**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE CIENCIAS FÍSICAS**

DEPARTAMENTO DE FÍSICA DE MATERIALES




# TRABAJO DE FIN DE GRADO

Código de TFG: FM10

Nano-dispositivos para sistemas de computación

Nano-devices for Computing

Supervisor/es: Miguel Romera y Lucas Pérez

**Adrián Benito Ayala**

Grado en Física

Curso académico 2025-26

Convocatoria Ordinaria de Junio

Calificación: 9.2

**Resumen:**

Este trabajo de fin de grado expone los fundamentos físicos y tecnológicos de la computación neuromórfica como solución a las crecientes limitaciones energéticas y estructurales de la tecnología CMOS y la arquitectura clásica de Turing-von Neumann. Inspirada en el funcionamiento del cerebro, esta alternativa tecnológica propone el procesamiento masivamente paralelo y la co-localización de memoria y unidades de procesamiento mediante nano-dispositivos electrónicos basados en óxidos de metales de transición: los neuristores y los sinaptores. Por un lado, los neuristores emulan la dinámica neuronal de integración y disparo con pérdidas (LIF) aprovechando la conmutación resistiva volátil (VRS) en aislantes de Mott, donde se inducen transiciones de fase metal-aislante reversibles. Por otro lado, los sinaptores replican la plasticidad y memoria de las sinapsis biológicas mediante la conmutación resistiva no volátil (NVRS). Estos últimos operan bajo los principios teóricos del memristor, alterando su conductancia mediante procesos de migración iónica y formación de filamentos para implementar reglas de aprendizaje fundamentales como la plasticidad dependiente del tiempo entre impulsos (STDP). El trabajo concluye que la física subyacente a estos materiales tiene potencial para permitir reproducir analógicamente las capacidades de las redes neuronales con una extraordinaria eficiencia energética, allanando el camino hacia una inteligencia artificial sostenible en hardware.

**Abstract:**

This final degree project presents the physical and technological foundations of neuromorphic computing as a solution to the growing energetic and structural limitations of CMOS technology and the classical Turing-von Neumann architecture. Inspired by the functioning of the brain, this technological alternative proposes massively parallel processing and the co-location of memory and computation through electronic nano-devices based on transition metal oxides: neuristors and synaptors. On the one hand, neuristors emulate the leaky integrate-and-fire (LIF) neuronal dynamics by exploiting volatile resistive switching (VRS) in Mott insulators, where reversible metal-insulator phase transitions are induced. On the other hand, synaptors replicate the plasticity and memory of biological synapses through non-volatile resistive switching (NVRS). The latter operate under the theoretical principles of the memristor, altering their conductance through ion migration and filament formation processes to implement fundamental learning rules such as spike-timing-dependent plasticity (STDP). The study concludes that the underlying physics of these materials show potential for the analog reproduction of neural network capabilities with extraordinary energy efficiency, paving the way toward sustainable artificial intelligence in hardware.

# Índice

**1. Introducción** **4**

**2. Objetivos** **5**

**3. Metodología** **5**

**4. Fundamentos de la Computación Neuromórfica** **6**

**5. Electrónica de óxidos: Fenomenología de la conmutación resistiva** **8**
5.1. Conmutación Resistiva Volátil (VRS) . . . . . 9
5.2. Conmutación Resistiva No Volátil (NVRS) . . . . . 11
5.2.1. NVRS Unipolar . . . . . 12
5.2.2. NVRS Bipolar . . . . . 13
5.3. Memristor . . . . . 14

**6. Implementaciones** **16**
6.1. Neuristor . . . . . 16
6.2. Sinaptor . . . . . 17

**7. Conclusiones** **20**

# 1. Introducción

Desde la invención del transistor en 1947, nuestra sociedad ha logrado grandes avances impulsados por el desarrollo y la implantación masiva del microprocesador. Atrás queda ya la famosa *"Ley de Moore"*[1], que vaticinaba la duplicación del número de transistores de un microprocesador cada dos años, de manera que nos encontramos cada vez más cerca del límite de las arquitecturas tradicionales de semiconductores de óxidos metálicos[2]. La crisis de la tecnología CMOS (*Complementary Metal-Oxide Semiconductor*) contemporánea presenta un origen dual. Desde una perspectiva mecanocuántica, la miniaturización de los transistores a escalas nanométricas exige reducir el espesor del dieléctrico de puerta a unas pocas capas atómicas[3]. Esta constricción geométrica dispara la probabilidad de que los electrones atraviesen la barrera de potencial por efecto túnel[4]. La consecuencia crítica de este fenómeno es la aparición de corrientes de fuga masivas a través del terminal de puerta: el transistor pierde su capacidad para bloquear el paso de corriente en el estado de reposo (*off-state*), degradando su comportamiento como interruptor ideal y provocando un consumo de potencia estática[4]. Por otro lado, desde una perspectiva termodinámica, el fracaso del escalado de Dennard[5] en los dispositivos MOSFET (*Metal–Oxide–Semiconductor Field-Effect Transistor*) implica que ya no es posible aumentar la frecuencia de reloj de los procesadores sin provocar un colapso térmico[5]. Todo esto conlleva que, cada vez más, el enfoque actual basado en producir chips más densos y con transistores más pequeños esté dejando de funcionar. A los límites físicos del hardware subyacente se suma una ineficiencia arquitectónica estructural. La inmensa mayoría de los sistemas informáticos actuales operan bajo la arquitectura de *Turing-von Neumann* (TvN). Esta arquitectura presenta la segmentación e interconexión de distintas unidades. Los dos componentes principales son la unidad central de procesamiento (CPU) y la memoria, que están conectados al usuario mediante dispositivos de entrada y salida. En este esquema, cualquier operación requiere transportar los datos y las instrucciones a través de un único canal de comunicación limitado[6].

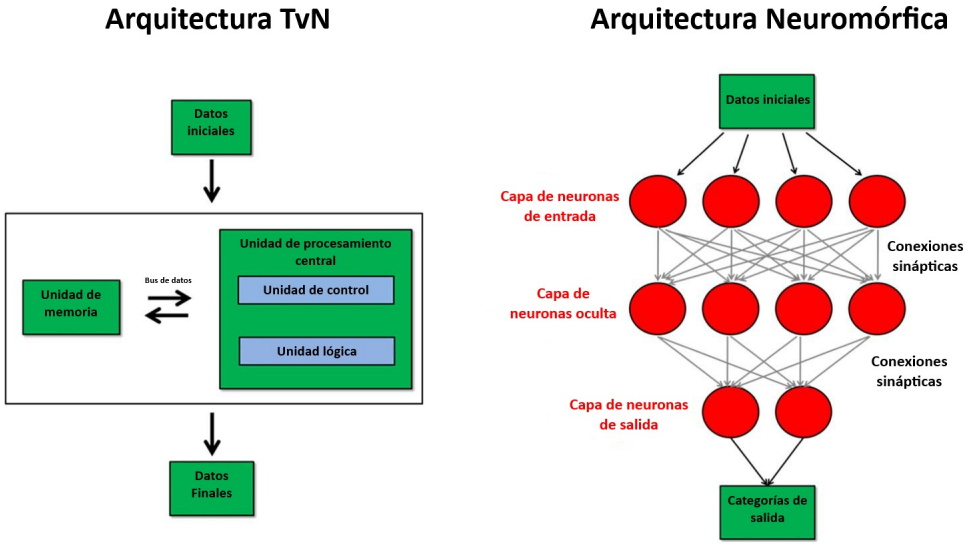



Figura 1: Comparación entre las arquitecturas Turing-von Neumann y neuromórfica. Figura traducida y adaptada de la referencia [7].

Esta ineficiencia arquitectónica se acentúa ante el auge de la Inteligencia Artificial. A pesar de que los algoritmos de aprendizaje profundo se fundamentan en la estructura de redes neuronales, su ejecución actual se reduce a una emulación por software usando dispositivos de tipo convencional. De este modo, el coste energético de procesar información compleja mediante emulación digital se ha convertido en el principal obstáculo para el desarrollo de una inteligencia artificial sostenible.

Para sortear estos límites fundamentales, surge la computación neuromórfica como una alternativa inspirada en el cerebro[8,9]. Si comparamos a este con un ordenador convencional, se puede observar que el primero es capaz de realizar tareas de una complejidad inmensa con una eficiencia energética inalcanzable incluso para los dispositivos más punteros del mundo[10]. Esta nueva forma de diseñar los dispositivos electrónicos presenta grandes diferencias con respecto a la arquitectura TvN. Como se puede observar en la Figura 1, la arquitectura neuromórfica (derecha) elimina la frontera entre el procesamiento lógico y el almacenamiento de datos. Inspirada en la biología, una red neuronal artificial se compone de unidades de procesamiento simples (neuronas) dispuestas habitualmente en capas y unidas mediante una red masiva de interconexiones (sinapsis). A nivel operativo, esto permite un modo de procesamiento intrínsecamente paralelo, como contraposición a los dispositivos convencionales, que lo hacen de manera secuencial. Computar en paralelo conlleva ventajas cruciales, como la tolerancia a los fallos físicos: en una arquitectura TvN convencional, la precisión se garantiza forzando cálculos con alrededor de 16 cifras significativas. Por el contrario, los sistemas neuromórficos poseen unidades que, aunque individualmente imprecisas o ruidosas, integran la respuesta de forma colectiva, lo que otorga una gran robustez estadística a nivel de red[7].

El objetivo de la computación neuromórfica es, entonces, simular las funciones más importantes y útiles del cerebro y conseguir adaptarlas a dispositivos experimentales, como una alternativa al paradigma de Turing-von Neumann para lograr avances en eficiencia energética imposibles con el paradigma actual.

