UNIVERSIDAD DE VALLADOLID

ESCUELA DE INGENIERÍAS INDUSTRIALES

Grado en Ingeniería electrónica industrial y automática

Memristor: funcionamiento y aplicaciones

Autor:

Jorge Grande de Miguel

Tutora:

Dra. Carmen Quintano Pastor

Departamento de tecnología electrónica

Índice de contenidos

Resumen 3
1 Introducción y objetivos 4
2 Fundamentos del memristor 6
2.1 La labor de Leon Chua: predicción, definición y análisis del cuarto elemento 6
2.2 Propiedades del dióxido de titanio no estequiométrico TiO(2-x) 13
2.2.1 Generalidades y propiedades mecánicas 13
2.2.2 Propiedades termoeléctricas 15
3 Modelos matemáticos y eléctricos 22
3.1 Bases de los modelos 22
3.1.2 Ecuaciones de las representaciones globales 22
3.1.2 Características del memristor 23
3.2 Modelos de simulación del memristor 25
3.2.1 Simulación en Simulink 25
3.2.2 Simulación en LabView 34
4 El memristor de HP 37
5 Aplicaciones del memristor 42
5.1 Ciberseguridad y sistemas caóticos 42
5.1.1 Números aleatorios en la seguridad 42
5.1.2 Generador basado en memristor de difusión 43
5.1.3 Encriptación mediante circuitos electrónicos memristivos 47
5.2 Análisis y construcción de redes neuronales 54
5.3 Memorias no volátiles 57
5.3.1 Contexto y conceptos generales 57
5.3.2 Funcionamiento básico. El problema de los caminos furtivos 58
5.3.3 Arquitecturas propuestas con solución a los caminos furtivos 61
6 Problemas inherentes al memristor de HP 68
7 Alternativas funcionales: el memristor “self-directed channel “ 71
Conclusiones 72
Bibliografía 73

Resumen

El memristor como nueva tecnología electrónica es ya una realidad, aunque presumiblemente presenta problemas de vida útil que deben ser superados para que ésta tecnología no se quede en el cajón de las curiosidades. Además, el mayor impacto que tendría este dispositivo, su aplicación en redes neuronales analógicas, es una rama de la computación y la electrónica tan nueva y compleja que sin duda llevará tiempo asegurar si el memristor será importante y realmente aprovechable. En un sentido más puramente científico, la historia de la predicción del memristor es muy interesante: en lugar de partir de los datos y la observación de lo que la realidad nos presenta para hacer hipótesis o aprovechar dichos fenómenos y su estudio en la construcción de aparatos y máquinas, el profesor Chua discurrió extrayendo conclusiones de lo que él veía como una falta de simetría o de armonía en la tecnología existente, pero no de la observación empírica de ningún fenómeno real. Todos estos puntos, además de otras posibles aplicaciones prácticas y las matemáticas fundamentales tanto de los modelos teóricos como del funcionamiento físico observado en el memristor son presentados en este trabajo.

1 Introducción y objetivos

Desde que a mediados del siglo XX aparecieron los primeros transistores y con ellos nació la electrónica digital, el camino más importante que ha conducido a la mejora de las tecnologías basadas en la electrónica ha sido la ley de Moore, según la cual cada dos años se conseguiría duplicar el número de transistores por unidad de superficie que se podrían implementar en procesadores y dispositivos digitales de todo tipo. Este aumento en la densidad de transistores trae, por un lado, un aumento en la potencia de funcionamiento de los dispositivos, y por otro una capacidad de disminuir el tamaño de los mismos, lo cual es una gran comodidad para el consumidor y una ventaja a la hora de transportar y almacenar los productos. Pero, como es evidente, esta ley no podrá cumplirse siempre. La frontera de lo atómico representa el límite que esa densidad de transistores podrá alcanzar, ya sea en 5, 10 o 20 años. ¿Será ese el final de la mejora del hardware de los dispositivos electrónicos digitales? Hasta hace unos años la respuesta más probable habría sido que sí. Sin embargo, un nuevo elemento ha abierto una puerta de posibles mejoras que nada tienen que ver con encoger transistores: se trata del memristor, el dispositivo fundamental que el profesor Leon Chua predijo y describió hace casi 50 años, y que nadie fue capaz de construir (incluso de pensar que fuera posible su construcción) hasta hace relativamente poco.

En 1995, el prestigioso científico especializado en nanotecnología Richard Stanley Williams empezó a trabajar en los laboratorios de HP para, junto con un gran equipo de profesionales reclutados por David Packard y colaborando con otros eruditos de la época como Phil Kuekes, intentar avanzar en el desarrollo del conocido *Teramac*, o *tera-operation-per-second multiarchitecture computer.*

Cuando construyeron y trabajaron con un modelo simplificado del *Teramac* de tipo *crossbar,* observaron que el funcionamiento que deseaban (capacidad de almacenar información de forma digital en las uniones aprovechando sus diferentes resistencias entre los estados on y off) se cumplía a veces. Otras no. Y cuando se cumplía, no sabían muy bien por qué. No podían modelarlo, ni predecirlo. En palabras del propio Williams "el éxito o fracaso del dispositivo nunca parecía depender del mismo factor" (Williams, 2014).

Fue entonces, ya en el año 2002, cuando un compañero de Williams le mostró un extraño paper de un profesor de ingeniería que se había publicado hacía ya 30 años y que apenas había tenido importancia, ni demasiadas citas. Pero ciertas gráficas se parecían sospechosamente a algunas de las que presentaban los dispositivos que estaban construyendo. Finalmente, en 2004, tras estudiar a fondo los conceptos del profesor Chua y realizar análisis y pequeñas "cirugías" a los electrodos de titanio y a las moléculas que los conectaban, llegaron por fin a la conclusión demostrable de que lo que habían construido era un memristor de dióxido de titanio.

Actualmente el memristor es ya una realidad, se han fabricado distintos tipos y su modelado eléctrico y matemático está muy desarrollado aunque aún no es un producto comercializado.

En este contexto, el objetivo de este trabajo es introducir esta nueva tecnología de forma ilustrativa, resumida y práctica, analizar las principales aplicaciones reales para las que se podrá emplear o ya se emplea este nuevo dispositivo, considerando también los obstáculos y dificultades que aún se deben salvar para hacerlo de manera óptima y rentable. Además explicaremos con un cierto nivel de profundidad los modelos matemáticos y justificaciones físicas de la memristancia en lo que probablemente sea un buen ejemplo de los contenidos y estructura que tendrán los puntos sobre el memristor en asignaturas de electrónica analógica en los próximos años.

2 Fundamentos del memristor

El camino hasta el descubrimiento de la realizabilidad física del memristor fue largo y complicado de describir, ya que se trató en parte de un evento fortuito.

A continuación, exponemos un resumen de los principales fundamentos teóricos y prácticos que han permitido construir este dispositivo tan prometedor: por un lado, todo el trabajo que su inventor llevó a cabo para sustentar su teoría, y predecir comportamientos que pudieran ayudar a distinguir el memristor del resto de dispositivos fundamentales. Por otro, las singulares propiedades del material clave en el artefacto de Williams y su equipo de HP, el dióxido de titanio TiO2.

2.1 La labor de Leon Chua: predicción, definición y análisis del cuarto elemento

En 1971 el profesor de ingeniería eléctrica de la Universidad de California, Leon Chua, publicó un paper en el que por primera vez se mencionaba y describía un cuarto elemento pasivo eléctrico, que él mismo bautizó como memristor, y su propia variable intrínseca, la memristancia. El profesor Chua se percató de una "relación perdida" en la teoría de circuitos, la cual le llevó a predecir la existencia de un cuarto elemento pasivo fundamental.

Repasemos pues los tres elementos pasivos bi-terminales fundamentales ya bien conocidos:

-Resistor: elemento con un valor numérico real en su magnitud intrínseca, la resistencia (R), que supone una relación de transferencia directa entre la tensión entre sus terminales y la corriente que lo atraviesa.

-Inductor: elemento con un valor numérico real en su magnitud intrínseca, la inductancia (L), que supone una relación de transferencia directa entre el campo magnético que genera y la corriente que lo atraviesa.

-Capacitor: elemento con un valor numérico real en su magnitud intrínseca, la capacidad (C), que supone una relación de transferencia directa entre la carga que almacena el elemento y la tensión entre sus terminales.

Estos tres dispositivos se denominan fundamentales porque no se pueden construir mediante combinaciones de los otros, sino que son independientes, y son pasivos porque para funcionar no necesitan ninguna fuente de alimentación externa.

Como vemos, estos dispositivos relacionan y controlan entre sí las cuatro principales magnitudes físicas que definen y determinan el comportamiento de cualquier circuito eléctrico y/o electrónico, a saber:

-Tensión o diferencia de potencial, medida en voltios en el sistema internacional (V).

-Carga eléctrica, medida en culombios en el sistema internacional (Q).

-Corriente eléctrica, medida en amperios en el sistema internacional (I).

-Flujo magnético, medido en webers en el sistema internacional (ϕ).

Podríamos hablar de otras magnitudes físicas que influyen en los dispositivos eléctricos y electrónicos, como es la temperatura, pero esta influencia se limita al estado de los materiales de los circuitos, no tanto a las magnitudes electromagnéticas, por lo que no tiene sentido considerarla en este momento.

Bien, estas cuatro magnitudes se relacionan entre ellas en algunos casos por definición física, y en otros casos mediante el funcionamiento de los tres elementos pasivos.

La intensidad es por definición la derivada de la carga respecto del tiempo, es decir la cantidad de carga que se mueve por el circuito; el flujo magnético ϕ se calcula como la integral de la tensión inducida sobre el tiempo, es decir, la tensión equivale a la derivada del flujo respecto del tiempo. Esto, junto con las definiciones de las magnitudes intrínsecas de los tres elementos pasivos fundamentales nos proporcionan las siguientes relaciones entre las magnitudes físicas del circuito, las cuales están experimentalmente confirmadas:

$$dq = I * dt \quad [1]$$

$$d\phi = -\varepsilon * dt \quad [2]$$

$$dV = R * dI \quad [3]$$

$$dq = C * dV \quad [4]$$

$$d\phi = L * dI \quad [5]$$

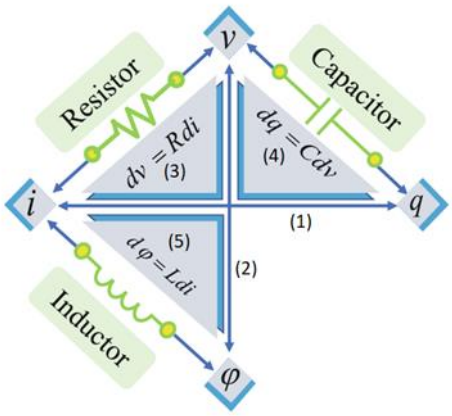


Ilustración 1: relaciones entre dispositivos fundamentales. Rahma, F. y Munean, S. 2018. Pág. 3

Como podemos ver en la ilustración 1, hay un claro desequilibrio en estas relaciones obtenidas con los tres elementos pasivos fundamentales, que no puede ser completado con combinaciones de éstos, ni con ecuaciones físicas: la relación entre carga y flujo magnético.

Las ecuaciones 3, 4 y 5 son relaciones llamadas constitutivas, ya que partiendo por ejemplo de una inductancia L en la que conocemos el flujo magnético y la corriente, podemos obtener la tensión en sus extremos y la carga que la ha atravesado hasta el instante actual (suponiendo conocidas las condiciones iniciales necesarias para realizar la integración correspondiente) mediante 1 y 2. Debe existir por tanto una última ecuación constitutiva que relacione la carga con el flujo magnético mediante una magnitud intrínseca a un cuarto elemento fundamental pasivo, y que una vez resuelta permite conocer la tensión y la corriente en dicho elemento mediante las dos ecuaciones físicas no constitutivas, 1 y 2. Esta magnitud se denomina memristancia (M), y este cuarto elemento fundamental pasivo, memristor. (Chua, 1971)

$$d\phi = M * dq \quad [6]$$

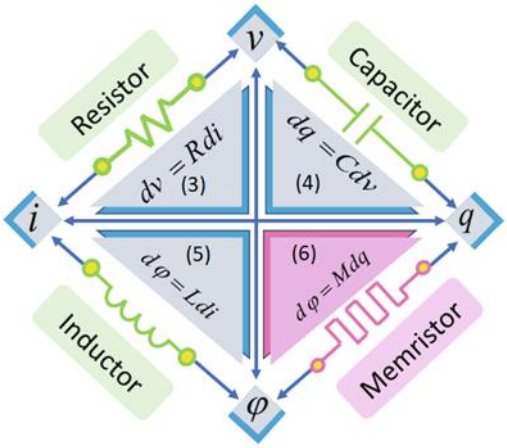


Ilustración 2: inclusión del memristor. Rahma, F. y Munean, S. 2018. Pág 4

Sustituyendo en 6 las ecuaciones 1 y 2 obtenemos la expresión del memristor ideal (Chua, 2019c)

$$V = R(q) * I \ ; \ R(q) \geq 0 \quad [7]$$

Donde R(q) es la memristancia del dispositivo, cuyas unidades en sistema internacional serán los ohmios, y que es una función real de la carga y estrictamente no-negativa.

Por tanto, el memristor puede asumir un rango de valores estables de resistencia, la cual varía con la carga, y que se puede calcular conociendo las condiciones iniciales adecuadas para integrar. El nombre de "resistor de memoria" viene de este hecho: el valor instantáneo de resistencia del dispositivo depende de su historial de corriente (tensión) y su estado inicial. Se deduce también que la relación entre tensión y corriente pasa a ser no-lineal ya que la variable de la que depende la resistencia es función a su vez de la integral de la intensidad sobre el tiempo.

Una expresión más general en la que asumimos que la memristancia puede ser una función de otras variables de estado que no dependan de tensión ni corriente es la siguiente, la cual define el memristor extendido (Chua, 2019b)

$$V = R(x, I) * I \ ; \ R(x, I) \geq 0 \ ; \ \frac{dx}{dt} = f(x, I) \ \ [8]$$

El profesor Chua describe, tras estas definiciones matemáticas, la "huella" que debe dejar un memristor, (ilustración 3) es decir, las condiciones prácticas que un dispositivo debe cumplir para ser considerado un memristor, las cuales son principalmente dos:

1-Para cualquier entrada de tensión o corriente periódica y alterna (bipolar), la curva de Lissajous obtenida al representar [V(t),I(t)] será una curva de histéresis estrechada o *pinched hysteresis loop*, la cual pasa por cero en el origen (V=0 cuando I=0) y se localiza solo en el primer y tercer cuadrante (R(x,I)>0).

2-La curva de histéresis se comprime cuando aumenta la frecuencia, convirtiéndose en una recta que pasa por cero en el origen cuando la frecuencia tiende a infinito, límite en el que el memristor pasa a comportarse como una resistencia pura.

