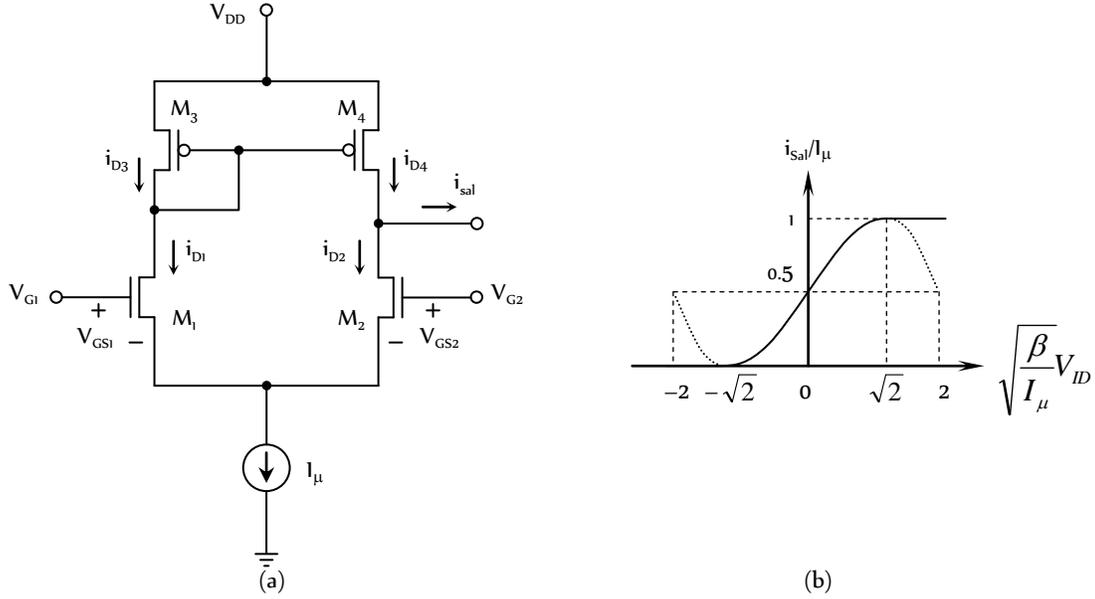


Fig. 3.1 a) Amplificador diferencial CMOS con transistores de entrada de canal N; b) Característica de transferencia.

Inicialmente, se parte de la relación que describe el comportamiento de gran señal, definida por:

$$V_{ID} = V_{GS1} - V_{GS2} = \left(\frac{2i_{D1}}{\beta} \right)^{1/2} - \left(\frac{2i_{D2}}{\beta} \right)^{1/2} \quad (3.1)$$

y la siguiente ecuación:

$$I_{\mu} = i_{D1} + i_{D2} \quad (3.2)$$

donde $\beta = \frac{K'W}{L}$, siendo K un parámetro tecnológico y W y L son las dimensiones del transistor, para lo cual se asume que M_1 y M_2 son iguales.

Al sustituir la ec. 3.2 en la 3.1, y realizando una manipulación algebraica, se llega a las siguientes soluciones para las corrientes i_{D1} e i_{D2} :

$$i_{D1} = \frac{I_{\mu}}{2} + \frac{I_{\mu}}{2} \left(\frac{\beta V_{ID}^2}{I_{\mu}} - \frac{\beta^2 V_{ID}^4}{4I_{\mu}^2} \right)^{1/2} \quad (3.3)$$

$$i_{D2} = \frac{I_{\mu}}{2} - \frac{I_{\mu}}{2} \left(\frac{\beta V_{ID}^2}{I_{\mu}} - \frac{\beta^2 V_{ID}^4}{4I_{\mu}^2} \right)^{1/2} \quad (3.4)$$

las cuales son válidas dentro del siguiente límite, fuera del cual, el valor de la corriente se considera que es constante:

$$|V_{ID}| < \left(\frac{2I_{\mu}}{\beta} \right)^{1/2} \quad (3.5)$$

Si se asume que las corrientes en el espejo de corriente (transistores canal P) son idénticas, entonces i_{sal} se encuentra restando i_{D2} de i_{D1} , con lo cual se obtiene la siguiente expresión:

$$i_{sal} = i_{D1} - i_{D2} = I_{\mu} \left(\frac{\beta V_{ID}^2}{I_{\mu}} - \frac{\beta^2 V_{ID}^4}{4I_{\mu}^2} \right)^{1/2} \quad (3.6)$$

Con este análisis se obtiene una corriente de salida en términos del voltaje diferencial de entrada V_{ID} . La función de transferencia resultante es como la que se muestra en la figura 3.1(b) anterior, donde se observa que la corriente máxima (correspondiente al 1 lógico) se determina por el valor de I_{μ} . La pendiente de esta curva, la cual resulta en la ecuación de transconductancia del amplificador diferencial, se obtiene derivando la ec. 3.6 con respecto a V_{ID} , de tal forma que:

$$g_m = \frac{\partial i_{Sal}}{\partial V_{ID}} = \left(\frac{\beta I_{\mu}}{4} \right)^{1/2} = \left(\frac{K' I_{\mu} W}{4L} \right)^{1/2} \quad (3.7)$$

Para obtener la forma completa de la función de membresía, se utiliza una configuración como la mostrada en la figura 3.2(a), basada en [17] y [18], y que básicamente consiste en dos amplificadores diferenciales entrelazados, lo cual permite obtener una función de transferencia como la que se observa en la figura 3.2(b).

La corriente de salida del circuito (i_s) es la suma de las corrientes de ambos amplificadores diferenciales (i_{O1} e i_{O2}) las cuales, al estar en sentidos contrarios, forman la función de transferencia deseada. Entonces, se tiene que:

$$i_s = i_{O1} + i_{O2} \quad (3.8)$$

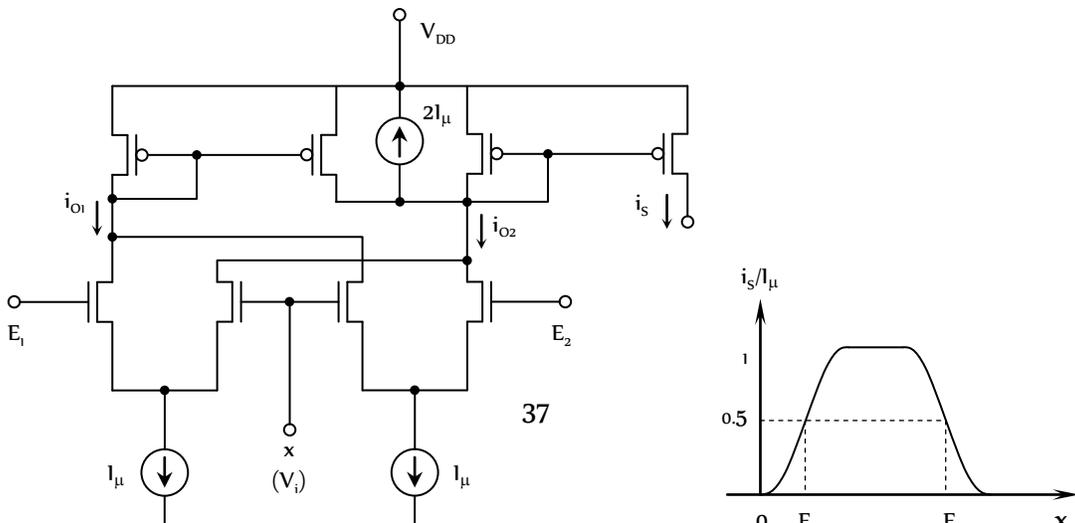
Los voltajes para cada amplificador se determinan por:

$$V_{ID1} = V_i - E_1 \quad (3.9)$$

$$V_{ID2} = E_2 - V_i \quad (3.10)$$

Tomando la ec. 3.6, y sustituyéndola en 3.8, para cada corriente, y considerando los voltajes de las ec. 3.9 y 3.10, se obtiene la siguiente expresión:

$$i_s = I_{\mu} \left(\frac{\beta V_{ID1}^2}{I_{\mu}} - \frac{\beta^2 V_{ID1}^4}{4I_{\mu}^2} \right)^{1/2} + I_{\mu} \left(\frac{\beta V_{ID2}^2}{I_{\mu}} - \frac{\beta^2 V_{ID2}^4}{4I_{\mu}^2} \right)^{1/2} \quad (3.11)$$



(a) (b)

Fig. 3.2 a) Configuración para aproximar la función de una campana generalizada; b) característica de transferencia.

La ecuación 3.11 es la de una función que aproxima la forma de una campana generalizada, permitiendo controlar su posición y su ancho de banda a través de sus puntos de cruce, los cuales se determinan directamente con los voltajes E_1 y E_2 . Sin embargo, para poder modificar las pendientes, se debe modificar la transconductancia de los transistores del par diferencial (ec. 3.7), lo cual se puede lograr de varias formas. La que aquí se considera, utiliza un arreglo de dos transistores que comparten sus drenadores y sus fuentes, y una batería que une sus compuertas, como se muestra en la figura 3.3 siguiente.

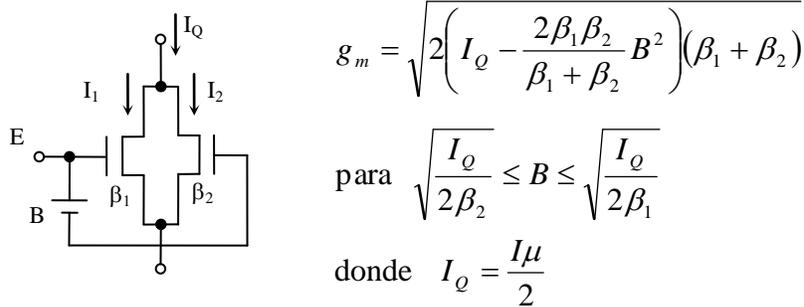


Fig. 3.3 Configuración para modificar la transconductancia de un transistor MOS.

Como se observa en la ecuación mostrada en la figura 3.3, puesto que las dimensiones de ambos transistores son fijas, al modificar el valor de la batería B se modifica la transconductancia global del arreglo de transistores. De esta forma, si se sustituye cada transistor del par diferencial por un arreglo como el de la figura 3.3, se logra el ajuste de las pendientes en función de un voltaje independiente (B).

3.2.2. Circuito para realizar la operación MIN.

En la segunda capa de ANFIS se requiere un circuito que realice una operación normal; en este trabajo se utiliza la operación más común que es la MIN, la cual es sencilla de generar con circuitos. Esta operación se logra efectuar tomando como punto de partida un circuito de competencia del tipo *el Ganador Toma Todo* propuesto inicialmente por Lazzaro [19], el cual lleva a cabo una operación MAX, con la cual se efectúa la operación MIN a partir del teorema de D' Morgan, como sigue: $w_i = \min(S_{i1}, S_{i2}, \dots, S_{iM}) = \overline{\max(\overline{S_{i1}}, \overline{S_{i2}}, \dots, \overline{S_{iM}})}$, donde las S_{ij} son corrientes y su complemento se obtiene usando la ley de Kirchoff de corrientes: $\overline{S_{ij}} = I_U - S_{ij}$, donde I_U es la corriente asociada con el valor lógico 1. En la figura 3.4(a) se muestra el circuito que realiza esta función, donde w_i es una corriente que corresponde a la corriente S_{ij} más

pequeña (o bien, a la corriente \bar{S}_{ij} más grande). A las entradas se les inyecta una corriente adicional I_B para evitar que el transistor por el cual circula la corriente \bar{S}_{ij} mayor, entre en inversión débil. La configuración de la figura 3.4(b) permite ajustar V_{ref} de tal forma que iguale al voltaje drenador-fuente de M_O con el del transistor de entrada que conduce la corriente máxima. En otras palabras, el voltaje V_{ref} se modifica de acuerdo a las variaciones de la corriente máxima.

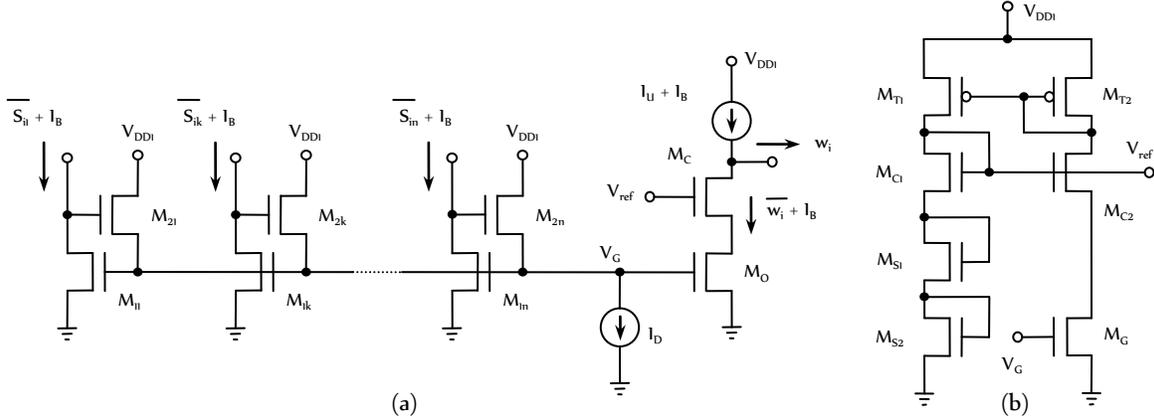


Fig. 3.4 (a) Circuito CMOS en modo de corriente para realizar la función MIN; (b) Circuito de polarización de (a).

Para analizar el funcionamiento del circuito de competencia, se considera una configuración con dos entradas, como en la figura 3.5, pudiéndose generalizar fácilmente el desarrollo para el caso de múltiples entradas. Por sencillez de análisis, se asumen características ideales de los componentes. Inicialmente se plantea la condición $I_1 = I_2 \equiv I_m$. Si todos los transistores tienen las mismas dimensiones, se tiene que $\beta_{11} = \beta_{12} = \beta_{21} = \beta_{22} \equiv \beta$. Los transistores M_{11} y M_{12} tienen voltajes de compuerta y fuente idénticos, y ambos conducen una corriente I_m ; así, los voltajes V_1 y V_2 deben ser iguales. Los transistores M_{21} y M_{22} tienen sus voltajes de drenador, compuerta y fuente iguales, por lo que deben conducir la misma corriente, de tal forma que $I_{c1} = I_{c2} = I_D/2$. Considerando que los transistores operan en la región de saturación, la corriente que circula en M_{11} y M_{12} se define por:

$$I_m = \frac{\beta}{2} (V_G - V_T)^2, \quad \text{si } 0 < (V_G - V_T) < V_m \quad (3.12)$$

De igual forma, la corriente de M_{21} y M_{22} está definida por:

$$\frac{I_D}{2} = \frac{\beta}{2} (V_m - V_G - V_T)^2, \quad \text{si } 0 < (V_m - V_G - V_T) < (V_{DD} - V_G) \quad (3.13)$$

Despejando V_G en 3.12, sustituyendo en 3.13 y despejando para V_m , se obtiene la siguiente expresión:

$$V_m = \left(\frac{I_D}{\beta} \right)^{1/2} + \left(\frac{2I_m}{\beta} \right)^{1/2} + 2V_T \quad (3.14)$$

De tal forma que, para corrientes de entrada iguales, el circuito produce voltajes de salida iguales, lo cual es un comportamiento deseado en un circuito de competencia.

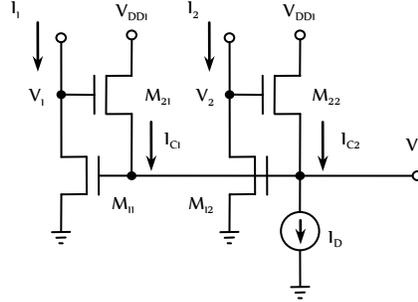


Fig. 3.5 Configuración para un circuito de competencia de dos entradas.

Ahora, se considera la condición $I_1 = I_m + \Delta i$ e $I_2 = I_m$ para ilustrar la acción inhibitoria del circuito. Bajo esta condición, M_{11} debe drenar una corriente Δi más que en el caso previo; como resultado, el voltaje de compuerta de M_{11} se incrementa. Al tener M_{12} el mismo voltaje de compuerta, también debería drenar una corriente $I_m + \Delta i$, pero sólo I_m está presente en su drenador. Esto origina que su voltaje de drenador (V_2) disminuya, con lo cual se sale de saturación hasta llevar a V_2 aproximadamente a 0V. Para valores grandes de Δi , se tiene que $I_{C2} \cong 0$ e $I_{C1} \cong I_D$. En esta situación, las ecuaciones de corriente de M_{11} y M_{21} serán, respectivamente, las dos siguientes:

$$I_m + \Delta i = \frac{\beta}{2}(V_G - V_T)^2 \quad (3.15)$$

$$I_D = \frac{\beta}{2}(V_1 - V_G - V_T)^2 \quad (3.16)$$

Al despejar V_G en la ec. 3.15, sustituyéndolo en 3.16 y despejando V_1 , se obtiene que:

$$V_1 = \left(\frac{2I_D}{\beta} \right)^{1/2} + \left[\frac{2(I_m + \Delta i)}{\beta} \right]^{1/2} + 2V_T \quad (3.17)$$

Esta ecuación predice la respuesta ganadora del circuito, ya que el voltaje mayor entre V_1 y V_2 , estará determinado por la rama en la cual circule la máxima corriente. Esto se hace más evidente si se despeja V_G en la ec. 3.16, con lo que se observa que las variaciones del voltaje de salida están determinadas por el voltaje presente en la rama ganadora:

$$V_G = V_1 - \left(\frac{2I_D}{\beta} \right)^{1/2} - V_T \quad (3.18)$$

La topología simétrica del circuito, asegura un comportamiento similar si se tiene un incremento de I_2 con respecto a I_1 .

3.3. SIMULACIÓN DE CELDAS

3.3.1. Celdas básicas.

Inicialmente, la simulación eléctrica de celdas individuales se llevó a cabo para el circuito que aproxima la función campana generalizada. Para ello se empleó el circuito mostrado en la figura 3.2, con la configuración indicada en la figura 3.3, obteniéndose el circuito de la figura 3.6.

El simulador eléctrico utilizado fue el programa PSpice. Todas las simulaciones se hicieron para una tecnología de $2\mu\text{m}$, pozo N, considerando el conjunto de parámetros de simulación proporcionados por la organización académica MOSIS, a la cual se enviaron los diseños para su fabricación; este conjunto de parámetros se describe en el apéndice C. Se emplearon los parámetros de Spice típicos de esta organización, y las simulaciones se hicieron utilizando el modelo BSIM de nivel 4 para transistores MOS. Con respecto a esto último, cabe mencionar que las ecuaciones del transistor MOS para el nivel 4, resultan complicadas de manejar para hacer diseño y, por ello, la metodología utilizada consistió en emplear inicialmente las ecuaciones simplificadas que se desarrollaron en los puntos anteriores. Con estas ecuaciones se obtuvo una primera aproximación de las dimensiones de los transistores y posteriormente, con ayuda del simulador, se hicieron ajustes hasta obtener una respuesta satisfactoria. De esta forma se llegó a los circuitos que se mencionan a continuación:

Las características a destacar en el circuito de la figura 3.6 son las siguientes: Dimensiones de los transistores (W/L), en donde el cociente indicado expresa dos cantidades en micras: $M_1 = M_2 = M_3 = M_4 = 10/2$, $M_5 = M_6 = M_7 = M_8 = 50/2$, $M_{5a} = M_{6a} = M_{7a} = M_{8a} = 3/12$; $I_{\mu} = 4\mu\text{A}$; $V_{DD} = 5\text{V}$; $2\text{V} \leq V_i \leq 5\text{V}$; $0 \leq B \leq 0.6\text{V}$ (ver fig. 3.3), $2\text{V} \leq (E_1 \equiv E_2) \leq 5\text{V}$. Los voltajes correspondientes a la batería B, mostrada en la figura 3.3, se suministran mediante voltajes independientes (E_{1A} , E_{2A} y E_{1A}), de tal forma que en todo momento se cumpla la condición $B = E_{1A} - E_1 = E_{2A} - E_2 = E_{1A} - V_i$, condición necesaria para generar campanas simétricas.

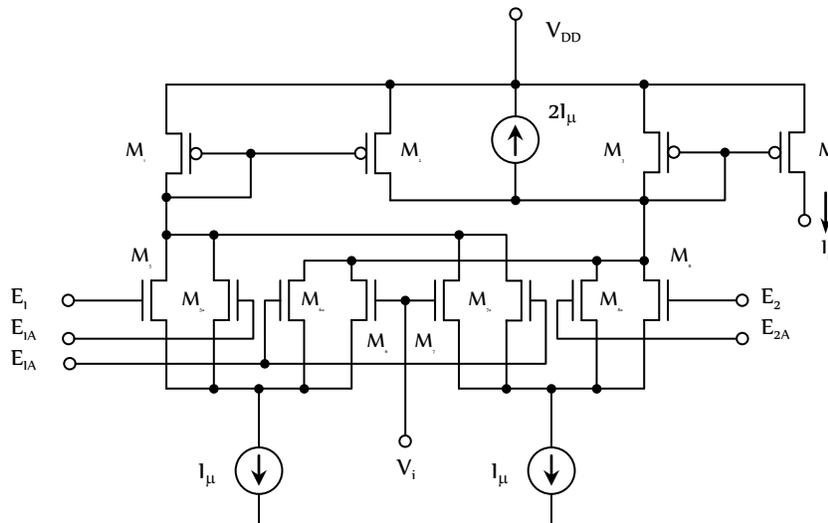


Fig. 3.6 Circuito para aproximar una función de tipo campana generalizada con pendientes ajustables.

La fig. 3.7 muestra la simulación hecha con PSpice del circuito generador de campanas para diferentes valores de la batería B. El intervalo de variación de las fuentes de voltaje, necesario para observar los cambios en las pendientes, es el definido anteriormente ($0 \leq B \leq 0.6\text{V}$). En este caso, la posición de la campana se define por $E_1 = 2.5\text{V}$ y $E_2 = 4.0\text{V}$. El

complemento mediante una corriente de $5\mu\text{A}$ en la figura, es tan solo para observarla sin inversión.

En la figura 3.8 se observan los resultados de la simulación de un circuito que realiza la operación MIN, basado en la configuración de la fig. 3.4. La figura muestra el resultado de la operación MIN entre dos funciones de membresía. Los voltajes que definen las posiciones de las campanas son los siguientes:

Campana 1:	$E_1 = 2.5\text{V}$	$E_2 = 3.5\text{V}$
Campana 2:	$E_1 = 3.5\text{V}$	$E_2 = 4.5\text{V}$

Los datos del circuito de la figura 3.4 son los siguientes: Dimensiones de los transistores (W/L) $M_{1k} = M_{2k} = 30/30$, $M_C = M_O = 25/22$, $M_{S1} = M_{S2} = M_G = M_{C1} = M_{C2} = 3/2$, $M_{T1} = M_{T2} = 10/2$, $I_B = 5\mu\text{A}$, $I_U = 10\mu\text{A}$, $I_D = 2\mu\text{A}$, $0 \leq S_{ik} \leq 4\mu\text{A}$.

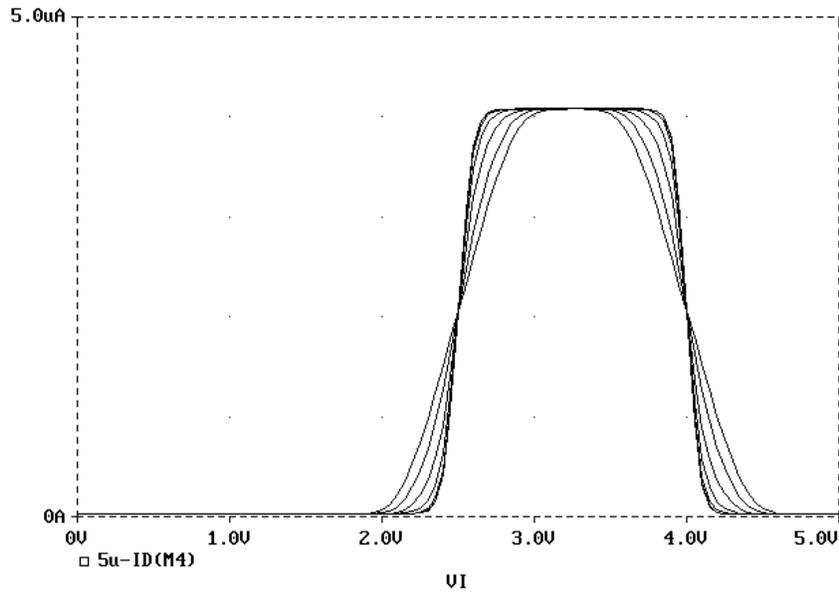


Fig. 3.7 Simulación del cambio de pendiente de la función de membresía.

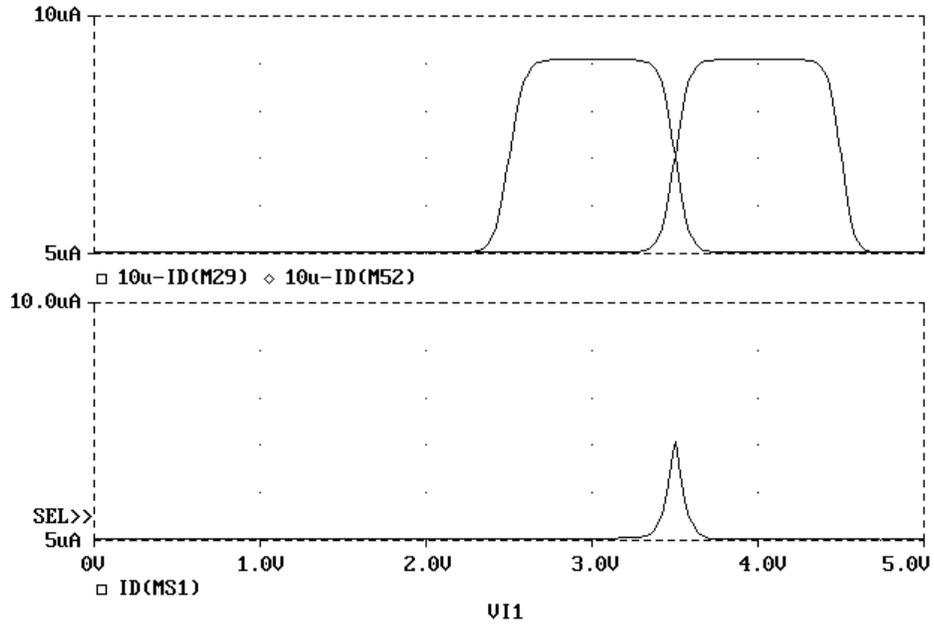


Fig. 3.8 Simulación del circuito que realiza la función MIN con dos funciones de membresía.

Una característica importante de este tipo de circuitos es su capacidad de discriminación, que se define como la habilidad del circuito para distinguir entre dos valores de entrada cercanos, esto es, la magnitud mínima de corriente en la cual una de las entradas debe ser mayor que las demás para ser reconocida como la mayor de todas ellas (la entrada ganadora). Esto puede expresarse como un porcentaje con respecto a la máxima amplitud de corriente que se aplica en sus entradas, siendo deseable obtener el menor valor posible. De acuerdo a los valores obtenidos en simulación, el mínimo incremento de corriente que puede detectar el circuito es de 15nA, lo cual corresponde a un incremento de 0.38% de una de las corrientes de entrada con respecto a la otra (considerando una amplitud máxima de 4µA).

3.3.2. Variación de temperatura.

Esta prueba se hace con el fin de garantizar el buen funcionamiento del circuito para distintas condiciones ambientales. El intervalo de temperatura considerado es el comercial, el cual se ha establecido para asegurar que los circuitos integrados operarán dentro de una variación de temperatura estandarizada, la cual va de 0 a 70°C.