## 2. Objetivos

El objetivo principal de este Trabajo de Fin de Grado es asentar las bases físicas de la computación neuromórfica y estudiar cómo el uso de nuevos materiales de estado sólido permite desarrollar hardware inspirado en el funcionamiento del cerebro.

Para lograrlo, el desarrollo de la memoria se estructura en torno a los siguientes objetivos específicos:

- Explicar los principios teóricos y la arquitectura de la computación neuromórfica, contrastando su forma de operar con la arquitectura tradicional de Turing-von Neumann.
- Describir la dinámica física necesaria para emular las funcionalidades biológicas clave, tales como la integración y el disparo neuronal (LIF) o los mecanismos de aprendizaje y plasticidad sináptica (LTP, LTD y STDP).
- Presentar la fenomenología de la conmutación resistiva dentro de la electrónica de óxidos, detallando los mecanismos termodinámicos y de transporte de carga subyacentes, tanto para el régimen no volátil (NVRS) como para el volátil (VRS).
- Analizar la física de los componentes neuromórficos elementales, concretamente los neuristores y sinaptores basados en óxidos, evaluando su capacidad para replicar los impulsos eléctricos de las neuronas biológicas y la retención de memoria de las sinapsis.

## 3. Metodología

La metodología empleada para la realización de este Trabajo de Fin de Grado se fundamenta en una revisión bibliográfica exhaustiva y en el análisis teórico de los fundamentos de la computación neuromórfica y la electrónica de óxidos. Al tratarse de un trabajo puramente teórico, la metodología usada está basada en la recopilación, el filtrado y la síntesis crítica de la literatura científica.

Para recopilar la información necesaria se han consultado las principales bases de datos y repositorios académicos en el ámbito de la física del estado sólido y la ingeniería de materiales. Destacan entre ellas *IEEE* y *Journal of Applied Physics* y las publicaciones de editoriales como *Nature* o la *American Physical Society*. La estrategia de búsqueda se ha formado combinando términos generales con palabras clave como conmutación resistiva, aislantes de Mott y memristor.

El proceso de selección de las fuentes se ha regido por dos criterios: la relevancia histórica y el estado del arte. Por un lado, se ha acudido a los artículos originales que fundamentaron las bases de la computación neuromórfica y los límites de la computación clásica. Por otro lado, para el estudio de los dispositivos basados en la electrónica de óxidos, se ha priorizado la literatura de la última década, consiguiendo con esto garantizar que la discusión sobre los mecanismos y conceptos estudiados refleje fielmente los problemas y el consenso de la física actual.

Finalmente, el trabajo se ha centrado en extraer y conectar los conceptos físicos que rigen el funcionamiento de los dispositivos neuromórficos. De esta forma, se consigue enlazar conceptos como la transición de fase aislante-metal o la migración de vacantes de oxígeno con el comportamiento macroscópico de integración, disparo y aprendizaje, tendiendo un puente entre la física de materiales emergentes y la topología de redes neuronales.

# 4. Fundamentos de la Computación Neuromórfica

El objetivo de la computación neuromórfica no reside en replicar con exactitud un sistema biológico, sino en extraer sus características operativas esenciales para el diseño de circuitos eficientes. De esta manera, para comprender los requisitos y características básicas que se pretenden replicar con los dispositivos neuromórficos, es necesario establecer una equivalencia directa entre los elementos biológicos fundamentales y sus contrapartes neuromórficas.

La neurona opera como la unidad activa y fundamental en el procesamiento de información del tejido nervioso, basando su funcionamiento dinámico en la recepción, integración y transmisión de impulsos eléctricos discretos[10]. Estructuralmente, la neurona recibe múltiples señales de entrada a través de ramificaciones denominadas dendritas, las cuales convergen en el cuerpo celular o soma para su procesamiento[7].

Físicamente, el comportamiento eléctrico de la membrana celular se formaliza mediante el modelo de *integración y disparo con pérdidas* (LIF, del inglés *Leaky Integrate-and-Fire*), el cual describe la membrana como un circuito análogo a un condensador en paralelo con una resistencia de fuga, propuesto originalmente por Lapicque[11]. Como se ilustra en la evolución del potencial de la Figura 2b, cuando las señales de entrada llegan a la neurona, inyectan una carga eléctrica que se integra de manera temporal, incrementando el potencial de la membrana, sin embargo, la neurona no es un integrador perfecto: el sistema disipa o pierde lenta y continuamente la carga acumulada con el transcurso del tiempo[7,12]. El rasgo definitorio de la dinámica neuronal es su respuesta *no lineal dependiente de la frecuencia*, operando bajo un mecanismo de umbral. Si los estímulos eléctricos de entrada llegan muy separados en el tiempo, la fuga constante domina el proceso y el potencial de membrana no logra acumularse significativamente. Por el contrario, si los impulsos incidentes tienen una frecuencia o intensidad lo suficientemente alta como para contrarrestar la relajación natural del sistema, el voltaje acumulado rebasa un umbral de potencial crítico[7,12]. Esto puede apreciarse en la Figura 2b: se puede ver cómo los impulsos sucesivos recibidos causan el aumento del potencial de la membrana celular. En el instante en el que se supera el umbral, el estado del sistema cambia abruptamente y la neurona "dispara", generando una señal eléctrica aguda denominada *potencial de acción*, visible como el pico de voltaje en la Figura 2b. Atendiendo a la Figura 2a, este impulso se propaga a lo largo de una única estructura de salida, el axón, actuando como vía de conducción para transmitir la señal hacia otras neuronas. La comunicación con las neuronas adyacentes tiene

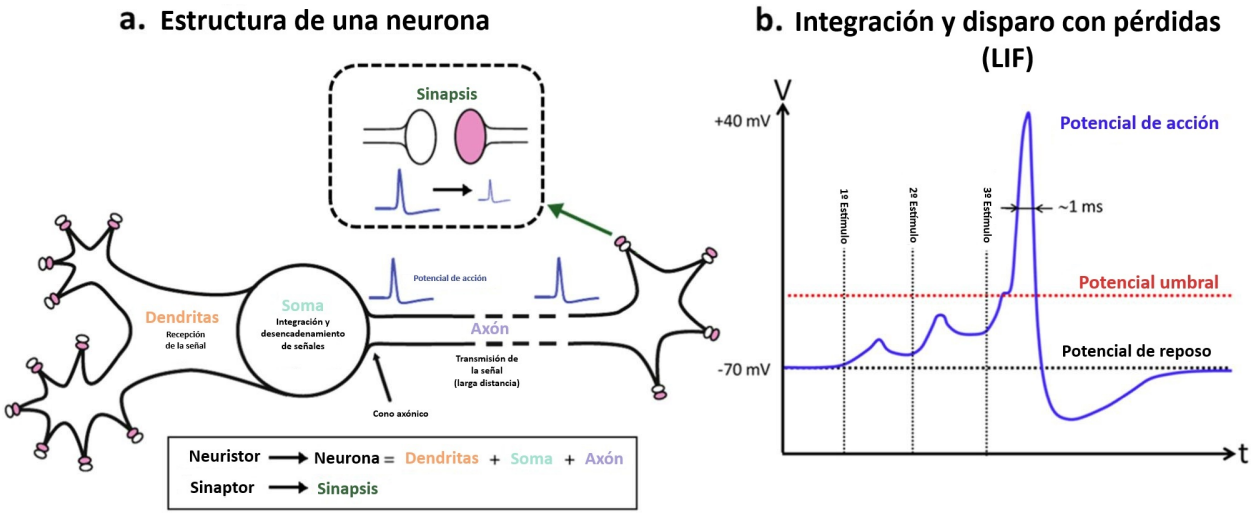



Figura 2: Estructura eléctrica equivalente (a) y generación de espigas de voltaje (b) de una neurona, ilustrando el modelo de integración y disparo con pérdidas (LIF) tras alcanzar un potencial umbral. Figura traducida y adaptada de [7] (adaptación original del panel (a) de van Dongen y Serdijn, 2016).

lugar en uniones funcionales microscópicas llamadas sinapsis, donde el terminal del axón de la neurona emisora (presináptica) se conecta con las dendritas de la neurona receptora (postsináptica), regulando la intensidad de la señal transferida. Inmediatamente después de la emisión del potencial de acción, el voltaje de la membrana experimenta una caída abrupta y es forzado a restablecerse a su valor de reposo inicial, un estado que se mantiene inactivo temporalmente durante el denominado periodo refractario antes de que la neurona pueda iniciar un nuevo ciclo de integración[7,12].

Las conexiones sinápticas no actúan como meros enlaces estructurales, sino que desempeñan un papel fundamental en la regulación de la fracción de la señal que se transfiere. En esta unión, la conductancia del nodo determina la intensidad de la interconexión, un parámetro que a nivel macroscópico se denomina *peso sináptico*. La modulación dinámica de estos pesos a lo largo del tiempo, conocida como plasticidad sináptica, es precisamente el mecanismo físico que dota a la red de su capacidad para el aprendizaje y el almacenamiento de memoria[7].

La evolución de estos pesos está gobernada por la correlación temporal entre los impulsos eléctricos (potenciales de acción) generados por las neuronas interconectadas. Si el impulso presináptico llega a la sinapsis justo antes de que la neurona postsináptica dispare, la fuerza de la conexión sináptica aumenta, lo que se traduce en una mayor conductancia. Este efecto se llama *potenciación a largo plazo* (LTP). Si el orden temporal se invierte y la neurona postsináptica dispara su potencial de acción antes de que llegue la activación presináptica, la eficacia de la sinapsis se reduce de forma persistente, es decir, disminuye la conductancia. A este efecto se le conoce como *depresión a largo plazo* (LTD)[7,13,14].

