



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

REDES NEURONALES ARTIFICIALES PARA EL
RECONOCIMIENTO DE CURVAS DE BRAGG

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

F Í S I C A

P R E S E N T A :

CAROLINA RENDÓN BARRAZA



DIRECTOR DE TESIS:

DR. JUAN JAIME VEGA CASTRO

2010



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Hoja de Datos del Jurado

1. Datos del alumno

Rendón
Barraza
Carolina
55 61 29 24
Universidad Nacional Autónoma de México
Facultad de Ciencias
Física
401005543

2. Datos del tutor

Dr.
Juan Jaime
Vega
Castro

3. Datos del sinodal 1

Dr.
Humberto Andrés
Carrillo
Calvet

4. Datos del sinodal 2

Dr.
Karo
Michaelian

5. Datos del sinodal 3

Dr.
Ernesto José María de la Salet
Belmont
Moreno

6. Datos del sinodal 4

Dra.
Elizabeth
Padilla
Rodal

7. Datos del trabajo escrito

Redes Neuronales Artificiales para el Reconocimiento de Curvas de Bragg
107 p
2010

Agradecimientos

- A la Universidad Nacional Autónoma de México, por su gente y por sus ideales. Por sus Pumas.
- A mi asesor, Dr. Juan Jaime Vega Castro por su inmejorable dedicación. Ha sido un honor y un placer aprender de ti.
- A mis sinodales, Dr. Karo Michaelian, Dr. Ernesto Belmont y Dra. Elizabeth Padilla, por la excelente disposición que siempre mostraron para apoyarme.
- Al Dr. Humberto Carrillo, por todos sus consejos y por todas las facilidades que siempre puso a mi disposición. Me siento privilegiada de haber sido parte del equipo.
- A todo el equipo de Dinámica no lineal y del Macroproyecto, especialmente a José Luis y Oreste por todos sus rescates técnicos y al Dr. Alejandro Raga por Oxum. Este trabajo no se habría concretado sin su ayuda.
- A mis maestros, especialmente René Ortega y Manuel López, por la confianza que siempre han tenido en mi. Los admiro por su ejemplo de trabajo constante, realizado con amor y humildad.
- A mis padres, por estar ahí. Sé que siempre podré contar con ustedes.
- A mis abuelos, por abrirme las puertas de su casa y asegurarse hasta hoy de que nunca me falte nada.
- A la Family, mis nenas y nenes, Cris, Diana, Eliza, Chowis, Mandis, Manzano, Pável, Tilón y Tío. Sin palabras.
- A todos mis compañeros y amigos en CU, Citla, la enana, Teresiux, Karlita, Eric (still missing you), mi Ñoñis, Tony Nava el Internacional, Zeus sabrosito, Georgie boy, Julito, Mario, David, Óscar, Juan(s), Pach, Pomis, Gordis. En general, a los ñoños, al kindergarden, a los chicos del nippon, a todos los papuchos (si, a todos) y al club de fans, por las risas y las lágrimas. Pero más por las risas.
- A los niños floridos, especialmente a Lulú y Becky. Nada sería posible sin ustedes.
- A Ricitos, por la cereza en el pastel... y por las fresas (fue sin querer...).

A Pau y Pedrito
... y Alice

See you in the other side

Índice general

1. Introducción	3
1.1. Antecedentes	3
1.2. Panorama general	5
1.2.1. Espectrómetro de masas de retroceso de Oak Ridge	6
1.2.2. Experimento Heidelberg-Moscow	8
1.2.3. CHIMERA	11
1.3. Espectroscopia de curva de Bragg	14
2. Análisis de forma de pulsos y espectroscopia de curva de Bragg	18
2.1. Panorámica del análisis digital de forma de pulsos	18
2.2. Los pulsos en la ECB	20
2.3. Representación de patrones	22
2.3.1. Curvas de Bragg ideales	23
2.3.2. Curvas de Bragg con ruido	25
3. Redes neuronales	27
3.1. Perspectiva histórica de las redes neuronales artificiales	27
3.2. Aprendizaje	30
3.2.1. Paradigmas de aprendizaje	31
3.2.1.1. Aprendizaje supervisado	31
3.2.1.2. Aprendizaje reforzado	31
3.2.1.3. Aprendizaje no supervisado	32
3.2.2. Reglas de aprendizaje	32
3.2.2.1. Aprendizaje por corrección de error	32
3.2.2.2. Aprendizaje basado en la memoria	33
3.2.2.3. Aprendizaje hebbiano	35
3.2.2.4. Aprendizaje competitivo	35
3.2.2.5. Aprendizaje de Boltzmann	36
3.3. Arquitecturas de redes neuronales artificiales	37
3.4. El perceptrón de multicapa y retro-propagación del error	41
3.4.1. Algoritmo de retro-propagación del error	42

4. Análisis digital de curvas de Bragg	47
4.1. Arquitectura de la red neuronal	48
4.2. Ley de aprendizaje	49
4.3. El ruido y los conjuntos de entrenamiento y de validación	50
4.4. Resultados	53
5. Conclusiones	79
Glosario	81
Bibliografía	88

Capítulo 1

Introducción

1.1. Antecedentes

Una visión puntual del objetivo que se plantea en este trabajo puede ser formulada como: explorar una poderosa alternativa que permita desarrollar la espectroscopia de curva de Bragg (ECB). La ECB es una forma de detección e identificación de iones pesados que, presumiblemente, son los productos emergentes de alguna reacción nuclear, estudiada mediante el bombardeo de un blanco usando haces de iones obtenidos en algún laboratorio de física nuclear equipado con un acelerador de partículas.

La ECB surge hace ya casi cerca de 30 años como una propuesta pionera, pero no aislada, planteada por C. R. Gruhn [91] (ver también [21, 69, 115, 193]). Si la señal que se obtiene del ánodo de un espectrómetro de curva de Bragg es adecuadamente formada usando dos canales de electrónica con constantes de tiempo larga y corta respectivamente, se puede obtener la carga total de ionización, así como la carga de ionización en torno al *pico de Bragg* (PB), o máximo de la *curva de Bragg* (CB). Estas dos señales nos dan, respectivamente, una medida de la energía total del ion frenado así como la carga nuclear del ion o número atómico.

La ECB tiene ciertas limitaciones. Por ejemplo, se requiere que los iones observados tengan energía por arriba del PB, el cual se presenta a una energía del ion medido aproximadamente igual a 1 MeV/amu. La resolución de las mediciones de la energía total del ion es del orden del 1% al 2% del valor de la energía medida, dependiendo de la masa del ion en cuestión. Adicionalmente, existen otros factores que se verán más adelante que hacen deseable la búsqueda de alternativas que puedan incidir en la mejora de la ECB.

Por otro lado, otro interesante antecedente para este trabajo es el siguiente. Poco más de dos décadas antes del desarrollo de la ECB [156, 157], fundamentalmente a lo largo de la década de los años sesenta, en diversos laboratorios de física nuclear, se empezó a desarrollar una poderosa técnica para poder extraer información muy relevante a partir de la forma de los pulsos de las corrientes proporcionadas por los detectores de radiación nuclear, en especial sistemas de detección basados en centelladores inorgánicos. Por razones obvias, esta técnica se ha dado en llamar análisis de forma de

pulsos (AFP). Gradualmente, el uso de esta técnica cayó en desuso, pues su desempeño fue superado por los detectores de estado sólido. Sin embargo, en la década de los ochentas, el interés por esta técnica comienza a recibir un importante impulso, particularmente en el campo de la física de iones pesados [7, 60, 84]. Esto se debe a que en los experimentos con iones pesados es común emplear un gran número de detectores con objeto de discriminar entre las partículas ligeras emitidas, lo cual requiere utilizar sistemas de detección sencillos y no muy costosos. Por ejemplo, los cristales de CsI(Tl) han sido seleccionados como los centelladores básicos en diversos sistemas de detección de diferentes laboratorios de física nuclear [9, 22, 26, 61, 62, 63, 80, 86, 190, 206]. También en diversos laboratorios [6, 56, 234] orientados al estudio de la materia oscura los cristales centelladores de CsI(Tl) han despertado gran interés en combinación con AFP [119, 127, 131, 163, 224] o cristales de NaI(Tl) [36, 96, 79, 82] o un líquido centellador como el xenón [20, 96, 164]. De hecho, la idea detrás de la técnica del AFP es muy poderosa, lo que la hace factible para que sea aplicada al procesamiento de señales obtenidas usando diversos sistemas de detección de radiación nuclear. Sin embargo, es conveniente ver otro antecedente muy importante para llevar a cabo aplicaciones de AFP otros campos.

A partir de la década de los ochentas [101] empezó una incipiente tendencia en relación a los detectores semiconductores en el sentido de demandar una buena resolución en energía con razones de conteo altas o también, sensibilidad a la posición, lo cual con el paso del tiempo, fue dando lugar a cada vez más complejos sistemas de detección, ya sea arreglos de muchos detectores, detectores con muchos electrodos, o ambos. Esto se sustentó en el desarrollo paralelo de sistemas de muchos canales de procesamiento de alto rendimiento pero a la vez relativamente económicos. Los sistemas electrónicos de procesamiento digital de la información han jugado un papel fundamental en este esfuerzo, lo cual se basa en su buen rendimiento, flexibilidad y su fácil conexión a una computadora de control. Estos sistemas se basan en el uso de dispositivos como convertidores analógico digitales, FPGAs, (del inglés Field Programmable Gate Array) y procesadores digitales de señales. Estos sistemas permiten registrar una copia digital de las señales provenientes de los diversos sistemas de detección de radiación nuclear. Desde luego, esto ha tenido impacto en la forma de procesar las señales provenientes de los sistemas de detección de la radiación nuclear. Con el paso del tiempo, estos sistemas se han ido desarrollando cada vez más. Así, para finales de los noventas se contaba con muy poderosos sistemas de digitalización de forma de pulsos [114, 201, 225, 226].

Actualmente, existen varias compañías que ofrecen poderosos sistemas digitalizadores de forma de pulso que incorporan la opción de cierto grado de procesamiento digital en línea de los pulsos adquiridos (XIA, CAEN, SkuTek, Analog Devices, etc). Sin embargo, en algunos casos el procesamiento de las señales requerido es complejo, por lo que no es factible llevarlo a cabo en línea. En tales casos, las señales digitalizadas se almacenan, y posteriormente, se requerirá de alguna opción para la extracción de la información relevante. En este punto, las ideas heredadas del AFP se combinan con el poder de las técnicas de digitalización de señales para permitir una muy buena extracción de la información relevante de las señales provenientes de los sistemas de detección de la radiación, dando lugar a las técnicas digitales de análisis de forma de pulsos (ADFP). La implementación de ADFP es una tarea que en general, representa un problema computacional fuerte.

En esta tesis se presentará el uso del paradigma neurocomputacional en combinación con técnicas de ADFP aplicadas a la espectroscopia de curva de Bragg. Esto es, se explorará una alternativa para mejorar la espectroscopia de Bragg a partir de contar con registro digitales de las curvas de Bragg, cuyas formas serán analizadas o reconocidas mediante el uso de redes neuronales artificiales (RNAs) especialmente entrenadas para esa tarea.

1.2. Panorama general

En física nuclear, el AFP surge de manera natural como una técnica de discriminación de forma de pulsos. En su forma más básica, la información inherente de un pulso lineal es su amplitud y su tiempo de ocurrencia. Sin embargo, en ciertas aplicaciones se ven involucradas mediciones para las cuales la forma particular de los pulsos cobra relevancia. La forma de estos pulsos está determinada en buena medida por la forma característica de la correspondiente corriente producida en el sistema de detección como resultado de la interacción de la radiación estudiada.

Algunas aplicaciones ampliamente usadas de análisis de forma de pulsos son:

1. Discriminación respecto de la radiación gamma de fondo en centelladores orgánicos usados como detectores de neutrones rápidos [4, 35, 49, 68, 75, 116, 117, 125, 140, 143, 149, 177, 204, 208, 220].
2. Separación de varios tipos de partículas en centelladores como el CsI(Tl) [9, 22, 30, 31, 72, 97, 125, 132, 133, 151, 164, 195, 202, 203, 224, 239].
3. Discriminación entre partículas de rango corto y rango largo en contadores proporcionales [8, 15, 23, 125, 145, 162, 209].
4. Análisis de forma de pulso en detectores de silicio o de germanio [1, 2, 10, 11, 14, 25, 27, 29, 66, 85, 95, 125, 126, 135, 136, 146, 150, 158, 159, 160, 161, 194, 197, 201, 227].
5. Separación de radiaciones de alcances diferentes en detectores phoswich [26, 52, 65, 73, 74, 77, 104, 105, 125, 238].
6. Desecho de pulsos de “pile-up”, los cuales pueden ocurrir en cualquier sistema de detección [28, 47, 64, 93, 112, 125, 128, 129, 153, 165, 168, 200, 232].
7. Análisis de forma de pulso en espectrómetros de curva de Bragg [17, 21, 51, 91, 120, 155, 180, 188, 198, 214, 216, 217, 218].

Como se ve, son bastos los trabajos que han hecho uso de alguna técnica de análisis de forma de pulsos. Los desarrollos tecnológicos cada vez hacen más atractivo el uso de la misma. De hecho, en muchos casos es indispensable, no hay otra alternativa o compite favorablemente con las técnicas alternas. A continuación, como ilustración y antes de entrar en materia, se presentan brevemente, descripciones de algunas aplicaciones recientes del análisis digital de forma de pulsos (ADFP).

1.2.1. Espectrómetro de masas de retroceso de Oak Ridge

En esta sección, se quiere ejemplificar el papel que juegan las novedosas técnicas de análisis digital de pulsos, las cuales permiten obtener información que anteriormente, usando las técnicas convencionales, eran inaccesibles.

El HRIBF (Holifield Radioactive Ion Beam Facility) es, como su nombre indica, la instalación de haces radioactivos del Laboratorio Nacional de Oak Ridge (ORNL), USA, donde se producen haces de nucleótidos radioactivos de corta vida de alta calidad para estudios de núcleos exóticos (algunos ejemplos se pueden encontrar en [24, 41, 83, 86, 181, 184]). La última parte de esta instalación se centra en un espectrómetro de masas de retroceso (RMS, por sus siglas en inglés). El HRIBF RMS [86] es un espectrómetro que combina un separador de momento con un separador de masas para hacer discriminación y transporte altamente eficiente de productos de una reacción inducidas por el haz primario (Fig. 1.1).

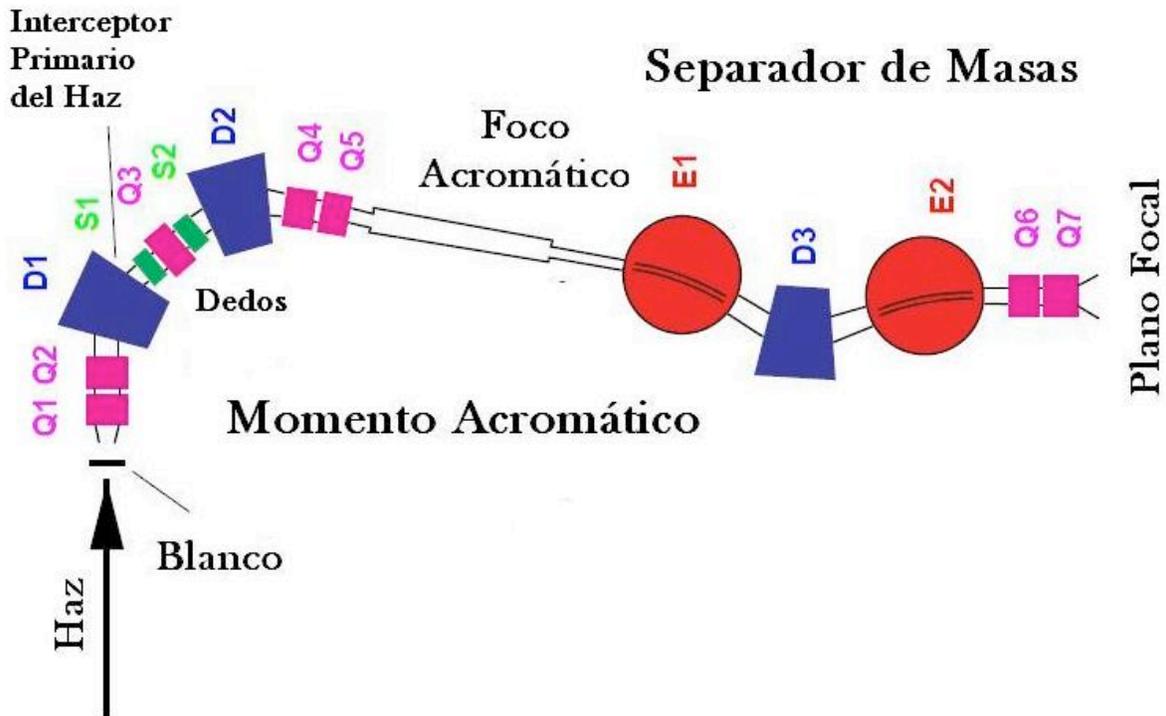


Figura 1.1: Tomada de [86]. Digitalización de la señal del preamplificador correspondiente a la implantación de un ion pesado seguida de un decaimiento rápido de éste por emisión de protón.

El RMS está diseñado para transportar iones con rigidez magnética de hasta $K = ME/Q^2 = 100$. Contiene siete cuadrupolos (Q), dos sextupolos (S), tres dipolos magnéticos (D) y dos dipolos electrostáticos (ED). Los primeros dos cuadrupolos enfocan los núcleos de retroceso y determinan el foco de los momentos que se encuentra dentro de Q3. El primer dipolo magnético separa por momento los núcleos de retroceso de las partículas del haz. La segunda mitad del separador de

momentos, D1-Q4-Q5, proporciona un segundo foco, el cual constituye el objeto de la sección del separador de masas del espectrómetro. Está posición, llamada acromática, es independiente de la energía cuando Q3 se ajusta correctamente. Los dos sextupolos proporcionan correcciones de segundo orden en el foco final. La sección del separador de masas contiene un dipolo magnético entre dos dipolos eléctricos. Los elementos eléctricos separan los núcleos de retroceso en función de la razón de su energía cinética a la carga (E/Q) y el dipolo magnético los separa en función de la razón de su momento a la carga (P/Q). El resultado neto son núcleos de retroceso separados en función de la masa y de la carga (A/Q). La magnitud de la dispersión de la masa se determina por el último par de cuadrupolos, Q6 y Q7.

Adicionalmente, se acondiciona al RMS un dispositivo llamado CLARION (clover array for radioactive ion beams), un arreglo de detectores de Ge que consiste de 11 detectores segmentados tipo *clover* y 10 elementos del viejo espectrómetro de ORNL con supresión Compton. El arreglo está instalado en la línea del haz del RMS. La mayoría de los detectores están localizados en el hemisferio posterior, aproximadamente a 25 cm del blanco. Esta distancia reduce el ensanchamiento Doppler y permite el uso de detectores auxiliares como HyBall (Hybrid- 4π), un pequeño arreglo de detectores de partículas cargadas que consiste en 95 cristales de CsI(Tl) acoplados a fotodiodos.

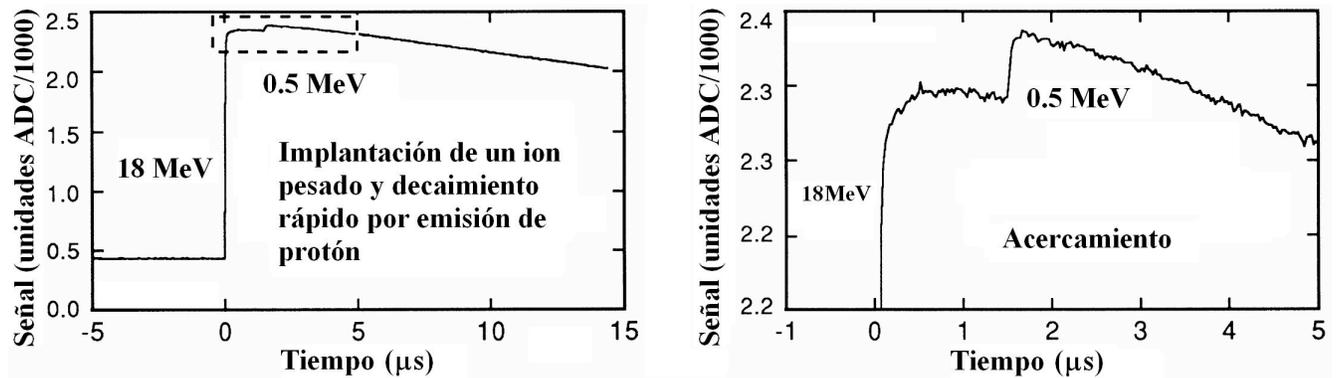


Figura 1.2: Tomada de [201]. Digitalización de la señal del preamplificador correspondiente a la implantación de un ion pesado seguida de un decaimiento rápido de éste por emisión de protón.

El interés por estudiar reacciones que involucren haces de iones radioactivos de corta vida media que decaen por emisión de protón dio lugar a una serie de experimentos que se han llevado a cabo en ORNL. Esta serie de experimentos demuestran el beneficio de usar ADFP para digitalizar las señales y obtener mejores resultados. Uno de éstos experimentos, que utilizó el RMS, es “Observation of the exotic nucleus ^{145}Tm via its direct proton decay” [24]. Los núcleos exóticos que decaen por emisión de protón depositan su energía cinética en un detector de Si de barrera superficial. La señal proveniente del detector es un gran pulso correspondiente a una energía de 18 a 30 MeV dependiendo de la reacción estudiada. La energía del correspondiente decaimiento por protón es del orden de 0.5 MeV. Este proceso de decaimiento por protón del núcleo exótico puede ocurrir en un tiempo muy corto, por ejemplo, en tan sólo $1.5 \mu\text{s}$ después de la implantación del ión. Por esta razón, cuando se usa electrónica tradicional la señal del decaimiento por protón

queda enmascarada por la señal correspondiente a la implantación, la cual, dependiendo del caso puede ser 30 o 60 veces mayor. Este enmascaramiento se debe a la sobrecarga de los circuitos analógicos por el gran pulso de implantación [24]. Con objeto de superar estas limitaciones, se ha desarrollado una nueva electrónica digital como el sistema DGF-4C [225], lo cual, debido al gran rango dinámico de esta electrónica digital, así como a su habilidad para registrar y procesar pulsos prácticamente sin tiempo muerto ha permitido implementar poderosas técnicas de ADFP, las cuales permiten observar claramente la señal del decaimiento de protón montada sobre la gran señal correspondiente a la implantación del ion en el detector de Si, la cual ocurre, como se dijo arriba, tan sólo $1.5 \mu\text{s}$ (Fig. 1.2). Se estima que podría ser posible estudiar tiempos de decaimiento dentro del régimen de submicrosegundos, que es un factor de 10 veces mejor que el que se logró en este experimento [201].

1.2.2. Experimento Heidelberg-Moscow

En esta sección, se desea ilustrar el uso de redes neuronales artificiales en combinación con técnicas de análisis de forma de pulsos para alcanzar altos niveles de discriminación del fondo.

El experimento Heidelberg-Moscow es una colaboración entre Rusia y Alemania a través de Max Planck Institut für Kernphysik, Heidelberg, Leningrad Institute of Nuclear Physics, Gatchina, y Kurchatov Institute, Moscú, para buscar el decaimiento doble beta sin neutrinos ($0\nu\beta\beta$) [136].

El decaimiento $\beta\beta$ consiste en dos transiciones β simultáneas en un núcleo. A través de la emisión de dos electrones y dos antineutrinos, un núcleo $A(Z, N)$ decae a un estado energéticamente más favorable $A(Z+2, N-2)$. Este proceso es observable sólo cuando el decaimiento β normal es prohibido debido a grandes diferencias entre los momentos angulares final e inicial (como en el caso del ^{48}Ca). El decaimiento $\beta\beta$ se presenta en varios modos:

$$\begin{aligned} 2\nu\beta\beta: & A(Z, N) \rightarrow A(Z+2, N-2) + 2e^- + 2\bar{\nu}_e \\ 0\nu\beta\beta: & A(Z, N) \rightarrow A(Z+2, N-2) + 2e^- \\ 0\nu(2)\chi\beta\beta: & A(Z, N) \rightarrow A(Z+2, N-2) + 2e^- + (2)\chi \end{aligned}$$

El primer caso está permitido por el Modelo Estándar y ha sido confirmado experimentalmente en varias ocasiones [53, 70]. Sin embargo, el decaimiento ($0\nu\beta\beta$) está prohibido por el Modelo Estándar y su observación traería importantes implicaciones en diferentes campos de investigación [122].

El experimento Heidelberg-Moscow [136] fue, en su momento, el experimento más sensible en búsqueda del decaimiento ($0\nu\beta\beta$). Consta de cinco detectores HPGe hechos de cristales enriquecidos en un 86-88 % de ^{76}Ge para medir vidas medias del orden de $T_{1/2}^{0\nu} = 1.5 \times 10^{25}$ años, requiriendo un ruido de fondo de ultra bajo nivel (0.2 cuentas/kg keV año) dentro del rango de energía de interés, entre 2000 keV y 2080 keV, donde yace el pico a 2038.56 keV (el valor Q del decaimiento $\beta\beta$). La medición de una vida media tan larga, requiere de reducir el fondo fuertemente, por lo cual, las primeras medidas en esta dirección consistieron, por un lado, en usar materiales hiperpuros para disminuir la radiación de fondo, y por otro lado, en llevar a cabo el experimento en un

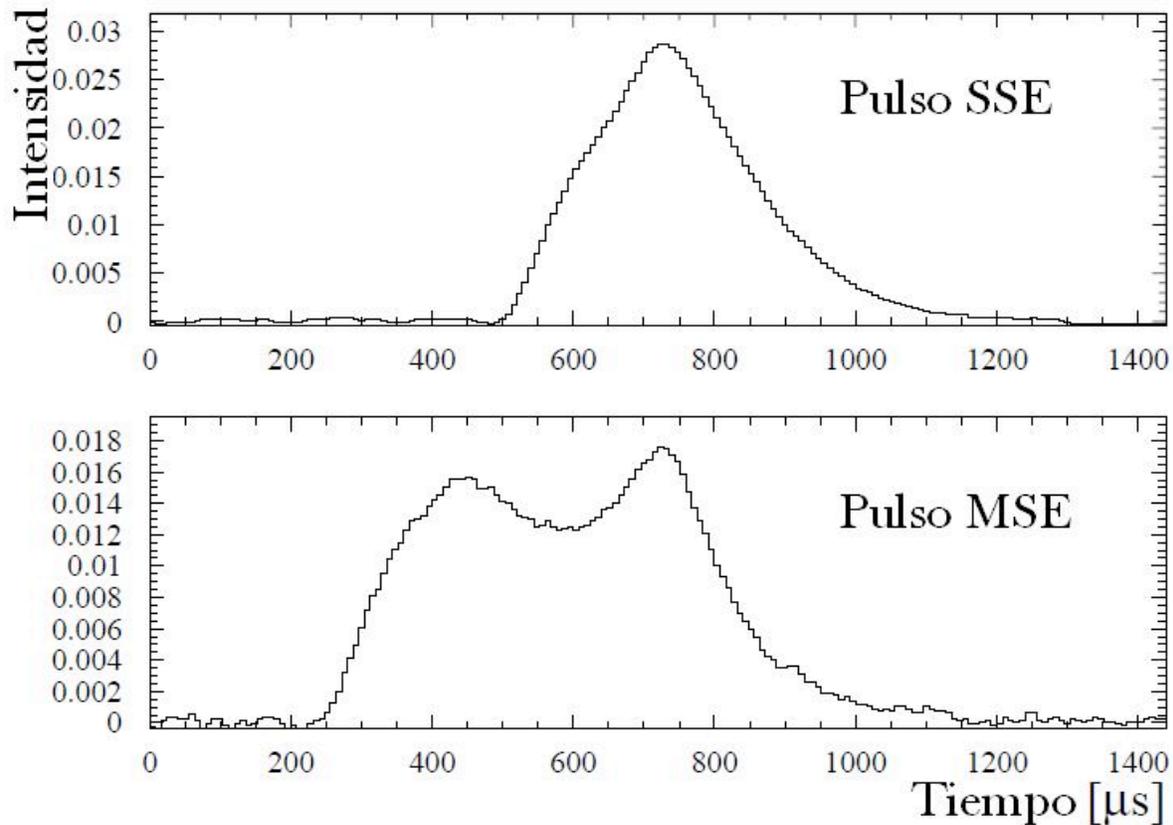


Figura 1.3: Tomada de [124]. Ejemplos de forma de pulsos del experimento Heidelberg-Moscow. La imagen superior muestra un evento SSE. Sólo se puede ver un máximo y el pulso tiene menor anchura. La imagen de abajo muestra un evento MSE típico. Dos máximos, aproximadamente de la misma energía, se pueden distinguir claramente, de tal manera que el ancho del pulso MSE es significativamente más grande que el del pulso SSE.

ambiente extremadamente blindado como el del laboratorio subterráneo Gran Sasso en Italia. A pesar de ser de gran ayuda, estas medidas no fueron suficientes, por lo que se enfocaron esfuerzos en tratar de discriminar la señal de interés con respecto a las señales de fondo. Para este propósito, se recurrió al uso del ADFP, basándose en las diferencias en las señales de salida correspondientes a diferentes distribuciones de las partículas que experimentan pérdida de energía dentro del detector. Así, se definieron eventos del tipo SSE (Single Site Events o eventos de un sólo sitio) y eventos del tipo MSE (Multiple Site Event o eventos de múltiple sitio) (Fig. 1.3). En los eventos SSE, la cercanía entre las partículas que depositan energía hace imposible que el detector pueda resolver espacialmente los eventos correspondientes. En los eventos MSE, la distancia entre las partículas que depositan la energía es mayor haciendo que el tiempo de recolección de la carga liberada sea mayor, así que los pulsos MSE son más anchos que los SSE. Por esta razón, se digitalizaron las señales de SSE y MSE con objeto de implementar una técnica de ADFP basada en un parámetro,

que tiene que ver con la anchura de los pulsos correspondientes [103]. Un decaimiento ($0\nu\beta\beta$) será del tipo SSE, lo que implica que los MSE se pueden considerar fondo.

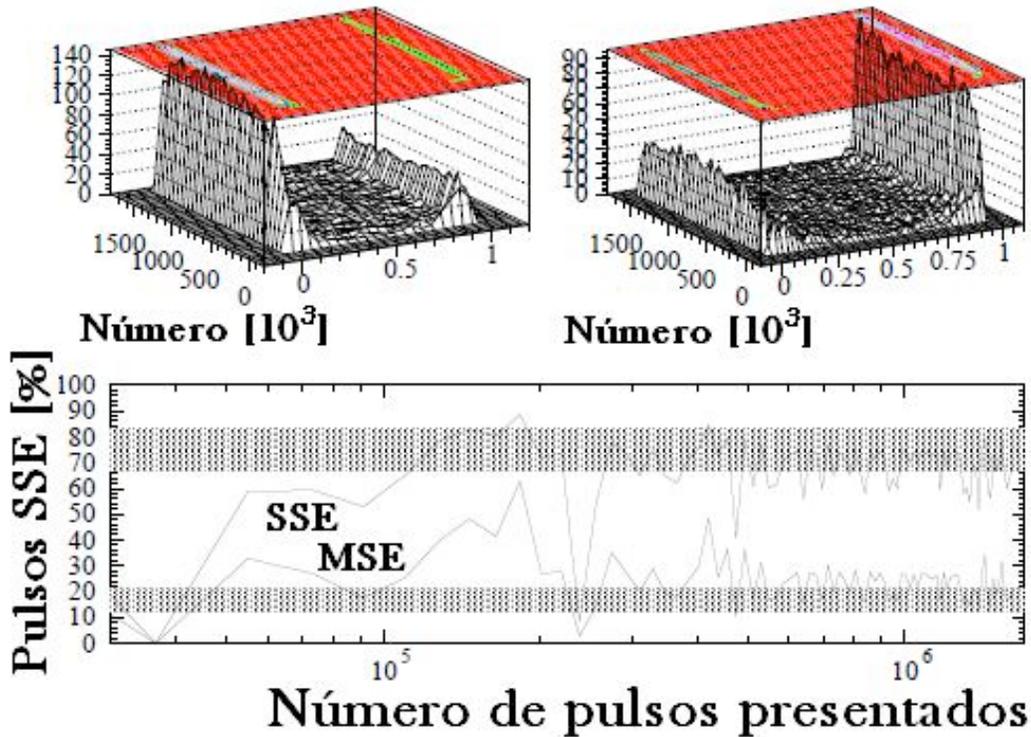


Figura 1.4: Tomada de [136]. Evolución de una red durante el proceso de entrenamiento. Panel superior: Las predicciones de la red neuronal se muestran como una función del número de pulsos presentados. Izquierda: Resultados para una biblioteca del tipo SSE preponderantemente (1592 keV). Derecha: Resultado para una biblioteca del tipo MSE preponderantemente (1621 keV). Panel inferior: Fracciones de pulsos SSE y MSE identificados en una biblioteca del tipo SSE preponderantemente, como función del número de pulsos presentados. Las áreas sombreadas corresponden al resultado esperado obtenido de una simulación.

Ya que la técnica de discriminación por forma de pulsos se puede comparar a una especie de reconocimiento de patrones, fue natural confeccionar una nueva técnica de ADFP basada en redes neuronales. Esta nueva técnica, en contraste con la anterior, donde sólo un parámetro (ancho del pulso) se utilizaba como criterio de distinción, hace uso de toda la información obtenida con la digitalización de las formas de los pulsos. Se guardaron bibliotecas de $\sim 20\,000$ eventos de cada tipo (SSE, MSE) con cada detector, los cuales se utilizaron para entrenar a la red neuronal asociada a cada detector. Se observó la evolución de la red durante el proceso de entrenamiento, como se muestra en la figura 1.4. Los autores concluyeron que está técnica es muy eficiente para clasificar entre SSE y MSE.

1.2.3. CHIMERA

En esta sección, se presenta, por un lado, una aplicación que hace uso de redes neuronales para el análisis eficiente y automático de espectros bidimensionales. Por otro lado, también se presenta una poderosa aplicación de análisis digital de forma de pulsos en relación a las señales provenientes de sistemas de detección, Si-CsI(Tl).

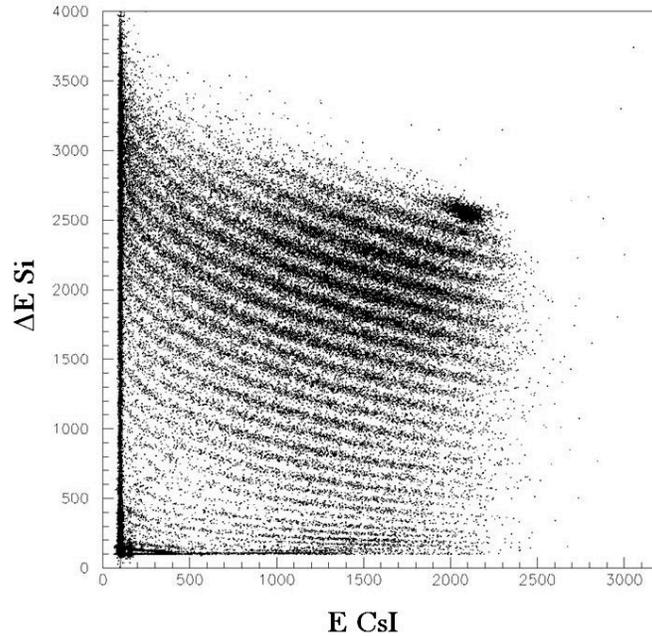


Figura 1.5: Tomada de [5]. En esta gráfica de dispersión 2D ΔE vs E , se puede reconocer a simple vista los cúmulos correspondientes a los productos de la reacción $^{58}\text{Ni} + ^{27}\text{Al}$ a 30 MeV/amu. Los autores usaron el método de Grossberg de redes neuronales de preatención para realizar la identificación de los diferentes iones pesados.

El interés por el estudio de las propiedades del núcleo bajo condiciones extremas de densidad y temperatura ha dado lugar al proyecto CHIMERA (Charge Heavy Ions Mass and Energy Resolving Array). Este es un multidetector que opera en el Laboratorio Nacional del Sur en Catania, el cual consta de 1192 celdas detectoras de Si-CsI(Tl). Este sistema detector permite la identificación, en carga y en masa, de los productos de reacciones nucleares mediante la medición de la pérdida de energía (ΔE) en los detectores de Si y la energía residual (E) en los detectores de CsI(Tl). Adicionalmente, este sistema mide el tiempo de vuelo de las partículas lo cual permite la determinación de su velocidad y su masa. Finalmente, el uso de sistemas detectores de Si-CsI(Tl) hacen posible la discriminación de partículas ligeras ($Z = 1, 2, 3$) mediante técnicas de análisis de forma de pulsos. En este proyecto, se ha dado una interesante aplicación de RNAs y ADFP para enfrentar diferentes aspectos en el análisis de los datos experimentales [5]. Este análisis consta de dos partes: la primera parte tiene que ver con la aplicación del método de Grossberg de redes neuronales de preatención [90] al análisis de gráficas de dispersión de dos dimensiones de los datos

experimentales de CHIMERA (ΔE vs E) para identificar iones pesados, y la segunda parte, usa técnicas de ADFP para identificar, en carga y en masa, partículas ligeras.

Las gráficas ΔE vs E contienen cúmulos de eventos, cada uno correspondiente a un producto de la reacción, los cuales tienen que ser integrados con objeto de obtener la información necesaria para la identificación de todos los iones detectados (Fig. 1.5). Generalmente, estas matrices de datos ΔE vs E son no densas y sus datos están caracterizados por una razón de señal a ruido baja. Adicionalmente, se observa una fuerte variación de la densidad y de la razón de señal a ruido. Esto hace poco eficiente el uso de algoritmos tradicionales para la extracción de la información. Sin embargo, los cúmulos pueden identificarse a simple vista. Esto ha sugerido usar los mismos mecanismos que a los sistemas visuales biológicos les permite la *percepción emergente*, es decir, la extracción de información oculta que está presente en el contexto pero que no está expresada explícitamente. Esto llevó a los autores al uso de RNAs para efectuar la tarea, en particular, a la utilización del método de redes neuronales de preatención de Grossberg.

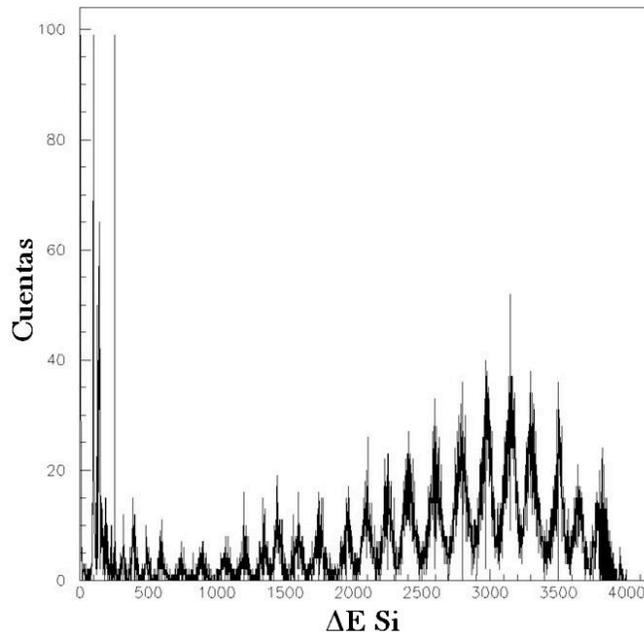


Figura 1.6: Tomada de [5]. Espectros unidimensionales que se obtienen a partir de la aplicación del método de Grossberg al análisis de los espectros de dispersión 2D ΔE vs E .

En la figura 1.6 se puede apreciar claramente como la aplicación del método de Grossberg al análisis de los espectros de dispersión 2D ΔE vs E da resultados muy satisfactorios que permiten obtener un espectro unidimensional el cual puede ser procesado usando los métodos convencionales para obtener la información correspondiente a cada uno de los productos de la reacción estudiada.

En relación al uso de ADFP para la identificación de partículas ligeras, en el proyecto CHIMERA se usó el hecho [205] de que la salida de luz ($L(t)$) de los cristales de CsI(Tl), (Fig 1.7),

está caracterizada por una combinación de dos funciones exponenciales con diferentes constantes de tiempo.

$$L(t) = (A/\tau_F)\exp(-t/\tau_F) + (B/\tau_S)\exp(-t/\tau_S) \quad (1.1)$$

En esta expresión, τ_F corresponde a la constante de tiempo rápida y τ_S a la constante de tiempo lenta. Las características de la salida de luz $L(t)$ son:

1. La razón $R = B/(A + B)$ crece cuando la densidad de ionización de la partícula incidente decrece.
2. τ_F crece cuando la densidad de ionización de la partícula incidente decrece.
3. τ_S tiene un valor cercano a $7 \mu\text{s}$ independientemente de la naturaleza de la partícula incidente.

Estas propiedades son las que permiten realizar una discriminación de pulsos en base a la forma de la salida de luz de un detector de CsI(Tl).

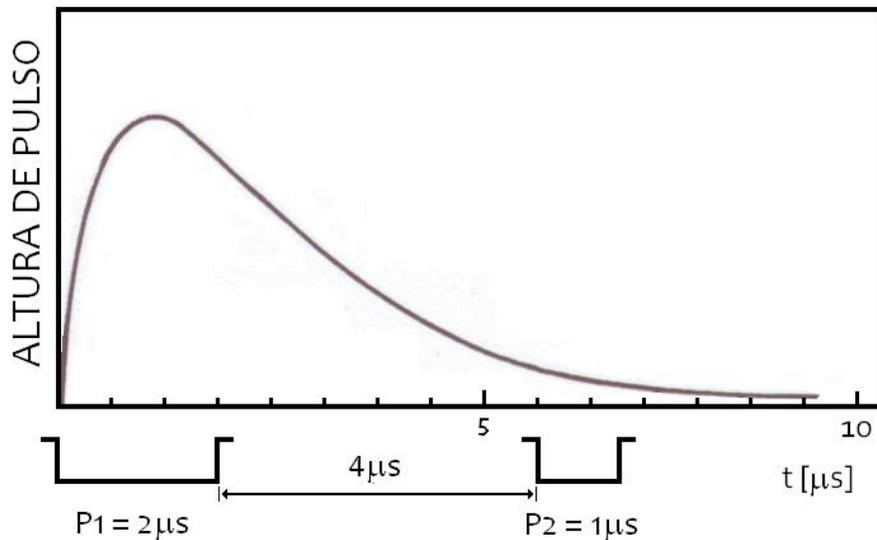


Figura 1.7: Tomada de [5]. Ilustración de la salida de luz $L(t)$ proveniente de un detector de CsI(Tl) acoplado a un fotomultiplicador. Como se indica en la figura, la señal fue integrada en carga usando tres diferente compuertas, siendo P_1 la rápida y P_2 , la lenta.

Por ejemplo, en [5], los autores efectuaron dos integraciones de $L(t)$. Una integración hizo uso de una compuerta rápida y la otra, de una compuerta lenta, dando como resultados las señales P_1 y P_2 , respectivamente (Fig. 1.7).

1.3. Espectroscopia de curva de Bragg

En este trabajo el interés se centrará en el análisis digital de forma de pulsos obtenidos con espectrómetros de curva de Bragg. En forma más específica, dado que la extracción de la información contenida en las curvas de Bragg digitalizadas requiere de mucho tiempo de procesamiento, se recurrirá a un procesamiento digital asistido por redes neuronales artificiales, las cuales brindan una poderosa alternativa como identificadores de patrones. Esta idea ha sido ampliamente probada en diversos campos de las ciencias y la tecnología, como ya hemos ilustrado arriba en las secciones precedentes.