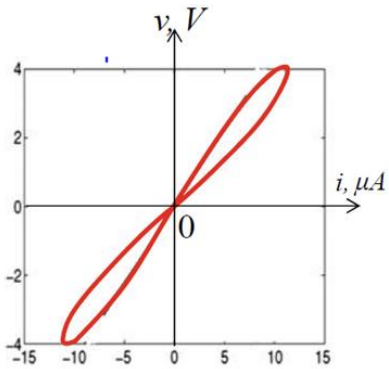


Ilustración 3: "Pinched hysteresis loop". Chua, L. 2019. Pág. 28

Veamos ahora el modelo diseñado por el profesor Chua para emular un memristor, que llamó "mutador resistor-memristor", basado en transistores y operacionales. Su símbolo y su esquema son los siguientes:

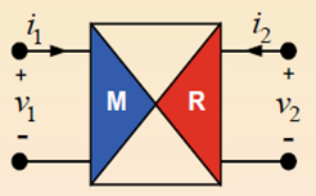


Ilustración 4: símbolo del mutador. Chua, L. 2018. Pág. 22

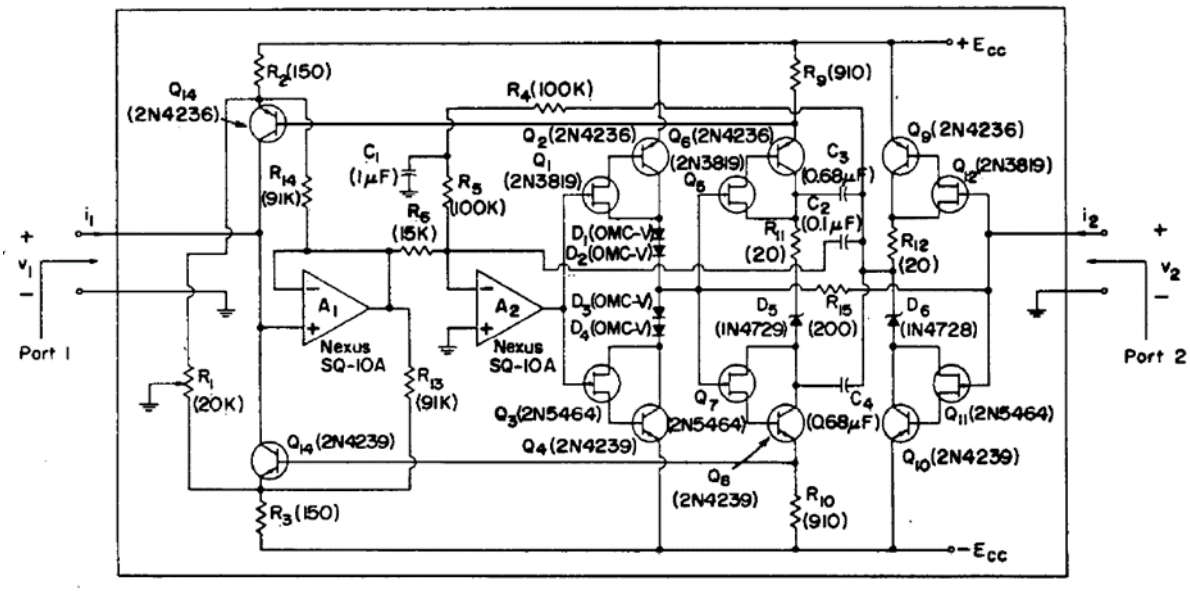


Ilustración 5: esquema eléctrico del mutador. Chua, L. 1971. Pág. 509

Este circuito permite "convertir" una resistencia no lineal conectada a los terminales del puerto 2 en un memristor, en el sentido de que al introducir una señal senoidal V1 se puede medir una curva I2 = f(V2) en el resistor no lineal que coincide virtualmente con la curva q1 = f(ϕ1) medida en la entrada. Es decir, que mirando desde el puerto 1 e introduciendo todos elementos tras este puerto en una caja negra, podríamos deducir la existencia de un memristor. Obviamente, para su correcto funcionamiento tendríamos que asegurar que los rangos dinámicos y la frecuencia de todas las señales de tensión y corriente en el interior del circuito estén dentro de los rangos lineales de funcionamiento de los transistores y operacionales presentes. Es evidente que este circuito no puede ser construido usando solo resistencias, condensadores y bobinas, pero no estamos ante un elemento fundamental ya que necesita de una fuente de energía interna para funcionar: se trata de una emulación de memristor no pasivo.

Para analizar de forma más completa la teórica existencia de este elemento, el profesor Chua desarrolla también su postulado desde un punto de vista electromagnético, y no solo desde el punto de vista de la teoría de circuitos (caso particular de la teoría del campo electromagnético). Todo el desarrollo matemático que se verá a continuación es un resumen del que presentó el profesor Chua en su artículo (Chua, 1971) en las páginas 514 y 515.

Recordemos las archiconocidas ecuaciones de Maxwell, las cuales describen las fuentes escalares y vectoriales de los campos eléctrico y magnético, o lo que es lo mismo, las divergencias y rotacionales de dichos campos vectoriales:

$$\nabla \vec{E} = {}^{\rho}/_{\varepsilon_0} \quad [9]$$

$$\nabla \vec{B} = 0 \quad [10]$$

$$\nabla \mathrm{X} \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad [11]$$

$$\nabla \mathrm{X} \vec{B} = \mu_0 * (\vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t}) \quad [12]$$

Como la inducción magnética B es igual a la intensidad del campo magnético por la permeabilidad magnética del medio (en este caso se considera el vacío), tenemos entonces:

$$\nabla \mathrm{X} \vec{H} = \vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \quad [13]$$

Se define la llamada "familia de tiempos" función del tiempo como $\tau = \alpha * t$ donde alfa es el parámetro llamado tasa de tiempo. Reescribiendo [11] y [12] en función de esta nueva variable:

$$\nabla \mathrm{X} \vec{E} = -\alpha \frac{\partial \vec{B}}{\partial \tau} \quad [14] \qquad \nabla \mathrm{X} \vec{H} = \vec{J} + \varepsilon_0 \alpha \frac{\partial \vec{E}}{\partial \tau} \quad [15]$$

Si se desarrollan estas cuatro expresiones diferenciales vectoriales en series de potencias formales en alfa, obtendríamos las ecuaciones de Maxwell de orden n, las cuales proporcionan soluciones llamadas campos cuasi estáticos para muchas situaciones electromagnéticas con sus versiones de orden 0 y 1.

Ecuaciones de Maxwell de orden cero:

$$\nabla \mathrm{X} \overrightarrow{E_0} = 0 \quad [16] \qquad \nabla \mathrm{X} \overrightarrow{H_0} = \overrightarrow{J_0} \quad [17]$$

Ecuaciones de Maxwell de orden uno:

$$\nabla \mathrm{X} \overrightarrow{E_1} = -\frac{\partial \overrightarrow{B_0}}{\partial \tau} \quad [18] \qquad \nabla \mathrm{X} \overrightarrow{H_1} = \overrightarrow{J_1} + \varepsilon_0 \frac{\partial \overrightarrow{E_0}}{\partial \tau} \quad [19]$$

Así, los tres elementos fundamentales de los circuitos se pueden identificar con sistemas electromagnéticos cuyas soluciones corresponden a ciertas combinaciones de las ecuaciones de orden cero y orden uno. Por ejemplo, un resistor es un sistema electromagnético cuyos campos de primer orden son despreciables respecto a sus campos de orden cero, resultando en una relación instantánea (sin memoria) entre los campos E_0 y H_0 del dispositivo. Un inductor sin embargo es un sistema que presenta un campo magnético de primer orden despreciable, y un capacitor presenta solo su campo eléctrico de primer orden nulo. Si hablamos de elementos no ideales, la representación

correcta del capacitor/inductor sería el de capacitor/inductor ideal con un resistor en serie. Queda por tanto una combinación que no encuentra su paralelismo con un dispositivo eléctrico fundamental, la que tuviera los dos campos de primer orden no nulos; esta combinación es precisamente la que se daría en un memristor.

Si suponemos la existencia de un dispositivo que cumpla las siguientes dos condiciones:

1-Sus dos campos de orden cero son despreciables respecto a los de primer orden, que denominaremos E_1 , H_1 , D_1 , B_1 , J_1 , donde D es la densidad de flujo eléctrico que equivale al valor del campo eléctrico E multiplicado por la permitividad eléctrica del medio.

2-El material con el que está fabricado es no lineal. LDR´s o diodos Zener son algunos ejemplos de materiales con resistencias no lineales en función de alguna variable de estado o externa.

Entonces, este dispositivo presentará las siguientes relaciones físicas no lineales genéricas entre sus variables eléctricas:

$$J_1 = \mathcal{J}\,(E_1) \;\; [20] \qquad B_1 = \mathcal{B}(H_1) \;\; [21] \qquad D_1 = \wp(E_1) \;\; [22]$$

Si combinamos las ecuaciones 19 y 20 tenemos:

$$\nabla \times \overrightarrow{H_1} = \mathcal{J}\,(E_1) \;\; [23]$$

Lo cual significa que E_1 es función de H_1; siendo f (H_1) dicha función, entonces tenemos finalmente que:

$$D_1 = \wp \; o \; f \; o \; \mathcal{B}^{-1}(B_1) = g\,(B_1) \;\; [24]$$

Según la ley de Gauss: $\oint D\, dS = f(q)$; $D = f'(q)$ [25]

Llegando así a la buscada relación entre carga eléctrica y flujo magnético:

$$q = g'(B_1) \;\; [26]$$

Hemos alcanzado así un punto en el que cabe cada vez menos escepticismo sobre la existencia del memristor, salvo el que todo buen científico debe siempre mantener. Parece que solo queda la espina de la realización efectiva de este cuarto dispositivo eléctrico fundamental, aunque precisamente es este el punto crucial ya que toda teoría necesita sustentarse en datos físicos medibles, de lo contrario está condenada a desaparecer ya que una teoría errónea es casi igual de útil que una teoría cuya veracidad no puede contrastarse ni evidenciarse en la experiencia. Además, ninguna de las hipotéticas ventajas y aplicaciones servirán nunca para nada por mucho que se avance su estudio teórico, si no se puede construir físicamente un memristor.

2.2 Propiedades del dióxido de titanio no estequiométrico TiO(2-x)

2.2.1 Generalidades y propiedades mecánicas

El dióxido de titanio TiO2 presenta, a presión atmosférica, tres estructuras cristalinas distintas: la brookita (ortorrómbica), la anatasa y el rutilo (ambas tetragonales centradas en el cuerpo). Se trata de un semiconductor intrínseco de tipo n, con una banda de energías prohibidas en el entorno de los 3 eV.

Su diagrama binario de fases entre Ti puro y TiO2 es el siguiente:

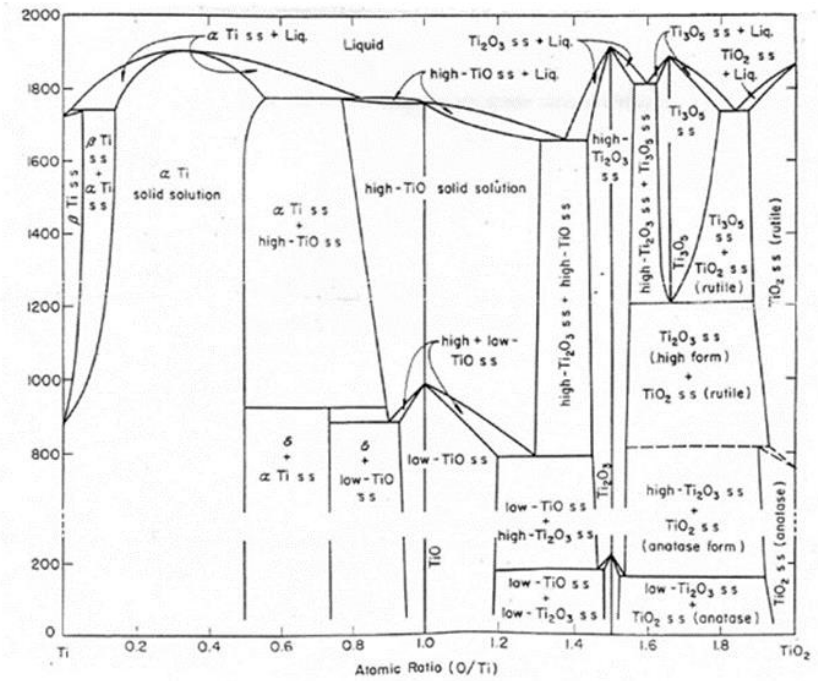


Ilustración 6: diagrama de equilibrio de fases del TiO2. Moreno, Carballo, Jurado y Chicharro. 2009. Pág. 322

Se explican los siguientes efectos mecánicos en función de las vacantes de oxígeno y de la temperatura realizando test de esfuerzo-deformación en muestras de cristal de rutilo (Gardos, 1988), tanto en vacío como a presión atmosférica y a temperaturas entre 600 y 1300 grados Celsius:

-Aparecen pequeñas deformaciones plásticas en el entorno de los 600ºC.

-Las medidas de esfuerzo confirman que en la superficie siempre hay más densidad de vacantes en vacío que a presión atmosférica.

-El rutilo muestra su menor resistencia a cizalladura en su superficie en un rango de reducción de aproximadamente $0.02<X<0.07$

Estas conclusiones fueron obtenidas por un lado interpretando los resultados experimentales y teniendo también en cuenta relaciones entre temperatura, presión parcial de oxígeno y densidad de vacantes (Farb, Johnson y Gibbs, 1965), y por otro lado a partir de análisis de los efectos de las vacantes de oxígenos, la presión y la temperatura en las dislocaciones, formación y movimientos de planos cristalográficos y límites de grano del dióxido de titanio (Ashbee y Smallman, 1963).

Mostramos una figura rescatada del estudio de O.W. Johnson para ilustrar gráficamente la relación entre vacantes de oxígeno y presión parcial del mismo (Gardos, 1988). Se aprecia como una menor presión parcial de oxígeno en el aire y una mayor temperatura facilitan la aparición de una mayor densidad de vacantes de oxígeno, aunque dentro del mismo orden de magnitud de 10^{20} vacantes por centímetro cúbico. La relación inversamente proporcional con la presión parcial del oxígeno era esperable, ya que cuanto mayor sea la concentración de oxígeno en el aire menor será su capacidad reductora.

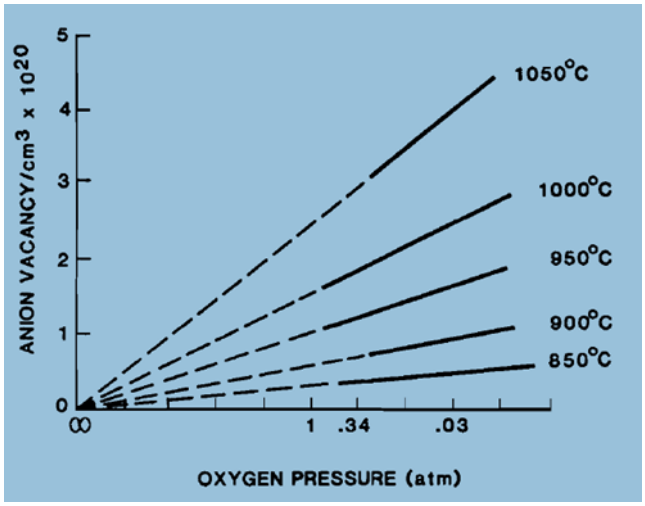


Ilustración 7: relación entre densidad de vacantes y presión parcial de oxígeno. Gardos, M.N. 1988. Pág. 433

Si bien el comportamiento mecánico no será determinante, no debemos olvidar estos estudios que podrán ser de gran ayuda a medida que se desarrolle el papel del TiO2 en los memristores y a la hora de predecir, entender, y/o resolver problemas derivados de altas temperaturas, esfuerzos de tracción, compresión o cizalladura (si los hubiera).

2.2.2 Propiedades termoeléctricas

Analicemos ahora los efectos termoeléctricos de la reducción del dióxido de titanio, el cual por tanto presenta vacantes de oxígeno ya que al reducirse ha oxidado el medio o agente reductor, modificando así su densidad, configuración electrónica, y otras variables de estado que influyen directamente en el comportamiento de los materiales.

Definamos primero las dos propiedades que más nos interesan:

1-Resistividad eléctrica: la inversa de la conductividad eléctrica, la resistencia específica de un material o conductor. Se define como la resistencia de un material por su sección entre su longitud.

$$\rho = R * \frac{S}{l} \quad [27]$$

Sus unidades en el sistema internacional son ohms por metro (Ω*m)

2-Coeficiente de Seebeck:

El efecto Seebeck fue descubierto por el físico alemán Thomas Seebeck. Observó la desviación de una brújula cuando se formaba un circuito cerrado con dos metales a diferente temperatura, lo cual dedujo que debía producir un flujo magnético provocado por una corriente eléctrica en los metales, definiendo así el efecto Seebeck como la diferencia de potencial eléctrico que aparece en la unión de dos conductores de diferente material cuando estos se encuentran a diferentes temperaturas y forman un circuito cerrado entre ellos o con otros conductores intermedios.

$$V = (S_2 - S_1) * (T_2 - T_1) \quad [28]$$

Donde V es la diferencia de potencial medida en la unión, S_i es el coeficiente de Seebeck del semiconductor i, y T_i es la temperatura absoluta del semiconductor i.

El coeficiente de Seebeck por tanto mide el cociente entre tensión y diferencia de temperatura, y tiene unidades en el sistema internacional de voltios entre grados Kelvin.

En el estudio de Lu, Hirohashi y Sato (2006) se emplearon placas de titanio con un 99.9% de pureza; estas se oxidaron en aire a 1173 grados Kelvin durante 110 horas para obtener el dióxido de titanio. Un análisis con difracción de rayos X confirmó que la microestructura del cristal así obtenido era rutilo. El tratamiento de reductor consistió en un calentamiento en horno eléctrico de las placas colocadas en tubos de alúmina y cubiertas con polvo de carbono.

$$TiO_2 \leftrightarrow TiO_{2-x} + \frac{X}{2}\, O_2 \quad [29]$$

Las medidas de resistencia y de tensión-temperatura (coeficiente Seebeck) se tomaron en las probetas de TiO(2-x), las cuales después se volvieron a oxidar durante 5 horas en aire a una temperatura de 1273 K para tomar el resto de medidas (incremento de masa, etc.).