Para entender mejor el proceso de modificación de los pesos sinápticos, resulta conveniente revisar de nuevo la potenciación a largo plazo. Para que esta ocurra, el impulso de la neurona presináptica debe llegar a la sinapsis justo antes de que la neurona postsináptica genere y emita su propio disparo[7,13,14]. Al generarse este impulso postsináptico, la señal viaja físicamente a través de la neurona hacia la sinapsis y confluye con el impulso presináptico. Es precisamente esta coincidencia espacio-temporal la que permite a la sinapsis modificar su conductancia[7,13]. Es importante remarcar también que, a pesar de que el potencial emitido por la neurona postsináptica llegue en efecto a la sinapsis, no hay una transmisión del impulso: las sinapsis actúan de manera direccional. La función de este pulso postsináptico en la sinapsis es la de modular el peso[14,15]. De esta manera,

la magnitud y dirección del cambio de los pesos sinápticos dependen de dos factores. Primero, el orden de llegada de los impulsos, lo que determina si se va a producir depresión o potenciación. Segundo, el tiempo entre impulsos: cuanto más cercanos en el tiempo lleguen a la sinapsis estos dos impulsos, más abrupto será el cambio en conductancia de la sinapsis[13,14,15].

Este fenómeno de regulación de los pesos sinápticos se denomina *plasticidad dependiente del tiempo entre impulsos* (STDP, por sus siglas en inglés *Spike-Timing-Dependent Plasticity*), y constituye un mecanismo esencial que los dispositivos neuromórficos deben emular para reproducir fielmente el aprendizaje biológico[7,14,13].

Establecido este marco teórico, es posible clasificar los componentes fundamentales que estructuran el hardware neuromórfico. Los *neuristores* son los elementos activos encargados de emular la dinámica neuronal mediante una respuesta no lineal. Deben replicar el modelo de integración y disparo con pérdidas (LIF). Los *sinaptores* operan como las sinapsis electrónicas que controlan la interconexión entre neuristores. La conductancia del dispositivo actúa como el peso sináptico, mientras que su retención de estado (no volatilidad) proporciona la memoria necesaria para implementar la plasticidad de la red, incluyendo procesos como LTP, LTD y STDP.

## 5. Electrónica de óxidos: Fenomenología de la conmutación resistiva

Para materializar los componentes requeridos por la computación neuromórfica, una de las ramas de estudio más prominentes es el uso de materiales que presentan *conmutación resistiva* (*Resistive Switching*, RS)[7]. Este fenómeno se caracteriza por la modulación de la resistencia eléctrica de un dispositivo en respuesta tanto al voltaje instantáneo como al historial de tensión aplicada. Si bien la RS se manifiesta en múltiples sistemas materiales, los óxidos de metales de transición (TMO) destacan como los candidatos idóneos para estas aplicaciones[7,16]. Estos son de especial interés debido a su amplio rango de propiedades físicas, que incluyen superconductividad, magnetorresistencia, termoelectricidad, ferromagnetismo, transiciones aislante-metal y, la más importante para la computación neuromórfica, la conmutación resistiva[7].

La naturaleza de los orbitales $d$ en los metales de transición da lugar a múltiples estados de valencia y a diferentes coordinaciones con el oxígeno en la red cristalina. Esta flexibilidad química y electrónica actúa por dos vías físicas distintas: por un lado, permite cambios puramente electrónicos en la ocupación de las bandas de conducción (lo que propicia la conmutación volátil) y, por otro, facilita la creación y migración de defectos estructurales metaestables (la base de la conmutación no volátil). Es precisamente esta doble capacidad la responsable directa de que los óxidos de metales de transición destaquen por presentar de forma intrínseca ambos tipos de conmutación resistiva[7].

La conmutación del estado de alta resistencia (HRS) al estado de baja resistencia (LRS) se denomina *set*, y el cambio del estado de baja resistencia al estado de alta resistencia se conoce como *reset*.

En la *conmutación resistiva no volátil* (*Non-Volatile Resistive Switching*, NVRS), la alteración de la resistencia del material tiene un carácter semipermanente, reteniéndose incluso en ausencia del campo eléctrico que ha inducido el cambio. Para revertir este estado resistivo se requiere aplicar un nuevo estímulo externo. Dependiendo de la polaridad de este segundo pulso de voltaje con respecto al pulso inicial se diferencian dos tipos de NVRS: la unipolar y la bipolar. La NVRS unipolar implica que la dirección de la conmutación depende de la amplitud del voltaje aplicado, pero no de la polaridad de dicho voltaje. Por ende, los procesos de *set* y *reset* pueden ocurrir con la misma polaridad. Por otro lado, la NVRS bipolar implica que la dirección de la conmutación depende de la polaridad del voltaje aplicado y, al contrario que la unipolar, el proceso de *set* ocurre con una polaridad y el proceso de *reset* ocurre con la polaridad inversa[17,16,7]. Se puede ver en

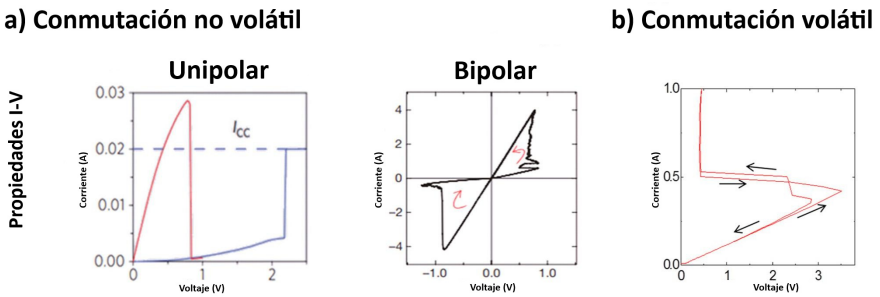



Figura 3: Curvas características de corriente-voltaje (I-V) de los tres regímenes principales de conmutación resistiva. El panel (a) ilustra los comportamientos no volátiles: el régimen unipolar (izquierda) y el régimen bipolar (derecha). El panel (b) muestra el régimen volátil, donde el material retorna espontáneamente a su estado de alta resistencia al disminuir el voltaje. Figura adaptada y traducida de la referencia [7].

la Figura 3 la curva característica de intensidad-voltaje de la NVRS unipolar (izquierda), en este caso los procesos de *set* (línea azul) y *reset* (línea roja) se producen de forma independiente a la polaridad del estímulo, de modo que ocurren en el mismo cuadrante de la gráfica. Por el contrario, la conmutación no volátil bipolar (centro) se caracteriza por una curva $I - V$ asimétrica que traza un lazo de histéresis cruzando por el origen. En este caso, transitar entre estados exige alternar obligatoriamente el signo del estímulo eléctrico. Estas características permiten diseñar sinaptores: la conductancia del dispositivo determina el peso sináptico, mientras que la retención del estado imita la memoria y la plasticidad buscadas[7].

En la *conmutación resistiva volátil* (*Volatile Resistive Switching*, VRS), se induce el cambio entre los estados resistivos cuando se aplica un voltaje superior a un umbral crítico ($V_{th}$, *threshold voltage*), y el material vuelve de forma natural a su estado basal aislante cuando el pulso eléctrico termina. Esto está esquematizado de nuevo en la Figura 3: al principio el material ofrece una gran resistencia, lo que hace que la intensidad crezca lentamente; cuando llega al umbral crítico se produce la conmutación, la resistencia baja y se produce conducción a bajo voltaje. Este comportamiento permite diseñar neuristores, ya que la transición abrupta hacia el estado conductor proporciona la no linealidad necesaria para emular el disparo neuronal, mientras que la relajación espontánea reproduce físicamente la disipación de energía y el periodo refractario intrínsecos al modelo de integración y disparo con pérdidas[18].

## 5.1. Conmutación Resistiva Volátil (VRS)

En el contexto de los dispositivos de conmutación resistiva volátil basados en óxidos de metales de transición, resulta imperativo abordar la física de los *aislantes de Mott*. Estos materiales se caracterizan por ser aislantes que poseen la capacidad de convertirse en conductores bajo determinadas condiciones.

De acuerdo con la teoría de bandas convencional basada en el teorema de Bloch, los materiales cuyo nivel de Fermi se sitúa dentro de una banda prohibida (es decir, poseen una banda de valencia completamente llena separada de una banda de conducción vacía) se caracterizan por su comportamiento aislante. Por el contrario, aquellos materiales que presentan bandas de energía parcialmente llenas deben exhibir intrínsecamente propiedades metálicas. A diferencia de los aislantes de bandas convencionales, los *aislantes de Mott* poseen electrones desapareados en su estado fundamental y, sin embargo, exhiben un comportamiento aislante. La condición estricta para que emerja este estado es que el material debe presentar un número entero de electrones localizados por cada sitio de la red[19]. Al estar todos los sitios atómicos sometidos a este llenado entero, la conducción de carga requiere el salto (*hopping*) de un electrón hacia un sitio vecino, forzando allí una

configuración de doble ocupación local. Este proceso conlleva una enorme penalización energética debido a la fuerte repulsión coulombiana interelectrónica, definida por el *parámetro de Hubbard* $U$. Si este coste repulsivo $U$ domina sobre la energía cinética de deslocalización de los electrones (típicamente representada por el ancho de banda $W$), el sistema no dispone de energía suficiente para generar sitios doblemente ocupados. En este escenario, la banda semillena se separa en dos bandas llamadas *banda inferior de Hubbard* (LHB) y *banda superior de Hubbard* (UHB), con una separación $E_g \approx U - W$ (*brecha de Mott*) y el material se comporta como un aislante[19].

Para el desarrollo de los dispositivos de conmutación resistiva volátil, es necesario estudiar los diferentes mecanismos que producen estas transiciones. Las transiciones pueden estar impulsadas térmica o eléctricamente y presentar distintos cambios estructurales. Una *ruptura de la simetría estructural* (espacial) ocurre cuando un material experimenta un cambio abrupto en la red de Bravais y la base atómica por la que pasa de tener una estructura cristalina muy ordenada (alta simetría) a una estructura donde algunos de esos elementos de orden se pierden o se degradan (baja simetría). Por otro lado, una *ruptura de la simetría magnética* (inversión temporal) ocurre cuando los espines de los electrones (los momentos magnéticos locales) se ordenan de forma cooperativa, pasando, por ejemplo, de un estado paramagnético (desordenado) a un estado ferromagnético o antiferromagnético (ordenado).