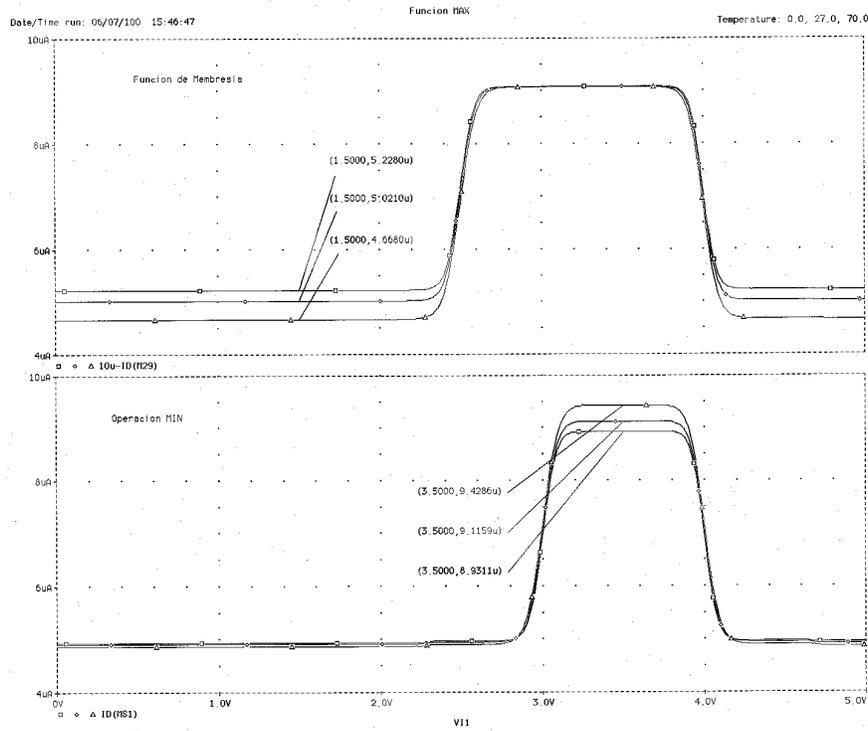


Fig. 3.9 Variación de la función de membresía y la función MIN para valores de temperatura de 0, 27 y 70°C, que cubren el intervalo comercial.

En la figura 3.9 se presentan las variaciones que se obtienen en las funciones de membresía y en la corriente correspondiente a la operación MIN, para los valores de temperatura considerados en el intervalo comercial. Considerando 27°C como el valor de referencia (temperatura nominal), las variaciones obtenidas en la simulación para las funciones de membresía son: 4.12%, para 0°C y 7.03% para 70°C; en el caso de la operación MIN los cambios obtenidos son: 2.03% para 0°C y 3.43% para 70°C. Como se observa en la figura, el mayor cambio de la función de membresía se tiene en la base de la campana, mientras que para la operación MIN, está en la parte superior. Si todas las campanas varían en igual proporción, el error máximo en la amplitud se debe casi exclusivamente al error del circuito MIN, esto es, 3.43%. Como se observa, las posiciones de las campanas no se ven afectadas por la temperatura: La variación es menor a 0.1% con respecto a la temperatura nominal.

3.3.3. Capas 1 y 2 de ANFIS.

El circuito total del que se ocupa esta tesis incluye las celdas necesarias para construir las capas 1 y 2 de una arquitectura ANFIS como la que se mostró en la figura 2.3(a), la cual produce una partición del espacio difuso como la de la figura 2.3(b). Esto implica que el circuito integrado a fabricarse debe contener 6 celdas de funciones de membresía, como la mostrada en la figura 3.6, y 9 celdas como la que se observa en la figura 3.4(a) y (b), con $n = 2$. Cada función de membresía requiere 5 voltajes para controlar su posición y sus pendientes.

Debido a la limitación en el número de terminales del encapsulado del circuito integrado (40 terminales), los voltajes para ajustar los parámetros de las campanas se introducen en forma serial. Para ello, se utilizan registros de corrimiento que controlan compuertas de transmisión,

las cuales permiten o no el almacenamiento de los voltajes adecuados en capacitores hechos con polisilicio. Este método requiere que continuamente se refresquen dichos voltajes. La figura 3.10 muestra el diagrama del circuito que realiza esta función, el cual proporciona los cinco voltajes que requiere una campana. La señal V_{CAP} es una señal periódica, que se determina por los diferentes niveles de voltaje que definen a la campana. Para poder generar esta señal, se emplean divisores de voltaje contruidos con potenciómetros variables multivuelta, en los cuales se fijan los niveles de voltaje adecuados y, para reunirlos en una única señal, se utiliza un circuito multiplexor analógico comercial.

Considerando el valor de la capacidad por unidad de área entre las capas Poly1 y Poly2 (correspondientes a los dos niveles de polisilicio disponibles en esta tecnología), $C_{P1P2} = 478aF/\mu m^2$, y las dimensiones de los capacitores: $60\mu m \times 78\mu m$, se obtiene un valor de $2.2pF$ para cada capacitor. Las compuertas lógicas U_6 y U_7 de la figura 3.10 introducen un retardo para asegurar que el voltaje deseado en la compuerta de transmisión correspondiente, se habrá estabilizado al momento de habilitar la compuerta para cargar al capacitor. Las compuertas U_1 a U_5 son para que el voltaje de entrada se transmita únicamente cuando el reloj esté en estado alto.

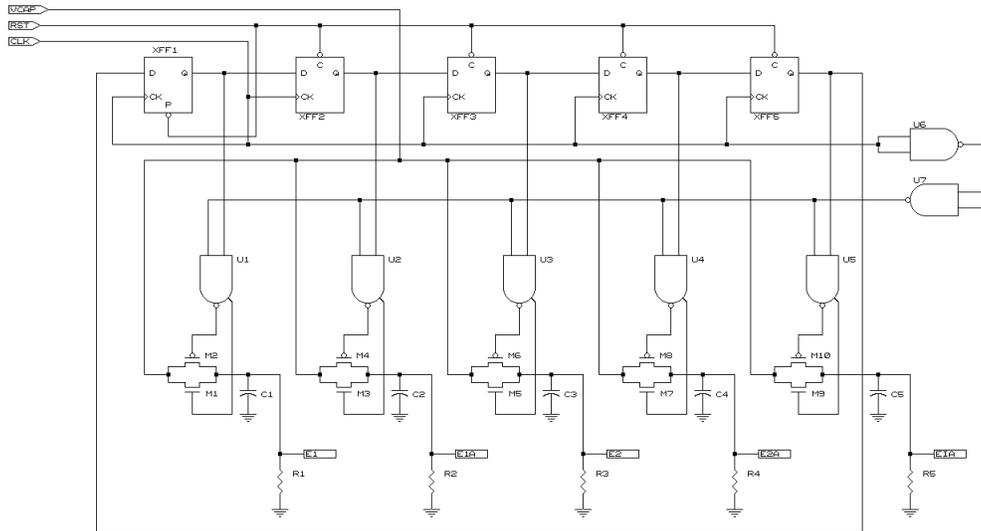


Fig. 3.10 Registros en serie para obtener los voltajes de los parámetros de las campanas, empleando carga de capacitores.

En la figura 3.11 se observan los resultados de simulación de dos funciones de membresía, utilizando este tipo de registro de corrimiento, para determinar si su posición y pendientes o el MIN entre ellas se ven afectados. El ruido presente en las funciones de membresía, se debe a que se hace un análisis transitorio que combina señales de alta frecuencia (la señal del reloj y el voltaje dinámico que se introduce a los capacitores) con una de baja frecuencia (el barrido para generar la propia función de membresía). Así, la resolución de las señales de alta frecuencia se mantiene baja, para mantener los tiempos de simulación en un nivel aceptable; sin embargo, se observa que el ruido no afecta la salida del circuito MIN (corriente $ID(ML5)$). El voltaje de entrada hace un barrido de 0 a 5V en un tiempo de 10 a $200\mu s$. Las posiciones de las funciones de membresía mostradas en la simulación, están dadas por los siguientes voltajes:

$$\text{Campana 1: } E_1 = 2.0V \qquad E_2 = 3.5V$$

Campana 2: $E_1 = 2.5V$ $E_2 = 4.0V$

En este caso, ambas campanas tienen pendientes diferentes, el reloj opera con una frecuencia de 1MHz, mientras que el voltaje de barrido para las funciones de membresía varía de 0 a 5V en un tiempo de 10 a 200 μ S. La señal de inicialización (Reset) para los registros tiene una duración de 2 μ S. El periodo máximo de refrescamiento para conservar una variación máxima de 2% en los voltajes de los capacitores, es de 200 μ s, ya que una variación mayor afecta en forma indeseable las posiciones y pendientes de las campanas.

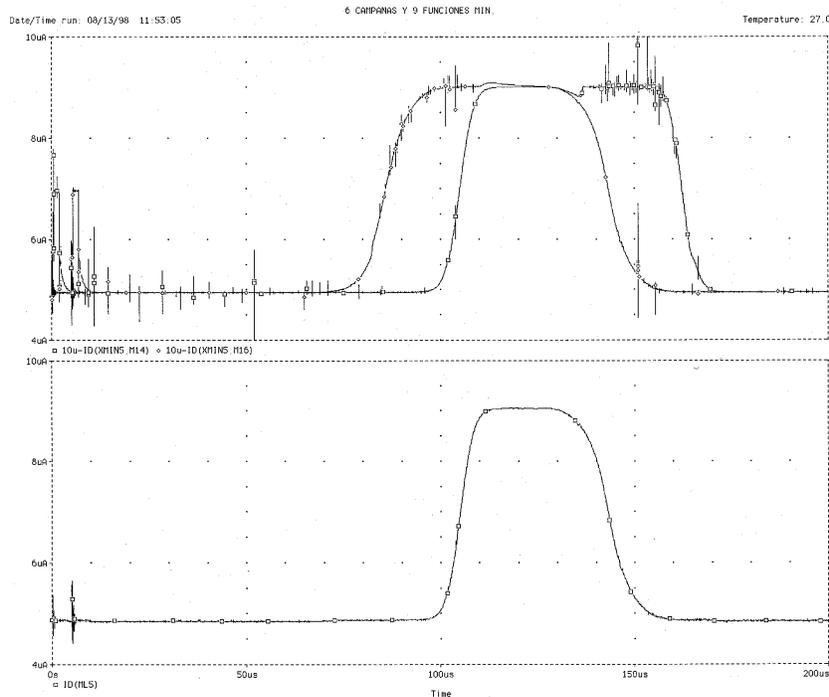


Fig. 3.11 Simulación de 6 funciones de membresía y 9 MIN, empleando registros en serie.

Puesto que en total son 6 funciones de membresía las que se utilizan en el circuito completo, se requerirían asimismo 6 configuraciones como la de la figura 3.10. Lo que se hizo en este caso fue construir tres arreglos, cada uno con 10 registros en serie (equivalente a dos configuraciones de la fig. 3.10), de tal forma que cada arreglo proporcionara los voltajes de dos campanas.

3.4. RESUMEN

Este capítulo describe los circuitos diseñados para generar las funciones requeridas para construir las capas 1 y 2 de la arquitectura ANFIS. Para ello se utilizan circuitos CMOS. En la primera capa de la arquitectura, es necesario un circuito que aproxime una función del tipo campana generalizada, lo cual se logra aprovechando la característica de transferencia de un

amplificador diferencial. Usando una combinación de dos de estos amplificadores, se obtiene un circuito que realiza la función deseada.

Para realizar la operación MIN, requerida en la capa 2, se emplea un circuito de competencia del tipo “el ganador toma todo”. Este circuito en realidad lleva a cabo una operación MAX, por lo que, empleando las leyes de De Morgan, se obtiene la función MIN que se necesita. Esto resulta sencillo de hacer al trabajar con circuitos que operan en modo de corriente, ya que el complemento de una corriente se obtiene mediante una simple resta de corrientes.

Se realizaron simulaciones de estos circuitos para verificar su correcto funcionamiento y establecer sus límites de operación. Asimismo, considerando la limitación en el número de terminales del circuito integrado, determinado por el fabricante del circuito, se propone el uso de registros de corrimiento para introducir todos los voltajes necesarios para los ajustes de las funciones de membresía, en forma serial.

CAPITULO 4

DISEÑO GEOMETRICO DE LOS CIRCUITOS CMOS

“Como puede suceder con cualquier persona, un científico puede estar entregado durante años a un trabajo que, en tanto útil, también es predecible y común. Sin embargo, como casi no sucede con nadie más, un científico puede encontrarse en cualquier momento con lo inesperado y descubrirse a sí mismo(a) como un agente de transformación en la sociedad, o del punto de vista humano respecto al universo.”

(Isaac Asimov, *La receta del Tiranosaurio*)

4.1. INTRODUCCION

El patrón a implantarse en un circuito integrado al momento de su fabricación, parte de un diseño geométrico (*layout*) del circuito eléctrico que se desea integrar, en el cual se representan todos los dispositivos, considerando sus interconexiones y dimensiones físicas. Para esta tesis, los diseños geométricos de las celdas construidas se realizan basándose en las reglas de diseño especificadas por la organización académica MOSIS, para una tecnología de $2\mu\text{m}$, dos niveles de polisilicio, dos niveles de metal y pozo N, ajustándose a un tipo de encapsulado llamado *Tiny-Chip*, que consta de 40 terminales y un área de silicio contenida en un rectángulo de $2220\mu\text{m} \times 2250\mu\text{m}$. Para la elaboración del diseño geométrico se utiliza el programa de edición gráfica llamado **L-Edit**, creado por la compañía Tanner Research, Inc.

Las reglas de diseño se refieren al conjunto de geometrías permitidas que puede usar el diseñador para asegurar que los elementos de su circuito tienen las dimensiones adecuadas para su correcto funcionamiento. Ellas establecen restricciones en los valores mínimos permisibles para ciertos *anchos*, *separaciones*, *extensiones* y *traslapes* de los objetos geométricos definidos en el sistema.

MOSIS, nombre que corresponde a *MOS Implementation System*, es una institución que proporciona el servicio de fabricación de circuitos integrados a sus abonados, mediante su envío a fábricas de silicio ubicadas en Estados Unidos. En una corrida de multiproyecto, ellos entregan 4 circuitos integrados prototipo.

El diseño que se envía a MOSIS es un archivo en un formato conocido como CIF (Caltech Intermediate Form). Este formato sirve como estándar para describir un diseño, sin importar el método de descripción original empleado para crearlo, tal como un lenguaje de descripción simbólica de geometrías, o un programa de diseño interactivo. El propio programa L-Edit genera el archivo en este formato.

Dentro del ambiente gráfico de L-Edit, los diferentes materiales que forman los elementos de los circuitos, se representan con distintos colores, como se detallan a continuación:

 Difusión (Activo)	 Contacto
 Poly 1	 Vía
 Metal 1	 Pozo N
 Metal 2	 N Select
 Poly 2	 P Select

Algunas observaciones de los materiales utilizados son: 1) Hay tres tipos de contactos que se representan con color negro: Contactos entre Metal 1 y Poly 1, contactos entre Metal 1 y Poly 2, y contactos entre Metal 1 y Activo, los cuales son claramente diferenciables dentro del programa; 2) La Vía, que es un contacto entre Metal 1 y Metal 2, es puramente blanca (sin ningún contorno); 3) El Pozo N, N Select y P Select se representan únicamente por contornos de los colores indicados; 4) Las capas N Select y P Select indican las zonas donde serán colocados los transistores canal N y canal P, respectivamente.

4.2. DISEÑO GEOMETRICO

Tomando como base las celdas y arreglos propuestos en el capítulo 3, el número de transistores MOS empleados en el circuito completo, se resume en la tabla 4.1. En el conteo de transistores por celda, se incluyen los transistores correspondientes a las fuentes de corriente

mostradas en forma simbólica en los diagramas del capítulo 3, así como las fuentes de corriente correspondientes a I_U e I_B y algunos espejos de corriente.

El programa L-Edit incluye bibliotecas de celdas pre-diseñadas de circuitos comunes, tanto analógicos como digitales, los cuales se pueden utilizar en un diseño en particular para lograr un ahorro en el tiempo de captura dentro del ambiente gráfico del programa. Otro tipo de celdas que se incluyen en las bibliotecas, son las correspondientes a las terminales (Pads) que permiten conectar el dado de silicio con las terminales del encapsulado mediante alambres de oro. En el circuito diseñado se emplearon cuatro de estas celdas: BareIOPad, CornerPad, VDDPad y GNDPad. La primera de ellas es una celda de entrada/salida que cuenta con diodos de protección contra descargas electrostáticas para las compuertas de los transistores MOS; la segunda se emplea en las esquinas del circuito como entrada/salida y no tiene diodos de protección. Las dos últimas son las que permiten introducir los voltajes de alimentación a todo el circuito.

Tipo de celda	No. de transistores	No. de celdas	Total de transistores
Generador de funciones de membresía	28	6	168
Circuito que realiza la operación MIN	16	9	144
Circuitos Auxiliares (I_B , I_U , espejos de corriente)	126	1	126
Compuerta de transmisión	2	30	60
Compuerta NAND	4	6	24
Compuerta NAND c/complemento	6	30	180
Flip Flop D c/Preset	28	3	84
Flip Flop D c/Clear	29	27	783
			1569

Tabla 4.1 Transistores utilizados en el circuito completo que integra las capas 1 y 2 de ANFIS.

En el circuito también se emplearon algunas de las celdas digitales de las bibliotecas de L-Edit. Las celdas utilizadas fueron: Inversor lógico, compuerta NAND de dos entradas con salida complementada, compuerta de transmisión (Transmission Gate), Flip Flop tipo D con *Preset* y Flip Flop tipo D con *Clear*. Todas ellas se usaron en la circuitería auxiliar para introducir los voltajes de ajuste de las campanas.

4.2.1. Celdas básicas.

Función de membresía.

La celda del circuito que genera la función de membresía tiene un área igual a $25200\mu\text{m}^2$ ($140\mu\text{m} \times 180\mu\text{m}$). El diseño geométrico de esta celda se muestra en la figura 4.1. Los nombres de las terminales mostradas, corresponden a los del circuito de la figura 3.6 del capítulo anterior.

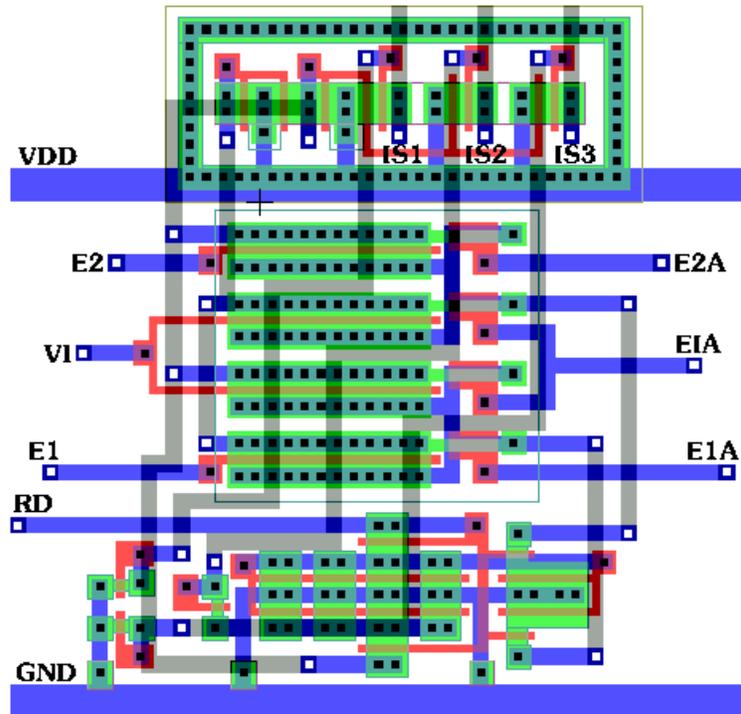


Fig. 4.1 Diseño geométrico del circuito generador de funciones de membresía.

Es necesario disponer de tres copias de la corriente I_s , pues ella se requiere en cada una de las tres reglas difusas en las que participa cada función de membresía, acorde con la construcción indicada en la figura 2.3(a). La terminal mostrada como R_D se conecta hacia un resistor exterior, con el cual se establece el valor de la fuente de corriente I_{μ} , así como las copias de ella que se requieren en esta celda.

Función MIN

El diseño geométrico de la siguiente celda básica es el correspondiente al circuito que realiza la operación MIN, basado en el circuito de competencia del tipo “el ganador toma todo”, definido por los circuitos de la figura 3.4(a) y (b), para el caso de $n = 2$. Las dimensiones de esta celda son $220\mu\text{m} \times 100\mu\text{m}$ y su diseño es el que se observa en la figura 4.2.

Los nodos IN_1 e IN_2 son las dos entradas donde se aplican las corrientes I_s producidas por las funciones de membresía, así como las corrientes de desplazamiento I_B , mencionadas en el capítulo previo. El nodo R_D se conecta a un resistor externo que define el valor de la corriente I_D .

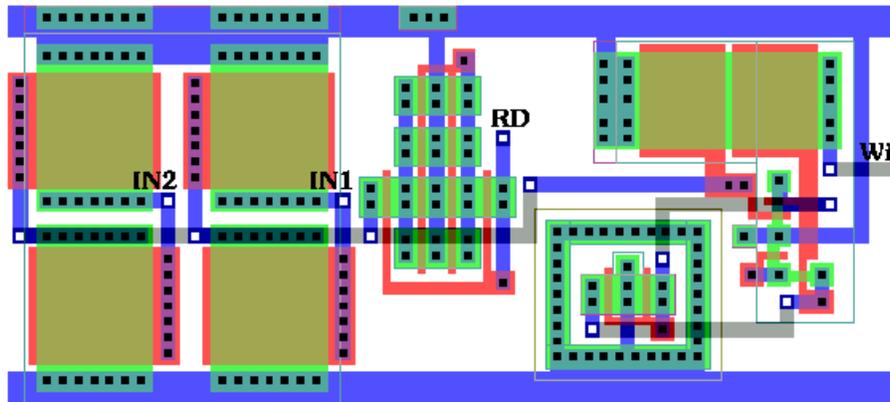


Fig. 4.2 Circuito que realiza la operación MIN.

En una corrida inicial se envió a fabricar un circuito integrado que contenía un circuito MIN de tres entradas y tres funciones de membresía. En este primer circuito integrado, las funciones de membresía tenían conectados todos sus nodos de control y su salida (a través de un espejo de corriente) a terminales exteriores, para poder realizar diferentes pruebas de su operación. Posteriormente, se envió a fabricar un segundo circuito integrado con las seis funciones de membresía y los nueve circuitos MIN de la arquitectura propuesta.

4.2.2. Construcción de la arquitectura (capas 1 y 2).

Las celdas básicas anteriores fueron hechas con un diseño geométrico que permite una sencilla interconexión entre ellas (entre celdas del mismo tipo), tomando en consideración que la alimentación (V_{DD} y GND) debe llegar a todas las etapas del circuito. El arreglo de celdas que incluye las seis funciones de membresía y los nueve circuitos MIN de ANFIS, se muestra en la figura 4.3 siguiente. En la parte inferior de la imagen se observan los circuitos de las funciones de membresía (A); sobre ellos están los circuitos MIN (B); la celda del lado derecho contiene las fuentes de corriente correspondientes a I_U e I_B (C). Todo este diseño abarca un área aproximada de silicio definida por $980\mu\text{m} \times 500\mu\text{m}$, lo cual toma en consideración el área que ocupan las interconexiones de metal entre celdas.

Las salidas de los circuitos MIN, que representan las salidas de todo el conjunto, se pasan a través de espejos de corriente. Esto se hace debido a que el circuito defusificador (capas 3, 4 y 5) al cual se deben conectar, requiere como entradas tanto las salidas MIN individuales, como una salida equivalente a la suma de todos los MIN. Estos espejos son los que se observan en la figura 4.4 que se muestra a continuación. La suma de los MIN se obtiene fácilmente mediante una suma de corrientes, conectando las salidas de los espejos a un nodo común (I_{SUM} en la figura).

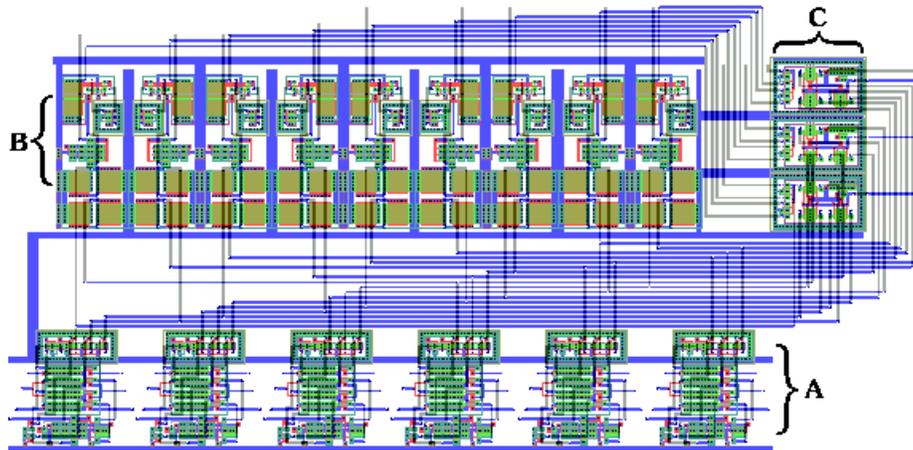


Fig. 4.3 Arreglo de 6 funciones de membresía y 9 funciones MIN

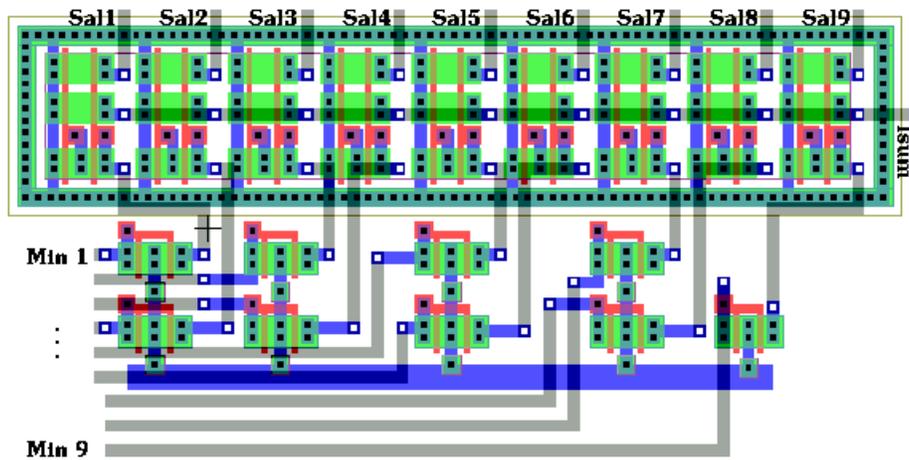


Fig. 4.4 Espejos de corriente de salida.

4.2.3. Estructuras auxiliares.

Como ya se ha mencionado, las funciones de membresía requieren de 5 voltajes para modificar los parámetros que controlan su posición, ancho y pendientes. La limitación del número de terminales del encapsulado en el que se fabrica el diseño, hace necesario que esos voltajes se introduzcan en forma serial. Ello se logra mediante un arreglo de registros de corrimiento como se explicó en el capítulo anterior, empleando capacitores y compuertas de transmisión (ver figura 3.10). Los registros se hicieron con 10 elementos de memoria cada uno (Flip Flops), para controlar dos campanas. Las celdas empleadas para construir este arreglo se tomaron de la biblioteca digital de L-Edit y son las siguientes: Inversor lógico (figura 4.5(a); dimensiones: $32\mu\text{m} \times 76\mu\text{m}$), compuerta NAND de dos entradas con salida complementada (figura 4.5(b); dimensiones: $40\mu\text{m} \times 76\mu\text{m}$), compuerta de transmisión (figura 4.5(c); dimensiones: $38\mu\text{m} \times 76\mu\text{m}$), Flip Flop tipo D con Clear (figura 4.6; dimensiones: $126\mu\text{m} \times 76\mu\text{m}$) y Flip Flop tipo D con Preset (figura 4.7; dimensiones: $122\mu\text{m} \times 76\mu\text{m}$). El inversor lógico se modificó para tener dos inversores en la misma celda, a fin de economizar área de silicio.