Todas las medidas que no indican otros valores, fueron tomadas a una temperatura de unos 550 K. La densidad de portadores n_e , y la reducción del ratio atómico oxígeno/titanio X han sido calculadas siguiendo las siguientes ecuaciones (Lu, Hirohashi, y Sato, 2006):

$$n_e = \frac{\Delta W}{W_{mol}^{o}} \frac{2N_A}{V_{TiO2}} \quad [30]$$

$$x = \frac{\Delta W}{W_{mol}^{o}} \frac{(2W_{mol}^{o} + W_{mol}^{Ti})}{(W_o + \Delta W)} \quad [31]$$

Donde:

ΔW: valor absoluto de la variación de masa de la probeta de TiO(2-x) tras la reducción

W_{mol}^{o} : peso molecular del oxígeno.

W_{mol}^{Ti} : peso molecular del titanio.

W_o : masa inicial de la probeta de TiO(2-x) tras la reducción.

V_{TiO2} : volumen de la probeta de TiO(2-x), calculado como la masa inicial de la probeta entre la densidad del TiO2.

N_A : número de Avogadro.

Nótese que el parámetro X es adimensional, y es directamente proporcional al incremento de masa.

Detengámonos para entender el cálculo de la densidad de portadores:

Cuando el dióxido de titanio se reduce, y cede átomos de oxígeno (valencia -2) a un agente reductor externo (en este caso el polvo de carbono), cada átomo de oxígeno que abandona el cristal deja un átomo de titanio enlazado con un solo oxígeno; por tanto, como el titanio actúa con valencia +4 en este compuesto, cada átomo de oxígeno vacante deja en el cristal dos portadores positivos de carga, que son precisamente los dos orbitales vacíos que quedan en el átomo de titanio. El cálculo del número de portadores tras este análisis es sencillo: tras la reducción la variación de la masa de la probeta coincide con la masa de oxígeno perdida, y como cada átomo de oxígeno ausente genera dos huecos, la cantidad de portadores será dos veces la masa de oxígeno perdida multiplicada por el número de Avogadro y dividida entre el peso atómico del oxígeno.

Este método por tanto podría generar unos valores algo inexactos ya que no podemos asegurar que en todo el cristal de TiO2 los átomos de titanio que pierden un átomo de oxígeno mantienen su valencia +4, por lo que el cálculo de la densidad de portadores puede desviarse de la realidad si una cantidad importante del titanio pasa a actuar tras la reducción con valencia +3 o +2, lo cual disminuiría la densidad real de portadores de carga.

A pesar de estas dudas razonables, el método parece adecuado ya que los autores compararon sus resultados con otros experimentos similares empleando TiO(2-x) reducido con otros métodos y los valores de resistencia y de densidad de portadores eran muy similares.

Presentamos los resultados más interesantes de este experimento en forma tabulada para su mejor comprensión y posterior análisis. La resistividad eléctrica fue medida a temperatura ambiente.

Tabla 1: resumen de resultados del trabajo de (Lu, Hirohashi, y Sato, 2006)

Temperatura de reducción (K)	Resistividad eléctrica (Ω*mm)	Wo (mg)	ΔW (mg)	n_e $(\frac{huecos*10^{17}}{mm^{-3}})$	2-x
837	-	201.97	0.50	6.5	1.99
1073	160.0	202.41	2.21	28.3	1.95
1173	17.0	196.97	3.70	48.6	1.91
1273	9.3	201.04	3.72	47.9	1.91
1373	2.2	198.63	4.41	57.5	1.89

Estos datos muestran que relación entre temperatura de reducción y el resto de variables es clara y de proporción directa: a mayor temperatura de reducción, mayor es el incremento de masa posterior, lo que quiere decir que la subestequiometría alcanzada en el cristal es más pronunciada y por tanto la densidad de portadores conseguida es mayor.

Esta mayor cantidad de portadores por milímetro cuadrado es sin duda la responsable del aumento de la conductividad del material, o lo que es lo mismo, la reducción de la resistividad del mismo en dos órdenes de magnitud con una diferencia de unos 300 grados en el proceso de reducción.

Observemos dos interesantes gráficos, ambos presentados en el ya citado estudio de Lu, Hirohashi y Sato (2006):

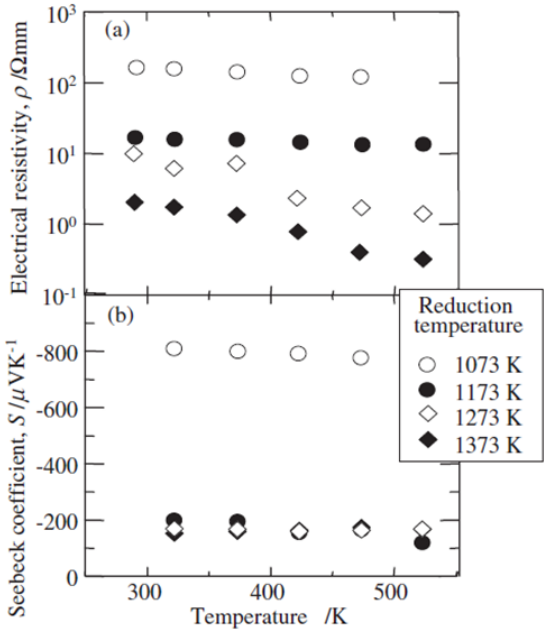


Ilustración 8: resistividad eléctrica y coeficiente de Seebeck en el TiO2

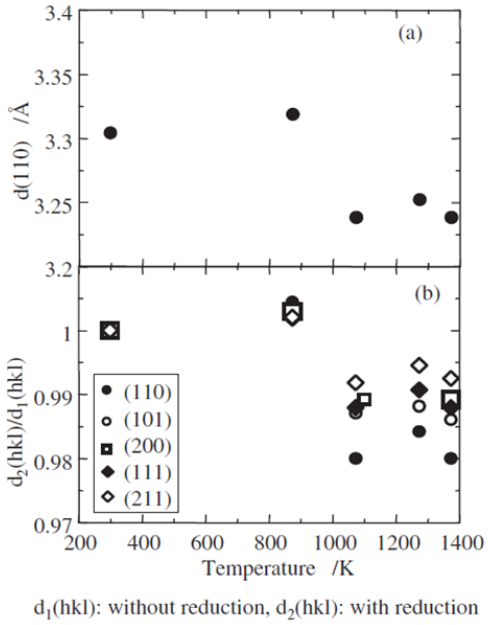


Ilustración 9: reducción de la distancia interplanar en el TiO2 reducido

Vemos en la ilustración 8 que la resistividad eléctrica disminuye con la temperatura de reducción, el coeficiente de Seebeck solo cae en el intervalo de los 1000 a los 1100 K aproximadamente, y posteriormente parece alcanzar un valor mínimo estable y entre los -100 y los -200 microvoltios por grado Kelvin. Además, podemos ver cómo, a partir de reacciones de reducción a 1200 K y en adelante, la resistividad eléctrica no solo disminuye con la temperatura de reducción, sino también con la temperatura del cristal en el momento de la medida. Por tanto, en cualquier aplicación se tendrá que controlar no solo la temperatura de reducción y reoxidación del TiO2, sino también su temperatura de trabajo como material semiconductor.

En la ilustración 9 observamos como la reducción y las vacantes de oxígeno que ésta provoca, implica también un cambio en la estructura cristalográfica del material; en concreto observamos que, en reducciones a temperaturas de 1000 K en adelante, la distancia interplanar del cristal se reduce respecto a la que presentaba antes de la reducción. Esta frontera de temperatura a partir de la cual los planos se acercan parece coincidir con el punto en el que la temperatura de trabajo afecta de forma significativa a la resistividad del material, lo cual podemos hipotetizar que se debe a una relación causal la cual sería interesante estudiar de forma teórica profundizando en el comportamiento eléctrico de los cristales en relación con su microestructura y de forma práctica en el laboratorio.

Para completar este análisis del comportamiento eléctrico del dióxido de titanio, veamos el experimento y las conclusiones del estudio de Tang, Prasad, Sanjinés, Schmid, y Levy (1994), en el cual se midieron las resistividades no solo del TiO2 rutilo, sino también en su forma de anatasa. Además, las temperaturas de reducción fueron mucho menores y los rangos te temperatura en los que se tomaron las medidas fueron inferiores a temperatura ambiente, por lo que servirá para complementar y ver con mejor perspectiva los resultados hasta ahora observados, los cuales vienen de medidas que se tomaron a temperaturas siempre por encima de los 200 K y solo con muestras de rutilo.

Definimos la nomenclatura de las distintas probetas (capas finas obtenidas por deposición):

A1, A2: anatasa pura sin dopar, cada una obtenida con un método de deposición diferente.

A3: anatasa dopada con silicio y germanio.

R1: rutilo puro sin dopar.

Mostramos las gráficas más interesantes de este trabajo: primero la representación de la resistividad medida en las distintas muestras en tres estados diferentes, a saber:

1-Muestras sin modificar.

2-Reducidas en horno de vacío a 400 ºC durante dos horas.

3-Reducidas en horno de vacío a 450 ºC durante dos horas.

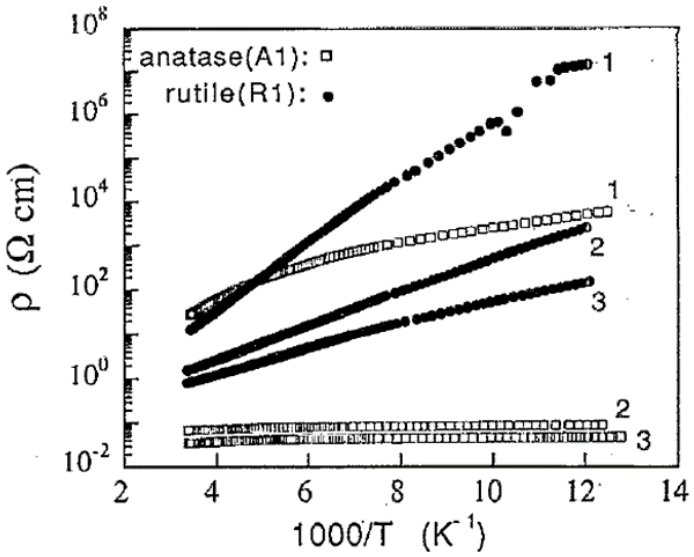


Ilustración 10: resistividad del rutilo y la anatasa en función de mil veces el inverso de T

En estas medidas se controló mediante difracción de rayos X la estructura cristalina de las muestras, las cuales no cambiaron para los tratamientos térmicos aplicados (todos inferiores a 500 ºC)

Como se puede ver en la ilustración 10, la resistividad del rutilo disminuye a medida que la temperatura de la reducción aumenta. También se advierte que dicha resistividad también varía con la temperatura de trabajo de forma inversamente proporcional. Es interesante como este comportamiento ya ha sido observado (Lu, Hirohashi, y Sato, 2006) pero solo ocurría a partir de temperaturas de reducción altas (por encima de los 1200 K), mientras que para temperaturas de reducción bajas la resistividad no variaba con la temperatura. Podemos suponer por tanto que hay un rango de temperaturas (o de relaciones entre temperatura y tiempo total de reacción) de reducción del TiO2 en los que el material obtenido presenta una resistividad constante con la temperatura de trabajo, mientras que para reducciones a temperaturas por encima o por debajo de dicho rango la resistividad del material disminuirá a medida que aumentamos la temperatura.

En cuanto a la anatasa, el comportamiento es muy diferente: cuando se produce la reducción, la resistividad no solo es independiente de la temperatura de trabajo, sino que apenas varía con la temperatura de reducción, sobre todo si lo comparamos con las curvas de 400 y 450 ºC que presenta el rutilo. Además, ciertos valores como la movilidad de Hall y la energía de activación (energía necesaria para mover un electrón desde el borde de la banda de conducción hasta un nivel de movilidad significativo) presentaban fuertes diferencias entre las dos polimorfias del dióxido de titanio.

La principal conclusión para explicar estos fenómenos es que la anatasa, a diferencia del rutilo, sufre una reacción llamada transición de Mott cuando la reducción genera la suficiente densidad de portadores de carga (Tang et al., 1994). Esta reacción, bastante común en los óxidos de metales de transición, consiste en que el material en cuestión pasa a comportarse como un metal en cuanto a la conductividad eléctrica debido a ciertas interacciones entre las fuerzas eléctricas (ley de Coulomb) y el nivel de localización de los portadores de carga, las cuales modifican la distribución y anchura de las diferentes bandas de energía de los portadores de carga que explican este comportamiento metálico.

Este efecto se aprecia también en la siguiente figura, en la que se representa la resistividad tanto del rutilo, como de los tres tipos de anatasa empleados (los cuales todos parecen sufrir la transición metálica de Mott) y para un rango de temperaturas mayor, desde unos 200 K hasta por debajo de los 10 K.

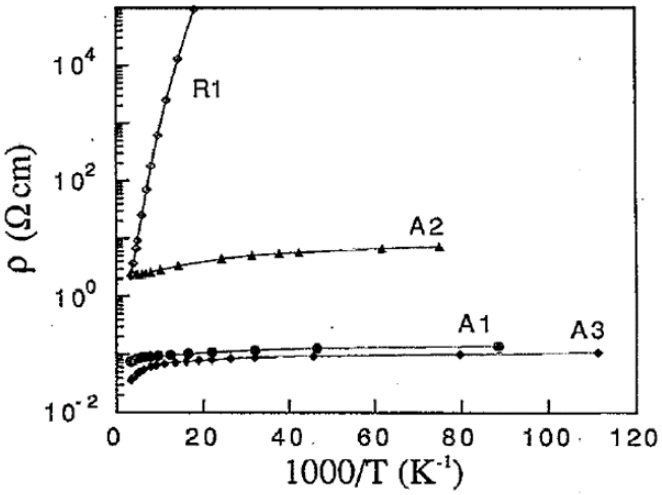


Ilustración 11: resistividad del rutilo y los tres tipos de anatasa estudiados. Tang et al., 1994

El hecho de que el rutilo no presente transición de Mott indica que el TiO2 presenta, en su estructura de anatasa, un radio de donadores de electrones muy grande en comparación con el del rutilo, el cual tiene un radio de donadores pequeño.

Tras este punto, estamos preparados para abordar los distintos modelos del memristor, desde sus formas teóricas hasta sus realizaciones físicas, en las cuales entenderemos la importancia del análisis que se ha hecho del dióxido de titanio.

3 Modelos matemáticos y eléctricos

Nos disponemos en este apartado a estudiar modelos matemáticos y eléctricos que se pueden emplear para la simulación del memristor, así como para su fabricación y posterior verificación de propiedades. Es importante resaltar que hay muchos tipos de memristores físicos diferentes (el memristor de TiO2, el memristor polimérico, el ferroeléctrico, los sistemas memristivos de sodio y potasio…) por tanto existen también gran cantidad de modelos diferentes que se pueden desarrollar tanto matemáticamente como con herramientas de software. Nos centraremos en los más generales, y principalmente, en los aplicables al primer memristor construido por HP basado en el dióxido de titanio.

3.1 Bases de los modelos

Analicemos las representaciones matemáticas generales de las tres posibles formas de memristor. Nótese que todas las ecuaciones describen el comportamiento de un memristor controlado por corriente, es decir, en el que la señal viene de una fuente de corriente y la variable dependiente es por tanto la tensión. De forma análoga se podría describir un memristor controlado por tensión, es decir, en el que la señal viene de una fuente de tensión y la variable dependiente es por tanto la corriente. En este segundo caso, no tendríamos más que trabajar con la memductancia (inversa de la memristancia), cuyas unidades son los siemens.

3.1.2 Ecuaciones de las representaciones globales

-Memristor extendido (Chua, 2019b)

$$V = R(x,i) * i \;\; ; \;\; R(x,0) \neq \infty \; ; \; \frac{dx}{dt} = f(x,i) \;\; [32]$$

La memristancia R(x,i), o bien la memductancia G(x,v), en un memristor extendido es función no solo de variables de estado del dispositivo, sino también de la corriente que lo atraviesa (tensión entre sus terminales). La memristancia debe tener un valor real finito cuando i=0.

-Memristor genérico (Chua, 2019b)

$$V = R(x) * i \;\; ; \;\; \frac{dx}{dt} = f(x,i) \;\; [33]$$

La memristancia en este caso es función tan solo de variables de estado del dispositivo, las cuales presentan una variación temporal que depende siempre o en ciertos intervalos de la corriente que lo atraviesa.