El primer grupo de interés es el de las *transiciones impulsadas por temperatura*. Dentro de este se distinguen dos casos. La primera categoría consta de las transiciones *sin cambios estructurales*. Un ejemplo representativo de material que presenta estas transiciones puramente térmicas es el $V_2O_3$ dopado con Cr[7,19]. Esta transición obedece a la relación entre la temperatura y la fuerza de correlación electrónica del material ($U/W$). En sistemas con $U/W \approx 1{,}15$, la ligera inclinación de la frontera de Mott en el diagrama de fases permite que un incremento térmico la atraviese, induciendo una transición de metal a aislante[19]. La segunda categoría de transiciones impulsadas por la temperatura engloba aquellas que van acompañadas de *cambios estructurales*. Un material que experimenta este tipo de transiciones es el $VO_2$[7,19].

El segundo grupo de interés es el de las *transiciones impulsadas eléctricamente*. En este caso existe un debate considerable en la literatura científica respecto a su verdadero mecanismo[7]. Primeramente, es necesario aclarar que cuando se aplica un voltaje o una corriente a un dispositivo, la transición puede desencadenarse por dos vías distintas. La explicación más sencilla sugiere que la corriente genera un calentamiento de Joule que eleva la temperatura local del material por encima de su punto crítico, induciendo así la transición de aislante a metal. Bajo este supuesto, el mecanismo sigue siendo impulsado por la temperatura. Sin embargo, aunque este modelo térmico es válido para ciertos dispositivos, no puede explicar la fenomenología de algunos materiales, los cuales carecen de una transición de fase impulsada por temperatura[7,19]. Para estos casos, se ha demostrado que la transición es puramente electrónica: es la intensidad del campo eléctrico la que desestabiliza directamente la fase aislante, forzando la transición a la fase metálica y produciendo la conmutación resistiva[7,19]. Las *transiciones de ruptura por avalancha electrónica* ocurren cuando un campo eléctrico intenso inyecta portadores a través de la brecha de Mott. Al ganar suficiente energía cinética, estos electrones inyectados colisionan contra los electrones fuertemente localizados en la red, desencadenando un proceso de ionización por impacto. Esta multiplicación de portadores en cascada incrementa drásticamente la densidad electrónica local, lo que apantalla la repulsión coulombiana de Hubbard ($U$) y provoca el colapso abrupto de la brecha de energía. De este modo, se destruye el estado aislante de Mott sin necesidad de alcanzar la temperatura crítica de la transición térmica[19]. Un material que presenta este tipo de transiciones es el $GaTa_4Se_8$, que se discutirá más adelante.

Es importante remarcar que el gran atractivo de los aislantes de Mott para la conmutación resistiva volátil (VRS) reside en su capacidad para transicionar rápida y reversiblemente de un estado a otro mediante cambios puramente electrónicos. Sin embargo, como se explorará en la siguien-

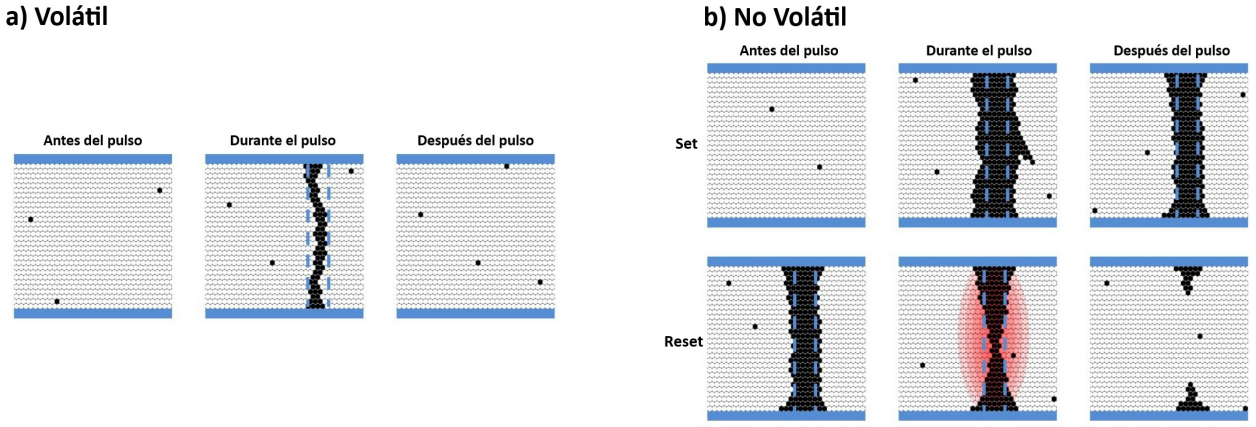



Figura 4: Ilustración esquemática de la evolución del filamento antes, durante y después de la aplicación de un pulso eléctrico. Se muestran tres casos: una transición volátil (a), una transición no volátil (b) *set* (arriba) y una transición no volátil *reset* (abajo). Los dominios blancos y negros representan, respectivamente, las regiones de aislante de Mott y de metal correlacionado del material. Los electrodos superior e inferior se representan en azul, y las líneas discontinuas azules indican el radio crítico de estabilidad para el filamento. Figura adaptada y traducida de la referencia [19].

te sección, numerosos aislantes de Mott también pueden exhibir conmutación resistiva no volátil (NVRS). En gran parte de los óxidos de transición clásicos, esta retención del estado resistivo ya no se fundamenta en la transición de fase electrónica (aislante-metal), sino en alteraciones semipermanentes de la estructura química y atómica del cristal. En estos casos, la viabilidad neuromórfica está gobernada por el movimiento físico de los átomos, a través de mecanismos como la deriva de vacantes de oxígeno o la formación de filamentos termoquímicos[7]. No obstante, cabe destacar que la frontera no es estricta: bajo ciertas condiciones, el propio cambio de estado electrónico puede estabilizarse y volverse no volátil, demostrando la enorme versatilidad de estos materiales[19]. Para ilustrar esta transición entre regímenes, resulta de gran utilidad observar la dinámica espacial de los dominios en el interior de un aislante de Mott puro. Como se esquematiza en la Figura 4, la frontera entre el comportamiento volátil y el no volátil se rige por un principio de tamaño crítico. Durante un evento de conmutación volátil, el estímulo eléctrico genera un filamento conductor compuesto por regiones de metal correlacionado dentro de la matriz del aislante de Mott. Sin embargo, este filamento inicial es demasiado estrecho: al no lograr superar el radio crítico de estabilidad, la estructura colapsa de forma espontánea devolviendo el material a su estado aislante cuando cesa el pulso eléctrico. Por el contrario, para conseguir un estado no volátil o *set*, es necesario aplicar estímulos que fuercen una mayor acumulación de sitios metálicos, logrando así que el filamento se ensanche. Al expandirse más allá del radio crítico, la estructura conductora adquiere la estabilidad suficiente para sobrevivir de forma persistente tras el pulso. Por último, para borrar esta memoria o realizar la transición *reset*, los modelos teóricos actuales proponen que el mecanismo subyacente es una disolución térmica del filamento. Esta hipótesis está fuertemente respaldada por evidencias experimentales: la aplicación de un pulso de mayor duración logra el autocalentamiento local de Joule necesario para destruir la ruta conductora. Además, dado que los electrodos del dispositivo actúan como sumideros de calor, la temperatura máxima se concentra en la región central del filamento, provocando que sea allí donde la estructura colapsa y se disuelve para restaurar la alta resistencia original[7,19].

## 5.2. Conmutación Resistiva No Volátil (NVRS)

El origen físico de la conmutación resistiva no volátil reside principalmente en la deriva iónica, aunque puede tener otros orígenes como se explicó con anterioridad. Este es un mecanismo facilitado

por la multiplicidad de estados de valencia del metal de transición. Una particularidad crítica de este fenómeno es que requiere la creación inicial de defectos estructurales y la posterior inducción de su migración bajo un campo eléctrico. Por consiguiente, la NVRS no está inducida por una transición termodinámica de fase global de la red cristalina, sino por el atrapamiento de los iones en diferentes configuraciones espaciales, dando lugar a una multiplicidad de estados metaestables[7,19]. Dado que la formación y ruptura de estas rutas de conducción iónica es un proceso estocástico, la conmutación entre los estados de alta y baja resistencia presenta una variabilidad intrínseca. Los dos casos distintos de NVRS surgen de cambios estructurales microscópicos diferenciados[7].

La conmutación resistiva no volátil unipolar está gobernada por el *mecanismo termoquímico* (TCM). Tras una formación inicial de los filamentos conductores, se produce una alternancia entre la desconexión de estos filamentos debido al extremo calentamiento de Joule (*reset*) y su reconexión mediante una ruptura dieléctrica localizada (*set*). Al ser tanto el calor de Joule generado en la desconexión como la ruptura dieléctrica producida en la reconexión fenómenos independientes de la polaridad, la modulación de la resistencia del dispositivo también lo es[7,16]. En claro contraste con la difusión térmica que caracteriza al mecanismo termoquímico, la NVRS bipolar presenta una dependencia estricta de la polaridad del campo eléctrico, lo que indica un transporte iónico direccional (deriva iónica)[16]. Este comportamiento se manifiesta de dos formas principales. Por un lado, según lo descrito en ciertos óxidos complejos conductores, puede originarse en las interfaces con los electrodos, donde la migración iónica modula la altura de barreras altamente resistivas (barreras de Schottky)[7]. Por otro lado, en los óxidos binarios (como el $HfO_2$ o el $Ta_2O_5$), la conmutación bipolar está dominada por el *mecanismo de cambio de valencia* (VCM). La diferencia fundamental entre el VCM y el TCM explicado anteriormente radica en su fuerza impulsora: mientras que en el TCM la ruta conductora se disuelve mediante reacciones redox impulsadas por el intenso calentamiento, en el VCM el campo eléctrico arrastra direccionalmente las vacantes de oxígeno cargadas para construir o retraer el filamento conductor nanométrico a través de la matriz aislante, alterando la estequiometría local del material[19,16].