El otro componente importante de este arreglo son los capacitores hechos con los niveles Poly 1 y Poly 2. La celda individual tiene dimensiones de $72\mu\text{m} \times 82\mu\text{m}$ y el área de traslape tiene dimensiones de $60\mu\text{m} \times 78\mu\text{m}$.

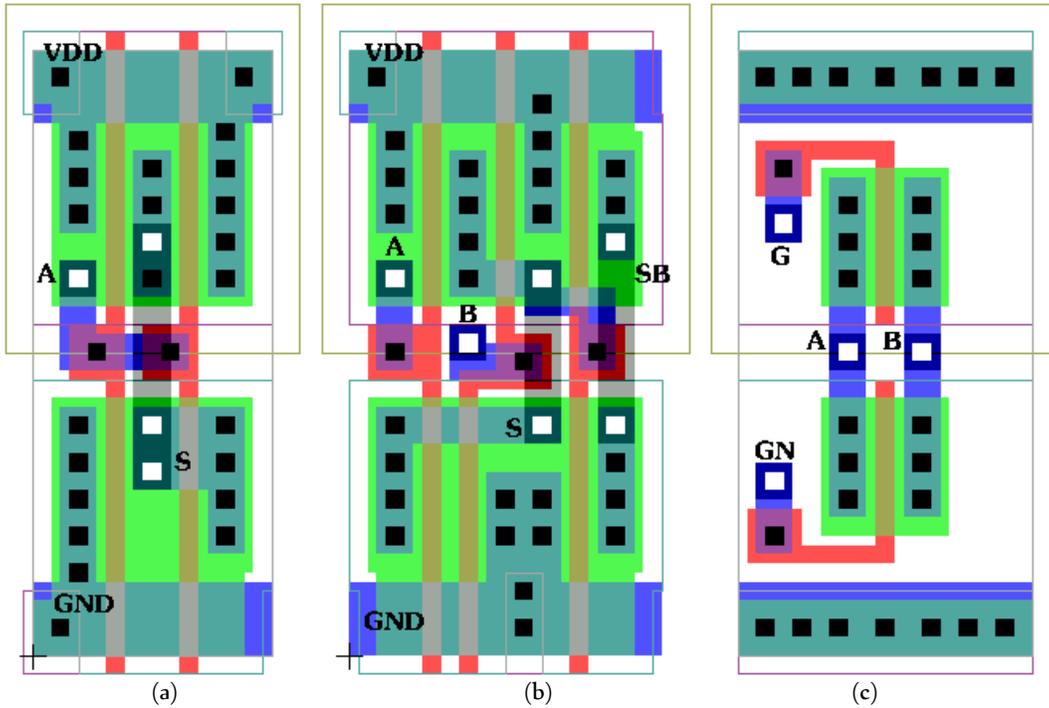


Fig. 4.5 Celdas de la biblioteca digital de L-Edit: a) Inversor lógico (doble);
 b) Compuerta NAND con salida complementada;
 c) Compuerta de transmisión.

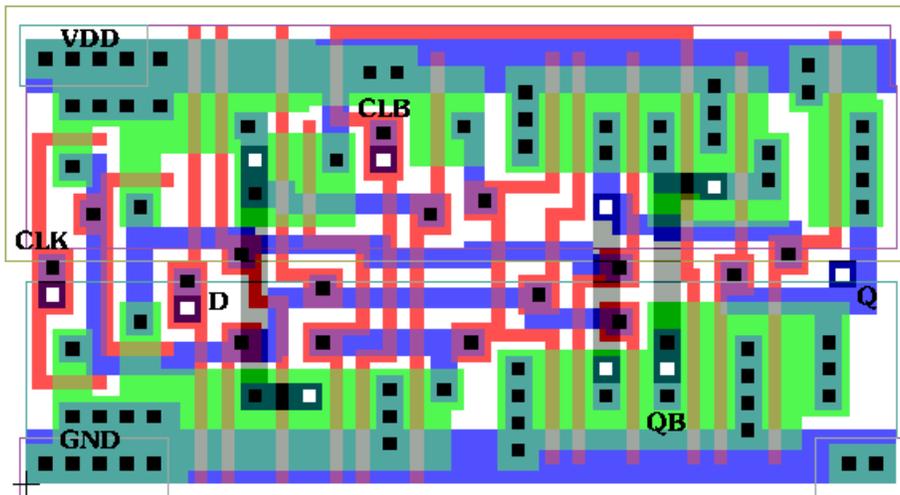


Fig. 4.6 Flip Flop tipo D con *Clear*.

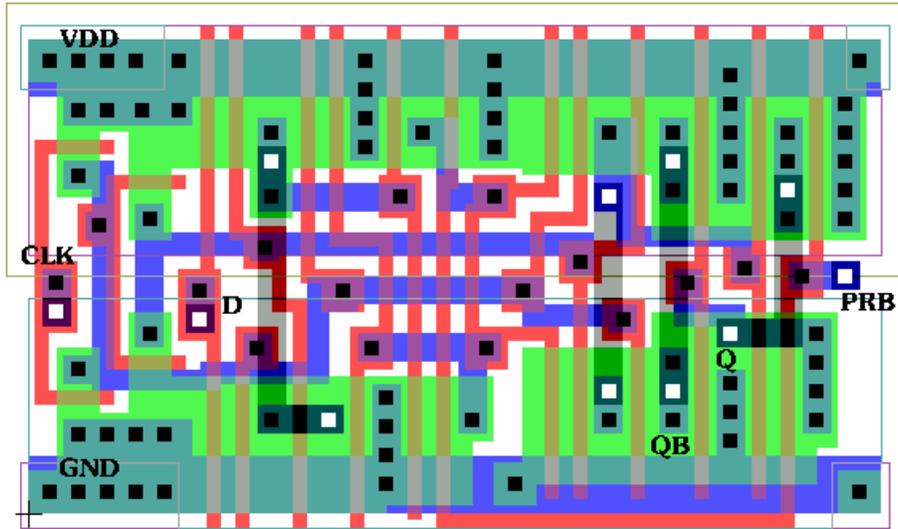


Fig. 4.7 Flip Flop tipo D con *Preset*.

El registro de corrimiento completo (10 elementos de memoria) tiene dimensiones aproximadas de $1078\mu\text{m} \times 322\mu\text{m}$, tomando en cuenta sus interconexiones internas. El diseño geométrico de este arreglo se muestra en la figura 4.8, donde los capacitores son los que se observan en la línea central (C_1 a C_{10}).

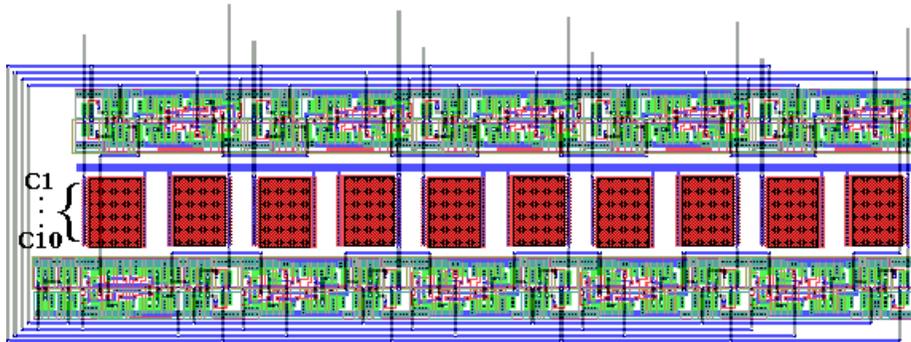


Fig. 4.8 Arreglo de registros de corrimiento para proporcionar los voltajes de ajuste de las funciones de membresía.

4.2.4. Circuito integrado completo.

Todos los circuitos anteriores, se conjuntan para formar el diseño geométrico completo del circuito integrado, que es resultado del diseño de la presente tesis. Este incluye las dos primeras capas de la arquitectura ANFIS, para 6 funciones de membresía y 9 funciones MIN. El diseño geométrico creado en L-Edit se muestra en la figura 4.9. La figura 4.10 muestra la fotografía al microscopio del circuito integrado que entregó MOSIS.

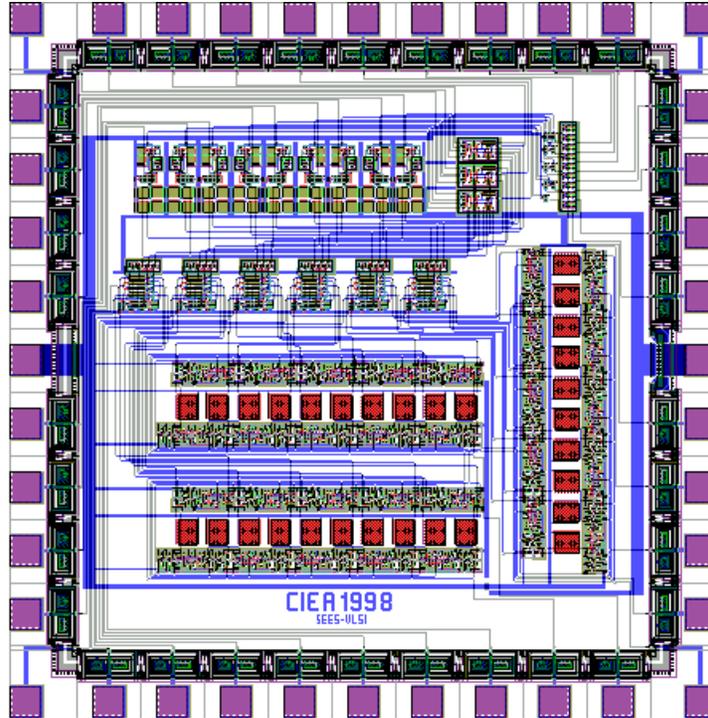


Fig. 4.9 Diseño geométrico completo del circuito integrado.

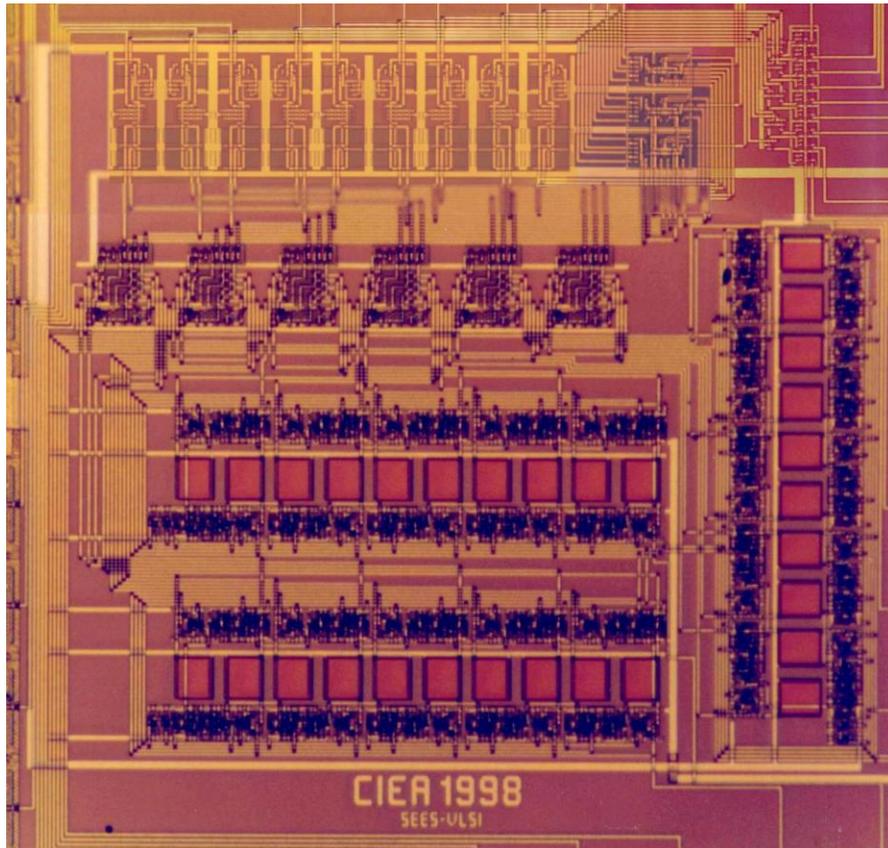


Fig. 4.10 Microfotografía del circuito integrado fabricado.

4.3. RESUMEN

En este capítulo se presenta el diseño geométrico de los diferentes circuitos utilizados para reproducir las funciones de las capas 1 y 2 de la arquitectura ANFIS, basado en el diseño del capítulo anterior. El número total de transistores en el circuito integrado es de 1569.

El programa utilizado para el dibujo de todas las celdas es L-Edit de Tanner Research Inc., con el cual se generó el archivo en formato CIF, necesario para la fabricación del circuito integrado. Esta se llevó a cabo mediante la organización académica MOSIS, para la tecnología de 2 μ m, pozo N, doble polisilicio y doble metal.

CAPITULO 5

APLICACION PROPUESTA

“Educar no es dar ‘carrera’ para vivir, sino temprar el alma para las dificultades de la vida.”
(J. Krishnamurti)

“Estudiar más, no para saber más, sino para saber mejor.”
(Lucio Anneo Séneca)

“El hombre que hace que las cosas difíciles parezcan fáciles es el educador.”
(R. U. Emerson)

5.1. INTRODUCCION

La clasificación de información en la forma de datos es un área de gran aplicación en la actualidad, en la cual se han utilizado diferentes metodologías, entre las que se cuentan: Métodos estadísticos, árboles de decisión, redes neuronales y algoritmos de agrupamiento (clustering). También es posible emplear un sistema difuso para la clasificación de datos, basado en reglas difusas del tipo *si ... entonces*.

Un clasificador difuso no es un reemplazo de los métodos arriba mencionados, sino una forma alternativa de lograr el mismo objetivo; sin embargo, un clasificador difuso presenta algunas ventajas adicionales: Puede manejar conocimiento vago; la clasificación se puede hacer manejando valores lingüísticos, lo cual da una mayor facilidad de interpretación de los resultados; desde el punto de vista de aplicación, el clasificador difuso es fácil de construir, de usar y entender.

Dentro de este esquema, en este trabajo se propuso el empleo del sistema ANFIS para la clasificación de señales electrocardiográficas (ECG), como un ejemplo de aplicación de dicho sistema. El problema que aquí se aborda es el de clasificación de arritmias cardíacas; sin embargo, debido a la amplia variedad de arritmias que se pueden presentar en una señal ECG, se determinó enfocar la detección hacia un tipo específico de arritmia, el cual, bajo ciertas condiciones, es signo de que en el paciente se puede desencadenar un problema serio del corazón.

El tipo de arritmia considerado es el que se genera debido a una contracción ventricular prematura (PVC), originada en un foco ectópico del sistema ventricular, la cual es identificable en el ECG. Estos latidos ectópicos, se detectan mediante un método llamado *predicción lineal*, con el cual se extraen rasgos característicos de la señal ECG. Los rasgos así extraídos forman el conjunto de datos que se utiliza para realizar el entrenamiento de ANFIS. La clasificación que realiza el sistema, permite reconocer entre tres condiciones de la señal ECG: Latidos con condición normal, latidos que presentan PVC y latidos no clasificados. Las señales ECG utilizadas, son tomadas de la base de datos de arritmias MIT-BIH [20], del Instituto Tecnológico de Massachusetts, la cual se ha considerado como una referencia para evaluar el desempeño del sistema desarrollado.

5.2. PRINCIPIOS DE ELECTROCARDIOGRAFIA

5.2.1. Antecedentes.

El corazón es una bomba muscular dividida en cuatro cámaras: Dos aurículas localizadas en la parte superior y dos ventrículos en la parte inferior. El corazón tiene la propiedad de contraerse automática y rítmicamente, así como la capacidad para iniciar y conducir los impulsos que provocan la contracción muscular. Estas propiedades se encuentran principalmente desarrolladas en un tejido neuromuscular especializado, el cual recibe el nombre de *sistema de conducción de impulsos del corazón*. A este sistema lo componen (Fig. 5.1): 1) El nodo sinoauricular; 2) el nodo aurículo-ventricular; 3) el haz de His; 4) las ramas derecha e izquierda del haz de His; y 5) el sistema de Purkinje. En un individuo normal, cada latido se inicia en la aurícula derecha, donde el grupo de células especializadas que componen el nodo sinoauricular, o marcapasos natural, envía una señal eléctrica que se propaga a través de las aurículas hacia el área entre las aurículas llamada el nodo aurículoventricular (AV). Este nodo conecta un grupo de

trayectorias especiales que conducen la señal debajo hacia los ventrículos. Conforme la señal viaja a través del corazón, éste se contrae: Primero se contraen las aurículas, bombeando sangre hacia los ventrículos y, una fracción de segundo después, los ventrículos se contraen, bombeando sangre hacia todo el cuerpo. Los potenciales eléctricos desarrollados durante este proceso, se indican en la gráfica del lado derecho de la figura 5.1. La suma de todos ellos genera la señal ECG que se puede detectar con un electrocardiógrafo, colocando electrodos en diferentes lugares de la superficie del cuerpo. Normalmente la señal ECG se registra en papel milimétrico, en el cual 5mm en el eje horizontal representan 0.2seg y 1cm en el eje vertical representa 1mV de amplitud.

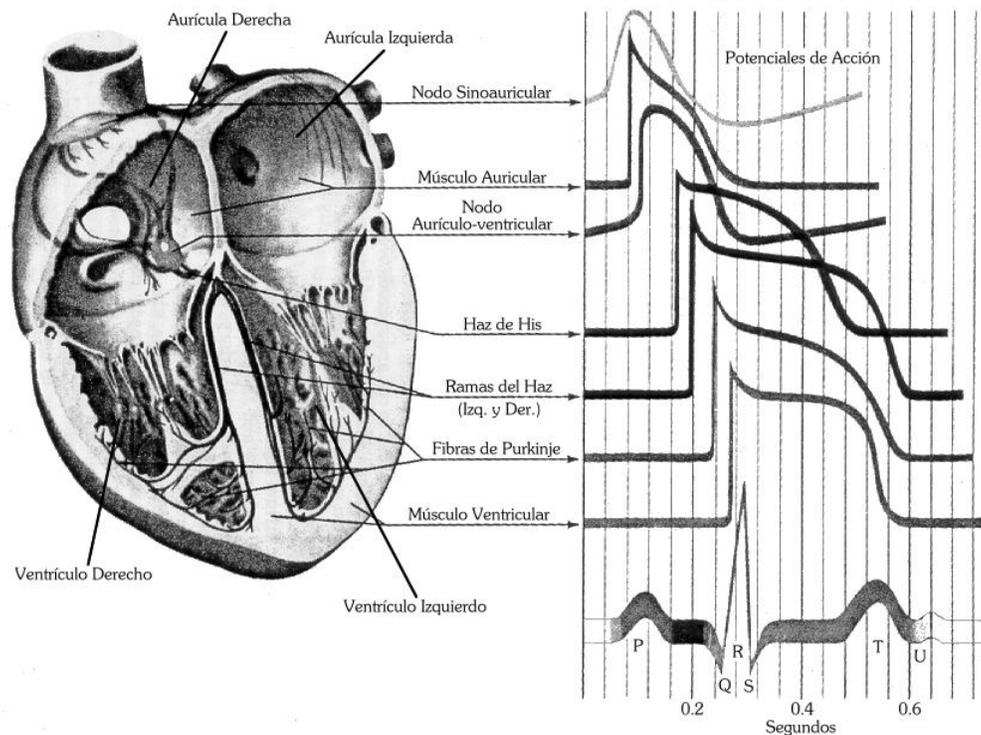


Fig. 5.1 Anatomía del sistema de conducción del corazón.

El ECG es de ayuda para el diagnóstico en los siguientes casos: a) Hipertrofia ventricular y auricular; b) isquemia e infarto del miocardio; c) pericarditis; d) padecimientos generalizados que afectan al corazón; e) evaluación del efecto de fármacos cardíacos; f) trastornos electrolíticos, en especial del potasio; y g) valorar la función de marcapasos cardíacos. Además es un auxiliar para la evaluación del retardo de la conducción de impulsos eléctricos ventriculares y auriculares, y para determinar el origen y comportamiento de las arritmias.

5.2.2. Derivaciones.

Las derivaciones estándar bipolares (I, II y III) son las derivaciones elegidas originalmente por Einthoven para registrar los potenciales eléctricos en el plano frontal del cuerpo [21]. Para ello se colocan electrodos en el brazo izquierdo (LA), brazo derecho (RA) y pierna izquierda (LL). La pierna derecha funciona como tierra y no realiza ninguna función en la producción del trazo electrocardiográfico. Las derivaciones bipolares muestran diferencias de potencial entre dos electrodos seleccionados: 1) Derivación I es la diferencia de potencial entre el brazo izquierdo y el

brazo derecho (LA-RA); 2) Derivación II, es la diferencia de potencial entre la pierna izquierda y el brazo derecho (LL-RA); y 3) Derivación III, es la diferencia de potencial entre la pierna izquierda y el brazo izquierdo. Estas derivaciones se muestran en la figura 5.2.

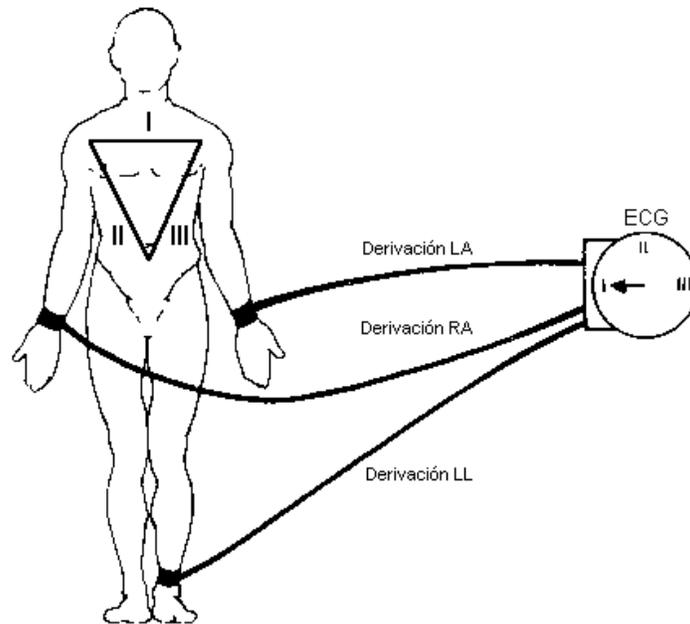


Fig. 5.2 Derivaciones estándar I, II y III. La relación entre ellas se expresa algebraicamente por la ecuación de Einthoven:
 derivación II = derivación I + derivación III

Las derivaciones unipolares (VR, VL y VF), las precordiales (V) y las esofágicas (E) fueron introducidas por Wilson en 1932; las unipolares intracardíacas lo fueron en 1942. Las derivaciones unipolares guardan una relación matemática bien definida con las bipolares estándar. Mediante una técnica automática incorporada en los aparatos electrocardiográficos modernos, se puede aumentar la amplitud de las señales VR, VL y VF en aproximadamente 50%. Estas son llamadas derivaciones unipolares aumentadas, y se designan como aVR, aVL y aVF. Las derivaciones precordiales se obtienen monitoreando varios sitios del precordio, como se muestra en la figura 5.3.

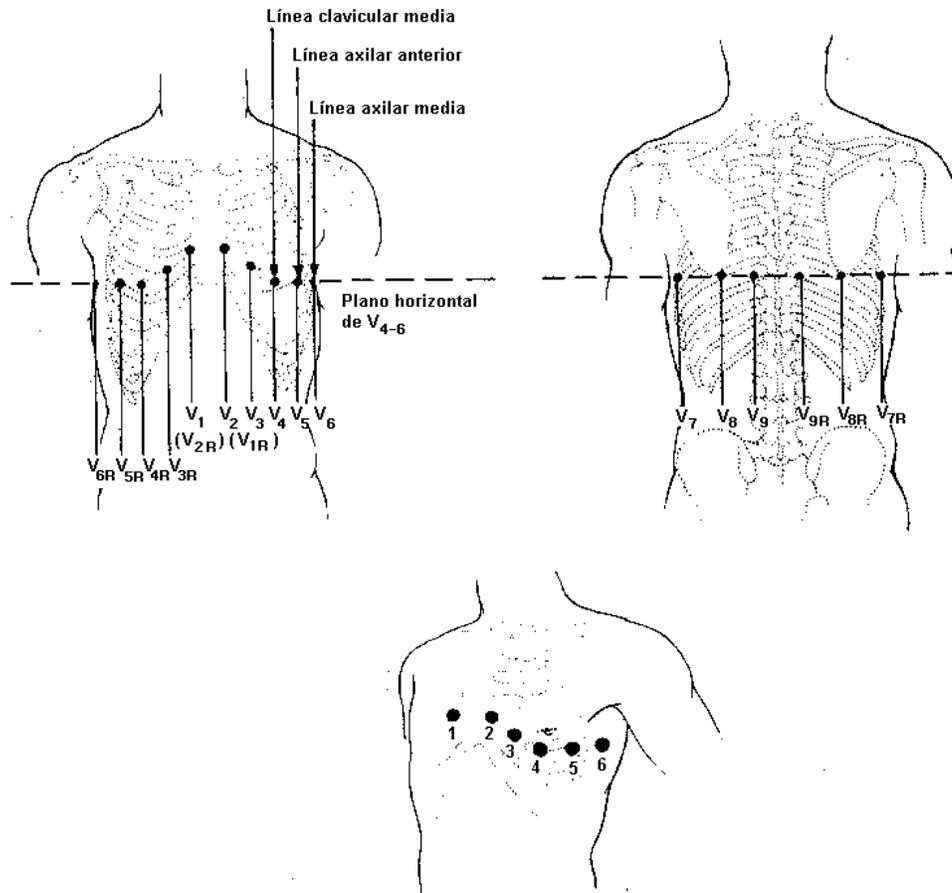


Fig. 5.3 Localización de las derivaciones precordiales unipolares.

5.2.3. El electrocardiograma normal.

Para entender en qué consisten las anomalías en el funcionamiento del corazón, primero se debe conocer cómo es el funcionamiento normal del mismo.

La generación del ECG depende de cuatro fenómenos electrofisiológicos: 1) Formación del impulso en el marcapaso principal del corazón, que normalmente es el nodo sinoauricular; 2) transmisión del impulso a través de fibras especializadas en la conducción; 3) activación (despolarización) del miocardio; y 4) recuperación (repolarización) del miocardio. El ritmo normal del corazón es entre 60 y 100 contracciones por minuto; cada contracción representa un latido.

La señal ECG se compone de complejos, intervalos, uniones, segmentos y ondas (figura 5.4). La forma y el período de ellos permiten interpretar el ECG para su diagnóstico. El segmento horizontal que precede a la primera deflexión se conoce como línea de base, basal o de isopotencial.

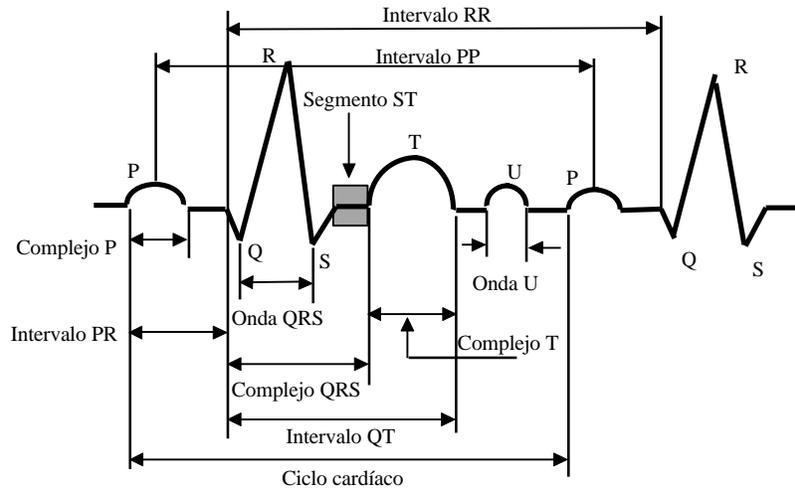


Fig. 5.4 Esquema de los complejos, intervalos y segmentos de la señal ECG.