-Memristor ideal (Chua, 2019b)

$$v = R(q) * i \quad ; \quad \frac{dq}{dt} = i \; ; \; R(q) = \frac{d\varphi(q)}{dq} \quad [34]$$

Se define así en el memristor ideal la memristancia R como la derivada del flujo magnético respecto del tiempo, es decir que no es función de ninguna variable de estado interna. Por definición y teniendo en cuenta la ecuación 2 tenemos:

$$\frac{d\varphi}{dt} = R(q)\frac{dq}{dt} \; ; \; \varphi = \varphi_0 + \int_0^q R(q)\, dq \quad [35]$$

La llamada relación constitutiva del memristor ideal controlado por carga.

3.1.2 Características del memristor

Partiendo de toda la teoría presentada hasta el momento (y de las características que ya hemos mencionado en su forma más básica y resumida) se pueden resumir las características del comportamiento eléctrico de los memristores en varios axiomas que en la literatura suelen llamar "*fingerprints*", las huellas que nuestro dispositivo deja como prueba de que se trata de un memristor y no un dispositivo que de forma puntual esté presentando un perfil parecido. Numeraremos y explicaremos a continuación las más importantes para definir las condiciones necesarias y suficientes para que cualquier modelo de simulación o dispositivo real sea considerado un memristor (Biolek y Biolek, 2014). Al referirnos a la gráfica que adopta la tensión del memristor en función de la corriente, la curva de histéresis pellizcada o "*pinched hysteresis loop*", la nombraremos como PHL.

1. En estado estacionario de DC (corriente continua), el memristor disipará una potencia que tenderá a infinito ya que al ser la memristancia directamente proporcional a la carga que ha atravesado el dispositivo (o la memductancia directamente proporcional al flujo magnético), en corriente continua la carga tenderá a infinito por tanto la memristancia también lo hará, lo que eventualmente supondría la destrucción del dispositivo.

2. En estado cuasiestacionario de DC, el memristor se comporta como una resistencia lineal.

3. Si una señal periódica de componente continua nula se aplica al memristor, las variables de estado del memristor y la señal de salida medida adoptan directamente un estado periódico sin pasar por ningún estado transitorio. Es importante resaltar que un modelo de simulación del memristor puede presentar un estado transitorio que no se produciría en el dispositivo real y que se deba a consideraciones numéricas del funcionamiento del programa.

4. Si una señal periódica de componente continua nula se aplica al memristor y su frecuencia tiende a infinito, el área de la PHL tiende a cero, es decir, la curva tensión vs corriente se convierte en una recta por lo que el memristor pasa a comportarse como una resistencia lineal cuyo valor es igual a la memristancia que presentaba al inicio de la excitación.

5. Si una señal periódica con simetría impar se aplica al memristor, entonces la PHL de la curva tensión vs corriente presenta simetría impar.

6. Si una señal periódica con simetría impar se aplica al memristor, la PHL será de tipo degenerada si la curva de la memristancia M(x) es simétrica respecto al eje vertical que pasa por el punto medio del intervalo de valores que toma la variable de estado X.

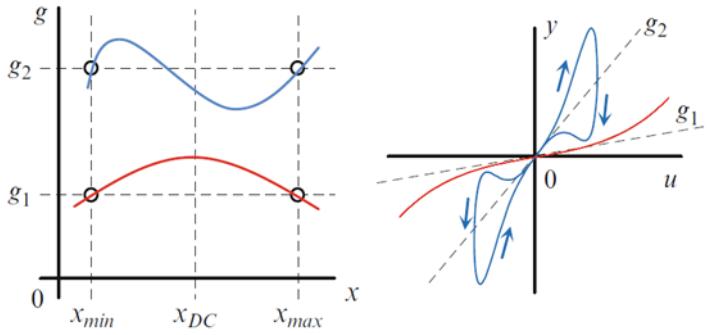


Ilustración 12: PHL de tipo degenerada (g1) y de tipo tangencial. Biolek, 2014. Pág. 120

7. Las curvas i(t) y v(t) en un memristor (cuando una de ellas es la señal aplicada y es periódica sin componente continua) presentan siempre un desfase nulo entre sus puntos de paso por cero, lo que implica que la PHL siempre pasa por cero en el origen. (Chua, 2019a)

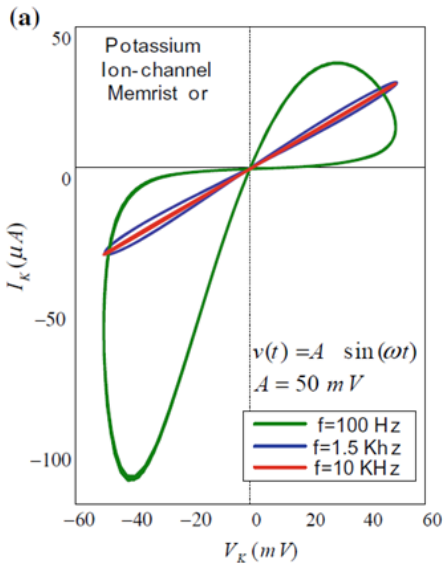


Ilustración 13: evolución del PHL con la frecuencia y paso por cero. Adhikari, Kim, et al. 2019. Pág. 177

Es interesante resaltar que la primera característica es consistente con el comportamiento de otros dos elementos fundamentales: un condensador acaba destruyéndose si lo conectamos a una fuente de corriente continua, igual que una bobina acaba destruyéndose si la conectamos a una fuente de tensión continua.

3.2 Modelos de simulación del memristor

Analizaremos varios modelos de simulación de memristores. El primero ha sido implementado por el autor del presente trabajo mediante la herramienta Matlab Simulink, de los demás presentaremos los resultados de los investigadores para comparar y extraer posibles conclusiones.

3.2.1 Simulación en Simulink

En el artículo de Valsa, Biolek y Biolek (2010) se presenta un modelo eléctrico del memristor que procedemos a implementar mediante el programa de simulación PSIM, una vez lo hayamos desgranado y comprendido.

Este modelo representa un memristor controlado por tensión, por lo tanto aunque hablemos de memristancia para ser más coherentes deberíamos trabajar con la memductancia, inversa de la memristancia.

La idea principal es recrear el comportamiento del memristor mediante un integrador que produce la tensión en función del flujo magnético, y esa tensión controla la conductancia de salida gracias a un transistor tipo JFET. El circuito que presentan, y que construyen físicamente, es el siguiente:

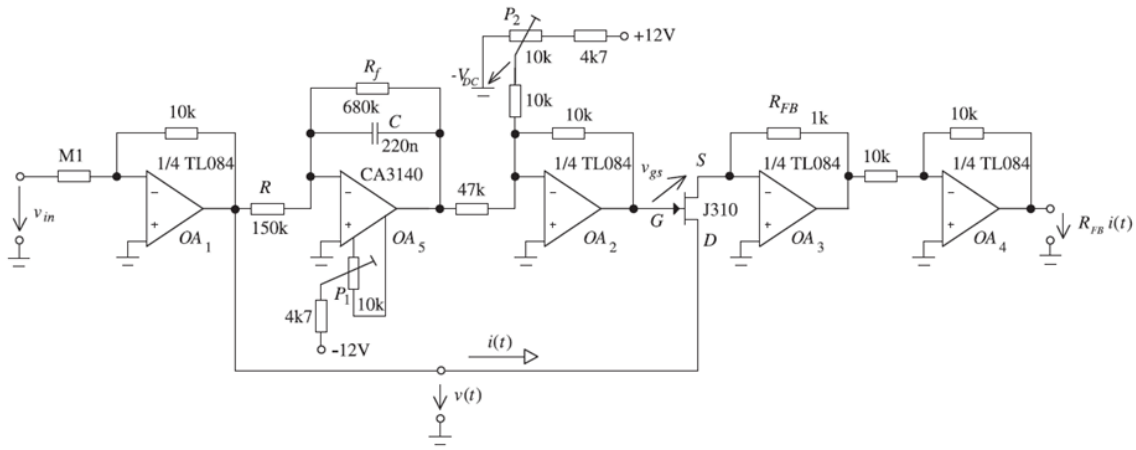


Ilustración 14: modelo eléctrico memristor. Valsa, Biolek y Biolek. 2010.

El primer bloque (OA1) es tan solo una etapa para controlar el nivel de v(t) colocando un potenciómetro en M1. La segunda etapa (OA5) es un integrador, seguido de un amplificador sumador (OA2) que asegura que el nivel de tensión que llega al terminal G del transistor JFET sea adecuado para tener la conductancia inicial que se requiera. Al drenador del JFET llega directamente la señal de tensión aplicada v(t), y la corriente que pasará será proporcional a la conductancia que presente y que está controlada por la integral sobre el tiempo de la tensión, es decir, el flujo magnético. Los últimos bloques proporcionan una salida de tensión en lugar de corriente y corrigen el signo. Debido a que solo pretendemos simular este modelo y no construirlo físicamente, lo adaptamos y simplificamos al implementarlo en Simulink, quedando como se muestra a continuación:

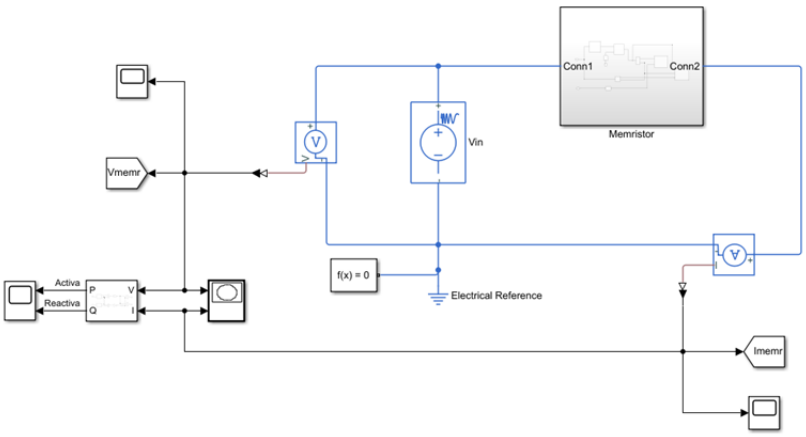


Ilustración 15: modelo en Simulink

Para comprobar la curva de Lissajous, introducimos las señales de tensión y corriente en un bloque "XY Graph". El bloque "Memristor" contiene el modelo como tal:

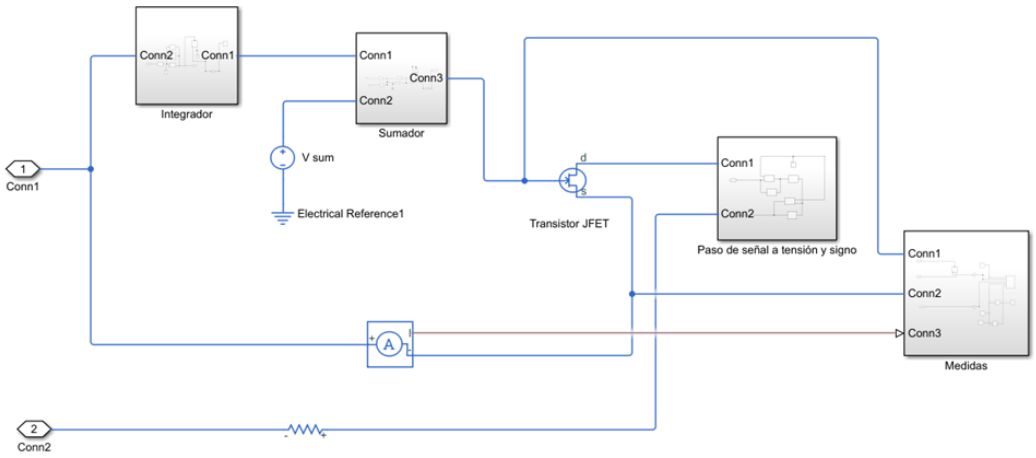


Ilustración 16: modelo en Simulink, bloque Memristor

Hemos sustituido la señal para el sumador obtenida con una referencia de 12 voltios y un divisor de tensión con potenciómetro por una fuente de tensión DC que pondremos directamente al valor que obtendríamos con el montaje original. En el bloque "Medidas" obtenemos y representamos las gráficas de las siguientes magnitudes del circuito:

a) Tensión aplicada al memristor Vin.

b) Tensión entre puerta y surtidor del transistor.

c) Potencia consumida por el dispositivo.

d) Corriente de drenador del transistor.

e) Corriente que atraviesa el memristor.

f) Memristancia (tensión partido por corriente en función del tiempo).

g) Carga que atraviesa el dispositivo.

A continuación, exponemos y analizamos los resultados obtenidos al ejecutar este modelo para distintas condiciones iniciales y entradas de tensión.

-Amplitud Vin (AC senoidal): 0.5 V

-Frecuencia Vin: 2 Hz

-Tensión del sumador: 0.4 V

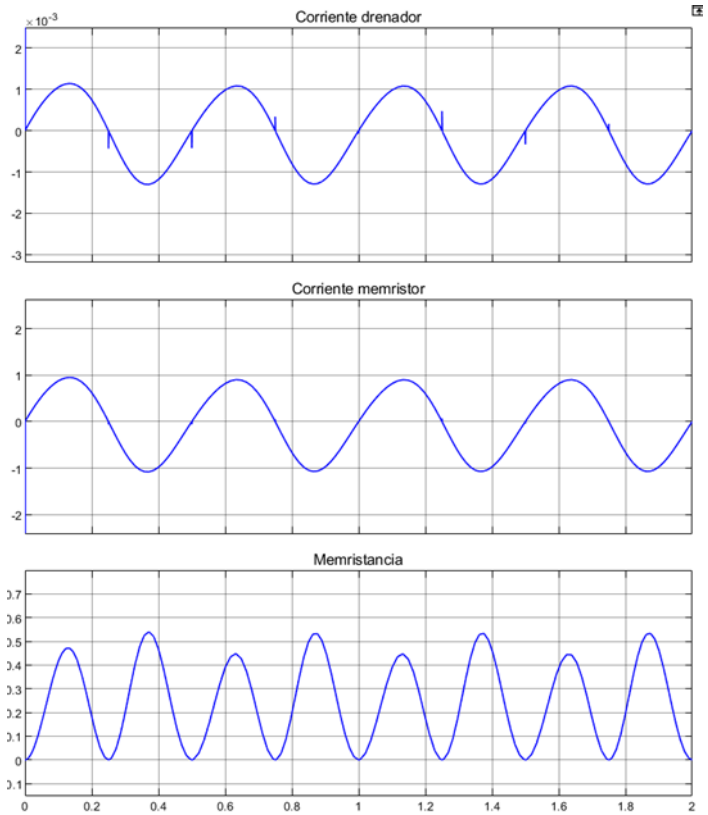


Ilustración 17: gráficas modelo Simulink 1

La curva corriente vs tensión obtenida es:

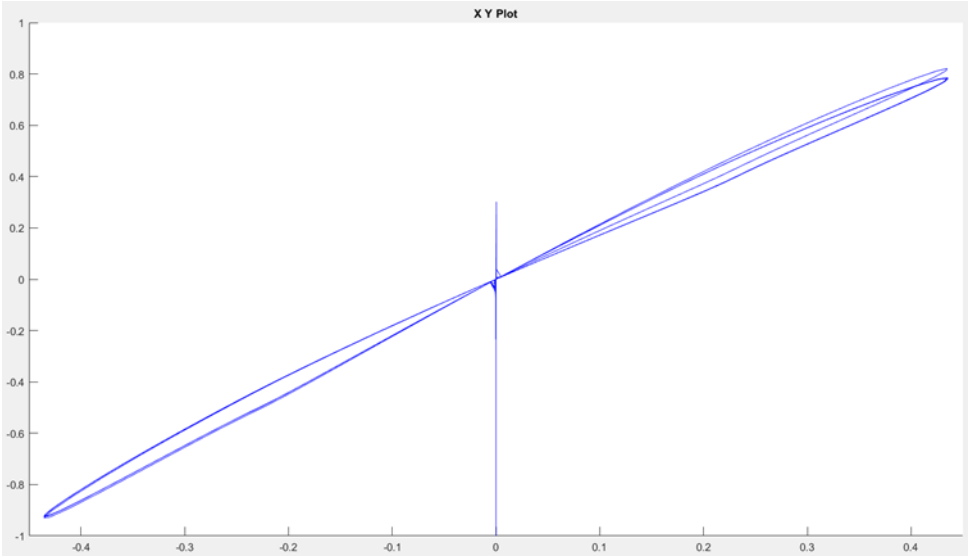


Ilustración 18: PHL modelo Simulink 1

Como vemos, la memristancia cumple la condición de ser estrictamente no negativa. Además, se obtiene el bucle de histéresis pellizcado PHL en la curva corriente en función de la tensión, principal característica del funcionamiento del memristor.