### 5.2.1. NVRS Unipolar

La NVRS de tipo unipolar es característica de los óxidos de metales de transición binarios tales como $TiO_2$, $NiO$, $CuO$ o incluso $HfO_2$ y $Ta_2O_5$ bajo condiciones específicas de operación. Dado que en su estequiometría ideal estos compuestos son aislantes, la puesta en marcha de los dispositivos requiere un paso previo de activación conocido como *electroformación*. Este proceso exige aplicar voltajes superiores a los de un ciclo normal de *set* para generar campos eléctricos en el rango de los MV/cm. Estos son tan elevados que fuerzan una ruptura dieléctrica suave en el seno del aislante, desencadenando una migración masiva de especies iónicas, predominantemente vacantes de oxígeno, que establecen los primeros canales conductores[7]. Se ha demostrado que el tiempo necesario para completar este formado depende exponencialmente del voltaje aplicado. Esto confirma que la electroformación no ocurre de manera espontánea al alcanzar un voltaje crítico, sino que es un proceso acumulativo impulsado por la paulatina generación de defectos estructurales[17]. En la Figura 5 se ilustra el estado inicial antes del proceso de electroformación y la diferencia entre estados *set* y *reset*. Durante el proceso de *reset*, la disolución de los filamentos es un fenómeno impulsado fundamentalmente por la corriente a través del extremo calentamiento de Joule[7]. La resistencia local $R$ de un segmento del filamento viene dada por la ley de Pouillet:

$$R = \rho \frac{l}{A} \tag{1}$$

donde $\rho$ es la resistividad intrínseca del filamento, $l$ es la longitud del segmento y $A$ es su sección transversal. Como se observa, la resistencia es inversamente proporcional al área. De acuerdo con

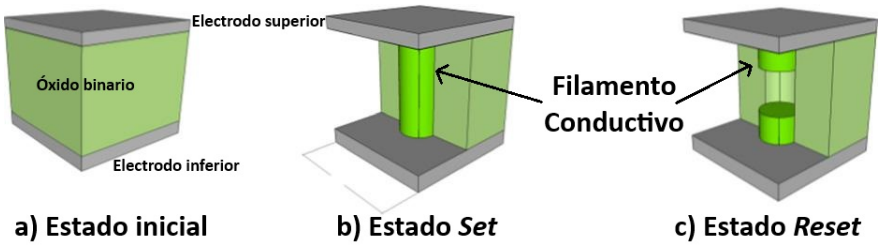



Figura 5: Estructura y funcionamiento de un dispositivo RRAM. (a) Apilamiento metal-aislante-metal (MIM) inicial. (b) Formación del filamento conductor tras la electroformación (estado LRS o *set*). (c) Ruptura del filamento tras el pulso eléctrico (estado HRS o *reset*). Figura adaptada y traducida de la referencia [20] (Copyright © 2014 John Wiley & Sons, Inc.).

la ley de Joule, la potencia total disipada en forma de calor por dicho segmento es:

$$P = I^2 R \tag{2}$$

Sustituyendo la resistencia, se deduce que el calentamiento se maximiza en las zonas más estrechas. Analizando la densidad de potencia térmica disipada $p$ (potencia por unidad de volumen, donde $V = lA$):

$$p = \frac{P}{V} = \frac{I^2 \left(\rho \frac{l}{A}\right)}{lA} = \rho \frac{I^2}{A^2} = \rho J^2 \tag{3}$$

donde $J$ es la densidad de corriente. Esta ecuación muestra que el calor generado por unidad de volumen escala de forma inversamente proporcional al cuadrado del área transversal ($p \propto 1/A^2$). En consecuencia, al forzar el paso de una alta corriente de *reset*, el calentamiento se concentra en la constricción más estrecha del filamento. Al superarse el umbral crítico de temperatura (que puede rondar los 900 K), los iones de oxígeno de la matriz circundante adquieren una alta movilidad térmica. Esto provoca una rápida reoxidación local de esa zona específica, disolviendo térmicamente el puente conductor metálico y devolviendo el dispositivo a su estado de alta resistencia (HRS)[7,20].

Durante el proceso de *set*, y dado que la brecha aislante creada durante el *reset* es de dimensiones nanométricas, la práctica totalidad de la caída de tensión se concentra en esta minúscula región, generando un campo eléctrico local inmenso[7]. Este intenso campo induce un nuevo evento de ruptura dieléctrica localizada que empuja las vacantes de oxígeno de vuelta hacia la brecha. Como resultado, el óxido se reduce localmente y las estructuras filamentarias vuelven a conectarse.

En conclusión, la dinámica de la NVRS unipolar queda definida por un proceso de *reset* impulsado por la corriente (mediante la difusión térmica por calentamiento de Joule) y un proceso de *set* impulsado por el voltaje (mediante ruptura dieléctrica local)[7]. Puesto que el calor disipado es proporcional al cuadrado de la corriente ($P \propto I^2$) y la ruptura dieléctrica depende exclusivamente de la magnitud del campo eléctrico y no de su dirección, ambos fenómenos físicos son inherentemente independientes del signo del estímulo eléctrico. Es esta simetría física subyacente la que demuestra y justifica el comportamiento no polar o unipolar del mecanismo termoquímico[7,16].

### 5.2.2. NVRS Bipolar

A diferencia del comportamiento unipolar, la NVRS bipolar exhibe una mayor complejidad fenomenológica y se manifiesta en diversas familias de materiales. Entre los más estudiados destacan los óxidos complejos como $Pr_{0,7}Ca_{0,3}MnO_3$ (PCMO), $YBa_2Cu_3O_{7-x}$ (YBCO) y $SrTiO_3$ (STO), así como las versiones no estequiométricas de óxidos binarios como $TiO_{2-x}$, $HfO_{2-x}$ y $Ta_2O_{5-x}$[7]. El principio físico ineludible en todos estos sistemas es el transporte direccional de especies iónicas impulsado por un campo eléctrico, exigiendo que dicho movimiento sea reversible al invertir la

polaridad para recuperar el estado resistivo original. La magnitud de este campo eléctrico local, generado por el voltaje aplicado entre los electrodos, se rige por la ley de Ohm generalizada:

$$\vec{E} = \rho(x)\vec{J} \tag{4}$$

Esta relación pone de manifiesto que el campo eléctrico $\vec{E}$ no es uniforme a lo largo del material, sino que es directamente proporcional a la resistividad local $\rho(x)$. Como consecuencia, las mayores caídas de tensión y, por ende, los campos eléctricos más intensos, se localizan de forma natural en las regiones de mayor resistividad. En los dispositivos basados en estos óxidos conductores, dichas regiones críticas corresponden a las *interfases de Schottky* conformadas en el contacto con los electrodos metálicos. Físicamente, el origen de estas barreras interfaciales altamente resistivas radica en el fenómeno de *curvatura de bandas* (*band bending*), un mecanismo necesario para equilibrar las fuerzas electrostáticas y los potenciales químicos en la unión metal-óxido[7].

Para explicar la modulación resistiva que ocurre en estas interfaces hay varias teorías, pero la atención se centrará en la *teoría del cambio de valencia* (VCM).

En el grupo de los óxidos complejos este mecanismo se puede ejemplificar con el PCMO mencionado anteriormente. El PCMO actúa como un semiconductor *tipo p* y se ha demostrado que bajo un campo eléctrico externo (voltaje) las vacantes de oxígeno migran a través de la interfaz metal-óxido[16]. La alteración en la concentración de oxígeno en la interfaz tiene un impacto inmediato: fuerza un cambio en el estado de valencia de los cationes del metal de transición[16]. En el caso del PCMO, la valencia del manganeso (Mn) es el parámetro fundamental para controlar las características del dispositivo. Mediante medidas experimentales se ha demostrado que la acumulación de vacantes de oxígeno genera una capa deficiente en oxígeno donde la valencia del Mn disminuye significativamente en las proximidades de la frontera con el electrodo[16]. Este cambio de valencia y la falta de oxígeno producen un ensanchamiento de la banda prohibida. Este aumento de la brecha de energía actúa como una barrera efectiva que interfiere en la conducción de los huecos, estableciendo el estado de alta resistencia (HRS)[16]. Por el contrario, cuando el campo eléctrico aleja las vacantes de la interfaz, el óxido recupera oxígeno y la banda prohibida se reduce. En estas condiciones, la región de vaciamiento se vuelve lo suficientemente delgada como para que los portadores de carga (huecos) puedan atravesar la interfaz por efecto túnel cuántico, provocando el estado de baja resistencia (LRS). Este grupo de materiales presenta conmutación resistiva no volátil bipolar en toda la superficie de contacto[16].

### 5.3. Memristor

Mientras que el modelo previamente expuesto detalla la física microscópica, el comportamiento macroscópico de la NVRS a nivel de circuito se describe mediante un componente fenomenológico: el *memristor*. Este concepto fue postulado teóricamente en 1971 por Leon O. Chua a partir de un argumento de simetría[21]. En la teoría de circuitos clásica interactúan cuatro magnitudes fundamentales: la corriente $I$, el voltaje $V$, la carga eléctrica $q = \int I\,dt$ y el flujo de enlace $\varphi = \int V\,dt$. De las seis posibles relaciones matemáticas binarias entre ellas, cinco ya estaban bien establecidas: dos provienen de las definiciones integrales mencionadas, y las tres restantes definen axiomáticamente a los componentes pasivos clásicos (el resistor, el capacitor y el inductor). De esta manera, faltaba una relación directa entre el flujo de enlace y la carga eléctrica. Chua introdujo un cuarto elemento, al que bautizó como *memristor* (contracción de *memory resistor*). Su ecuación se postula como:

$$d\varphi = M\,dq$$

donde el factor de proporcionalidad $M$ es la *memristancia*, con unidades de Ohmios ($\Omega$).