Cada onda del ECG está relacionada con una acción específica del sistema cardíaco que genera diferentes potenciales, como se mostró en la gráfica de la figura 5.1 anterior:

Activación auricular.

La onda P es la deflexión producida por la despolarización auricular. La onda T_a (localizada entre las ondas P y Q) es la deflexión producida por la repolarización auricular. Esta deflexión no suele observarse en el ECG de 12 derivaciones.

Activación ventricular.

La onda Q (q) es la deflexión inicial negativa producida por la despolarización ventricular. La onda R (r) es la primera deflexión positiva producida por la despolarización ventricular. La onda S (s) es la primera deflexión negativa de despolarización ventricular que sigue a la primera onda positiva (R). Una onda QS es una deflexión negativa que no sobrepasa la línea basal.

Repolarización ventricular.

La onda T es la deflexión producida por la repolarización ventricular. La onda U es la deflexión (casi siempre positiva) que sigue a la onda T y precede a la siguiente onda P; su causa exacta no se conoce, pero se cree que resulta de la repolarización lenta del sistema de conducción intraventricular (Purkinje).

Intervalos normales.

El intervalo RR tiene lugar entre dos ondas R consecutivas. Con ritmo ventricular regular, al dividir 60 (seg) entre el intervalo entre dos ondas R consecutivas, medido en segundos, se obtiene la frecuencia cardíaca por minuto. Si el ritmo ventricular es irregular, se debe contar el número de ondas R en un determinado periodo de tiempo y la cuenta convertirla en número de ondas R por minuto. Los intervalos RR se miden al inicio de los complejos QRS y no en los vértices de las ondas R.

Al intervalo PP lo definen dos ondas P consecutivas. Con ritmo sinusal regular, el intervalo PP es igual al intervalo RR. Sin embargo, con ritmo ventricular irregular, o cuando las

frecuencias auricular y ventricular son regulares pero diferentes entre sí, el intervalo PP debe medirse desde el mismo punto en dos ondas P sucesivas y calcular la frecuencia auricular por minuto, en la misma forma que se calcula la frecuencia ventricular por minuto.

El intervalo PR está comprendido entre el comienzo de la onda P y el inicio del complejo QRS. Mide el tiempo de conducción aurículoventricular (AV) e incluye el tiempo necesario para la despolarización auricular, el retardo normal de la conducción en el nodo AV (cerca de 0.07seg) y el paso del impulso a través del haz de His y sus dos ramas, hasta el comienzo de la despolarización ventricular. Normalmente varía entre 0.12 y 0.2seg.

El intervalo QRS representa el tiempo de despolarización ventricular. Se mide desde el comienzo de la onda Q (o de la onda R si no existe la Q) hasta el final de la onda S. El límite normal superior es de 0.1seg en las derivaciones del plano frontal y de 0.11seg en las precordiales.

El tiempo de activación ventricular (TAV) es el tiempo necesario para que un impulso atraviese el miocardio, desde el endocardio hasta el epicardio. Corresponde al intervalo medido desde el comienzo de la onda Q al vértice de la onda R.

El intervalo QT se mide desde el comienzo de la onda Q hasta el final de la onda T; indica la duración de la sístole eléctrica. Varía con la frecuencia cardíaca y por efecto del sistema nervioso autónomo. A veces el final de la onda T no está bien definido, o puede haber una onda U superpuesta a la T. En estos casos no es posible medir el intervalo QT correctamente.

Segmentos normales.

El segmento PR es la porción del ECG comprendida entre el final de la onda P y el comienzo del complejo QRS; normalmente es isoelectrico.

El Punto J (unión RST), es el punto donde termina el complejo QRS y comienza el segmento ST.

El segmento ST es la parte del ECG que está entre el punto J y el comienzo de la onda T. Este segmento suele ser isoelectrico, pero en las derivaciones precordiales puede variar un poco. Se define como elevado o deprimido al relacionarlo con la porción de la línea basal entre el final de la onda T y el comienzo de la onda P (segmento TP), o al relacionarlo con el segmento PR. Si ninguno de los dos segmentos es isoelectrico, entonces se puede usar el comienzo del complejo QRS.

El segmento TP es la parte del ECG entre el final de la onda T y el comienzo de la siguiente onda P. Con frecuencia cardíaca normal, el segmento suele ser isoelectrico; con frecuencias rápidas, la onda P se superpone a la onda T, con lo cual desaparece el segmento TP isoelectrico.

5.2.4. Variaciones del electrocardiograma normal.

Las mediciones del ECG consideradas normales se determinan estudiando a muchos individuos clínicamente normales. Los límites arbitrarios de "normalidad" se encuentran entre el 95 y 98%, de tal forma que, para cualquier medición electrocardiográfica, de 2 a 5% de personas normales presentará un ECG anormal. Entre las causas conocidas que hacen que un ECG cambie, están: Edad, sexo, peso corporal, posición del corazón en el interior de la caja torácica, configuración del tórax, raza, temperatura ambiental, consumo de alimentos, ejercicio, tabaquismo, hiperventilación y colocación de los electrodos precordiales.

El término "variante normal" de un ECG indica una señal obtenida en una persona normal con algunas variaciones en la morfología P-QRST. Un ECG anormal registrado en un

individuo sano y asintomático no se considera necesariamente como cardiópata. Puesto que algunos patrones anormales de ECG que se presentan en individuos normales pueden simular varias formas de cardiopatía, para hacer una interpretación correcta del ECG, se debe realizar una correlación clínica precisa.

Tipos de arritmias.

Una arritmia es el cambio en el ritmo normal del corazón. Las arritmias se identifican por el lugar del corazón donde se originan (aurículas o ventrículos) y por lo que le sucede al ritmo del corazón cuando ellas ocurren. Algunos de los tipos de arritmias se describen brevemente a continuación.

1) Originadas en las aurículas.

- ✧ Arritmia sinusal: Cambios cíclicos en el ritmo sinusal (normal) del corazón durante la respiración, con aumento y disminución alternantes en la frecuencia de los latidos. Es común en niños y frecuente en adultos.
- ✧ Taquicardia sinusal: Genera estímulos en el seno auricular con una frecuencia mayor a 100 por minuto. Esta es una reacción fisiológica normal al esfuerzo, la emoción, la digestión, algunos estimulantes del tipo de la adrenalina, la cafeína y algunos tóxicos. En el ECG simplemente se aprecia el aumento en la frecuencia cardíaca y P, QRS y T pueden ser normales, dependiendo de la causa que provoca la taquicardia.
- ✧ Bradicardia sinusal: Resulta de la iniciación de estímulos por el seno a una frecuencia menor a 60 por minuto. Puede tratarse de una reacción fisiológica al reposo, debido a la acción de ciertas drogas o al aumento de la presión intracraneana. El ECG muestra complejos P, QRS y T esencialmente normales y la disminución en la frecuencia antes mencionada.
- ✧ Síndrome del seno enfermo: El nodo del seno no dispara su señal adecuadamente, de tal forma que el ritmo del corazón disminuye. A veces el ritmo cambia arriba y abajo, entre un ritmo rápido (taquicardia) y uno lento (bradicardia).
- ✧ Contracción supraventricular o contracción auricular prematura (PAC): Un latido ocurre anticipadamente en la aurícula, originando que el corazón se contraiga antes del siguiente latido regular.
- ✧ Taquicardia supraventricular (SVT), taquicardia auricular parasistólica (PAT): Una serie de latidos anticipados en la aurícula aceleran el ritmo del corazón. En la PAT, periodos repetidos de latidos muy rápidos inician y terminan repentinamente.
- ✧ Aleteo auricular. Señales disparadas rápidamente hacen que los músculos en la aurícula se contraigan velozmente, llevando a un ritmo cardíaco muy rápido y estable.
- ✧ Fibrilación auricular: En las aurículas se disparan señales eléctricas en forma muy rápida y descontrolada. A los ventrículos les llegan señales eléctricas de un modo completamente irregular, de tal forma que los latidos del corazón son completamente irregulares.
- ✧ Síndrome de Wolff-Parkinson-White: Trayectorias anormales entre las aurículas y los ventrículos hacen que lleguen señales eléctricas a los ventrículos en forma adelantada y entonces se transmiten de nuevo hacia las aurículas. Como consecuencia, se pueden desarrollar ritmos cardíacos muy rápidos durante los *rebotes* entre las aurículas y los ventrículos.

2) Originadas en los ventrículos.

- ❖ Complejos ventriculares prematuros (PVC): Una señal eléctrica de los ventrículos origina un latido anticipado. El corazón, entonces, parece tener una pausa hasta que ocurre el siguiente latido del ventrículo en forma normal.
- ❖ Taquicardia ventricular: El corazón late rápido debido a señales eléctricas que llegan de los ventrículos (en vez de llegar de las aurículas).
- ❖ Fibrilación ventricular: En los ventrículos se disparan señales eléctricas de una forma muy rápida y descontrolada, haciendo que el corazón se estremezca en vez de latir y bombear sangre.

La aplicación propuesta para el sistema ANFIS, se enfoca hacia la detección del tipo de arritmia PVC, por lo que a continuación se hará una breve semblanza de las características que distinguen esta cardiopatía.

Arritmias ventriculares.

Los complejos ventriculares prematuros (PVC) pueden surgir de un foco ectópico situado en cualquier parte del miocardio ventricular o por reentrada de un impulso (ventricular o supraventricular) a una zona de tejido ventricular. La característica principal que se manifiesta en el ECG es la presencia de un complejo QRS que difiere notablemente de los otros QRS del registro por su forma y su tamaño. Los complejos ventriculares son anchos y de aspecto anormal, con una duración superior a 0.12seg (figura 5.5). El segmento ST y la onda T suelen estar desplazados en sentido opuesto al de la deflexión principal de QRS. A veces, un complejo PVC se produce entre dos impulsos sinusales sin modificar el ritmo sinusal, lo que se conoce como complejo PVC interpolado (figura 5.6).

Los PVC pueden tener morfologías similares o diferentes en una misma derivación del ECG. Los PVC uniformes (monomorfos o unimorfos) se pueden originar en un mismo foco o debido a un mecanismo de reentrada (figura 5.7); los de diferente morfología (multiformes, multimorfos o polimorfos) pueden surgir de varios focos (multifocales) o, más común, de un solo foco, pero conducidos por diferentes caminos al despolarizar el miocardio ventricular, originando diferentes morfologías de QRS (figura 5.8).

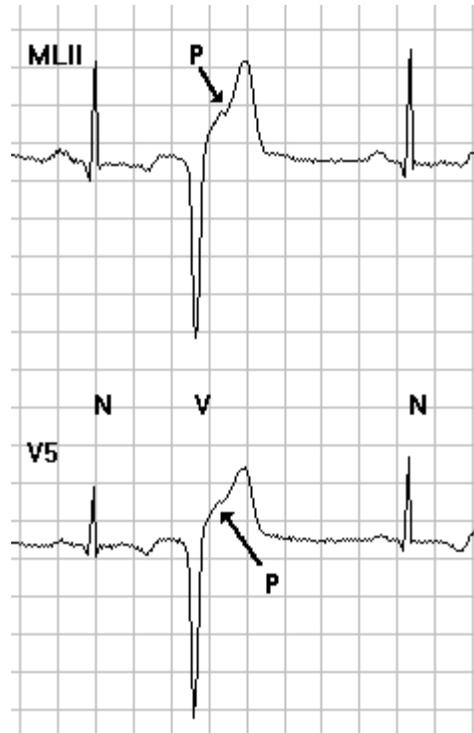


Fig. 5.5 Complejos ventriculares prematuros. Los complejos son anchos, alargados y anormales; les siguen ondas P invertidas.

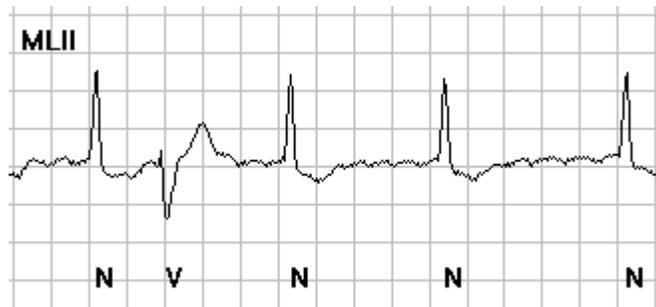


Fig. 5.6 Complejo PVC interpolado. Después del primer complejo QRS de origen sinusal, hay un PVC, ancho y anormal, que no modifica el ritmo sinusal.



Fig. 5.7 Complejos PVC uniformes. Existe un intervalo de acoplamiento fijo entre los latidos sinusales y los PVC.

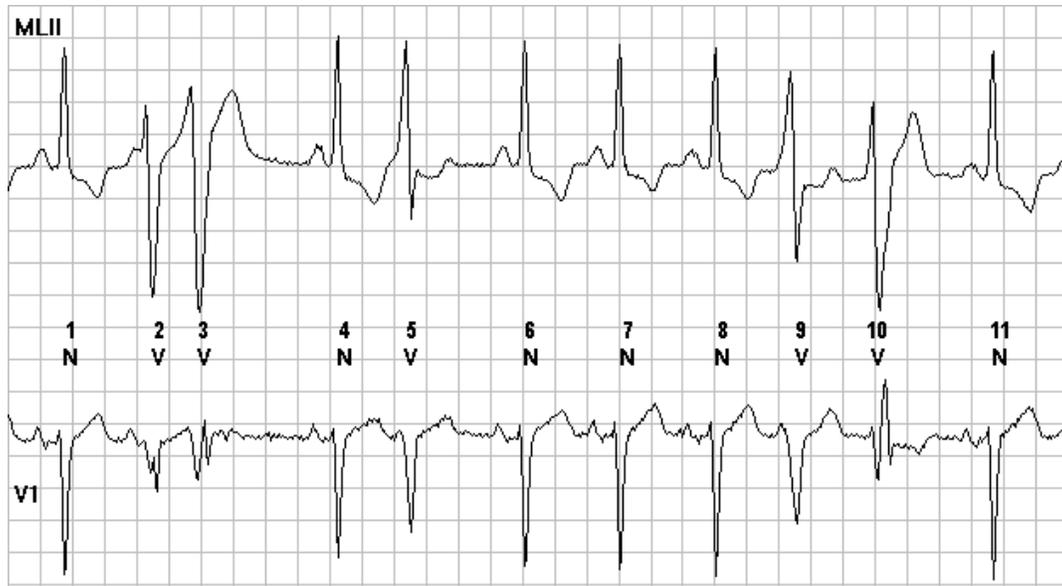


Fig. 5.8 Complejos PVC multiformes: Los complejos 2, 3 y 10 se dirigen hacia atrás (ondas S profundas en MLII), lo que sugiere un foco de origen en el ventrículo derecho. Los complejos 5 y 9 se dirigen hacia adelante (ondas R altas en MLII), lo que sugiere un foco de origen en el ventrículo izquierdo.

Puesto que el intervalo QT es aproximadamente igual al periodo refractario del tejido ventricular, los complejos PVC suelen inscribirse después de la onda T del latido precedente. Sin embargo, a veces el PVC se puede presentar en el vértice de la onda T o en su descenso (figura 5.9). Los PVC con R en T pueden ocurrir en el *periodo vulnerable* del tejido ventricular, lo cual desencadena latidos ventriculares repetidos (taquicardia o fibrilación); por ello, estos latidos se consideran potencialmente peligrosos (figura 5.10).

Por lo antes mencionado, resulta importante la temprana detección de los PVC, ya que, por ejemplo, una fibrilación ventricular puede resultar fatal; el único tratamiento eficaz para ésta es la desfibrilación eléctrica, aunque hay casos contados en que la fibrilación desaparece espontáneamente.

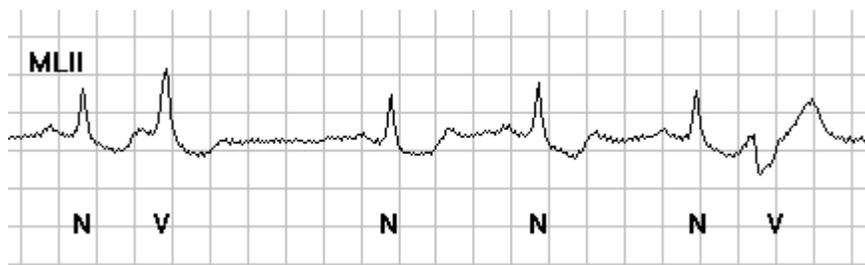


Fig. 5.9 Complejos PVC con R en T. Los PVC se producen cerca del vértice de la onda T, cuando los ventrículos podrían ser más vulnerables a una taquicardia o fibrilación.

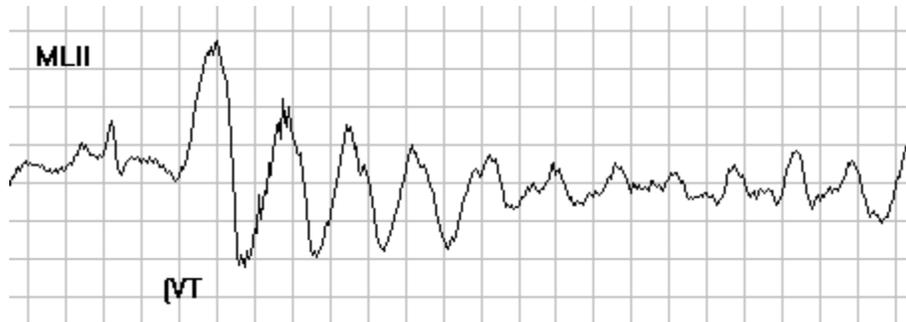


Fig. 5.10 Complejo PVC con R en T que ocurre en el vértice de la onda T del latido sinusal precedente. El PVC cae durante el periodo vulnerable del miocardio ventricular y desencadena una taquicardia ventricular polimorfa.

Debido a la importancia que tienen los complejos PVC dentro de las arritmias cardíacas, se pretende utilizar el sistema ANFIS para la identificación de tales complejos. El método empleado para ello, se describe en el siguiente apartado.

5.3. OBTENCION DE LOS DATOS DE ENTRENAMIENTO PARA EL SISTEMA ANFIS

El conjunto de datos que se utiliza para entrenar al sistema ANFIS, debe ser de tal forma que esté contenido dentro del espacio difuso bidimensional establecido como universo de discurso para las dos variables de entrada del sistema (X y Y). En este ejemplo de aplicación, los datos de entrenamiento se generan a partir rasgos característicos obtenidos de señales de la base de datos de arritmias MIT-BIH, para los casos que presentan PVC.

5.3.1. Preprocesamiento de la señal ECG.

El preprocesamiento de la señal se realiza con el objetivo de reducir interferencia debida al ruido, y extraer el mínimo conjunto de parámetros que representen adecuadamente a cada segmento QRS, sin sacrificar el desempeño del clasificador neurodifuso.

El paso inicial es el filtrado de la señal ECG. Para ello se utiliza un filtro pasa altas con una frecuencia de corte de 1Hz, seguido por un filtro pasa bajas con una frecuencia de corte de 20Hz y rechazo de banda en 60Hz, con lo cual se obtiene un pasa banda con ranura en 60Hz. Este primer paso permite reducir el ruido de la señal, que puede originarse por la línea de energía eléctrica (60Hz), la respiración del paciente (-0.2Hz), contracciones musculares (desde C.D. hasta 10KHz), contacto con los electrodos del electrocardiógrafo, etc.

El filtro pasa altas se realizó empleando el Toolbox de Procesamiento de Señales de MATLAB. Se construyó un filtro tipo Chebyshev de orden 2 con una frecuencia de corte de 1Hz, basado en una frecuencia de muestreo $f_s = 360\text{Hz}$, correspondiente a la frecuencia a la cual están digitalizadas las señales de la base MIT-BIH. Los coeficientes obtenidos para este filtro tienen los siguientes valores:

$$\begin{aligned} a_{1 \dots 3} &= 1.0000, \quad -1.9761, \quad 0.9764 \\ b_{1 \dots 3} &= 0.9881, \quad -1.9762, \quad 0.9881 \end{aligned}$$

El filtro pasa bajas se basó en un filtro de coeficientes enteros, también diseñado para una frecuencia $f_s = 360\text{Hz}$. La función de transferencia está dada por la siguiente expresión:

$$L(z) = \frac{1 - 2z^{-6} + z^{-12}}{1 - 2z^{-1} + z^{-2}} \quad (5.1)$$

Utilizando ambos filtros en cascada, se construye el filtro pasa banda completo, cuya respuesta en magnitud y fase se muestra en la figura 5.11 siguiente. El filtro tiene una ganancia de 32dB en la banda de interés, lo cual se debe compensar dividiendo la señal, una vez que ella se ha filtrado, por una constante de valor adecuado. Las frecuencias donde la ganancia máxima decae 3dB, se localizan aproximadamente en 1 y 20Hz, correspondientes a las frecuencias de corte del circuito.

Con la señal ECG ya filtrada, lo que sigue es la localización de los complejos QRS. En la base de datos MIT-BIH, las señales incluyen archivos de anotaciones en los cuales se ha registrado, latido por latido, el tipo de afección que presenta cada paciente, así como cambios en su ritmo cardíaco. Cuando se emplea esta base, se pueden usar estos archivos de anotaciones para localizar con precisión los complejos QRS; en el caso de utilizar cualquier otra señal que no tenga anotaciones (por ejemplo, la señal proveniente de un electrocardiógrafo), se requiere de un algoritmo de detección de complejos QRS (por ejemplo, el propuesto en [22]).

Una vez que se conoce la ubicación de los complejos, para cada segmento que contiene un complejo QRS, se generan dos coeficientes de predicción lineal (LPC) y se calcula su valor cuadrático medio, esto es: $\overline{X^2} = E(x^2)$, donde E es la esperanza matemática y x son los valores de las muestras en cada segmento. Ambos coeficientes junto con el valor cuadrático medio formarán el conjunto de datos de entrenamiento de ANFIS.

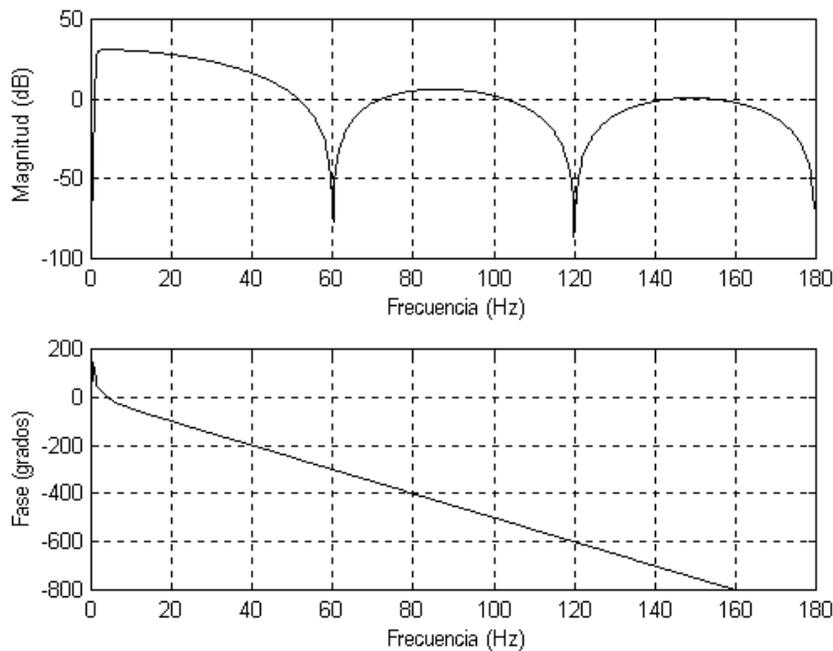


Fig. 5.11 Respuesta en magnitud y fase del filtro pasabanda.

Para generar los segmentos de los cuales se van a extraer los coeficientes LPC, se toma un cierto número de muestras antes y después del centro de cada complejo QRS, de tal forma

que se asegure que en tal segmento esté contenido un complejo completo. La duración normal de un complejo QRS típico es de 0.1seg, mientras que un complejo PVC puede durar más del doble de este tiempo; así, un valor inicial para la longitud de los segmentos es de 100 muestras, con el pico de la onda R en la muestra 51, lo cual es aproximadamente 0.28seg. Después de probar con varios valores de la longitud de los segmentos, los mejores resultados (en cuanto a la distribución gráfica de los coeficientes LPC, descrita más adelante) se obtuvieron para una longitud de 80 muestras.

Antes de que se calculen los coeficientes LPC, primero cada segmento QRS se debe centrar en su media (esto es, se calcula la media para cada segmento y se resta de las muestras del segmento) en el eje vertical. A continuación, los valores de las muestras de cada segmento son escalados para quedar entre el intervalo de +1 y -1, con el fin de reducir variaciones en la amplitud de los complejos. El último paso es aplicar una ventana de Hamming a cada segmento para reducir discontinuidades y resaltar la señal R. La figura 5.12 muestra el resultado de este proceso para un segmento que contiene un complejo QRS.

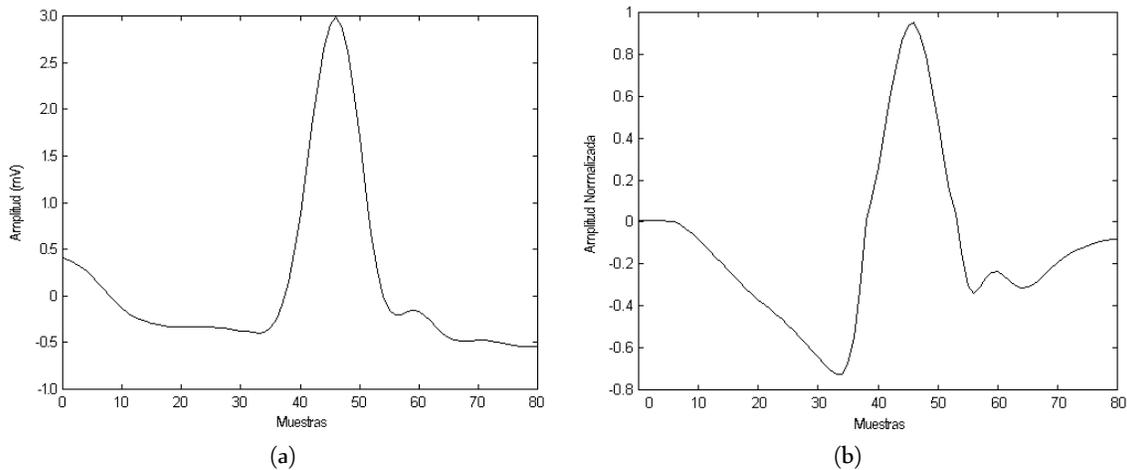


Fig. 5.12 a) Segmento que contiene un complejo QRS después del filtro pasa banda; b) El mismo segmento después de aplicarle la ventana de Hamming.