Las diversas tareas de investigación relacionadas con el manejo o estudio de las radiaciones han dado lugar al desarrollo de diversos conceptos de detectores de radiación. Uno de estos conceptos, por cierto bastante antiguo, está basado en el efecto producido por el paso de una partícula cargada a través de un gas. Las interacciones primordiales que están involucradas en este proceso son la ionización y la excitación de las moléculas gaseosas a lo largo de la trayectoria de la partícula. La mayoría de los detectores de gas derivan la señal relevante a partir de medir la ionización directa creada por el paso de la radiación. Dependiendo de la forma particular usada para obtener la señal relevante se pueden tener diferentes tipos de detectores gaseosos, siendo los más típicos los siguientes [125, 211]: cámaras de ionización, contadores proporcionales y los tubos o contadores Geiger.

De los tres tipos de detectores gaseosos, las cámaras de ionización [125, 176, 211, 231] son, en principio, los más sencillos. Su operación normal se basa en la recolección de todas las cargas creadas por ionización directa dentro del gas del detector mediante la aplicación de un campo eléctrico. Como ocurre con otros tipos de detectores las cámaras de ionización se pueden operar tanto en *modo de corriente* como en *modo de pulso*. Un tipo de cámara de ionización muy comúnmente usada para la detección de radiación fuertemente ionizante (iones pesados energéticos) son los espectrómetros de curva de Bragg (ECB). A la gráfica del *poder de frenado específico* de un ion, $S(E) = dE/dx$, como función de la energía o de la penetración del ion en el medio gaseoso se le denomina *curva de Bragg*. En la figura 1.8 se ve que esta curva se caracteriza por alcanzar una amplitud máxima, lo cual ocurre para cualquier ion a una energía cinética aproximadamente igual a 1 MeV/amu. Este máximo de la curva se denomina *pico de Bragg*. Conceptualmente, para la identificación de las partículas y de su energía, un ECB hace uso de parte de la información disponible en la CB del ion pesado que se frena. De la definición de CB se ve que la energía de la partícula medida está dada como el área bajo esta curva, es decir, la integral de la CB. Por otro lado, resulta que el PB es una función monótona creciente del *número atómico* (Z) de la partícula que se frena. De tal suerte que midiendo el PB se puede identificar el número atómico del ion observado.

Cuando en un ECB un ion pesado, presumiblemente emergente de una reacción nuclear, es frenado, se produce un perfil espacial de carga dado por la CB (Fig. 1.8). Conforme a esto, la tarea que corresponde al ECB es de medir, de alguna forma, la información contenida en la CB producida por el ion estudiado. Para esto, se selecciona un gas, por ejemplo, P-10 (10 % metano más 90 % de argón). La presión de este gas se selecciona de tal suerte que los iones de interés más

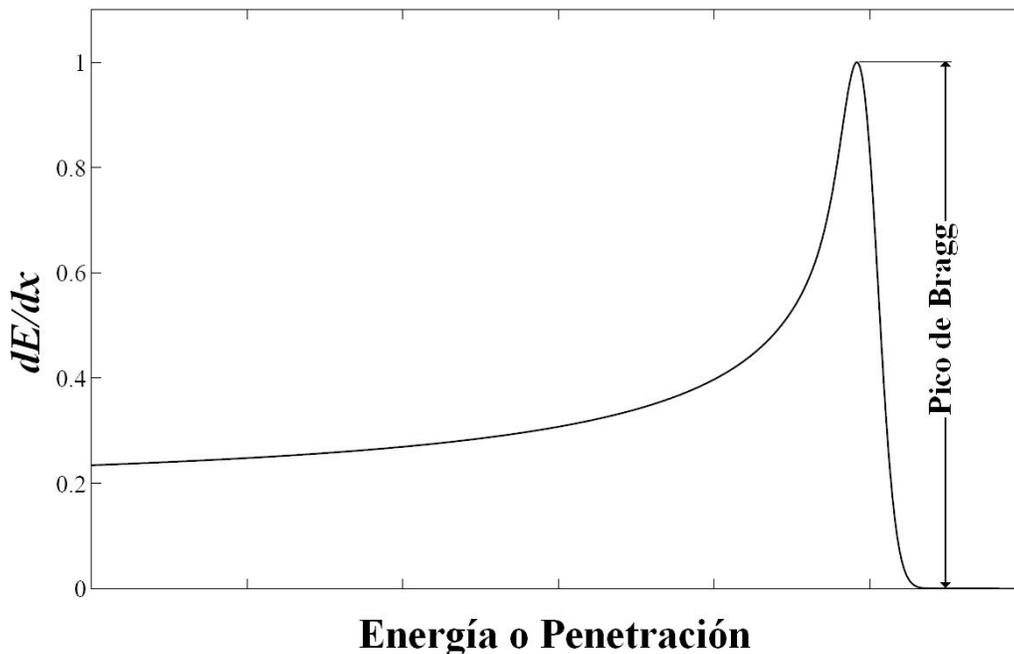


Figura 1.8: Gráfica del poder de frenado específico de un ión, $S(E) = dE/dx$, al atravesar un medio gaseoso. Esta curva se conoce como *curva de Bragg* y a la amplitud máxima se le denomina *pico de Bragg*.

energéticos alcancen a ser frenados en el gas contenido en el ECB. En la figura 1.9 se muestra un ECB y en la figura 1.10 se muestra un esquema descriptivo del mismo.

Entre el cátodo y el ánodo de una cámara de ionización cilíndrica se establece una diferencia de potencial de unos 1500 V, siendo desde luego, el ánodo el electrodo más positivo. Esto hará que los iones positivos y los electrones de los pares iónicos resultantes de la ionización de las moléculas del gas por el ion incidente deriven, hacia el cátodo los iones positivos, y hacia el ánodo los electrones; siendo la velocidad de deriva de los electrones unas mil veces más rápida que la de los iones positivos. Cerca del ánodo se encuentra una rejilla de Frisch [76], la cual se mantiene a un potencial intermedio entre el del cátodo y el ánodo y es diseñada para que sea lo más transparente posible a los electrones que derivan hacia el ánodo. La señal de interés es derivada entre el ánodo y la rejilla de Frisch. De esta manera, los electrones de cada par iónico pasarán por la misma diferencia de potencial y contribuirán en la misma cantidad a la amplitud de la señal de interés. Esto es, la amplitud total será proporcional al número original de pares iónicos formados, independientemente de su posición original de formación del otro lado de la rejilla de Frisch. El campo eléctrico uniforme de deriva se establece usando una serie de anillos formadores del campo espaciados uniformemente y conectados a una cadena de resistores divisora de voltaje.

La forma tradicional de extraer la información relevante de la señal de salida de un ECB consiste

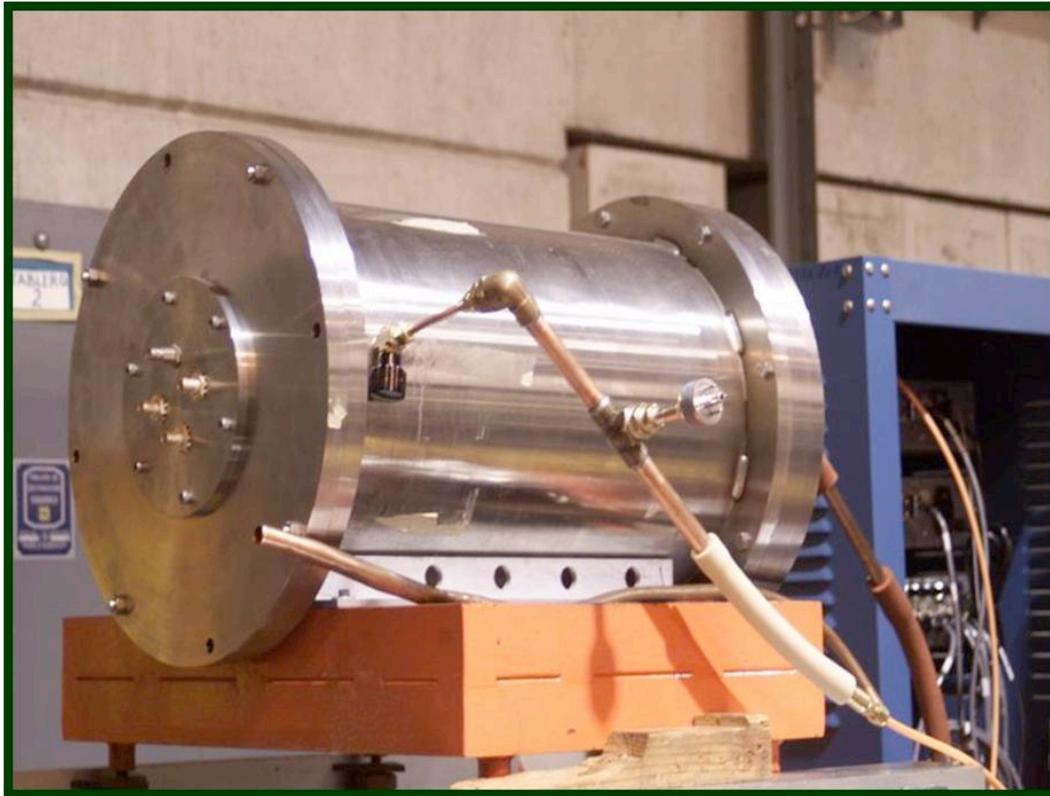


Figura 1.9: Espectrómetro de curva de Bragg. Se observa en la figura la entrada del gas P-10, el cual debe fluir continuamente para minimizar los efectos nocivos de la desgasificación de las componentes internas del ECB. También se observan los atravesadores de vacío para el manejo de las señales de salida y los voltajes de polarización requeridos.

en preamplificar la señal proveniente del ánodo del ECB y alimentar esta señal a dos cadenas de electrónica, una con un amplificador formador con una *constante de tiempo* larga $\approx 10 \mu\text{s}$ que dará una señal proporcional a la energía del ion medido y la otra a un amplificador formador con una *constante de tiempo* corta $\approx 0.5 \mu\text{s}$ que dará una señal proporcional al PB.

Una forma alternativa de procesar la señal anódica de un ECB que permite implementar técnicas digitales de análisis de forma de pulsos, consiste en digitalizar la señal que se obtiene directamente del preamplificador conectado al ánodo del ECB, es decir, digitalizar la CB. Esta imagen digitalizada de la CB deberá ser analizada con objeto de extraer la información relevante, al menos el área bajo la curva o energía cinética del ion incidente y la amplitud del PB o número atómico del ion. Otra posible información interesante que, en principio, podría ser extraída es el *alcance* del ion. La tarea de extraer la información relevante de la CB digitalizada no es una tarea trivial. De hecho, en términos generales, requiere de gran poder computacional, por lo cual, la selección adecuada del método para hacer este procesamiento digital de la información, es decir, el ADFP de la CB, es de suma importancia para poder llevar a cabo la tarea de una manera eficiente. Una

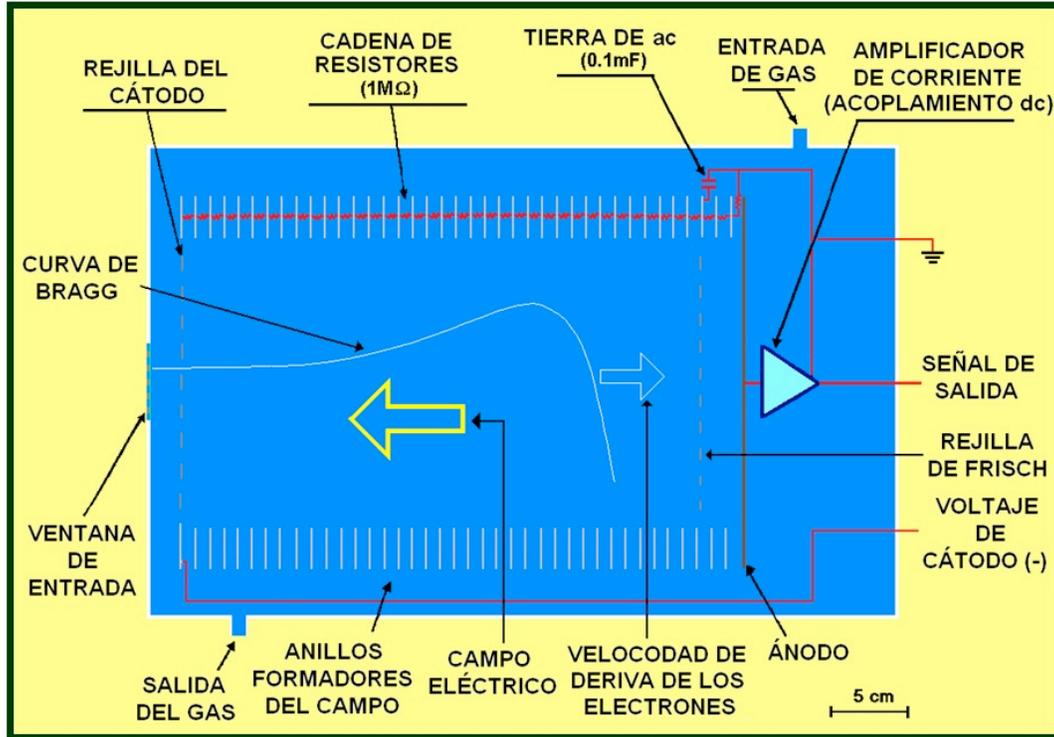


Figura 1.10: Esquema de un ECB. El perfil espacial de carga creado por el frenado de un ion que penetra el ECB se desplazará, debido a un campo eléctrico axial y uniforme, a velocidad constante hacia el ánodo.

poterosa alternativa para realizar el ADFP de las curvas de Bragg (CBs) lo proporcionan los sistemas conexionistas o redes neuronales artificiales. Estos sistemas conexionistas serán presentados más adelante en el capítulo tres, baste aquí con decir que el uso de RNAs replantea el problema del ADFP como un problema de reconocimiento de patrones, para lo cual es bien sabido que las RNAs son muy eficientes [16, 42, 50, 98, 152, 171].

Capítulo 2

Análisis de forma de pulsos y espectroscopia de curva de Bragg

2.1. Panorámica del análisis digital de forma de pulsos

En esta primera sección, se comienza presentando una breve panorámica del marco general del análisis de forma de pulso, dentro de la cual se inserta, de manera natural, la alternativa que más adelante se propone en este trabajo para el análisis de las señales provenientes de un espectrómetro de curva de Bragg.

En física nuclear, el típico espectrómetro analógico moderno procesa simultáneamente la señal del preamplificador usando un canal rápido y un canal lento. El canal rápido detecta pulsos y da indicaciones de apilamiento mientras que el canal lento da resolución de energía. Cuando un pulso válido se detecta, el pico del canal lento se captura, extiende y digitaliza con un convertidor analógico - digital (ADC, por sus siglas en inglés) y se procesa en un analizador multicanal para incluirlo en un espectro. La arquitectura de un procesador digital de pulsos (PDP) es similar, consistiendo la principal diferencia en que después de acondicionar la señal del preamplificador (circuito acondicionador), ésta se digitaliza inmediatamente, y todas las operaciones de los canales lento y rápido se llevan a cabo por filtros digitales. Es debido a estas diferencias que podremos entender las ventajas que el PDP introduce sobre la opción de procesamiento analógico de las señales.

El análisis de forma de pulsos digital es una técnica de procesamiento de señal en la cual las señales del detector (la salida del preamplificador) son directamente digitalizadas y procesadas para extraer cantidades de interés. Esta nueva técnica tiene varias ventajas significativas comparada con la técnica analógica tradicional de formado de señales:

1. Se pueden desarrollar análisis que toman en cuenta diferencias pulso a pulso, como al hacer la compensación por déficit balístico [142].
2. Las señales transientes de inducción de carga que no depositan carga neta en un electrodo, se pueden analizar para dar, por ejemplo, información muy precisa de la posición de interacción

dentro del detector [59].

3. Los tiempos muertos de señales transientes sobrecargadas se reducen grandemente, de decimas de μs a cientos de ns [166], lo cual se debe a que en los sistemas digitales la recuperación de los efectos de una sobrecarga transiente se determina por la velocidad del circuito acondicionador de la señal, la cual es rápida, mientras que en los circuitos analógicos esta recuperación se determina por el lento circuito de filtrado de energía.
4. Dado que las señales quedan capturadas permanentemente, permite que la realización de los análisis más complejos se lleven a cabo solamente para los eventos de interés.
5. Debido al registro digital de las señales, la implementación de coincidencias entre los diferentes detectores o diferentes partes del mismo detector resulta fácilmente realizable [59], lo que es difícil o imposible de implementar en un circuito analógico.

A nivel conceptual, el procesamiento de señales digital y analógico comparten varios propósitos comunes, que incluyen: (1) extraer la información de interés (altura del pulso, forma del pulso, tiempo de llegada, etc.) de un flujo de señales entrantes; (2) suprimir la información no esencial (i.e. ruido); (3) reducir el flujo de información entrante a un nivel manejable; y (4) ordenar o presentar la información extraída de una manera clara. Cada una de estas funciones es requerida en los diferentes sistemas de detección, ya sea con un simple y sencillo detector de rayos γ usado en un experimento de conteo o con un arreglo de cientos de detectores con escudos anti-Compton en grandes experimentos de física nuclear.

Una segunda diferencia es que los procesadores digitales de pulso tiene una habilidad mejorada para retrasar las señales mientras preservan la información de tiempos entre diferentes operaciones de filtración. Una línea de retraso digital, por ejemplo, es solo una memoria “first-in-first-out” (FIFO) que puede durar con facilidad decenas de μs con una fidelidad de señal perfecta. La información de los tiempos se puede conservar porque todos los filtros corren en sincronía con un reloj común y requieren un número fijo de ciclos de ejecución. Por ejemplo, ya que el tiempo entre el disparo del discriminador de un canal rápido y la llegada del pico de la señal en el canal lento es un número conocido de ciclos de reloj, la *captura del pico* en el PDP se implementa simplemente utilizando un registro controlado por un contador.

Varias alternativas para la implementación del PDP se han desarrollado. La primera fue el hacer pasar la información a través de una FIFO y, cuando un pulso es detectado, vaciar el contenido de la FIFO a un procesador digital de señales (PDS) para aplicar los algoritmos de filtración deseados. Esta alternativa es capaz de dar la más alta resolución pero es de una lentitud inherente debido al número de operaciones de procesamiento requeridas por pulso. La segunda alternativa fue implementar los filtros digitales en hardware y no en software. Esta alternativa es capaz de soportar razones de datos muy altas pero los circuitos que se requieren son complejos, caros y extremadamente consumidores de energía. La alternativa patentada por XIA, desarrollada para soportar razones de datos muy altas y a un bajo costo en aplicaciones de multicanal, divide el filtrado en dos partes. En el primer paso, una unidad de procesamiento en tiempo real (UPTR) se implementa usando lógica combinatoria, procesa la señal del preamplificador a la razón de

procesamiento de datos del ADC, usando operaciones no más complicadas que las sumas y restas. En el segundo paso, manteniendo la rapidez de la razón de datos, los valores de salida de la señal del UPTR se capturan y refinan por un PDS; produciendo buenos valores para la energía.

Una idea que subyace en el ADFP es la de permitir una gran flexibilidad en el manejo de las señales. La compañía XIA logra esto implementado su UPTR usando un FPGA (Field Programmable Gate Array), al cual, al comienzo, se le carga su programación en firme (firmware), lo mismo que el software del PDS. Esto habilita la utilización de algoritmos de procesamiento digital tanto en firmware como en software, es decir, algoritmos fácilmente modificables como para incorporar mejoras o aplicaciones especializadas, sin requerir de algún cambio de hardware.

Dentro de esta enorme perspectiva que se abre para el ADFP, en la próxima sección, se presentará una forma de asistir el ADFP mediante una novedosa alternativa computacional, la cual no se basa en el uso de algoritmos para el procesamiento digital, sino que hace uso de un novedoso paradigma computacional basado en un sistema conexionista adaptivo o red neuronal artificial, el cual aprende a extraer la información relevante a través de un proceso de aprendizaje o adaptación a un conjunto de datos representativo del ambiente de información de interés. En esta alternativa, la extracción de la información relevante se visualiza como un problema de reconocimiento de patrones.

2.2. Los pulsos en la ECB

Tradicionalmente, las *curvas de Bragg* (Fig. 1.8) han sido una técnica analítica simplificada a dos parámetros (E_{Tot} , PB) [148, 169, 213, 223]. En esta expresión, E_{Tot} representa el total de la *energía cinética* inicial del ion detectado y PB es el *pico de Bragg*, es decir, la amplitud o valor máximo de la *curva de Bragg* o *poder de frenado específico* del ion ($S(E) = dE/dx \equiv$ curva de Bragg) en un medio gaseoso. Estas dos señales se obtienen de la salida de un espectrómetro de curva de Bragg usando dos ramas de electrónica diferentes; una con una gran tiempo de integración o deformación (señal de E_{Tot}) y la otra con un tiempo de integración corto (señal de PB). La señal que se obtiene de la rama de electrónica con *constante de tiempo* larga es una medida del área bajo la CB, es decir, es una señal proporcional a la *energía cinética* total del ion frenado, es decir, E_{Tot} . Así mismo, la señal obtenida de la rama de electrónica con *constante de tiempo* corta es proporcional al valor de la amplitud máxima de la CB, es decir, el PB.

Esta técnica tiene algunas limitaciones. Por ejemplo, su límite de baja energía se establece en un valor de energía correspondiente, aproximadamente a 1 MeV/amu. Lo cual se debe a que independientemente del *número atómico* del ión, estos alcanzan su PB aproximadamente a una energía de 1 MeV/amu. De tal suerte que a energías menores, esta espectroscopia no es aplicable. El PB resulta fundamental en la ECB dado que su amplitud es una función monótona creciente del *número atómico* del ion de frenado, por lo cual su medición, resulta muy adecuada para extraer el valor de esta última cantidad.

Del párrafo anterior se desprende que la ECB es esencialmente una técnica de análisis de forma

de pulsos. En la sección anterior, se dejó clara la conveniencia de usar a ADFP en lugar de recurrir a las técnicas analógicas tradicionales. Con objeto de poder sacar ventaja del ADFP en la ECB, en este trabajo se seguirá una técnica basada en mediciones multiparamétricas consistentes en registrar y analizar la forma de pulso de la CB completa, lo cual requiere de digitalizar estas curvas. Hace unos pocos años, este procedimiento estaba muy limitado debido a la falta de electrónica adecuada, sistemas de adquisición de datos y limitaciones en el procesamiento automático de la información. Recientes adelantos en todas estas áreas, han ayudado a abrir nuevas oportunidades al mismo.

Por un lado, en este trabajo se propone utilizar un nuevo y poderoso ADFP basado en técnicas de digitalización de pulsos como las usadas en: [182, 189, 201, 214, 215, 221]. Por otro lado, en lugar de utilizar técnicas basadas en algoritmos para los PDS, se recurrirá al uso de técnicas basadas en el uso de RNAs para el análisis de la forma de los pulsos. Conforme a este paradigma computacional, la ECB se contempla como equivalente a un problema de reconocimiento de patrones, esto es, dado un patrón o CB, este debe de ser reconocido en términos de los parámetros E_{Tot} y PB.

Son diversos los factores que contribuyen a determinar la forma de las CBs experimentales. En la figura 1.8 se muestra la forma de una CB ideal, sin embargo, las CBs reales se ven modificadas por:

1. El ruido asociado a los equipos de medición.
2. Fluctuaciones de la línea base originados por un filtrado parcial de los sistemas rectificadores de los sesenta ciclos asociados a la alimentación de la electrónica.
3. La forma particular en que el sistema de medición puede seguir el cambio inicial de la señal al pasar de cero al valor finito correspondiente al comienzo de la CB, es decir, extremo izquierdo de la CB de la figura 1.8.
4. Distorsión de los pulsos por apilamiento “pile-up”.
5. Posibles cambios lentos de la forma de las CBs debidos a cambios lentos en las condiciones de operación del espectrómetro de CB.

De todos estos factores, el ruido representa el efecto más importante. De hecho, tratar de diseñar una red neuronal artificial (RNA) capaz de tomar en cuenta todos estos efectos es un problema muy complejo. Ante esta situación, se decidió enfrentar los problemas uno a uno, empezando en este trabajo, por el más importante, es decir, el efecto del ruido. La manera de llevar esta tarea a cabo será la de generar CBs sintéticas ideales modificadas por una componente de ruido, dejando para un futuro la consideración de los otros efectos. En la figura 2.1 se muestra una CB en la cual el único efecto distorsionador es una componente de ruido. Una descripción más detallada de esta curva se presentará más adelante.

Hay ventajas y desventajas tanto al digitalizar las CBs usando muchos parámetros como al digitalizar éstas usando pocos parámetros. Si se usan muchos parámetros, se contará con una imagen más fiel de la CB, pero el análisis digital correspondiente será más pesado. En caso de usar

pocos parámetros para la digitalización, los dos factores anteriores se revierten. Un inconveniente de cuando se usa una digitalización con muchos parámetros, es decir, una digitalización más rápida, es que menos carga corresponderá a cada punto de la digitalización, con un consecuente aumento del error estadístico, es decir, la razón ruido a señal crece. Como resultado al considerar este compromiso, se encontró conveniente digitalizar las CBs en una 81-tupla de parámetros $\{S(i)\}_{i=0,80}$. El análisis de estas señales es un proceso altamente consumidor de tiempo. Por esta razón, es esencial asistir el análisis mediante el uso de paradigmas computacionales nuevos y rápidos. Las RNAs proveen una buena opción para este propósito.

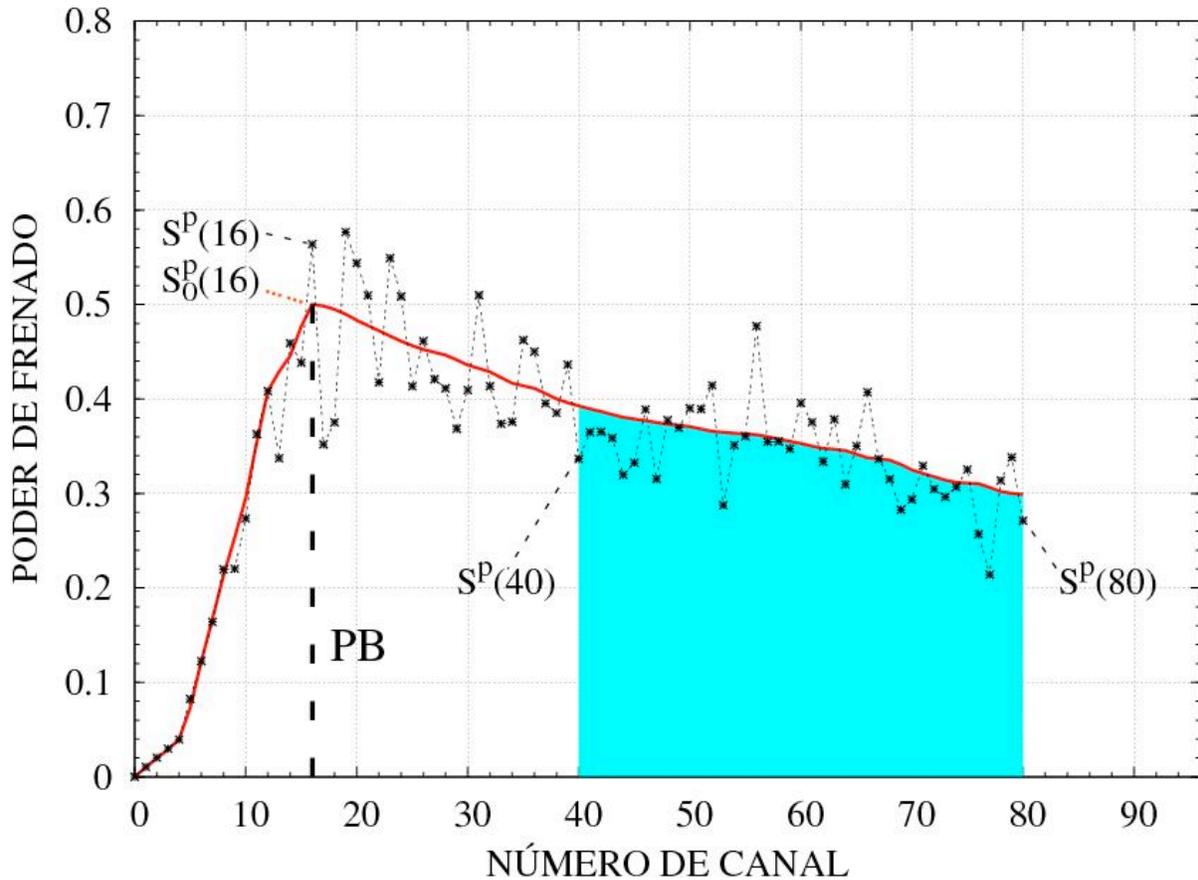


Figura 2.1: Gráfica de CBs sintéticas. La curva roja es una CB ideal y la curva punteada es una CB con ruido. En esta última el único efecto distorsionador es una componente con un tamaño de ruido igual al 10%, lo cual se describe más adelante.

2.3. Representación de patrones

Como ya se mencionó arriba, el tamaño de los patrones, o longitud de las curvas de Bragg sintéticas, es de 81 puntos $\{S(i)\}_{i=0,80}$. Estas CBs sintéticas fueron generadas de la siguiente

manera. Se usa una CB ideal lisa (línea sólida en la figura 2.1), a la cual se le añade una componente de ruido punto a punto, lo cual genera la CB sintética (línea punteada en la figura 2.1). La CB ideal se obtuvo al promediar un número grande (40) de CBs, las cuales fueron obtenidas usando partículas alfa de la misma energía, es decir, CBs de la misma longitud. Dado que la componente de ruido de cada una de las CBs es al azar, entonces esta componente promediará cero en la superposición de CBs.

2.3.1. Curvas de Bragg ideales

En esta sección se describirá la forma en que se representan las CBs ideales.

En un problema de reconocimiento de patrones es bien sabido que conviene normalizar los patrones que se quieren identificar, de tal manera que tanto los valores de las entradas como las salidas de la RNA sean cantidades entre cero y uno. Esto significa que el valor de cada punto de la CB ideal descrita en el párrafo anterior fue escalado, usando como factor de escala un medio del inverso de la amplitud del PB de la CB ideal, el cual corresponde al punto $i = 16$ de la CB ideal (Fig. 2.1). Esto es,

$$S^{nueva}(i) = \frac{1}{2} \frac{S^{vieja}(i)}{S^{vieja}(16)}, \quad i = 0, 1, \dots, 80. \quad (2.1)$$

En adelante, por comodidad, se omitirá el superíndice *nueva*, pero se entenderá que se trata de CBs normalizadas.

Las CBs de diferente *energía cinética* inicial o longitud de la digitalización se simulan definiendo $S(i) = 0$ a partir de cierto valor de i . Dado que la ECB se basa en determinar la amplitud del PB, el cual se presenta en el canal $i = 16$, $PB = S(16)$ (Fig. 2.1), entonces se decidió analizar CBs con una longitud mínima (*energía cinética* inicial mínima) de 41 puntos. En principio, usando ADFP pudiera intentarse extender la ECB a energías menores a la correspondiente al PB, el cual como ya se mencionó, ocurre a una energía aproximadamente igual a 1 MeV/amu. Sin embargo, esta sería una tarea bastante más compleja para una RNA, por lo cual esta tarea no se considerará en este trabajo. Conforme a esto, una CB de una longitud de n puntos, correspondiente a una *energía cinética* total igual a $E_{Tot}(n)$, está definida como:

$$S(i; E_{Tot}(n)) = \begin{cases} S(i) & \text{si } i \leq n \\ 0 & \text{si } i > n, \end{cases} \quad (2.2)$$

donde $n = 40, 41, \dots, 80$, es decir, 41 CBs que difieren en su longitud. Las energías $E_{Tot}(n)$ se determinaron usando una interpolación de la CB discreta, $\{S(i)\}_{i=0,80}$, usando un *spline* cúbico para ajustar los puntos. Con esto se obtiene una función continua $S^{spline}(x)$, de tal manera que la energía queda definida como:

$$E_{Tot}(n) = \int_n^0 S^{spline}(x) dx, \quad n = 40, 41, \dots, 80. \quad (2.3)$$

Para simular diferentes CBs correspondientes a iones de diferentes *números atómicos*, Z , se hizo uso del hecho de que la amplitud del PB es una función monótona creciente de la Z del ion frenado, entonces para simular diferentes valores Z se procedió a simplemente escalar cada una de las 41 CBs recién descritas, es decir, cada una de las curvas $S(i; E_{Tot}(n))$ con $n = 40, 41, \dots, 80$ da lugar a 10 CBs adicionales, las cuales simulan Z menores. Para simular CBs correspondientes a 11 diferentes *números atómicos* Z , se usaron los siguientes factores de escala: 1.00, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70, 0.65, 0.60, 0.55 y 0.50. Lo cual, tomando en cuenta la ecuación 2.1 implica que las respectivas amplitudes del PB serán:

$$PB = 0.500, 0.475, 0.450, 0.425, 0.400, 0.375, 0.350, 0.325, 0.300, 0.275 \text{ y } 0.250 \quad (2.4)$$

Esta manera de definir las CBs de las diferentes Z implica que las áreas bajo las respectivas CBs, *energías cinéticas* iniciales, se escalarán de la misma forma.

De acuerdo a lo anterior, una CB ideal (sin ruido) de un ion pesado con un *número atómico* Z , que le corresponda un valor PB y un valor de la *energía cinética* total $E_{Tot}(n)$, es decir, una digitalización de longitud n , será denotada como:

$$S_0(i; E_{Tot}(n), PB), \quad i = 0, 1, \dots, 80. \quad (2.5)$$

En la ecuación anterior, los valores son iguales a cero si $i > n$. Se puede notar un superíndice cero en la S, la razón de esto es para indicar que se trata de una CB ideal, es decir, sin ruido o lisa.

Conviene hacer notar que los valores de $E_{Tot}(n)$ en la ecuación anterior dependen del valor de PB . Por otro lado, la normalización de los valores de $E_{Tot}(n)$ es conveniente ante cualquier problema de reconocimiento de patrones. La forma en que se realizó esta normalización es la siguiente:

$$E_{Tot}^{escalada}(n, PB) = \frac{1}{2} \frac{E_{Tot}(n, PB)}{E_{Tot}(n = 80, PB = 0.5)}, \quad n = 40, 41, \dots, 80. \quad (2.6)$$

A manera de ejemplo,

$$E_{Tot}^{escalada}(n = 80, PB = 0.5) = 0.5 \quad (2.7)$$

y

$$E_{Tot}^{escalada}(n = 40, PB = 0.25) = 0.12655 \quad (2.8)$$

También con la idea de simplificar la notación, de ahora en adelante ya no se indicará el superíndice *escalada*. Todos los demás valores $E_{Tot}(n, PB)$ caerán entre los dos valores anteriores. En la figura 2.2 se despliegan en el plano $PB - E_{Tot}$ los 451 valores reescalados de los pares (PB, E_{Tot}) (451 CBs ideales) correspondientes a los 11 posibles valores de PB y a los 41 posibles valores $E_{Tot}(n, PB)$ que se originan de los diferentes tamaños de la digitalización, $n = 40, 41, \dots, 80$.

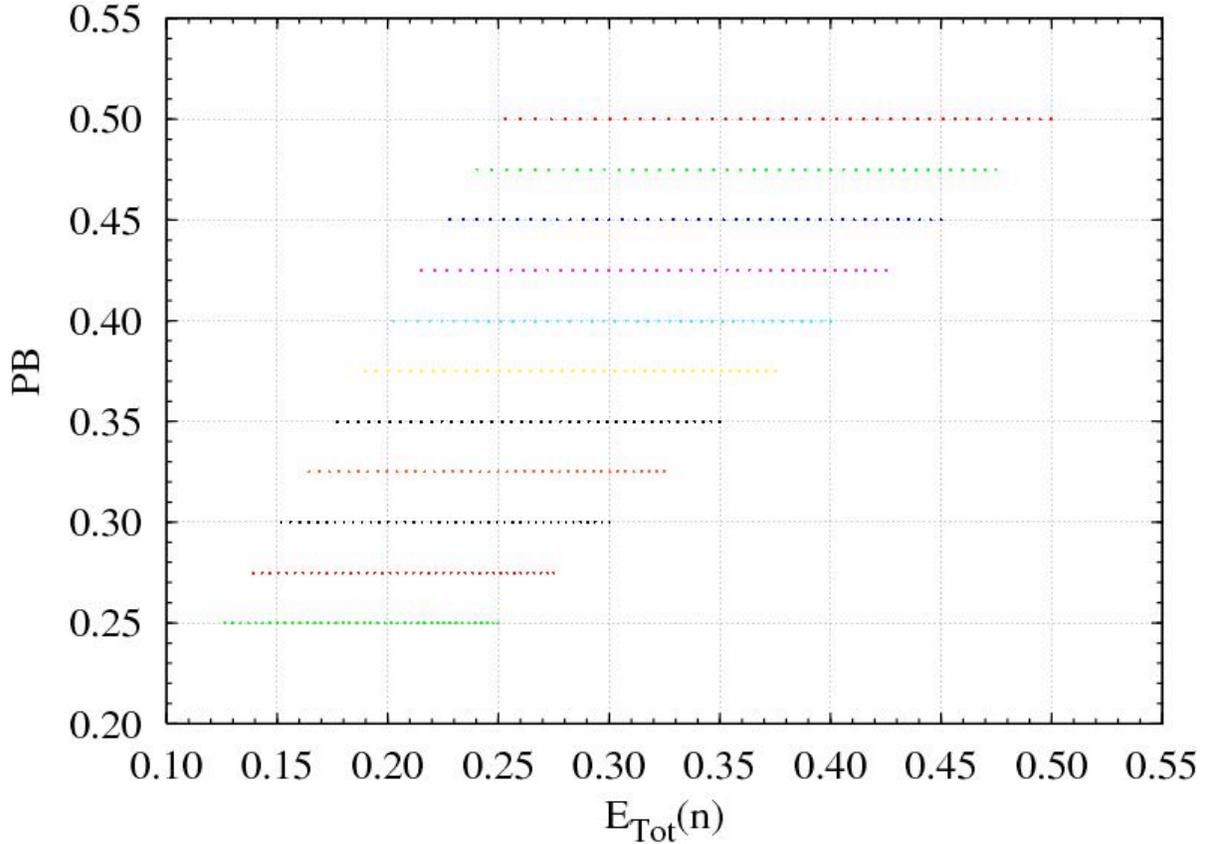


Figura 2.2: Gráfica en el plano $PB - E_{Tot}$ de los 451 valores reescalados de los pares (PB, E_{Tot}) correspondientes a las 451 CBs ideales. Cada una de estas curvas corresponde a cada una de las combinaciones de los 11 posibles valores de PB con cada uno de los 41 posibles valores $E_{Tot}(n, PB)$ que se originan de las diferentes longitudes de las CBs.

2.3.2. Curvas de Bragg con ruido

Como ya se mencionó en la sección 2.2, el efecto distorsionador de las CBs más importante es el debido al ruido experimental. Debido a eso y con objeto de no complicar demasiado la tarea de reconocimiento de patrones que llevarán a cabo las RNAs, en este trabajo nos limitaremos a estudiar el reconocimiento de patrones de CBs distorsionadas únicamente por la presencia de una componente de ruido experimental.

En vista de lo anterior se decidió estudiar CBs modificadas por la adición de ruidos de diferente tamaño, lo cual se implementó de la siguiente manera. Construimos CBs experimentales sintéticas, es decir, con ruido, denotadas como $S_r(i; E_{Tot}(n), PB)$, de la siguiente manera:

$$S_r(i; E_{Tot}(n), PB) = S_0(i; E_{Tot}(n), PB) + S_G(1; E_{Tot}(n), PB) \quad (2.9)$$

donde, como se describe arriba, $S_0(1; E_{Tot}(n), PB)$ representa una CB ideal (línea sólida en la

figura 2.1) y $S_G(1; E_{Tot}(n), PB)$, una componente de ruido que sigue una distribución Gaussiana con una desviación estándar dependiente de la energía e igual a $RS_0(1; E_{Tot}(n), PB)$ y un valor medio igual a cero (línea punteada en la figura 2.1). En donde R , o razón de ruido a señal, toma los valores: 0.01, 0.02, ... y 0.10. De esta manera, se pueden simular componentes de ruido Gaussianas de media igual a cero y desviación estándar iguales a 1%, 2%, ... y 10% de la amplitud de $S_0(1; E_{Tot}(n), PB)$. Nótese que de esta manera, la amplitud la razón ruido a señal en cada punto, depende de la energía.

De la misma manera que las CBs ideales, las longitudes digitales de las CBs con ruido se seleccionan entre $n = 40$ y $n = 80$, es decir,

$$S_r(i; E_{Tot}(n), PB) = \begin{cases} S_0(i; E_{Tot}(n), PB) + S_G(1; E_{Tot}(n), PB) & \text{si } i \leq n \\ 0 & \text{si } i > n, \end{cases} \quad (2.10)$$

en donde los valores de $E_{Tot}(n)$ y de PB se definen como antes, ecuaciones 2.4 y 2.6, respectivamente.

En la figura 2.1, se muestra un ejemplo de una CB con ruido (curva punteada), correspondiente a $PB = 0.50$, $n = 80$ y $R = 0.10$, es decir, a un ruido gaussiano cuyo tamaño (desviación estándar) es igual, en cada punto, al 10% de la amplitud de la CB ideal, $S_0(i; E_{Tot}(n))$.

Finalmente, la tarea que mediante ADFP se desea llevar a cabo, es decir, la espectroscopia de curva de Bragg digital, se puede formular como:

“Dada una CB con ruido (Ec. 2.10) se desea determinar los valores desconocidos $E_{Tot}(n)$ y PB correspondientes.”

En el próximo capítulo, se presentará un novedoso paradigma computacional basado en el uso de redes neuronales artificiales, el cual tendrá por objetivo asistir al ADFP en la tarea de implementar la espectroscopia de curva de Bragg digital.

Capítulo 3

Redes neuronales

Las redes neuronales artificiales son sistemas *conexionistas adaptivos* capaces de aprender a partir de ser expuestos ante individuos pertenecientes a algún *ambiente de información*, cuyo estudio, por algún motivo, resulta de interés. Esta breve caracterización de las RNAs está inspirada por su contraparte biológica, es decir, el funcionamiento del cerebro humano. En alguna forma en los inicios de esta aventura [55, 147], se pensó que las RNAs podrían usarse para modelar, modestamente, el funcionamiento del cerebro. Sin embargo, con el paso del tiempo, los usuarios de las RNAs se han alejado cada vez más de verlas o usarlas como modeladoras del funcionamiento del cerebro, dando lugar al desarrollo de un novedoso y poderoso paradigma computacional que puede ser muy útil en la solución de problemas complejos para los cuales los antiguos paradigmas computacionales algorítmicos no resultan adecuados. De hecho, esta forma en la que se ha dado el desarrollo de las RNAs está enmarcado en dos visiones complementarias acerca de la inteligencia artificial (IA): una como una disciplina ingenieril que tiene que ver con la creación de máquinas inteligentes; la otra como una ciencia empírica que tiene que ver con el modelado computacional de la inteligencia humana. Por ejemplo, el problema que nos interesa en este trabajo, el ADFP de CBs, cae dentro de la primera visión de la IA, es decir, es justamente uno de esos problemas ingenieriles complejos para los cuales no existe un algoritmo que de una manera eficiente lo resuelva. La característica fundamental de las RNAs que, para la solución de ciertos problemas, las hace ser una muy poderosa alternativa, es precisamente el que son sistemas capaces de responder en forma *adaptiva* al ser expuestas a un ambiente de información.