-Amplitud Vin (AC senoidal): 0.5 V

-Frecuencia Vin: 15 Hz

-Tensión del sumador: 0.4 V

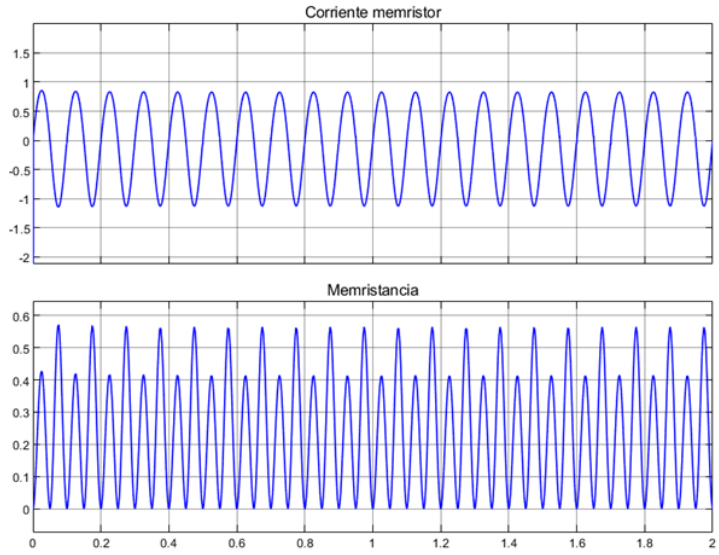


Ilustración 19: gráficas modelo Simulink 2

En esta segunda ejecución confirmamos un comportamiento que ya se apreciaba en la primera: la corriente del memristor (que es la misma que la corriente de drenador pero amplificada por la última etapa) tiene componente continua no nula y negativa, lo cual hace que la curva corriente vs tensión no sea simétrica y que la carga no se mantenga acotada. En los semiciclos de corriente negativa por tanto la memristancia alcanza un máximo menor que en los semiciclos de corriente positiva. Observamos el PHL obtenido en este caso:

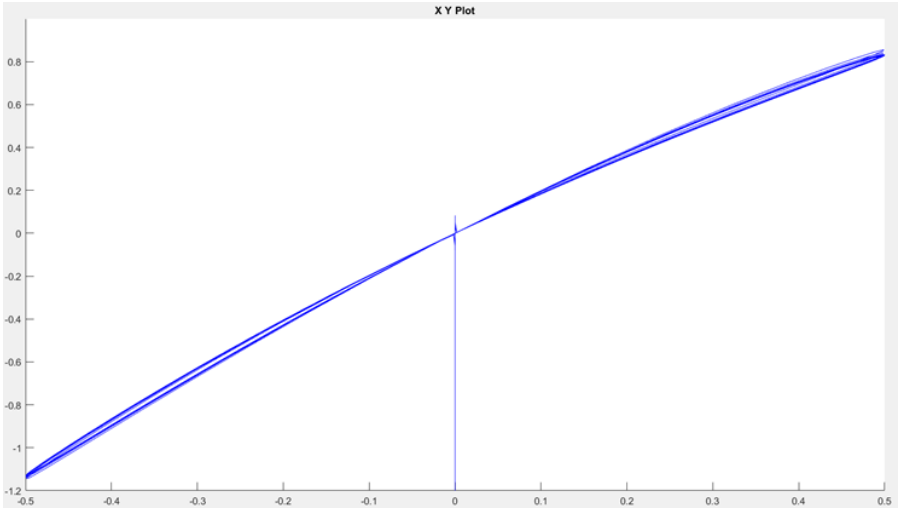


Ilustración 20: PHL modelo Simulink 2

En estas dos ejecuciones del modelo hemos podido comprobar como se obtiene el deseado bucle de histéresis pellizcado PHL en la representación de la corriente en función de la tensión al aplicar señal alterna. Además, también vemos la tendencia a una línea recta o disminución del área del PHL cuando aumentamos la frecuencia de la tensión aplicada, otra de las importantes características de comportamiento de los memristores.

Aplicaremos ahora una señal con frecuencia modulada en forma de rampa decreciente para observar la respuesta transitoria del modelo, y representaremos también la potencia consumida por el dispositivo. Es importante señalar que nos referimos a la "potencia pasiva", es decir, que no tenemos en cuenta la potencia consumida para la alimentación de los operacionales que forman parte del modelo del memristor. Introducimos una señal senoidal cuya frecuencia varía de forma constante desde los 10 Hz hasta 0.2 Hz, obteniendo los siguientes resultados:

Potencia para frecuencia de 10 Hz:

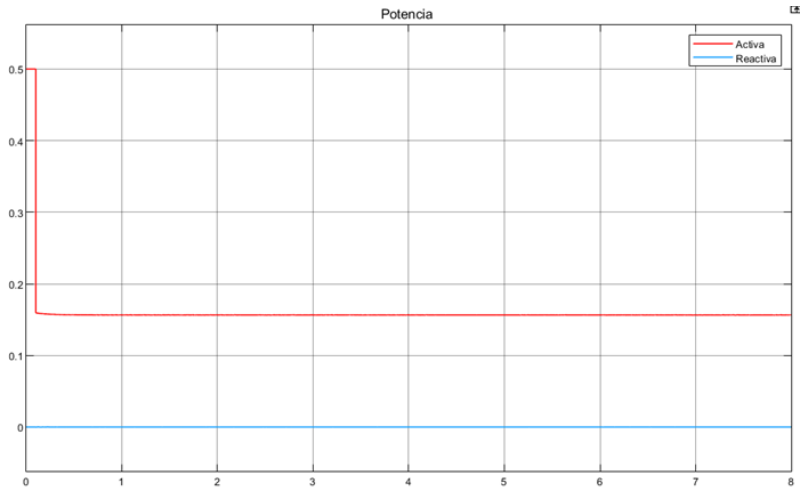


Ilustración 21: potencia disipada modelo SImulink

Como ya habíamos visto, no existe desfase entre tensión y corriente (alcanzan sus máximos y pasan por cero en los mismos instantes) por lo que era de esperar que la potencia reactiva fuese nula, como así es. El valor de la potencia activa muestra un salto debido a condiciones iniciales internas del bloque "Power". Su valor es constante ya que no se proporciona la potencia instantánea, la cual es relativamente poco útil, sino la potencia media.

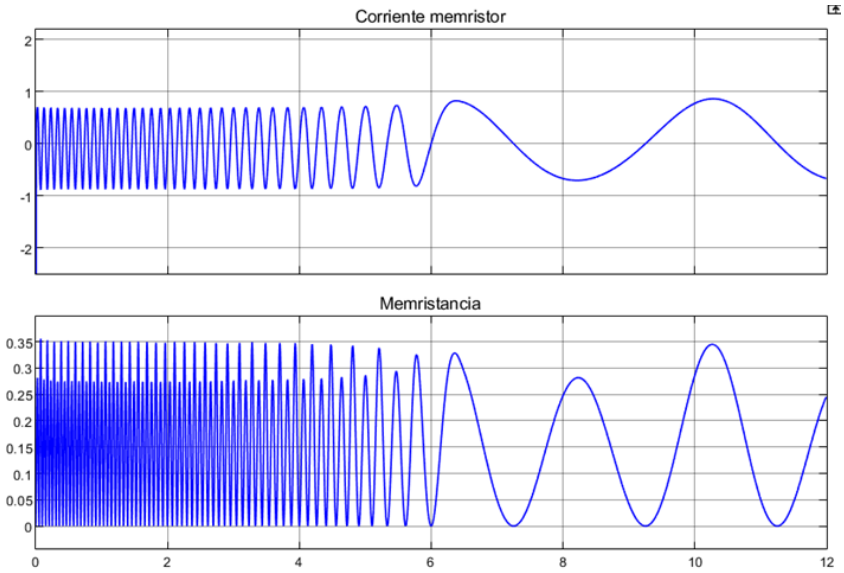


Ilustración 22: gráficas modelo Simulink 3

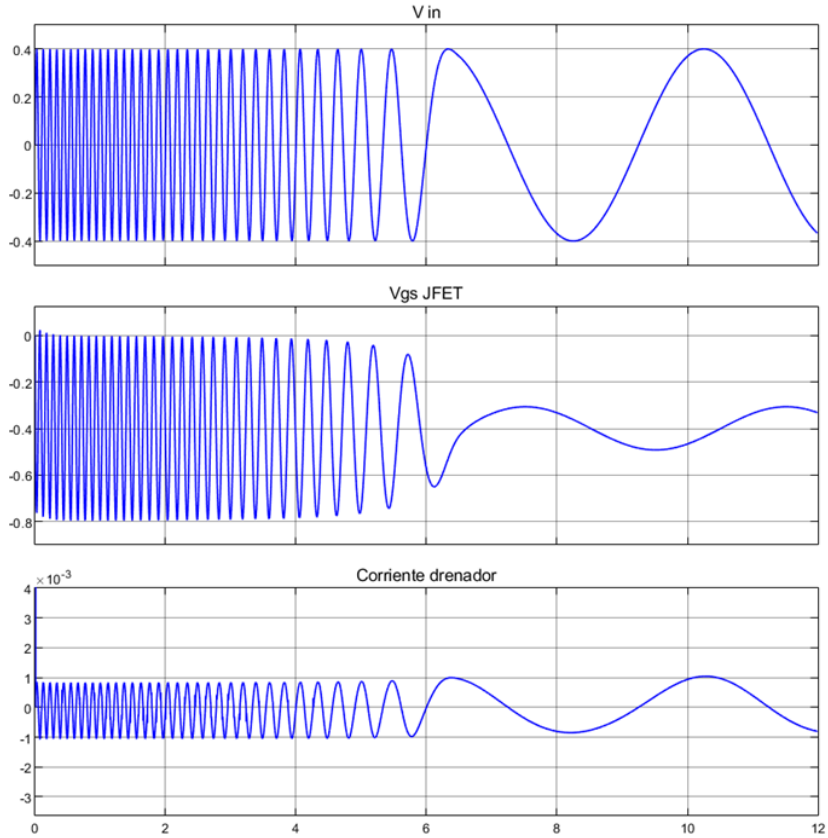


Ilustración 23: gráficas modelo Simulink 4

En las ilustraciones 22 y 23 se aprecia como para una entrada alterna de amplitud constante pero periodo creciente que termina estabilizándose, las magnitudes que nos interesan del memristor tales como la memristancia y la corriente que lo atraviesa no sufren ningún tipo de retardo y presentan un régimen transitorio sin saltos ni discontinuidades hasta alcanzar el nuevo estado estable.

Por último, procedemos a incluir otro bloque "XY Graph" en el modelo, en este caso la entrada X será la función carga q(t) y la entrada Y será la integral sobre el tiempo de la tensión alterna aplicada Vin, que se corresponde físicamente con el flujo magnético. Ejecutamos la simulación con una frecuencia de dos herzios en la tensión Vin solamente durante un ciclo, es decir durante 0.5 segundos, y observamos la gráfica de la memristancia en función del tiempo:

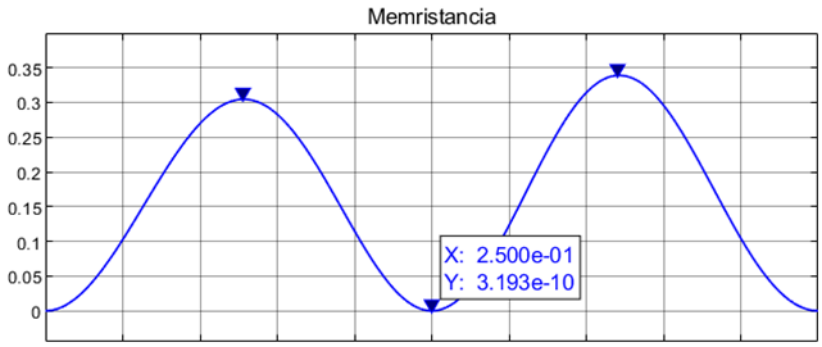


Ilustración 24: gráficas modelo SImulinnk 5

Según las definiciones que hemos visto para un memristor ideal, el punto marcado en la figura superior al ser un paso por cero de la función memristancia se correspondería con un extremo relativo de la función $\varphi(q)$. Vemos en la gráfica de la carga q(t) (ilustración 25) que para el instante t=0.25 segundos la carga alcanza un valor de q(0.25)=0.123 C; si observamos la gráfica flujo vs corriente de la ilustración 26, podemos ver que la correspondencia $\varphi(q)$ alcanza un máximo precisamente en un punto poco por encima de los 0.12 culombios.

Aunque matemáticamente esta afirmación es totalmente incorrecta, ya que esta función al ser multivaluada no es derivable, nos sirve para poder prever el comportamiento que esperaremos de un memristor real o de un modelo más adecuado, así como para confirmar que este modelo aunque esta lejos de ser perfecto, se ajusta bastante a las definiciones y características dadas en el punto anterior.

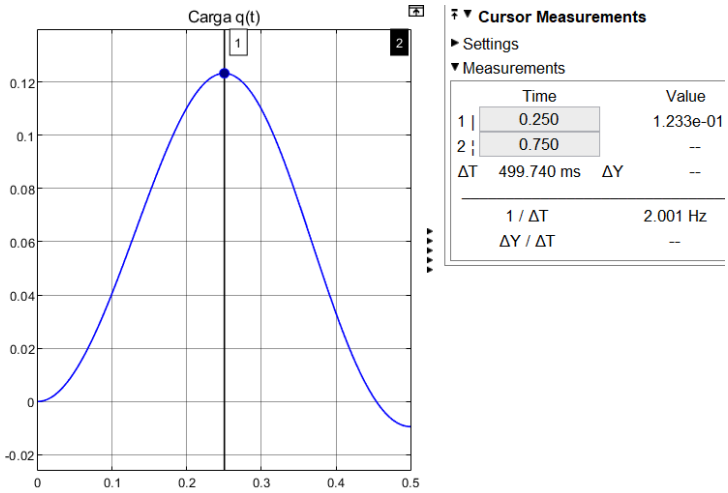


Ilustración 25: gráficas modelo Simulink 6

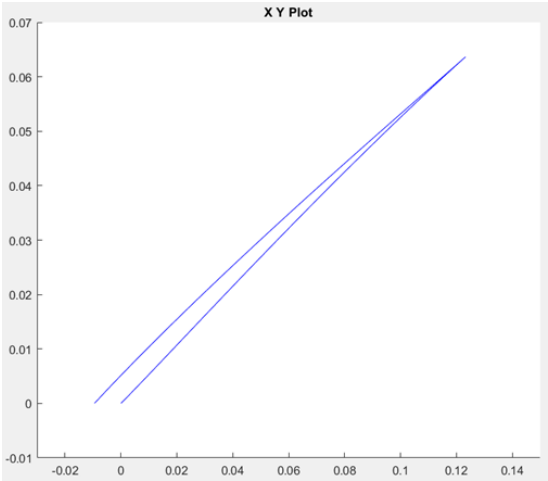


Ilustración 26: gráficas modelo Simulink 7. Flujo vs corriente

3.2.2 Simulación en LabView

Mientras que el modelo presentado en el anterior punto se basa en un circuito con elementos eléctricos y electrónicos clásicos que en conjunto y combinados de la manera adecuada imitan el comportamiento de un memristor, exponemos ahora un modelo de simulación que representa el comportamiento matemático de las variables internas y eléctricas en un memristor real. Dicho modelo es implementado en el software de programación gráfica Labview, y fue desarrollado tanto en su parte teórico/matemática como en el diseño del programa por Sahin y Güler (2017).

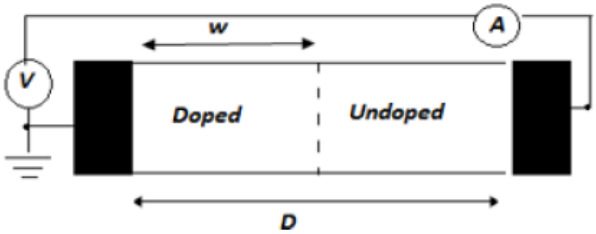


Ilustración 27: parámetros dimensionales del memristor. Sahin y Guler. 2017

Aunque el funcionamiento físico del memristor construido por el equipo de HP se detallará en el siguiente punto, debemos avanzar un pequeño resumen para entender el diseño del presente modelo: el memristor presenta dos zonas diferenciadas de dióxido de titanio, una dopada con portadores de carga en forma de vacantes de oxígeno, y otra sin vacantes. Como hemos visto en los estudios del comportamiento eléctrico del TiO(2-x), estas vacantes aumentan la conductividad del material y además van aumentando en densidad (aumenta el número de vacantes, en presencia de un agente reductor externo) cuanto más tiempo esté sometido a una tensión eléctrica, lo que permite delimitar una frontera entre el material con/sin vacantes que se va desplazando. Esta frontera se encuentra a unan distancia W del terminal a la que se aplica la polaridad positiva de la tensión. La distancia total entre terminales será D, y por tanto el valor de W oscilará entre 0 y D.