5.3.2. Conjunto de datos de entrenamiento.

Una vez normalizados los segmentos de los complejos QRS, se calculan los coeficientes LPC. La predicción lineal modela cada muestra sucesiva de una señal como una combinación lineal de muestras previas. Para obtener los coeficientes, se utiliza el método de máxima entropía de Burg, del cual, una descripción detallada se encuentra en [23]. La predicción óptima de la muestra presente, x_n , está dada por:

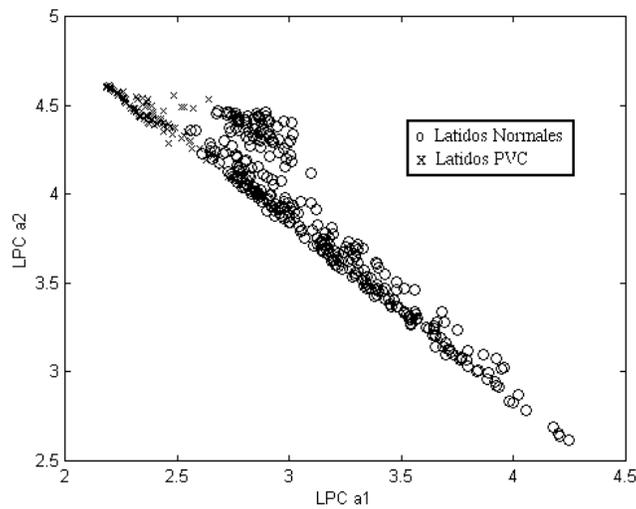
$$x_n = \sum_{i=1}^p a_i x_{n-i} \tag{5.2}$$

donde x_n es la serie en el tiempo de la entrada real, a_i es el i -ésimo coeficiente de predicción lineal y p es el orden del denominador polinomial $a(z)$ (también llamado predictor lineal), esto es, $a=[1, a(1), \dots, a(p+1)]$; el problema consiste en encontrar el valor de la matriz a . De acuerdo con [24], cuando se realiza el análisis de predicción lineal a una señal ECG, es suficiente con un

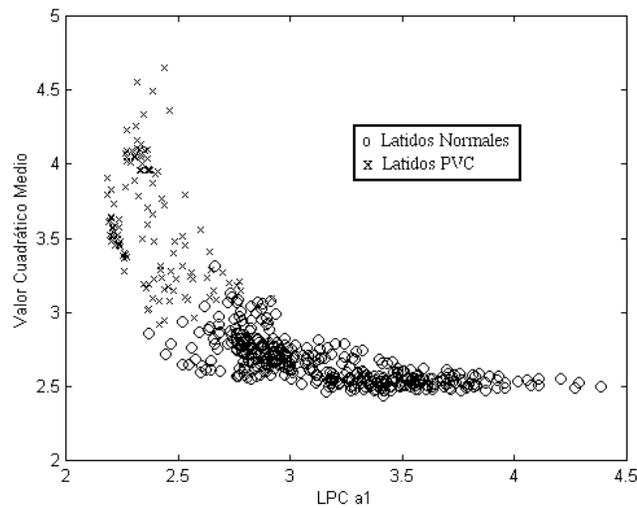
predictor de 2^{do} orden, esto es, $p=2$ en la ec. 5.2 anterior; en ese trabajo se señala que al incrementar el orden del predictor no se reduce significativamente el error de predicción, lo cual significa que el orden no debe ser mayor que 2 para detección de arritmias cardíacas.

Como se mencionó anteriormente, el valor cuadrático medio de cada complejo QRS está dado por $\bar{X}^2 = E(x^2)$, lo cual se aplica para todas las muestras de cada segmento. Con los dos coeficientes y el valor cuadrático de cada complejo, se obtiene el conjunto de datos que se utilizarán posteriormente para entrenar y probar al sistema ANFIS. La figura 5.13(a) muestra la gráfica de ambos coeficientes, a_1 y a_2 , para varios complejos QRS de señales que presentan latidos PVC; la figura 5.13(b) muestra la gráfica de la media cuadrática contra el coeficiente LPC a_1 , para los mismos complejos. Los valores de los coeficientes y del valor cuadrático medio se han escalado en ambas gráficas, para ajustarlos a los valores del espacio difuso que se obtienen al emplear los circuitos CMOS diseñados. Los intervalos originales obtenidos para los coeficientes LPC y el valor cuadrático medio son: $a_1 = [0.135, 0.18]$, $a_2 = [0.75, 1]$, $\bar{X}^2 = [0.05, 0.3]$.

El conjunto de datos se generó a partir de señales de la base de datos MIT-BIH. Se tomaron siete pacientes que presentan latidos PVC, los cuales se identifican en la base con los números: 100, 114, 116, 119, 208, 221 y 233. El registro de cada paciente en la base tiene una duración aproximada de 30 minutos, con alrededor de 2000 complejos QRS por registro, llegando a más de 3000 complejos por registro en algunos casos. Por ello, y para mayor claridad, en las figuras 5.13 únicamente se tomaron algunos latidos que representen la distribución que se obtiene para los coeficientes LPC y el valor cuadrático medio en el espacio difuso. En este caso se consideraron 100 latidos PVC y 500 latidos normales. Sin embargo, para hacer el entrenamiento se consideró un número mayor de complejos.



(a)



(b)

Fig. 5.13 Distribución de los coeficientes LPC obtenidos para latidos normales y PVC de diferentes señales ECG: (a) Gráfica de a_2 vs. a_1 ; (b) Gráfica del valor cuadrático medio vs. a_1 .

5.3.3. Entrenamiento de ANFIS con los datos obtenidos.

La metodología a seguir para poder hacer el entrenamiento de ANFIS a partir del conjunto de datos obtenido previamente, requiere de una función auxiliar para agrupar los datos obtenidos de acuerdo a su condición (normal, PVC o no clasificado). Para lograr este propósito, existen diferentes técnicas de agrupamiento (clustering), algunas de las cuales emplean una forma de medir la distancia entre un dato dado y los diferentes grupos (clusters) para asignarle un grado de pertenencia a cada uno de ellos.

En este trabajo se hicieron pruebas usando el método denominado la *Media C difusa* (Fuzzy C-Mean) y el método de *Máxima Verosimilitud* (Maximum Likelihood), siendo este último el que produjo los mejores resultados. La fig. 5.14 se obtuvo al aplicar el criterio de máxima verosimilitud a los datos de entrenamiento obtenidos anteriormente, mostrados en la fig. 5.13(b), a fin de asignarles un grado de pertenencia tanto a los datos que representan una condición normal, como a los que representan la condición PVC. Para ello, se tomaron aleatoriamente 1500 latidos de 7 pacientes de la base MIT-BIH (1000 para la condición normal y 500 para la condición PVC); de estos latidos, la mitad se utilizó para entrenamiento del sistema, y la otra mitad fue para probar su respuesta. En la figura 5.14 se observa un valle, el cual corresponde al grado de pertenencia de los datos que representan la condición PVC, mientras que la cima mostrada en la misma figura, corresponde al grado de pertenencia de los datos que representan latidos normales; la parte plana (central) corresponderá a latidos no clasificados. La elección de la distribución de la figura 5.13(b) sobre la 5.13(a) se debe a que aquella presenta una distribución de los datos más adecuada para la técnica de agrupamiento que se utilizó.

El ajuste de las funciones de membresía se realizó en 50 épocas de entrenamiento. Se observó que un aumento en el número de épocas no modificaba significativamente el error del sistema. Como ya se mencionó en otro capítulo, el entrenamiento genera los valores óptimos de los parámetros de las funciones de membresía, así como los coeficientes de los polinomios en la capa 4 de ANFIS. Puesto que a este trabajo únicamente corresponden las dos primeras capas de

la arquitectura, tan sólo se presentan las funciones de membresía obtenidas después de entrenar al sistema. Ellas se muestran en las figuras 5.15(a), para la variable de entrada X, y 5.15(b), para la variable de entrada Y, siendo éstas las dos entradas del sistema.

Los parámetros de las campanas dibujadas en la figura 5.15 (a) y (b) son los siguientes:

Variable X:

Campana 1: $a = 0.375$, $b = 3.567$, $c = 2.117$

Campana 2: $a = 0.450$, $b = 4.450$, $c = 3.385$

Campana 3: $a = 0.630$, $b = 2.301$, $c = 5.0$

Variable Y:

Campana 1: $a = 0.295$, $b = 2.940$, $c = 2.348$

Campana 2: $a = 0.261$, $b = 1.631$, $c = 3.251$

Campana 3: $a = 0.697$, $b = 2.523$, $c = 5.0$

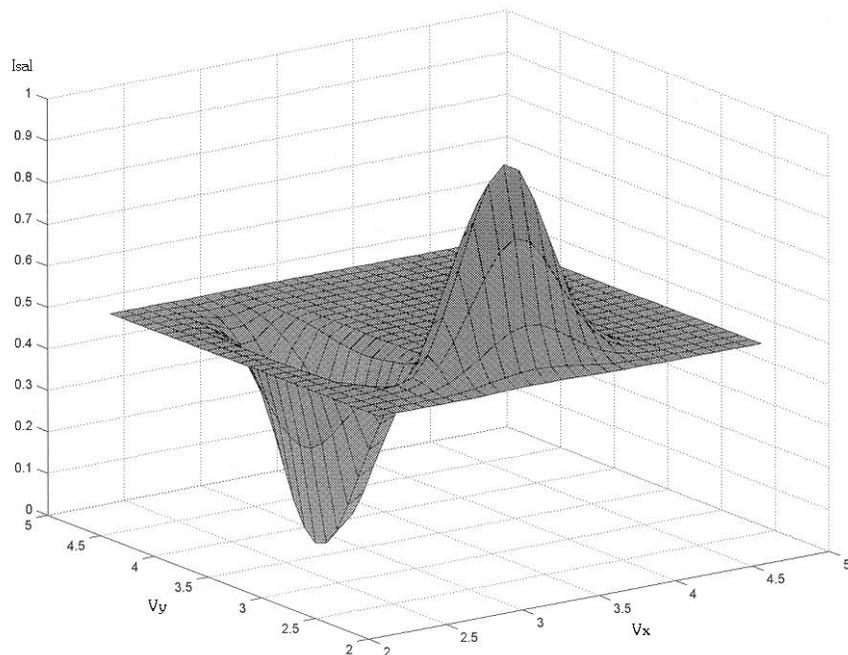
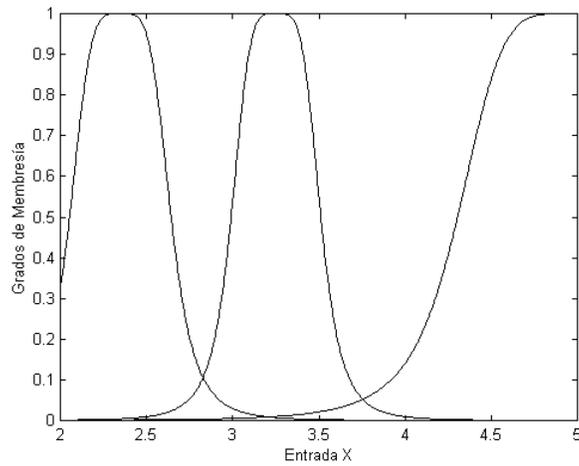
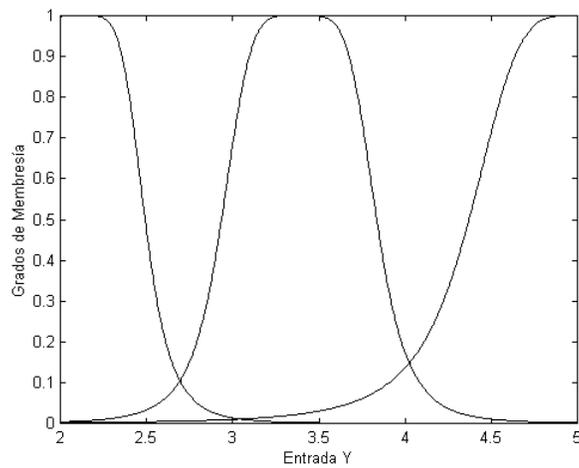


Fig. 5.14 Superficie obtenida mediante el criterio de Máxima Verosimilitud aplicado al conjunto de datos de entrenamiento.



(a)



(b)

Fig. 5.15 Funciones de Membresía obtenidas después de 50 épocas de entrenamiento.

La fig. 5.16 muestra la superficie de control obtenida en Matlab con el sistema, una vez realizado su entrenamiento. Tanto en este caso, como en el de la figura 5.14, los valores de entrada fueron escalados para ajustarse a los niveles aceptados por el circuito correspondiente a las capas 1 y 2; el intervalo de entrada es de 2 a 5 volts para cada variable. Como se observa, la aproximación obtenida parece no ser muy buena aunque, debido a la limitación en el número de conjuntos difusos que se tiene para cada variable (3 conjuntos por variable), se puede esperar una mejor aproximación si se aumenta dicho número de conjuntos. Sin embargo, observando la distribución de los datos de la figura 5.13(b), se puede anticipar que el error presente en la superficie generada por el sistema, no tendrá gran impacto en los resultados de la clasificación.

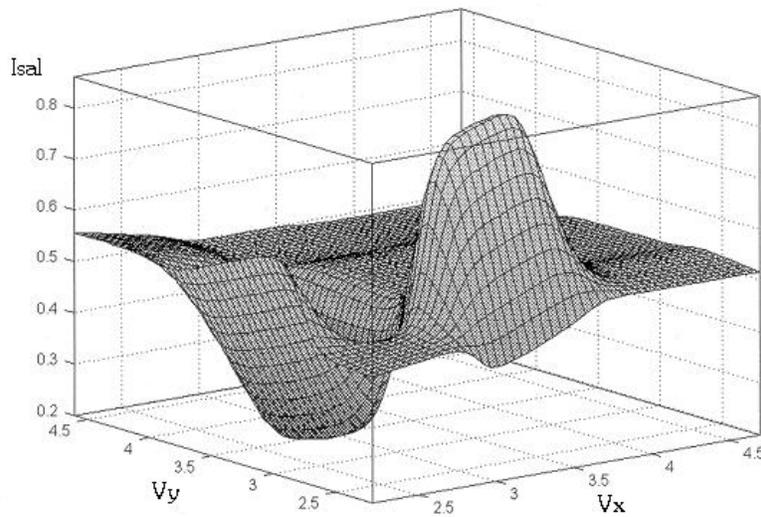


Fig. 5.16 Superficie obtenida con ANFIS para aproximar a la generada mediante el criterio de máxima verosimilitud.

En el capítulo siguiente se presentarán los resultados del sistema realizado con los circuitos integrados fabricados, ajustado en base a los parámetros obtenidos en este capítulo.

5.4. RESUMEN

En el presente capítulo se propuso una aplicación para el sistema ANFIS desarrollado. En esta aplicación se planteó emplear ANFIS para la clasificación de arritmias cardíacas. Debido a la amplia gama de arritmias existentes, el enfoque es hacia el tipo específico que se denomina PVC, el cual en ocasiones puede desencadenar situaciones de peligro para el paciente.

Se expusieron los principios básicos de electrocardiografía, para explicar cómo se obtiene un electrocardiograma, las derivaciones típicas de interés para los médicos, así como las ondas, complejos, segmentos e intervalos que componen a la señal ECG. Se explican las características de una señal ECG normal, así como de diferentes tipos de arritmia que se pueden presentar y, puesto que resultan de interés para este trabajo las arritmias ventriculares, se da mayor información sobre este tema.

Finalmente, se describió el procedimiento seguido para obtener un conjunto de datos que permitan entrenar al sistema ANFIS, a partir de rasgos característicos extraídos de señales ECG de la base de datos MIT-BIH. Los datos se obtienen empleando una técnica conocida como predicción lineal, con la que se obtienen dos coeficientes para cada complejo QRS de la señal. Se empleó un método de agrupamiento para generar una superficie que defina la pertenencia de los datos a una condición normal, PVC o no clasificada y, con esta superficie y el conjunto de datos, se hizo el entrenamiento de ANFIS en Matlab para generar los parámetros de ajuste de las funciones de membresía que se utilizarán en los circuitos con los que se construirá el sistema.

CAPITULO 6

RESULTADOS EXPERIMENTALES

“... Y como el geómetra que afanado en medir el círculo, no halla en su pensamiento el principio que necesita, tal estaba yo con aquella nueva representación: Quería ver cómo se adaptaba al círculo la imagen y cómo se identificaban sus naturalezas; pero no hubieran podido mis alas encumbrarse tanto, a no haber iluminado mi mente un resplandor que dejó satisfecho mi deseo.

Aquí perdí el sublime vigor de mi fantasía; mas ya daba impulso a mi anhelo y mi voluntad, como a una rueda que gira por igual, el Amor que mueve el Sol y las demás estrellas.”

(Dante Alighieri, *La Divina Comedia*)

6.1. INTRODUCCION

Las mediciones y pruebas de los circuitos diseñados y fabricados, se realizaron considerando básicamente tres etapas: Celdas individuales, sistema ANFIS completo y aplicación propuesta.

Para llevar a cabo esta labor, fue necesario elaborar algunos circuitos auxiliares que permiten alimentar, aplicar estímulos y generar las diferentes polarizaciones que requieren los circuitos integrados. La tarea principal consiste en diseñar los circuitos que permitan reproducir los voltajes, corrientes y señales generadas en las simulaciones realizadas con PSpice.

Las pruebas de celdas individuales se realizaron en una primera tarjeta de circuito impreso diseñada con OrCAD-PCB. Estas incluyeron la comprobación de la forma de la función de membresía, verificación del control de posición y ancho de la campana, y ajuste de las pendientes de las mismas.

Las mediciones del sistema completo, así como los resultados de la aplicación propuesta, se realizaron en otra tarjeta también diseñada en OrCAD. Esta tarjeta incluyó el circuito integrado desarrollado en este trabajo (capas 1 y 2 de ANFIS) y otro circuito integrado que forma parte de otro trabajo de tesis (capas 3, 4 y 5), así como los circuitos para generar los polinomios de primer orden que se utilizan en la etapa defusificadora, como se describió en el capítulo 2. En este caso, se verificó que las zonas del espacio difuso, correspondientes a las nueve reglas propuestas, funcionaran adecuadamente; se revisó nuevamente el control sobre los parámetros de las funciones de membresía, cuyo ajuste se realizó empleando los registros de corrimiento y capacitores, descritos en el capítulo 3; se comprobó la correspondencia entre los resultados de ANFIS obtenidos en Matlab y los de la tarjeta; y, finalmente, se hicieron las mediciones del sistema para la clasificación de arritmias de señales ECG, basadas en el conjunto de parámetros obtenidos a partir de rasgos característicos de complejos QRS de señales de la base de datos MIT-BIH.

6.2. PRUEBAS Y MEDICIONES DE CELDAS INDIVIDUALES

Con el fin de realizar pruebas a los circuitos diseñados y simulados en el capítulo 3, inicialmente se mandó a fabricar un circuito integrado que contenía un circuito de competencia y tres funciones de membresía, las cuales se construyeron con la configuración mostrada en la figura 3.3, de tal forma que se tuvieran pendientes ajustables.

En este primer circuito se midieron los límites de operación de las celdas, tomando en consideración los valores obtenidos previamente mediante simulación. Básicamente se comprobó el ajuste, mediante voltajes, de las posiciones de las funciones de membresía, lo cual está directamente vinculado con el control de su anchura. Se verificaron los ajustes que se pueden lograr en las pendientes de las campanas obtenidas, lo cual también se hace mediante voltajes de control. Se hicieron pruebas para comprobar el funcionamiento correcto del circuito de competencia, con el cual se realiza la función MIN, y obtener su respuesta dinámica cuando se le aplican los barridos de dos o más funciones de membresía.

Finalmente, se determinaron los retardos de propagación de las celdas en conjunto (funciones de membresía y función MIN), para conocer la velocidad de procesamiento que puede esperarse del circuito cuando esté en operación en el sistema completo.

A continuación se detallan algunos de los resultados más representativos que se obtuvieron de las mediciones hechas con este primer circuito integrado fabricado.

Con base en las mediciones que se realizaron a las celdas, se observa que el ancho mínimo que puede tener una campana, sin que pierda altura, es de 0.5V; este valor se incrementa si las pendientes de las campanas son muy suaves.

De acuerdo a la simulación con PSpice, los valores extremos de los voltajes E_1 y E_2 (descritos en la figura 3.2), para los cuales una campana no muestra deformación, son 2.2V y 4.8V; de acuerdo con las mediciones, el intervalo puede ser aún mayor, aumentando en el límite inferior, de tal forma que se tiene buena respuesta desde 2.0V. Sin embargo, si se ocupan campanas con pendientes más suaves, se requiere de un intervalo más reducido de los valores de E_1 y E_2 , el cual va de 2.1V hasta 4.2V. Esta reducción no es necesaria si las funciones de membresía utilizadas en los extremos del espacio difuso son abiertas por la izquierda y por la derecha, respectivamente. Los valores obtenidos son suficientes para construir las capas 1 y 2 de una arquitectura ANFIS como la que se mostró en la figura 2.3(a).

La figura 6.1 muestra la variación que se obtiene en las pendientes de las funciones de membresía al ajustar los voltajes correspondientes. Las pendientes obtenidas varían desde 0.96 hasta 5.68, que corresponde, respectivamente, a 43.8° y 80.0° , para variaciones de voltaje de 0.5V hasta 1V en los valores de las baterías descritas en el capítulo 3.

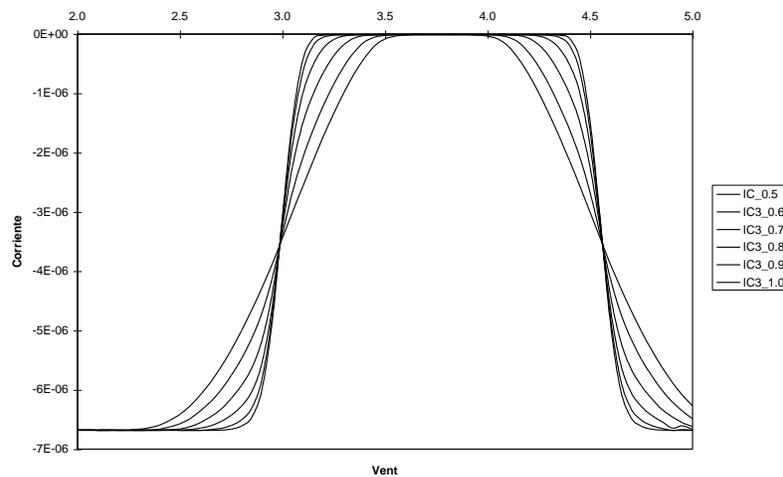


Fig. 6.1 Ajuste de las pendientes en las funciones de membresía.

La amplitud medida en las campanas, es mayor que la obtenida en simulación debido al tamaño de los transistores que se ocuparon como cargas (el circuito comercial CD4007); sin embargo, esto no representa un mayor problema, puesto que fácilmente se puede cambiar el valor de la corriente que se tomará como 1 lógico, sin modificar los resultados.

En las figuras 6.2 y 6.3 se presentan resultados obtenidos usando la función MIN para funciones de membresía con diferentes parámetros. En el primer caso (figura 6.2) se tienen campanas con pendientes grandes y, en el segundo caso (figura 6.3), las pendientes fueron modificadas, para obtener la operación MIN entre dos campanas con pendientes suaves.

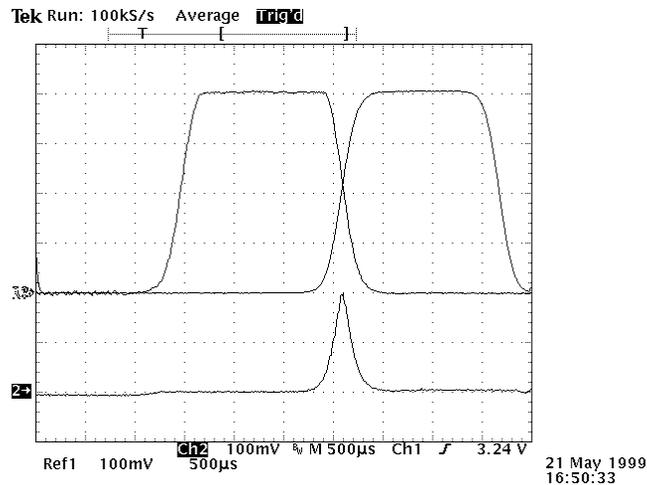


Fig. 6.2 Corriente correspondiente a la operación MIN entre dos campanas de pendientes grandes.

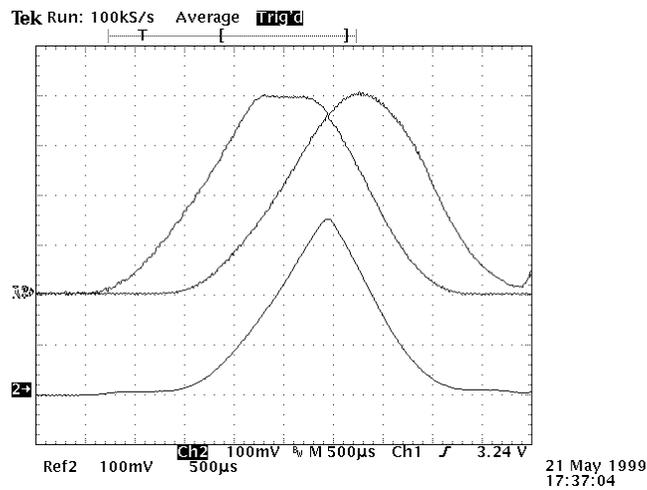


Fig. 6.3 Operación MIN entre dos funciones de membresía con pendientes suaves.

Una característica importante que permite cuantificar el desempeño de un sistema difuso, es el número de inferencias que realiza por unidad de tiempo (FIPS). Una máquina de inferencias opera con una serie de declaraciones conocida como *evaluación de reglas*, la cual conecta antecedentes con consecuentes, premisas con conclusiones o condiciones con acciones. En el circuito, el valor correspondiente a la evaluación de reglas se obtiene a la salida de la celda que realiza la operación MIN. El tiempo en que se lleva a cabo este proceso, es el tiempo que tarda el circuito en generar una salida estable, contado desde el momento en que se presenta un cambio en las variables de entrada.

Por ello, una de las pruebas que se le hicieron al circuito integrado, fue medir su retardo de propagación para conocer la máxima frecuencia de operación del circuito. La figura 6.4 muestra la medición de dicho retardo. Si se toma el peor caso de los mostrados en la figura (708ns), se obtiene una frecuencia máxima de operación correspondiente a 1.41MHz; en otras palabras, el circuito realizará 1.41MFIPS. El consumo total de potencia es de 4.38mW, con un consumo máximo de corriente de 876µA.

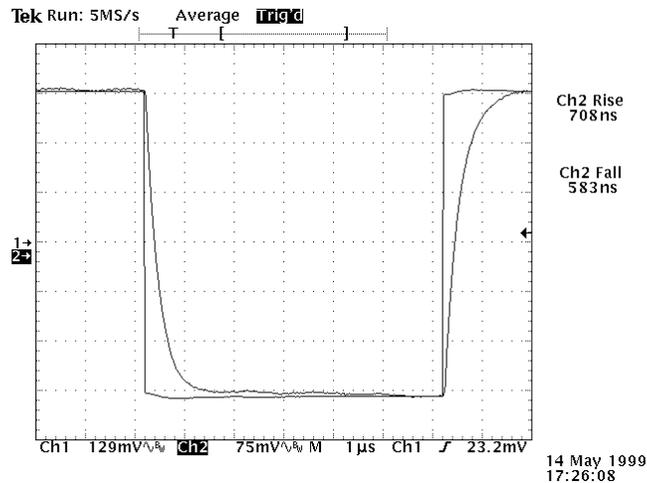


Fig. 6.4 Retardo de propagación medido para las capas 1 y 2 del sistema ANFIS.

Se puede esperar que las celdas tengan un mejor desempeño al integrar todo el sistema ANFIS en un solo circuito integrado, pues al realizar las mediciones en estas condiciones, se tienen algunos efectos indeseables debidos principalmente a las capacitancias de las terminales internas (*pads*) de entrada y salida del circuito integrado, así como la capacitancia del instrumento de medición, en este caso las puntas del osciloscopio digital utilizado.