En este capítulo, con objeto de entender más cabalmente la forma en que se emplearán las RNAs para asistir el ADFP en relación a la espectroscopia de curva de Bragg, se presentará una breve perspectiva acerca de las RNAs.

3.1. Perspectiva histórica de las redes neuronales artificiales

Como se dijo arriba, las RNAs se inspiraron en su contraparte biológica, pero el cerebro y sus componentes son tan complejos que la correlación entre las RNAs y el cerebro termina tan pronto se consideran a las neuronas en detalle. Por ejemplo, las RNAs fueron tema de estudio por primera vez, cuando en 1943, Warren McCulloch y Walter Pitts [139] introdujeron su modelo

de la neurona, como una síntesis de neurofisiología y lógica donde se plantea la correspondencia de la capacidad del cerebro con los límites de computabilidad de una máquina de Turing [212]. Este trabajo dio lugar a un largo camino de investigaciones en diversas áreas del conocimiento, como por ejemplo la IA, en particular, en la neurocomputación. Estos autores parten de sencillos elementos computacionales sugeridos por el conocimiento fisiológico y psicológico que se tenía en aquella época, de las neuronas y sus conexiones, para explicar lo que hace el sistema nervioso de una manera sencilla. De hecho, los autores se centraron en los aspectos computacionales o lógicos, y no en los aspectos fisiológicos o morfológicos de las neuronas. Esto dio lugar a que el lenguaje usado por los autores en sus modelos (neurona de McCulloch-Pitts) sea un lenguaje matemático. El modelo de la neurona de McCulloch y Pitts, así como el interés que McCulloch mostró a lo largo de su carrera por investigaciones sobre el cerebro, lo colocan, junto con personajes como Norbert Wiener, John von Neumann, Heinz von Foerster y otros, como los fundadores de la cibernética.

Otra contribución fundamental que, junto con el trabajo de McCulloch y Pitts, se puede considerar como fundador de las investigaciones sobre redes neuronales artificiales y sobre los modelos cognitivos *conexionistas*, son las ideas expuestas en 1949, por Donald Olding Hebb en su libro “The organization of behavior: a neurophysiological theory” [99]. Su oposición al conductismo radical y su énfasis por entender qué ocurre entre el estímulo y la respuesta (percepción, aprendizaje, pensamiento) ayudaron a despejar el paso a la revolución cognitiva. De su gran contribución destacaremos su modelo de *aprendizaje* (hebbiano):

“Cuando un axón de una célula A está lo suficientemente cercano a una célula B como para excitarla y participa repetida o persistentemente en su disparo, ocurre algún proceso de crecimiento o cambio metabólico en una o en ambas células de modo tal que aumenten tanto la eficiencia de A como la de una de las distintas células que disparan a B.”

Una manera alternativa sería:

“Una sinapsis se refuerza si la neurona de entrada (o presináptica) y la neurona de salida (o postsináptica) son activadas de manera continua. De esta forma, las conexiones que se usan son las que se refuerzan.”

La hipótesis de Hebb significa que la neurona de McCulloch y Pitts se debía modificar si es que se deseaba una mayor cercanía con respecto a su contraparte biológica. La solución obvia fue que la fortaleza de las sinapsis se modelaría incluyendo pesos para ponderar las entradas de la neuronas artificiales. Las ideas de McCulloch y de Hebb constituyen un gran impulso inicial para el desarrollo de las investigaciones en el campo de las RNAs. Concretamente, se sugiere la posibilidad de buscar arquitecturas de RNAs, la cuales, con reglas de aprendizaje adecuadas, puedan convertirse en “*máquinas inteligentes*”. El primer intento optimista en esta dirección lo constituye el trabajo del perceptrón de Frank Rosenblatt [174]. La perspectiva optimista de este trabajo la plantea Rosenblatt de la siguiente manera:

“Para responder las preguntas de cómo la información acerca del mundo físico es sentida, de que manera se recuerda la información, y como la información retenida en la memoria influye nuestro

reconocimiento y comportamiento, se desarrolla una teoría para un sistema nervioso hipotético llamado perceptrón. La teoría sirve como puente entre la biofísica y la psicología.”

Con el paso del tiempo, Rosenblatt fue perfeccionando su perceptrón [175], pero su carácter innovador tiene que ver con la inclusión de pesos numéricos ponderadores, así como un patrón específico de interconexión neuronal [98]. Rosenblatt usó su perceptrón para resolver algunos problemas de reconocimientos de imágenes.

Hasta aquí, el avance de las RNAs se veía muy prometedor, sin embargo, las ideas propuestas por Marvin Lee Minsky y Seymour Papert en su libro “Perceptrons: an introduction to computational geometry” [144] tuvieron un gran impacto negativo sobre la comunidad en torno a la inteligencia artificial, de tal suerte que los correspondientes fondos asignados por agencias dedicadas al apoyo de las investigaciones científicas fueron reducidos drásticamente para proyectos en el área de RNAs. Es decir, el continuo duelo entre la escuela *simbólica* y la escuela *conexionista*, desde el punto de vista del financiamiento, se decidió, por una larga temporada, a favor de la escuela *simbólica*.

Hubo que esperar largos años, hasta que algo realmente novedoso ocurriera en el área de las RNAs para que el interés por éstas reviviera. A este papel glorioso, le tocó desempeñarlo al físico John Joseph Hopfield, el cual, empleando una visión de las RNAs típica de un físico, propuso un modelo [113] de *red neuronal asociativa* o *red de Hopfield*. Estas redes funcionan como *memorias asociativas* o *dirigidas por contenido*, es decir, son capaces de recuperar un patrón almacenado en memoria al ser estimuladas por la presentación de una versión parcial o ruidos del patrón. Las redes de Hopfield pueden ser visualizadas [98] como *sistemas dinámicos* con “energía” cuyo espacio fase contiene un conjunto de *puntos fijos estables* que representan las memorias fundamentales del sistema. La formalización empleada por Hopfield, permite visualizar al entrenamiento de una red de Hopfield como consistente en reducir la función de energía de los estados que la red debe “recordar”.

La próxima contribución fundamental para el desarrollo de las RNAs se debe a David Everett Rumelhart y James Lloyd McClelland quienes editaron el libro *Parallel distributed processing* [178], el cual es considerado por mucho colegas como la “biblia” de las ciencias cognitivas. En este libro se conjuntan muchos resultados técnicos importantes de investigadores del grupo “Parallel Distributed Processing” (PDP), de enfoque *conexionista*, que impulsaron la revolución de las RNAs durante los 1980s. La orientación del grupo PDP se basa en la convicción de que varios aspectos de la actividad cognitiva pueden ser visualizadas en términos de un procesamiento masivo y en paralelo. En particular, en el capítulo 8, *Learning internal representations by error propagation*, escrito por David Everett Rumelhart, Geoffrey E. Hinton, y Ronald J. Williams, quienes conjuntamente descubrieron el algoritmo de *error back-propagation* o *retro-propagación del error*, se desarrolló una nueva fundamentación matemática para explicar la cognición humana basada en modelos con variables continuas y *sistemas dinámicos* (sistemas *conexionistas*) en lugar de usar programas de computadora basados en procesamiento lógico y simbólico (ver también [179]). La contribuciones de este grupo se puede decir que ayudaron a revertir el desbalance en IA entre, por un lado, la escuela *simbólica* y, por el otro, la escuela *conexionista*, el cual se había originado, como ya se dijo, desde la publicación del libro, en 1969, “Perceptrons: an introduction to computational geometry”

por Minsky y Papert.

En realidad, el exitoso modelo computacional basado en aprendizaje por *retro-propagación del error* fue descrito con anterioridad, en 1974, por Paul J. Werbos en su tesis doctoral en Harvard [229]. Sin embargo, fue hasta 1986, con el trabajo de Rumelhart, Hinton y Williams, cuando se enteró la comunidad del modelo de aprendizaje por *retro-propagación del error*.

Con objeto de dar una mejor descripción del modelo de *retro-propagación del error* es conveniente presentar primero un breve resumen acerca de los métodos o paradigmas de aprendizaje.

3.2. Aprendizaje

La habilidad de aprender es un rasgo fundamental de la inteligencia. Aunque es difícil formular una definición precisa del aprendizaje, en el contexto de las RNAs se puede ver como el problema de actualizar la arquitectura de la red y los pesos de las conexiones utilizando patrones de entrenamiento disponibles. El desempeño se mejora con el tiempo al actualizar iterativamente los pesos en la red. La habilidad de las RNAs para aprender automáticamente dada una muestra, las hace tanto atractivas como excitantes. En lugar de seguir una serie de reglas específicas asignadas por humanos expertos, las RNAs parecen aprender reglas subyacentes (como las relaciones entrada-salida) de una colección provista de muestras representativas. Esta es la mayor de las ventajas de las redes neuronales sobre los sistemas expertos tradicionales.

Para entender o diseñar procesos de aprendizaje, se debe tener primero un modelo del ambiente de información en el que la red neuronal operará, esto es, se debe saber qué información estará disponible para la red. A esto se refiere comúnmente como paradigma de aprendizaje. Después, se debe entender cómo los pesos de la red se actualizan, es decir, qué reglas de aprendizaje gobiernan el proceso de actualización. Un algoritmo de aprendizaje es el procedimiento en el cual se basan las reglas de aprendizaje para ajustar los pesos.

En los sistemas adaptivo como los conexionistas, el aprendizaje descansa en la presentación iterativa al sistema de muestras o ejemplos tomados del correspondiente ambiente de información de interés. Dentro de esta perspectiva, la teoría de aprendizaje distingue claramente tres aspectos importantes en relación a las muestras tomadas del ambiente de información para entrenar al sistema:

- Capacidad de aprendizaje del sistema
- Complejidad de las muestras o ejemplos
- Complejidad computacional del aprendizaje

La capacidad de aprendizaje del sistema se refiere a cuántos patrones pueden ser almacenados por éste, y qué funciones y fronteras de decisión puede formar. La complejidad de la muestra determina el tamaño o número de patrones de entrenamiento que se necesitan para asegurar que

el sistema alcance un grado de generalización aceptable. Un conjunto de entrenamiento pequeño puede dar lugar a un entrenamiento deficiente por no ser representativo de la diversidad de los patrones, es decir, del ambiente de información. La complejidad computacional del aprendizaje se refiere al tiempo requerido para que un algoritmo de aprendizaje encuentre una solución aceptable. En general, los algoritmos de aprendizaje tienen una muy alta complejidad computacional. Diseñar algoritmos eficientes para el aprendizaje de las redes neuronales es un tema de investigación actual muy activo.

3.2.1. Paradigmas de aprendizaje

Los paradigmas de aprendizaje se pueden clasificar en tres tipos básicos:

- Aprendizaje supervisado
- Aprendizaje reforzado
- Aprendizaje no supervisado

3.2.1.1. Aprendizaje supervisado

El aprendizaje supervisado requiere de lo que se ha dado en llamar un “maestro” externo. Lo que esto significa es que para el entrenamiento se requiere coleccionar y presentar a la RNA una serie de ejemplos de entrada. Las salidas o predicciones de la RNA se comparará con el correspondiente valor deseado para determinar una medida de su discrepancia o diferencia, es decir, el error cometido por la RNA. En este caso, lo que representa el “maestro” es el conocimiento del *ambiente de información*, es decir, un conjunto de ejemplos *entrada-salida*. Los parámetros de la RNA se ajustarán a partir de la *señal de error*. Este ajuste será en forma iterativa con la mira de que la RNA logre emular óptimamente a su “maestro”. En este caso, el aprendizaje de la RNA equivale a la transferencia del conocimiento que el “maestro” tiene del *ambiente de información* a la RNA.

3.2.1.2. Aprendizaje reforzado

En el aprendizaje reforzado no se conocen las salidas deseadas para cada ejemplo de entrada. Lo único de lo que se dispone es de una “crítica” es decir, del conocimiento de si la salida es correcta o incorrecta. La RNA ha de actualizarse sólo en base esta información. El aprendizaje reforzado tiene sus orígenes en estudios experimentales de psicología acerca del aprendizaje en animales [33], aunque los psicólogos no usan el término aprendizaje reforzado. De hecho, el aprendizaje reforzado tiene como antecedente la *Ley de efecto* de Thorndike [210], la cual es expresada por Sutton y Barto [207, 32] como:

“Si una acción tomada por un sistema de aprendizaje es seguida por un estado de cosas satisfactorio o por una mejora del estado de cosas, entonces la tendencia a producir esa acción se refuerza. Por el contrario, si una acción es seguida por un estado de cosas insatisfactorio, entonces la tendencia a producir esa acción se debilita.”

Finalmente, se hace notar que, de los tres paradigmas de aprendizaje, el correspondiente al aprendizaje reforzado es sin lugar a duda el que tiene mayor plausibilidad biológica.

3.2.1.3. Aprendizaje no supervisado

El aprendizaje no supervisado o auto-organizado se emplea cuando se desconocen los valores precisos deseados correspondientes a los ejemplos que se le presentan a la RNA. Para esta situación se acostumbra pensar que no hay un “maestro” ni “crítica” disponible para la RNA, por lo cual ésta debe auto-organizarse de tal manera que pueda darle un sentido a los ejemplos que se le presenta como, por ejemplo, la formación de cúmulos de ejemplos. Para este paradigma de aprendizaje, se requiere de una medida de la calidad de la representación de la RNA de los ejemplos de entrada, la cual sea independiente de la tarea. Los parámetros libres de la RNA serán optimizados con respecto a esa medida. Una vez que la RNA se adapta a las regularidades estadísticas de los ejemplos de entrada, desarrolla la habilidad para formar representaciones internas para codificar las características de las entradas y, por consiguiente, para crear automáticamente clases nuevas [34].

3.2.2. Reglas de aprendizaje

El tipo o regla de aprendizaje está determinado por la manera en que se cambian los parámetros adaptables de la RNA. Se pueden distinguir cinco tipos básicos de reglas de aprendizaje:

- Aprendizaje por corrección de error
- Aprendizaje basado en la memoria
- Aprendizaje hebbiano
- Aprendizaje competitivo
- Aprendizaje de Boltzmann

3.2.2.1. Aprendizaje por corrección de error

En el aprendizaje por corrección de error, al tiempo n , para cada ejemplo o entrada k , se usa la *señal de error* $e_k(n)$, determinada como la diferencia entre la señal deseada $d_k(n)$ y la señal de salida $y_k(n)$ para, paso a paso, accionar un “mecanismo de control” con objeto de acercar la salida $y_k(n)$ a la respuesta deseada $d_k(n)$. La idea es la de minimizar la función de costo

$$J = E \left[\frac{1}{2} \sum_k e_k^2(n) \right] \quad (3.1)$$

donde E representa el valor de expectación estadístico, y la suma es sobre todas los valores de las neuronas de la capa de salida de la RNA. Dado que normalmente se desconocen las propiedades

estadísticas que subyacen al ambiente de información en la cual se embebe a la RNA, entonces la alternativa es minimizar el *valor instantáneo* de la suma de errores cuadrados:

$$\varepsilon(n) = \frac{1}{2} \sum_k e_k^2(n) \quad (3.2)$$

La minimización de la función de costo J , o en su caso, del valor instantáneo $\varepsilon(n)$ es respecto a los parámetros de la RNA, es decir, los pesos sinápticos y los umbrales de las neuronas. Este proceso de minimización de la función de costo J respecto a sus parámetros da lugar al método de descenso por el gradiente [42, 98, 173]. En realidad, fue en un trabajo de Bernard Widrow y de su estudiante de doctorado Marcian Edward Hoff [230] en donde se fundamentó por primera vez, esta manera de adaptar los pesos de una RNA. Esta regla de aprendizaje de Widrow se conoce como: algoritmo LMS (least square algorithm) o regla delta de Widrow-Hoff. Los autores desarrollaron esta regla para una RNA que consistía en un *combinador lineal adaptivo*, el cual, como su nombre lo sugiere, tiene por salida una combinación lineal de sus entradas. Esto se puede expresar como:

$$y = \vec{\omega} \cdot \vec{x} = \sum_{i=0}^n \omega_i x_i, \quad (3.3)$$

donde $x_0 = 1$ y ω_0 es un término bias o de polarización. Usando esta sencilla forma para las salidas de un *combinador lineal adaptivo*, junto con la idea de minimizar el error que comete la red, se obtiene la siguiente forma de la regla delta:

$$\vec{\omega}(k+1) = \vec{\omega}(k) + \eta [d(k) - y(k)] \vec{x}(k), \quad (3.4)$$

donde $k+1$ numera la etapa de actualización a partir de los valores de la etapa k , $\vec{\omega}$ es el vector de pesos sinápticos de la RNA, η es una constante positiva que determina la razón de convergencia del algoritmo al mínimo, es decir, la razón de aprendizaje, $d(k)$ es la salida deseada, $y(k)$ es la salida de la RNA y $x(k)$ es el ejemplo de entrada que se presenta a la RNA.

3.2.2.2. Aprendizaje basado en la memoria

En el aprendizaje basado en la memoria, se almacenan en memoria todas las experiencias pasadas de ejemplos correctos de entrada-salida. Esta información es usada junto con técnicas basadas en métodos locales de estimación no-paramétrica de la densidad de probabilidad que subyace al ambiente de información correspondiente. Ejemplos de estos modelos locales son: vecino más cercano, promedios pesados o regresión localmente pesada. Cada uno de estos modelos locales combina las salidas de los puntos cercanos a un punto de interés para estimar la salida correspondiente. Este tipo de aprendizaje resulta muy adecuado para tareas que involucren estudios de lenguajes naturales [57], de robótica [192] o de minería de datos [199].

En el aprendizaje basado en la memoria se distinguen dos elementos fundamentales:

- Una métrica o distancia
- Un modelo local

Hay diversas opciones que se pueden usar como modelo local:

- El número de vecinos cercanos
- Una función de ponderación (regresión de kernel)
- Regresiones pesadas localmente

Una métrica o distancia es el punto de interés para el cual se desea hacer una predicción acerca de sus correspondientes posibles atributos de salida se denomina *punto de indagación* (query point). Lo primero que se requiere hacer cuando se desea realizar una predicción usando aprendizaje basado en memoria es buscar, en la base de datos en memoria, los datos cuyos atributos de entrada sean similares a los del *punto de indagación*. Para poder decir cuando dos puntos son similares, se necesita contar con una métrica o función distancia que nos indique que tan cerca están esos dos puntos.

Un modelo local sencillo consiste en calcular la predicción para el *punto de indagación* como un promedio sencillo de los puntos cercano éste. En este modelo se ve natural recurrir a un promedio de los k vecinos más cercanos. El valor de k se escoge, por un lado, lo suficientemente grande como para eliminar lo más posible el efecto de ruido en los datos, pero por otro lado, no debe ser tan grande como para borrar, por la acción del promedio, la presencia de posibles estructuras de interés en los datos. Obviamente el valor de k o número de vecinos dependerá del problema en cuestión. Una función de ponderación (regresión de kernel) es cuando se usa el promedio sencillo involucrado en la técnica de los k vecinos más cercanos, se obtiene como aproximación una función escalón discontinua. Una alternativa más poderosa de modelo local que permita alisar la predicción es usar un promedio ponderado de los puntos. Esta técnica se conoce como regresión de kernel. En esta técnica, cada punto del conjunto de datos en memoria recibirá un peso entre 0 y 1, en base a que tan cerca esté del *punto de indagación*. Una función ponderadora muy usada es la función gaussiana. El peso ω_i correspondiente al punto x_i se calculan como:

$$\omega_i = \exp\left(\frac{-d(x_i, \text{punto de indagación})}{K_\omega}\right), \quad (3.5)$$

donde $d(x_i, \text{punto de indagación})$ es la distancia entre el punto x_i y el *punto de indagación*. K_ω se denomina anchura del kernel.

La predicción y de la salida para el *punto de indagación* queda expresada como:

$$y = \frac{\sum_i \omega_i y_i}{\sum_i \omega_i} \quad (3.6)$$

Las regresiones pesadas localmente se usan si hay problemas para los cuales las alternativas que brindan tanto el uso de un número de vecinos cercanos como regresión de kernel no son suficientes como modelos locales para ajustar los datos, por lo que se necesita un modelo local más sofisticado. Por ejemplo, si se considera conveniente usar un modelo lineal para ajustar localmente a los datos, entonces, en lugar de usar una regresión global lineal, es factible obtener un mejor ajuste si se ajusta el modelo lineal localmente.

3.2.2.3. Aprendizaje hebbiano

La regla de aprendizaje de Hebb [99] es la más antigua y más famosa de todas las reglas de aprendizaje; siendo la de mayor plausibilidad biológica. Esta regla, la cual es un ejemplo de aprendizaje reforzado, descansa en dos postulados hebbianos acerca de las sinapsis neuronales:

- Si las dos neuronas de los lados de una sinapsis son activadas simultáneamente, entonces la fortaleza de esa sinapsis aumenta selectivamente.
- Si las dos neuronas de los lados de una sinapsis son activadas asincrónamente, entonces esa sinapsis se debilitará o eliminará selectivamente.

Estas dos suposiciones permiten clasificar a las sinapsis hebbianas como aquellas que usan un mecanismo dependiente del tiempo, altamente local y fuertemente interactivo, para incrementar la eficiencia sináptica como función de la correlación entre las actividades presináptica y postsináptica.

La formulación matemática de la regla de aprendizaje hebbiano se puede expresar en términos de los pesos ω_{kj} y de las actividades presináptica y postsináptica x_j y y_k , respectivamente, de la siguiente manera:

$$\Delta\omega_{kj}(n) = F(y_k(n), x_{jk}(n)), \quad (3.7)$$

donde $F(\cdot, \cdot)$ es una función tanto de la actividad presináptica como de la postsináptica. De acuerdo a esta formulación, la forma más sencilla para la actualización de los pesos sinápticos sería:

$$\Delta\omega_{kj}(n) = \eta y_k(n) x_{jk}(n), \quad (3.8)$$

donde η es una constante positiva que determina la razón de aprendizaje. La ecuación 3.8 suele denominarse como regla de producto de actividades.

Para evitar que, debido a la correlación positiva de las señales presináptica y postsináptica, se dé un crecimiento exponencial que llegara a saturar al peso sináptico ω_{kj} , se acostumbra introducir un factor no lineal de olvido, o sea:

$$\Delta\omega_{kj}(n) = \eta y_k(n) x_{jk}(n) - \alpha y_k(n) \omega_{kj}(n) \quad (3.9)$$

donde α es una constante positiva. La ecuación 3.9 suele denominarse *regla generalizada de producto de actividades*.

3.2.2.4. Aprendizaje competitivo

En el aprendizaje competitivo, las neuronas de la capa de salida de una RNA compiten entre sí para llegar a ser la neurona activa. De acuerdo a esta regla, a un tiempo dado, sólo una neurona de la capa de salida puede estar activa. Esto hace al aprendizaje competitivo apto para descubrir las características estadísticas relevantes para la clasificación de un conjunto de patrones de entrada. Los orígenes del aprendizaje competitivo se remontan a los trabajos de [87, 88, 89, 137, 78, 233].

El algoritmo de aprendizaje competitivo se puede expresar en términos del conjunto normalizado de ejemplos de entrada $X = \{x_1, x_2, \dots, x_l\}$ perteneciente a un espacio n -dimensional, el cual se desea clasificar en k cúmulos diferentes. La RNA consistirá de una capa de entrada de n unidades con umbral cero y una capa de salida de k unidades. El algoritmo será:

Comienzo:

Se generan al azar los vectores de pesos $\omega_1, \dots, \omega_k$, de componentes positivas y normalizados, es decir,

$$\sum_i \omega_{ji} = 1 \text{ para toda } j. \quad (3.10)$$

Prueba:

Se selecciona al azar un vector de ejemplo $x_j \in X$

Se determina cuál es la *neurona ganadora toma todo* (digamos que la m -ésima).

Actualización:

Se sustituye ω_{mi} por $\omega_{mi+\eta(x_i-\omega_{mi})}$ para toda i , y se normaliza ω_{mi} .

Continuar con prueba.

El efecto de esta regla de aprendizaje es mover el vector del peso sináptico ganador, ω_m , hacia el patrón de entrada x_j .

Si m es la *neurona ganadora toma todo*, entonces su nivel de actividad interna ν_m para un patrón dado de entrada x_j debe ser el más grande de entre todas las neuronas de salida de la RNA. En ese caso, la salida de la *neurona ganadora toma todo* m , y_m , se define como igual a uno; las salidas de las demás neuronas que pierden la competencia se definen como iguales a cero.

3.2.2.5. Aprendizaje de Boltzmann

La regla de aprendizaje de Boltzmann, es un algoritmo de aprendizaje estocástico [3, 111], el cual fue nombrado en honor a Ludwig Edward Boltzmann. La razón de esto se debe a la equivalencia formal entre el trabajo original de Boltzmann en termodinámica estadística y el comportamiento dinámico de una máquina de Boltzmann, que es como se ha dado en llamar a este tipo de RNA que usan la regla de aprendizaje de Boltzmann. La máquina de Boltzmann, lo mismo que la red de Hopfield, comparten las siguientes características:

- Sus unidades procesadoras asumen estados con valores binarios ($s = \pm 1$).
- Todas las conexiones sinápticas entre sus unidades son simétricas.
- Para la actualización de las unidades éstas se seleccionan una por una y al azar.

- Las unidades no tienen auto-retroalimentación.

Por otro lado difieren en:

- A diferencia de la red de Hopfield, la máquina de Boltzmann permite el uso de *neuronas ocultas*.
- La máquina de Boltzmann usa *neuronas estocásticas* junto con un mecanismo probabilístico de disparo. Por su parte, la red de Hopfield usa neuronas basadas en el modelo de McCulloch y Pitts junto con un mecanismo determinista de disparo.
- La red de Hopfield opera de una manera no supervisada, mientras que la máquina de Boltzmann, además de esa opción, también puede ser entrenada por supervisión de una forma probabilística.

La manera en que una máquina de Boltzmann opera es minimizando una función de energía, E , expresable en términos los estados binarios de las neuronas y de los pesos sinápticos:

$$E = -\frac{1}{2} \sum_{i \neq j} \sum_j \omega_{ji} s_j s_i, \quad (3.11)$$

donde la suma doble corre sobre todas las neurona de la red: entrada, salida y ocultas. El umbral θ_j de la j -ésima neurona se introducen definiendo alguno de los pesos sinápticos igual a θ_j y alimentando ese peso particular usando una entrada fija e igual a -1 .

Para esta minimización, a lo largo del proceso de aprendizaje, se selecciona al azar una neurona (la j -ésima), a la cual, durante el proceso de *recocido* (concepto tomado del símil termodinámico de *recocido*) se le cambiará su estado de s_j a $-s_j$ con una probabilidad dada por:

$$Prob(s_j \rightarrow -s_j) = \frac{1}{1 + \exp\left(-\frac{\Delta E_j}{T}\right)}, \quad (3.12)$$

donde T es la “temperatura” durante el ciclo de *recocido* y ΔE_j es el cambio de energía producto del cambio de estado. Si se itera la aplicación de esta regla probabilística a las neuronas, la red alcanzará un equilibrio termodinámico. En equilibrio, los estados de las neuronas seguirán una *distribución de Boltzmann*.

3.3. Arquitecturas de redes neuronales artificiales

No obstante que con el paso del tiempo, el desarrollo en RNAs ha sido considerable, actualmente, el problema del diseño de una arquitectura adecuada de una RNA para resolver algún problema es un problema abierto. Existen diferentes alternativas para la arquitectura o topología de una RNA. En una primera instancia, estas diferentes alternativas para las arquitecturas de las RNAs pueden ser clasificadas en dos grupos:

- Redes de alimentación hacia adelante
- Redes recurrentes

La diferencia esencial entre estas dos clases de arquitecturas de RNAs estriba en que, en las redes de alimentación hacia adelante, las interconexiones de las neuronas de la red no forman ciclos, lo cual si está permitido en las redes recurrentes. Esta posibilidad, en las redes recurrentes, de formar de ciclos, les permite presentar un comportamiento dinámico temporal [67]. En las figuras 3.1 y 3.2 se muestran esquemas de una RNA de alimentación hacia adelante y de otra recurrente, respectivamente.

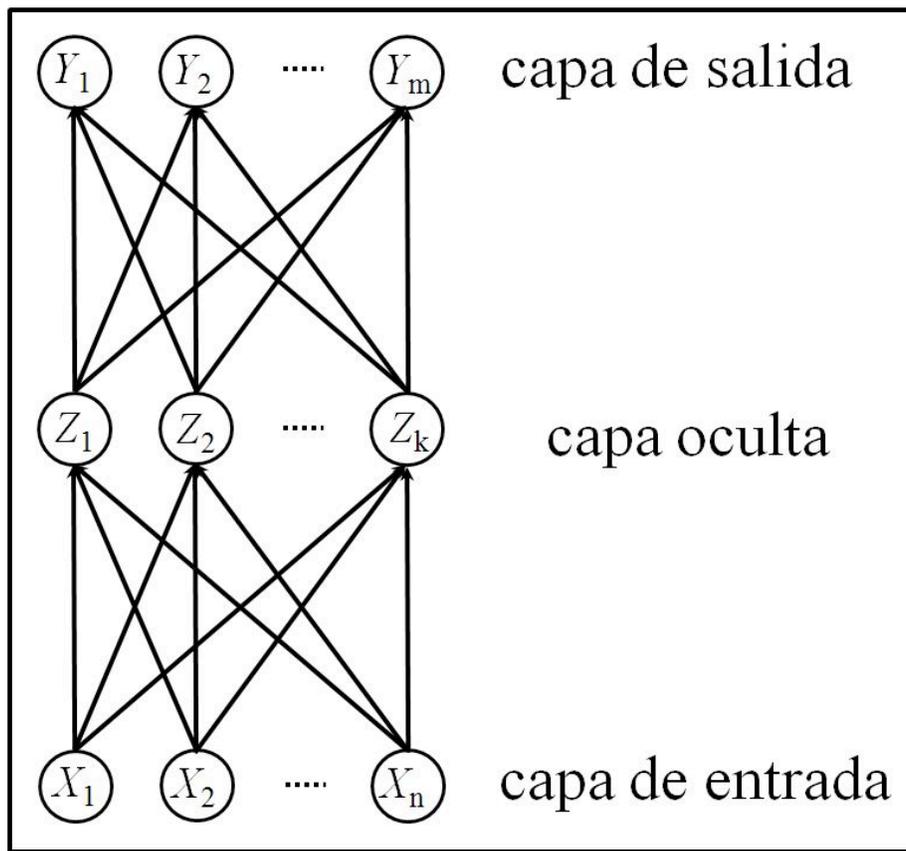


Figura 3.1: Red neuronal de alimentación hacia adelante. Se muestra una red que consta de capa de entrada, capa de salida y, en este caso, de una sola capa oculta. Como se puede apreciar en la figura, las conexiones no dan lugar a ciclos.

En vista de que en este trabajo se hará uso de una red de alimentación hacia adelante, ya no se abundará más acerca de las redes recurrentes.

De acuerdo a Bishop [42], una RNA de alimentación hacia adelante se puede definir como:

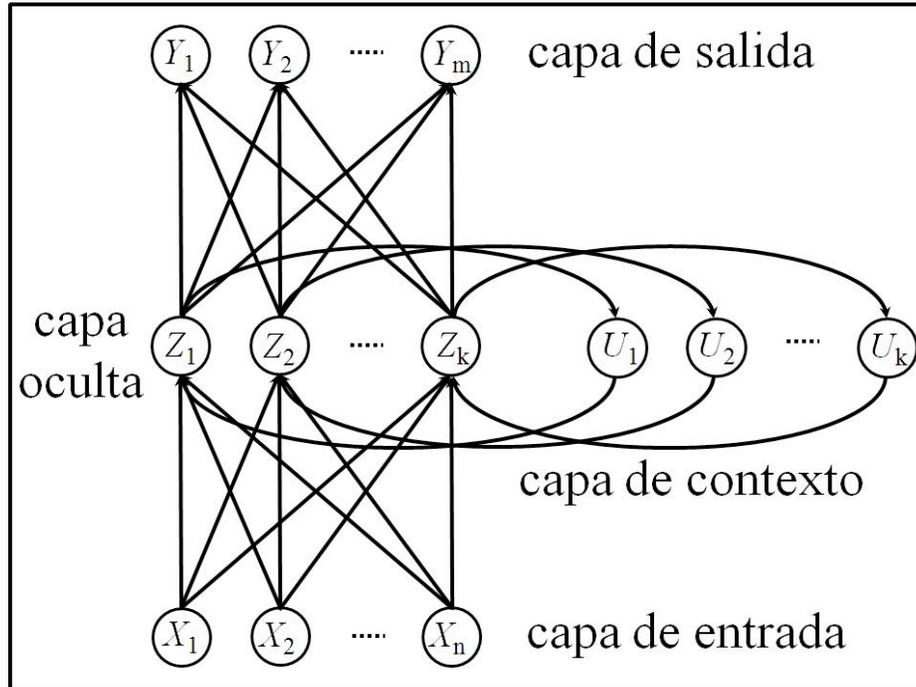


Figura 3.2: La red neuronal de Elman es una sencilla red neuronal recurrente desarrollada por Jeffrey L. Elman en 1990 [67]. Esta red consiste de una capa de entrada, una capa oculta, una capa de contexto y una capa de salida. La capa de contexto es alimentada, sin pesos, con la salida de la capa oculta. De esta manera, la red de Elman recuerda o guarda estos valores para la siguiente evaluación que realice la red neuronal. Estos valores son enviados, usando conexiones de pesos entrenables, de regreso a la capa oculta. Esto hace que las redes de Elman sean muy útiles en la predicción de secuencias, dado que, la capa de contexto las dota de una memoria de corto plazo.

“Una red en la cual es posible asignar números sucesivos a todas las neuronas de entrada lo mismo que a todas las neuronas ocultas y de salida de tal manera que cada neurona sólo recibe conexiones de neuronas de entrada o de neuronas que tienen un número menor.”

Las redes de alimentación hacia adelante tienen la propiedad de que sus salidas pueden ser expresadas como funciones deterministas de las entradas, de tal suerte que la red representa un mapeo funcional no lineal de varias variables.

Como se puede ver en la figura 3.1, las redes de alimentación hacia adelante constan de una capa de entrada, una capa de salida y de un cierto número de capas ocultas, el cual dependerá de la complejidad del problema que se considera. Cada capa consiste de una o más neuronas. Las flechas entre las neuronas indican la dirección del flujo de información de una neurona a otra. Las neuronas de la capa de entrada son pasivas, es decir, no alteran la información. Cada una de ellas recibe una entrada, la cual es distribuida tal cual por cada una de sus salidas. A diferencia, las

neuronas de las capas ocultas y de salida son activas, es decir, éstas son capaces de modificar los datos (Fig. 3.3).

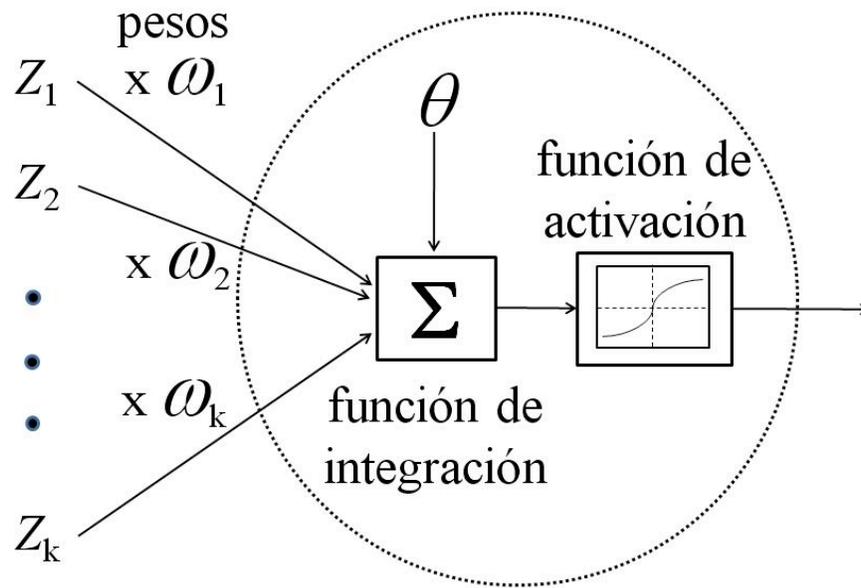


Figura 3.3: Este es un diagrama de flujo de un nodo activo de una capa oculta o de la capa de salida de la red neuronal. Como se indica, cada entrada (Z_1, Z_2, \dots, Z_k) es ponderada por un peso ($\omega_1, \omega_2, \dots, \omega_k$). Posteriormente, estas señales ponderadas se integran (se suman) y se comparan con un umbral θ . Finalmente, esta integración se pasa a través de una función de activación (por ejemplo, una función sigmoide), la cual determina la salida de la neurona. En caso de tratarse de una neurona de una capa oculta de una red de alimentación hacia adelante, esta salida es alimentada a las neuronas de la siguiente capa.

Cuando, en una red de alimentación hacia adelante, las salidas de las neuronas de una capa son alimentadas a cada una de las neuronas de la capa siguiente, se dice que la estructura de la red es completamente conectada. Ese es el caso ejemplificado en la figura 3.1. En este caso, las neuronas de las capas diferentes a la de entrada reciben las salidas de cada una de las neuronas de la capa anterior (Z_1, Z_2, \dots, Z_k), las cuales son ponderadas usando pesos ($\omega_1, \omega_2, \dots, \omega_k$), (Fig. 3.3). Estos productos son integrados o sumados y la cantidad obtenida es comparada con un umbral θ . En la figura 3.3, estos se denotan como Σ . Finalmente, para producir la salida o activación de la neurona, esta cantidad es pasada a través de una función no lineal denominada sigmoide. La tarea de esta función sigmoide es la de que para entradas a la neurona entre $-\infty$ y $+\infty$, la correspondiente salida está limitada a tomar un valor entre 0 y 1. Lo anterior se puede resumir expresando la salida

o activación de la neurona como:

$$Z_j^L = g(a_j^L), \quad (3.13)$$

donde

$$a_j^L = \sum_{i=0}^k \omega_{ji}^{L-1} z_i^{L-1} \quad (3.14)$$

La función $g(\cdot)$ representa a una función de activación sigmoide. La suma ponderada que realiza la neurona está denotada como a_j^L , la correspondiente salida o activación se denota como Z_j^L , los pesos ω_{ji}^{L-1} para $i = 1, \dots, k$, corresponden a las conexiones entre las neuronas $i = 1, \dots, k$ de la capa anterior a la neurona j a la cual se le calcula su activación. De la misma manera, las activaciones z_i^{L-1} para $i = 1, \dots, k$, corresponden a las neuronas $i = 1, \dots, k$ de la capa anterior. El valor del peso sináptico ω_{j0}^{L-1} para el cual se define la entrada fija $z_0^{L-1} = -1$, es igual al umbral θ de la neurona j . De esta manera, la comparación de la suma ponderada con el umbral queda incorporada en la suma sobre i desde 0 a k . Conviene notar que en este trabajo se sigue la convención de considerar capas de pesos más que de neuronas, es decir, el índice L designa la L -ésima capa de pesos y no la L -ésima capa de neuronas. También conviene aclarar, que en este trabajo las salidas de las neuronas de la capa de entrada se denotarán como (X_1, \dots, X_n) , las de las neuronas de las capas ocultas como (Z_1^L, \dots, Z_k^L) , y las de la capa de salida como (Y_1, \dots, Y_m) .

3.4. El perceptrón de multicapa y retro-propagación del error

El perceptrón de multicapa (PMC) es una de las arquitecturas de RNAs más importantes pues ha demostrado su utilidad para resolver problemas que involucran a muy diversas disciplinas. En particular, esto resulta muy evidente cuando el PMC se combina con una ley de entrenamiento supervisado conocida como algoritmo de retro-propagación del error. De esta manera, se obtiene una ley de actualización del error que se puede considerar como una generalización de la regla delta de Widrow-Hoff. Una manera de expresar la operación de procesamiento de la información que un PMC entrenado por retro-propagación del error debe realizar es [100]: “La aproximación de un mapeo acotado o función $f : A \subset \mathbf{R}^n \rightarrow \mathbf{R}^m$, de un subconjunto compacto A de un espacio euclideo n -dimensional a un subconjunto acotado $f(A)$ de un subespacio euclideo m -dimensional, mediante el entrenamiento por ejemplos $(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k), \dots$ correspondientes al mapeo, donde $y_k = f(x_k)$.”

En la literatura acerca de la neurocomputación, el término retro-propagación se ha usado frecuentemente de una manera algo vaga, pero que puede resultar conveniente o cómoda. En ocasiones, a la arquitectura de un PMC suele denominársele como red de retro-propagación. El término también suele usarse para denominar el entrenamiento de un PMC por aprendizaje por corrección de error usando descenso por el gradiente.

La mayoría de los algoritmos de entrenamiento de una RNA involucran un procedimiento iterativo para minimizar la función de error, mediante el ajuste por pasos de los pesos sinápticos

de la RNA. La forma de llevar a cabo esta tarea usando retro-propagación consiste en dos pasadas a través de las capas del PMC: un paso hacia delante y un paso hacia atrás. En el *paso hacia adelante*, un patrón de actividades (el ejemplo de entrada) es aplicado a las neuronas de la capa de entrada del PMC, propagando su efecto a través de la red. Al final, como respuesta de la red, se obtendrá una salida. Durante este paso hacia delante, los pesos sinápticos de la red permanecen fijos. Por el contrario, durante el *paso hacia atrás*, todos los pesos sinápticos son ajustados conforme a una regla de corrección de error. Para comenzar este paso hacia atrás, se calcula la *señal de error* definida como la diferencias entre la señal deseada menos la salida real que proporciona la red. Esta señal de error será propagada hacia atrás a través de la red, de donde se origina el nombre de retro-propagación del error. Los pesos sinápticos de la red serán modificados con objeto de acercar la salida de la red al valor deseado. Rumelhart *et al.* [178] propuso una alternativa para este propósito basada en el descenso por el gradiente.

3.4.1. Algoritmo de retro-propagación del error

El algoritmo de retro-propagación se deriva para una red general teniendo una topología de alimentación hacia adelante arbitraria, funciones de activación diferenciables no lineales arbitrarias y una función de error diferenciable arbitraria.

En una red general de alimentación hacia adelante (Fig. 3.4), cada unidad calcula una suma ponderada de sus entradas de la forma:

$$a_j^L = \sum_i \omega_{ji}^{L-1} z_i^{L-1} \quad (3.15)$$

donde z_i^{L-1} es la activación de una unidad, o entrada, la cual manda una conexión a la unidad j , y ω_{ji}^{L-1} es el peso asociado a esa conexión. La suma corre sobre todas las unidades que tienen conexión con la unidad j . Los umbrales se pueden incluir en esta suma introduciendo una unidad extra, o entrada, con la activación fija en $+1$. Por lo tanto, no se necesita lidiar con los umbrales explícitamente. La suma ponderada a_j^L se transforma con una función de activación no lineal $g(\cdot)$, de tal manera que la activación z_j^L de la unidad j se expresa como:

$$z_j^L = g(a_j^L). \quad (3.16)$$

Nótese que las variables z_i^{L-1} en la ecuación 3.15 pueden ser entrada a la red, en cuyo caso se denotarán por x_i . De manera similar, la unidad j en la ecuación 3.16 podría ser una unidad de salida, en cuyo caso se denotará la activación por y_i .