Aplicando la regla de la cadena:

$$\frac{d\varphi}{dt} = \frac{d\varphi}{dq} \cdot \frac{dq}{dt}$$

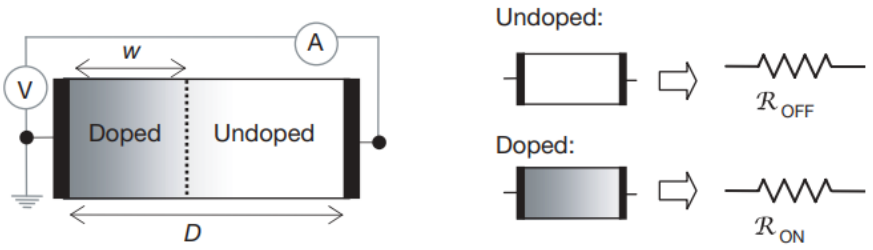



Figura 6: Diagrama simplificado con un circuito equivalente. Figura adaptada y traducida de la referencia [22].

y sabiendo que $v = d\varphi/dt$ e $i = dq/dt$, se obtiene la ecuación:

$$v(t) = M(q(t)) \cdot i(t)$$

Físicamente, esta ecuación revela que el memristor ideal obedece una versión generalizada de la ley de Ohm[22]. Se comporta como una resistencia ordinaria, con la particularidad vital de que su valor $M$ depende explícitamente de la cantidad total de carga $q$ que ha fluido a través del dispositivo en el pasado. En consecuencia, en el instante en que cesa la corriente ($i = 0$), la evolución temporal se detiene y la memristancia queda congelada en su último estado. Es precisamente esta dependencia del histórico de carga la que le confiere su capacidad de memoria y su naturaleza intrínsecamente no volátil[21,22].

A pesar del sólido desarrollo analítico de Chua, la teoría de los sistemas memristivos y el estudio físico de la conmutación resistiva en dieléctricos avanzaron por caminos paralelos durante décadas. La unificación de ambos campos se produjo en 2008, cuando Strukov *et al.*[22] publicaron en la revista *Nature* la demostración de que la migración asimétrica de vacantes de oxígeno a escala nanométrica en un óxido de metal de transición ($TiO_2$) obedece de forma natural a las ecuaciones de estado de Chua.

Como se puede observar en la Figura 6, el dispositivo se modela como una película delgada de óxido semiconductor de espesor $D$ delimitada por dos electrodos metálicos. La película se divide en dos regiones acopladas en serie: una región de ancho $w$ dopada con vacantes de oxígeno (un sub-óxido metálico $TiO_{2-x}$), y una región prístina de ancho $D - w$ puramente aislante. Se definen $R_{ON}$ y $R_{OFF}$ como las resistencias límite del dispositivo entero si estuviese completamente dopado ($w = D$) o completamente sin dopar ($w = 0$, donde $R_{OFF} \gg R_{ON}$)[22].

Asumiendo una aproximación óhmica para el transporte de electrones, la resistencia total del dispositivo en función de la variable geométrica $w(t)$ es la suma de las resistencias de ambas subregiones[22]:

$$R_{MEM}(t) = R_{ON}\frac{w(t)}{D} + R_{OFF}\left(1 - \frac{w(t)}{D}\right)$$

La dinámica estructural del dispositivo está gobernada por la deriva iónica bajo la acción del campo eléctrico local en la región dopada. Suponiendo un modelo de deriva lineal, la velocidad de deriva de las vacantes de oxígeno es proporcional a dicho campo ($v = \mu_v E$). Expresando el campo eléctrico en función de la corriente aplicada $i(t)$ y la resistividad del estado dopado, se obtiene la ecuación diferencial de estado del sistema[22]:

$$\frac{dw(t)}{dt} = \mu_v \frac{R_{ON}}{D} i(t)$$

Al integrar esta velocidad en el tiempo suponiendo una anchura inicial nula, se obtiene que $w(t)$ depende directamente de la carga total $q(t)$ que ha fluido por el circuito. Sustituyendo esta solución en la ecuación de la resistencia total, y aplicando la aproximación $R_{OFF} \gg R_{ON}$, Strukov dedujo

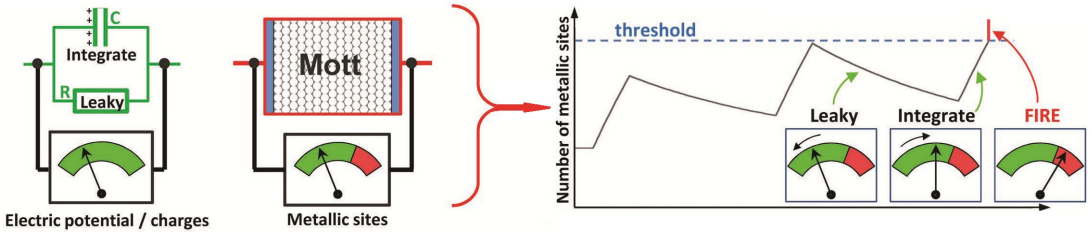



Figura 7: Las neuronas artificiales LIF basadas en el modelo de Lapicque reproducen la evolución del potencial de membrana gracias a un circuito de Resistencia-Capacitancia (RC) que acumula cargas eléctricas. La neurona artificial de Mott esquematizada en la izquierda reproduce el comportamiento LIF gracias a la acumulación de sitios metálicos correlacionados. Figura adaptada y traducida de la referencia [18].

la memristancia analítica del dispositivo[22]:

$$M(q) = R_{OFF}\left(1 - \frac{\mu_v R_{ON}}{D^2} q(t)\right)$$

Este resultado analítico resultó de vital importancia por dos motivos fundamentales. En primer lugar, demostró que la anchura de la región conductora $w$ operaba exactamente como la variable de estado interna propuesta por el modelo generalizado de Chua. En segundo lugar, el factor $1/D^2$ reveló que la magnitud del efecto memristivo escala inversamente con el cuadrado del espesor de la película. Esta dependencia geométrica explica físicamente por qué el memristor no pudo ser identificado empíricamente de forma consistente hasta que las técnicas de nanolitografía permitieron fabricar películas de óxido con espesores del orden de unos pocos nanómetros, donde los campos eléctricos son lo suficientemente intensos para inducir una deriva iónica mensurable a temperatura ambiente[22].

# 6. Implementaciones

## 6.1. Neuristor

Uno de los mayores retos en el diseño de hardware neuromórfico es la excesiva cantidad de componentes convencionales, como transistores, condensadores y resistencias, necesarios para emular la dinámica de una única neurona. Frente a esta limitación arquitectónica, es posible condensar todas las funciones del modelo de *integración y disparo con pérdidas* (LIF) utilizando un único nanodispositivo de dos terminales. Para ilustrar el funcionamiento físico de un neuristor real, en esta sección se analizarán los resultados experimentales obtenidos por Stoliar *et al.*[18] basados en un aislante de Mott de brecha estrecha ($GaTa_4Se_8$). Este compuesto presenta VRS, más concretamente, presenta transiciones metal aislante de ruptura por avalancha electrónica impulsadas eléctricamente. De manera esquemática en la Figura 7 se puede observar la fenomenología de este dispositivo.

La dinámica del nanodispositivo está determinada por dos escalas temporales intrínsecas del material. En primer lugar, si se aplica un pulso de voltaje continuo que supere el umbral crítico, la transición al estado conductor no es instantánea: existe un tiempo de retardo dependiente del voltaje, denominado $t_{FIRE}$, que el sistema requiere para mostrar la conmutación de la resistencia. En segundo lugar, dada la naturaleza volátil de esta transición, al interrumpirse el suministro eléctrico, el material recobra espontáneamente su estado aislante de alta resistencia. Este proceso sigue una dinámica exponencial gobernada por el tiempo de relajación natural, $\tau$. Para el nanodispositivo de $GaTa_4Se_8$, la resistencia sigue entonces la ecuación[18]:

$$R(t) = R_0\left(1 - e^{-t/\tau}\right) \tag{5}$$

donde $R_0$ es la resistencia de la muestra en su estado aislante antes de la conmutación. Es precisamente la coexistencia del tiempo de retardo para la activación ($t_{FIRE}$) y del tiempo de relajación ($\tau$) lo que habilita la emulación del modelo de integración y disparo con pérdidas.

Para evaluar el funcionamiento de este neuristor y emular una neurona, se somete al material a un tren de pulsos de voltaje caracterizado por una duración de estímulo $t_{ON}$ y un intervalo de reposo $t_{OFF}$. Los estímulos deben diseñarse de modo que $t_{ON} < t_{FIRE}$ y $t_{OFF} < \tau$. Por una parte, al aplicar pulsos de voltaje cuya duración $t_{ON}$ sea estrictamente menor que el tiempo de retardo $t_{FIRE}$, se garantiza que un único pulso no es capaz por sí solo de producir el cambio resistivo. En su lugar, el efecto de múltiples pulsos cortos se vuelve acumulativo. Por otro lado, durante el intervalo de reposo entre los estímulos $t_{OFF}$, la resistencia comienza su lenta recuperación exponencial hacia el estado inicial de alta resistencia dictada por la constante $\tau$. Es precisamente esta recuperación de la resistencia a lo largo del tiempo lo que equivale a la "fuga"del modelo neuronal, ya que el material va perdiendo progresivamente el efecto de los pulsos anteriores si el tiempo de reposo es demasiado prolongado.