6.3. PRUEBAS Y MEDICIONES DEL SISTEMA ANFIS COMPLETO

Como ya se ha mencionado, este trabajo forma parte de un proyecto para desarrollar una arquitectura ANFIS funcional. Con este propósito, se construyó una tarjeta de desarrollo en un circuito impreso, en la cual se incluyó la circuitería necesaria para dar soporte a todas las etapas de la arquitectura mencionada.

La tarjeta se diseñó para cumplir con dos tareas básicas: La primera, hacer pruebas de forma individual en cada uno de los dos circuitos integrados que se diseñaron (uno de los cuales, el correspondiente a las capas 1 y 2 de ANFIS, forma parte de este trabajo de tesis); la segunda, desarrollar el sistema de inferencias difusas, propiamente dicho, para hacer pruebas y mediciones a nivel funcional. Los principales componentes de esta tarjeta son fuentes de corriente (construidas por medio de amplificadores operacionales), referencias de voltaje y algunos circuitos digitales para generar señales de control.

Las señales de alimentación para la tarjeta son: +5V para el circuito integrado, y $\pm 9V$ para el grupo de amplificadores operacionales; el circuito integrado correspondiente a las capas 3, 4 y 5, utiliza una alimentación de $\pm 5V$.

Las dos variables de entrada a la tarjeta se dimensionan en voltajes. Para generar la superficie de control en el circuito, en una de las entradas se emplea un barrido de voltaje, con un generador y, en la otra, se aplica un voltaje que se varía con ajustes manuales, generándose así un conjunto de curvas, las cuales posteriormente se juntan en una sola gráfica tridimensional.

Una vez que el sistema ha sido entrenado en Matlab, los parámetros así generados se introducen a los circuitos en forma de ajustes a las fuentes de corriente y las referencias de voltaje.

La primera prueba consistió en verificar que funcionaran correctamente las nueve regiones del espacio difuso, correspondientes a las nueve reglas difusas del sistema. Para ello, se creó una superficie que mostrara una elevación en cada región (figura 6.5).

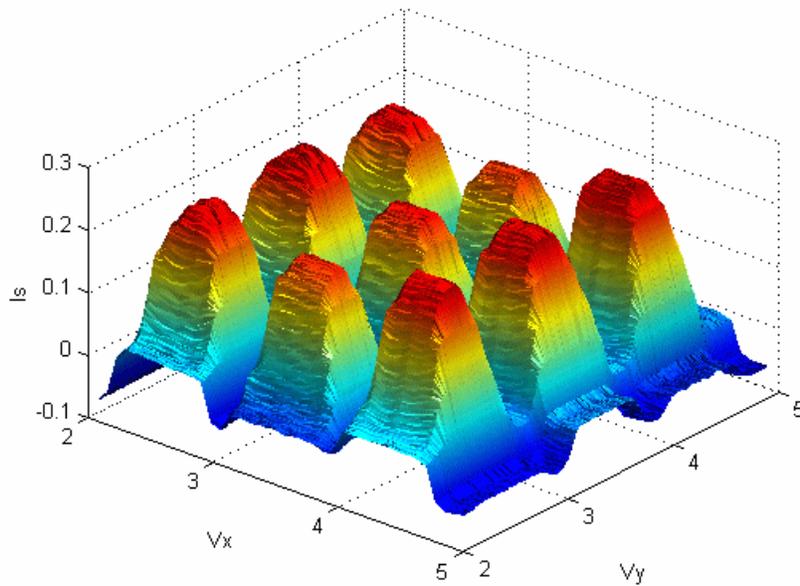
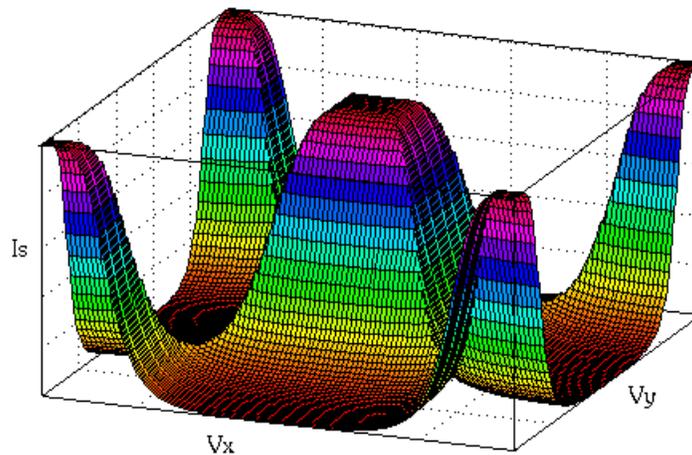


Fig. 6.5 Superficie de prueba de las regiones correspondientes a las 9 reglas difusas del sistema.

En este caso no es necesario realizar ningún entrenamiento, pues las posiciones de las funciones de membresía simplemente se distribuyen uniformemente en el espacio difuso; se utilizan *impulsos difusos* en las funciones de salida, pues el objetivo únicamente es comprobar el correcto funcionamiento del circuito integrado.

Las figuras 6.6(a) y 6.6(b) muestran dos superficies generadas por medio del sistema ANFIS. La primera se obtuvo empleando un sistema ANFIS construido en software, con ayuda de Matlab y, la segunda, utilizando la tarjeta con los circuitos fabricados. El objetivo de generar estas dos figuras es realizar una comparación desde el punto de vista de las dos metodologías (software y hardware), y poder comprobar qué tan buena reproducción se tiene con el sistema construido en la tarjeta de desarrollo.



(a)

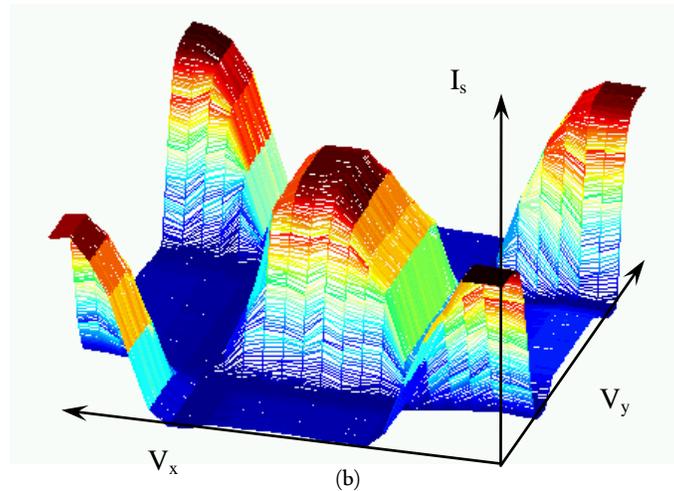


Fig. 6.6 Superficies de control generadas por el sistema ANFIS: (a) Generada en MatLab; (b) Generada con la tarjeta de desarrollo.

El error máximo que se presentó en esta prueba fue cercano a 9.5%. En este punto cabe aclarar que el error no es uniforme, y es menor para los valores máximos de I_s , donde se obtiene un error con un valor menor al 1%. Para un sistema difuso, estas magnitudes de error se consideran aceptables.

Otras pruebas consistieron en reproducir diferentes superficies de control, con las cuales se determinaron los límites de operación de los circuitos para producir los polinomios del sistema. También se verificó que los valores de voltajes y corrientes en la tarjeta, correspondientes a los diferentes parámetros entregados por Matlab para el correcto ajuste del sistema, concordaran con los encontrados en las pruebas de celdas individuales.

La figura 6.7 es una fotografía de la tarjeta de desarrollo y del equipo de alimentación y medición utilizados para las pruebas del sistema. Se observa un osciloscopio digital, dos fuentes de alimentación y dos generadores de señales. La figura 6.8 muestra la tarjeta de desarrollo con mayor detalle, donde el circuito integrado que se observa en la parte superior derecha, es el correspondiente a las capas 1 y 2 desarrolladas en este trabajo.

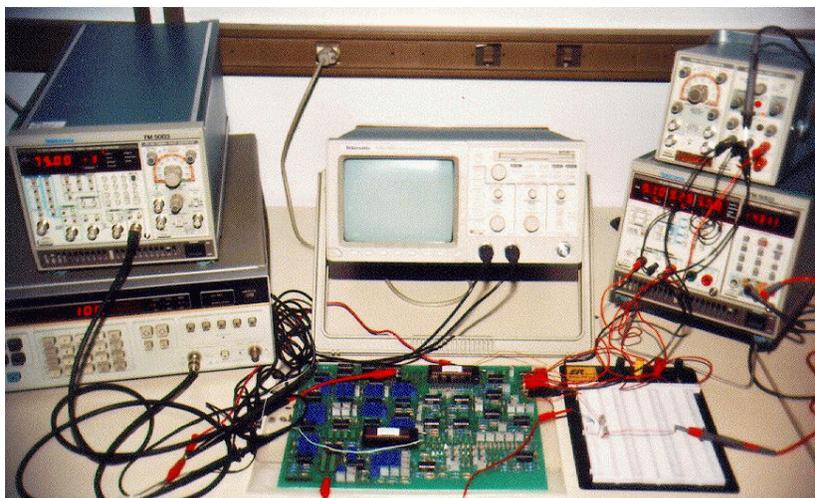


Fig. 6.7 Sistema de pruebas completo que muestra las fuentes de alimentación, generadores de señales, osciloscopio digital y tarjeta de desarrollo del sistema ANFIS.

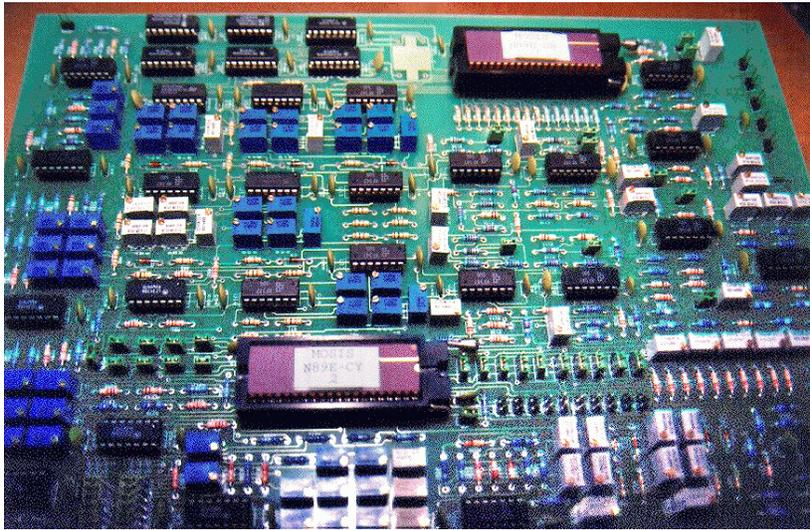


Fig. 6.8 Detalle de la tarjeta de desarrollo del sistema ANFIS construido.

El tamaño de la tarjeta resulta algo grande debido a que fue diseñada para poder realizar ajustes individuales de todos los parámetros (voltajes y corrientes) utilizados por los circuitos integrados, lo cual se hizo con circuitos discretos comerciales, así como incluir circuitería para realizar pruebas individuales de cada circuito integrado diseñado. Sin embargo, para un sistema funcional que excluya los circuitos para hacer pruebas, y buscando una solución más óptima en vez de emplear circuitos comerciales, se puede esperar una reducción considerable en el tamaño de la tarjeta.

En la sección siguiente, se presenta la aplicación del sistema construido para el caso de clasificación de señales ECG, lo cual, visto de una forma general, también entra en la parte correspondiente a pruebas y mediciones del sistema.

6.4. MEDICIONES Y RESULTADOS DE LA APLICACION PROPUESTA

Como se describió en el capítulo anterior, la aplicación propuesta para valorar el desempeño del sistema ANFIS, fue la clasificación del tipo de arritmia cardíaca denominada PVC, para señales ECG. En aquel capítulo se obtuvo el conjunto de datos para realizar el entrenamiento de ANFIS a partir de una superficie de control, la cual se obtuvo mediante el método de máxima verosimilitud aplicado a la gráfica de un coeficiente de predicción lineal (a_i) y el valor cuadrático medio de complejos QRS de señales de la base MIT-BIH.

El entrenamiento que se realizó en Matlab, produjo los valores de los parámetros de ajuste, que fueron utilizados en el sistema de desarrollo, para generar la superficie de control propuesta.

En la figura 5.16 se mostró la superficie obtenida con ANFIS, construido en software y empleando Matlab, para el ejemplo de aplicación. La figura 6.9 muestra la superficie de control que se obtuvo con el sistema desarrollado. Para generarla, se hicieron los ajustes de los parámetros de los circuitos (voltajes y corrientes), con base en los valores arrojados por el entrenamiento en Matlab y, posteriormente, se procedió a capturar la información de la

superficie propiamente dicha. Para ello, a la variable de entrada V_x se le aplicó una rampa de voltaje con un generador de señales, cubriendo el intervalo de 2 a 5V; a la variable V_y se le aplicó una referencia de voltaje, ajustable manualmente, para hacer capturas de los barridos de V_x en diferentes posiciones de V_y , también dentro del intervalo de 2 a 5V; estas capturas se almacenaron en archivos de datos que el osciloscopio utilizado permite crear y que, posteriormente, pueden ser leídas en Matlab para generar una gráfica tridimensional de la superficie. La superficie de la figura 6.9 se obtuvo haciendo 55 capturas en el eje V_y , las cuales tienen una distribución irregular, de tal forma que se tenga mejor resolución en las partes donde la superficie tenga una mayor variación; en el eje V_x se tiene una resolución de 1000 puntos distribuidos uniformemente para todo el intervalo medido, correspondiente a la resolución propia del osciloscopio para almacenar una señal en un archivo.

La salida total de la tarjeta de aplicación es una corriente que se obtiene mediante la suma de todas las salidas del circuito integrado que contiene las capas 3, 4 y 5 de la arquitectura. Puesto que en el osciloscopio sólo se pueden medir voltajes, la corriente de salida se pasa a través de una carga de valor conocido a fin de generar una caída de potencial que pueda medirse y, posteriormente, obtener el valor de corriente equivalente.

La generación de la superficie de control de la figura 6.9, es el paso inicial para poder evaluar la clasificación que realiza el sistema para la aplicación propuesta. El siguiente paso consiste en aplicarle al circuito, ya ajustado con los valores de entrenamiento, diferentes valores de voltajes de entrada, que correspondan a los valores de los datos obtenidos para los siete pacientes de la base MIT-BIH que se consideraron en el capítulo anterior (mostrados en la figura 5.13(b) como el valor cuadrático medio contra el coeficiente LPC a_1).

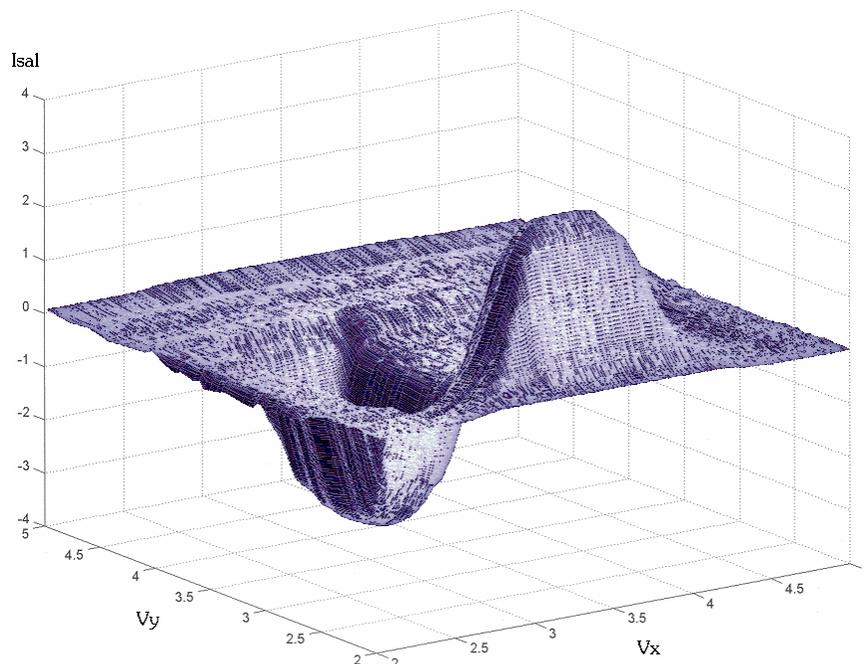


Fig. 6.9 Superficie de control obtenida con la tarjeta de desarrollo, para el ejemplo de clasificación de arritmias cardíacas.

Para esta evaluación, se plantean tres escenarios: 1) Aplicar valores de entrada que correspondan a datos con los que se hizo el entrenamiento del sistema en Matlab; 2) Aplicar valores de entrada correspondientes a datos que se utilizaron como comprobación en Matlab,

una vez entrenado el sistema; 3) Aplicar otros datos, tomados de los siete pacientes, no considerados en los dos casos anteriores.

Las mediciones de los tres escenarios se resumen en la tabla 6.1, en donde se muestran los resultados obtenidos para cada tipo de latido (normal o PVC). La primera columna muestra los tres casos considerados en la evaluación y el número de latidos que se utilizó en cada uno de ellos. La segunda columna indica la cantidad de latidos normales detectados correctamente por el sistema, y latidos normales que se clasificaron como PVC (falsos positivos). La tercera columna representa la cantidad de latidos PVC clasificados correctamente y latidos PVC identificados como normales (falsos negativos). Finalmente, la cuarta columna representa latidos a los que no se les pudo asignar una pertenencia bien definida a alguno de los dos grupos anteriores.

Casos de prueba	Normal ND / NV	PVC VD / VN	No Clasificado NNC / VNC
Escenario 1 (N=500, V=250)	485 / 7	241 / 4	8 / 5
Escenario 2 (N=500, V=250)	484 / 6	240 / 6	10 / 4
Escenario 3 (N=500, V=300)	480 / 8	289 / 6	12 / 5

Tabla 6.1 Resultados de la clasificación del sistema de desarrollo. (N≡Latidos normales, V≡Latidos PVC, ND≡Normales detectados, VD≡PVC detectados, NV≡Falsos positivos, VN≡Falsos negativos, NNC≡Normales no clasificados, VNC≡PVC no clasificados).

Para determinar la pertenencia de los latidos a alguno de los grupos considerados, se establecieron dos umbrales de asignación, esto es: Dadas las características de la superficie de control obtenida (figura 6.9), se fijó un valor positivo de salida, arriba del cual, un latido se considera como Normal; asimismo, se fijó otro umbral negativo de salida, por debajo del cual, un latido se identifica como PVC; finalmente, los latidos situados entre ambos umbrales, conforman los No Clasificados.

Si se considera la superficie de control obtenida inicialmente mediante el criterio de máxima verosimilitud (figura 5.14), no es necesario establecer estos umbrales, ya que un valor positivo de salida representa un latido normal; un valor negativo, un latido PVC; y, un valor igual a cero es un latido no clasificado. Sin embargo, las pequeñas fluctuaciones que tiene la superficie medida en las zonas donde debería tener un valor igual cero, hacen necesario el empleo de tales umbrales. Los resultados de la tabla 6.1 se obtuvieron considerando un umbral positivo aproximadamente equivalente al 4% del valor máximo positivo de salida, mientras que el umbral negativo se definió como 8.5% del valor mínimo de salida.

Con el objetivo de valorar los resultados de la tabla 6.1, el siguiente paso es realizar el cálculo del error global del sistema, así como algunas otras características que cuantifican su desempeño para este problema en particular. Las expresiones empleadas para ello son las siguientes, considerando las mismas literales definidas en la tabla 6.1:

Error total (%):

$$\%ET = \frac{NV + VN + NNC + VNC}{(N + V)} * 100 \quad (6.1)$$

Sensitividad (%):

$$\%S = \frac{V - VN - VNC}{V} * 100 \quad (6.2)$$

Detección de latidos normales (%):

$$\%DN = \frac{N - NV - VN - NNC}{N} * 100 \quad (6.3)$$

Detección de latidos PVC (%):

$$\%DV = \frac{V - VN - NV - VNC}{V} * 100 \quad (6.4)$$

Con los resultados de la tabla 6.1 y con ayuda de las expresiones anteriores, se elabora la tabla 6.2 siguiente, en la cual se presentan las evaluaciones individuales para cada uno de los tres escenarios mencionados, así como la evaluación considerando los tres escenarios conjuntamente (resultado global).

Casos de prueba	%ET	%S	%DN	%DV
Escenario 1	3.2	96.4	96.2	93.6
Escenario 2	3.47	96.0	95.6	93.6
Escenario 3	3.88	96.33	94.8	93.67
Evaluación global	3.52	96.25	95.53	93.63

Tabla 6.2 Desempeño del sistema de desarrollo para clasificación de señales ECG.

Si los resultados de la tabla 6.2 se comparan con los entregados por otras metodologías desarrolladas en programación (software) (por ejemplo [25]), tal vez no resulten muy satisfactorios; sin embargo, se debe considerar que la construcción del sistema desarrollado ha sido con circuitos (hardware), y el sistema ANFIS construido se ha limitado en cuanto al número de funciones de membresía y reglas difusas, debido al área de silicio disponible para la fabricación de los circuitos integrados diseñados y la tecnología de $2\mu\text{m}$ utilizada. En este sentido, se puede esperar un mejor desempeño si, posteriormente, al disponer de una tecnología de tamaño característico menor, se realice una arquitectura con un mayor número de funciones de membresía por variable de entrada, y con un mayor número de reglas difusas.

6.5. RESUMEN

Los temas tratados en este capítulo resaltan las mediciones y pruebas eléctricas llevadas a cabo para evaluar el funcionamiento de los circuitos desarrollados a lo largo de este trabajo.

Se mencionaron las pruebas hechas a las celdas básicas para verificar su correcto funcionamiento, de acuerdo con las simulaciones hechas anteriormente, así como para determinar sus límites de operación.

Posteriormente, se describieron los resultados obtenidos de las pruebas que se hicieron a nivel funcional de las capas construidas de la arquitectura ANFIS; con el sistema completo, construido en una tarjeta de desarrollo, se verificó el correcto funcionamiento de las zonas correspondientes a las nueve reglas difusas del sistema, y la aproximación lograda con respecto a una superficie obtenida en Matlab.

Finalmente, con ayuda de la tarjeta de desarrollo, se efectuaron diferentes mediciones para evaluar el problema de aplicación propuesto, correspondiente a la identificación del tipo de arritmia denominado PVC, basado en una superficie de control generada a partir de datos obtenidos de diferentes pacientes de la base de datos MIT-BIH.

CAPITULO 7

CONCLUSIONES

“564. Luego Corazón del Cielo, a estos cuatro primeros hombres, les echó vaho en los ojos y se los empañó, así como un espejo se empaña cuando se sopla.

Y de esa manera ya sólo pudieron ver lo que estaba cerca, y comprender sólo aquello que era evidente y claro.

565. Así fue perdida la sabiduría y entendimiento de los cuatro primeros hombres.”

(*Popol Vuh*)

“Canek dijo:

- En buena hora, niño Guy, nos acompaña el pájaro *Pujuy*. Hay que seguir adelante, vencer el cansancio, el miedo y el deseo. La fatiga disfraza sus intenciones. La fatiga es sueño, curiosidad y desgano en los caminantes. Abre bien los ojos, hijo, y sigue al pájaro *Pujuy*. El no se equivoca. Su destino es como el nuestro: Caminar para que otros no se pierdan.”

(Emilio Abreu Gómez, *Canek*)

Conclusiones.

Tomando como base el desarrollo teórico y práctico realizado a lo largo del presente trabajo de tesis, y considerando los resultados obtenidos, tanto de simulaciones a nivel funcional como a nivel eléctrico de las diferentes etapas que lo integran, así como de las pruebas y mediciones hechas al sistema de desarrollo construido, se llega a los siguientes puntos sobresalientes que se dan como las conclusiones más importantes de la tesis:

- ✧ Los sistemas neurodifusos emplean las cualidades propias de los sistemas difusos para incorporar el conocimiento de expertos en forma de reglas del tipo *si...entonces*, junto con la propiedad de las redes neuronales, que los hace ser adaptables y, de esta forma, pueden ajustar los valores de los parámetros de sus premisas o de sus consecuentes basados en un conjunto de pares de datos de entrada/salida. Estas características los hacen muy adecuados para tratar con el problema que se plantea en este trabajo, correspondiente a la clasificación de señales.
- ✧ Entre los diferentes sistemas neurodifusos existentes, se consideró a la arquitectura ANFIS como la parte fundamental de la metodología en la cual se basaría el desarrollo de este trabajo, debido, principalmente, al algoritmo de aprendizaje híbrido que emplea para el ajuste de sus parámetros adaptables. Este algoritmo es muy eficiente para minimizar la medida del error de salida del sistema cuando se ajustan los parámetros de sus funciones de membresía.
- ✧ Se hicieron los diseños de las celdas básicas, en tecnología CMOS, que permiten aproximar las funciones requeridas para construir las capas 1 y 2 de ANFIS. La ventaja de construir el sistema en circuitos CMOS, es que permite hacer efectivamente procesamiento en paralelo, que es una característica que hace poderosos a los sistemas neurodifusos. Este tipo de diseño es novedoso y aún es tema abierto de investigación.
- ✧ Los resultados obtenidos con las celdas básicas, se tomaron como punto de partida para construir las capas 1 y 2 de una arquitectura ANFIS de dos entradas, 3 funciones de membresía por entrada, 9 reglas difusas y una salida. Así, se construyó un sistema de desarrollo para poder evaluar el funcionamiento de ANFIS operando con circuitos (hardware). El sistema construido es un sistema difuso de tipo Sugeno de 1er orden; en este sentido, hay que resaltar que otros sistemas reportados en hardware [6], [7] son de orden cero, esto es, no trabajan con polinomios de salida, sino únicamente con impulsos difusos.
- ✧ En el desarrollo de la tesis, la realización de la arquitectura ANFIS requirió de la utilización de dos circuitos integrados, debido, por una parte, a las dimensiones de la tecnología utilizada (CMOS de $2\mu\text{m}$), y por otra parte, a la necesidad de contar con un determinado número de terminales para la caracterización de múltiples celda que conforman el circuito. Con los resultados obtenidos en este trabajo, es factible que, utilizando una tecnología CMOS de $1.2\mu\text{m}$, la cual permita un mayor número de dispositivos, se puedan incluir en un sólo circuito integrado las cinco capas de la arquitectura ANFIS así como la instrumentación correspondiente al aprendizaje en línea. Asimismo, es factible pensar en introducir una mayor cantidad de funciones de membresía y reglas difusas, aunque la migración tecnológica

también incluya la utilización de un encapsulado que contenga un mayor número de terminales.

- ✧ Como ejemplo de aplicación, se hizo un estudio de señales electrocardiográficas (ECG) y se utilizó el sistema construido para la clasificación de señales que presentan contracción ventricular prematura (PVC), las cuales se tomaron de la base de datos de arritmias MIT-BIH. La metodología utilizada para esta tarea arrojó buenos resultados, obteniéndose un error considerado pequeño (~3.5%) y un índice de detección de arritmias mayor al 93%. Desarrollando un sistema ANFIS con las características mencionadas en el punto anterior, se esperaría una detección más precisa del tipo de cardiopatía seleccionada. Aquí cabe mencionar, que la utilización de ANFIS para clasificación de arritmias en señales ECG, se ha propuesto únicamente a nivel de programación [26] (software), y no se ha reportado el empleo de un sistema ANFIS en hardware para tal aplicación.
- ✧ La aplicación que se propuso en este trabajo es en un problema del área biomédica, pero la generalidad de los circuitos construidos permite que fácilmente puedan ocuparse en otros campos de aplicación diferentes.
- ✧ Una posible comparación entre el trabajo aquí desarrollado y otros realizados con estructuras similares es difícil de llevar a cabo, ya que el proyecto del que esta tesis forma parte contempla, en una primera etapa, la realización por separado de las distintas capas que conforman ANFIS. Además, hay que recordar que en este trabajo se introdujeron otros conceptos distintos a aquellos realizados por otros investigadores con los que podría equipararse. En efecto, nuestro circuito presenta algunas características como la ya mencionada de que el sistema que se ha implantado es un sistema de tipo Sugeno de 1er orden, lo cual no es el caso de las redes publicadas a la fecha. Sin embargo, a manera de comparación de algunas características relevantes, en la tabla 7.1 se muestran los resultados obtenidos en este trabajo junto con los publicados por otros autores.