Como anteriormente, se busca determinar los valores adaptables de los pesos de la red minimizando la apropiada función de error. Aquí, se consideran las funciones de error que pueden ser escritas como la suma, sobre todos los patrones en el conjunto de entrenamiento, de errores definidos para cada patrón por separado, o sea:

$$E = \sum_n E^n \quad (3.17)$$

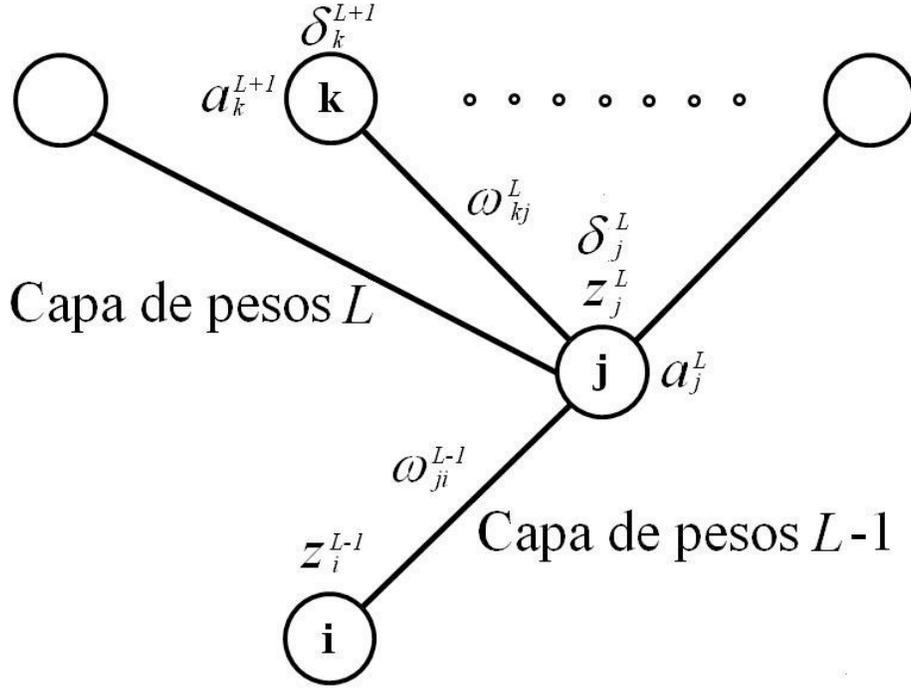


Figura 3.4: Esquema de las entradas y salidas de una neurona correspondiente a una arquitectura de alimentación hacia adelante. L designa el número de capa (capa de pesos), ω_{ji}^{L-1} el peso que conecta la neurona i con la neurona j , z_i^{L-1} es una de las entradas de la neurona j , a_j^L es la suma ponderada de las entradas para la neurona j , z_j^L es la salida de la neurona j . Finalmente, δ_j^L es el error correspondiente a la neurona j .

donde n es la etiqueta de los patrones. Casi todas las funciones de error de interés práctico toman esta forma. Se supone que el error E^n se puede expresar como una función diferenciable de las variables de salida de la red de tal manera que

$$E^n = E^n(y_1, \dots, y_c). \quad (3.18)$$

Dado que cada una de las salidas y_i es función de todos los pesos y los umbrales sinápticos representados por el vector $\vec{\omega}$, entonces la ecuación 3.18 se puede reescribir como:

$$E^n = E^n(\vec{\omega}). \quad (3.19)$$

De la misma manera, la ecuación 3.17 se puede reescribir como:

$$E = E(\vec{\omega}) = \sum_n E^n(\vec{\omega}) \quad (3.20)$$

Con esto se ve que el error $E(\vec{\omega})$, contemplado como una función del espacio de los pesos y los umbrales sinápticos a los reales, define una superficie, la *superficie de error*. Dentro de este

contexto, la minimización del error corresponde a encontrar el mínimo de la superficie de error $E(\vec{\omega})$, para lo cual, como es bien sabido, se requiere desarrollar un procedimiento para encontrar el punto donde se anulan las derivadas de la función de error $E(\vec{\omega})$ con respecto a los pesos y los umbrales de la red. Utilizando la ecuación 3.20 se pueden expresar estas derivadas como sumas sobre los patrones del conjunto de entrenamiento de las derivadas correspondientes a cada patrón por separado. De aquí en adelante se considerará un patrón a la vez.

Para cada patrón se provee a la red del vector de entrada correspondiente y se calculan las activaciones de todas las unidades ocultas y de salida de la red usando las ecuaciones 3.15 y 3.16. Este proceso se conoce como *propagación o paso hacia adelante* ya que se puede considerar como un flujo de información hacia adelante a través de la red.

Después se evalúa la derivada de E^n con respecto a un peso ω_{ji}^{L-1} . Las salidas de las unidades dependerán del patrón de entrada particular n . Sin embargo, para mantener la notación, se omite el índice n de las variables de entrada y de activación. Nótese que E^n depende del peso ω_{ji}^{L-1} sólo a través de la entrada a_j^L a la unidad j . Por esto, se aplica la regla de la cadena para derivadas parciales para obtener

$$\frac{\partial E^n}{\partial \omega_{ji}^{L-1}} = \frac{\partial E^n}{\partial a_j^L} \frac{\partial a_j^L}{\partial \omega_{ji}^{L-1}} \quad (3.21)$$

Ahora se introducirá una notación útil

$$\delta_j^L = \frac{\partial E^n}{\partial a_j^L} \quad (3.22)$$

donde δ_j^L se asocia a los errores. Utilizando la ecuación 3.15 se ve que

$$\frac{\partial a_j^L}{\partial \omega_{ji}^{L-1}} = z_i^{L-1} \quad (3.23)$$

Sustituyendo las ecuaciones 3.22 y 3.23 en la 3.21 se obtiene

$$\frac{\partial E^n}{\partial \omega_{ji}^{L-1}} = \delta_j^L z_i^{L-1}. \quad (3.24)$$

Este resultado hace notar que la derivada requerida se obtiene simplemente multiplicando el valor δ_j^L de la unidad correspondiente al extremo de la salida del peso ω_{ji}^{L-1} por el valor de z_i^{L-1} correspondiente al extremo de entrada del peso ω_{ji}^{L-1} (donde $z = 1$ en el caso del umbral). Así, para evaluar las derivadas, necesitamos únicamente calcular el valor de δ_j^L para cada unidad oculta y de salida en la red y luego aplicar la ecuación 3.24.

Para las unidades de salida la evaluación de δ_k^{out} es directa. De la definición (Ec. 3.22) se tiene

$$\delta_k^{out} = \frac{\partial E^n}{\partial a_k^{out}} = g'(a_k^{out}) \frac{\partial E^n}{\partial y_k} \quad (3.25)$$

donde se usó la ecuación 3.16 con z_k^{out} denotada por y_k . Para evaluar la ecuación 3.25 se sustituyen las expresiones apropiadas de $g'(a_k^{out})$ y de $\partial E^n / \partial y_k$.

Para evaluar δ_j^L para las unidades ocultas se hace uso nuevamente de la regla de la cadena

$$\delta_j^L = \frac{\partial E^n}{\partial a_j^L} = \sum_k \frac{\partial E^n}{\partial a_k^{L+1}} \frac{\partial a_k^{L+1}}{\partial a_j^L} \quad (3.26)$$

donde la suma corre sobre todas las unidades k con las que la unidad j tiene conexión. Al escribir la ecuación 3.26 se hace uso del hecho que las variaciones en a_j^L dan lugar a variaciones en la función de error sólo a través de variaciones en las variables a_k^{L+1} . Si se sustituye la δ_k^{L+1} de acuerdo a la definición dada por la ecuación 3.22 en la ecuación 3.26 y haciendo uso de las ecuaciones 3.15 y 3.16, se obtiene la siguiente fórmula de retro-propagación del error

$$\delta_j^L = g'(a_j^L) \sum_k \omega_{kj}^L \delta_k^{L+1} \quad (3.27)$$

Este último resultado indica que el valor de δ_j^L para una unidad oculta particular se puede obtener propagando hacia atrás las δ_k^{L+1} de las unidades de la capa inmediatamente arriba en la red (Fig. 3.4). Ya que se conocen los valores δ_k^{out} de las unidades de salida, aplicando la ecuación 3.27, se pueden evaluar las δ 's de todas las unidades ocultas en una red de alimentación hacia adelante, sin importar su topología.

Resumiendo, el procedimiento de retro-propagación para evaluar las derivadas del error E^n con respecto a los pesos consiste en cuatro pasos:

1. Se presenta un vector de entrada \mathbf{x}^n a la red y se propaga hacia adelante a través de la red utilizando las ecuaciones 3.15 y 3.16 para encontrar las activaciones de todas las unidades ocultas y de salida.
2. Se evalúa δ_k^{out} para las unidades de salida usando la ecuación 3.25.
3. Se retro-propagan las δ 's usando la ecuación 3.27 para obtener δ_j^L para cada unidad oculta en la red.
4. Se evalúan las derivadas requeridas usando la ecuación 3.24.

La derivada del error total E^T se puede obtener repitiendo los pasos anteriores para cada patrón del conjunto de entrenamiento, y luego sumarlo sobre todos los patrones, es decir:

$$\frac{\partial E^T}{\partial \omega_{ji}^L} = \sum_n \frac{\partial E^n}{\partial \omega_{ji}^L}. \quad (3.28)$$

En la derivación de arriba se asumió que cada unidad oculta o de salida en la red tiene la misma función de activación $g(\cdot)$. La derivación se generaliza fácilmente para permitir que diferentes unidades tengan diferentes funciones de activación, simplemente conociendo qué clase de $g(\cdot)$ corresponde a cada unidad.

Cuando se usa entrenamiento por retro-propagación del error, es muy común usar el método de *entrenamiento incremental*. En este método, la actualización o adaptación de los pesos sinápticos se realiza después de presentar cada ejemplo o patrón a la red, siguiendo la dirección del gradiente negativo. De tal suerte que los pesos sinápticos cambian de la siguiente manera:

$$\Delta\omega_{ji}^L(\text{incremental}) = -\eta \frac{\partial E^n}{\partial \omega_{ji}^L} = -\eta \delta_j^{L+1} z_i^L \quad (3.29)$$

donde $\eta > 0$ es la razón de aprendizaje, la cual sirve para determinar el tamaño del paso en la dirección del gradiente negativo.

Otra alternativa corresponde al *aprendizaje por lotes*. En éste, se actualizan los pesos después de presentar un cierto número de patrones y de sumar las derivadas correspondientes. Para escribir la forma del cambio de los pesos para el caso de aprendizaje por lotes, denotaremos las cantidades correspondientes a un patrón o ejemplo n como: $\delta_j^{L+1}(n)$ y $z_i^L(n)$. De tal suerte que:

$$\Delta\omega_{ji}^L(\text{por lotes}) = -\eta \sum_{n=n_1}^{n=n_{lote}} \frac{\partial E^n}{\partial \omega_{ji}^L} = -\eta \sum_{n=n_1}^{n=n_{lote}} \delta_j^{L+1}(n) z_i^L(n), \quad (3.30)$$

donde n_{lote} denota el tamaño del lote. De acuerdo a Warren S. Sarle [191] el aprendizaje por lotes corresponde al caso en que n_{lote} es igual al tamaño del conjunto de entrenamiento. El caso en que n_{lote} es menor, lo denomina *aprendizaje por mini-lote*.

Comparando la ecuación 3.29 o la 3.30 con la 3.18 se ve que la corrección por retro-propagación del error puede ser considerada como una generalización de la regla delta de Widrow-Hoff.

Normalmente, cuando se usa aprendizaje por mini-lotes, el valor de la razón de aprendizaje η se acostumbra escalar dividiéndolo por el tamaño del mini-lote, es decir, $\eta = \eta_0/n_{lote}$. En donde el valor del parámetro η_0 sería comparable al valor de la razón de aprendizaje si se usara entrenamiento incremental. De tal manera que si se rescribe la ecuación 3.30 en términos de η_0 se obtiene:

$$\begin{aligned} \Delta\omega_{ji}^L(\text{por lotes}) &= -\eta \sum_{n=n_1}^{n=n_{lote}} \frac{\partial E^n}{\partial \omega_{ji}^L} = -\frac{\eta_0}{n_{lote}} \sum_{n=n_1}^{n=n_{lote}} \frac{\partial E^n}{\partial \omega_{ji}^L} \\ &= -\eta_0 \frac{1}{n_{lote}} \sum_{n=n_1}^{n=n_{lote}} \frac{\partial E^n}{\partial \omega_{ji}^L} = -\eta_0 \frac{1}{n_{lote}} \sum_{n=n_1}^{n=n_{lote}} \delta_j^{L+1}(n) z_i^L(n), \end{aligned} \quad (3.31)$$

En esta última ecuación, se ve claramente que el aprendizaje por mini-lotes actualiza los pesos sinápticos usando una estimación del gradiente que es el valor promedio sobre todo el mini-lote, por lo cual es de esperarse que sea una mejor estimación del valor real del gradiente. Sin embargo, también es claro que este método será más lento pues requiere de un mayor esfuerzo computacional.

Capítulo 4

Análisis digital de curvas de Bragg

Como ya se dijo anteriormente, el objetivo de este trabajo es el de explorar nuevas alternativas neurocomputacionales para asistir el ADFP aplicado a la ECB. De manera muy concreta, se desea comparar la calidad y características de la identificación de CBs, contempladas éstas como patrones a los cuales, usando una RNA, se les desea identificar mediante la determinación, por un lado, de la energía total del ion que las produce, cantidad determinada por el área bajo la CB correspondiente y, por otro lado, del número atómico del ion, el cual está determinado por la amplitud del pico de Bragg.

Un esfuerzo en esta dirección ha sido emprendido por Vega et al. [214, 215, 216, 217, 218, 219], de hecho este trabajo es una continuación de ese esfuerzo. En los trabajos citados, las RNAs se adaptaron mediante un método de *aprendizaje supervisado* basado en la *retro-propagación del error*, usando el método de *entrenamiento incremental* (Ec. 3.29). En este trabajo, la adaptación de las RNAs seguirá otra alternativa, es decir retro-propagación, pero usando *entrenamiento por mini-lotes* (Ec. 3.30). La elección entre entrenamiento incremental o entrenamiento por lotes ha sido de interés cuando se quiere resolver un problema empleando una RNA [130], en esencia, este problema es el de seleccionar la mejor estrategia para el proceso de minimización de la función de error. También desde un punto de vista teórico, el problema del aprendizaje incremental y del aprendizaje por lotes es de gran interés [46, 107, 185, 187, 222, 228, 235, 236, 240].

En términos de la mecánica estadística, el aprendizaje ha sido formulado ya sea como el problema de un sistema en equilibrio termodinámico o como un sistema fuera del equilibrio, lo cual dependerá de la estrategia o método de aprendizaje [40, 167, 185, 186, 196, 222, 228]. Se pueden distinguir diversos regímenes intermedios, pero los extremos se representan por el método de *entrenamiento incremental* (proceso fuera del equilibrio) y el de *entrenamiento por lotes* (formulación en equilibrio). En el método por lotes se parte de construir una *función de error* (Ecs. 3.19 y 3.20) o *función de energía*, la cual depende de todos los datos de entrenamiento (en la ecuación 3.19 se suma sobre todos los patrones) por lo cual este es un proceso de mapeo iterativo determinista. El aprendizaje se basa en un proceso de descenso por el gradiente de la superficie de error en el espacio de parámetros (los pesos sinápticos y los umbrales de las neuronas), proceso que está sujeto al efecto del ruido (térmico), es decir, el ruido asociado al proceso de medición, el cual, en cierta

medida le permite escapar de mínimos locales. Después de cierto tiempo, estos modelos garantizan el alcance de un “*equilibrio térmico*”, para cuando esencialmente toda la información posible del conjunto de entrenamiento ha sido extraída por el algoritmo de aprendizaje. El sistema estará descrito por una distribución de probabilidad (Gibbs) definida sobre el espacio de parámetros.

Como se dijo arriba, en el otro extremo se encuentra el caso del aprendizaje incremental. En este caso, la *función de energía* o *función de error* depende de sólo un único ejemplo o patrón (Ecs. 3.18 y 3.19), el cual se escoge al azar para cada paso del aprendizaje dinámico, por lo cual este proceso es estocástico (cadena de Markov) [12, 138]; de ahí que se suele hablar de *aprendizaje estocástico*. El aprendizaje en línea también se basa en descenso por el gradiente, pero ahora, la presentación de patrones es al azar, lo que implica que cada paso de aprendizaje se usa una función de error diferente. Es posible ver que esta alternativa da lugar a un buen desempeño. En particular, para sistemas en los que el ambiente de información no es estacionario, el entrenamiento en línea resulta muy adecuado, el cual puede ser modelado usando ideas de mecánica estadística [38, 39, 121]. En relación con todo este esfuerzo teórico por entender el proceso de aprendizaje en las redes neuronales, podemos decir que los resultados aportados por los análisis de problemas complejos, como los abordados en este trabajo, resultan muy valiosos para validar los modelos teóricos desarrollados.

4.1. Arquitectura de la red neuronal

El objetivo de este trabajo es hacer una comparación del proceso de aprendizaje basado en el método de *entrenamiento incremental* [219] con el proceso de aprendizaje basado en el método de *entrenamiento por mini-lotes*. En ambos casos, el aprendizaje es por retro-propagación del error. Esto determina que la arquitectura de la RNA será la usada en [219]. La RNA usada será de alimentación hacia adelante. El tamaño de la capa de entrada queda determinado por el tamaño más grande de las digitalizaciones de las CBs analizadas, 81 canales, es decir, la capa de entrada constará de 81 neuronas, las cuales no son neuronas activas, su única función es la de alimentar cada uno de los valores de entrada a cada una de las neuronas de la primera capa oculta de la RNA. Por su lado, la capa de salida constará de dos neuronas: una proporcionará una estimación de la amplitud del PB de la correspondiente CB, la otra neurona de salida proporcionará una estimación de la energía del ion correspondiente a la CB analizada, es decir, una estimación del área bajo la CB. El número de capas ocultas usadas será el mismo que en [219], es decir, 5. Cada una de estas capas ocultas constará de nueve neuronas. Hacemos notar que, en [219], tanto el número de capas ocultas como el número de neuronas de estas capas se determinó en base a un proceso de prueba y error. La función de activación (Ec. 3.13) que se usará para todas las neuronas será una función logística:

$$g(a) = \frac{1}{1 + \exp(a)} \quad (4.1)$$

Más explícitamente, usando la ecuación 3.14, se tiene que:

$$g(a_j^L) = \frac{1}{1 + \exp(\sum_{i=0}^k \omega_{ji}^{L-1} z_i^{L-1})} = \frac{1}{1 + \exp(\sum_{i=1}^k \omega_{ji}^{L-1} z_i^{L-1} - \theta_j^{L-1})} \quad (4.2)$$

Los valores iniciales de los pesos sinápticos lo mismo que los valores de los umbrales de las neuronas de la red usada se seleccionaron uniformemente al azar en intervalos de la forma $[-b, b]$. Para el parámetro b se estudió el efecto de dos valores: 0.5 y 0.03. La red neuronal usada será completamente conectada, esto es, cada neurona de cada capa tiene una conexión sináptica con cada neurona de la próxima capa.

4.2. Ley de aprendizaje

El aprendizaje se llevará a cabo usando el algoritmo de *retro-propagación del error*. Se usará el método de *entrenamiento por mini-lotes* (Ec. 3.30). Se estudiarán dos tamaños para los mini-lotes: 16 y 100 patrones. La implementación de software que se usará es la del paquete SNNS (Stuttgart Neural Network Simulator) versión 4.2 [241]. A la razón de aprendizaje η que aparece en la ecuación 3.30 se le asignará un valor inicial igual a 0.3. La mayoría de las corridas se realizarán usando el valor de 0.3 durante todo el entrenamiento, sin embargo, en algunos casos, después de cierto número de épocas de entrenamiento, cuando las curvas de error se alisaban, se incrementó el valor del parámetro η . Aquí conviene aclarar que, para el entrenamiento por mini-lotes, el paquete SNNS automáticamente divide el valor del parámetro η entre el número de patrones del mini-lote. Esto implica que el aprendizaje usando el método por mini-lotes será lento en comparación con el aprendizaje incremental. En las tablas 4.1 y 4.2 se resumen las ventajas de las dos alternativas de aprendizaje [130].

Ventajas del aprendizaje incremental	
1	El aprendizaje incremental generalmente es más rápido que el aprendizaje por mini-lotes.
2	El aprendizaje incremental a menudo resulta en mejores soluciones.
3	El aprendizaje incremental puede ser usado para ambientes de información cambiantes.

Tabla 4.1: Ventajas del aprendizaje cuando se usa el método de aprendizaje incremental.

Ventajas del aprendizaje por mini-lotes	
1	En el aprendizaje por mini-lotes se entienden mejor las condiciones de convergencia.
2	Muchas opciones de aceleración (gradiente conjugado) sólo operan en aprendizaje por mini-lotes.
3	Los análisis teóricos de la dinámica de los pesos sinápticos y de razón de convergencia son más simples.

Tabla 4.2: Ventajas del aprendizaje cuando se usa el método de aprendizaje por mini-lotes.

Estas observaciones presentadas por Yann LeCun [130], están formuladas para entrenamiento por lotes y no precisamente por mini-lotes. Sin embargo, se puede esperar que para tamaños de

mini-lotes no muy pequeños, estas observaciones puedan seguir siendo válidas. En última instancia, los resultados de este trabajo, entre otras cosas, serán útiles para evaluar algunas de estas observaciones. Por ejemplo, trataremos de estimar que tanto más rápido es el aprendizaje incremental en comparación al aprendizaje por mini-lotes. Trataremos de determinar si en efecto las soluciones que proporciona el método incremental son mejores que la de mini-lotes. En relación al punto 3 de la tabla 4.1, cabe aclarar que el entrenamiento incremental puede ser en línea o fuera de línea. En caso del entrenamiento incremental en línea, cada patrón de entrada, una vez procesado con el fin de que la RNA adapte sus pesos, es descartado. Es decir, la red siempre estará siendo entrenada con patrones nuevos, por lo cual es factible que ésta pueda seguir los cambios de un ambiente de información cambiante con el tiempo. En relación al punto 1 de la tabla 4.2, la razón por la cual los trabajos teóricos permiten entender mejor las condiciones de convergencia para el aprendizaje por mini-lotes se debe a que, como se dijo arriba, la formulación mecánico estadística del aprendizaje por lotes ha sido formulado como el problema de un sistema en equilibrio térmico [222] y no fuera del equilibrio como sucede en el caso del aprendizaje incremental. En relación al punto 2 de la tabla 4.2, podemos decir que, con objeto de aumentar la lenta convergencia del aprendizaje por lotes o mini-lotes, para esta aprendizaje se han desarrollado poderosas técnicas que, por ejemplo, no requieren, como si ocurre en el aprendizaje incremental, del tedioso ajuste de las razones de aprendizaje (parámetro η en la ecuación 3.29) o de otros parámetros que se usan en esa técnica de aprendizaje. En este trabajo no se hará uso de esas alternativas avanzadas, sólo mencionaremos algunas de éstas: Quickpro [71], RPROP [170] y de manera muy especial las alternativas basadas en métodos para optimización de funciones objetivo (funciones de error más generales) no lineales como gradientes conjugados y Levenberg-Marquardt [141]. En relación al punto 3 de la tabla 4.2, las razones son similares a las presentadas en relación al punto 1 de la misma tabla.

4.3. El ruido y los conjuntos de entrenamiento y de validación

Uno de los objetivos fundamentales de este trabajo es determinar el efecto del tamaño ruido en las mediciones sobre el desempeño del ADFP aplicado a la ECB. Este tamaño del ruido es definido como la razón de ruido a señal o parámetro R definido en la sección 2.3.2. Es de esperarse que en la medida en que aumenta el valor de la *razón de ruido a señal* R de los patrones perteneciente a un determinado ambiente de información, irá disminuyendo la cantidad de información extraíble mediante el entrenamiento de una RNA que use patrones tomados de ese ambiente de información ruidoso. Esto es, el *error de generalización* de la RNA irá en aumento en la medida en que el parámetro R crezca [219].

El error de generalización de una RNA se define de la siguiente manera. Para el proceso de aprendizaje o adaptación de una RNA, normalmente se usa el método de “*hold-out*” [42], es decir, se seleccionan dos conjuntos de patrones tomados del ambiente de información correspondiente. Uno de ellos, denominado *conjunto de entrenamiento*, es el que realmente se usa para la adaptación de la RNA, es decir, *seleccionar el modelo*. Por ejemplo, en el caso de aprendizaje por retro-propagación, el conjunto de entrenamiento es el conjunto para el cual se retro-propaga el error definido como el desacuerdo entre el valor predicho por la RNA con respecto al valor deseado para

cada patrón del conjunto de entrenamiento. A lo largo de ese proceso de retro-propagación, la RNA aprende o se adapta, ya sea usando el aprendizaje incremental, por lotes o por mini-lotes. El otro conjunto empleado durante el proceso de aprendizaje se denomina *conjunto de validación*. También se calcula el error para los patrones de este conjunto de validación, pero esta vez el error no se retro-propaga, es decir, no se usa para adaptar los valores de los pesos sinápticos de la RNA, simplemente se usa para hacer una estimación no sesgada del error de generalización cometido por la RNA, es decir, *validar el modelo*. Esto contrasta con el error que se obtiene usando el conjunto de entrenamiento, el cual subestima el error de generalización, pues éste da una estimación sesgada del error de generalización, puesto que los parámetros de la RNA fueron ajustados usando justamente los patrones de este conjunto de entrenamiento, por lo cual sus valores pueden estar sesgados a favor de esos patrones, es decir, es un *modelo sesgado*, en particular, por el ruido de esos patrones.

De acuerdo a esta heurística, la manera de estimar el error de generalización es usando la *curva de error de validación* $E^V(t)$ (el error calculado sobre el conjunto de validación como función del número de épocas de entrenamiento (t) de la RNA) junto con la técnica de *detención temprana* [42] del entrenamiento. Con esta técnica se controla la complejidad necesaria de la RNA que se entrena, para que pueda resolver el problema que se desea de manera óptima. Durante una sesión típica de entrenamiento de una RNA, las curvas de error de entrenamiento $E^T(t)$ (el error calculado sobre el conjunto de entrenamiento como función del número de épocas de entrenamiento (t) de la RNA) y de validación $E^V(t)$ tienen la forma conceptual mostrada en la figura 4.1. En este trabajo la forma explícita de calcular estas curvas de error será:

$$E^T(t) = \frac{1}{K^T} \sum_{p \in C^T} \left| \bar{y}^d(\vec{x}_p) - \vec{f}(\vec{x}_p; \vec{\omega}(t)) \right|^2 \quad (4.3)$$

y

$$E^V(t) = \frac{1}{K^V} \sum_{p \in C^V} \left| \bar{y}^d(\vec{x}_p) - \vec{f}(\vec{x}_p; \vec{\omega}(t)) \right|^2 \quad (4.4)$$

donde C^T y C^V son los conjuntos de entrenamiento y de validación, el primero con K^T patrones y el segundo con K^V . Como se verá más adelante, estos dos números se tomarán iguales. $\bar{y}^d(\vec{x}_p)$ representa el valor deseado para el patrón p , representado por el vector de entrada \vec{x}_p . Es decir, $\bar{y}^d(\vec{x}_p)$ es un arreglo de dimensión 2, que tiene por primera coordenada el valor deseado de la energía total del ion E_{tot}^d , correspondiente al patrón p o curva de Bragg \vec{x}_p , y por segunda coordenada tiene el valor deseado de la amplitud del pico de Bragg, PB^d , de esta CB. Las correspondientes predicciones que la RNA hace para estas dos cantidades están representadas por el arreglo bidimensional $\vec{f}(\vec{x}_p; \vec{\omega}(t))$. Estas predicciones las hace la RNA usando el valor que tiene el vector de pesos y umbrales $\vec{\omega}(t)$ después de t épocas de entrenamiento.

Como se indica esquemáticamente en la figura 4.1, la curva de error de entrenamiento $E^T(t)$ decrece monótonamente, mientras que la curva del error de validación $E^V(t)$ presenta un mínimo después de un cierto número de épocas de entrenamiento t_{min} . Cuando se usa la técnica de detención temprana, se escoge como RNA óptima a la correspondiente al mínimo de la curva de error de validación. A partir de ese punto, lo que ocurre es sobre-entrenamiento de la RNA, lo cual acarrea

pérdida de *poder de generalización* de la RNA, es decir, a partir de ese punto, la red seguirá disminuyendo el error que comete al presentársele patrones del conjunto de entrenamiento, pero irá aumentando el error que comete al presentársele patrones que pertenecen al conjunto de validación, lo cual, por ejemplo, en casos de un ambiente de información ruidoso, se debe a que la red empieza a aprender el ruido presente en los patrones del conjunto de entrenamiento, el cual, desde luego, es diferente al ruido presente en los patrones del conjunto de validación.

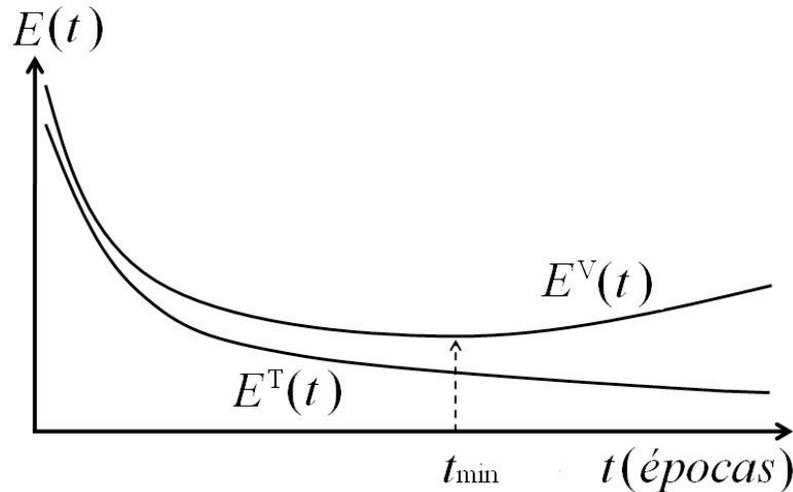


Figura 4.1: Forma general de la *curva de error de entrenamiento* $E^T(t)$ y de la *curva error de validación* $E^V(t)$. La técnica de *detención temprana* consiste en elegir como RNA óptima a la correspondiente al mínimo del error de validación, el cual ocurren en t_{min} épocas de entrenamiento.

La existencia de este mínimo en la curva del error de validación, $E^V(t)$, también se puede explicar en términos de la descomposición de este error de generalización en dos partes: el *sesgo* y la *varianza* [81]. El sesgo es una medida de en cuánto la salida de la red, promediada sobre el conjunto de entrenamiento, difiere de los valores deseados. La varianza es una medida de cuánto la salida de la red varía entre diferentes conjuntos de entrenamiento. Al comienzo del entrenamiento de una red, el sesgo es grande puesto que la salida de ésta aún está lejos de los valores deseados. La varianza es pequeña puesto que los datos aún tienen poca influencia en la red. Hacia el final del entrenamiento de la red, el sesgo es pequeño puesto que ésta ha aprendido los valores deseados correspondientes al ambiente de información estudiado. Sin embargo, si el entrenamiento se prolonga demasiado, entonces la red empezará a aprender el ruido propio del conjunto de entrenamiento, lo cual se acostumbra denominar como *sobre-entrenamiento*. En tal caso, la varianza será grande puesto que el ruido varía al cambiar el conjunto de entrenamiento. El mejor desempeño de la red queda determinado por este compromiso entre el sesgo y la varianza, lo cual se manifiesta como un mínimo del error de generalización o de validación $E^V(t)$. Es en este mínimo en donde se suspende el entrenamiento cuando se usa la técnica de la detención temprana.

La forma en que en este trabajo se determinará el efecto del tamaño del ruido R en las mediciones sobre el desempeño del ADFP aplicado a la ECB será estudiando la dependencia del error

de generalización, es decir, del mínimo del error de validación, $E^V(t_{min})$, con el valor del tamaño de la razón ruido a señal, R . Con tal objetivo, usando la Ec. 2.10, se generarán conjuntos de entrenamiento $C^T(R)$ y de validación $C^V(R)$ correspondientes a ambientes de información con los siguientes valores de la razón de ruido a señal:

$$R = 1\%, 2\%, \dots, 10\% \quad (4.5)$$

En cada caso se determinará el error de generalización como el valor del error de validación mínimo correspondiente, es decir, $E^V(R, t_{min})$.

En el capítulo 2 se mencionó que las CBs consideradas en este trabajo corresponden a 11 diferentes valores de la amplitud del pico de Bragg (PB^d) (simulando 11 diferentes números atómicos) y a 41 diferentes posibles valores de la energía total inicial del ion (E_{Tot}^d); todos los 41 valores garantizan que se alcanza el PB. Cada uno de los conjuntos $C^T(R)$ como los $C^V(R)$ constan de 100 representantes para cada uno de las diferentes $451 = 11 \times 41$ clases de CBs, es decir, ambos conjuntos constan de 45,100 CBs. Cada uno de las 45,100 CBs pertenecientes a $C^T(R)$ o a $C^V(R)$ se genera con una razón de ruido a señal igual a R , donde R toma los valores definidos en la ecuación 4.5.

En la figura 2.1 se muestra una CB correspondientes a $R = 10\%$ (curva punteada). En esta figura también se muestra la correspondiente CB ideal, es decir, sin ruido (curva roja). Los correspondientes PBs de ambas curvas también se señalan en la figura. Como se puede ver en la figura el tamaño de la digitalización es de 81 canales (de 0 a 80). Este tamaño de digitalización es el que determinó cual sería el tamaño de la capa de entrada de la RNA usada. Los 41 valores de la energía total de las 451 diferentes clases analizadas corresponden a CBs con tamaño de digitalización de al menos 41 canales (del canal 0 al 40 mínimamente), región indicado en color cian en la figura 2.1. De esta manera se garantiza que todas las CBs alcanzan su correspondiente PB, el cual, como se indica en la figura 2.1, se localiza en el canal 16.

4.4. Resultados

Como se dijo arriba, se estudiaron dos tamaños de mini-lotes para la implementación de la ley de aprendizaje por retro-propagación del error, $n_{lote} = 16$ y $n_{lote} = 100$. También se realizaron estudios para dos intervalos para la inicialización de los pesos sinápticos y de los umbrales, es decir: $[-0.5, 0.5]$ y $[-0.03, 0.03]$. Para $n_{lote} = 16$ y el intervalo $[-0.5, 0.5]$, se estudiaron los ambientes de información correspondientes a los 10 valores de la razón ruido a señal listados en la ecuación 4.5. En la tabla 4.3 se resumen todos los casos estudiados y se indica el número total de épocas de entrenamiento para cada uno. Además, se muestra con un asterisco los casos en los que se alcanzó el mínimo de las curvas de error de validación, $E^V(R, t)$ vs t .

En la tabla 4.4 se resumen los parámetros usados para el entrenamiento de las RNAs. El valor que se reporta para la razón de aprendizaje η es el valor inicial de este parámetro. En algunos casos, debido a lo liso de las curvas de error y al lento aprendizaje, se optó por incrementar el valor de η , valores que se indicarán en el momento conveniente. Cada patrón del conjunto de entrenamiento

$R(\%)$	$n_{lote} = 16$		$n_{lote} = 100$	
	$\eta = 0.5$	$\eta = 0.03$	$\eta = 0.5$	$\eta = 0.03$
1	30,000,000*	3,000,000	3,000,000	3,000,000
2	26,000,000*			
3	26,000,000*			
4	26,000,000*			
5	30,000,000*	25,000,000*	4,000,000	4,000,000
6	20,000,000*	40,000,000		
7	20,000,000*			
8	20,000,000*			
9	20,000,000*			
10	30,000,000*	30,000,000*	35,000,000*	35,000,000*

Tabla 4.3: Resumen de todos los casos usados para la adaptación de las RNAs usando el aprendizaje supervisado por retro-propagación del error por mini-lotes. En cada caso se indica el número total de épocas de entrenamiento y se indica con un asterisco los casos en los que se alcanzó el mínimo de la curva de error de validación $E^V(R, t)$ vs t .

es presentado una vez durante cada época de entrenamiento. El orden de los patrones en que, en cada época, éstos se presentan a la red, se escogió que sea barajado, es decir, al azar, pero sólo una vez cada uno. La función de activación seleccionada para todas las neuronas es la función logística (Ec. 4.1 y 4.2). Finalmente, el cálculo de tanto las activaciones como de las salidas de las neuronas de la red se realiza en orden serial, es decir, el proceso de cálculo comienza con la primera neurona y termina con la última. Hacemos notar que las neuronas se numeran empezando con las neuronas de la capa de entrada, siguiendo con las de las capas oculta, y terminando con las neuronas de la capa de salida.

En las figuras 4.2 a 4.11, se presentan las curvas de error $E^T(R, t)$ y $E^V(R, t)$ para los 10 valores estudiados de R (Ec. 4.5). En estas figuras, se indica la posición del mínimo de la curva $E^V(R, t)$, el cual determina el error de generalización, es decir, el mínimo del error de validación $E^V(R, t_{min})$.

En las figuras 4.2 a 4.5, correspondientes a $R = 1\%$, 2% , 3% y 4% , se observa que después de cierto número de épocas de entrenamiento, las curvas de error son lisas y disminuyen muy lentamente, por lo cual se incrementó el aprendizaje aumentando el valor de la constante de aprendizaje η . En todos los casos, el valor inicial fue $\eta = 0.3$. En el caso $R = 1\%$, la η se incrementó en dos ocasiones: en $t = 1,000,000$ épocas a $\eta = 6.0$, y en $t = 10,000,000$ épocas a $\eta = 30.0$. En los otros casos, $R = 2\%$, 3% y 4% , η se incrementó en $t = 20,000,000$ épocas a $\eta = 6.0$. En todos los casos, los cambios en la razón de aprendizaje producen una inestabilidad en las curvas de error, las cuales, eventualmente, se amortiguan y la RNA se sigue adaptando, es decir, siguen disminuyendo su valor. Estos incrementos en la razón de aprendizaje son tolerados por las RNAs, lo cual se debe a que el entrenamiento previo las ha llevado a zonas en las cuales las correspondientes superficies de error son ya muy lisas, por lo que cambian muy lentamente y

Parámetro	Valor
Ley de aprendizaje	Retro-propagación por mini-lotes
Tamaño de la RNA	Capa de entrada: 81 neuronas Cinco capas ocultas: 9 neuronas cada una Capa de salida: 2 neuronas Completamente conectada
η = razón de aprendizaje	Valor inicial: 0.3
Intervalo de inicialización de los pesos	[-0.5, 0.5] y [-0.03, 0.03]
Orden de presentación de los patrones	Barajado
Función de activación	Función logística
Orden de actualización de las neuronas	Orden serial
Conjuntos de entrenamiento C^T y C^V	100 muestras de cada una de las 451 clases: (11 valores de PBs) \times (41 valores de ETot)

Tabla 4.4: Valores de los parámetros para el entrenamiento de las RNAs.

pueden tolerar pasos grandes en la dirección del gradiente negativo en la búsqueda del mínimo de la curva de validación. La razón fundamental para hacer el incremento del valor de la razón de aprendizaje η en los casos anteriores es que las correspondientes RNAs puedan llegar al mínimo en un tiempo razonable. En la tabla 4.5 se puede ver el número de épocas equivalente ($\eta = 0.3$) que hubiera tomado a cada RNA para llegar a su mínimo.

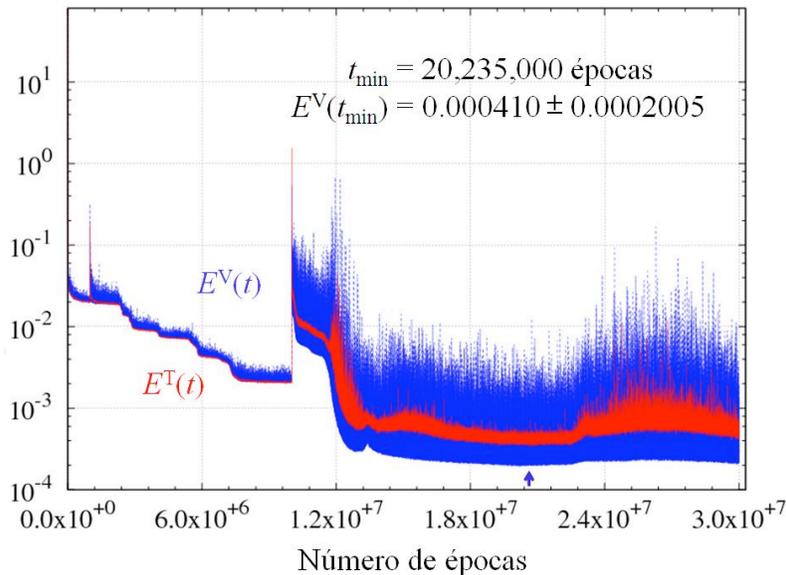


Figura 4.2: Curvas del logaritmo del error de entrenamiento (roja) y validación (azul) para $R = 1\%$. El valor de η se cambió de 0.3 a 6.0 después de 1,000,000 épocas de entrenamiento, y de 6.0 a 30.0 después de 10,000,000 épocas de entrenamiento.

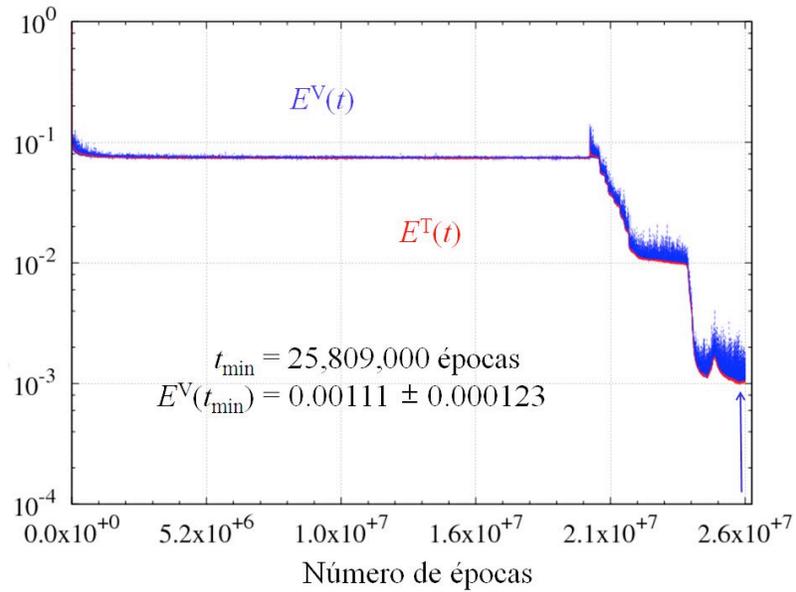


Figura 4.3: Curvas del logaritmo del error de entrenamiento (roja) y validación (azul) para $R = 2\%$. El valor de η se cambió de 0.3 a 6.0 después de 20,000,000 épocas de entrenamiento.