Los resultados experimentales pueden apreciarse en la Figura 8: al cambiar los parámetros $t_{ON}$ y $t_{OFF}$ que definen el tren de pulsos eléctricos recibidos cambia el número de pulsos necesarios para producir el disparo ($N_{FIRE}$).

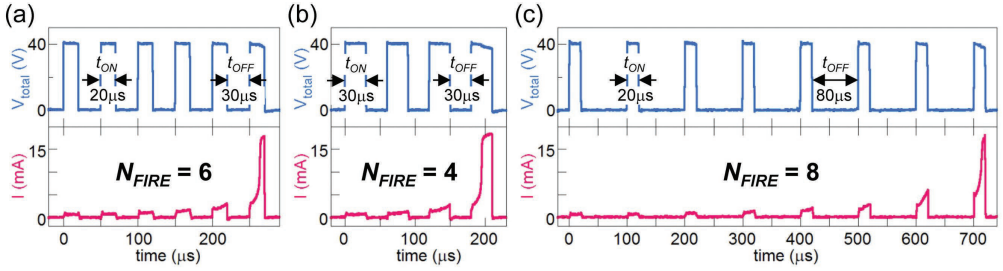



Figura 8: Conmutación resistiva experimental obtenida mediante la aplicación de trenes de pulsos cortos con diversos $t_{\mathrm{ON}}$ y $t_{\mathrm{OFF}}$. a) El uso de $t_{\mathrm{ON}} = 20$ $\mu$s y $t_{\mathrm{OFF}} = 30$ $\mu$s da lugar a $N_{\mathrm{FIRE}} = 6$. b) Incrementar $t_{\mathrm{ON}}$ a 30 $\mu$s conduce a una disminución de $N_{\mathrm{FIRE}}$ a 4. c) Incrementar $t_{\mathrm{OFF}}$ a 80 $\mu$s conduce a un aumento de $N_{\mathrm{FIRE}}$ a 8. Figura adaptada y traducida de la referencia [18].

## 6.2. Sinaptor

En paralelo al desarrollo de neuristores para emular la dinámica celular, la implementación del hardware neuromórfico requiere replicar la plasticidad y la memoria de las conexiones neuronales. Para este propósito, la conmutación resistiva no volátil (NVRS) en óxidos de metales de transición se ha posicionado como la tecnología candidata ideal para el diseño de *sinaptores* (sinapsis electrónicas). Para ilustrar el funcionamiento físico y la implementación de un sinaptor real, en esta sección se analizarán los resultados obtenidos por Yu *et al.*[15].

El neuristor usado por Yu *et al.* está basado en una estructura de matriz de óxido formada por electrodos de nitruro de titanio y platino (TiN y Pt, respectivamente), entre los que se intercala una bicapa dieléctrica de $HfO_x/AlO_x$. Como se puede observar en la comparación esquemática entre este dispositivo y una sinapsis biológica (Figura 9), dicho elemento presenta conmutación resistiva no volátil bipolar gobernada por el mecanismo de cambio de valencia (VCM) filamentario[15].

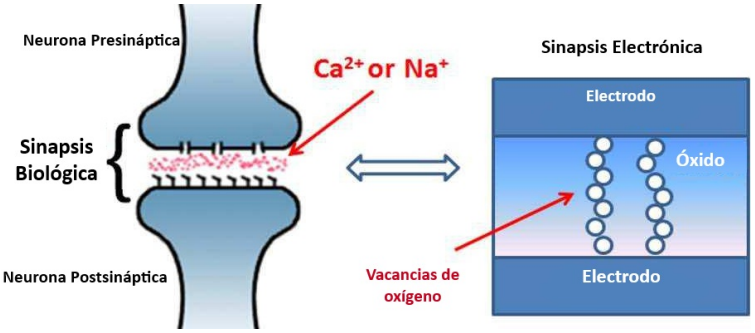



Figura 9: Analogía estructural entre la sinapsis biológica y la sinapsis electrónica. La contraparte electrónica emplea una estructura de película delgada metal-aislante-metal, donde el comportamiento sináptico emerge de la dinámica de las vacantes de oxígeno en el óxido. Figura adaptada y traducida de la referencia [15].

Físicamente, el sinaptor altera su resistencia a través de la generación y migración de vacantes de oxígeno impulsadas por los pulsos eléctricos incidentes. En este marco, los procesos de aumento de conductancia (reducción de la resistencia o *set*) y disminución de conductancia (aumento de la resistencia o *reset*) de la ruta filamentaria equivalen, respectivamente, a la potenciación y depresión sináptica biológica[15].

Sin embargo, para que una red neuronal artificial sea capaz de aprender, una memoria estrictamente binaria (que salte bruscamente del estado LRS al HRS) es insuficiente. El logro fundamental de estos dispositivos es su capacidad para retener múltiples niveles de conductancia intermedios y continuos. El control del peso sináptico se logra modulando la amplitud de pulsos de voltaje ultracortos (del orden de 50 ns)[15]. Al aplicar una secuencia de pulsos con amplitudes progresivamente crecientes o decrecientes, se ajusta con precisión nanométrica la densidad de defectos en el filamento conductor. Esta modulación de amplitud reduce significativamente el estrés eléctrico sobre la matriz del óxido, lo que permite estabilidades operativas superiores a los $10^5$ ciclos de programación[15].

A nivel de arquitectura de red, esta capacidad de modulación gradual por amplitud es la clave para implementar de forma nativa la *plasticidad dependiente del tiempo entre impulsos* (STDP). En la biología, la magnitud del cambio en el peso sináptico depende estrictamente de la diferencia temporal ($\Delta t = t_{\text{post}} - t_{\text{pre}}$) entre el disparo de la neurona presináptica y la postsináptica. Como ilustra la curva característica de la Figura 10, el aprendizaje asociativo se rige por la causalidad. Si el impulso presináptico antecede al postsináptico ($\Delta t > 0$) la sinapsis se refuerza incrementando su conductancia (potenciación a largo plazo o LTP, zona roja). A la inversa, si el orden se invierte ($\Delta t < 0$), la conexión se debilita, provocando una caída en la conductancia (depresión a largo plazo o LTD, zona azul). Además, la fenomenología biológica plasmada en la gráfica evidencia que esta actualización de pesos no es lineal: la magnitud de la modificación ($\Delta G$) alcanza su valor máximo cuando los impulsos son casi simultáneos y decae de forma exponencial a medida que aumenta el desfase temporal. Este comportamiento asintótico define una ventana crítica de unos pocos milisegundos fuera de la cual el aprendizaje deja de ser efectivo[15].

Para lograr replicar este efecto biológico, se recurre a un esquema de codificación temporal mediante el diseño estructurado de las formas de onda. En lugar de emitir un único potencial de acción continuo, las unidades neuronales generan señales discretizadas, dividiendo la duración total del impulso en ranuras de tiempo constantes. A cada una de estas ventanas temporales se le asigna un pulso con una amplitud de voltaje distinta, dando lugar a un tren de pulsos escalonado (como se muestra en la Figura 11a).

Dado que el sinaptor se ubica en las intersecciones del circuito, el campo eléctrico neto que experimenta el dieléctrico depende de la diferencia de potencial instantánea entre el impulso presináptico y el postsináptico ($V_{\text{sinapsis}} = V_{\text{pre}} - V_{\text{post}}$). Únicamente cuando ambos trenes de pulsos

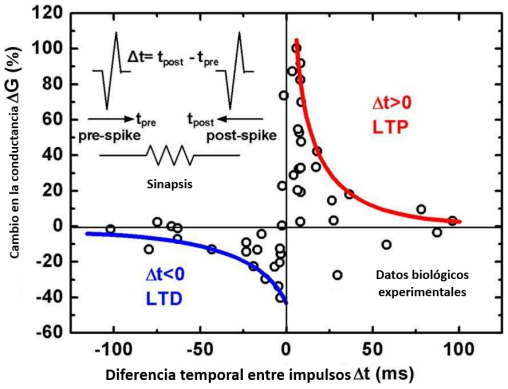



Figura 10: Curva biológica de la plasticidad dependiente del tiempo entre impulsos (STDP). Si el potencial de acción presináptico antecede al postsináptico ($\Delta t > 0$), la conductancia aumenta y se produce una potenciación a largo plazo (LTP). En el escenario inverso ($\Delta t < 0$), la conductancia disminuye, lo que resulta en una depresión a largo plazo (LTD). Cuanto menor es la diferencia temporal, mayor es la magnitud relativa de la variación resistiva. Figura adaptada y traducida de la referencia [15].

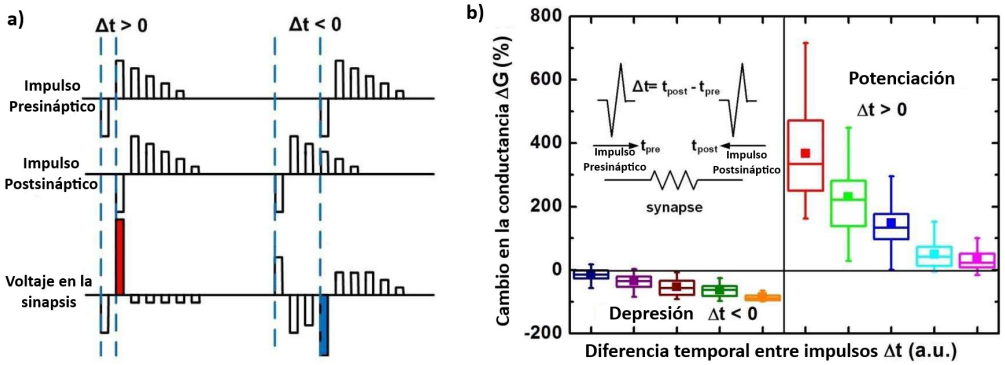



Figura 11: a) Esquemas de implementación de STDP desarrollados mediante TDM y modulación de amplitud de pulso. b) Curva de tipo STDP calculada empleando los esquemas de señal de a). Figura traducida y adaptada de la referencia [15].

se solapan en el tiempo, la superposición de sus voltajes logra superar el umbral crítico de programación del dispositivo. Cuanto menor es el retraso temporal $\Delta t$, mayor es la amplitud del pulso resultante superpuesto, lo que induce un salto de conductancia drástico. A medida que los impulsos se separan en el tiempo, la amplitud efectiva de la superposición decae, replicando la curva exponencial de aprendizaje STDP biológica[15], como se puede observar en la Figura 11b.