Entonces, conociendo los valores límite de la memristancia (ohmnios), es decir el valor que presenta el memristor para W=0 (valor máximo de resistencia y mínima conductividad: off) y para W=D (valor mínimo de resistencia y máxima conductividad: on) se podrá ajustar el comportamiento de la memristancia respecto al tiempo en función precisamente de la posición de la frontera. Esta variable será adimensional , el tanto por uno de la anchura del memristor que presenta alta densidad de vacantes de oxígeno, y su expresión es:

$$x(t) = \frac{W(t)}{D} \in [0,1] \quad [36]$$

$$M(t) = Roff + (Ron - Roff) * x(t) \quad [37]$$

Para encontrar la expresión de esa posición de la frontera x(t) se parte de la expresión de la velocidad de desplazamiento de dicha frontera, función de la resistencia de la zona dopada, de la intensidad de carga y de su propio valor x:

$$\frac{dx}{dt} = K * i(t) * f(x) \;\; ; \;\; K = \mu_v * \frac{Ron}{D^2} \quad [38]$$

Donde $\mu_v = 10^{-14}\, m^2/(s * V)$ es el parámetro de movilidad iónica.

Los autores del estudio obtienen una expresión válida para f(x) empleando funciones ventana, pudiendo así resolver la ecuación diferencial 38 para obtener la expresión buscada:

$$x(t) = 1 - \frac{1}{A * e^{4Kq(t)} + 1} \;\; ; \;\; A = \frac{Roff - Ro}{Ro - Ron} \quad [39]$$

Solo resta entonces programar el memristor como un dispositivo de resistencia variable con las funciones obtenidas (37 y 39), dependientes del tiempo y de la intensidad o tensión aplicada (la carga es la integral sobre el tiempo de la intensidad, la cual es directamente proporcional a la tensión aplicada)

Las gráficas más importantes obtenidas tras simular el comportamiento de este dispositivo al aplicarle una señal de corriente de amplitud 0.5 mA y frecuencia de 5 Hz se muestran a continuación:

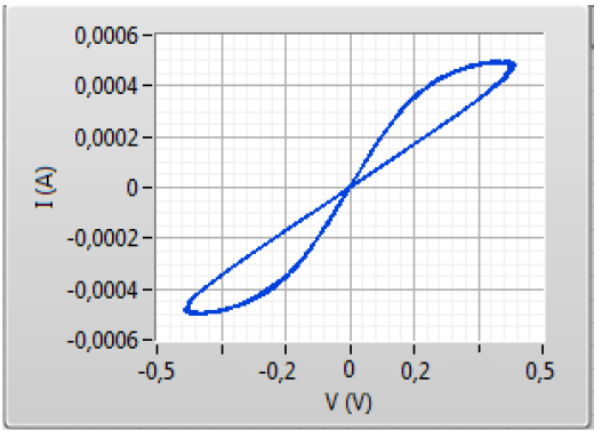


Ilustración 28: PHL 1 modelo LabView. Sahin y Guler. 2017

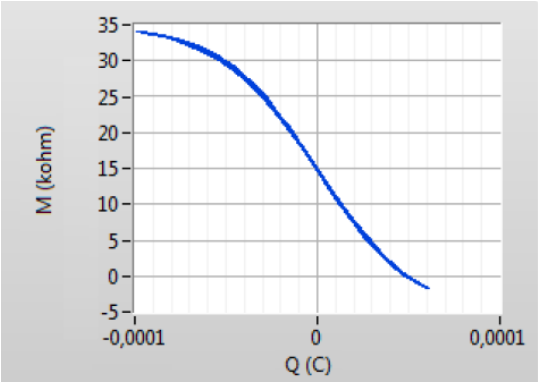


Ilustración 29: memristancia frente a carga modelo LabView. Sahin y Guler. 2017

Se aprecia el ya famoso *pinched hysteresis loop* en la ilustración 28. Además, en la ilustración 29 vemos la curva de tipo no degenerada de la memristancia en función de la carga que ha atravesado el dispositivo. Además, en la ilustración 30 se muestra la relación fuertemente no lineal entre la memristancia y la tensión aplicada. Por último vemos como, también en concordancia con la teoría expuesta y con el modelo implementado en Simulink, el PHL se estrecha a medida que aumentamos la frecuencia, tendiendo al comportamiento de resistor lineal.

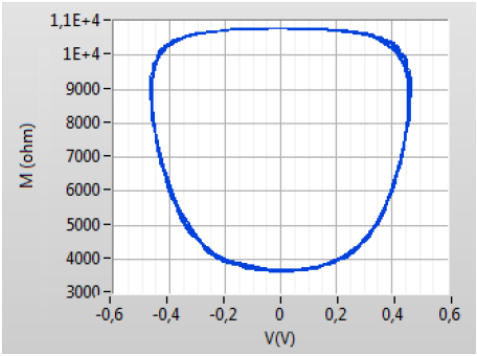


Ilustración 30: memristancia vs tensión modelo LabView. Sahin y Guler. 2017

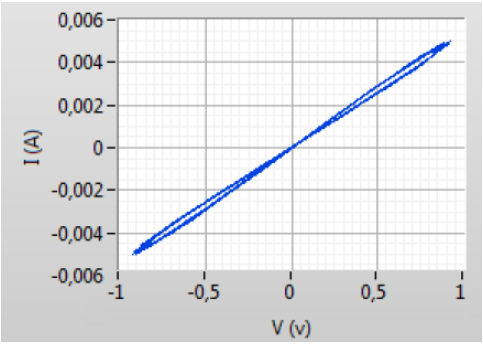


Ilustración 31: PHL 2 modelo LabView. Sahin y Guler. 2017

4 El memristor de HP

Procedemos a exponer las principales caracterizaciones, experiencias, datos graficados obtenidos de ellas y conclusiones extraíbles que se llevaron a cabo en los laboratorios de HP con el primer memristor reconocido como tal. Dichos resultados fueron publicados en la prestigiosa revista *Nature nanotechnology* con la autoría de Williams, Yang, Pickett, y otros (2008).

El nanodispositivo que construyeron y que, según las definiciones vistas y teniendo en cuenta los datos experimentales obtenidos, presenta un comportamiento memristivo, se muestra a continuación:

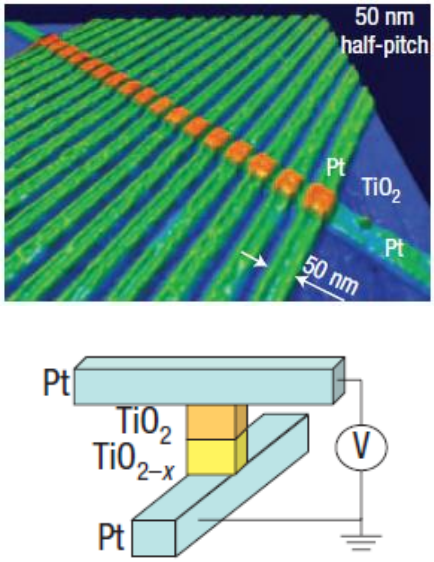


Ilustración 32: el dispositivo memristivo de HP. Williams, Yang, et al. 2008.

Un conjunto de filamentos conductores de platino coplanares se toman como electrodo superior, y un filamento conductor de platino perpendicular a los primeros y por debajo de ellos se toma como electrodo inferior, y se conecta a la masa del circuito eléctrico que se forma al alimentar con una fuente de tensión tal y como se ve en la ilustración 32. Entre los electrodos se encuentra la unión que presenta la conmutación entre on y off (en el sentido de la conductividad) con características memristivas, y que está formada por cristal de dióxido de titanio TiO2. Este cristal fue recocido en una atmósfera de 95% de nitrógeno y 5% de hidrógeno a 550 grados Celsius durante 2 horas para generar una capa de vacantes de oxígeno (reducción del TiO2), que actuarán como dopantes de tipo n.

El comportamiento al aplicar tensiones variantes (alternas) y bipolares se ve graficado en la ilustración 33, donde vemos la representación de la intensidad en función de la tensión aplicada. Se aprecia la alta repetibilidad (que podemos asociar con determinismo) de la variación de la resistividad/conductividad del dispositivo.

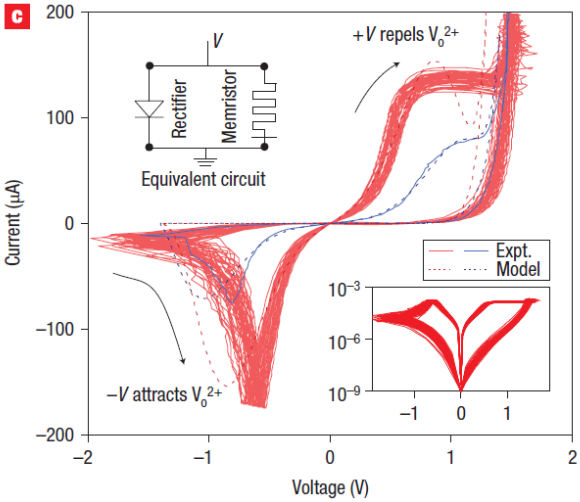


Ilustración 33: PHL del memristor de HP. Williams, Yang, et al. 2008.

En la zona on o de conducción, la corriente en función de la tensión es exponencial en los dos cuadrantes y similar a una curva de efecto túnel. La zona off o de no conducción presenta un comportamiento rectificador que hace pensar en un transporte de tipo Schottky en una de las uniones metal-óxido. Los contactos entre material metálico y semiconductor son típicamente óhmicos , es decir con una relación lineal entre tensión e intensidad, cuando hay fuerte dopaje en el semiconductor. Sin embargo, con niveles bajos de dopaje la unión suele ser rectificadora (tipo Schottky).

Para profundizar mejor en el comportamiento eléctrico del dispositivo, se construye una probeta con la siguiente distribución:

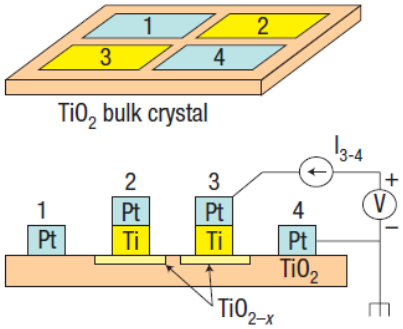


Ilustración 34: probeta TiO2. Williams, Yang, et al. 2008.

Los electrodos 1 y 4 son películas de platino, mientras que los electrodos 2 y 3 son películas de titanio más platino. La capa de titanio de los electrodos 2 y 3 ayuda a crear una zona con alta concentración de vacantes de oxígeno en la capa de dióxido de titanio común que une los cuatro pads. Entre los electrodos de platino-titanio se mide una resistencia de unos 40 ohmios; entre los electrodos de platino la resistencia es de hasta 4 órdenes de magnitud superior. Las curvas de corriente-tensión obtenidas al aplicar voltajes del mismo rango entre los distintos pares de electrodos posibles se muestran a continuación:

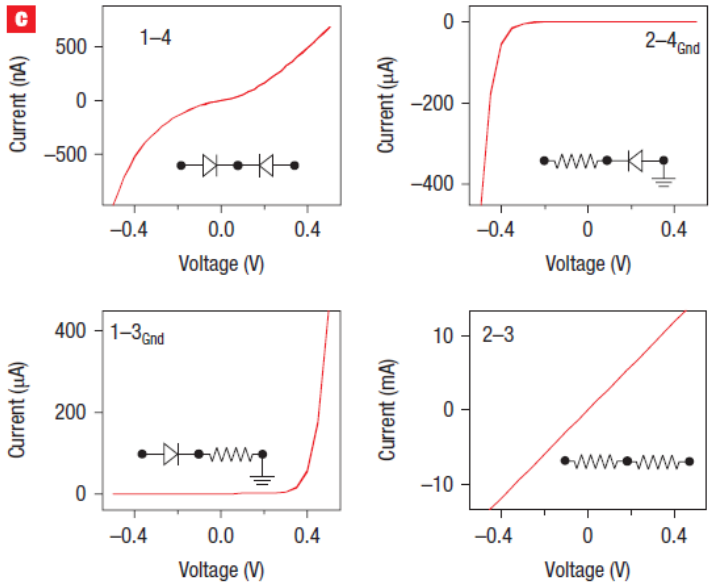


Ilustración 35: curvas I-V entre los electrodos de la probeta. Williams, Yang, et al. 2008.

Al aplicar tensión entre dos electrodos de titanio (figura 2-3 de la ilustración 35) se aprecia una inequívoca relación lineal entre tensión e intensidad, por tanto las dos superficies de unión son óhmicas. Sin embargo, entre los electrodos de platino (figura 1-4 de la ilustración 35) la relación V-I es claramente no lineal. Al aplicar la tensión al electrodo de platino y llevar a masa el electrodo de platino-titanio (figura 1-3 de la ilustración 35) se observa un comportamiento rectificador que se corresponde claramente con un modelo de circuito de diodo y resistencia en serie.

La baja concentración de vacantes bajo el electrodo de platino mantiene la barrera de tipo Schottky responsable del comportamiento rectificador que hay en la unión Pt-TiO2. La alta concentración de vacantes bajo el electrodo de platino-titanio rompe la barrera de tipo Schottky, ya que estas vacantes de oxígeno son positivas eléctricamente, y presentan movilidad suficiente para desplazarse al ser atraídas por la polaridad negativa de la tensión a la que está conectada el electrodo de platino.

Para averiguar si este fenómeno ocurre de forma uniforme o por el contrario de forma localizada, el electrodo 4 fue separado en dos mitades (4_1 y 4_2), ambas conectadas a masa o tensión negativa de la fuente de tensión.

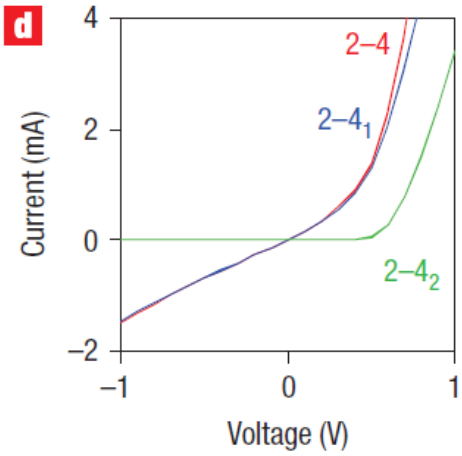


Ilustración 36: distribución irregular de la conmutación. Williams, Yang, et al. 2008.

La relación corriente-tensión entre el electrodo de Pt-Ti y los dos semi-electrodos de Pt se observa en la ilustración 36. Como vemos, el semi-electrodo 4_2 aún presenta un comportamiento altamente rectificador, mientras que en el 4_1 la conmutación ya ha comenzado, y se ha alcanzado un nivel de conductividad alto para las dos polaridades de tensión aplicada. Por tanto este cambio en la conductividad no se produce de forma global en toda la superficie del contacto entre electrodo y TiO2, sino que se produce de forma local y se va extendiendo con el tiempo. La conmutación se completa en tiempos del orden de 50 nanosegundos.

En general se concluye que la conmutación entre on y off del dispositivo ocurre en la superficie superior, que cambia de un comportamiento no óhmico rectificador a un estado de conducción con histéresis. El circuito más sencillo que permite modelar este comportamiento constaría tan solo de un memristor ideal en paralelo con un diodo. La conexión en directa o inversa del diodo genera el mismo efecto que el orden que presenten las capas de TiO2 con o sin dopantes en el dispositivo. Todo esto se visualiza en la ilustración 37.

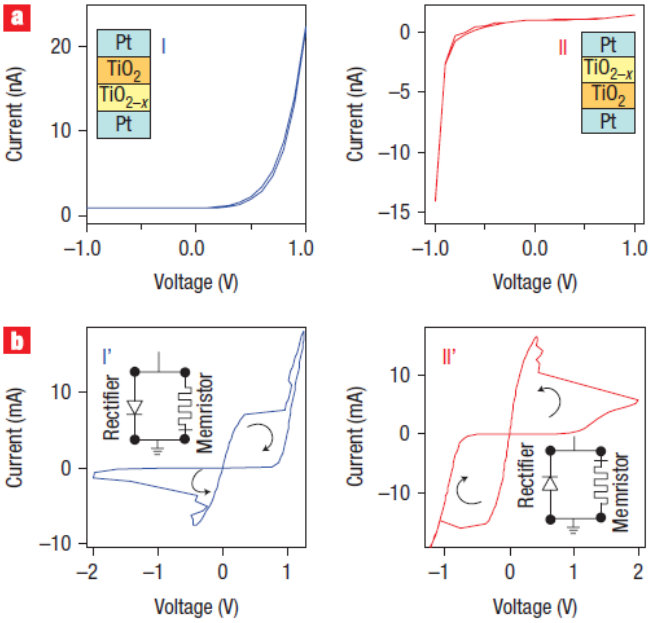


Ilustración 37: verificación del comportamiento del dispositivo. Williams, Yang, et al. 2008.