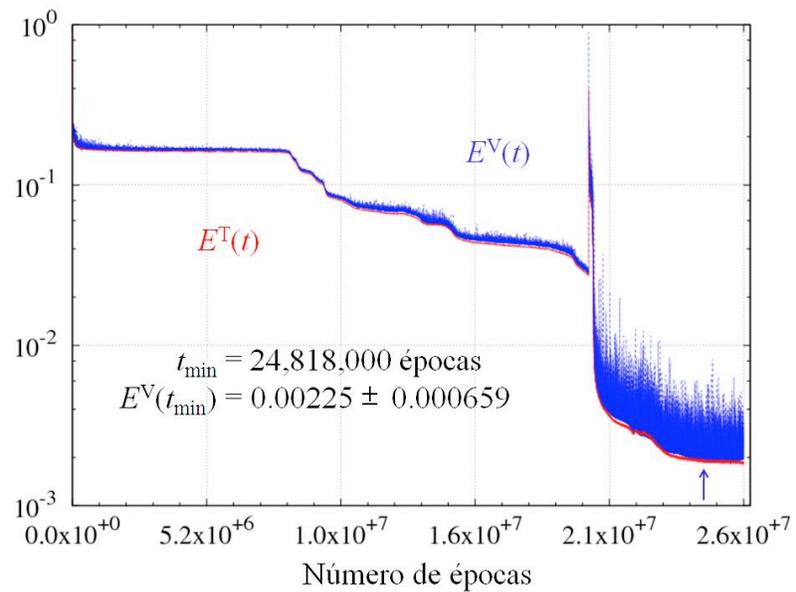


Figura 4.4: Curvas del logaritmo del error de entrenamiento (roja) y validación (azul) para $R = 3\%$. El valor de η se cambió de 0.3 a 6.0 después de 20,000,000 épocas de entrenamiento.

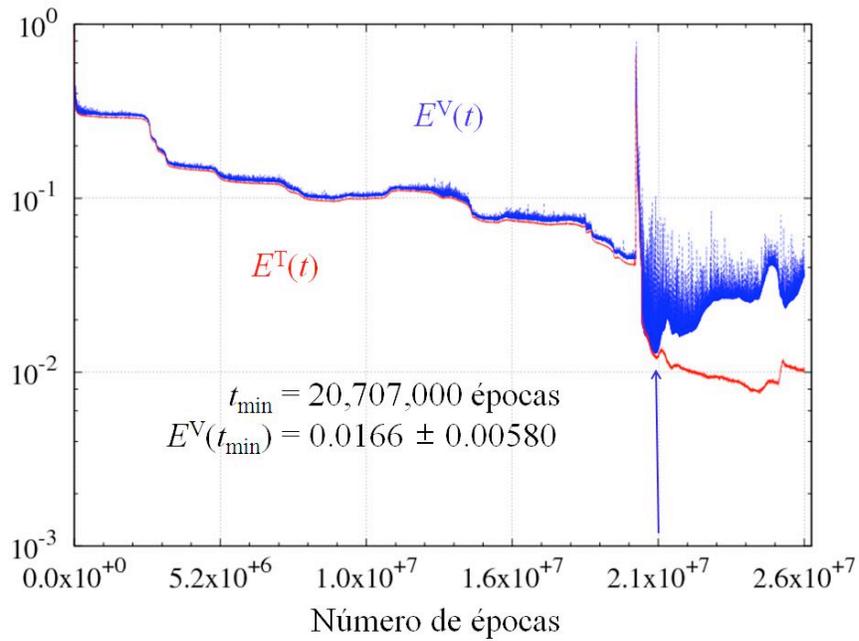


Figura 4.5: Curvas del logaritmo del error de entrenamiento (roja) y validación (azul) para $R = 4\%$. El valor de η se cambió de 0.3 a 6.0 después de 20,000,000 épocas de entrenamiento.

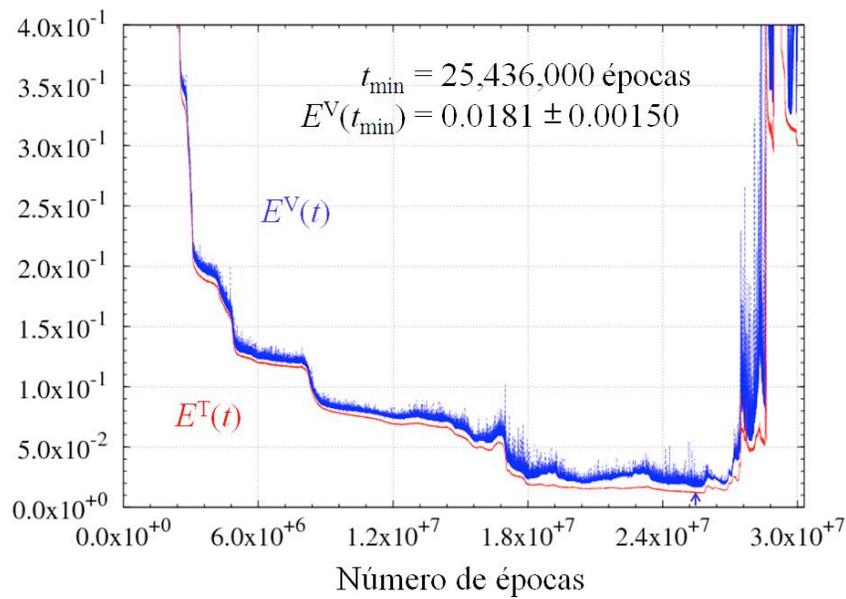


Figura 4.6: Curvas de error de entrenamiento (roja) y validación (azul) para $R = 5\%$.

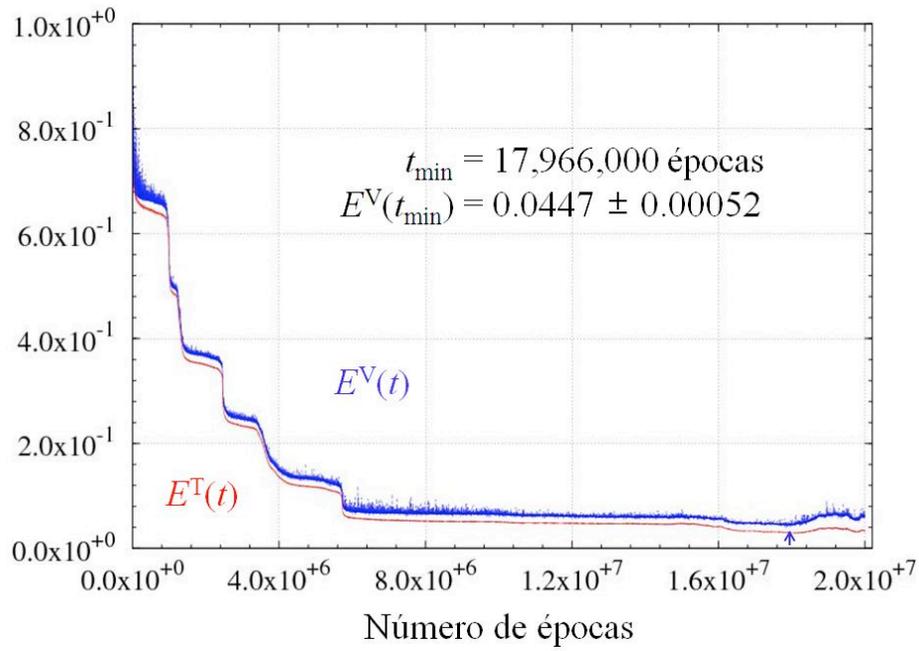


Figura 4.7: Curvas de error de entrenamiento (roja) y validación (azul) para $R = 6\%$.

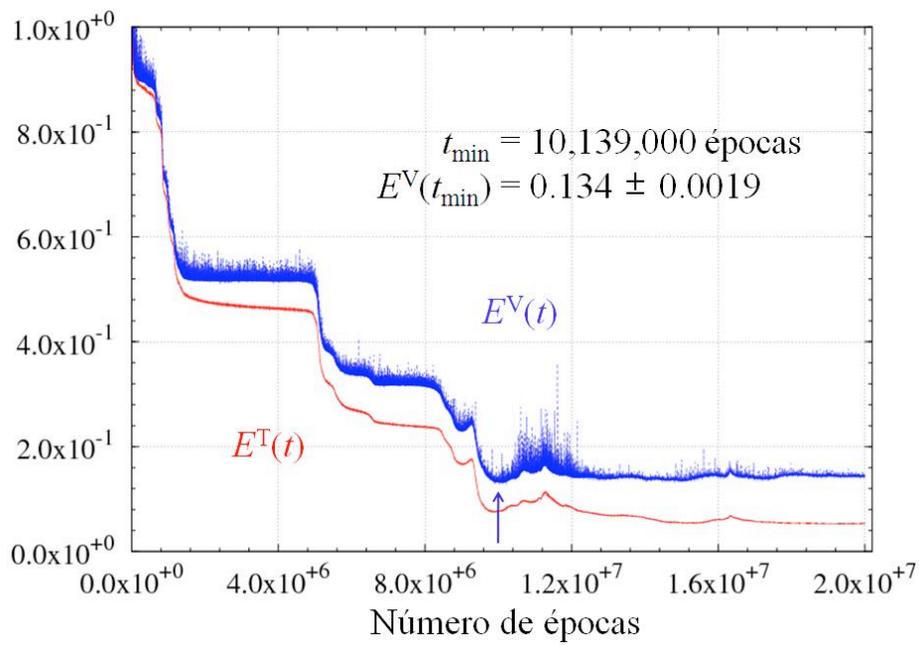


Figura 4.8: Curvas de error de entrenamiento (roja) y validación (azul) para $R = 7\%$.

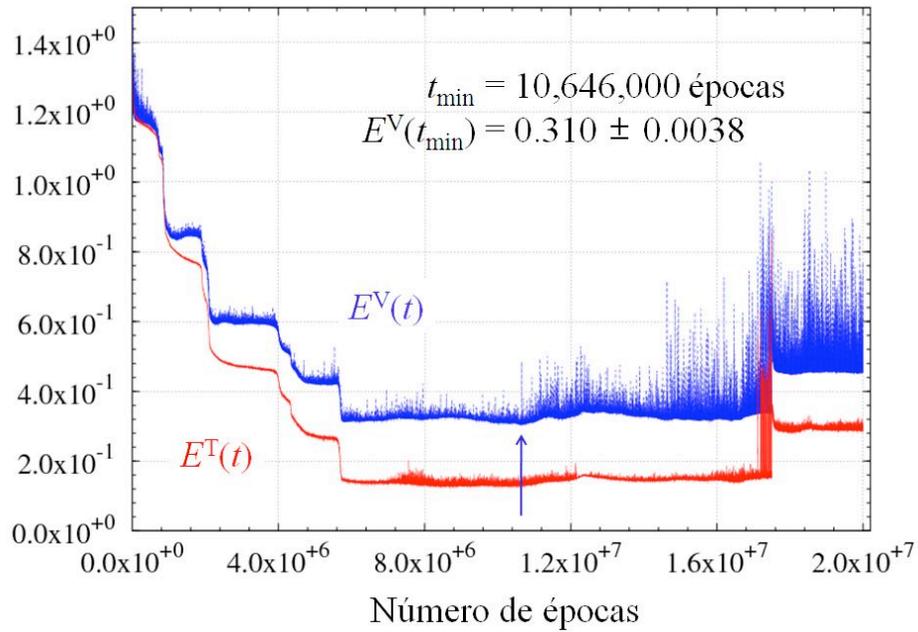


Figura 4.9: Curvas de error de entrenamiento (roja) y validación (azul) para $R = 8\%$.

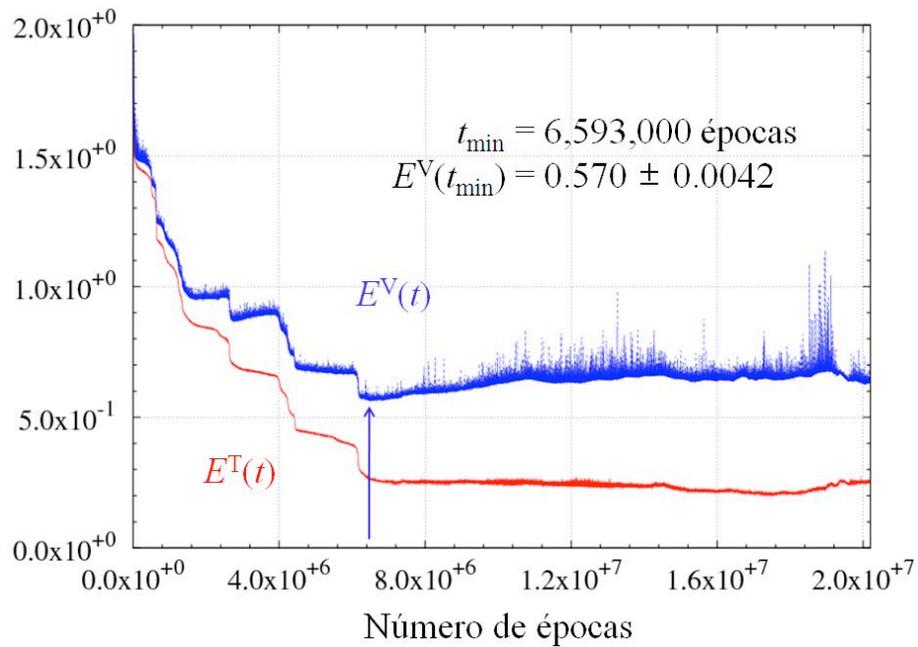


Figura 4.10: Curvas de error de entrenamiento (roja) y validación (azul) para $R = 9\%$.

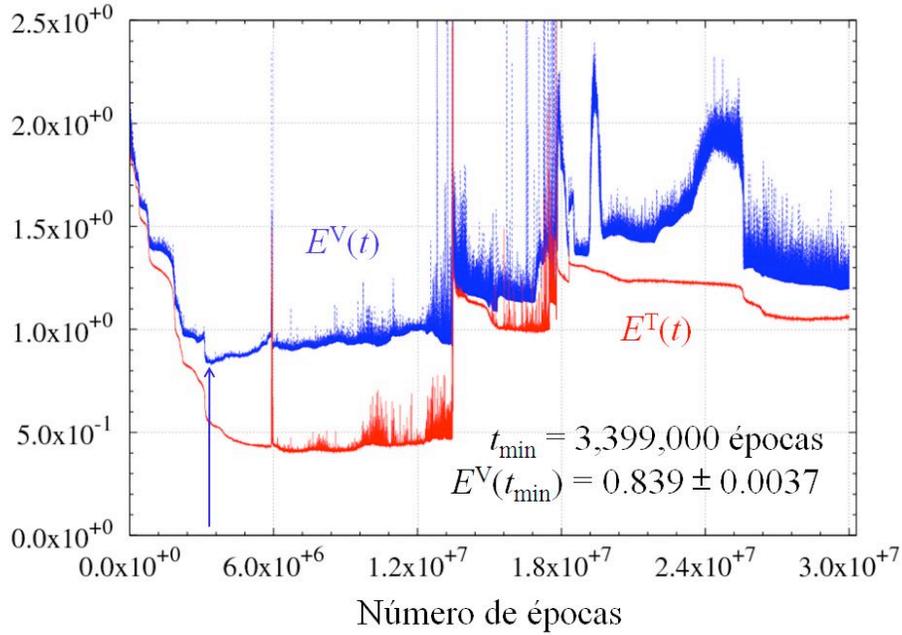


Figura 4.11: Curvas de error de entrenamiento (roja) y validación (azul) para $R = 10\%$.

En la tabla 4.5, para cada uno de los 10 casos analizados, $R = 1\%$, ..., 10% , se resume la información relevante en la búsqueda del mínimo de la curva de validación $E^V(R, t)$. Se indica el número de épocas equivalente de entrenamiento (ver adelante definición), t_{min} , para llegar al mínimo, el valor del mínimo $\vec{E}^V(R, t_{min})$, la desviación estándar de este último valor y la norma en el mínimo, ω_{min} , del vector de pesos sinápticos de las RNAs.

Con objeto de que el valor del mínimo sea más representativo de la tendencia de los valores y dependa menos de las rápidas fluctuaciones de los valores de $E^V(R, t)$ observadas en las figuras 4.2 a 4.11, el valor reportado en la tabla 4.5 de $\vec{E}^V(R, t_{min})$ se calcula como un promedio local de 21 valores de $E^V(R, t)$, espaciados cada uno por 10,000 épocas, en una vecindad en torno a t_{min} . La desviación estándar reportada en esta tabla es la correspondiente a estos 21 valores de $E^V(R, t)$. En relación al número de épocas de entrenamiento necesarias para llegar al mínimo, t_{min} , se hace notar que debido al proceso de adaptación o entrenamiento de la RNA, éstas son una medida de lo que, sobre la superficie de error, se avanza durante la búsqueda del mínimo, y que dado que con objeto de acelerar este procesos de búsqueda, en algunos casos se incrementó el valor del parámetro de razón de entrenamiento η (lo cual hace que se avance más rápido sobre la superficie de error), se decidió escalar el número de épocas conforme al valor usado del parámetro η , tomando como referencia el valor $\eta_0 = 0.3$, es decir:

$$\Delta t^{escalada}(i) = \frac{\eta_i}{\eta_0} \Delta t(i), \quad (4.6)$$

en donde $\Delta t(i)$ es el número de épocas durante las cuales se usó un valor de la razón de aprendizaje

R (%)	t_{min} (épocas)	$\vec{E}^V(R, t_{min})$	$DE[E^V(R, t_{min})]$	ω_{min}
1	1,204,500,000	0.000410	0.0002005	124.5
2	136,180,000	0.00111	0.000123	106.6
3	116,360,000	0.00225	0.000659	110.1
4	34,140,000	0.0166	0.00580	192.7
5	25,436,000	0.0181	0.00150	185.4
6	17,966,000	0.0447	0.00052	182.2
7	10,139,000	0.134	0.0019	136.7
8	10,646,000	0.310	0.0038	139.2
9	6,593,000	0.570	0.0042	154.7
10	3,399,000	0.839	0.0037	95.5

Tabla 4.5: Valores de los parámetros relevantes en la búsqueda del mínimo de las curvas de validación $E^V(R, t)$, correspondientes a los 10 valores estudiados de la razón de ruido a señal R . En los casos $R = 1\%$, 2% , 3% y 4% se reportan las épocas equivalentes ($\eta = 0.3$) que hubiera tomado para llegar al mínimo, en caso de que la η no se hubiera incrementado como se indica en el texto.

igual a η_i . El valor reportado de t_{min} será:

$$t_{min} = \sum_{i=0} \Delta t^{escalada}(i). \quad (4.7)$$

Este escalamiento del número de épocas es lo que en la tabla 4.5 se denominó como número de épocas equivalente. Por ejemplo, en el caso $R = 1\%$, la RNA se entrenó de 1 hasta 1,000,000 de épocas usando $\eta = 0.3$, de 1,000,000 hasta 10,000,000 usando $\eta = 6.0$, y finalmente, de 10,000,000 hasta 30,000,000 épocas usando $\eta = 30.0$, encontrándose el mínimo en 20,235,000, lo cual determina que el mismo se ubique en:

$$t_{min} = 1,000,000 + \frac{6.0}{0.3}9,000,000 + \frac{30.0}{0.3}10,235,000 = 1,204,500,00 \quad (4.8)$$

Con objeto de visualizar la calidad del desempeño de cada una de las 10 RNAs de las figuras 4.2 a 4.11, en las figuras 4.12 a 4.21, se muestran gráficas de dispersión de puntos BP vs. E_{tot} correspondientes a los mínimos del error de validación $E^V(R, t_{min})$. De estas figuras es claro que, como era de esperarse, las gráficas de dispersión de puntos para valores pequeños de la razón ruido a señal, R , son de más calidad que las correspondientes a valores grandes de R . Esto es, en estas gráficas, el traslape de puntos de clases diferentes (lo cual da lugar a una identificación defectuosa de las CBs) es menor a medida el valor de R disminuye. Por la referencia [219], se sabe que el número de épocas que le toma a las curvas $E^V(R, t)$ llegar a su mínimo crece a medida que decrece el valor de la razón ruido a señal, R . Resultado aparentemente paradójico, pues nos dice que a medida que el ruido disminuye, la RNA tarda más en aprender su tarea, es decir, llegar al mínimo de la curva del error de validación. Sin embargo, como se explica en [219], esto se debe a que las tareas de aprendizaje que cada RNA realiza en su búsqueda del mínimo no son equivalentes.

Podemos pensar que las CBs tienen una serie de *características jerarquizadas* de acuerdo a su relevancia, es decir, la más conspicua ocupará la más alta jerarquía y la irán disminuyendo sucesivamente. Es de esperarse que, en el proceso de aprendizaje de las RNAs, éstas aprendan primero la característica de más alta jerarquía de las CBs y, posteriormente, irán aprendiendo paso a paso, las características de menor jerarquía. Este proceso de aprendizaje será perturbado por la presencia de ruido en las señales o CBs. Esto es, aquellas características de jerarquías bajas corresponden a características de las CBs que podrán ser opacadas u ocultadas por la presencia de una componente de ruido en las señales. La RNA no será capaz de aprender nunca esas características de baja jerarquía. De tal suerte que, a medida que la razón de ruido a señal, R , aumente, menor será el número de características aprendibles por la correspondiente RNA. Esto significa que la red neuronal que corresponde a un ambiente de información con $R = 1\%$, será la RNA que más características extraerá o aprenderá del medio, por lo cual también será la que más tarde en llegar al mínimo. Por ejemplo, esta tendencia se ve claramente reflejada en la tabla 4.5, en donde se ve como la función $t_{min}(R)$, es decir, el número de épocas que le toma a cada curva de error $E^V(R, t)$ para llegar a su mínimo, es una función monótonamente decreciente de R .

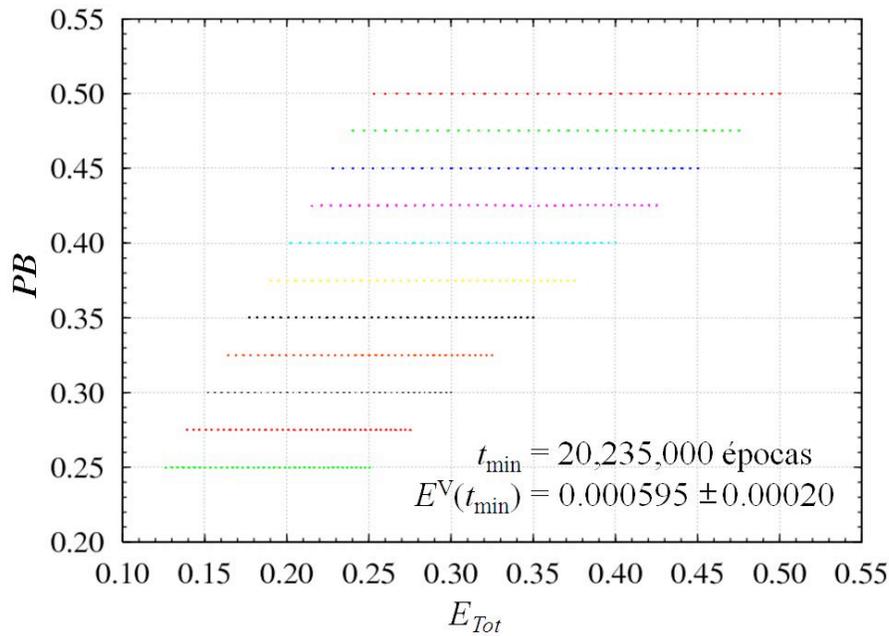


Figura 4.12: Gráfica de dispersión de puntos de validación BP vs E_{Tot} en el mínimo para $R = 1\%$.

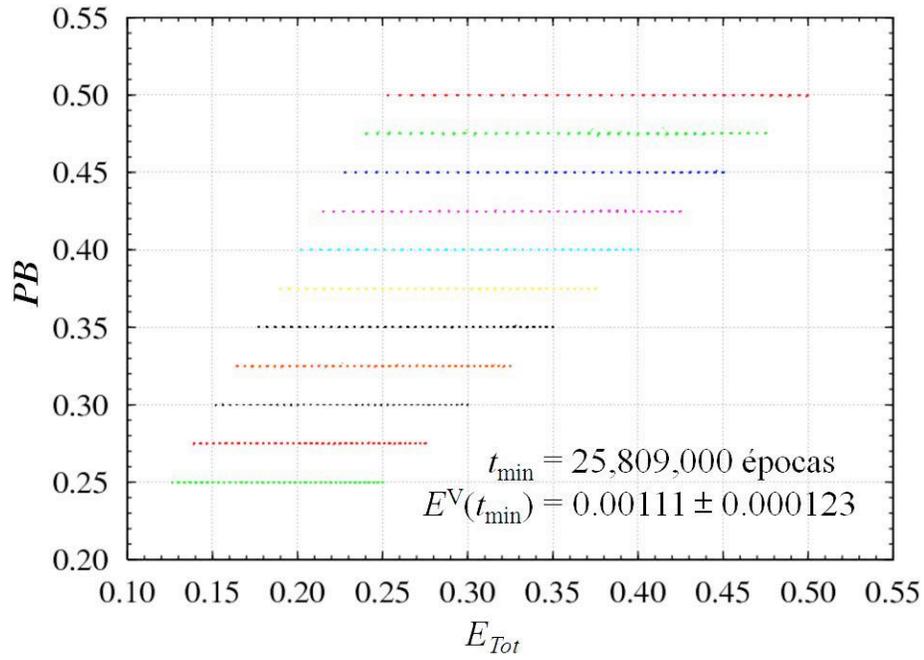


Figura 4.13: Gráfica de dispersión de puntos validación BP vs E_{Tot} en el mínimo para $R = 2\%$.

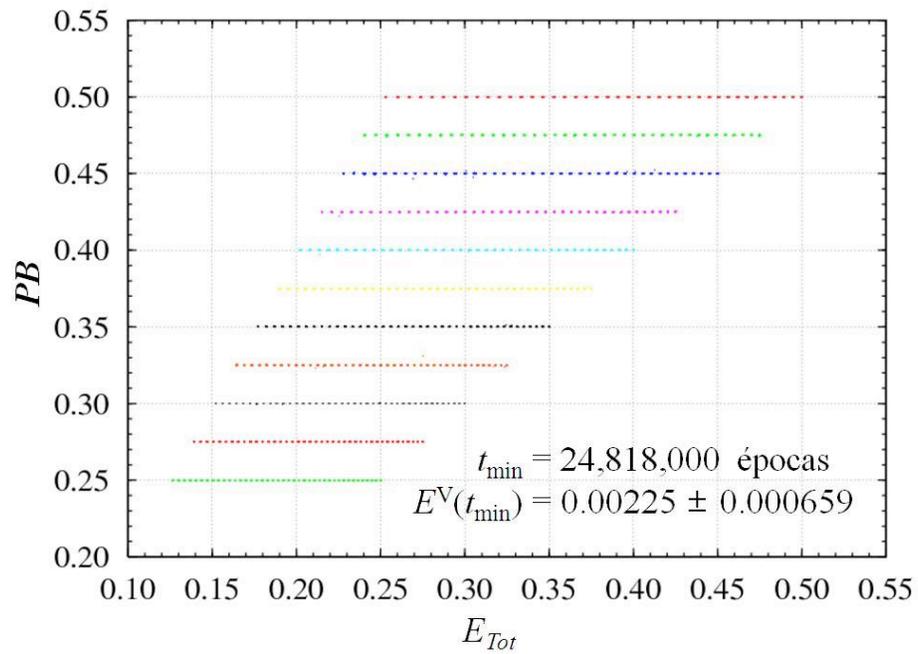


Figura 4.14: Gráfica de dispersión de puntos validación BP vs E_{Tot} en el mínimo para $R = 3\%$.

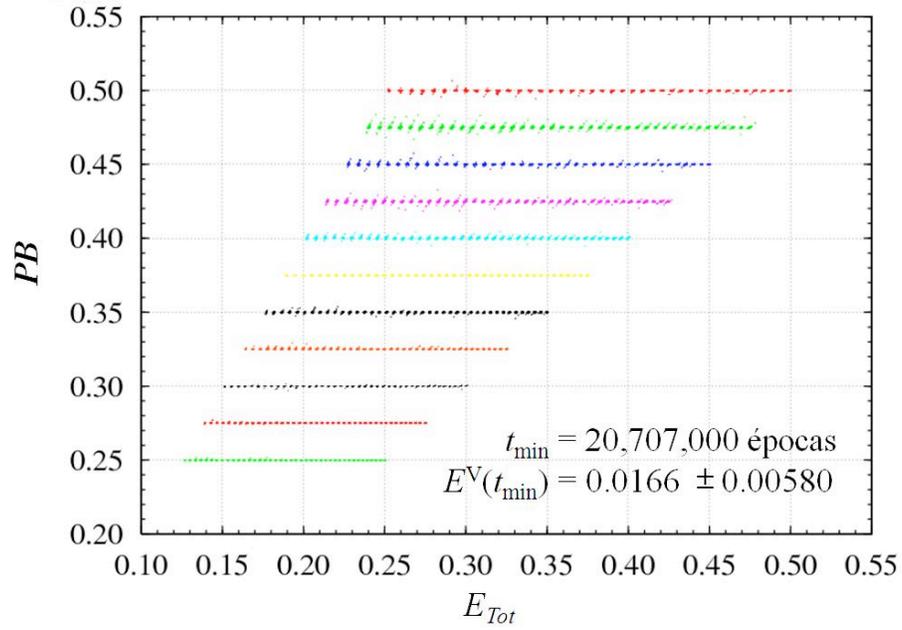


Figura 4.15: Gráfica de dispersión de puntos validación BP vs E_{Tot} en el mínimo para $R = 4\%$.

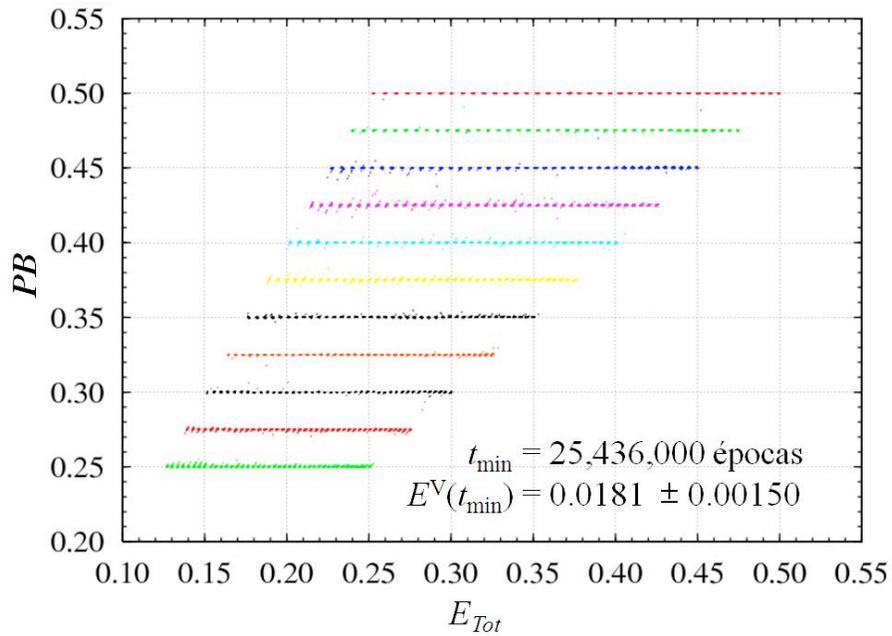


Figura 4.16: Gráfica de dispersión de puntos validación BP vs E_{Tot} en el mínimo para $R = 5\%$.

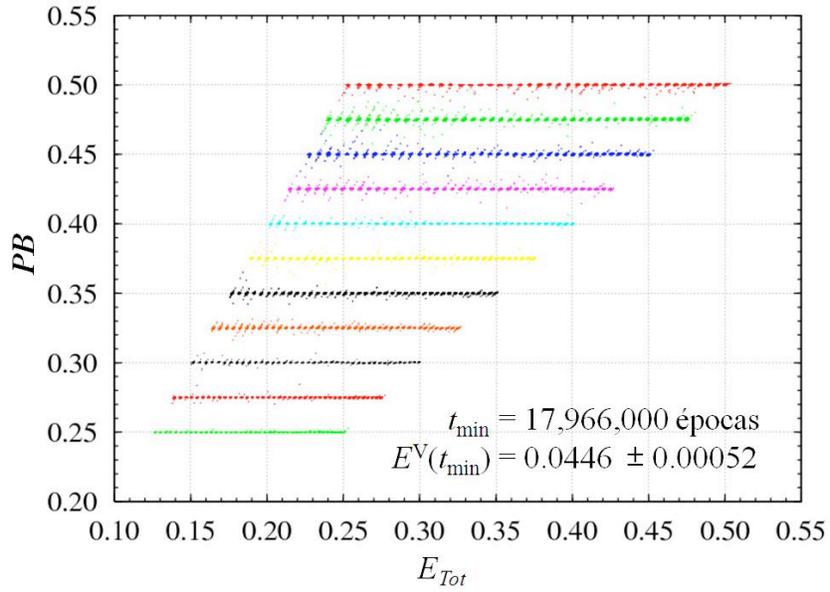


Figura 4.17: Gráfica de dispersión de puntos validación BP vs E_{Tot} en el mínimo para $R = 6\%$.

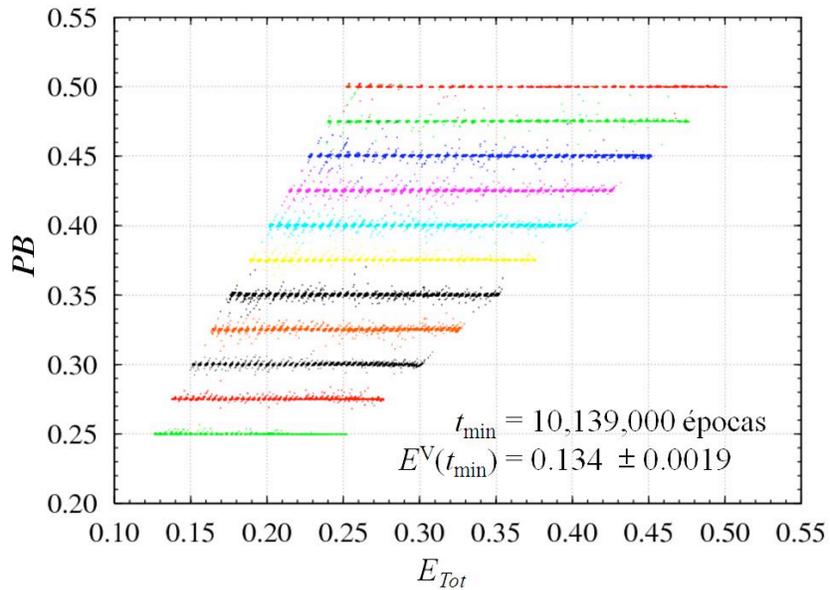


Figura 4.18: Gráfica de dispersión de puntos validación BP vs E_{Tot} en el mínimo para $R = 7\%$.

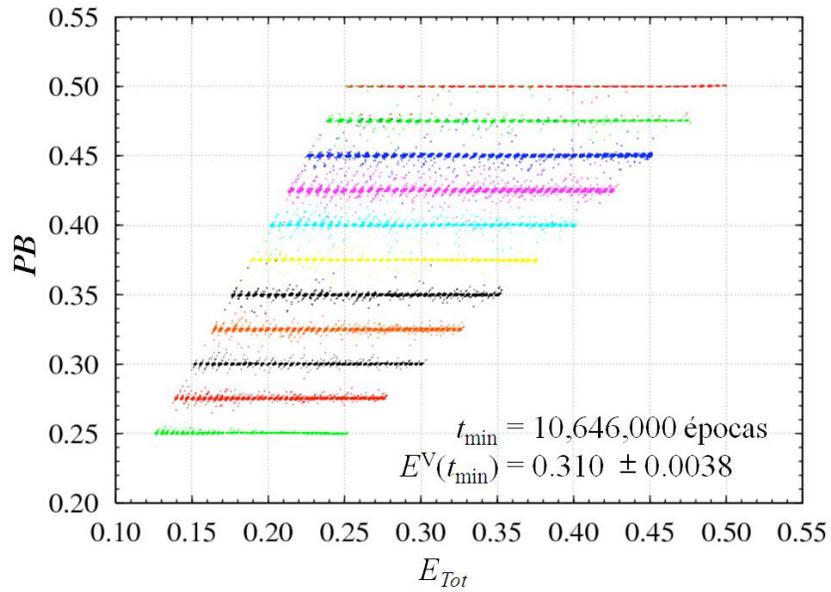


Figura 4.19: Gráfica de dispersión de puntos validación BP vs E_{Tot} en el mínimo para $R = 8\%$.

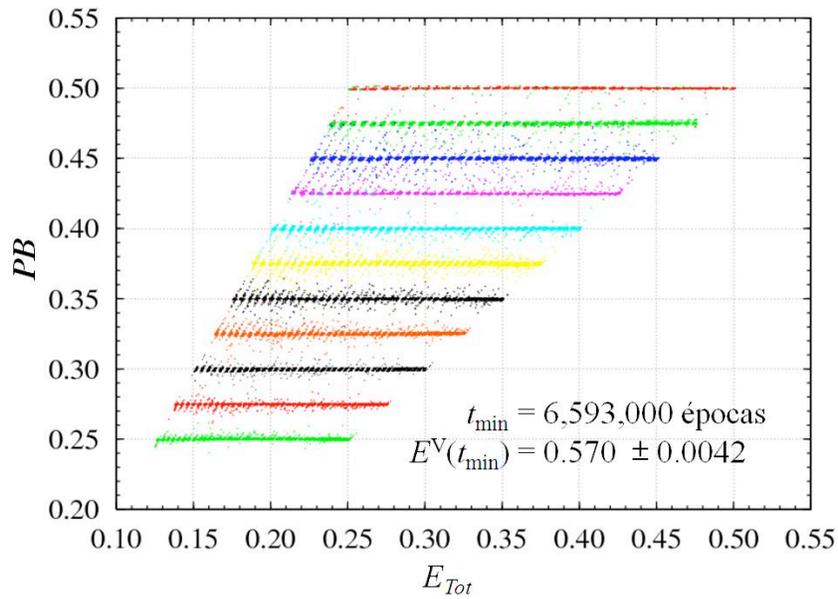


Figura 4.20: Gráfica de dispersión de puntos validación BP vs E_{Tot} en el mínimo para $R = 9\%$.

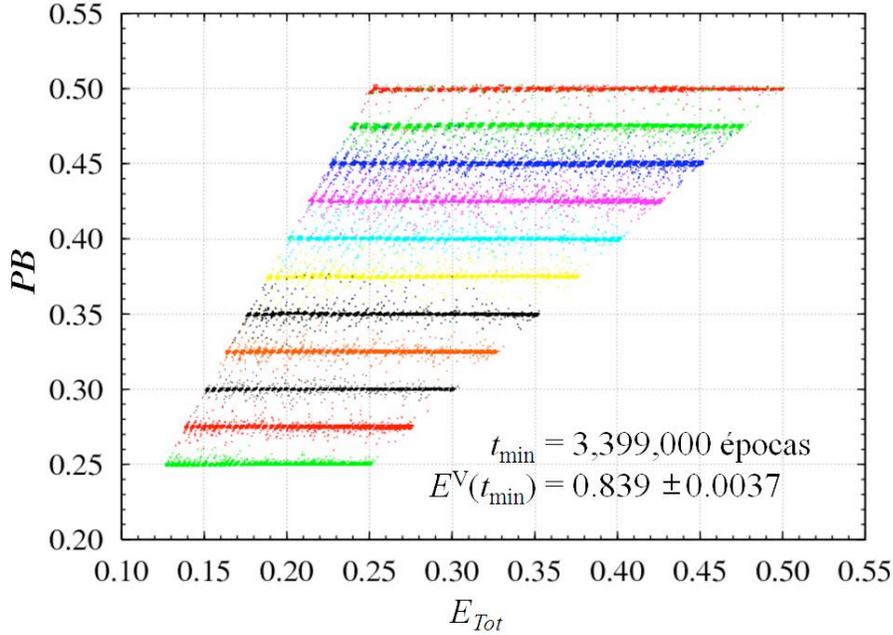


Figura 4.21: Gráfica de dispersión de puntos validación BP vs E_{Tot} en el mínimo para $R = 10\%$.

Con objeto de averiguar el efecto limitante que el ruido tiene sobre la capacidad de aprendizaje de una RNA, en [219], usando entrenamiento incremental en lugar de entrenamiento por mini-lotes como en este trabajo, se determinó la dependencia del tamaño del mínimo de la curva de error de validación como función del tamaño de la razón ruido a señal, R , esto es, se determinó la dependencia funcional $E^V(R, t_{min})$. En la figuras 4.22 y 4.23, tomadas de [219], se muestran las gráficas de $\ln[E^V(R, t_{min})]$ y de $E^V(R, t_{min})$ para los 10 valores estudiados de R . En la figura 4.22, también se presenta un ajuste por mínimos cuadrados, de una recta a los datos (línea sólida roja). Se indican tanto la banda de confianza al 95 % (línea punteada azul) como la banda de predicción al 95 % (línea a guiones negra). El resultado del ajuste por mínimos cuadrados es:

$$\ln[E^V(R, t_{min})] = \beta_0 + \beta_1 R, \quad (4.9)$$

donde $\beta_0 = -8.075 \pm 0.071$ y $\beta_1 = 0.822 \pm 0.030$. El valor del coeficiente de regresión del ajuste por mínimos cuadrados es $r = 0.995$.

En las figuras 4.22 y 4.23, los autores muestran unas barras de incertidumbre, las cuales sirven para dar una idea de la *variabilidad local* de $E^V(R, t)$ en una vecindad en torno a t_{min} . Los autores también hacen notar la existencia de una *variabilidad global* de $E^V(R, t_{min})$ en torno a su valor promedio, es decir, del conjunto de valores mínimos obtenidos cuando se repite todo el entrenamiento de la RNA varias veces. El caso $R = 10\%$ es el único para el cual los autores repitieron el entrenamiento de la correspondiente RNA en 13 ocasiones, lo cual les permitió hacer una estimación del valor promedio así como de su desviación estándar, esta última cantidad brinda

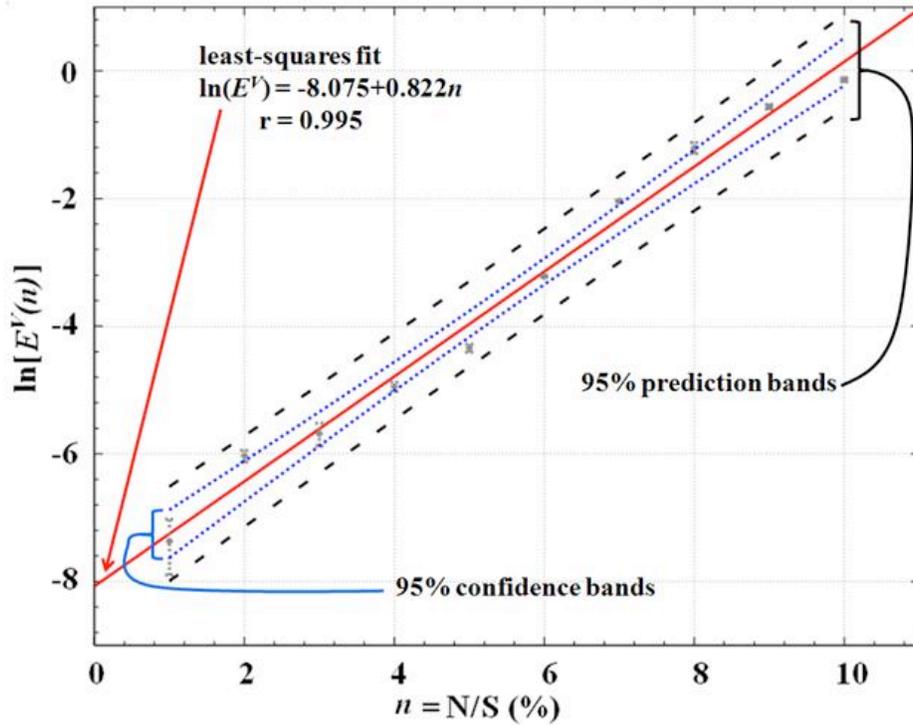


Figura 4.22: Se despliegan los datos correspondientes a los 10 valores de la razón ruido a señal analizados, es decir, $R = n = 1\%$, 2% , ..., 10% . El ajuste a una recta por mínimos cuadrados (línea sólida roja) es $\ln[E^V(R, t_{min})] = -8.075 + (0.822)R$. El coeficiente de correlación del ajuste por mínimos cuadrados es $r = 0.995$. También se muestran la banda de confianza al 95% (líneas punteadas azules) y la banda de predicción al 95% (líneas a guiones negras). (Tomado de [219]).

una medida de la variabilidad global de $E^V(R, t_{min})$. Los valores globales del promedio y de la variabilidad obtenido de esa manera para el caso $R = 10\%$ son iguales a 0.908 ± 0.052 . A estos resultados los comparan con los valores locales obtenidos en el caso de los 13 estudiados, que presentó el aprendizaje más rápido para $R = 10\%$. Estos valores se obtuvieron usando el conjunto de 21 valores de $E^V(R, t)$ en torno a su mínimo ($t_{min} = 196,000$ épocas), es decir, usaron los 21 valores correspondientes a $t = 186,000, 187,000, \dots, 196,000, \dots, 206,000$ épocas. El valor así definido para la incertidumbre local arroja el valor de 0.872 ± 0.027 .