Finalmente, el factor decisivo para la viabilidad de estos sistemas es el consumo energético. Mientras que una sinapsis biológica humana consume alrededor de 1 a 10 fJ por evento, la alta velocidad de conmutación de los dispositivos de $HfO_x$ (con tiempos de transición de aproximadamente 10 ns) ha permitido reducir el consumo nominal de energía al rango de los sub-picojulios (sub-pJ) por operación[15]. Esta extraordinaria eficiencia energética confirma que los sinaptores basados en memoria resistiva son un pilar tecnológico viable para superar el cuello de botella de la arquitectura de von Neumann y posibilitar sistemas cognitivos a gran escala.

# 7. Conclusiones

La revisión y el análisis de los fundamentos físicos subyacentes a la computación neuromórfica evidencian que la electrónica de óxidos ofrece una alternativa prometedora frente a los límites estructurales y energéticos de la arquitectura clásica de Turing-von Neumann[7,10]. Si bien la superación del cuello de botella de la memoria y la ineficiencia térmica exigen una transición hacia sistemas paralelos donde el procesamiento lógico y el almacenamiento de datos coexistan espacialmente[10], la implementación física a gran escala aún debe resolver desafíos críticos. Entre ellos destacan la variabilidad estocástica en la formación de los filamentos conductores ciclo a ciclo y la corta vida útil de los componentes ($10^5$ ciclos en el caso de los sinaptores)[15,20].

Este paradigma biológico se logra reproducir a nivel de hardware mediante el uso de nanodispositivos basados en óxidos de metales de transición. A través del estudio de la conmutación resistiva volátil (VRS), se ha establecido que las transiciones de fase en materiales fuertemente correlacionados, como los aislantes de Mott (ej. $GaTa_4Se_8$), son capaces de emular la dinámica de la membrana celular de forma directa[18]. La física subyacente a la ruptura de la brecha de Mott por avalancha electrónica proporciona el mecanismo físico de umbral necesario para generar el potencial de acción[18,19]. Simultáneamente, la relajación espontánea de los filamentos conductores hacia el estado basal aislante reproduce la disipación temporal y el periodo refractario dictados por el modelo de integración y disparo con pérdidas (LIF)[7,18]. El neuristor actúa así como un elemento activo capaz de condensar toda la complejidad temporal de una neurona en un solo dispositivo de dos terminales[18].

En lo que respecta a la interconexión y la memoria de la red, la conmutación resistiva no volátil (NVRS) sustenta el funcionamiento de las sinapsis electrónicas[7,15]. Se ha detallado cómo el mecanismo de cambio de valencia (VCM) y la deriva direccional de vacantes de oxígeno bajo un campo eléctrico, presentes en estructuras como las bicapas de $HfO_x/AlO_x$, permiten la modulación analógica de la conductancia del material[15]. Esta capacidad de retener múltiples estados resistivos persistentes se traduce en un control ultrapreciso del peso sináptico (LTP y LTD)[14,15]. Además, la respuesta de estos sinaptores ante la superposición temporal de voltajes dependientes de la forma de onda logra recrear físicamente la plasticidad dependiente del tiempo entre impulsos (STDP)[15]. En consecuencia, la causalidad y correlación temporal de las espigas biológicas quedan codificadas como alteraciones controladas y semipermanentes en la red cristalina del óxido[15,7].

Como conclusión, la convergencia entre la física de la materia condensada y la arquitectura de computación subraya que el futuro de la inteligencia artificial neuromórfica está estrictamente condicionado por la optimización de los fenómenos de transporte en la red cristalina[20]. Dominar la termodinámica de la migración iónica y la cinética de las transiciones de fase electrónicas en estos óxidos de metales de transición es el requisito ineludible para suprimir la variabilidad estocástica intrínseca de los materiales[7]. Solo un control físico riguroso sobre estas variables de estado permitirá consolidar redes analógicas a gran escala con capacidades de retención y aprendizaje estables[7,20].

# Bibliografía


[1] Gordon Moore. «Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff». En: *Solid-State Circuits Newsletter, IEEE* 11 (oct. de 2006), págs. 33-35. DOI: 10.1109/N-SSC.2006.4785860.

[2] Igor Markov. «Limits on fundamental limits to computation». En: *Nature* 512 (ago. de 2014), págs. 147-54. DOI: 10.1038/nature13570.

[3] Mark Bohr. «A 30 Year Retrospective on Dennard's MOSFET Scaling Paper». En: *IEEE Solid-State Circuits Society Newsletter* 12.1 (2007), págs. 11-13. DOI: 10.1109/N-SSC.2007.4785534.

[4] D.J. Frank et al. «Device scaling limits of Si MOSFETs and their application dependencies». En: *Proceedings of the IEEE* 89.3 (2001), págs. 259-288. DOI: 10.1109/5.915374.

[5] Mark Horowitz. «1.1 Computing's energy problem (and what we can do about it)». En: *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*. 2014, págs. 10-14. DOI: 10.1109/ISSCC.2014.6757323.

[6] Wm. A. Wulf y Sally A. McKee. «Hitting the Memory Wall: Implications of the Obvious». En: *Computer Architecture News* 23 (ene. de 1996). DOI: 10.1145/216585.216588.

[7] Javier del Valle et al. «Challenges in materials and devices for resistive-switching-based neuromorphic computing». En: *Journal of Applied Physics* 124.21 (dic. de 2018), pág. 211101. ISSN: 0021-8979. DOI: 10.1063/1.5047800.

[8] C. Mead. «Neuromorphic electronic systems». En: *Proceedings of the IEEE* 78.10 (1990), págs. 1629-1636. DOI: 10.1109/5.58356.

[9] Carver Mead. *Analog VLSI and Neural Systems*. Addison-Wesley, 1989. ISBN: 978-0201059922.

[10] Kaushik Roy, Akhilesh Jaiswal y Priyadarshini Panda. «Towards spike-based machine intelligence with neuromorphic computing». En: *Nature* 575.7784 (2019), págs. 607-617. DOI: 10.1038/s41586-019-1677-2.

[11] Nicolas Brunel y Mark C. W. van Rossum. «Lapicque's 1907 paper: from frogs to integrate-and-fire». En: *Biological Cybernetics* 97.5-6 (2007), págs. 337-339. DOI: 10.1007/s00422-007-0190-0.

[12] A. N. Burkitt. «A Review of the Integrate-and-fire Neuron Model: I. Homogeneous Synaptic Input». En: *Biological Cybernetics* 95.1 (2006), págs. 1-19. ISSN: 1432-0770. DOI: 10.1007/s00422-006-0068-6.

[13] Guo-qiang Bi y Mu-ming Poo. «Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type». En: *Journal of Neuroscience* 18.24 (1998), págs. 10464-10472. DOI: 10.1523/JNEUROSCI.18-24-10464.1998.

[14] Patrick D. Roberts y Curtis C. Bell. «Spike timing dependent synaptic plasticity in biological systems». En: *Biological Cybernetics* 87.5 (2002), págs. 392-403. ISSN: 1432-0770. DOI: 10.1007/s00422-002-0361-y.

[15] Shimeng Yu et al. «An Electronic Synapse Device Based on Metal Oxide Resistive Switching Memory for Neuromorphic Computation». En: *IEEE Transactions on Electron Devices* 58.8 (2011), págs. 2729-2737. DOI: 10.1109/TED.2011.2147791.

[16] Isao H. Inoue y Akihito Sawa. «Resistive Switchings in Transition-Metal Oxides». En: *Functional Metal Oxides*. John Wiley & Sons, Ltd, 2013. Cap. 16, págs. 443-463. ISBN: 9783527654864. DOI: https://doi.org/10.1002/9783527654864.ch16.

[17] H.-S. Philip Wong et al. «Metal–Oxide RRAM». En: *Proceedings of the IEEE* 100.6 (2012), págs. 1951-1970. DOI: 10.1109/JPROC.2012.2190369.

[18] Pablo Stoliar et al. «A Leaky-Integrate-and-Fire Neuron Analog Realized with a Mott Insulator». En: *Advanced Functional Materials* 27.11 (2017), pág. 1604740. DOI: https://doi.org/10.1002/adfm.201604740.

[19] Etienne Janod et al. «Resistive Switching in Mott Insulators and Correlated Systems». En: *Advanced Functional Materials* 25.40 (2015), págs. 6287-6305. DOI: https://doi.org/10.1002/adfm.201500823.

[20] Daniele Ielmini. «Resistive switching memories based on metal oxides: mechanisms, reliability and scaling». En: *Semiconductor Science and Technology* 31.6 (mayo de 2016), pág. 063002. DOI: 10.1088/0268-1242/31/6/063002.

[21] L. Chua. «Memristor-The missing circuit element». En: *IEEE Transactions on Circuit Theory* 18.5 (1971), págs. 507-519. DOI: 10.1109/TCT.1971.1083337.

[22] Dmitri B. Strukov et al. «The missing memristor found». En: *Nature* 453 (2008), págs. 80-83. DOI: 10.1038/nature06932.