La siguiente ecuación describe el comportamiento de la corriente en el dispositivo reflejado en la ilustración 33:

$$I = W^n * \beta * \sinh(\alpha V) + X * (\exp(\gamma V) - 1) \quad [40]$$

El primer término representa un memristor ideal. El seno hiperbólico es la aproximación al estado on que corresponde con un efecto túnel en los electrones a través de una barrera residual. α y β son constantes de ajuste, y W es la variable interna del memristor proporcional a la integral del voltaje aplicado al dispositivo. Está normalizado para tomar valores entre 0 (estado off) y 1 (estado on). El segundo término representa la aproximación I-V de un rectificador. Tiene también sus constantes de ajuste (X, γ) que permiten calibrar el estado inicial del dispositivo. El exponente n de la variable de estado responde a la necesidad de que el modelo se aproxime lo máximo posible a los datos experimentales. La mejor correspondencia entre la ecuación 40 y los valores medidos y representados se alcanza para valores de n entre 14 y 22.

5 Aplicaciones del memristor

El comportamiento novedoso de estos dispositivos los hace muy interesantes teóricamente para desarrollar nuevas tecnologías o mejorar las existentes. Además, los distintos tipos de memristores que se han construido físicamente en los últimos años muestran unas características prometedoras a la hora de llevar a la práctica esas posibilidades teóricas, como una excelente escalabilidad, compatibilidad con tecnologías CMOS, y altas velocidades de conmutación. En este apartado explicaremos las principales aplicaciones tecnológicas reales que actualmente se están desarrollando.

5.1 Ciberseguridad y sistemas caóticos

Aunque hay más de una propuesta de circuito generador de números aleatorios basados en memristores nos centraremos, tras una introducción al tema, en un trabajo en concreto en el que se presenta un generador de números aleatorios que por primera vez en el caso de los basados en memristores ha superado sin ningún tipo de procesamiento adicional los quince tests estandarizados de medida de aleatoriedad propuestos en la norma 800-22 del *National institute of standars and technology*. Por último, analizaremos un método de encriptación basado en dispositivos memristivos.

5.1.1 Números aleatorios en la seguridad

En los sistemas informáticos en general, la seguridad y la garantía de privacidad en el acceso a información es fundamental. En particular, los sistemas que se emplean para aplicaciones de banca, almacenamiento de archivos y datos personales, o gestión de acceso y de control de personas físicas a un edificio o instalación tienen como un pilar principal y crítico el proporcionar la máxima seguridad posible ante intentos de robo de datos o acceso y manejo no permitido del sistema (lo que se conoce coloquialmente como "hackear").

En este sentido, generar series de números verdaderamente aleatorios permite al administrador crear contraseñas totalmente arbitrarias, lo que disminuye las probabilidades de que dicha contraseña sea conocida o deducida por nadie más que el usuario. También se pueden utilizar las series de números aleatorios para generar claves de encriptación seguras, que permiten enviar información codificada de tal forma que solo un usuario que conozca dicha clave pueda "traducir" la información, quitándole así peso a la seguridad de la comunicación en sí. Una de las formas de medir la incertidumbre de una fuente de información muy empleada en el criptoanálisis es analizar la entropía condicional, que emplea cálculos probabilísticos. Cuando en la relación entre el mensaje cifrado y la clave o entre el mensaje cifrado y el mensaje se alcanza un valor de entropía condicional nulo, quiere decir que no hay incertidumbre y que uno se puede inferir del otro.

Hay dos grandes tipos de generadores. Por un lado, los generadores de números aleatorios reales (TRNG) que suelen emplear un elemento físico con comportamiento estocástico como fuente para generar las series aleatorias midiendo los valores de su magnitud física, normalizándolos y acotándolos al rango deseado. El ruido térmico o los saltos entre orbitales de los electrones de un átomo son algunos de estos métodos más importantes. Por otro lado, los generadores de números pseudoaleatorios (PRNG) que emplean algoritmos o ecuaciones recursivas que, siendo deterministas, pueden generar series de números cuyos patrones de repetición son difíciles de detectar o presentan periodos muy largos. La calidad de estos métodos recae principalmente en la aleatoriedad y la frecuencia de cambio de la semilla, es decir del primer término a partir del cual se calcula el resto de la serie.

Los distintos sistemas operativos tienen sus propios métodos para mantener la entropía de sus generadores de números pseudoaleatorios lo más alta posible. Caben destacar ciertos sistemas basados en Linux, los cuales emplean las entradas en el teclado o los movimientos del ratón como fuentes de aleatoriedad. Pero en cualquier caso, es fundamental mejorar la incertidumbre de los métodos tanto de generación de semillas como de generación de series aleatorias completas, y en esa tarea se centran actualmente muchos profesionales y estudios tecnológicos y científicos, como el que exponemos en este trabajo.

5.1.2 Generador basado en memristor de difusión

Presentamos los resultados y el funcionamiento del circuito generador de números aleatorios reales o TRNG presentado en (Jiang, Belkin, et al., 2017)

El fundamento principal de este dispositivo es el comportamiento de un memristor de plata y dióxido de silicio ante entradas de tensión de onda cuadrada de altas frecuencias. Cuando la tensión que se le aplica al memristor pasa de 0 a un valor suficiente para provocar la conmutación pero a una velocidad tan alta como para considerar que aplicamos un salto virtualmente instantáneo, el memristor tarda un tiempo determinado cuyo valor presenta, a medida que repetimos ciclos de aplicación de la señal de tensión cuadrada, un comportamiento estocástico y no predecible. Sin embargo, cuando la onda cuadrada termina y la tensión cae a 0, el retardo mostrado es mucho más constante, entorno a 100 microsegundos. La corriente cae entonces a su valor de estado off el cual es tan pequeño que se puede despreciar (orden de picoamperios).

El comportamiento estocástico del tiempo de retardo en la conmutación al estado on es modelado y justificado por los autores mediante simulaciones de dinámica de nanopartículas en los que aplicaron consideraciones eléctricas, químicas y termodinámicas. Gracias a estos modelos matemáticos y al control de la composición y métodos de fabricación del dispositivo llegan a la conclusión de que la aleatoriedad reside básicamente en los procesos de formación de canales de conducción en la plata.

En cada ciclo (tren de pulsos de onda cuadrada de tensión aplicada al memristor) la distribución de nanopartículas de plata que forman canales de conducción es diferente, lo cual provoca una fuerte anisotropía en el flujo de portadores que pueden atravesar el material afectando directamente a la conductancia. Cuando la distribución de nanopartículas es lo bastante uniforme en el electrodo, entonces la conductividad es mayor y por tanto la intensidad que atraviesa el memristor aumenta más fácilmente, llegando más rápido al estado on de máxima conducción si se sigue aplicando la tensión. Sin embargo, en los ciclos en los que la distribución de la que se parte está muy localizada (picos) en ciertas zonas, entonces la intensidad en el memristor crecerá más lentamente, y el retardo observado aumentará.

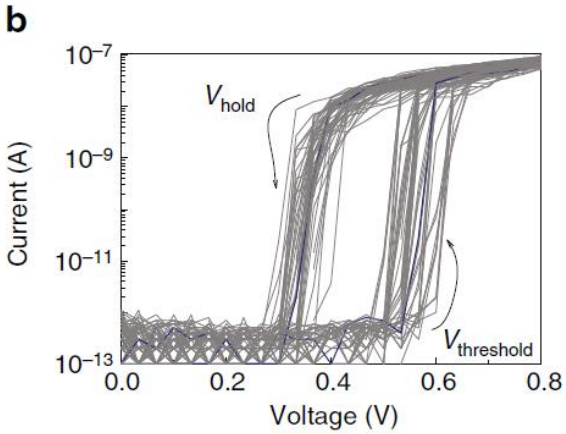


Ilustración 38: corriente para ciclos de tensión de onda cuadrada en memristor Ag-SiO2. Jiang, et al. 2017

Vemos en la siguiente figura el retardo en la conmutación del memristor de Ag-SiO2. Este retardo es aleatorio porque, como hemos dicho, la distribución y densidad de formación de canales de conducción en la plata dan lugar a valores aleatorios de tensión umbral para la que el memristor comienza a conducir. La resistencia en serie de orden de MΩ sirve para limitar la corriente a un valor lo bastante pequeño para asegurar la supervivencia del dispositivo.

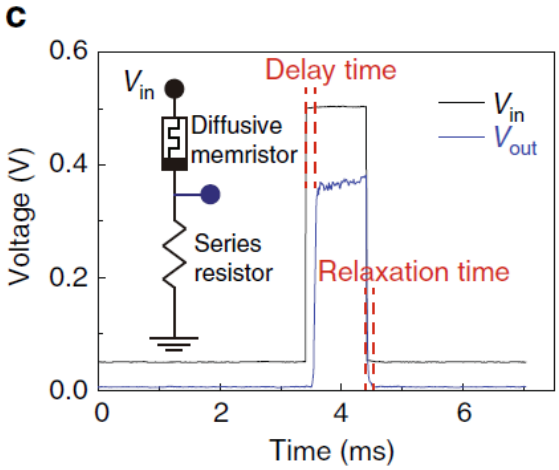


Ilustración 39: retraso en un solo ciclo. Jiang, et al. 2017

Veamos el sencillo circuito propuesto que emplea este efecto para funcionar como un TRNG:

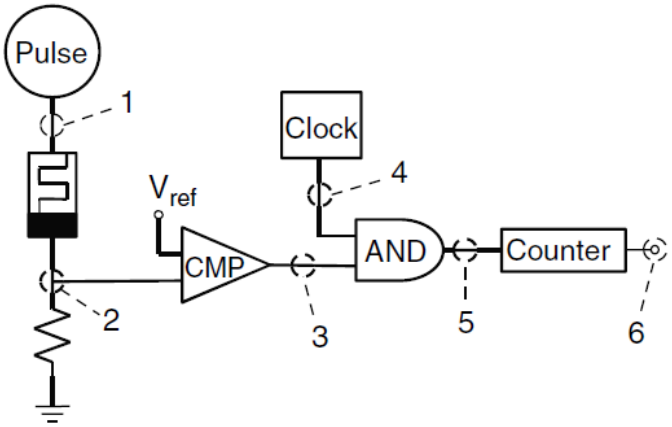


Ilustración 40: circuito TRNG basado en memristor Ag-SiO2. Jiang, et al. 2017

La fuente de señal de tensión "Pulse" general ondas cuadradas de amplitud 0.5 V y ancho de pulso de 300 microsegundos. La frecuencia de esta señal es de 1 KHz. Mientras la señal está a cero, el memristor se encuentra en off (corriente despreciable) por lo que la tensión en el punto (2) respecto a masa es también despreciable. La salida del comparador por tanto es 0 ya que la tensión de referencia es positiva y mayor que 0, lo que implica que la puerta AND dará un 0 en la salida. Cuando llega la onda cuadrada a 1, el memristor comenzará a conducir pero con un retardo de valor diferente en cada ciclo, por lo que la tensión en (2) aumentará de forma proporcional y en fase con dicha corriente del memristor. En el instante en que la tensión 2 supere la tensión de referencia, el comparador da un 1 en la salida (3). Entonces la puerta AND, al tener un 1 fijo en una de sus entradas, transmitirá el valor de la otra, que es un reloj o tren de pulsos cuadrados (4) normalizados entre valores 0 y 1 lógicos. Esa salida de pulsos se lleva a un contador de flanco ascendente, que conmuta el valor lógico de su salida de 0 a 1 cuando detecta un flanco de subida en su entrada (5). Cuando la onda cuadrada de entrada (1) termina, se almacena el último valor del bit de salida del contador (6) y hasta que llegue la próxima onda dicho bit se almacena en un microprocesador. Así, si queremos generar un número aleatorio binario de 1 Kbyte de tamaño, tendremos que ejecutar este proceso durante 8000 ciclos, es decir durante 8 segundos. El carácter estocástico radica en que, ya que el tiempo de retardo en la conmutación es intrínsecamente aleatorio, también lo es el número de ciclos de reloj que se envían al contador, y por tanto es impredecible el número de veces que el bit de salida cambia de valor y en qué valor va a terminar en cada ciclo (que será el valor del que partirá en el siguiente ciclo, disminuyendo aún más el determinismo del sistema).

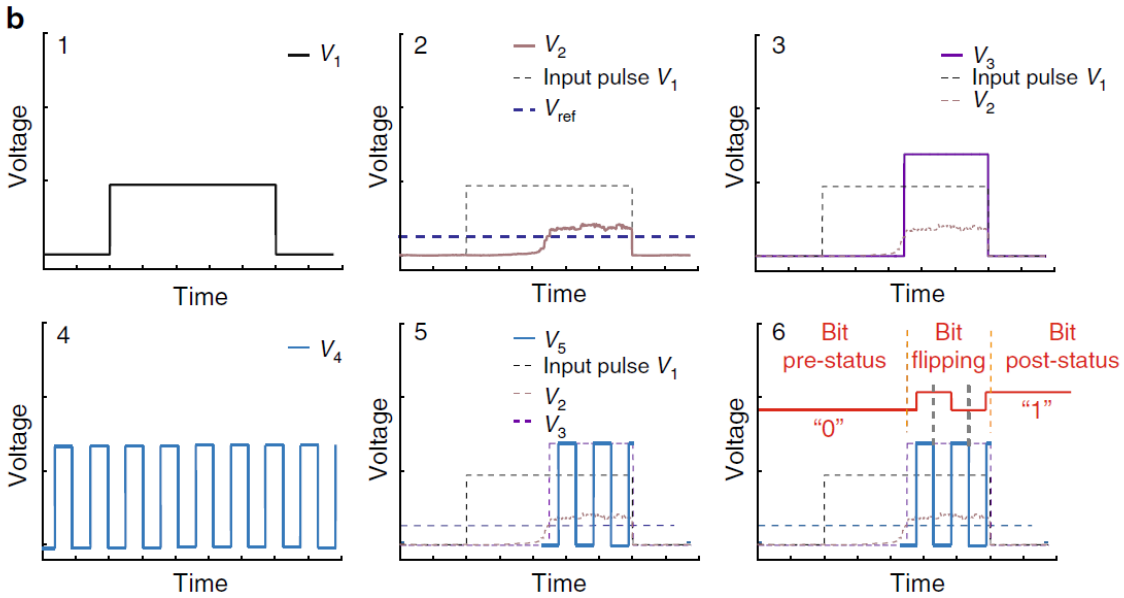


Ilustración 41: gráficas del funcionamiento del TRNG. Jiang, et al. 2017

El ritmo de generación de bits aleatorios se puede incrementar fácilmente usando un contador multi-bit. Los autores emplearon este circuito para generar 76 millones de bits las cuales almacenaron en un microcontrolador en 76 secuencias diferentes, por tanto cada una de 1 millón de bites de tamaño. Aplicando a dichas secuencias los 15 tests de aleatoriedad de NIST, observaron que todas superaron todos los test con éxito, por lo que se pueden considerar como verdaderamente aleatorias.

Este método, aunque ofrece una velocidad de generación moderada, es sencillo, escalable, funcional, y muy robusto ante cambios de temperatura o perturbaciones externas como radiación electromagnética (uno de los principales problemas de los TRNG basados en oscilación de electrones).

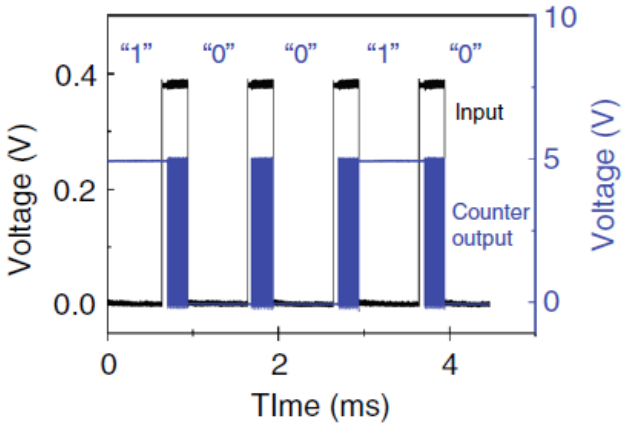


Ilustración 42: generación de 5 bits aleatorios. Jiang, et al. 2017

5.1.3 Encriptación mediante circuitos electrónicos memristivos

El interesante método de encriptación que nos disponemos a exponer fue diseñado y construido por Rahma y Muneam y presentado en su libro *Memristive non linear electronic circuits* (2019).