Lo que aquí se ha dado en llamar incertidumbre global de $E^V(R, t_{min})$ se origina a partir del proceso de inicialización al azar de los pesos y umbrales sinápticos de la RNA, lo cual determina globalmente la trayectoria que la red seguirá en su adaptación por aprendizaje en búsqueda del mínimo del error de validación. Es decir, si el punto de inicialización del vector de pesos y umbrales sinápticos cae sobre una parte donde la superficie de error tiene una gran pendiente, la evolución hacia el mínimo se verá favorecida, pero, si por el contrario, este punto de inicialización cae en una parte donde la superficie de error es muy lisa, la evolución hacia el mínimo será lenta. En contraste, la incertidumbre local de $E^V(R, t_{min})$ en torno a t_{min} se origina a partir una componente de

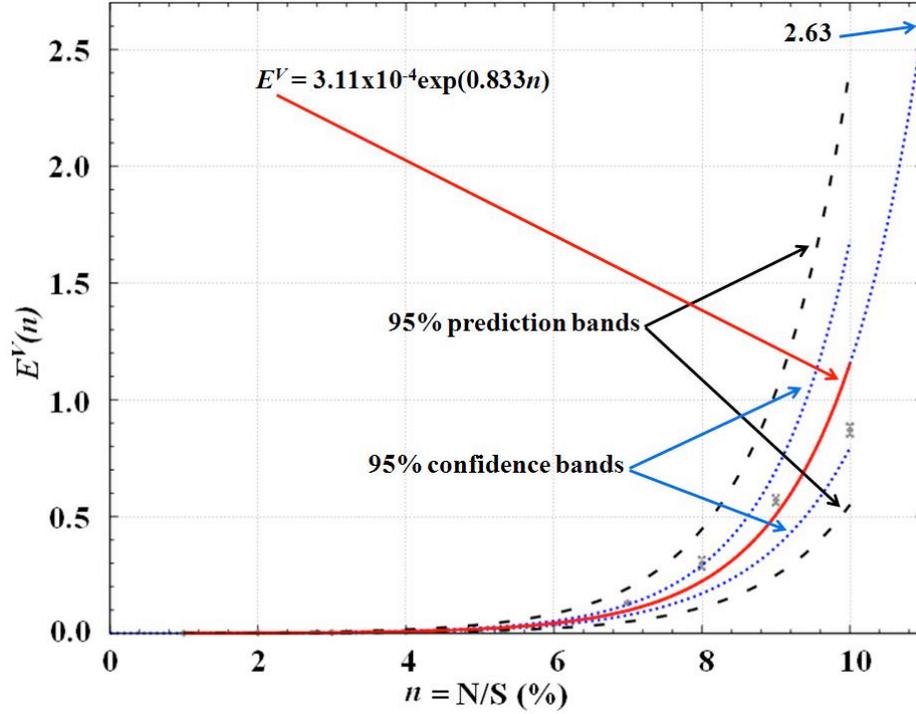


Figura 4.23: Se despliegan los datos correspondientes a los 10 valores de la razón ruido a señal analizados, es decir, $R = n = 1\%$, 2% , ..., 10% . Se muestra la gráfica de $E^V(R, t_{min}) = 3.11 \times 10^{-4} \exp[0.822R]$ que se obtiene del ajuste de una línea recta por mínimos cuadrados a los datos de $\ln[E^V(R, t_{min})]$ vs. R . También se muestran las banda de confianza al 95 % (líneas punteadas azules) y la banda de predicción al 95 % (líneas a guiones negras). (Tomado de [219]).

fluctuación al azar en la adaptación de los pesos y umbrales sinápticos, la cual no es capaz de alterar la forma general de la trayectoria seguida por la red en su búsqueda del mínimo. Esta componente de fluctuaciones al azar se debe a la naturaleza estocástica de la actualización de los pesos y umbrales sinápticos involucrada en el método de aprendizaje por retro-propagación del error, lo cual es más fuerte en el caso del entrenamiento incremental que en el caso del entrenamiento por mini-lotes. Dado que estas dos fuentes de incertidumbre pueden ser consideradas como independientes, entonces la incertidumbre total estará dada por las raíz cuadrada de las suma de cuadrados de ambas incertidumbres, lo cual arroja un valor total de la incertidumbre igual a ± 0.059 .

La línea recta $\ln[E^V(R, t_{min})] = \beta_0 + \beta_1 R$ describe como el ruido R presente en los patrones o CBs por identificar contribuye a fijar un límite a la capacidad de aprendizaje de la RNA. Se puede regresar y mirar a la forma funcional de la figura 4.23, es decir,

$$E^V(R, t_{min}) = \exp[\beta_0 + \beta_1 R] = A \exp[\beta_1 R] \quad (4.10)$$

donde $A = \exp(\beta_0) = 3.11 \times 10^{-4}$. En la figura 4.23, la curva de la ecuación 4.10 corresponde a la curva sólida roja, la cual se extrapola hasta $R = 11\%$, lo cual deja ver que, debido al fuerte

crecimiento exponencial, es muy factible que la RNA ya no pudiera llevar a cabo la tarea de identificación de CBs para $R = 11\%$ o mayores. Las líneas punteadas en azul indican la banda de confianza al 95%. Las líneas a guiones en negro indican la banda de predicción al 95%.

Con objeto de comparar mejor los valores de los mínimos de las curvas de error de validación reportados en [219], obtenidos con aprendizaje por entrenamiento incremental, con los valores obtenidos en este trabajo (Tabla 4.5), obtenidos por aprendizaje por entrenamiento por mini-lotes, se presenta a continuación la tabla 4.6 con los valores de los mínimos de las curvas de error de validación reportados en [219].

R (%)	t_{min} (épocas)	$\vec{E}^V(R, t_{min})$	$DE[E^V(R, t_{min})]$
1	13,433,333	6.3×10^{-4}	2.6×10^{-4}
2	3,050,000	2.4×10^{-3}	2.2×10^{-4}
3	1,409,000	3.4×10^{-3}	6.0×10^{-4}
4	1,874,000	7.1×10^{-3}	5.1×10^{-4}
5	960,000	1.3×10^{-2}	8.2×10^{-4}
6	1,308,000	4.0×10^{-2}	6.0×10^{-4}
7	758,000	1.3×10^{-1}	2.0×10^{-3}
8	777,000	3.0×10^{-1}	2.6×10^{-2}
9	2,515,000	5.7×10^{-1}	2.0×10^{-2}
10	196,000	8.7×10^{-1}	2.7×10^{-2}

Tabla 4.6: Valores de los mínimos de las curvas de validación $E^V(R, t)$ y sus *errores locales*, correspondientes a los 10 valores estudiados de la razón de ruido a señal R en [219].

En las figuras 4.24 y 4.25, se muestran para nuestros datos de la tabla 4.5, las curvas $\ln[E^V(R, t_{min})]$ y $E^V(R, t_{min})$ como función de R . En estas figuras, de la misma manera que en las figuras 4.22 y 4.23, también se grafican la banda de confianza y la de predicción al 95%. Con objeto de comparar la forma de estas dos curvas con las obtenidas en [219], en la figura 4.22, se incluye el ajuste a una recta de estos datos (línea a guiones verde), los cuales ya fueron presentados en la figura 4.22. En la figura 4.25 se incluye la correspondiente curva exponencial (línea a guiones verde) de la figura 4.23.

El hecho de que la recta de ajuste por mínimos cuadrados de los datos de [219] (Fig. 4.24) esté enteramente contenida dentro de la banda de confianza de nuestros datos (líneas punteadas azules), deja ver que la adaptación de la red por retro-propagación del error, tanto usando el método de entrenamiento incremental como usando el método de entrenamiento por mini-lotes, determinan estadísticamente la misma forma funcional para la dependencia del mínimo del error de la curva de validación, $E^V(R, t)$, como función de la razón de ruido a señal, R , presente en los datos. Este es un resultado interesante puesto que, en la literatura al respecto, se ha argumentado a favor tanto del método de entrenamiento incremental [44, 45, 47, 48, 109, 130, 154] como del método de entrenamiento por mini-lotes [130, 191]. No obstante, con el resultado anterior, se ve que ambos métodos proporcionan valores similares para la curva $E^V(R, t_{min})$ vs. R . Sin embar-

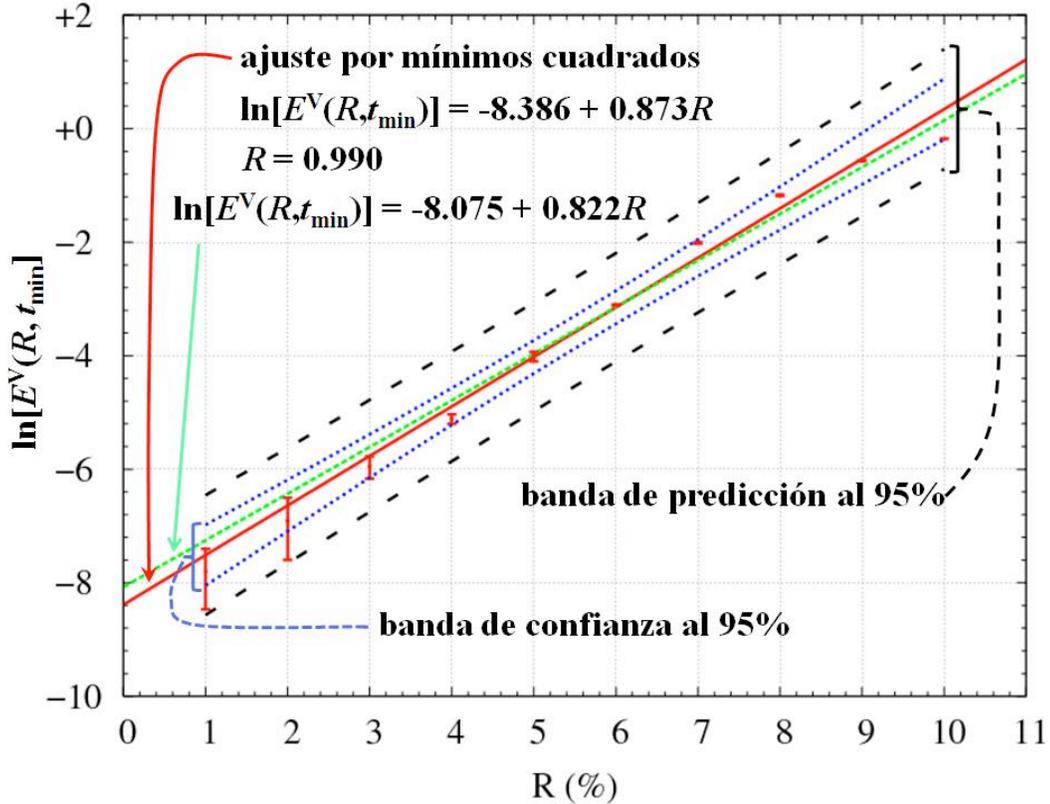


Figura 4.24: Se despliegan los datos correspondientes a los 10 valores de la razón ruido a señal analizados (en rojo), es decir, $R = 1\%$, 2% , ..., 10% . El ajuste a una recta por mínimos cuadrados (línea sólida roja) es $\ln[E^V(R, t_{\min})] = (-8.386 \pm 0.102) + (0.873 \pm 0.043)R$. El coeficiente de correlación del ajuste por mínimos cuadrados es $r = 0.990$. También se muestran la banda de confianza al 95% (líneas punteadas azules) y la banda de predicción al 95% (líneas a guiones negras). Se incluye el ajuste a una recta por mínimos cuadrados de los datos de [219] (línea a guiones verde).

go, como se dijo arriba, en el proceso de aprendizaje existe una componente de fluctuaciones al azar, la cual se debe a la naturaleza estocástica de la actualización de los pesos y los umbrales sinápticos, involucrada en el método de aprendizaje por retro-propagación del error, la cual es más fuerte en el caso del entrenamiento incremental que en el caso del entrenamiento por mini-lotes. Esto quiere decir que, aunque en promedio, ambos métodos de entrenamiento son equivalentes, es de esperarse una diferencia en las fluctuaciones que en ambos métodos se observen en torno a la forma promedio de evolución de la curva $E^V(R, t)$ vs. t para todos los valores de la razón de ruido a señal R . De hecho, en el caso del entrenamiento incremental o estocástico, se piensa que precisamente este efecto estocástico en el aprendizaje resulte en mejores soluciones. El razonamiento es el siguiente: las RNAs no lineales normalmente se espera que tengan múltiples mínimos locales de diferentes profundidades. El objetivo del entrenamiento es el de localizar uno de estos mínimos. En el entrenamiento por lotes, se descubrirá el mínimo de la cuenca en la que se inicialice el vector

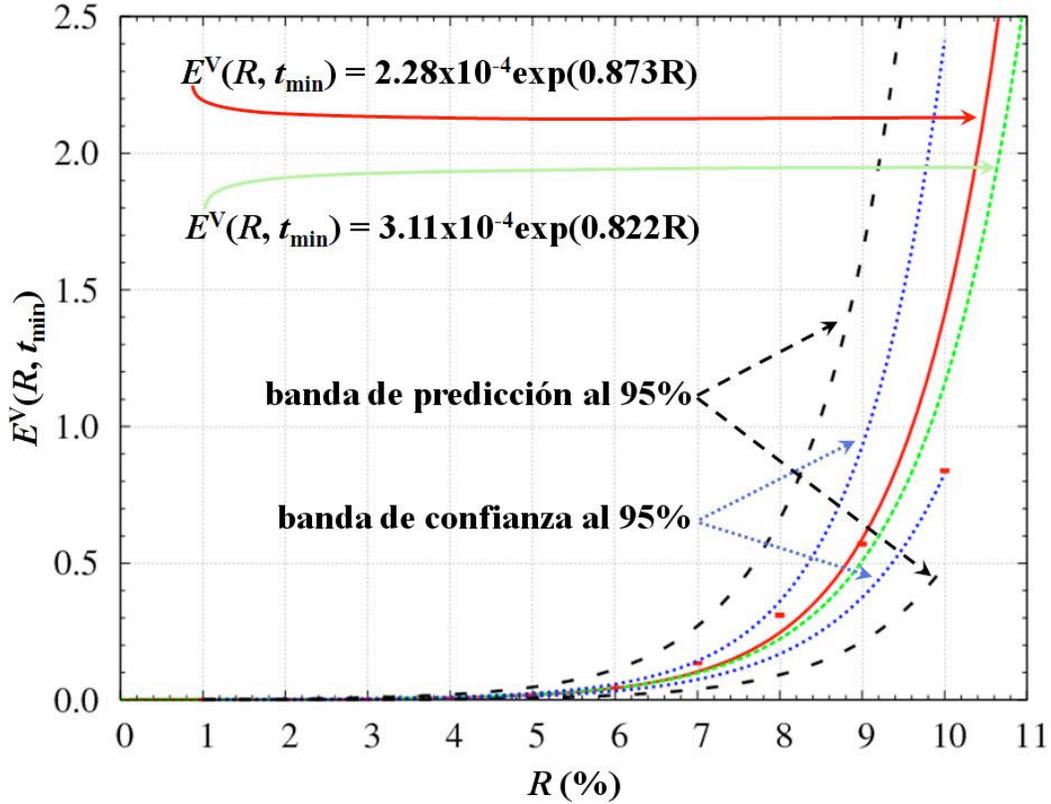


Figura 4.25: Gráfica de $E^V(R, t_{min}) = 2.28 \times 10^{-4} \exp(0.873R)$ (línea sólida roja) obtenida del ajuste por mínimos cuadrados a una recta de los datos de $\ln[E^V(R, t_{min})]$ vs. R . Se despliegan los datos corresponden a los 10 valores de la razón ruido a señal analizados, es decir, $R = 1\%$, 2% , ..., 10% (en rojo). También se muestran la banda de confianza al 95% (líneas punteadas azules) y la banda de predicción al 95% (líneas a guiones negras). Se incluye la curva exponencial correspondiente al ajuste a una recta por mínimos cuadrados de los datos de [219] (línea a guiones verde).

de pesos y umbrales, $\vec{\omega}(0)$. En el aprendizaje incremental o estocástico, el ruido presente en las actualizaciones del vector $\vec{\omega}$ puede hacer que este vector salte a la cuenca de otro mínimo local, posiblemente más profundo.

Veamos con nuestros resultados si se cumplen algunas observaciones que se plantean en la literatura. Primero, en [37, 45, 54, 107, 108, 134, 172, 191, 236] se menciona que, en el entrenamiento por lotes, el aprendizaje por retro-propagación eventualmente converge a un mínimo local, si es que existe alguno. En relación al entrenamiento incremental, estos autores mencionan que retro-propagación no converge a un punto estacionario de la superficie de error. Los autores proponen, para que en este último caso se garantice la convergencia, que la razón de aprendizaje, η , se reduzca lentamente, lo cual se ha dado en llamar *aproximación estocástica* o *recocido*. Más explícitamente, los autores proponen que, conforme a la aproximación estocástica, la condición que garantiza la

convergencia en el caso de entrenamiento por lotes, la cual también la asumimos válida para el entrenamiento por mini-lotes, es:

$$\sum_{t=1}^{\infty} \eta(t) = \infty, \quad \lim_{t \rightarrow \infty} \eta(t) = 0 \quad \text{o} \quad \sum_{t=1}^{\infty} \eta^2(t) < \infty \quad (4.11)$$

El primer requerimiento es necesario para garantizar que la red pueda alcanzar cualquier estado a partir del estado inicial $\vec{\omega}(0)$, el segundo es para evitar las fluctuaciones asintóticas.

De acuerdo a nuestros resultados para los 10 valores estudiados de la razón de ruido a señal, R , ninguna de las dos alternativas de entrenamiento, incremental o por mini-lotes, tuvo problema para converger a un mínimo, lo cual se puede constatar en las figuras 4.2 a 4.11 y en la tabla 4.5 para nuestros resultados, es decir, usando entrenamiento por mini-lotes, o en la tabla 4.6 para los datos de [219], es decir, usando entrenamiento incremental. En lo relativo a la necesidad de introducir una lenta reducción de la razón de aprendizaje para garantizar la convergencia cuando se usa aprendizaje incremental, es decir, usar la técnica de recocido, en nuestro caso no sólo no fue necesario tal plan de recocido, sino que, por el contrario, en los casos correspondientes a $R = 1\%$, 2% , 3% y 4% , debido a lo liso de las curvas del error de validación, $E^V(R, t)$ vs. t , se optó por aumentar el valor de la razón de aprendizaje, η , como ya se explicó anteriormente, lo cual aceleró considerablemente la convergencia al mínimo de la curva en todos los casos. Como se ve en las figuras 4.2 a 4.5, el incremento del valor de η no afecta al valor promedio pero si afecta a las fluctuaciones en torno a la tendencia promedio de la curva $E^V(R, t)$. Sin embargo, aunque en todos los casos el entrenamiento por mini-lotes pudo llegar a un mínimo, conviene analizar con más cuidado algunos casos. En primer lugar, el caso que corresponde a $R = 4\%$. En la tabla 4.5, se ve que para entrenamiento por mini-lotes $E^V(R = 4\%, t_{min} = 34, 140, 000) = 0.0166 \pm 0.00580$, mientras que en la tabla 4.6, se ve que para entrenamiento incremental $E^V(R = 4\%, t_{min} = 1, 874, 000) = 0.0071 \pm 0.00051$, es decir, este último valor pareciera corresponder a un mínimo más profundo. En este único caso, pareciera ser que, en efecto, el entrenamiento incremental o estocástico permitió, presumiblemente, que el vector $\vec{\omega}$ en su adaptación, tuviera la oportunidad de saltar a un mínimo más profundo que el que se alcanzó usando entrenamiento por mini-lotes. Tal vez este efecto sería más notable, si en lugar de usar entrenamiento por mini-lotes se usara entrenamiento por lotes, es decir, que el lote fuera igual al conjunto de entrenamiento. Sin embargo, en general, el entrenamiento por mini-lotes para un mini-lote de tamaño 16, como uno de los usados en este trabajo, es casi tan eficiente, en términos de su habilidad para encontrar el mínimo del error de validación, como el método de entrenamiento incremental o estocástico. Pues como se ha visto en este trabajo, sólo en el caso de $R = 4\%$ fue superado por el método incremental en cuanto a la profundidad del mínimo de la curva del error de validación o generalización.

En los casos $R = 2\%$ y 3% , resulta que los mínimos alcanzados por el entrenamiento por mini-lotes son más profundos que los alcanzados por entrenamiento incremental. Esto pareciera contradecir el argumento a favor de la mayor estocasticidad en el entrenamiento incremental, lo cual, se arguye, puede ayudar a que el vector $\vec{\omega}$, ayudado por el ruido térmico o estocástico, pueda saltar y escapar de mínimos locales en su búsqueda del mínimo global más fácilmente que en el entrenamiento por mini-lotes, en el cual la estocasticidad está reducida. En relación con esto,

conviene hacer notar que en el caso del entrenamiento por mini-lotes, las RNAs se entrenaron un total de 120,000,000 épocas equivalentes, es decir, 20,000,000 épocas usando $\eta = 0.3$ y 5,000,000 épocas usando $\eta = 6.0$. Mientras que en el caso del entrenamiento incremental la RNA de $R = 2\%$ se entrenó 5,000,000 épocas y la de $R = 3\%$, 2,000,000 épocas. Cabe la posibilidad de que si estas dos últimas RNAs se entrenaran más, podrían alcanzar un menor mínimo. El caso de $R = 1\%$ se analizará un poco más adelante, donde se verá que el valor del mínimo es el mismo en ambos métodos de aprendizaje.

Los resultados anteriores parecieran no ser consistentes con los requerimientos de los modelos de aprendizaje para redes neuronales derivados de la *teoría de aproximación estocástica* (Ec. 4.11). Sin embargo, es conveniente notar que, la estocasticidad de éstos modelos tiene que ver con el mecanismo de aprendizaje, el cual, por ejemplo, para retro-propagación (Ecs. 3.29 y 3.31) contiene un elemento aleatorio en cuanto a la selección al azar del patrón, o patrones en el caso de entrenamiento por mini-lotes, que se le presentan a la RNA, en su capa de entrada, para que esta sea actualizada durante el entrenamiento. Es esta estocasticidad la que se explota en los modelos de la aproximación estocástica. Pero no se debe olvidar que los patrones o CBs que se usan en este trabajo, pertenecen a ambientes de información ruidosos, con valores de razón de ruido a señal $R = 1\%$, 2% , ..., 10% . De hecho, es este ruido el que determina, tanto para el entrenamiento incremental como para el entrenamiento por mini-lotes, la forma de las curvas $E^V(R, t)$ vs. t (Figs. 4.23 y 4.25), las cuales, como ya se vio arriba, coinciden dentro de los límites de significación estadística. La forma de manifestarse de ambas fuentes de estocasticidad es mediante fluctuaciones en las correspondientes curvas $E^V(R, t)$ vs. t . Claro, la componente de las fluctuaciones debido a un valor finito de la razón de aprendizaje η , tienen el efecto positivo de implicar, en la medida de su tamaño, una aceleración de la convergencia hacia el valor óptimo del vector de pesos y umbrales $\vec{\omega}$. Mientras que la componente de las fluctuaciones debidas a un valor no nulo de R no tienen ese efecto. Esto es, en el primer caso, un paso grande en la búsqueda del mínimo, η grande, implica fluctuaciones de la curva $E^V(R, t)$ vs. t , por lo cual, en la medida que la búsqueda se acerca al mínimo, el valor de η se debe hacer cada vez menor, para minimizar las correspondientes fluctuaciones, y lograr una determinación del valor óptimo de $\vec{\omega}$ más precisa, aunque eso implique la reducción de la rapidez de avance en el espacio de los pesos y umbrales $\vec{\omega}$. En el segundo caso, el ruido presente en las señales o CBs presentadas a la RNA para su entrenamiento, no tiene ninguna implicación sobre la rapidez de avance del punto representativo de la RNA en el espacio de los pesos y umbrales $\vec{\omega}$, sin embargo, hace que el punto $\vec{\omega}$ representativo de la red siga una trayectoria fluctuante en su búsqueda del mínimo, lo cual, aunque hace el proceso de búsqueda del valor óptimo de $\vec{\omega}$ menos preciso, pudiera tener el efecto benéfico de permitir que la RNA pueda escapar de mínimos locales debido a las fluctuaciones que se originan por el ruido de las CBs presentadas a la red para su adaptación.

A continuación, se presentarán algunos comentarios que confirman la necesidad, en nuestro caso, de recurrir al uso de la *aproximación estocástica o recocido*. El hecho de que en el caso de entrenamiento incremental no haya sido necesario hacer uso de un plan de recocido, es decir, hacer tender la razón de aprendizaje η a cero, para que la curva de error de generalización $E^V(R, t)$ alcance su mínimo, es probable que se deba a que el valor inicial de η era suficientemente pequeño

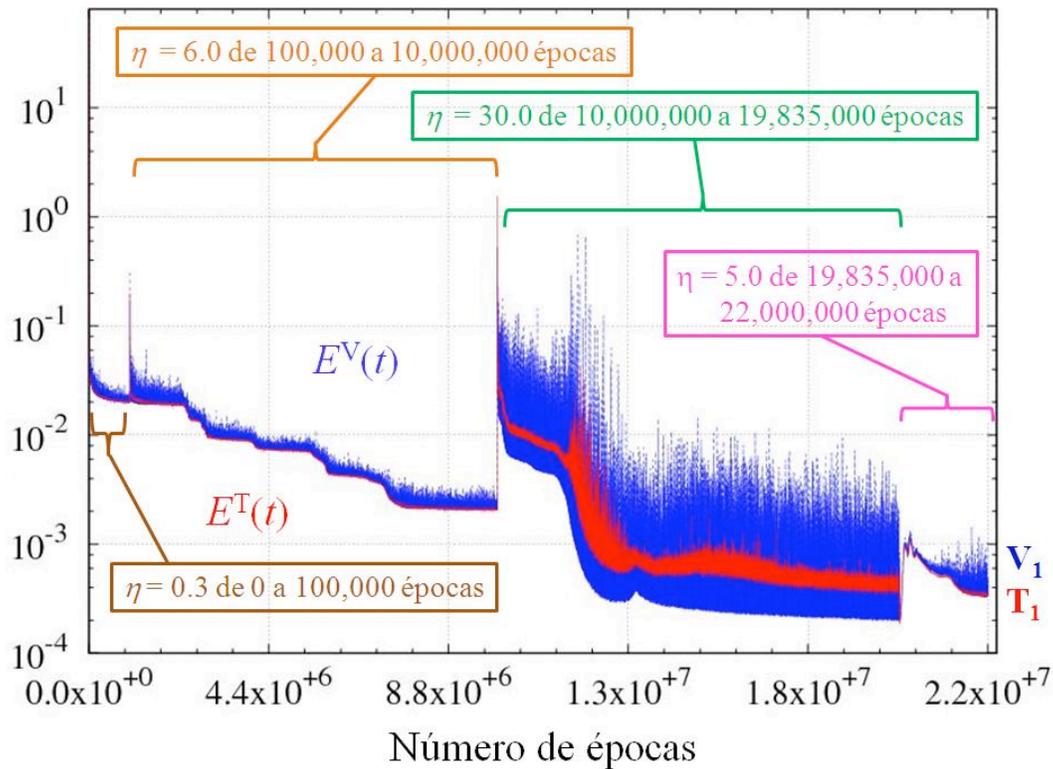


Figura 4.26: Curvas del logaritmo del error de validación o generalización $E^V(R = 1\%, t)$. El valor de la razón de aprendizaje η se cambió de 30.0 a 5.0 a partir de 19,835,000 épocas. Se ve claramente la disminución del tamaño de las fluctuaciones a partir de este cambio.

como para permitir que el error de generalización alcance su mínimo. De hecho, en los casos $R = 1\%$, 2% , 3% y 4% , una vez que el proceso de entrenamiento lleva a valores de $\vec{\omega}$ para los cuales el gradiente es suficientemente pequeño, entonces es factible incrementar el valor de η para acelerar la convergencia al mínimo. El costo de esto es, conforme a la teoría de aproximación estocástica, que el tamaño de las fluctuaciones de las respectivas curvas del error de generalización aumenta, sin que esto impida, en estos casos, alcanzar el mínimo. Esto se puede observar en las figuras 4.2, 4.3, 4.4 y 4.5 correspondientes a $R = 1\%$, 2% , 3% y 4% . Lo que esto significa es que el efecto de la estocasticidad inherente tanto al entrenamiento incremental como, en menor grado, al entrenamiento por mini-lotes, introduce fluctuaciones en la evolución de la curva del error de generalización, pero no altera la tendencia promedio, la cual, en el aprendizaje por retropropagación, está determinada por el valor del gradiente de la función de error (Ecs. 3.29 y 3.31).

Con objeto de observar como estas fluctuaciones térmicas o estocásticas son menos prominentes en el caso del entrenamiento por mini-lotes en comparación con el entrenamiento incremental, se reentrenó la RNA correspondiente a entrenamiento por mini-lotes y a un tamaño de ruido $R = 1\%$. Dado que en este caso la RNA alcanza su mínimo después de ser entrenada 1,000,000 épocas con

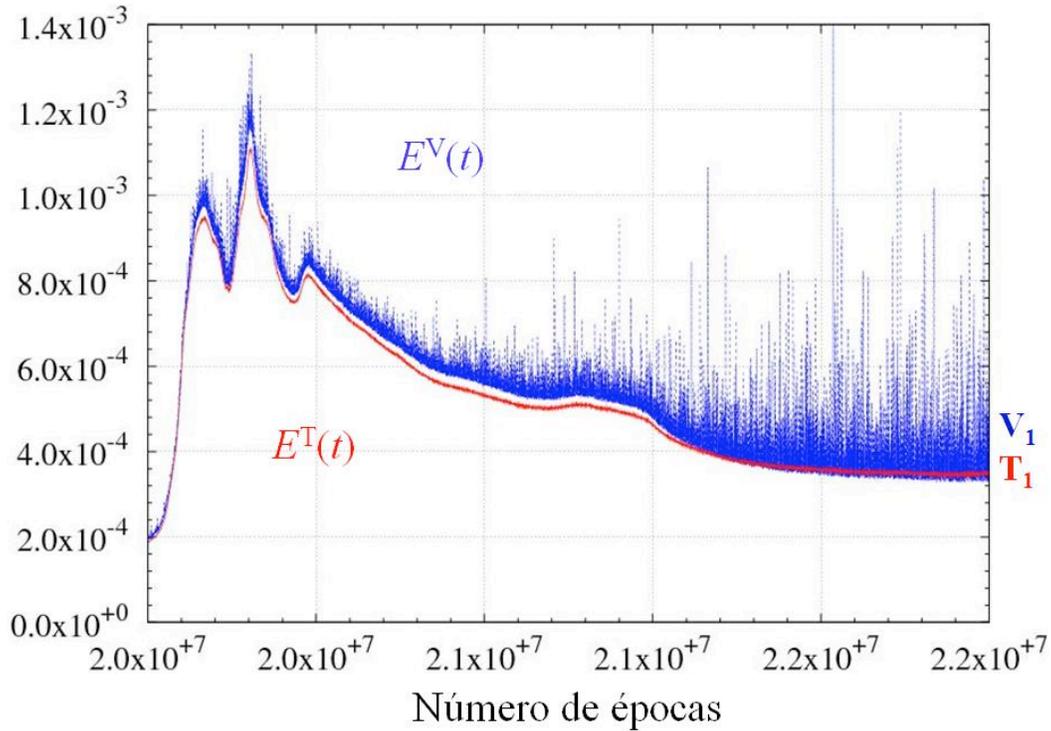


Figura 4.27: Curvas del error de validación o generalización $E^V(R = 1\%, t)$ a partir de 19,835,000 épocas. En esta gráfica sólo se muestra el aprendizaje con $\eta = 5.0$ con el propósito de mostrar más claramente el comportamiento de la curva $E^V(R = 1\%, t)$ con dicho valor.

$\eta = 0.3$ más 9,000,000 épocas con $\eta = 6.0$ más 10,235,000 épocas con $\eta = 30$, es decir, un total de 20,235,000 épocas, entonces se decidió recomenzar el reentrenamiento 400,000 épocas antes de este número, esto es, después de 19,835,000 épocas. En la figura 4.26 se presenta la curva del error de validación o generalización y en la figura 4.27 se muestra la misma curva en la fase final donde $\eta = 5.0$, con el propósito de percibir mejor el comportamiento de la misma.

Como se puede ver en la figuras 4.26 y 4.27, después de 2,165,000 épocas de entrenamiento con $\eta = 5.0$, la curva $E^V(R = 1\%, t)$ aún no ha alcanzado su valor mínimo pero, presumiblemente, se encuentra cerca. No obstante aún se observa una ligera tendencia decreciente de la curva. Dado que el propósito es estimar el tamaño de las fluctuaciones respecto a la tendencia local, se decidió estimar éstas a partir de una línea base ajustando una recta a la curva $E^V(R = 1\%, t)$ en el intervalo de 21,760,000 a 22,000,000 épocas. El valor de la fluctuación respecto a esta línea base es de 0.000023 y el valor promedio de los errores en ese mismo intervalo es de 0.00035. Los valores reportados en la tabla 4.5 correspondientes a un valor de $\eta = 30.0$ son 0.00020 y 0.00041 respectivamente. Vemos que los valores promedio son muy similares, pero el valor de la fluctuación disminuyó apreciablemente, casi un factor de 9, al cambiar la razón de aprendizaje de $\eta = 30.0$ a $\eta = 5.0$. Lo cual está de acuerdo con lo que predice teoría de aproximación estocástica. También

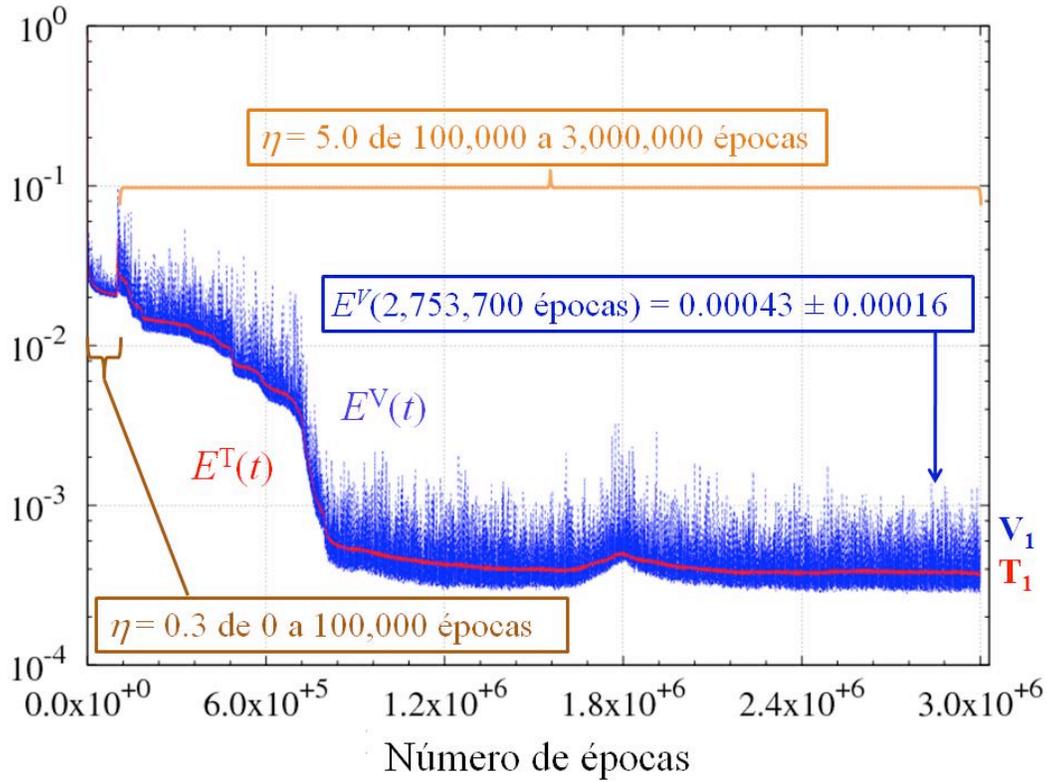


Figura 4.28: Curvas del logaritmo del error de entrenamiento (roja) y de validación (azul) para $R = 1\%$ para el caso de entrenamiento incremental de la referencia [219]. Se continuó el entrenamiento de la correspondiente curva de 900,000 a 3,000,000 épocas. El valor de η se cambió de 0.3 a 5.0 después de las primeras 100,000 épocas de entrenamiento.

podemos comparar estos dos valores con los correspondientes al caso de entrenamiento incremental para $R = 1\%$ reportados en la tabla 4.6, los cuales corresponden a $\eta = 5.0$. El valor de la fluctuación para este caso es 0.00026 y el valor promedio de los errores es de 0.00063. Como predice la teoría de aproximación estocástica el valor de la fluctuación es mayor para el caso del entrenamiento incremental por ser mayor su estocasticidad en comparación a la correspondiente al entrenamiento por mini-lotes. Estos valores se resumen en la tabla 4.7.

Dado que la curva de error de generalización $E^V(R, t)$ de la referencia [219] correspondiente a $R = 1\%$ y al entrenamiento incremental muestra una ligera tendencia decreciente, en este trabajo decimos continuar el entrenamiento incremental de la correspondiente RNA a partir de 900,000 épocas hasta un total de 3,000,000 épocas. Con esto se determinó que el mínimo queda ubicado en $E^V(2,753,700 \text{ épocas}) = 0.00043 \pm 0.00016$. Estos valores se incluyen en la tabla 4.7 para su comparación con los otros resultados. Aún en estas condiciones, el tamaño de las fluctuaciones correspondientes al entrenamiento por lotes y con una razón de aprendizaje $\eta = 5.0$ es el menor. Sin embargo, el valor del mínimo es esencialmente el mismo. La correspondiente curva $E^V(R, t)$

	Promedio	Fluctuación
Entrenamiento mini-lotes ($\eta = 30.0$)	4.1×10^{-4}	2.0×10^{-4}
Entrenamiento mini-lotes ($\eta = 5.0$)	3.5×10^{-4}	2.3×10^{-5}
Entrenamiento incremental ($\eta = 5.0$)	6.3×10^{-4}	2.6×10^{-4}
Continuación entrenamiento incremental ($\eta = 5.0$)	4.3×10^{-4}	1.6×10^{-4}

Tabla 4.7: Comparación de los valores promedio y de las fluctuaciones en los correspondientes mínimos para dos casos de entrenamiento por mini-lotes y dos casos de entrenamiento incremental. Todos los casos corresponden a un valor de la razón de ruido a señal $R = 1\%$. Los datos de entrenamiento incremental corresponden a la referencia [219].

vs. t se despliega en la figura 4.28.

Una vez hechas estas consideraciones, se puede concluir que los resultados presentados en este trabajo están de acuerdo con las predicciones de la *teoría de aproximación estocástica*.

Finalmente, en la tabla 4.5, es interesante observar que el módulo del vector de pesos $\omega_{min} = |\vec{\omega}_{min}|$ correspondiente al mínimo de la curva del error de generalización muestra una tendencia decreciente a medida que la razón de ruido a señal R crece. Esto pudiera deberse a que a medida que decrece R , es decir, el ruido, el número de características que la RNA correspondiente puede aprender de los datos de entrada aumenta, puesto que el ruido decreciente cada vez permite a la RNA aprender características que antes (valores mayores de R) eran ocultadas por el ruido. De hecho, esta es la razón por la cual se obtiene una dependencia exponencial para la función $E^V(R, t_{min})$ vs. R (Figs. 4.22 y 4.24. Esto es consistente con la creencia generalizada de que un problema más complejo (aprender más características de los patrones de entrada debido a un menor ruido) requiere de un vector de pesos de mayor tamaño.

Capítulo 5

Conclusiones

En este trabajo se presenta una aplicación novedosa de análisis digital de forma de pulsos en relación a la identificación de curvas de Bragg sintéticas. Este ADFP es asistido por el uso de RNAs en su calidad de buenas reconocedoras de patrones. Los resultados obtenidos son muy alentadores en el sentido de que esta alternativa digital permita desarrollar la espectroscopia de CB. Las tareas planteadas a las RNAs en este trabajo, fueron desempeñadas satisfactoriamente, inclusive hasta el caso límite de una razón de ruido a señal $R = 10\%$, el cual es muy raro que se dé, por ejemplo, en los laboratorios de física nuclear que hacen uso de la espectroscopia de Bragg para el estudio de reacciones nucleares.

Para continuar el desarrollo de esta alternativa de ADFP aplicada a la espectroscopia de CB se requiere, como una tarea para el futuro que, en la modelación de CB sintéticas, se incluyan algunos efectos que incidan en la determinación de la forma de las CBs. Por ejemplo, la fluctuación de la línea base respecto a la cual se determinan las CBs. Es bien sabido que los sistemas de detección electrónicos, por más cuidado que se tenga en la rectificación de sus voltajes de alimentación, padecen de una componente residual proveniente de la variación periódica proveniente del voltaje de alimentación de 60 ciclos. Estas fluctuaciones de la línea base, a las escalas de tiempos relevantes a la duración de una curva de Bragg (menos de $10 \mu s$), se ven como fluctuaciones de la línea base. Estas fluctuaciones de la línea base modifican el área bajo las CBs, es decir, la energía del ion frenado, así como la amplitud del máximo o pico de Bragg.

Otros efectos presentes y que también pueden cambiar la forma de las CBs son, por ejemplo, una muy lenta deriva, la cual puede ser oscilatoria, de los voltajes de alimentación del sistema electrónico de amplificación usado. Otra razón por la cual las formas de las CBs pueden cambiar lentamente se debe a que, los espectrómetros de CB, se operan bajo una atmosfera controlada de un gas de frenado, el cual es renovado continuamente para garantizar su pureza (ausencia de trampas de cargas como, por ejemplo, vapor de agua). Con el tiempo, esto hace que la atmosfera de gas del ECB se vaya liberando cada vez más de trampas de iones, lo cual aumenta la velocidad de deriva de los electrones en el ECB, modificando las formas de las CBs lentamente.