CMOS	Manaresi [27]	Guo [28]	Rodríguez [7]	Baturone [6]	Desarrollado
Complejidad	9 reglas, 2 entradas, 2 salidas	13 reglas, 3 entradas, 1 salida	16 reglas, 2 entradas, 1 salida	12 reglas, 2 entradas, 1 salida	9 reglas, 2 entradas, 1 salida
Tecnología	0.7 μ m CMOS	2.4 μ m CMOS	1 μ m CMOS	2.4 μ m CMOS	2 μ m
Consumo de potencia	44mW@5V	550mW@10V	8.6mW@5V	9.5mW@5V	4.38mW@5V
Retardo de propagación	570ns	160ns	471ns	450ns	708ns
Precisión	ND	ND	6.5%	2.5%	3.52% (aplicación propuesta)
Interfaces	Voltajes /	Voltajes /	Voltajes /	Voltajes /	Voltajes /

(entrada / salida)	Voltajes	Voltajes	Corrientes	Corrientes	Corrientes
Area de silicio	1.9mm ²	16.2mm ²	1.6mm ²	0.98mm ²	4.995mm ²
Modelo difuso	Mamdani	Mamdani	Sugeno de orden 0	Sugeno de orden 0	Sugeno de 1er orden

Tabla 7.1 Cuadro comparativo de distintos trabajos publicados y el desarrollado en esta tesis.

Comparando las características de este trabajo con las de Guo [28], quizá el mayor factor de desventaja sea en el retardo de propagación, aunque el consumo de potencia es evidentemente mejor en nuestro caso. En las otras aproximaciones [27] y [7], es difícil extrapolar las características a las de nuestro trabajo ya que, como se observa, ellos utilizan tecnologías de 0.7µm, 1µm, respectivamente y, además, hay que recordar que esta tesis corresponde únicamente a la integración de las dos primeras capas de la arquitectura propuesta. En el caso de [6], puesto que ellos utilizan funciones trapezoidales en sus conjuntos de entrada, se tiene una menor precisión para aproximar funciones. La ventaja clara que presenta nuestra aproximación es el hecho de desarrollar un sistema de tipo Sugeno de 1er orden, lo cual proporciona un grado de flexibilidad mucho mayor que el alcanzado por las otras redes.

Trabajo futuro.

Los resultados obtenidos en el presente trabajo, pueden servir como motivación para realizar trabajos posteriores que busquen elaborar un sistema ANFIS que incluya algunas características adicionales a las aquí desarrolladas. Entre ellas se pueden mencionar las siguientes:

- ❖ Hacer que el sistema sea adaptable en Hardware: En este momento, el entrenamiento de ANFIS se hace empleando el programa Matlab en una computadora personal y, posteriormente, los valores de los parámetros así obtenidos, se ajustan manualmente en forma de voltajes y corrientes; sin embargo, resulta de interés que el aprendizaje se pueda realizar en la propia tarjeta de desarrollo, para lo cual se tendrían que incluir los circuitos necesarios para realizar, en hardware, las operaciones del algoritmo de aprendizaje híbrido descrito en el texto. En este contexto, también se puede pensar que el sistema sea adaptable *en línea*.
- ❖ Cuando se tenga la posibilidad de elaborar los CI's con una tecnología más pequeña, se podrán incluir más funciones de membresía y más reglas difusas. Aquí, el trabajo consistirá en integrar todo el sistema de desarrollo que se construyó en este trabajo, en un sólo CI, que deberá incluir las 5 capas de ANFIS, los generadores de los polinomios requeridos por el sistema de tipo Sugeno de 1er orden y, tal vez, los circuitos propuestos en el punto previo.
- ❖ Las funciones de membresía que utiliza ANFIS (funciones tipo campana generalizada), se realizaron en el circuito mediante una configuración, que buscó ser sencilla, para aproximar la forma de ellas. Esto origina un error al pasar de los resultados obtenidos en Matlab hacia la tarjeta de desarrollo, puesto que la ecuación matemática de una campana no se puede aproximar exactamente con la configuración utilizada. Para reducir este problema se tienen dos propuestas, que pueden dar lugar a trabajos de investigación: 1) Desarrollar alguna otra

configuración con transistores MOS (que no debe llegar a ser muy compleja), la cual aproxime de una mejor manera la ecuación matemática de una campana; de esta forma se obtiene una mayor concordancia entre el sistema ANFIS desarrollado originalmente (en software) y el que se pueda construir en hardware, usando dichas configuraciones. 2) Modificar el algoritmo de entrenamiento que emplea ANFIS para que, en lugar de usar funciones de membresía del tipo campana generalizada, aplicara la Regla de Aprendizaje Híbrida para optimizar funciones de membresía basadas en las ecuaciones del par diferencial doble, descritas en el capítulo 3; esto es, se crearía un nuevo sistema ANFIS, cuyo entrenamiento arrojaría resultados aplicables en forma directa al ajuste de los parámetros de circuitos MOS.

Publicaciones en congresos.

A continuación se presenta la lista de trabajos a los que ha dado lugar la presente tesis, con los cuales se ha participado en congresos nacionales e internacionales.

- ❖ Oliverio Arellano Cárdenas, Herón Molina Lozano, Felipe Gómez Castañeda, José A. Moreno Cadenas y Luis M. Flores Nava.
 “Desarrollo de un Sistema de Clasificación de Arritmias en Tecnología CMOS Basado en ANFIS”, Cuarto Taller Iberoamericano de Micro-electrónica y sus Aplicaciones, IV Workshop Iberchip, pp. 209-219, Mar del Plata, Argentina, Marzo 11-13, 1998.
 - ❖ Oliverio Arellano-Cardenas, José A. Moreno-Cadenas, Felipe Gómez-Castañeda y Luis M. Flores-Nava.
 “Extracción de Grupos de Clases en Señales Electrocardiográficas”, Cuarta Conferencia de Ingeniería Eléctrica CIE'98, pp. 54-56, CINVESTAV-IPN, México, D.F., Sep. 9-11, 1998.
 - ❖ Oliverio Arellano Cardenas, Herón Molina Lozano, José A. Moreno Cadenas, Felipe Gómez Castañeda y Luis Martín Flores Nava.
 “Sistemas Analógico Neurodifuso CMOS Universal basado en ANFIS”, Quinta Conferencia de Ingeniería Eléctrica, CIE'99, pp. 283-292, CINVESTAV-IPN, México, D.F., Sep. 8-10, 1999.
 - ❖ O. Arellano, H. Molina, J.A. Moreno y F.Gómez.
 “Analog Test and Simulation of Functional CMOS Building Blocks for Fuzzy Systems”, 3rd Workshop on Simulation and Characterization Techniques in Semiconductors, pp 17-22, CINVESTAV-IPN, México, D.F., Sep. 6-7, 1999.
 - ❖ Molina H., Arellano O., Flores L.M., Moreno J.A. y Gómez F.
 “Prototipo de Sistema Neuro-Difuso ANFIS en CMOS” *, SOMI XIV Congreso de Instrumentación, pp 712-716, Tonantzintla, Puebla, México, Oct. 4-8, 1999.
Distinción: Mejor trabajo en Diseño e Ingeniería del área de Redes Neuronales y Sistemas Borrosos.
- * Versión en inglés sometida para publicación en la revista: Instrumentation and Development de la Sociedad Mexicana de Instrumentación.
- ❖ Oliverio Arellano-Cardenas, José A. Moreno-Cadenas, Felipe Gómez-Castañeda y Luis M. Flores-Nava.
 “Extracción de Grupos de Clases en Señales Electrocardiográficas”, Revista CIENTIFICA de ESIME-IPN, pp. 29-31, Vol. 11, 1999.
 - ❖ O. Arellano-Cárdenas, H. Molina-Lozano, J. Moreno Cadenas, F. Gómez-Castañeda, L. Flores-Nava.
 “CMOS Analog Neurofuzzy Prototype based on ANFIS”, IEEE International Symposium on Circuits and Systems, pp. 726-729, Geneva, Switzerland, May 28-31, 2000.

- ✧ Oliverio Arellano-Cárdenas, Herón Molina-Lozano, Luis M. Flores-Nava, José A. Moreno-Cadenas and Felipe Gómez-Castañeda.
“Classification of Analog Signals in Control Systems by a CMOS ANFIS Prototype”, Third International Symposium on Soft Computing for Industry. World Automation Congress, Maui, Hawaii, June 11-16, 2000. (CD-ROM).

APENDICES



“Nuestra tarea no es imaginar o suponer, sino descubrir lo que la Naturaleza hace...”
Francis Bacon

“Se tienen diez pilas de monedas, cada una con diez monedas del mismo valor. Toda una pila es de monedas falsas, pero no se sabe cuál es. Se conoce el peso de una moneda auténtica, y se sabe que una moneda falsa pesa un gramo más de lo debido. Se cuenta con una báscula y el problema consiste en identificar el grupo de las monedas falsas en una sola pesada. ¿Cómo se hace?”
(M. Gardner, *Acertijos Matemáticos*)

Apéndice A: Método descendiente abrupto.

La finalidad de este método es minimizar una función real E , definida en un espacio de entrada n -dimensional $\theta = [\theta_1, \theta_2, \dots, \theta_n]^T$. Para ello, la labor consiste en encontrar un punto mínimo $\theta = \theta^*$ (posiblemente local), que minimice a $E(\theta)$.

En general, la función objetivo E puede tener una forma no lineal con respecto a un parámetro ajustable θ . Debido a la complejidad de E , frecuentemente se recurre a un algoritmo iterativo para explorar eficientemente el espacio de entrada. En los métodos descendientes iterativos, el punto siguiente θ_{sig} está determinado por un paso hacia abajo del punto actual θ_{act} sobre un *vector de dirección* \mathbf{d} :

$$\theta_{\text{sig}} = \theta_{\text{act}} + \eta \mathbf{d} \quad (\text{A.1})$$

donde η es un *tamaño de paso* que regula a lo que se extiende en la dirección de \mathbf{d} . En la literatura neurodifusa, el término **razón de aprendizaje** se utiliza para definir el tamaño de paso η . Por conveniencia, eventualmente se empleará la siguiente fórmula:

$$\theta_{k+1} = \theta_k + \eta_k \mathbf{d}_k \quad (k=1, 2, 3, \dots) \quad (\text{A.2})$$

donde k denota el número actual de iteración y $\theta_{\text{act}}, \theta_{\text{sig}}$ representan dos elementos consecutivos en una secuencia generada de candidatos a solución $\{\theta_k\}$. Se pretende que θ_k converja a un mínimo (local) θ^* .

Los métodos descendientes iterativos calculan el k -ésimo paso $\eta_k \mathbf{d}_k$ mediante dos procedimientos: Primero se determina la *dirección* \mathbf{d} , y entonces se calcula el *tamaño de paso* η . El punto siguiente θ_{sig} debe satisfacer la siguiente desigualdad:

$$E(\theta_{\text{sig}}) = E(\theta_{\text{act}} + \eta \mathbf{d}) < E(\theta_{\text{act}}) \quad (\text{A.3})$$

La diferencia principal entre varios algoritmos descendientes, consiste en el primer procedimiento para determinar las direcciones sucesivas. Una vez que se ha tomado la decisión, todos los algoritmos se mueven hacia un punto mínimo (local) sobre la línea determinada por el punto actual θ_{act} y la dirección \mathbf{d} . Esto es, para el segundo procedimiento, el tamaño óptimo de paso se puede determinar mediante *minimización en línea*:

$$\eta^* = \arg \min_{\eta > 0} \phi(\eta) \quad (\text{A.4})$$

donde:

$$\phi(\eta) = E(\theta_{\text{act}} + \eta \mathbf{d}) \quad (\text{A.5})$$

La búsqueda de η^* se efectúa mediante métodos de *búsqueda en línea* (o búsqueda unidimensional).

Métodos basados en Gradiente.

Cuando la dirección \mathbf{d} , que apunta directamente hacia abajo (el punto mínimo), se determina con base en el **gradiente** (\mathbf{g}) de una función objetivo E , a tales métodos descendientes se les llama *métodos descendientes basados en gradiente*.

El gradiente de una función diferenciable E (tal que $R^n \rightarrow R$ en θ), es el vector de la primera derivada de E , denotado como \mathbf{g} . Esto es:

$$\mathbf{g}(\theta) (= \nabla E(\theta)) \equiv \left[\frac{\partial E(\theta)}{\partial \theta_1}, \frac{\partial E(\theta)}{\partial \theta_2}, \dots, \frac{\partial E(\theta)}{\partial \theta_n} \right]^T \quad (\text{A.6})$$

Por sencillez, se utiliza simplemente \mathbf{g} , omitiendo el argumento $\boldsymbol{\theta}$ en $\mathbf{g}(\boldsymbol{\theta})$.

En general, tomando como base un gradiente dado, la dirección hacia abajo cumple la siguiente *condición para posibles direcciones descendientes* (lo cual no garantiza convergencia de los algoritmos):

$$\phi'(0) = \left. \frac{dE(\boldsymbol{\theta}_{act} + \eta \mathbf{d})}{d\eta} \right|_{\eta=0} = \mathbf{g}^T \mathbf{d} = \|\mathbf{g}^T\| \|\mathbf{d}\| \cos(\xi(\boldsymbol{\theta}_{act})) < 0 \quad (\text{A.7})$$

donde ξ es el ángulo entre \mathbf{g} y \mathbf{d} , y $\xi(\boldsymbol{\theta}_{act})$ es el ángulo entre \mathbf{g}_{act} y \mathbf{d} en el punto actual $\boldsymbol{\theta}_{act}$. Empleando expansión en series de Taylor para E :

$$E(\boldsymbol{\theta}_{act} + \eta \mathbf{d}) = E(\boldsymbol{\theta}_{act}) + \eta \mathbf{g}^T \mathbf{d} + \mathcal{O}(\eta^2) \quad (\text{A.8})$$

El segundo término del lado derecho de la igualdad dominará sobre el tercero y otros términos de orden mayor de η cuando $\eta \rightarrow 0$. Con esos valores pequeños de η , la desigualdad A.3 claramente se mantiene cuando $\mathbf{g}^T \mathbf{d} < 0$.

Un tipo de métodos basados en el *descendiente del gradiente* tienen la siguiente forma fundamental, en la cual las posibles direcciones en descenso se pueden determinar *desviando* los gradientes mediante una multiplicación por \mathbf{G} (llamados *gradientes desviados*).

$$\boldsymbol{\theta}_{sig} = \boldsymbol{\theta}_{act} - \eta \mathbf{G} \mathbf{g} \quad (\text{A.9})$$

para algún tamaño de paso positivo η y alguna matriz positiva y definida \mathbf{G} . Claramente, cuando $\mathbf{d} = -\mathbf{G} \mathbf{g}$, la condición de dirección del gradiente (A.7) se mantiene, puesto que $\mathbf{g}^T \mathbf{d} = -\mathbf{g}^T \mathbf{d} \mathbf{G} < 0$. Muchas otras variantes de métodos basados en gradiente (por ejemplo, el método de Newton y el de Levenberg-Marquardt) tienen la forma arriba mencionada de ajustar la dirección negativa del gradiente ($-\mathbf{g}$) para una mejor elección.

Idealmente, se desea encontrar un valor de $\boldsymbol{\theta}_{sig}$ que satisfaga la siguiente relación:

$$\mathbf{g}(\boldsymbol{\theta}_{sig}) = \left. \frac{\partial E(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{sig}} = 0 \quad (\text{A.10})$$

En la práctica, sin embargo, resulta difícil resolver analíticamente la ecuación A.10. Es por ello que, a fin de minimizar la función objetivo, los procedimientos descendientes típicamente se repiten hasta que se satisface alguno de los siguientes criterios:

- ✧ El valor de la función objetivo es suficientemente pequeño.
- ✧ La longitud del vector gradiente \mathbf{g} es menor que un valor especificado.
- ✧ Se ha excedido el tiempo de cómputo especificado.

El método descendiente abrupto.

También conocido como el *método del gradiente*, esta es una técnica antigua para minimizar una función dada que está definida en un espacio de entrada multidimensional. A pesar de su lenta convergencia, esta es una técnica empleada frecuentemente en optimización no lineal, debido a su simplicidad.

Cuando $\mathbf{G} = \eta \mathbf{I}$, para algún valor positivo de η y la matriz identidad \mathbf{I} , la ecuación A.9 resulta ser la fórmula del *descendiente abrupto*:

$$\boldsymbol{\theta}_{sig} = \boldsymbol{\theta}_{act} - \eta \mathbf{g} \quad (\text{A.11})$$

En las ecuaciones A.7 y A.8 se observa que, si $\cos\xi = -1$ (esto es, \mathbf{d} apunta en la misma dirección de la dirección del gradiente negativo $-\mathbf{g}$), la función objetivo se puede decrementar localmente por la suma máxima en el punto actual θ_{act} . Esto implica que la dirección del gradiente negativo ($-\mathbf{g}$) apunta en la dirección localmente más abrupta hacia abajo. Desde una perspectiva global, dirigirse en la dirección del gradiente negativo, puede no ser lo más óptimo para alcanzar el punto mínimo θ^* .

Si el método descendiente abrupto utiliza minimización en línea (ec. A.4), esto es, si el punto mínimo η^* en una dirección \mathbf{d} se obtiene para cada iteración, se tiene lo siguiente:

$$\phi(\eta) = \frac{dE(\theta_{act} + \eta\mathbf{g}_{act})}{d\eta} = \nabla^T E(\theta_{act} + \eta\mathbf{g}_{act})\mathbf{g}_{act} = \mathbf{g}_{sig}^T \mathbf{g}_{act} = 0 \quad (A.12)$$

donde \mathbf{g}_{sig} es el vector gradiente en el punto siguiente. Esta ecuación indica que el vector gradiente siguiente \mathbf{g}_{sig} siempre es ortogonal al vector gradiente actual \mathbf{g}_{act} .

En la ec. A.9, el tamaño del paso η determina la eficiencia del proceso completo de minimización. En la ec. A.11 se utiliza un tamaño de paso pequeño η que es fijo. Una versión ligeramente diferente se puede obtener al normalizar el gradiente:

$$\theta_{sig} = \theta_{act} - \kappa \frac{\mathbf{g}}{\|\mathbf{g}\|} \quad (A.13)$$

donde κ es el tamaño de paso *real*, que indica la distancia euclidiana de la transición de θ_{act} a θ_{sig} :

$$\kappa = \|\theta_{sig} - \theta_{act}\| \quad (A.14)$$

A la ec. A.11 se le llama *el descendiente abrupto simple*, y la ec. A.13 es la *versión normalizada del descendiente abrupto simple*.

La magnitud del paso $\eta\mathbf{g}$ en A.11, con η fijo, cambia automáticamente con cada iteración, debido a los diferentes gradientes de \mathbf{g} . Si el punto mínimo cae en una zona plana (meseta), entonces \mathbf{g} tiende a ser infinitesimalmente pequeño, y el descendiente abrupto tiene una convergencia lenta. Por otro lado, la versión normalizada del descendiente abrupto, con κ fija, siempre tiene la misma longitud de paso, independientemente de lo abrupto de la pendiente.

El método hasta aquí descrito, se puede emplear, con sus diferentes variaciones, en el *aprendizaje neurodifuso* de sistemas como el que se utilizó en este trabajo de tesis (ANFIS). La principal diferencia entre los algoritmos basados en el descendiente del gradiente (Newton-Raphson, Levenberg-Marquard, Gauss-Newton, etc.), reside en seleccionar direcciones descendientes sucesivas. Una vez que se ha determinado la dirección hacia abajo, todos los algoritmos requieren dar un paso hacia el valor mínimo sobre la línea correspondiente.

Apéndice B: Método de mínimos cuadrados.

Estimador de mínimos cuadrados.

El problema general del método de mínimos cuadrados toma como punto de partida la salida de un modelo lineal y , definido por la siguiente expresión lineal parametrizada:

$$y = \theta_1 f_1(\mathbf{u}) + \theta_2 f_2(\mathbf{u}) + \dots + \theta_n f_n(\mathbf{u}) \quad (B.1)$$

donde $\mathbf{u} = [u_1, \dots, u_p]^T$ es el vector de entrada del modelo, f_1, \dots, f_n son funciones conocidas de \mathbf{u} , y $\theta_1, \dots, \theta_n$ son parámetros desconocidos a ser estimados. En estadística, la labor de ajustar datos empleando un modelo lineal se denomina *regresión lineal*. De esta forma, la ec. B.1 también se conoce como la *función de regresión*, y los coeficientes θ_i se llaman *coeficientes de regresión*.

Para identificar los parámetros desconocidos θ_i , normalmente se deben realizar experimentos para obtener un *conjunto de datos de entrenamiento*, compuesto de pares de datos $\{(\mathbf{u}_i, y_i), i=1, \dots, m\}$; éstos representan los pares deseados de entrada/salida del sistema a ser modelado. Al sustituir cada par de datos en la ec. B.1, se obtiene un conjunto de m ecuaciones lineales:

$$\begin{cases} f_1(\mathbf{u}_1)\theta_1 + f_2(\mathbf{u}_1)\theta_2 + \dots + f_n(\mathbf{u}_1)\theta_n = y_1 \\ f_1(\mathbf{u}_2)\theta_1 + f_2(\mathbf{u}_2)\theta_2 + \dots + f_n(\mathbf{u}_2)\theta_n = y_2 \\ \vdots \\ f_1(\mathbf{u}_m)\theta_1 + f_2(\mathbf{u}_m)\theta_2 + \dots + f_n(\mathbf{u}_m)\theta_n = y_m \end{cases} \quad (\text{B.2})$$

Usando notación matricial, la ecuación anterior se puede escribir en una forma concisa:

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{y} \quad (\text{B.3})$$

donde \mathbf{A} es una matriz de $m \times n$ (algunas veces llamada la *matriz de diseño*), definida como:

$$\mathbf{A} = \begin{bmatrix} f_1(\mathbf{u}_1) & \dots & f_n(\mathbf{u}_1) \\ \vdots & & \vdots \\ f_1(\mathbf{u}_m) & \dots & f_n(\mathbf{u}_m) \end{bmatrix} \quad (\text{B.4})$$

$\boldsymbol{\theta}$ es un vector de $n \times 1$, de parámetros desconocidos:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \quad (\text{B.5})$$

\mathbf{y} es un vector de salida de $m \times 1$:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \quad (\text{B.6})$$

La i -ésima fila de la matriz de datos $[\mathbf{A}; \mathbf{y}]$, denotada por $[\mathbf{a}_i^T; y_i]$, se relaciona con el i ésimo par de datos de entrada/salida (\mathbf{u}_i, y_i) mediante la siguiente relación:

$$\mathbf{a}_i^T = [f_1(\mathbf{u}_i), \dots, f_n(\mathbf{u}_i)] \quad (\text{B.7})$$

Para encontrar una solución única del vector desconocido $\boldsymbol{\theta}$, es necesario que $m \geq n$. Si \mathbf{A} es una matriz cuadrada y no singular, se puede resolver $\boldsymbol{\theta}$ de la ec. B.3 como sigue:

$$\boldsymbol{\theta} = \mathbf{A}^{-1} \mathbf{y} \quad (\text{B.8})$$

Sin embargo, m es normalmente mayor que n , indicando que se tienen más pares de datos que parámetros ajustables. En este caso, una solución exacta que satisfaga las m ecuaciones no siempre es posible, debido a que los datos pueden tener ruido, o el modelo puede no ser adecuado para describir al sistema objetivo. Así, la ecuación B.3 se debe modificar para incorporar un vector de error \mathbf{e} a fin de considerar el ruido aleatorio o el error de modelado, de tal forma que:

$$\mathbf{A}\boldsymbol{\theta} + \mathbf{e} = \mathbf{y} \quad (\text{B.9})$$

De esta forma, en lugar de encontrar la solución exacta de la ec. B.3, se busca obtener una $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ que minimice la suma del error cuadrado, definida por:

$$E(\boldsymbol{\theta}) = \sum_{i=1}^m (y_i - \mathbf{a}_i^T \boldsymbol{\theta})^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) \quad (\text{B.10})$$

donde $\mathbf{e} = \mathbf{y} - \mathbf{A}\boldsymbol{\theta}$ es el vector de error producido para una elección en particular de $\boldsymbol{\theta}$. Debe notarse que $E(\boldsymbol{\theta})$ está en forma cuadrática, y tiene un único mínimo en $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. El siguiente teorema establece una condición necesaria que debe satisfacer el estimador de mínimos cuadrados $\hat{\boldsymbol{\theta}}$.

Teorema B.1: El estimador de mínimos cuadrados.

El error cuadrático en la ec. B.10 se minimiza cuando $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, llamado el estimador de mínimos cuadrados (LSE), el cual satisface la *ecuación normal*:

$$\mathbf{A}^T \mathbf{A} \hat{\boldsymbol{\theta}} = \mathbf{A}^T \mathbf{y} \quad (\text{B.11})$$

Si $\mathbf{A}^T \mathbf{A}$ no es singular, $\hat{\boldsymbol{\theta}}$ es único y está dado por:

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (\text{B.12})$$

En la literatura existen varias formas de probar esta última relación, por lo que no se detallará en este espacio.

Empleando algunas propiedades de las matrices y, considerando $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, la ec. B.10 se transforma en:

$$E(\hat{\boldsymbol{\theta}}) = (\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\theta}})^T (\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\theta}}) = \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (\text{B.13})$$

Si $\mathbf{A}^T \mathbf{A}$ es singular, entonces el LSE no es único y se debe utilizar el concepto de *inverso generalizado* para encontrar $\hat{\boldsymbol{\theta}}$. Sin pérdida de generalidad, aquí se asumirá que $\mathbf{A}^T \mathbf{A}$ no es singular.

La expresión anterior (B.13) considera que cada elemento del vector de error \mathbf{e} tiene el mismo peso al evaluar el error cuadrático global. Una generalización de dicha expresión, considera que cada término del error puede ponderarse de una manera diferente. Específicamente, si se toma a \mathbf{W} como la matriz de pesos deseados, la cual es simétrica y positiva, entonces, el error cuadrático ponderado es:

$$E_w(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{A}\boldsymbol{\theta})^T \mathbf{W} (\mathbf{y} - \mathbf{A}\boldsymbol{\theta}) \quad (\text{B.14})$$

Al minimizar $E_w(\boldsymbol{\theta})$ con respecto a $\boldsymbol{\theta}$, se obtiene el *estimador de mínimos cuadrados ponderado* $\hat{\boldsymbol{\theta}}_w$:

$$\hat{\boldsymbol{\theta}}_w = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \mathbf{y} \quad (\text{B.15})$$

Obviamente, $\hat{\boldsymbol{\theta}}_w$ se reduce a $\hat{\boldsymbol{\theta}}$ cuando \mathbf{W} se elige como una matriz identidad.

Estimador de mínimos cuadrados recursivo.

La ecuación B.12 se puede expresar como:

$$\boldsymbol{\theta}_k = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (\text{B.16})$$

donde, por sencillez, no se utiliza el símbolo \wedge . Se asume que el número de filas de \mathbf{A} y \mathbf{y} es k ; este subíndice se ha agregado para denotar el número de pares de datos empleados para el estimador $\boldsymbol{\theta}$. También se puede interpretar al subíndice k como una medida del tiempo si los pares de datos están disponibles en orden secuencial. Así, por ejemplo, se tiene la tarea de encontrar una forma de utilizar un par de datos $(\mathbf{a}^T; y)$, introducido en el tiempo presente, para actualizar el valor de $\boldsymbol{\theta}_k$ en forma adecuada para encontrar $\boldsymbol{\theta}_{k+1}$. Este problema se conoce como *identificación recursiva de mínimos cuadrados*.