Una tarea ambiciosa para explorar en el futuro es la de intentar llevar a cabo el reconocimiento de CBs aun en los casos en que, por lo pequeño de la energía inicial del ion detectado, la CB

correspondiente no alcanza su pico de Bragg. Esta es una tarea que, en principio, es factible pero cuya implementación práctica es realmente complicada. Sin embargo, con los elementos que brinda el ADFP es tentador intentarlo, puesto que ampliaría la ECB por el lado de las energías pequeñas, lo cual sería muy beneficioso para todos aquellos que estudian reacciones nucleares con iones pesados a bajas energías.

Considerando estos efectos y otros adicionales, es posible plantear un refinamiento aún mayor del ADFP aplicado a la ECB, sin embargo, también es bueno pensar en términos del desarrollo de técnicas más poderosas de ADFP asistida por RNA. Por ejemplo, es posible recurrir al uso de técnicas de regularización, es decir, a la modificación de la función de error empleada, mediante la adición a ésta de un término extra, de tal manera que la minimización de esta nueva función de error, que es lo que se lleva a cabo al entrenar a la RNA, permita encontrar una solución mejor, es decir, con un menor error y con una mejor generalización.

Existe también la interrogante de, dado un problema a resolver, cuál es la arquitectura óptima de la RNA usada. En este trabajo, se hizo un estudio de entrenamiento por mini-lotes que se comparó con un estudio similar [219], el cual empleó el método de entrenamiento incremental. Por tal motivo, se hizo uso de la misma arquitectura de RNAs que las empleadas en [219], en donde, para la determinación de la arquitectura, se llevó a cabo un intenso trabajo de prueba y error, lo cual es muy común en la literatura al respecto. Un desarrollo ambicioso en esta dirección sería el usar algoritmos genéticos para evolucionar ensambles de RNAs y determinar, de esta manera, la arquitectura óptima de la RNA requerida para la solución un problema dado.

Es importante hacer notar que los resultados obtenidos en este estudio de ADFP asistido por RNAs y aplicado a la ECB son, en términos generales, consistentes con las predicciones propuestas por la teoría de aprendizaje estocástico. Un punto que, entre otros, queda pendiente, sería el estudio de la relación de las diferentes soluciones encontradas ($\vec{\omega}_{min}(R)$), es decir, las soluciones encontradas para cada uno de los diez problemas estudiados: $R = 1\%$, 2% , ..., 10% . En la tabla 4.5, se puede ver que los diferentes vectores solución $\vec{\omega}_{min}(R)$ son diferentes, es decir, difieren considerablemente en sus módulos. Otro punto, el cual, a nuestro juicio es muy importante, es que, hasta donde sabemos, en la literatura no existe un análisis teórico que prediga la forma exponencial de la curva $E^V(R, t_{min})$ vs. R , la cual fue determinada experimentalmente en [219] para entrenamiento incremental (Figs. 4.22 y 4.23) y verificadas, en este trabajo, para entrenamiento por mini-lotes (Figs. 4.24 y 4.25), coincidiendo estadísticamente ambas dependencias funcionales.

Finalmente, es conveniente remarcar que los desarrollos aquí logrados en el campo del ADFP asistido por RNA son de aplicabilidad general, es decir, no se restringen, como en nuestro caso, a la espectroscopia de curva de Bragg. Esto es, estos avances pueden ser usados cuando se desee o se requiera enfrentar un problema de reconocimiento de patrones que se hayan digitalizado para su estudio, lo cual es un problema muy común en laboratorios que hacen uso de sistemas de adquisición de datos en los cuales se digitalizan las señales estudiadas.

Glosario

ADC Un convertidor analógico digital o ADC por sus siglas en inglés es un dispositivo que convierte señales (voltajes o corrientes) continuas a números digitales discretos.

Alcance Las partículas cargadas, en su viaje a través de la materia, ionizan las moléculas de la materia y van perdiendo su energía cinética en muchas de estas colisiones, hasta que finalmente agotan su energía. La distancia del punto inicial al punto final se denomina alcance de la partícula. El alcance depende del tipo de partícula, de su energía inicial y del material por el cual se desplaza la partícula.

Apilamiento de pulsos (Pulse pile-up) Ocurre cuando algunos pulsos llegan más cercanos en el tiempo que el tiempo de resolución de pulsos del sistema, de tal suerte que el sistema no puede medir correctamente la altura del pulso. Si los pulsos están muy cerca en el tiempo, el sistema simplemente registrará los dos pulsos como un único evento con una altura de pulso combinada (Pulso1 + Pulso 2), lo cual se conoce como apilamiento pico (peak pile-up). Si los dos pulsos están separados un poco más, el sistema puede aceptar ambos eventos pero registrar sus alturas incorrectamente, esto se conoce como apilamiento de cola (tail pile-up). En cualquier caso el pulso terminará en un canal de energía incorrecto, contaminando el espectro.

Aprendizaje Es un proceso por el cual los parámetros libres de un sistema adaptivo (red neuronal) toman valores óptimos a través de un proceso continuo de estimulación por un ambiente determinado en el cual el sistema está embebido.

Banda de confianza Si a partir de datos con ruido se construye una estimación puntual $\hat{f}(x)$ del valor $f(x)$ que toma una función en el punto x , entonces, la banda de confianza, $\hat{f}(x) \pm w(x)$ al $(1 - \alpha)\%$ satisface, para cada x :

$$Pr[\hat{f}(x) - w(x) \leq f(x) \leq \hat{f}(x) + w(x)] = 1 - \alpha$$

Banda de predicción Si a partir de datos con ruido se construye una estimación puntual $\hat{f}(x)$ del valor $f(x)$ que toma una función en el punto x , entonces, la banda de predicción, $\hat{f}(x) \pm w(x)$ al $(1 - \alpha)\%$ satisface, para cada x :

$$Pr[\hat{f}(x) - w(x) \leq y \leq \hat{f}(x) + w(x)] = 1 - \alpha$$

donde y representa un valor del proceso observado que corresponde al valor x , el cual es independiente de los datos usados tanto para construir el estimador puntual $\hat{f}(x)$ como para construir el intervalo de confianza $w(x)$.

Cadena de Markov Una cadena de Markov es un proceso estocástico discreto para el cual el comportamiento aleatorio de una cadena de variables (sucesión de variables) queda determinado únicamente por la dependencia entre variables aleatorias sucesivas, es decir, el valor aleatorio de la variable estocástica X_{n+1} sólo depende del valor de la variable estocástica X_n . Se suele decir que esto equivale a que el sistema no tenga memoria.

Cámara de ionización Es un detector gaseoso en el cual su voltaje de operación (usado para coleccionar los portadores de carga) no da lugar a que se produzca multiplicación de carga. La señal de salida es proporcional a la energía que la partícula disipa en el detector, posibilitando la medición de la energía de la partícula. Se emplean para la medición de partículas de fuerte ionización: alfas, protones, iones pesados y fragmentos de fisión.

Centellador Es un material que exhibe luminiscencia cuando por él pasa radiación ionizante (electrones, positrones u otras partículas o iones más pesados). Esto es producido porque el material absorbe parte de la energía de la partícula incidente y la re-emite en forma de un corto destello de luz, típicamente en el rango de la luz visible. Si esta re-emisión es rápida (menos de 10^{-8} s), el fenómeno se conoce como fluorescencia. Si la excitación es metaestable y dura de microsegundos a horas, se denomina fosforescencia. La dependencia temporal del número $N(t)$ de fotones de centelleo emitidos en un solo evento se puede describir en muchas ocasiones como una superposición de uno o dos decaimientos exponenciales. Para el caso bastante común de dos decaimientos se tiene:

$$N(t) = A \exp\left(-\frac{t}{\tau_f}\right) + B \exp\left(-\frac{t}{\tau_s}\right)$$

donde τ_f y τ_s son las constantes de decaimiento rápida (pronta) y lenta (retardada) respectivamente. Generalmente, la componente rápida domina, las amplitudes relativas A y B de las dos componentes dependen del material centellador. También ocurre que ambas componentes pueden depender de la razón de pérdida de energía dE/dx de la partícula incidente. Cuando esta dependencia con dE/dx es fuerte, la dependencia temporal del decaimiento varía apreciablemente con el tipo de partícula incidente. Esos centelladores permiten la implementación de la discriminación por forma de pulso, es decir, la identificación de partículas con base en las características de decaimiento del pulso eléctrico proveniente del tubo fotomultiplicador acoplado al centellador. Por ejemplo, en el caso del centellador BaF_2 , los rayos γ excitan típicamente la componente rápida, mientras que las partículas α excitan la componente lenta, permitiendo la identificación de ellas en base a el tiempo de decaimiento de la señal del tubo fotomultiplicador.

Conexionismo Es un enfoque común a diversas disciplinas: inteligencia artificial, ciencias cognitivas, psicología cognitiva, neurociencia y filosofía de la mente, las cuales modelan ya sea los fenómenos mentales o los conductuales como procesos emergentes de redes interconectadas

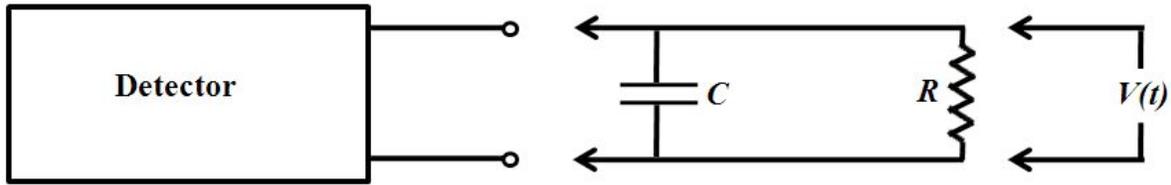
constituidas por elementos sencillos. Los modelos de redes neuronales son los ejemplos más comunes de sistemas conexionistas.

Contador proporcional Es un tipo de detector gaseoso en el cual una partícula ionizante que entra en él con una energía suficiente, libera electrones de los orbitales de los átomos del gas, dejando atrás un electrón y un átomo cargado positivamente, a los cuales se les denomina como par iónico. A medida que la partícula ionizante se desplaza por la cámara del detector, deja una estela de pares de iones a lo largo de su trayectoria. Los electrones creados en este proceso son arrastrados hacia el electrodo de lectura, conocido como ánodo, bajo la influencia de un campo eléctrico aplicado. Simultáneamente, a una mucha menor velocidad, los iones positivos son arrastrados por el mismo campo eléctrico hacia el cátodo. En dispositivos prácticos, los correspondientes tiempos de arrastre o de deriva se miden en microsegundos y milisegundos. Un contador proporcional se diferencia de una cámara de ionización en cuanto a que el voltaje de operación del primero es suficientemente grande como para que los electrones de deriva ganen suficiente energía a lo largo de un camino medio libre como para crear nuevos pares de iones en sus colisiones con otros átomos neutros del gas. Los nuevos electrones así creados son arrastrados hacia el ánodo, pudiendo crear ellos mismos nuevos pares iónicos, de tal suerte que se crea una cascada de pares iónicos, denominada avalancha Townsend. Si el voltaje de operación se escoge adecuadamente, cada proceso de avalancha ocurre independientemente de las otras avalanchas simultáneas originadas por el mismo evento ionizador inicial. Así, no obstante que el número total de electrones liberados puede crecer exponencialmente con la distancia, la cantidad total de carga creada permanece proporcional a la cantidad de carga liberada en el evento original. Si se mide la carga total (la integral de la corriente eléctrica del ánodo), se puede determinar la energía cinética de la radiación ionizante, ya que el número total de pares iónicos creados por la partícula ionizante cargada incidente es proporcional a su energía.

Detector La función de un detector es la de producir una señal para cada partícula que entra en él. Cada detector opera usando alguna interacción de las partículas con la materia. En un modelo muy sencillo de un detector prototipo ocurre que como respuesta a una partícula o cuanto de radiación se produce en el detector una corriente $i(t)$ que fluye por un tiempo t_c igual al tiempo de recolección de la carga Q producida por la radiación. La integral sobre el tiempo t_c de la corriente $i(t)$ debe ser igual a Q . Los detectores se pueden clasificar en dos grandes grupos:

- i) Sistemas que operan en modo de pulso. Su salida consiste de pulsos de voltaje, un pulso por cada partícula detectada.
- ii) Sistemas que operan en modo de corriente. Su salida es un valor promedio, que resulta de la detección de muchas partículas.

Cuando se opera en modo de pulso, la forma de los pulsos o señales producidas por un evento particular dependen de las características del circuito al cual el detector es conectado, que generalmente, es un preamplificador. El circuito equivalente se puede representar como:



R representa la resistencia de entrada del circuito, y C representa la capacitancia equivalente del detector mismo más la del circuito medidor. Si el detector estuviera acoplado a un preamplificador, R sería la resistencia de entrada del preamplificador y C la suma de las siguientes capacitancias: la del cable usado para conectar el detector al preamplificador más la capacitancia de entrada del preamplificador. En la mayoría de los casos el voltaje dependiente del tiempo $V(t)$, a través de la resistencia de carga R , representa la señal de voltaje fundamental sobre la cual se basa la operación en modo de pulso. Dependiendo del valor de la constante de tiempo del circuito medidor, $\tau = RC$, en relación al tiempo de recolección de carga, t_c , en el detector, se distinguen dos casos límite de operación, es decir, $\tau \ll t_c$ y $\tau \gg t_c$. En el caso $\tau \ll t_c$, la constante del circuito medidor es mucho menor que el tiempo de recolección de carga en el detector, así que la corriente que fluye por la resistencia de carga R es esencialmente igual al valor instantáneo de la corriente que fluye en el detector. Bajo estas condiciones, la señal de voltaje $V(t)$ tiene una forma prácticamente idéntica a la dependencia temporal de la corriente producida en el detector. En el otro caso límite $\tau \gg t_c$, la constante de tiempo del circuito medidor es mucho mayor que el tiempo de recolección de carga en el detector. En este caso, muy poca corriente fluirá en la resistencia de carga durante el tiempo de recolección de carga y la corriente del detector será integrada momentáneamente en la capacitancia C . Si se asume que el tiempo entre pulsos consecutivos es suficientemente largo, la capacitancia C se descargará a través de la resistencia R , retornando el voltaje a través de la resistencia de carga a cero.

Detector semiconductor Son dispositivos de estado sólido que en esencia operan como una cámara de ionización. Los portadores de carga en los semiconductores no son electrones y iones, como en un contador de gas, sino electrones y agujeros. Los detectores semiconductores más exitosos están hechos de silicio y de germanio pero existen otros materiales que se pueden usar.

Doble decaimiento beta Es un proceso en el que dos neutrones de un núcleo atómico decaen a dos protones, dos electrones y dos neutrinos de electrón. Para algunos núcleos, como el de ^{76}Ge , sucede que el núcleo con un número atómico una unidad más grande pero con el mismo número de masa, ^{76}As , tiene una energía de amarre menor, lo cual impide que se produzca decaimiento beta a él. Sin embargo, el núcleo con número atómico dos unidades más grande y con el mismo número de masa, es decir, el ^{76}Se , tiene una energía de amarre mayor, lo cual permite que se dé el proceso de doble decaimiento beta a él. Si el neutrino fuera un fermión del tipo Majorana, es decir, su propia antipartícula, entonces sería posible observar el doble decaimiento beta sin emisión de neutrino.

Drip line (línea de goteo) De acuerdo a la física nuclear, un exceso de neutrones en el núcleo atómico producirá en éste un “goteo” de los neutrones de exceso. Por lo tanto, en una carta de núclidos, la línea de goteo de neutrones representa la frontera de estabilidad para los núcleos ricos en neutrones. Por oposición, la línea de goteo de protones representa la frontera para la estabilidad del núcleo ante un exceso de protones.

Epifenómeno Término aplicable para designar la conciencia como reflejo pasivo del contenido material (o ideal) del mundo.

Espectroscopia En términos generales es la medición de alguna cantidad como función ya sea de la frecuencia o de la energía. La gráfica de la cantidad medida contra la frecuencia o la energía se conoce como espectro. El instrumento usado para tales medidas se denomina espectrómetro.

Fermion Uno de los dos tipos fundamentales de partículas de la naturaleza que obedecen la estadística de Fermi-Dirac y tienen spin semi-entero. En contraposición, se tienen los bosones, los cuales son partículas que obedecen la estadística de Bose-Einstein y tienen spin entero. Característico de los fermiones es que sólo un fermion puede ocupar un estado cuántico a un momento dado, lo cual se conoce como Principio de Exclusion de Pauli.

Flash ADC Es un convertidor analógico digital muy rápido en el cual la señal analógica se compara simultáneamente con $2n - 1$ voltajes de referencia diferentes, donde n es la resolución del ADC.

Fotodiodo Es un detector de fotones capaz de convertir luz en corriente o voltaje, dependiendo del modo de operación.

Fermion de Majorana o de Dirac En física de partículas, los fermiones pueden ser su propia antipartícula, en cuyo caso se denominan fermiones de Majorana. La otra posibilidad, es decir, que no sean su propia antipartícula da lugar a los fermiones de Dirac. Actualmente no se conocen fermiones de Majorana.

Impulso nervioso Es una onda de naturaleza eléctrica que recorre toda la neurona y que se origina como consecuencia de un cambio transitorio de la permeabilidad en la membrana plasmática, secundario a un estímulo.

Inteligencia artificial simbólica También conocida como inteligencia artificial tradicional, se distingue por dos características. 1) Subdivide el problema dado, de manera recursiva, en una serie de sub-problemas, los cuales supuestamente, son más fáciles de resolver. 2) Se basa en el conocimiento, es decir, descansa en una descripción simbólica del mundo, como por ejemplo, un conjunto de reglas.

Mitocondria Es un organelo encargado de suministrar la mayor parte de la energía para la actividad celular, para lo cual sintetizan adenosín trifosfato (ATP del inglés adenosine triphosphate) por medio de la fosforilación oxidativa.

Montículo del axón Es el segmento que conecta al axón con el soma.

Neurotransmisor Es una biomolécula, sintetizada generalmente por las neuronas, que se vierte, a partir de vesículas existentes en la neurona presináptica, hacia la brecha sináptica, produciendo un cambio en el potencial de acción de la neurona postsináptica.

Phoswich Un detector phoswich (phosphorus sandwich) es una combinación de centelladores con diferentes características de forma de pulso que están acoplados ópticamente entre sí lo mismo que a un fotomultiplicador común. Usando técnicas de análisis de forma de pulsos, se pueden distinguir las señales de los dos centelladores, identificando el detector en el que ocurrió el evento. Los detectores phoswich fueron desarrollados para detectar rayos gamma y rayos X de baja intensidad y de baja energía, así como partículas alfa y beta eficientemente en un ambiente de fondo de alta energía. Algunos detectores se diseñan para medir e identificar todas las energías en forma simultánea.

Potencial de acción Es una onda de descarga eléctrica que viaja a lo largo de la membrana celular. Constituyen la vía fundamental de transmisión de códigos neurales.

Potencial de membrana Es la diferencia de potencial eléctrico entre la parte interna y externa de la membrana plasmática de una célula (por lo general -70 mV). La membrana de las células está polarizada debido a que hay un reparto desigual de cargas eléctricas entre el interior y el exterior de la célula. Esto crea una diferencia de potencial, siendo el exterior positivo respecto al interior.

Procesos estocásticos y deterministas Se distinguen procesos estocásticos y procesos deterministas. Los procesos estocásticos consisten de variables aleatorias. Los procesos deterministas consisten de variables que siempre darán el mismo resultado para una misma condición inicial.

Punto fijo estable En relación a una función, es un punto que es mapeado a él mismo por la función. Esto es, si x es un punto fijo de la función f , entonces $f(x) = x$. Si punto del espacio de estados de un sistema dinámico es un punto fijo estable o de equilibrio, entonces pequeñas perturbaciones darán lugar a un movimiento localizado, por ejemplo, pequeñas oscilaciones.

Radiación de fondo En virtud de que la radiación cósmica continuamente bombardea la atmósfera terrestre y de la existencia de la radioactividad natural ambiental, todos los detectores de radiación producen una señal de fondo. La naturaleza de esta señal de fondo varía grandemente con el tamaño y el tipo de detector y con el tamaño y características del blindaje del detector.

Ruido electrónico Es cualquier tipo de interferencia que tienda a “enmascarar” la cantidad que se desea observar. Generalmente, resulta del movimiento térmico de los portadores de carga en las componentes del sistema de detección (cables, resistores, el mismo material detector, etc.), el cual se manifiesta como un gran número de pulsos de pequeño nivel. No se debe confundir el ruido electrónico con pulsos de la radiación de fondo que se originan de fuentes de radiación que siempre se encuentran presentes, como, por ejemplo, los rayos cósmicos.

Receptor celular Es alguna proteína que permite la interacción de determinadas sustancias con los mecanismos del metabolismo celular. Los receptores son proteínas o glicoproteínas presentes en la membrana plasmática, en las membranas de los organelos o en el citosol celular, a las que se unen específicamente otras sustancias químicas llamadas moléculas señalizadoras, como las hormonas y los neurotransmisores.

Sistema dinámico Es un espacio de estados S y una regla de evolución R , la cual, para un intervalo de tiempo T , tales que $R : S \times T \rightarrow S$ describe la evolución temporal del sistema S . Un sistema dinámico se puede visualizar como un modelo que describe la evolución temporal de un sistema.

Vesícula celular Es un organelo en forma de un pequeño saco intracelular cerrado, limitado por una membrana que almacena, digiere o transporta productos y residuos dentro de una célula.

Bibliografía

- [1] C. E. Aalseth y H. S. Miley, “A novel method of pulse shape discrimination for germanium spectrometers”, *Nucl. Phys. B* 87 (2000) 491-492.
- [2] I. Abt, A. Caldwell, K. Kröniger, J. Liu, X. Liu y B. Majorovits, “Pulse shapes from electron and photon induced events in segmented high-purity germanium detectors”, *Eur. Phys. J. C* 52 (2007) 19-27.
- [3] D. H. Ackley, G. E. Hinton y T. J. Sejnowski, “A learning algorithm for Boltzmann machines”, *Cognitive Science* 9 (1985) 147-169.
- [4] J. M. Adams y G. White, “A Versatile Pulse Shape Discriminator for Charged Particle Separation and Its Application to Fast Neutron Time of Flight Spectroscopy”, *Nucl. Instrum. and Meth.* 156 (1968) 459-476.
- [5] S. Aiello, M. Alderighi, A. Anzalone, M. Bartolucci, M. Bruno, G. Cardella, S. Cavallaro, M. D’Agostino, E. De Filippo, E. Geraci, M. Geraci, F. Giustolisi, P. Guazzoni, C. M. Dacono Menno, G. Anzalone, G. Lozano, S. Loner, G. Manfred, A. Pagan, M. Papa, S. Prone, G. Politi, F. Porto, S. Russo, S. Sambataro, G. Sechi, M. L. Sperduto, C. M. Sutera y L. Zetta, *Proc. of Computing in High Energy and Nuclear Physics - CHEP2000, Padova, Italy, February 7-11 (2000)* 41.
- [6] D. Akimov, “Detectors for Dark Matter search (review)”, *Nucl. Instrum. and Meth. A* 598 (2009) 275-281.
- [7] J. Alarja, A. Dauchy, A. Giorni, C. Morand, E. Pollaco, P. Stassi, R. Billerey, B. Chambon, B. Cheynis, D. Drain y C. Pastor, “Charged particles identification with a CsI(Tl) scintillator”, *Nucl. Instrum. and Meth. A* 242 (1986) 352-354.
- [8] M. Altmann, F. V. Feilitzsch y U. Schanda, “A parameterized pulse shape analysis method for the GALLEX solar neutrino experiment”, *Nucl. Instrum. and Meth. A* 381 (1996) 398-412.
- [9] M. Alderighi, A. Anzalone, R. Bassini, I. Berceanu, J. Blicharska, C. Boiano, B. Borderie, R. Bougault, M. Bruno, C. Calí, G. Cardella, S. Cavallaro, M. D’Agostino, M. D’Andrea, R. Dayras, E. De Filippo, F. Fichera, E. Geraci, F. Giustolisi, A. Grzeszczuk, N. Guardone, P. Guazzoni, D. Guinet, C. M. Iacono-Manno, S. Kowalski, E. La Guidara, A. L. Lanchais, G. Lanzalone, G. Lanzaón, N. Le Neindre, S. Li, C. Maiolino, Z. Majka, G. Manfredi, D. Nicotra,

- T. Paduszynski, A. Pagano, M. Papa, C. M. Petrovici, E. Piasecki, S. Pirrone, G. Politi, A. Pop, F. Porto, M. F. Rivet, E. Rosato, G. Saccá, G. Sechi, V. Simion, M. L. Sperduto, J. C. Steckmeyer, A. Trifiró, M. Trimarchi, S. Urso, G. Vannini, M. Vigilante, J. Wilczynski, H. Wum, Z. Xiao, L. Zetta y W. Zipper, “Particle identification method in the CsI(Tl) scintillator used for the CHIMERA 4π detector”, *Nucl. Instrum. and Meth. A* 489 (2002) 257-265.
- [10] M. Alderighi, F. Amorini, A. Anzalone, N. Arena, L. Auditore, R. Bassini, C. Boiano, G. Cardella, S. Cavallaro, M. B. Chatterjee, M. D’Andrea, E. De Filippo, F. Fichera, E. Geraci, F. Giustolisi, N. Guardone, A. Grimaldi, P. Guazzoni, E. La Guidara, G. Lanzano, G. Lanzalone, C. Maiolino, D. Nicotra, A. Pagano, M. Papa, S. Pirrone, G. Politi, F. Porto, A. Pullia, F. Rizzo, S. Russo, P. Russotto, D. Santonocito, G. Sacca, M. Sassi, W. U. Schroder, G. R. Sechi, A. Trifiro, J. Toke, M. Trimarchi, S. Urso y L. Zetta, “Particle identification via pulse shape analysis for large-area silicon detectors of the CHIMERA array”, *IEEE Trans. on Nucl. Sci.* 52 (2005) 1624-1629.
- [11] M. Alderighi, F. Amorini, A. Anzalone, R. Bassini, C. Boiano, G. Cardella, S. Cavallaro, M. E. De Filippo, P. Guazzoni, E. La Guidara, G. Lanzano, G. Lanzalone, C. Maiolino, A. Pagano, M. Papa, S. Pirrone, G. Politi, F. Porto, F. Rizzo, S. Russo, G. Sacca, M. Sassi, W. G. Sechi, y L. Zetta, “Charge identification in large area planar silicon detectors, using digital pulse shape acquisition”, *IEEE Trans. on Nucl. Sci.* 53 (2006) 279-285.
- [12] S. Amari, “Theory of adaptive pattern classifiers”, *IEEE Trans. EC-16* (1967) 299-307.
- [13] M. Ambrosio et al. (The MACRO Collaboration), “Muon energy estimate through multiple scattering with the MACRO detector”, *Nucl. Instrum. and Meth. A* 492 (2002) 376-386.
- [14] C. A. J. Ammerlaan, R. F. Rumphorst y L. A. C. Koerts, “Particle identification by pulse shape discrimination in the p-i-n type semiconductor detector”, *Nucl. Instrum. and Meth.* 22 (1963) 189-200.
- [15] J. F. Amsbaugh, J. M. Anaya, J. Banar, T. J. Bowles, M. C. Browne, T. V. Bullard, T. H. Burritt, G. A. Cox-Mobrand, X. Dai, X. H. Deng, M. Di Marco, P. J. Doe, M. R. Dragowsky, C. A. Duba, F. A. Duncan, E. D. Earle, S. R. Elliott, E. I. Esch, H. Fergani, J. A. Formaggio, M. M. Fowler, J. E. Franklin, P. Geissbuehler, J. V. Germani, A. Goldschmidt, E. Guillian, A. L. Hallin, G. Harper, P. J. Harvey, R. Hazama, K. M. Heeger, J. Heise, A. Hime, M. A. Howe, M. Huang, L. L. Kormos, C. Kraus, C. B. Krauss, J. Law, I. T. Lawson, K. T. Lesko, J. C. Loach, S. Majerus, J. Manor, S. McGee, K. K. S. Miknaitis, G. G. Miller, B. Morissette, A. Myers, N. S. Oblath, H. M. O’Kee, R. W. Ollerhead, S. J. M. Peeters, A. W. P. Poon, G. Prior, S. D. Reitzner, K. Rielage, R. G. H. Robertson, P. Skensved, A. R. Smith, M. W. E. Smith, T. D. Steiger, L. C. Stonehill, P. M. Thornewell, N. Tolich, B. A. VanDevender, T. D. Van Wechel, B. L. Wall, H. W. C. Tseung, J. Wendland, N. West, J. B. Wilhelmy, J. F. Wilkerson y J. M. Wouters, “An array of low-background ^3He proportional counters for the Sudbury Neutrino Observatory”, *Nucl. Instrum. and Meth. A* 579 (2007) 1054-1080.
- [16] Y. Anzai, *Pattern Recognition and Machine Learning*, Academic Press, Inc., 1989.

- [17] L. N. Andronenko, M. N. Andronenko, A. A. Kotov, W. Neubert, L. A. Vaishnena, S. A. Pavlenko, D. M. Seliverstov, V. L. Stepanov y V. I. Yatsura, “A twin axial ionization chamber for studies of multiple indeterminate-mass-fragment production in 1 GeV proton-nucleus collision”, Nucl. Instrum. and Meth. A 312 (1992) 467-474.
- [18] L. Andronenko, M. Andronenko, A. Kotov, W. Neubert, G. Petrov, D. Seliverstov, L. Vaishnena y V. Yatsura, Preprint PNPI NP-3-(1998) Nr. 2217.
- [19] M. N. Andronenko y W. Neubert, Annual report 1998-1999, FZR-271 (1999) 67.
- [20] G. Angloher, M. Bauer, I. Bavykina, A. Bento, A. Brown, C. Bucci, C. Ciemniak, C. Coppi, G. Deuter, F. Von Feilitzsch, D. Hauff, S. Henry, P. Huff, J. Imber, S. Ingleby, C. Isaila, J. Jochum, M. Kiefer, M. Kimmerle, H. Kraus, J.-C. Lanfranchi, R. F. Lang, B. Majorovits, M. Malek, R. McGowan, V. B. Mikhailik, E. Pantic, F. Petricca, S. Pfister, W. Potzel, F. Pröbst, W. Rau, S. Roth, K. Rottler, C. Sailer, K. Schäffner, J. Schmalzer, S. Scholl, W. Seidel, L. Stodolsky, A. J. B. Tolhurst, I. Usherov y W. Westphal, “The XENON dark matter search experiment”, New Astronomy Reviews 49 (2005) 289-295.
- [21] J. M. Asselineau, J. Duchom, M. L’Haridon, P. Mosrin, R. Regimbart y B. Tamain, “Performance of a Bragg curve detector for heavy ion identification”, Nucl. Instrum. and Meth. 204 (1982) 109-115.
- [22] V. Avdeichikov, R. Guetti, P. Golubev, B. Jakobsson y N. Colonna, “Energy calibration of CsI(Tl) scintillator in pulse-shape identification technique”, Nucl. Instrum. and Meth. A 501 (2003) 505-513.
- [23] A. Bamberger, H. J. Brundiers, U. Fischer, D. Friedrich, R. Hartung, W. Heck, R. Maier, K. Runge, O. Schaile, B. Thomauske, C. Weber, A. Arefiev, A. Babaev, G. Eliseev, Yu. Galaktionov, Yu. Gorodkov, Yu. Kamishkov, V. Lubimov, W. Plyaskin, V. Pojedajev, V. Shevchenkov y M. Vlasov, “Particle identification through pulse shape analysis in proportional counters”, Phys. Scr. 23 (1981) 759-762.
- [24] J. C. Batchelder, C. R. Bingham, K. Rykaczewski, K. S. Toth, T. Davinson, J. A. McKenzie, P. J. Woods, T. N. Ginter, C. J. Gross, J. W. McConnell, E. F. Zganjar, J. H. Hamilton, W. B. Walters, C. Baktash, J. Greene, J. F. Mas, W. T. Milner, S. D. Paul, D. Shapira, X. J. Xu y C. H. Yu, “Observation of the exotic nucleus ^{145}Tm via its direct proton decay”, Phys. Rev. C 57 (1998) R1042-R1046.
- [25] L. Baudis, J. Hellmig, H. V. Klapdor-Kleingrothaus, Y. Ramachers, J. W. Hammer y A. Mayer, “High-purity germanium detector ionization pulse shapes of nuclear recoils, gamma interactions and microphonism”, Nucl. Instrum. and Meth. A 418 (1998) 348-354.
- [26] L. Bardelli, M. Bini, G. Poggi y N. Taccetti, “Application of digital sampling techniques to particle identification in scintillation detectors”, Nucl. Instrum. and Meth. A 491 (2002) 244-257.

- [27] D. Bazzacco, “The Advanced Gamma Ray Tracking Array AGATA”, Nucl. Phys. A 746 (2004) 248c-254c.
- [28] P. Bastia, G. Bertuccio, F. Borghetti, S. Caccia, V. Ferragina, F. Ferrari, D. Maiocchi, P. Malcovati, D. Martin, A. Pullia, y N. Ratti, “An integrated reset/pulse pile-up rejection circuit for pixel readout ASICs”, IEEE Trans. on Nucl. Sci. 53 (2006) 414-417.
- [29] S. Barlini, R. Bougault, Ph. Laborie, O. Lopez, D. Mercier, M. Parlog, B. Tamain, E. Vient, E. Chevallier, A. Chbihi, B. Jacquot y V. L. Kravchuk, “New digital techniques applied to A and Z identification using pulse shape discrimination of silicon detector current signals”, Nucl. Instrum. and Meth. A 600 (2009) 644-650.
- [30] H. O. Back, et al., “Pulse-Shape discrimination with the Counting Test Facility”, Nucl. Instrum. and Meth. A 584 (2008) 98-113.
- [31] L. Bardelli, M. Bini, P. G. Bizzeti, F. A. Danevich, T. F. Fazzini, V. V. Kobychyev, N. Krutyak, P. R. Maurenzig, V. M. Mokina, S. S. Nagorny, M. Pashkovskii, D. V. Poda, V. I. Tretyak y S. S. Yurchenko, “Pulse-shape discrimination with PbWO₄ crystal scintillators”, Nucl. Instrum. and Meth. A 584 (2008) 129-134.
- [32] A. G. Barto, “Reinforcement learning and adaptive critic methods”, Handbook of intelligent control: neural, fuzzy and adaptive approaches, ed. D. A. White y D. A. Sofg, New York: Van Nostrand Reinhold (1992) 469-491.
- [33] A. G. Barto, “Reinforcement learning control”, Current Opinion in Neurobiology 4 (1994) 888-893.
- [34] S. Becker, “Unsupervised learning procedures for neural networks”, The International Journal of Neural Systems 1 (1991) 17-33.
- [35] Z. Bell, “Tests on a digital neutron - gamma pulse shape discriminator with NE-213”, Nucl. Instrum. and Meth. A 188 (1981) 105-109.
- [36] R. Bernabei, P. Belli, A. Bussolotti, F. Cappella, R. Cerulli, C. J. Dai, A. D’Angelo, H. L. He, A. Incicchitti, H. H. Kuang, J. M. Ma, A. Mattei, F. Montecchia, F. Nozzoli, D. Prospero, X. D. Sheng y Z. P. Ye, “The DAMA/LIBRA apparatus”, Nucl. Instrum. and Meth. A 592 (2008) 297-315.
- [37] B. Bharath y V. S. Borkar, “Stochastic approximation: overview and recent trends”, Sadhana 24 (1999) 425-452.
- [38] M. Biehl y H. Schwarze, “Online learning of a time-dependent rule”, Europhysics Letters 20 (1992) 733-738.
- [39] M. Biehl y H. Schwarze, “Learning drifting concepts with neural networks”, J. Physics A: Mathematical and General 26 (1993) 2651-2665.

- [40] M. Biehl y H. Schwarze, “Learning by on-line gradient descent”, J. Phys. A 28 (1995) 643-656.
- [41] C. R. Bingham, J. C. Batchelder, K. Rykaczewski, K. S. Toth, C.-H. Yu, T. N. Ginter, C. J. Gross, R. Grzywacz, M. Karny, S. H. Kim, B. D. MacDonald, J. Mas, J. W. McConnell, P. B. Semmes, J. Szerypo, W. Weintraub y E. F. Zganjar, “Identification of a proton-emitting isomer in ^{151}Lu ”, Phys. Rev. C 59 (1999) R2984-R2988.
- [42] C. M. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1995.
- [43] M. Bolić y V. Drndarević, “Digital gamma-ray spectroscopy based on FPGA technology”, Nucl. Instrum. and Meth. A 482 (2002) 761-766.
- [44] L. Bottou, “Stochastic gradient learning in neural networks”, Proc. of the Neuro-Nimes 91, EC2, Nimes, France (1991).
- [45] L. Bottou, “Online learning and stochastic approximation”, en ed. D. Saad, Online Learning in Neural Networks, Cambridge University Press, Cambridge, UK (1998) 9-42.
- [46] S. Blös, “Statistical mechanics approach to early stopping and weight decay”, Phys. Rev. E 58 (1998) 833-844.
- [47] L. Bottou y N. Murata, “Stochastic Approximations and Efficient Learning”, The Handbook of Brain Theory and Neural Networks, Second edition,, Ed. by M. A. Arbib, The MIT Press, Cambridge, MA (2002).
- [48] L. Bottou y Y. LeCun, “Large Scale Online Learning”, Advances in Neural Information Processing Systems 16, Ed. S. Thrun, L. Saul y B. Schölkopf, MIT Press, Cambridge, MA (2004).
- [49] W. L. Bryan, C. L. Britton, J. T. Mihalcz, J. S. Neal, S. A. Pozzi y R. W. Tucker, “Fast neutron - gamma pulse shape discrimination of liquid scintillation signals for time correlated measurements”, Proc. IEEE Nucl. Sci. Symp. 2 (2003) 1192-1195.
- [50] G. A. Carpenter, S. Grossberg, Pattern Recognition by Self-organizing Neural Networks, The MIT Press, 1991.
- [51] D. A. Cebra, S. Howden, J. Karn, D. Kataria, M. Maier, A. Nadasen, C. A. Ogilvie, N. Stone, D. Swan, A. Vander Molen, W. K. Wilson, J. S. Winfield, J. Yurkon, G. D. Westfall y E. Norbeck, “Bragg curve spectroscopy in a 4π geometry”, Nucl. Instrum. and Meth. A 300 (1991) 518-521.
- [52] P. Chandrikamohan y T. A. DeVol, “Comparison of Pulse Shape Discrimination Methods for Phoswich and CsI:Tl Detectors”, IEEE Transactions on Nucl. Sci 54 (2007) 398-403.
- [53] O. Civitarese y J. Suhonen, “Is the single-state dominance realized in double- β -decay transitions?”, Phys. Rev. C 58 (1998) 1535-1538.
- [54] D. Clark y K. Ravishankar, “A convergence theorem for Grossberg learning”, Neural Networks 3 (1990) 87-92.

- [55] S. J. Cooper, “Donald O. Hebb’s synapse and learning rule: a history and commentary”, *Neuroscience and Biobehavioral Reviews*, 28 (2005) 851-874.
- [56] P. B. Cushman, “Dark matter searches: Technology and backgrounds”, *Nucl. Instrum. and Meth. A* 579 (2007) 437-442.
- [57] W. Daelemans y A. van den Bosch, “Memory-based language processing”, Cambridge University Press (2005).
- [58] J. Damgov y L. Litov, “Application of neural networks for energy reconstruction”, *Nucl. Instrum. and Meth. A* 482 (2002) 776-788.
- [59] M. A. Deleplanque, I. Y. Lee, K. Vetter, G. J. Schmid, F. S. Stephens, R. M. Clark, R. M. Diamond, P. Fallon y A. O. Macchiavelli, “GRETA: utilizing new concepts in γ -ray detection”, *Nucl. Instrum. and Meth. A* 430 (1999) 292-310.
- [60] P. A. Deyoung, R. L. McGrath y W. F. Piel Jr.r, “A single element NaI(Tl) light particle detector”, *Nucl. Instrum. and Meth.* 226 (1984) 555-557.
- [61] R. T. De Souza, N. Carlin, Y. D. Kim, J. Ottarson, L. Phair, D. R. Bowman, C. K. Gelbke, W. G. Gong, W. G. Lynch, R. A. Pelak, T. Peterson, G. Poggi, M. B. Tsang y H. M. Xu, “The MSU Miniball 4π fragment detection array”, *Nucl. Instrum. and Meth. A* 295 (1990) 109-122.
- [62] D. D. DiJulio, V. Avdeichikov, J. Cederkall, P. Golubev, B. Jakobsson, H. Johansson, C. Tintori, “Proton in-beam tests of the Lund R3B calorimeter prototype”, *Nucl. Instrum. and Meth. A* 612 (2009) 127-132.
- [63] D. Drain, A. Giorni, D. Hilscher, C. Ristori, J. Alarja, G. Barbier, R. Bertholet, R. Billerey, B. Chambon, B. Cheynis, J. Crançon, A. Dauchy, P. Désesquelles, A. Fontenille, L. Guyon, D. Heuer, A. Llères, M. Maurel, E. Monnard, C. Morand, H. Nifenecker, C. Pastor, J. Poux, H. Rossner, J. Saint-Martin, F. Schussler, P. Stassi, M. Tournier y J. B. Viano, “The particle detector array AMPHORA”, *Nucl. Instrum. and Meth. A* 281 (1989) 528-538.
- [64] V. Drndarević, “Elimination of distortions of nuclear energy spectra caused by pulse pile-up”, *Meas. Sci. Technol.* 5 (1994) 1573-1575.
- [65] J. H. Ely, C. E. Aalseth y J. I. McIntyre, “Novel beta-gamma coincidence measurements using phoswich detectors”, *J. of Rad. and Nucl. Chem.* 263 (2005) 245-250.
- [66] S. R. Elliot, “Pulse shape analysis in segmented detectors as a technique for background reduction in Ge double-beta decay experiments”, *Nucl. Instrum. and Meth.* 558 (2006) 504-510.
- [67] J. L. Elman, “Finding structures in time”, *Cognitive Science* 14 (1990) 179-211.
- [68] B. Esposito, L. Bertalot, D. Marocco, M. Riva, A. Zimbal, M. Reginatto, H. Schuhmacher, J. M. Adams y A. Murari, “Neutron measurements on Joint European Torus using an NE213 scintillator with digital pulse shape discrimination”, *Rev. of Sci. Instrum.* 75 (2004) 3550-3552.

- [69] D. Fabris, F. Gramegna, G. Nebbia, G. Prete, E. A. Ricci, G. Viesti, I. Iori y A. Moroni, “The Legnaro Bragg-curve spectroscopy ionization chamber for heavy-ion mass and charge identification”, *Lettere Al Nuovo Cimento* (1982) 115-120.
- [70] A. Faessler y F. Simkovic, “Double Beta Decay”, *J. Phys G.* 24 (1998) 2139-2178.
- [71] S. E. Fahlman, “Faster-Learning Variations on Back-Propagation: An Empirical Study”, en eds. D. Touretzky, G. Hinton y T. Sejnowski, *Proc. of the 1988 Connectionist Models Summer School*, Morgan Kaufmann (1989) 38-51.
- [72] A. Fallu-Labruyere, H. Tan, W. Hennig y W. K. Warburton, “Time resolution studies using digital constant fraction discrimination”, *Nucl. Instrum. and Meth A* 579 (2007) 247-251.
- [73] A. T. Farsoni y D. M. Hamby, “MCNP analysis of a multilayer phoswich detector for β -particle dosimetry and spectroscopy”, *Nucl. Instrum. and Meth. A* 555 (2005) 225-230.
- [74] A. T. Farsoni y D. M. Hamby, “A system for simultaneous beta and gamma spectroscopy”, *Nucl. Instrum. and Meth A* 578 (2007) 528-536.
- [75] M. Flaska y S. A. Pozzi, “Digital pulse shape analysis for the capture-gated liquid scintillator BC-523A”, *Nucl. Instrum. and Meth. A* 599 (2009) 221-225.
- [76] O. Frisch, *British Atomic Energy Report*, BR-49 (1944).
- [77] R. de la Fuente, B. de Celis, V. del Canto, J. M. Lumbreras, B. de Celis Alonso, A. Martín-Martín y J. L. Gutiérrez-Villanueva, “Low level radioactivity measurements with phoswich detectors using coincident techniques and digital pulse processing analysis”, *J. of Environ. Radioac.* 99 (2008) 1553-1557.
- [78] K. Fukushima, “Cognitron: A self-organizing multilayered neural network”, *Biological Cybernetics* 20 (1975) 121-136.
- [79] K. Fushimi, H. Ejiri, K. Hayashi, R. Hazama, T. Kishimoto, N. Kudomi, K. Kume y H. Ohsumi, “Limits on the annual modulation of WIMPS-nucleus scattering with large-volume NaI(Tl) scintillators”, *Astroparticle Physics* 12 (1999) 185-192.
- [80] A. Galindo-Uribarri, “Experiments with a 4π charged-particle detector and the 8π spectrometer”, *Prog. Part. Nucl. Phys.* 28 (1992) 463-472.
- [81] S. Geman, E. Bienenstock y R. Doursat, “Neural networks and the bias/variance dilemma”, *Neural Computation* 4 (1992) 1-58.
- [82] G. Gerbier, J. Mallet, L. Mosca, C. Tao, B. Chambon, V. Chazal, M. De Jésus, D. Drain, Y. Messous y C. Pastor, “Pulse shape discrimination and dark matter search with NaI(Tl) scintillator”, *Astroparticle Physics* 11 (1999) 287-302.
- [83] T.N. Ginter, “The new HRIBF recoil mass spectrometer”, Ph.D. Dissertation, Vanderbilt University, Nashville, Tennessee (1999).

- [84] W. G. Gong, Y. D. Kim, G. Poggi, Z. Chen, C. K. Gelbke, W. G. Lynch, M. R. Maier, T. Murakami, M. B. Tsang, H. M. Xu y K. Kwiatkowski, “Resolution tests of CsI(Tl) scintillators read out by pin diodes”, Nucl. Instrum. and Meth. A 268 (1988) 190-199.
- [85] D. González, J. Morales, S. Cebrián, E. García, I. G. Irastorza, A. Morales, A. Ortiz de Solórzano, J. Puimedón, M. L. Sarsa, J. A. Villar, C. E. Aalseth, F. T. Avignone, R. L. Brodzinski, W. K. Hensley, H. S. Miley, J. H. Reeves, I. V. Kirpichnikov, A. A. Vasenko, A. A. Klimenko, S. B. Osetrov, A. A. Smolnikov, S. I. Vasiliev, V. S. Pogosov y A. G. Tamanyan, “Pulse-shape discrimination in the IGEX experiment”, Nucl. Instrum. and Meth A 515 (2003) 634-643.
- [86] C. J. Gross, T. N. Ginter, D. Shapira, W. T. Milner, J. W. McConnell, A. N. James, J. W. Johnson, J. Mas, P. F. Mantica, R. L. Auble, J. J. Das, J. L. Blankenship, J. H. Hamilton, R. L. Robinson, Y. A. Akovali, C. Baktash, J. C. Batchelder, C. R. Bingham, M. J. Brinkman, H. K. Carter, R. A. Cunningham, T. Davinson, J. D. Fox, A. Galindo-Uribarri, R. Grzywacz, J. F. Liang, B. D. MacDonald, J. MacKenzie, S. D. Paul, A. Piechaczek, D. C. Radford, A. V. Ramayya, W. Reviol, D. Rudolph, K. Rykaczewski, K. S. Toth, W. Weintraub, C. Williams, P. J. Woods, C. H. Yu y E. F. Zganjar, “Performance of the Recoil Mass Spectrometer and its detector systems at the Holifield Radioactive ion Beam Facility”, Nucl. Instrum. and Meth. A 450 (2000) 12-29.
- [87] S. Grossber, “Neural expectation: Cerebellar and retinal analogs of cells fired by learnable or unlearnable pattern classes”, Kybertnetik 10 (1972) 49-57.
- [88] S. Grossber, “Adaptive pattern classification and universal recording: I. Parallel development and coding of neural detectors”, Biological Cybernetics 23 (1976) 121-134.
- [89] S. Grossber, “Adaptive pattern classification and universal recording: II. Feedback, expectation, olfaction, illusions”, Biological Cybernetics 23 (1976) 187-2002.
- [90] S. Grossberg, Mind and Language (Special Issue on Understanding Vision), 5 (1990) 411-456.
- [91] C. R. Gruhn, M. Binimi, R. Legrain, R. Loveman, W. Pang, M. Roach, D. K. Scott, A. Shotter, T. J. Symons, J. Wouters, M. Zisman R. Devries, Y. C. Peng y W. Sondheim, “Bragg curve spectroscopy”, Nucl. Instrum. and Meth. 196 (1982) 33-40.
- [92] D. Guinet, B. Chambon, B. Cheynis, A. Demeyer, D. Drain, X. C. Hu, C. Pastor, L. Vagneron y K. Zaid, “Using the combination CsI(Tl) and photodiode for identification and energy measurement of light particles”, Nucl. Instrum. and Meth. A 278 (1989) 614-616.
- [93] W. Guo, R. P. Gardne y C. W. Mayo, “A study of the real-time deconvolution of digitized waveforms with pulse pile up for digital radiation spectroscopy”, Nucl. Instrum. and Meth. A 544 (2005) 668-678.
- [94] S. B. Guetersloh, P. J. Taddei, T. B. Borak, B. B. Gersey, C. Zeitlin, L. Heilbronn, J. Miller, T. Murakami y Y. Iwata, “The response of a spherical tissue-equivalent proportional counter to different heavy ions having similar velocities”, Rad. Research 41 (2006) 1227-1234.

- [95] H. Hamrita, E. Raully, Y. Blumenfeld, B. Borderie, M. Chabot, P. Edelbruck, L. Lavergne, J. Le Bris, Th. Legou, N. Le Neindre, A. Richard, M. F. Rivet, J. A. Scarpaci, J. Tillier, S. Barbey, E. Becheva, F. Bocage, R. Bougault, R. Bzyl, S. Grévy, B. Carniol, D. Cussol, P. Désesquelles, D. Etasse, E. Galichet, D. Guinet, G. Lalu, G. Lanzalone, Ph. Lautesse, O. Lopez, G. Martinet, S. Pierre, G. Politi, E. Rosato, B. Tamain y E. Vient, “Charge and current-sensitive preamplifiers for pulse shape discrimination techniques with silicon detectors”, Nucl. Instrum. and Meth. A 531 (2004) 607-615.
- [96] S. P. Hart para Boulby Dark Matter Collaboration, “Status of Dark Matter Searches in the Boulby Mine”, Nucl. Phys. B (Proc. Suppl.) 110 (2002) 91-93.
- [97] K. H. Hastie, T. A. DeVol y R. A. Fjeld, “Development of a homogeneous pulse shape discriminating flow-cell radiation detection system”, Nucl. Instrum. and Meth. A 422 (1999) 133-138.
- [98] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, 1998.
- [99] D. O. Hebb, “The organization of behavior: a neurophysiological theory”, New York, Wiley, (1949).
- [100] R. Hecht-Nielsen, “Neurocomputing”, Addison-Wesley Publishing Company, Inc. (1990) 125.
- [101] E. H. M. Heijne, “1980, a revolution in silicon detectors, from energy spectrometer to radiation imager: Some technical and historical details”, Nucl. Instrum. and Meth. A 591 (2008) 6-13.
- [102] J. Hellmig, F. Petry y H.V. Klapdor-Kleingrothaus, Patent DE19721323A (1999).
- [103] J. Hellmig y H.V. Klapdor-Kleingrothaus, “Identification of single-site events in germanium detectors by digital pulse shape analysis”, Nucl. Instr. and Meth. A455 (2000) 638-644.
- [104] W. Hennig, H. Tan, W. K. Warburton y J. I. McIntyre, “Single channel beta-gamma coincidence detection of radioactive xenon using digital pulse shape analysis of phoswich detector signals”, IEEE Trans. on Nucl. Sci. 53 (2006) 620-624.
- [105] W. Hennig, H. Tan, A. Fallu-Labruyere, W. K. Warburton, J. I. McIntyre y A. Gleyzer, “A phoswich well detector for radioxenon monitoring”, Nucl. Instrum. and Meth. A 579 (2007) 431-436.
- [106] J. Hertz, A. Krogh y R. G. Palmer, “Introduction to the Theory of Neural Computation”, Addison Wesley, Redwood City, CA (1991).
- [107] T. M. Heskes y B. Kappen, “Learning processs in neural network”, Phys. Rev A 44 (1991) 2718-2726.
- [108] T. M. Heskes y B. Kappen, “Learning-parameter adjustment in neural networks”, Phys. Rev A 45 (1992) 8885-8893.

- [109] T. M. Heskes y B. Kappen, "On-line learning processes in artificial neural networks", en editorial J. G. Taylor, *Mathematical Approaches to Neural Networks*, Elsevier Amsterdam 51 (1993) 199-233.
- [110] T. M. Heskes y W. Wiegerinck, "A theoretical comparison of batch-mode, on-line, cyclic, and almost cyclic learning", *IEEE Trans. on nuclear networks* 7 (1996) 919-925.
- [111] G. E. Hinton y T. J. Sejnowski, "Optimal perceptual inference", *Proc. of the IEEE Computer Science, Conference on Computer Vision and Pattern Recognition* (1983) 448-453.
- [112] N. P. O. Homman y P. Kristiansson, "Improvements through Pulse Shape Analysis in X-Ray Spectrometry Using Si(Li) Detectors", *X-ray spectrometry* 25 (1996) 66-73.
- [113] J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", *Proc. of the National Academy of Sciences*, 79 (1982) 2554-2558.
- [114] B. Hubbard-Nelson, M. Momayezi y W. K. B. Warburton, "A module for energy and pulse shape data acquisition", *Nucl. Instrum. and Meth. A* 422 (1999) 411-416.
- [115] H. Ikezoe, T. Murakami, Y. Tomita y N. Shikazono, "Particle identification for heavy ions in a time-of-flight spectrometer", *Nucl. Instrum. and Meth.* 196 (1982) 215-218.
- [116] S. D. Jastaniah y P. J. Sellin, "Digital pulse-shape algorithms for scintillation-based neutron detectors", *IEEE Trans. on Nucl. Sci.* 49 (2002) 1824-1828.
- [117] S. D. Jastaniah y P. J. Sellin, "Digital techniques for n/ γ pulse shape discrimination and capture-gated neutron spectroscopy using liquid scintillators", *Nucl. Instrum. and Meth. A* 517 (2004) 202-210.
- [118] T. Kihm, V. F. Bobrakov y H. V. Klapdor-Kleingrothaus, "A digital multi-channel spectroscopy system with 100 MHz flash ADC module for the GENIUS-TF and GENIUS projects", *Nucl. Instrum. and Meth. A* 498 (2003) 334-339.
- [119] T. Y. Kim, I. S. Cho, D. H. Choi, J. M. Choi, I. S. Hahn, M. J. Hwang, H. K. Jang, R. K. Jain, U. K. Kang, H. J. Kim, J. H. Kim, S. C. Kim, S. K. Kim, S. Y. Kim, Y. D. Kim, Y. J. Kwon, H. S. Lee, J. H. Lee, M. H. Lee, S. E. Lee, S. H. Noh, H. Park, I. H. Park, E. S. Seo, E. Won, H. S. Won, H. Y. Yang, M. S. Yang y I. Yu, "Study of the internal background of CsI (Tl) crystal detectors for dark matter search", *Nucl. Instrum. and Meth. A* 500 (2003) 337-344.
- [120] K. Kimura, "Particle identification through the measurement of the Bragg curve centroid", *Nucl. Instrum. and Meth.* 212 (1983) 227-234.
- [121] O. Kinouchi y N. Caticha, "Lower bounds on generalization errors for drifting rules", *J. of Phys. A: Mathematical and General* 26 (1993) 6161-6171.

- [122] H. V. Klapdor-Kleingrothaus, L. Baudis, G. Heusser, B. Majorovits y H. Päs, “GENIUS: a Supersensitive Germanium Detector System for Rare Events”, Proposal, MPI-H-V26-1999 (1999).
- [123] H. V. Klapdor-Kleingrothaus y S. Stoica, “Critical view on double-beta decay matrix elements within Quasi Random Phase Approximation-based methods”, Nucl. Phys. A 694 (2001) 269-294.
- [124] H. V. Klapdor-Kleingrothaus, “Evidence for Neutrinoless Double Beta Decay”, Int. J. Mod. Phys. A 18 (2003) 4113-4128.
- [125] G. F. Knoll, Radiation detection and measurement, New York: John Wiley & Sons, Inc., 2000.
- [126] Th. Kröll y D. Bazzacco, “Simulation and analysis of pulse shapes from highly segmented HPGe detectors for the γ -ray tracking array MARS”, Nucl. Instrum. and Meth. A 463 (2001) 227-249.
- [127] V. A. Kudryavtsev, N. J. C. Spooner, D. R. Tovey, J. W. Roberts, M. J. Lehner, J. E. McMillan, P. K. Lightfoot, T. B. Lawson, C. D. Peak, R. Lüscher y J. C. Barton, “CsI(Tl) for WIMP dark matter searches”, Nucl. Instrum. and Meth. A 456 (2001) 272-279.
- [128] V. A. Kudryavtsev, N. J. C. Spooner, P. K. Lightfoot, J. W. Roberts, M. J. Lehner, T. Gamble, M. J. Carson, T. B. Lawson, R. Lüscher, J. E. McMillan, B. Morgan, S. M. Paling, M. Robinson, D. R. Tovey, N. J. T. Smith, P. F. Smith, G. J. Alner, S. P. Hart, J. D. Lewin, R. M. Preece, T. J. Sumner, W. G. Jones, J. J. Quenby, B. Ahmed, A. Bewick, D. Davidge, J. V. Dawson, A. S. Howard, I. Ivaniouchenkov, M. K. Joshi, V. Lebedenko, I. Liubarsky, J. C. Barton, G. Gerbier, J. Mallet, L. Mosca y C. Tao, “Study and suppression of anomalous fast events in inorganic scintillators for dark matter searches”, Astroparticle Physics 17 (2002) 401-408.
- [129] V. N. Lebedenko, H. M. Arajo, E. J. Barnes, A. Bewick, R. Cashmore, V. Chepel, D. Davidge, J. Dawson, T. Durkin, B. Edwards, C. Ghag, V. Graffagnino, M. Horn, A. S. Howard, A. J. Hughes, W. G. Jones, M. Joshi, G. E. Kalmus, A. G. Kovalenko, A. Lindote, I. Liubarsky, M. I. Lopes, R. Luscher, P. Majewski, A. St. J. Murphy, F. Neves, J. Pinto da Cunha, R. Preece, J. J. Quenby, P. R. Scovell, C. Silva, V. N. Solovov, N. J. T. Smith, P. F. Smith, V. N. Stekhanov, T. J. Sumnery, C. Thorne y R. J. Walker, “Result from the First Science Run of the ZEPLIN-III Dark Matter Search Experiment”, Phys. Rev. D 80 (2008).
- [130] Y. LeCun, L. Bottou, G. B. Orr y K.-R. Müller, “Efficient BackProp”, Lecture Notes in Computer Science, 1524 (1998) 9-50.
- [131] H. S. Lee, H. Bhang, J. H. Choi, I. S. Hahn, D. He, M. J. Hwang, H. J. Kim, S. C. Kim, S. K. Kim, S. Y. Kim, T. Y. Kim, Y. D. Kim, J. W. Kwak, Y. J. Kwon, J. Lee, J. H. Lee, J. I. Lee, M. J. Lee, J. Li, S. S. Myung, H. Park, H. Y. Yang y J. J. Zhu, “First limit on WIMP cross section with low background CsI(Tl) crystal detector”, Nucl. Instrum. and Meth. A 633 (2006) 201-208.

- [132] Y. Liu, C. P. Chen, H. B. Li, C. H. Tang, C. Y. Chang, L. Hou, W. P. Lai, J. Li, S. T. Lin, C. S. Luo, J. F. Qiu, H. Y. Sheng, C. C. Wang, M. Z. Wang, S. C. Wang, H. T. Wong, B. Xin, Q. Yue, D. X. Zhao, S. Q. Zhao, Z. Y. Zhou y B. A. Zhuang, “Studies of prototype CsI(Tl) crystal scintillators for low-energy neutrino experiments”, Nucl. Instrum. and Meth. A 482 (2002) 125-143.
- [133] W. H. Lippincott, K. J. Coakley, D. Gastler, A. Hime, E. Kearns, D. N. McKinsey, J. A. Nikkel y L. C. Stonehill, “Scintillation time dependence and pulse shape discrimination in liquid argon”, Phys. Rev. C 78 (2008) 035801-035812.
- [134] L. Ljung, “Analysis of recursive stochastic algorithms”, IEEE Trans. on Automatic Control AC-22 (1977) 551-575.
- [135] J. Lua, P. Figuera, F. Amorini, G. Cardella, A. DiPietro, A. Musumarra, M. Papa, G. Pappalardo, F. Rizzo y S. Tudisco, “Pulse shape discrimination of charged particles with a silicon stripdetector”, Nucl. Instrum. and Meth A 471 (2001) 374-379.
- [136] B. Majorovits y H.V. Klapdor-Kleingrothaus, “Digital pulse shape analysis by neural networks for the Heidelberg-Moscow-Double-Beta-Decay-Experiment”, Eur. Phys. J. A 6 (1999) 463-469.
- [137] C. Von Der Malsburg, “Self-organization of orientation sensitive cells in the striate cortex”, Kybernetik 14 (1973) 85-100.
- [138] C. W. H. Mace y A. C. C. Coolen, “Statistical mechanics analysis of the dynamics of learning in perceptrons”, Statistics and Computing 8 (1998) 55-88.
- [139] W. S. McCulloch y W. Pitts, “A logical calculus of the ideas immanent in neuron activity”, Bull. Math. Biophys., 5 (1943) 115-133.
- [140] S. Marrone, D. Cano-Ott, N. Colonna, C. Domingo, F. Gramegna, E. M. Gonzalez, F. Gunsing, M. Heil, F. Kappeler, P. F. Mastinu, P. M. Milazzo, T. Papaevangelou, P. Pavlopoulos, R. Plag, R. Reifarth, G. Tagliente, J. L. Tain y K. Wisshak, “Pulse shape analysis of liquid scintillators for neutron studies”, Nucl. Instrum. and Meth. A 490 (2002) 299-307.
- [141] T. Masters, Advanced Algorithms for Neural Networks: A C++ Sourcebook, NY: John Wiley and Sons, 1995.
- [142] W. Mengesha, C. E. Aalseth, D. S. Barnett, M. Bliss y C. Schaefer, “Pulse shape analysis for electron mobility study in cadmium zinc telluride gamma-ray detectors”, IEEE Nucl. Sci. Symposium Conf. Record, 5 (2003) 3498-3502.
- [143] L. F. Miller, J. Preston, S. Pozzi, M. Flaska y J. Neal, “Digital pulse shape discrimination”, Rad. Prot. Dos. 126 (2007) 645-647.
- [144] M. L. Minsky y S. Papert, “Perceptrons: an introduction to computational geometry”, MIT Press, (1969).

- [145] M. Mizuho y T. Mizoguchi, “Electronic collimation in proportional counter by rise time discrimination”, *J. Nucl. Sci. and Tech.* 7 (1979) 335-340.
- [146] M. Momayezi, W. K. Warburton y R. Kroeger, “Position resolution in a Ge-strip detector”, *SPIE* 3768 (1999) 530-537.
- [147] R. Moreno-Díaz y A. Moreno-Díaz, “On the legacy of W. S. McCulloch”, *BioSystems*, 88 (2007) 185-190.
- [148] A. Moroni, I. Iori, L. Z. Yu, G. Prete, G. Viesti, F. Gramegna y A. Dainelli, “Position sensitive and Bragg curve spectroscopy detector system”, *Nucl. Instrum. and Meth.* 225 (1984) 57-64.
- [149] M. Moszyński, G. J. Costa, G. Guillaume, B. Heusch, A. Huck y S. Mouatassim, “Study of n- γ discrimination with NE213 and BC501A liquid scintillators of different size”, *Nucl. Instrum. and Meth. A* 350 (1994) 226 -234.
- [150] B. K. Nayaki, E. T. Mirgule y R. K. Choudhury, “Application of pulse shape discrimination in Si detector for fission fragment angular distribution measurements”, *Pramana J. Phys.* 65 (2005) 1053-1059.
- [151] N. Le Neindre, M. Alderighi, A. Anzalone, R. Barná, M. Bartolucci, I. Berceanu, B. Borderie, R. Bougault, M. Bruno, G. Cardella, S. Cavallaro, M. D’Agostino, R. Dayras, E. De Filippo, D. De Pasquale, E. Geraci, F. Giustolisi, A. Grzeszczuk, P. Guazzoni y D. Guinet, “Mass and charge identification of fragments detected with the Chimera Silicon CsI(Tl) telescopes”, *Nucl. Instrum. and Meth. A* 490 (2002) 251-262.
- [152] A. Nigrin, *Neural Networks for Pattern Recognition*, The MIT Press, 1993.
- [153] T. Okuda, H. Yamazaki, M. Kawabata, J. Kasagi y H. Harada, “Performance of n- γ pulse-shape discrimination with simple pile-up rejection at high γ -ray count rates”, *Nucl. Instrum. and Meth. A* 426 (1999) 497-502.
- [154] G. B. Orr, “Dynamics and Algorithms for Stochastic Learning”, tesis doctoral, Oregon Graduate Institute (1995).
- [155] H. G. Ortlepp y A. Romaguera, “Real time digital processing of Bragg ionization chamber signals”, *Nucl. Instrum. and Meth. A* 276 (1989) 500-508.
- [156] R. B. Owen, “Pulse-shape discrimination identifies particle types”, *Nucleonics* 17 9 (1959) 92-95.
- [157] R. B. Owen, “Pulse-shape discrimination - A Survey of Current Techniques”, *IRE Trans. on Nucl. Sci.* NS-59 (1962) 285-293.
- [158] G. Pausch, W. Bohne y D. Hilscher, “Particle identification in solid-state detectors by means of pulse-shape analysis results of computer simulations”, *Nucl. Instrum. and Meth. A* 337 (1994) 573-587.

- [159] G. Pausch, H. G. Ortlepp, W. Bohne, H. Grawe, D. Hilscher, M. Moszyński, D. Wolski, R. Chubart, G. de Angelis y M. de Poli, “Identification of light charged particles and heavy ions in silicon detectors by means of pulse-shape discrimination”, *IEEE Trans. on Nucl. Sci.* 43 (1996) 1097-1101.
- [160] G. Pausch, M. Moszyński, W. Bohne, J. Cederkall, H. Grawe, W. Klamra, M. O. Lampert, P. Rohr, R. Schubart, W. Seidel y D. Wolski, “Limitations of the pulse-shape technique for particle discrimination in planar Si detectors”, *IEEE Trans. on Nucl. Sci.* 44 (1997) 1040-1045.
- [161] G. Pausch, H. Prade, M. Sobiella, H. Schnare, R. Schwengner, L. Käubler, C. Borcan, H. G. Ortlepp, U. Oehmichen, H. Grawe, R. Schubart, J. Gerl, J. Cederkäll, A. Johnson, A. Kerek, W. Klamra, M. Moszyński, D. Wolski, M. Kapusta, A. Axelsson, M. Weiszflog, T. Härtle, D. Pansegrau, G. de Angelis, S. Ashrafi, A. Likar, y M. Lipoglavsek, “RoSiB -a 4π silicon ball for charged-particle detection in EUROBALL”, *Nucl. Instrum. and Meth. A* 443 (2000) 304-318.
- [162] L. Pandola, C. Cattadori y N. Ferrari, “Neural network pulse shape analysis for proportional counters events”, *Nucl. Instrum. and Meth. A* 522 (2004) 521-528.
- [163] S. Pécourt, B. Chambon, M. de Jésus, D. Drain, G. Gerbier, J. Mallet, M. Massa, L. Mosca, C. Pastor, C. Tao y L. Vagneron, “Calibration of a CsI(Tl) crystal with nuclear recoils and pulse shape measurements for dark matter detection”, *Astroparticle Physics* 11 (1999) 457-462.
- [164] P. Peiffer, T. Pollmann, S. Schönert, A. Smolnikov y S. Vasiliev, “Pulse shape analysis of scintillation signals from pure and xenon-doped liquid argon for radioactive background identification”, *J. Inst.* 3 (2008) PO8007.
- [165] R. C. Pereira, J. Sousa, A. M. Fernandes, F. Patricio, B. Carvalho, A. Neto, C. A. F. Varrandas, G. Gorini, M. Tardocchi, D. Gin y A. Shevelev, “ATCA data acquisition system for gamma-ray spectrometry”, *Fusion Engineering and Design* 83 (2008) 341-345.
- [166] M. Pflützner, E. Badura, R. Grzywacz, Z. Janas, M. Momayezi, C. Bingham, B. Blank, M. Chartier, H. Geissel, J. Giovinazzo, M. Hellstrom, J. Kurcewicz, A. S. Lalleman, C. Mazzocchi, I. Mukha, C. Plettner, E. Roeckl, K. P. Rykaczewski, K. Schmidt, R. S. Simon, M. Stanoiu y J.-C. Thomas, “Charged-particle spectroscopy in the microsecond range following projectile fragmentation”, *Nucl. Instrum. and Meth. A* 493 (2002) 155-164.
- [167] M. Rattray y D. Saad, “Incorporating curvature information into on-line learning”, en ed. D. Saad, *Online Learning in Neural Networks*, Cambridge University Press, New York, NY (1999) 183-207.
- [168] S. J. B. Reed, “Pulse pile-up rejection in Si(Li) X ray detection systems”, *J. Phys. E* 5 (1972) 997-1000.
- [169] K. E. Rhem y F. L. Wolfs, “A focal plane detector for reactions with medium weight projectiles”, *Nucl. Instrum. and Meth. A* 273 (1988) 262-272.

- [170] M. Riedmiller y H. Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm", Proc. of the IEEE International Conference on Neural Networks 1993, San Francisco: IEEE (1993).
- [171] B. D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, 1996.
- [172] H. Ritter y K. Schulten, "Convergence properties of Kohonen's topology conserving maps: fluctuations, stability, and dimension selection", Biological Cybernetics 60 (1988) 59-71.
- [173] R. Rojas, "Neural Networks - A Systematic Introduction", Springer-Verlag, New York (1996).
- [174] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain", Psychological Review, 65 (1958) 386-408.
- [175] F. Rosenblatt, Principles of neurodynamics, New York: Spartan, (1962).
- [176] B. B. Rosi y H. H. Staub, ionization chambers and counters, McGraw-Hill, New York (1949).
- [177] M. L. Roush, M. A. Wilson, y W. F. Hornyak, "Pulse Shape Discrimination", Nucl. Instrum. and Meth. 31 (1964) 112-124.
- [178] D. E. Rumelhart y J. L. McClelland, "Parallel distributed processing", 2 volúmenes, MIT Press, Cambridge, MA, USA (1986).
- [179] D. E. Rumelhart, G. E. Hinton y R. J. Williams, "Learning representations by back-propagation errors", Nature 323 (1986) 533-536.
- [180] G. A. W. Rutgers y J. M. W. Milatz, "ionization and straggling of Po α -particles in Hydrogen and Helium", Physica 7 (1940) 508-514.
- [181] K. Rykaczewski, J. C. Batchelder, C. R. Bingham, T. Davinson, T. N. Ginter, C. J. Gross, R. Grzywacz, M. Karny, B. D. MacDonald, J. F. Mas, J. W. McConnell, A. Piechaczek, R. C. Slinger, K. S. Toth, W. B. Walters, P. J. Woods, E. F. Zganjar, B. Barmore, L. Gr. Ixaru, A. T. Kruppa, W. Nazarewicz, M. Rizea y T. Vertse, "Proton emitters ^{140}Ho and ^{141}Ho : Probing the structure of unbound Nilsson orbitals", Phys. Rev. C 60 (1999) 011301-011305.
- [182] K. Rykaczewski, J. W. McConnell, C. R. Bingham, R. Grzywacz, M. Karny, J. C. Batchelder, C. J. Gross, Z. Janas, M. Momayezi, J. Wahl, A. Piechaczek, E. F. Zganjar, T. N. Ginter, J. H. Hamilton, W. B. Walters, W. D. Kulp y J. A. Winger, Invited paper, Fifth International Conference on Radioactive Nuclear Beams, Divonne, France, 3-8 April 2000.
- [183] K. Rykaczewski, J. W. McConnell, C. R. Bingham, R. Grzywacz, M. Karny, J. C. Batchelder, C. J. Gross, Z. Janas, M. Momayezi, J. Wahl, A. Piechaczek, E. F. Zganjar, T. N. Ginter, J. H. Hamilton, W. B. Walters, W. D. Kulp y J. A. Winger, "Towards new proton radioactivities with radioactive beams and digital signal processing", Nucl. Phys. A 701 (2002) 179-183.

- [184] K. P. Rykaczewski, R. K. Grzywacz, M. Karny, J. W. McConnell, M. Momayez, J. Wahl, Z. Janas, J. C. Batchelder, C. R. Bingham, D. Hartley, M. N. Tantawy, C. J. Gross, T. N. Ginter, J. H. Hamilton, W. D. Kulp, M. Lipoglavsek, A. Piechaczek, E. F. Zganjar, W. B. Walters y J. A. Winger, “Towards digital spectroscopy of proton emitters”, Nucl. Phys. A 682 (2001) 270c-278c.
- [185] D. Saad y S. A. Solla, “On-line learning in soft committee machines”, Phys. Rev. E 52 (1995) 4225-4243.
- [186] D. Saad y S. A. Solla, “Exact solution for on-line learning in multilayer neural networks”, Phys. Rev. Lett. 74 (1995) 4337-4340.
- [187] D. Saad y M. Rattray, “Globally Optimal Parameters for On-Line Learning in Multilayer Neural Networks”, Phys. Rev. Lett. 79 (1997) 2578-2581.
- [188] T. Sanami, M. Hagiwara, T. Oishi, M. Baba y M. Takada, “A Bragg curve counter with an active cathode to improve the energy threshold in fragment measurements”, Nucl. Instrum. and Meth. 589 (2008) 193-201.
- [189] H. Sakai, A. Uritani, Y. Takenaka, C. Mori y T. Iguchi, “New pulse-shape analysis method with multi-shaping amplifiers”, Nucl. Instrum. and Meth. A 421 (1999) 316-321.
- [190] D. G. Sarantites, P. -F. Hua, M. Devlin, L. G. Sobotka, J. Elson, J. T. Hood, D. R. LaFosse, J. E. Sarantites y M. R. Maier, ““The microball” Design, instrumentation and response characteristics of a 4π -multidetector exit channel-selection device for spectroscopic and reaction mechanism studies with Gammasphere”, Nucl. Instrum. and Meth. 381 (1996) 418-432.
- [191] W. S. Sarle, (editor) comp.ai.neural-nets FAQ, parte 2, disponible en línea en: <ftp://ftp.sas.com/pub/neural/FAQ2.html> (2002).
- [192] S. Schaal y C. G. Atkeson, “Robot juggling: An implementation of memory-based learning”, Control Systems Magazine 14 (1994) 57-71.
- [193] C. Schiessl, W. Wagner, K. Hartel, P. Kienle, H. J. Körner, Waltraud Mayer y K. E. Rehm, “A bragg-curve spectroscopy detector”, Nucl. Instrum. and Meth. 192 (1982) 291-294.
- [194] G. J. Schmid, D. Beckedahl, J. J. Blair, A. Friensehner y J. E. Kammeraad, “HPGe Compton suppression using pulse shape analysis”, Nucl. Instrum. and Meth. A 422 (1999) 368-372.
- [195] S. Selvakumar, A. Arulchakkaravarthi, Rakesh Kumar, S. Muralithar, S. Sankar y K. Sivaji, “Enhanced energy resolution (^{22}Na , ^{137}Cs and ^{241}Am) and alpha-gamma ($^{241}\text{Am} + ^{60}\text{Co}$) and neutron-gamma (^{241}Am -Be) pulse shape discrimination properties of *p*-terphenyl crystals prepared by modified selective self-seeded vertical Bridgman technique (SSVBT)”, Sol. Stat. Comm. 145 (2008) 482-486.
- [196] H. S. Seung, H. Sompolinsky y N. Tishby, “Statistical mechanics of learning from examples”, Phys. Rev. A 45 (1992) 6056-6091.

- [197] P. Siffert, y A. Coche, “Low Energy α -p Pulse-Shape Discrimination with Silicon Surface Barrier Detectors”, IEEE Trans. on Nucl. Sci. 13 (1966) 757-761.
- [198] N. J. Shenhav y H. Stelzer, “The mass dependence of the signal peak height of a Bragg-curve ionization chamber”, Nucl. Instrum. and Meth. 228 (1985) 359-364.
- [199] C. K. Shin, U. T. Yun, H. K. Kim y S. C. Park, “A Hybrid Approach of Neural Network and Memory-Based Learning to Data Mining”, IEEE Trans. on Neural Networks 11 (2000) 637-646.
- [200] K. A. Sjöland y P. Kristiansson, “Pile-up and defective pulse rejection by pulse shape discrimination in surface barrier detectors”, Nucl. Instrum. Meth B 94 (1994) 333-337.
- [201] W. Skulski, M. Momayezi, B. Hubbard-Nelson, P. Grudberg, J. Harris y W. Warburton, “Towards digital γ -ray and particle spectroscopy”, Acta Phys. Pol. B 31 (2000) 47-57.
- [202] W. Skulski y M. Momayezi, “Particle identification in CsI(Tl) using digital pulse shape analysis”, Nucl. Instrum. and Meth A 458 (2001) 759-771.
- [203] N. J. T. Smith, “Dark matter detectors and the use of interaction location information for background suppression”, Nucl. Instrum. and Meth. A 513 (2003) 215-221.
- [204] P. A. Söderström, J. Nyberg y R. Wolters, “Digital pulse-shape discrimination of fast neutrons and gamma rays”, Nucl. Instrum. and Meth. A 594 (2008) 79-89.
- [205] R. S. Storey, W. Sack y A. Warp, Proc. Phys. Soc. London 72 (1958) 1.
- [206] D. W. Stracener, D. G. Sarantites, L. G. Sobotka, J. Elson, J. T. Hood, Z. Majka, V. Abenante, A. Chbihi y D. C. Hensley, “Dwarf Ball and Dwarf Wall: Design, instrumentation, and response characteristics of a 4π CsI(Tl) plastic phoswich multidetector system for light charged particle and intermediate mass fragment spectrometry”, Nucl. Instrum. and Meth A 294 (1990) 485-503.
- [207] R. S. Sutton, A. G. Barto y R. J. Williams, “Reinforcement learning is direct adaptive optimal control”, Proc. of the 1991 American control conference, Evanston, Illinois; American Automatic Control Council (1991) 2143-2146.
- [208] N. Takeda, K. Kudo, S. Koshikawa, H. Ohgaki, H. Toyokawa, y T. Sugita, Proc. of the Second International Workshop on EGS, 8-12 Agosto 2000, Tsukuba, Japan KEK 200-20 (2000) 168-175.
- [209] P. J. Taddei, T. B. Borak, S. B. Guetersloh, B. B. Gersey, C. Zeitlin, L. Heilbronn, J. Miller, T. Murakami y Y. Iwata, “The response of a spherical tissue-equivalent proportional counter to different heavy ions having similar velocities” Rad. Meas. 41 (2006) 1227-1234.
- [210] E. L. Thorndike, “Animal Intelligence”, Darien Connecticut: Hafner (1911).
- [211] N. Tsoufanidis, Measurement and detection of radiation, Taylor & Francis (1995).

- [212] A. M. Turing, "On computable numbers with an application to the entscheidungsproblem", Proc. of the London Math. Soc., Serie 2, 42 (1936) 230-265; una corrección en 43 (1937) 544-546.
- [213] J. J. Vega, J. J. Kolata, W. Chung, D. J. Henderson y C. N. Davids, in: M. Brandan (Ed.), Proceedings of XIV Symposium on Nuclear Physics, Cuernavaca, Mexico, 1991, World Scientific, Singapore (1991) 221.
- [214] J. J. Vega, M. R. Reynoso, M. Arias y L. Altamirano, "Bragg Curve Identification Using a Neural Network", Shun-Ichi Amari, C. Lee Giles, Marco Gori, Vincenzo Piuri (Eds.), Proc. of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, IJCNN 2000, Neural Computing: New Challenges and Perspectives for the New Millennium, Como, Italy, July 24-27, 4 (2000) 379-382.
- [215] J. J. Vega y M. R. Reynoso, in: N. Callaos, W. Lesso, B. Sánchez, S. Li (Eds.), Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics, Vol. XIV, Computer Science Engineering Applications, Orlando, FL, USA, 27-30 July 2003, 396.
- [216] J. J. Vega y R. Reynoso, "Application of neural networks to pulse-shape analysis of Bragg curves", Nucl. Instrum. and Meth. B 243 (2006) 232-240.
- [217] J. J. Vega y R. Reynoso, "Effect of noise on the identification of digitized Bragg Curves", Rev. Mex. Fís. S 53(3) (2007) 118-124.
- [218] J. J. Vega, R. Reynoso y H. Carrillo-Calvet, "Learning limits of an artificial neural network", Rev. Mex. Fís. S 54(1) (2008) 22-29.
- [219] J. J. Vega, M. R. Reynoso y H. Carrillo Calvet, "Effect of signal noise on the learning capability of an artificial neural network". Nucl. Instrum. and Meth. A 606 (2009) 693-699.
- [220] S. Venkataramanan, A. Gupta, K. S. Golda, H. Singh, R. Kumar, R. P. Singh y R. K. Bhowmik, "A compact pulse shape discriminator module for large neutron detector arrays", Nucl. Instrum. and Meth. A 596 (2008) 248-252.
- [221] K. Vetter, A. L. Kuhn, M. A. Deleplanque, I. Y. Lee, F. S. Stephens, G. J. Schmid, D. Beckedahl, J. J. Blair, R. M. Clark, M. Cromaz, R. M. Diamond, P. Fallon, G. J. Lane, J. E. Kammeraad, A. O. Macchiavelli y C. E. Svensson, "Three-dimensional position sensitivity in two-dimensionally segmented HP-Ge detectors", Nucl. Instrum. and Meth. A 452 (2000) 223-238.
- [222] R. Vicente, O. Kinouchi y N. Caticha, "Statistical Mechanics of Online Learning of Drifting Concepts: A Variational Approach", Machine Learning 32 (1998) 179-201.
- [223] M. F. Vinyard, B. D. Wilkins, D. J. Henderson, D. G. Kovar, C. Beck, C. N. Davids y J. J. Kolata, "Performance of a large Bragg-curve spectrometer", Nucl. Instrum. and Meth. A 255 (1987) 507-511.

- [224] M. Z. Wang, Q. Yue, J. R. Deng, W. P. Lai, H. B. Li, J. Lib, Y. Liu, B. J. Qi, X. C. Ruan, C. H. Tang, H. Q. Tang, H. T. Wong, S. C. Wu, B. Xinc y Z. Y. Zhou, “Nuclear recoil measurement in CsI(Tl) crystal for cold dark matter detection”, *Phys. Lett. B* 536 (2002) 203-208.
- [225] W. K. Warburton, M. Momayezi, B. Hubbard-Nelson y W. Skulski, “Digital pulse processing: new possibilities in nuclear spectroscopy”, *Appl. Rad. and Isotopes* 53 (2000) 913-920.
- [226] W. K. Warburton y P. M. Grufberg, “Current trends in developing digital signal processing electronics for semiconductor detectors”, *Nucl. Instrum. and Meth. A* 568 (2006) 350-358.
- [227] N. Warr, “MINIBALL: The first gamma-ray spectrometer using segmented, encapsulated germanium detectors for studies with radioactive beams”, *Eur. Phys. J. A* 20 (2004) 65-66.
- [228] T. L. H. Watkin y A. Rau, “The statistical mechanics of learning a rule”, *Rev. of Mod. Phys.* 65 (1993) 499-556.
- [229] P. J. Werbos, “The Roots of Backpropagation”, NY: John Wiley & Sons., incluye la tesis doctoral de Werbos en Harvard de 1974, “Beyond Regression” (1974/1994).
- [230] B. Widrow y M. E. Hoff, “Adaptive switching circuits”, 1960 IRE WESCON Conventioin Record, New York (1960) 96-104.
- [231] D. H. Wilkinson, *ionization chambers and counters*, Cambridge Univ. Press (1950).
- [232] O. M. Williams y W. J. Sandle, “A pile-up gate generator for removing distortion in multi-channel delayed coincidence experiments”, *J. Phys. E* 3 (1970) 741-743.
- [233] D. J. Willshaw y C. Von Der Malsburg, “How patterned neural connections can be set up by self-organization”, *Proc. Of the Royal Society of London, Serie B* 194 (1976) 431-445.
- [234] H. T. Wong, “Prospects of scintillating crystal detector in low-energy low-background experiments”, *Astroparticle Physics* 14 (2000) 141-152.
- [235] K. Y. M. Wong, S. Li y Y. W. Tong, “Many-body approach to the dynamics of batch learning”, *Phys. Rev. E*, 62 (2000) 4036-4032.
- [236] Z. -B. Xu, R. Zhang y W. -F. Jing, “When Does Online BP Training Converge?”, *IEEE Trans. on Neural Networks* 20 (2009) 1529-1539.
- [237] E. Yoshida, K. Shizuma, S. Endo y T. Oka, “Application of neural networks for the analysis of gamma-ray spectra measured with a Ge spectrometer”, *Nucl. Instrum. and Meth. A* 484 (2002) 557-563.
- [238] S. Yousefi y L. Lucchese, “Digital Pulse Shape Discrimination in Triple-Layer Phoswich Detectors Using Fuzzy Logic”, *IEEE Trans. on Nucl. Sci.* 55 (2008) 2739-2748.

- [239] S. Yousefi y L. Lucchese, “A wavelet-based pulse shape discrimination method for simultaneous beta and gamma spectroscopy”, *Nucl. Instrum. and Meth. A* 599 (2009) 66-73.
- [240] H. Zhang, C. Zhang y W. Wu, “Convergence of Batch Split-Complex backpropagation Algorithm for Complex-Valued Neural Networks”, *Discrete Dynamics in Nature and Society*, Vol. 2009 (2009) 329173.
- [241] A. Zell, G. Mamier, M. Vogt, N. Mache, R. Hübner, S. Döring, K. Herrmann, T. Soyez, M. Schmalzl, T. Sommer, A. Hatzigeorgiou, D. Posselt, T. Schreiner, B. Kett, G. Clemente, J. Wieland, J. Gatter, M. Reczko, M. Riedmiller, M. Seemann, M. Ritt, J. DeCoster, J. Biedermann, J. Danz, C. Wehrfritz, R. Werner, M. Berthold y B. Orsier, “SNNS, Stuttgart Neural Network Simulator”, Version 4.2, University of Stuttgart, Institute for Parallel and Distributed High Performance Systems y University of Tübingen, Wilhelm-Schickard-Institute for Computer Science, 1998.