Empleando el par de datos presente, $\boldsymbol{\theta}_{k+1}$ se puede expresar como:

$$\boldsymbol{\theta}_{k+1} = \left(\begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \mathbf{y} \\ y \end{bmatrix} \quad (\text{B.17})$$

Para simplificar la notación, se introducen dos matrices de dimensiones $n \times n$ (\mathbf{P}_k y \mathbf{P}_{k+1}), definidas por:

$$\mathbf{P}_k = (\mathbf{A}^T \mathbf{A})^{-1} \quad (\text{B.18})$$

$$\mathbf{P}_{k+1} = \left(\begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix}^T \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} = \left(\begin{bmatrix} \mathbf{A}^T \mathbf{a} \\ \mathbf{a}^T \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \mathbf{a}^T \end{bmatrix} \right)^{-1} = (\mathbf{A}^T \mathbf{A} + \mathbf{a} \mathbf{a}^T)^{-1} \quad (\text{B.19})$$

Estas dos matrices se relacionan entre sí como sigue:

$$\mathbf{P}_k^{-1} = \mathbf{P}_{k+1}^{-1} - \mathbf{a} \mathbf{a}^T \quad (\text{B.20})$$

Empleando las ecuaciones B.18 y B.19 para definir $\boldsymbol{\theta}_k$ y $\boldsymbol{\theta}_{k+1}$, se obtiene lo siguiente:

$$\boldsymbol{\theta}_k = \mathbf{P}_k \mathbf{A}^T \mathbf{y} \quad (\text{B.21})$$

$$\boldsymbol{\theta}_{k+1} = \mathbf{P}_{k+1} (\mathbf{A}^T \mathbf{y} + \mathbf{a} y) \quad (\text{B.22})$$

$\boldsymbol{\theta}_{k+1}$ se puede expresar en términos de $\boldsymbol{\theta}_k$ eliminando $\mathbf{A}^T \mathbf{y}$ en las ecuaciones B.21 y B.22. Así, se tiene lo siguiente:

$$\mathbf{A}^T \mathbf{y} = \mathbf{P}_k^{-1} \boldsymbol{\theta}_k \quad (\text{B.23})$$

Sustituyendo este resultado en la ec. B.22, y considerando B.20, se obtiene:

$$\begin{aligned} \boldsymbol{\theta}_{k+1} &= \mathbf{P}_{k+1} (\mathbf{P}_k^{-1} \boldsymbol{\theta}_k + \mathbf{a} y) \\ &= \mathbf{P}_{k+1} [(\mathbf{P}_{k+1}^{-1} - \mathbf{a} \mathbf{a}^T) \boldsymbol{\theta}_k + \mathbf{a} y] \\ &= \boldsymbol{\theta}_k + \mathbf{P}_{k+1} \mathbf{a} (y - \mathbf{a}^T \boldsymbol{\theta}_k) \end{aligned} \quad (\text{B.24})$$

De tal forma que $\boldsymbol{\theta}_{k+1}$ queda expresada como función del valor anterior $\boldsymbol{\theta}_k$ y del nuevo par de datos $(\mathbf{a}^T; y)$. El nuevo estimador $\boldsymbol{\theta}_{k+1}$ es igual al estimador anterior $\boldsymbol{\theta}_k$ más un término de corrección basado en el nuevo dato $(\mathbf{a}^T; y)$; este término de corrección es igual a un *vector de ganancia de adaptación* $\mathbf{P}_{k+1} \mathbf{a}$ multiplicado por el error de predicción producido por el estimador previo, esto es, $(y - \mathbf{a}^T \boldsymbol{\theta}_k)$.

Sin embargo, en la ec. B.19, para el cálculo de \mathbf{P}_{k+1} se debe obtener una matriz inversa de $n \times n$, lo cual tiene un alto costo computacional. Así, de la ec. B.20, se tiene que:

$$\mathbf{P}_{k+1} = (\mathbf{P}_k^{-1} + \mathbf{a} \mathbf{a}^T)^{-1} \quad (\text{B.25})$$

Utilizando la fórmula de inversión de una matriz, la cual establece que si \mathbf{A} e $(\mathbf{I} + \mathbf{CA}^{-1}\mathbf{B})$ son matrices cuadradas no singulares, entonces la siguiente igualdad es válida: $(\mathbf{A} + \mathbf{BC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1}$ y, considerando $\mathbf{A} = \mathbf{P}_k^{-1}$, $\mathbf{B} = \mathbf{a}$ y $\mathbf{C} = \mathbf{a}^T$, se obtiene la siguiente fórmula incremental para \mathbf{P}_{k+1} :

$$\begin{aligned}\mathbf{P}_{k+1} &= \mathbf{P}_k - \mathbf{P}_k \mathbf{a} (\mathbf{I} + \mathbf{a}^T \mathbf{P}_k \mathbf{a})^{-1} \mathbf{a}^T \mathbf{P}_k \\ &= \mathbf{P}_k - \frac{\mathbf{P}_k \mathbf{a} \mathbf{a}^T \mathbf{P}_k}{1 + \mathbf{a}^T \mathbf{P}_k \mathbf{a}}\end{aligned}\quad (\text{B.26})$$

En resumen, el *estimador de mínimos cuadrados recursivo* para el problema planteado por la expresión $\mathbf{A}\boldsymbol{\theta} = \mathbf{y}$ (donde la k -ésima ($1 \leq k \leq m$) fila de $[\mathbf{A}; \mathbf{y}]$, denotada por $[\mathbf{a}_k^T; y_k]$, se obtiene en forma secuencial), se puede calcular como sigue:

$$\begin{aligned}\mathbf{P}_{k+1} &= \mathbf{P}_k - \frac{\mathbf{P}_k \mathbf{a}_{k+1} \mathbf{a}_{k+1}^T \mathbf{P}_k}{1 + \mathbf{a}_{k+1}^T \mathbf{P}_k \mathbf{a}_{k+1}} \\ \boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k + \mathbf{P}_{k+1} \mathbf{a}_{k+1} (y_{k+1} - \mathbf{a}_{k+1}^T \boldsymbol{\theta}_k)\end{aligned}\quad (\text{B.27})$$

donde $0 \leq k \leq m - 1$ y el estimador de mínimos cuadrados global $\hat{\boldsymbol{\theta}}$ es igual a $\boldsymbol{\theta}_m$, que es el estimador que emplea todos los m pares de datos.

Para iniciar el algoritmo de la ec. B.27, se deben seleccionar los valores iniciales de $\boldsymbol{\theta}_0$ y \mathbf{P}_0 . Una forma de evitar el determinar estos valores iniciales es recolectar los primeros n puntos de datos y resolver $\boldsymbol{\theta}_n$ y \mathbf{P}_n directamente de:

$$\begin{aligned}\mathbf{P}_n &= (\mathbf{A}_n^T \mathbf{A}_n)^{-1} \\ \boldsymbol{\theta}_n &= \mathbf{P}_n \mathbf{A}_n^T \mathbf{y}_n\end{aligned}\quad (\text{B.28})$$

donde $[\mathbf{A}_n; \mathbf{y}_n]$ es la matriz de datos, compuesta de los primeros n pares de datos. Así, se pueden iniciar las iteraciones del algoritmo desde el $(n+1)$ -ésimo par de datos. Sin embargo, algunas veces es más conveniente usar las fórmulas recursivas de la ec. B.27 a través del proceso de identificación. Para ello, hay que notar que:

$$\mathbf{P}_k = (\mathbf{P}_0 + \mathbf{A}_k^T \mathbf{A}_k)^{-1}\quad (\text{B.29})$$

y el valor correspondiente de $\boldsymbol{\theta}_k$ es:

$$\boldsymbol{\theta}_k = \mathbf{P}_k (\mathbf{A}_k \mathbf{y}_k + \mathbf{P}_0^{-1} \boldsymbol{\theta}_0)\quad (\text{B.30})$$

donde $[\mathbf{A}_k; \mathbf{y}_k]$ es la matriz de datos, compuesta de k pares de datos. Si se selecciona $\mathbf{P}_0 = \alpha \mathbf{I}$, se tiene que:

$$\lim_{\alpha \rightarrow \infty} \mathbf{P}_0^{-1} = \lim_{\alpha \rightarrow \infty} \frac{1}{\alpha} \mathbf{I} = 0\quad (\text{B.31})$$

Por consiguiente, eligiendo un valor grande para α , se pueden forzar las ecuaciones B.29 y B.30 a ser arbitrariamente cercanas a la ec. B.27, independientemente del valor de $\boldsymbol{\theta}_0$. En la práctica, por conveniencia, $\boldsymbol{\theta}_0$ es normalmente una matriz igual a cero.

Apéndice C: Parámetros tecnológicos de Spice proporcionados por Mosis.

A continuación se da la lista del conjunto de parámetros para el modelo de transistores MOS de 2µm llamado BSIM nivel 4, tal como la proporciona Mosis, para la simulación eléctrica de dispositivos con algoritmos Spice:

MOSIS PARAMETRIC TEST RESULTS

RUN: N89E
TECHNOLOGY: SCNA20

VENDOR: ORBIT
FEATURE SIZE: 2.0 microns

INTRODUCTION: This report contains the lot average results obtained by MOSIS from measurements of MOSIS test structures on each wafer of this fabrication lot. SPICE parameters obtained from similar measurements on a selected wafer are also attached.

COMMENTS: Orbit Semiconductor 2.0 um n-well.

SPICE BSIM MODEL PARAMETERS (2 micron N-well CMOS process)

	Name	Units	P-channel	N-channel
1)	LEVEL	—	4	4
2)	TOX	meters	25.00E-9	25.00E-9
3)	VDD	volts	5.000	5.000
4)	LD	meters	176.6E-9	393.2E-9
5)	WD	meters	612.1E-9	321.2E-9
6)	NSUB	1/cm ³	1.000E+16	1.000E+16
7)	VFB	volts	-170.8E-3	-826.5E-3
8)	PHI	volts	689.6E-3	778.4E-3
9)	K1	volts ^{-.5}	396.8E-3	861.8E-3
10)	K2	—	-40.05E-3	1.947E-3
11)	ETA	—	-10.03E-3	-3.604E-3
12)	MUZ	cm ² /v ² s	187.9	531.7
13)	U0	1/v	134.5E-3	62.90E-3
14)	U1	um/v	-5.564E-3	-10.74E-3
15)	X2E	1/v	-760.7E-6	-1.338E-3
16)	X3E	1/v	1.474E-3	903.7E-6
17)	X2MZ	cm ² /v ² s	9.123	6.700
18)	MUS	cm ² /v ² s	195.5	525.0
19)	X2MS	cm ² /v ² s	8.685	7.794
20)	X3MS	cm ² /v ² s	316.3E-3	-1.045
21)	X2U0	1/v ²	7.428E-3	616.4E-6
22)	X2U1	um/v ²	-137.5E-6	175.9E-6
23)	X3U1	um/v ²	375.4E-6	-886.9E-6
24)	LVFB	Vm	284.6E-9	34.95E-9
25)	WVFB	Vm	1.011E-6	-789.6E-9
26)	LPHI	Vm	0.000	0.000
27)	WPHI	Vm	0.000	0.000
28)	LK1	m/v [^] .5	-81.21E-9	-41.57E-9
29)	WK1	m/v [^] .5	-8.732E-9	1.445E-6
30)	LK2	m	48.08E-9	24.81E-9
31)	WK2	m	-7.917E-9	147.8E-9
32)	LETA	m	37.73E-9	7.169E-9
33)	WETA	m	45.09E-9	9.957E-9
34)	LU0	m/v	105.9E-9	113.4E-9
35)	WU0	m/v	-132.4E-9	-30.11E-9
36)	LU1	m.um/v	165.9E-9	318.1E-9
37)	WU1	m.um/v	1.043E-22	6.887E-23
38)	LX2E	m/v	-695.3E-12	-1.588E-9
39)	WX2E	m/v	2.105E-9	-22.21E-12
40)	LX3E	m/v	-1.734E-9	-1.132E-9
41)	WX3E	m/v	-6.087E-9	-4.089E-9
42)	LX2MZ	m.cm ² /v ² s	-6.230E-6	-22.46E-6
43)	WX2MZ	m.cm ² /v ² s	4.086E-6	40.37E-6
44)	LMUS	m.cm ² /v ² s	101.3E-6	171.2E-6

45)	WMUS	m.cm2/v2s	-48.37E-6	28.27E-6
46)	LX2MS	m.cm2/v2s	-1.564E-6	-30.69E-6
47)	WX2MS	m.cm2/v2s	1.810E-6	17.63E-6
48)	LX3MS	m.cm2/v2s	7.101E-6	30.54E-6
49)	WX3MS	m.cm2/v2s	-2.432E-6	416.3E-9
50)	LX2U0	m/v2	-4.508E-9	-12.12E-9
51)	WX2U0	m/v2	2.429E-9	6.789E-9
52)	LX2U1	m.um/v2	4.103E-9	-5.210E-9
53)	WX2U1	m.um/v2	5.783E-25	1.397E-24
54)	LX3U1	m.um/v2	-11.19E-9	26.25E-9
55)	WX3U1	m.um/v2	-4.630E-24	-1.668E-23
56)	N0	—	839.1E-3	1.643
57)	NB	—	-69.83E-3	-15.44E-3
58)	ND	—	53.43E-3	427.8E-3
59)	LN0	m	81.23E-9	13.63E-9
60)	WN0	m	2.823E-6	570.2E-9
61)	LNB	m	94.84E-9	63.24E-9
62)	WNB	m	274.0E-9	556.8E-9
63)	LND	m	80.08E-9	-474.3E-9
64)	WND	m	-580.6E-9	-2.156E-6
65)	PB	volts	800.0E-3	800.0E-3
66)	MJ	—	394.7E-3	372.4E-3
67)	CJ	F/m2	394.8E-6	395.3E-6
68)	PBSW	volts	850.0E-3	850.0E-3
69)	MJSW	—	330.0E-3	941.4E-3
70)	CJSW	F/m	450.0E-12	550.0E-12

BSIM MODEL PARAMETER DESCRIPTIONS

Name	Units	Description
1) LEVEL	-	BSIM model level
2) TOX	meters	Oxide thickness
3) VDD	volts	Drain DC bias
4) LD	meters	Channel length shortening due to lateral diffusion
5) WD	meters	Channel width shortening
6) NSUB	1/cm3	Substrate doping concentration
7) VFB	volts	Flat-band voltage
8) PHI	volts	Surface inversion potential
9) K1	volts ^{-.5}	Body-effect coefficient
10) K2	-	Source and drain depletion charge sharing coefficient
11) ETA	-	Drain-induced barrier lowering coefficient
12) MUZ	cm2/v2s	Carrier mobility at Vds = 0
13) U0	1/v	Vertical field mobility degradation coefficient
14) U1	um/v	Velocity saturation coefficient
15) X2E	1/v	ETA dependence coefficient to Vbs
16) X3E	1/v	ETA dependence coefficient to Vds
17) X2MZ	cm2/v2s	Mobility dependence coefficient to Vbs (at Vds = 0)
18) MUS	cm2/v2s	Carrier saturation mobility (at Vds = Vdd)
19) X2MS	cm2/v2s	Saturation mobility dependence coefficient to Vbs
20) X3MS	cm2/v2s	Saturation mobility dependence coefficient to Vds
21) X2U0	1/v2	Dependence coefficient of U0 to Vbs
22) X2U1	um/v2	Dependence coefficient of U1 to Vbs
23) X3U1	um/v2	Dependence coefficient of U1 to (Vds - Vdd)
24) LVFB	Vm	Channel-length sensitivity of VFB
25) WVFB	Vm	Channel-width sensitivity of VFB
26) LPHI	Vm	Channel-length sensitivity of PHI
27) WPHI	Vm	Channel-width sensitivity of PHI
28) LK1	m/v ^{.5}	Channel-length sensitivity of K1
29) WK1	m/v ^{.5}	Channel-width sensitivity of K1
30) LK2	m	Channel-length sensitivity of K2
31) WK2	m	Channel-width sensitivity of K2
32) LETA	m	Channel-length sensitivity of ETA
33) WETA	m	Channel-width sensitivity of ETA
34) LU0	m/v	Channel-length sensitivity of U0
35) WU0	m/v	Channel-width sensitivity of U0
36) LU1	m.um/v	Channel-length sensitivity of U1

37)	WU1	m.um/v	Channel-width sensitivity of U1
38)	LX2E	m/v	Channel-length sensitivity of X2E
39)	WX2E	m/v	Channel-width sensitivity of X2E
40)	LX3E	m/v	Channel-length sensitivity of X3E
41)	WX3E	m/v	Channel-width sensitivity of X3E
42)	LX2MZ	m.cm2/v2s	Channel-length sensitivity of X2MZ
43)	WX2MZ	m.cm2/v2s	Channel-width sensitivity of X2MZ
44)	LMUS	m.cm2/v2s	Channel-length sensitivity of MUS
45)	WMUS	m.cm2/v2s	Channel-width sensitivity of MUS
46)	LX2MS	m.cm2/v2	Channel-length sensitivity of M2MS
47)	WX2MS	m.cm2/v2s	Channel-width sensitivity of M2MS
48)	LX3MS	m.cm2/v2s	Channel-length sensitivity of X3MS
49)	WX3MS	m.cm2/v2s	Channel-width sensitivity of X3MS
50)	LX2U0	m/v2	Channel-length sensitivity of X2U0
51)	WX2U0	m/v2	Channel-width sensitivity of X2U0
52)	LX2U1	m.um/v2	Channel-length sensitivity of X2U1
53)	WX2U1	m.um/v2	Channel-width sensitivity of X2U1
54)	LX3U1	m.um/v2	Channel-length sensitivity of X3U1
55)	WX3U1	m.um/v2	Channel-width sensitivity of X3U1
56)	N0	-	Bias independent subthreshold slope coefficient
57)	NB	-	V _{bs} dependence of subthreshold slope coefficient
58)	ND	-	V _{ds} dependence of subthreshold slope coefficient
59)	LN0	m	Channel-length sensitivity of N0
60)	WN0	m	Channel-width sensitivity of N0
61)	LNB	m	Channel-length sensitivity of NB
62)	WNB	m	Channel-width sensitivity of NB
63)	LND	m	Channel-length sensitivity of ND
64)	WND	m	Channel-width sensitivity of ND
65)	PB	volts	Bulk junction potential
66)	MJ	-	Bulk junction bottom grading coefficient
67)	CJ	F/m2	Zero-bias bulk junction bottom capacitance per square meter of junction area
68)	PBSW	volts	Bulk sidewall junction potential
69)	MJSW	-	Bulk junction sidewall grading coefficient
70)	CJSW	F/m	Zero-bias bulk junction sidewall capacitance per meter of junction

Referencias.

- [1] J. S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Logic Inference System", IEEE Trans. on Systems, Man, and Cybernetics, Vol. 23, No. 3, pp. 665-685, May 1993.
- [2] Bart Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice Hall, NJ, 1992.
- [3] M. T. Hagan, H. B. Demuth and M. Beale, *Neural Network Design*, PWS Publishing Company, Boston, MA, 1995.
- [4] M. Sugeno, and G. T. Kang, "Structure Identification of Fuzzy Model", Fuzzy Sets and Systems, vol. 28, pp. 15-33, 1988.
- [5] J.-S. R. Jang, "Neuro-Fuzzy Modeling: Architectures, Analyses, and Applications", Ph.D. Dissertation, EECS Department, Univ. of California at Berkeley, July, 1992.
- [6] I. Baturone, S. Sánchez-Solano, A. Barriga, J.L. Huertas, "Design Issues for the VLSI Implementation of Universal Approximator Fuzzy Systems in Computational Intelligence and Applications", Nikos E. Mastorakis, (Ed.), World Scientific Engineering Society, Vol. 25, pp. 150-155, 1999.
- [7] A. Rodríguez-Vázquez, R. Navas, M. Delgado-Restituto and F. Vidal-Verdú, "A Modular Programmable CMOS Analog Fuzzy Controller Chip", IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing, Vol. 46, No.3, pp. 251-265, March 1999.
- [8] G. J. Klir and T. A. Folger, *Fuzzy Sets, Uncertainty and Information*. Prentice Hall, 1992.
- [9] J. Bezdek, "Fuzzy Models – What Are They, and Why?", IEEE Tran. on fuzzy systems, vol. 1, No. 1, Feb. 1993.
- [10] J. S. R. Jang, C. T. Sun and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall, 1997.
- [11] E. H. Mamdani and S. Assilian, "An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller", International Journal of Man-Machine Studies Vol. 7, No. 1, pp. 1-13, 1975.
- [12] T. Takagi, and M. Sugeno, "Fuzzy Identification of Systems and its applications to Modeling and Control", IEEE Tran. on Sys., Man, and Cybernetics, vol. 15, pp. 116-132, 1985.
- [13] Y. Tsukamoto, "An Approach to Fuzzy Reasoning Method", in Madan M. Gupta, Rammohan K. Ragade and Ronald R Yager, editors, *Advances in fuzzy sets theory and applications*, pp. 137-149. North Holland, Amsterdam, 1979.
- [14] J. S. R. Jang and C. T. Sun, "Neuro-Fuzzy Modeling and Control", Proc. of the IEEE, Vol. 83, No.3, pp. 378-406, March 1995.
- [15] P. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences", Ph.D. dissertation, Harvard Univ., Cambridge. MA, 1974.
- [16] L. A. Zadeh, "Fuzzy Sets", Information and Control, Vol. 8, pp. 338-353, 1965.
- [17] J. W. Fattaruso and R. G. Meyer, "MOS Analog Function Synthesis", IEEE Journal of Solid State Circuits, Vol. SC-22, no. 6, pp. 1056-1063, Dec. 1987
- [18] F. Vidal-Verdú and A. Rodríguez-Vázquez, "Using Building Blocks to Design Analog Neuro-Fuzzy Controllers", IEEE Micro, pp. 49-57, Aug. 1995.
- [19] J. Lazzaro et al., "Winner-Take-All Networks of $O(N)$ Complexity", Advances in Neural Information Processing Systems, vol. 1, D.S. Touretzky, de., Morgan Kaufmann, San Francisco, 1989, pp. 703-711.
- [20] "MIT-BIH arrhythmia database directory" (CD-ROM), Harvard Univ. and Mass. Inst. of Tech. Div. of Health Sciences and Tech., Cambridge, MA, July 1992.
- [21] N. Goldschlager and M. J. Goldman, Principios de Electrocardiografía Clínica. El Manual Moderno, S. A. de C. V., 1991.
- [22] Tompkins, W. J., *Biomedical Digital Signal Processing*, Prentice Hall, 1993.
- [23] S. J. Orfanidis, *Optimum Signal Processing-An Introduction*, 2nd ed., New York: McGraw-Hill, 1988.
- [24] K. P. Lin and W. H. Chang, "QRS Feature Extraction Using Linear Prediction", IEEE Tran. on Biomed. Eng., vol. 36, no. 10, pp. 1050-1055, Oct. 1989.
- [25] F. M. Ham and S. Han, "Classification of Cardiac Arrhythmias Using Fuzzy ARTMAP", IEEE Trans. Biomed. Eng., vol. 43, no. 4, pp. 425-430, Apr. 1985.
- [26] J. Usher, D. Campbell, J. Vohra and J. Cameron, "Fuzzy Classification of Intra-Cardiac Arrhythmias", Eng. in Medicine and Biology Society, 18th Annual International Conference of the IEEE, Vol. 3, pp. 997-998, 1997.
- [27] N. Manaresi, R. Rovatti, E. Franchi, R. Guerrieri and G. Baccarani, "A Silicon Compiler of Analog Fuzzy Controllers: From Behavioral Specifications to Layout", IEEE Tran. on Fuzzy Systems, Vol. 4, No. 4, pp. 418-428, Nov., 1996.

- [28] S. Guo, L. Peters and H. Surmann, "Design and Application of an Analog Fuzzy Logic Controller", IEEE Tran. on Fuzzy Systems, Vol. 4, No. 4, pp. 429-438, Nov., 1996.

Bibliografía.

- 1) Allen, P. E. and D. R. Holberg, *CMOS Analog Circuit Design*, Harcourt Brace Jovanovich College Publishers, 1987.
- 2) Baturone, I., S. Sánchez-Solano, A. Barriga, J. L. Huertas, "Implementation of CMOS Fuzzy Controllers as Mixed-Signal Integrated Circuits", Trans. Fuzzy Systems, Vol. 5, No. 1, pp. 1-19, Feb. 1997.
- 3) Ciaccio, E. J., S. M. Dunn and M. Akay, "Biosignal Pattern Recognition and Interpretation Systems", IEEE Eng. in Medicine and Biology, parts 1 of 4, Set. 1993; 2 of 4 Dec. 1993; 3 of 4, Feb./Mar. 1994; 4 of 4 Apr./May 1994.
- 4) Haykin, S. S., *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice Hall, second de., 1991.
- 5) Jang, J. S. R. and N. Gulley, *Fuzzy Logic Toolbox*, Matlab Manual, The Math Works, Inc., 1995.
- 6) Jang, J. S. R. and C. T. Sun, "A Neuro-Fuzzy Classifier and Its Applications", Proc. of the Second IEEE International Conference on Fuzzy Systems, pp. 94-98, 1993.
- 7) Jang, J. S. R. and C. T. Sun, "Adaptive Network Based Fuzzy Classification", Japan/USA Symposium on Flexible Automation, Vol. 2, pp. 885-888, ASME 1992.
- 8) Krauss, T. P., L. Shure and J. N. Little, *Signal Processing Toolbox*, Matlab Manual, 3rd printing, The Math Works, Inc., 1994.
- 9) Nauck, D. and F. Klawonn, "Neuro-Fuzzy Classification Initialized by Fuzzy Clustering", Proc. 4th European Congress on Intelligent Techniques and Soft Computing (EUFIT96), Aachen 1996.
- 10) Pan, J. and W. J. Tompkins, "A Real-Time QRS Detection Algorithm", IEEE Trans. Biomed. Eng., vol. BME-32, no. 3, pp. 230-236, Mar. 1985.
- 11) Rojas, I., J.L. Bernier, M. Rodríguez-Alvarez, A. Prieto (University of Granada), "What are the Main Functional Block Involved in the Design of Adaptive Neuro-Fuzzy Inference Systems?", Proceedings of the IEEE-INNS-ENNS, International Joint Conference on Neural Networks (IJCNN'00), Oct. 30, 2000.
- 12) Suzuki, Y., "Self-Organizing QRS-Wave Recognition in ECG Using Neural Networks", IEEE Trans. on Neural Networks, Vol. 6, No. 6, Nov. 1995.
- 13) Tomazou, C. et al., eds., *Analog IC Design: The Current Mode Approach*, Peter Peregrinus, London, 1990.
- 14) Vidal-Verdú, F. and A. Rodríguez-Vázquez, "Circuits and Algorithms for Adaptive Neuro-Fuzzy Analog Chips", Proc. of the Fourth International Conference on Microelectronics for Neural Networks and Fuzzy Systems, pp. 331-338, Set. 26-28, 1994.
- 15) Vidal-Verdú, F. and A. Rodríguez-Vázquez, "Learning under hardware restrictions in CMOS fuzzy logic controllers able to extract rules from examples", Proc. IFSA'95, Sao Paulo, Brazil, pp. 189-192, July 1995.
- 16) Vidal-Verdú, F., A. Rodríguez-Vázquez, B. Linares-Barranco and E. Sánchez-Sinencio, "A Basic Building Block Approach to CMOS Design of Analog Neuro/Fuzzy System", Proc. of the Third IEEE Conference on Fuzzy Systems, IEEE World Congress on Computational Intelligence, Vol. 1, pp. 118-123 1994.