Partiremos del diagrama de bloques fundamental del circuito para explicar el algoritmo general, después su base matemática y por último sus detalles de implementación y funcionamiento electrónicos y resultados.

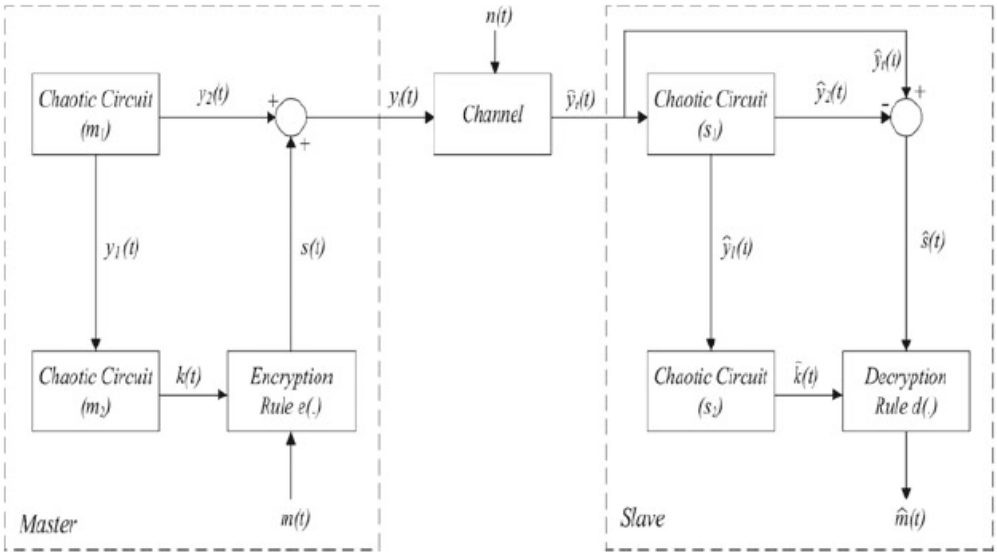


Ilustración 43: diagrama de bloques encriptador. Rahma y Muneam. 2019

Sea un mensaje o trama de información m(t). A dicha trama se le aplica una función de encriptación e(.), la cual emplea valores de una secuencia llave o *keystream*. Dicha función k(t) se genera mediante un sistema caótico m2, es decir está formada por valores totalmente aleatorios aunque acotados en un cierto rango, en función del tiempo. A su vez, este sistema m2 toma como valores de entrada los dados por la función y1 que ha sido generada por otro sistema caótico m1, aumentando así aún más la impredecibilidad de la secuencia llave k(t). Del sistema m1 obtenemos también la función aleatoria y2, que junto con la función obtenida de la encriptación s(t) forman la trama de información encriptada final, y que se enviará al canal de comunicación correspondiente para que llegue al receptor. Para modelar el posible ruido o incertidumbres en la transmisión, se puede aplicar una función ruido n(t) al mensaje.

En el lado del receptor, tenemos otros dos sistemas caóticos, s1 y s2, los cuales están sincronizados con sus pares m1 y m2 y tienen una correspondencia en su dinámica interna, parámetros y realimentación, tal que cuando a s1 le introducimos el mensaje y(t) este puede estimar (la aleatoriedad intrínseca del proceso genera incertidumbres inevitables) la función y2 que generó su par m1. Por tanto, si al mensaje recibido le restamos el mensaje estimado y2 (ya que en el emisor lo que hicimos fue sumar y2 con el mensaje cifrado) obtendremos la estimación del mensaje cifrado s(t). De igual manera que con el sistema s1 podemos estimar y2, también estimamos la secuencia y1. Ésta alimentará el sistema caótico s2 que gracias a la sincronización y a la mencionada similitud con el sistema caótico m2, estimará la secuencia llave k(t). Por último, aplicando la inversa de la función de encriptación e(.), la cual denominamos d(.), a las funciones estimadas s(t) y k(t), obtenemos el mensaje original desencriptado con un cierto nivel de error debido a que por supuesto también se trata de una estimación. Tras un proceso de filtrado adecuado, siempre que los errores no sean muy grandes, obtendremos el mensaje m(t). Podríamos suponer en principio que este método funcionará mejor para señales digitales ya que son mucho más robustas ante el ruido y las incertidumbres que las analógicas. Con este método, un usuario o entidad que acceda sin autorización al mensaje no podrá "traducirlo" si no conoce la dinámica, métodos de integración y valores de parámetros de los cuatro sistemas caóticos implicados, además de la función de encriptación/desencriptación aplicada. Cuanto mayor sea la aleatoriedad real de los sistemas, y la frecuencia de cambio de los valores de los parámetros, más complicado será que se descifre la clave por mucha potencia computacional que tenga el intruso, o por muchas tramas o mensajes a las que consiga acceder para analizar.

Describimos a continuación el algoritmo de cifrado desde el punto de vista del funcionamiento matemático concreto de los sistemas.

-Encriptado del mensaje:

Se aplica una función que genera un nuevo valor a partir del valor del mensaje (x) y el valor caótico correspondiente de la secuencia llave (k). Es implícito por tanto que el número de bits de la llave debe coincidir con el número de bits del mensaje, o en caso de trabajar analógicamente, la frecuencia de muestreo de ambas señales debe ser la misma. Para un mensaje y llave acotados al intervalo (-h,h) tendremos:

$$f_1(x,k) = \begin{cases} (x+k)+2h, & -2h \leq (x+k) \leq -h \\ (x+k), & -h < (x+k) < h \\ (x+k)-2h, & h \leq (x+k) \leq 2h \end{cases}$$

Ilustración 44: función de encriptación. Rahma y Muneam. 2019

Al obtener en el receptor la estimación del mensaje s(t) que contiene por tanto los valores de f1(x,k) estimada, la desencriptación consiste en despejar de la misma ecuación mostrada en la ilustración 44 los valores de x. Conocemos los valores k de la llave gracias a la estimación del sistema s2.

-Sistemas caóticos y secuencias obtenidas:

$$m_1 : \begin{cases} \dot{x}_{1m}(t) = \sigma x_{2m} \\ \dot{x}_{2m}(t) = \beta h(x_{1m}, x_{4m}) + \alpha x_{3m} - \gamma x_{2m} \\ \dot{x}_{3m}(t) = 1 - \alpha h(x_{3m}, x_{5m}) - x_{1m} \\ \dot{x}_{4m}(t) = x_{1m} - x_{1m} x_{4m} - \delta x_{4m} \\ \dot{x}_{5m}(t) = x_{3m} - x_{3m} x_{5m} - \delta x_{5m} \\ y_1(t) = x_{1m} \\ y_2(t) = x_{2m} \\ y_t(t) = y_2 + s(t) \end{cases}$$

Ilustración 45: sistema caótico m1. Rahma y Muneam. 2019

Las relaciones dinámicas de m1 generan dos secuencias aleatorias y1 e y2, las cuales se obtienen por integración de dos variables internas cuyas derivadas temporales dependen en un caso del valor de la otra, y en el otro del valor de ella misma, de la otra, y de otras tres variables internas (explícita o implícitamente), además de ciertos parámetros cuyo valor es dado previamente. El sistema obtenido así es altamente no lineal, oscilante y poco determinista.

La secuencia y2 permanecerá a la espera hasta que se genere el mensaje cifrado para sumarse a él antes de ser enviado. La secuencia y1 entra al sistema m2:

$$m_2 : \begin{cases} \dot{z}_{1m}(t) = F\big((y_1(t))_{\tau_m}\big) - a_1 h(z_{1m}, z_{2m}) + l_2(z_{1m} - y_1(t)) \\ \dot{z}_{2m}(t) = z_{1m} - z_{1m} z_{2m} - a_2 z_{2m} \\ k(t) = z_{1m} \end{cases}$$

Ilustración 46: sistema caótico m2. Rahma y Muneam. 2019

De la cual, también integrando una variable interna, se obtiene la secuencia de valores de la llave, k(t).

Los sistemas caóticos del receptor (simétricos a los del emisor) s1 y s2 se muestran a continuación. Hay que destacar que el desenmarañamiento que llevan a cabo las ecuaciones del sistema s1 (o las de s2) para obtener de manera aproximada los valores de y1 (o de k(t)) a partir del mensaje encriptado y(t) es, matemáticamente, muy complejo, y no nos centraremos en ese aspecto al caer totalmente fuera de los objetivos del presente trabajo.

$$s_1 : \begin{cases} \dot{x}_{1s}(t) = \sigma x_{2s} + c_0(y_1 + n(t) - x_{1s}) \\ \dot{x}_{2s}(t) = \beta h(x_{1s}, x_{4s}) + \alpha x_{3s} - \gamma x_{2s} + c_0(y_t + n(t) - x_{2s}) \\ \dot{x}_{3s}(t) = 1 - \alpha h(x_{3s}, x_{5s}) - x_{1s} \\ \dot{x}_{4s}(t) = x_{1s} - x_{1s}x_{4s} - \delta x_{4s} \\ \dot{x}_{5s}(t) = x_{3s} - x_{3s}x_{5s} - \delta x_{5s} \\ \hat{y}_1(t) = x_{1s} \\ \hat{y}_2(t) = x_{2s} \end{cases}$$

$$s_2 : \begin{cases} \dot{z}_{1s} = F\left(\left(\hat{y}_1(t)\right)_{\tau_s}\right) - a_1 h(z_{1s}, z_{2s}) + l_2\left(z_{1m} - \hat{y}_1(t)\right) \\ \dot{z}_{2s} = z_{1s} - z_{1s}z_{2s} - a_2 z_{2s} \\ \hat{k} = z_{1s} \end{cases}$$

Ilustración 47: sistemas caóticos s1 y s2. Rahma y Muneam. 2019

Para entender el papel del memristor en este método, analizaremos la implementación física del sistema m2. Omitimos la implementación del resto de sistemas y del circuito encriptador por cuestiones de brevedad.

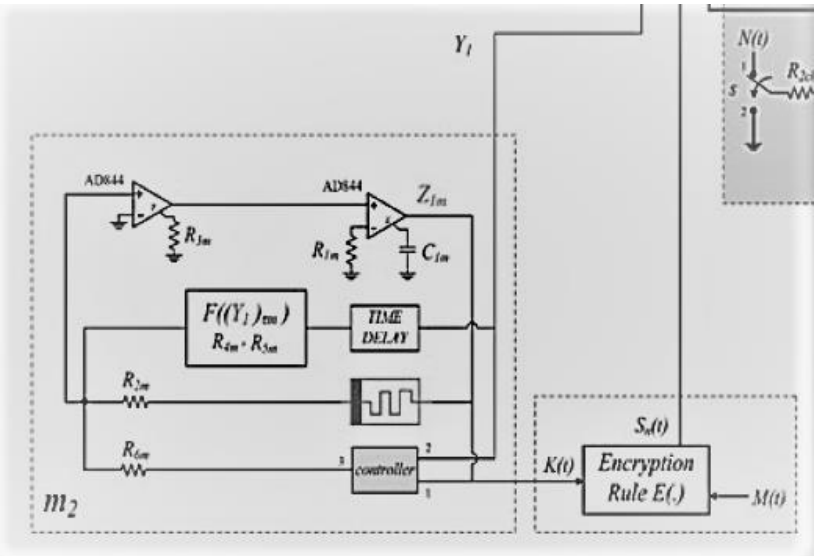


Ilustración 48: implementación del sistema m2. Rahma y Muneam. 2019

El bloque "controller" es un restador, del que sale el término K(t)-Y(t), que nombrado como variables internas del sistema es Z1m-X1m. Esta señal, se suma a otras dos en el primer operacional (la salida de la función F(y1) y la tensión entre el memristor y la resistencia R2m), las cuales forman la expresión de la derivada temporal de la secuencia llave k(t). Por tanto esta señal se lleva a un integrador (segundo operacional) del que se obtiene la función k(t). La realimentación y el comportamiento aleatorio llega aquí, ya que la tensión entre los terminales del memristor (la cual se lleva al sumador y por tanto su variación afectara a la salida k(t)) depende tanto de la misma señal k(t) (o Z1m) como de la variable interna del memristor la cual nombran como Z2m. Nótese que el memristor en serie con R2m forman un divisor de tensión, en uno de sus extremos tenemos masa debido al operacional del sumador, y en el otro la tensión de la señal k(t), por tanto la tensión entre el memristor y la resistencia, que es la que afectará a la salida del sumador, variará según varíe la memristancia, como si se tratara de un potenciómetro de ajuste. La relación entre esa tensión y sus variables es la h(z1m, z2m) y como hemos visto, es una función no lineal, con histéresis, y además presenta comportamiento estocástico a nivel de tensión de umbral y de derivada de la corriente con el tiempo especialmente cuando la tensión en la entrada varía en forma de escalón, lo cual ocurrirá si la información procesada es digital. Todo esto aporta la fluctuación aleatoria a la señal k(t), que cuanto mayor sea, producirá una señal cifrada más segura cuando se aplique k(t) al circuito encriptador.

$$\dot{Z}_{1m}(t) = \frac{1}{R_{1m}C_{1m}}\left(F\left((X_{1m})_{\tau_m}\right) - \frac{R_{3m}}{R_{2m}}h(Z_{1m}, Z_{2m}) + \frac{R_{3m}}{R_{6m}}(Z_{1m} - X_{1m})\right)$$

Ilustración 49: derivada de la secuencia llave k(t). Rahma y Muneam. 2019

Veamos los resultados obtenidos por los autores para varios tipos de mensajes y diferentes parámetros en el circuito de encriptado. Para los diferentes casos, vemos primero el mensaje original m(t), después la señal codificada transmitida y(t), y por último el mensaje desencriptado final (estimación de m(t)). Los dos primeros corresponden a la transmisión encriptada de información analógica, en concreto la señal producida por la onda sonora de una voz humana hablando. En los dos últimos, se trata de un mensaje binario de unos pocos bits.

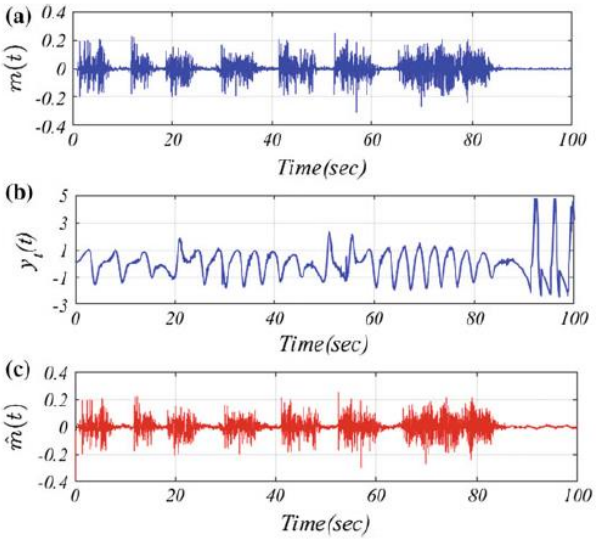


Ilustración 51: mensaje analógico sin ruido. Rahma y Muneam. 2019

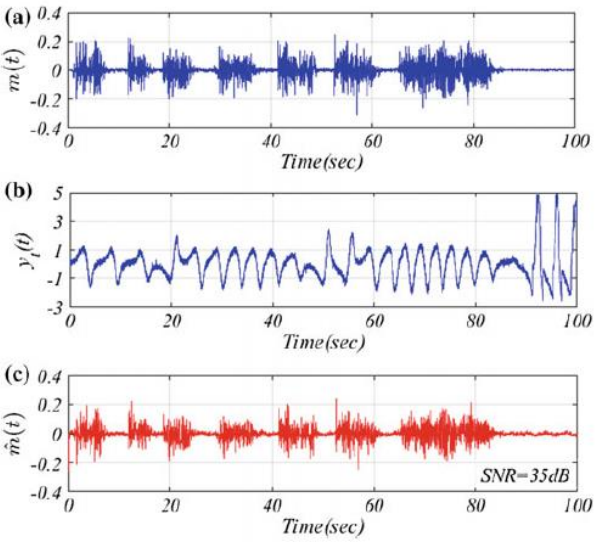


Ilustración 50: mensaje analógico con ruido. Rahma y Muneam. 2019

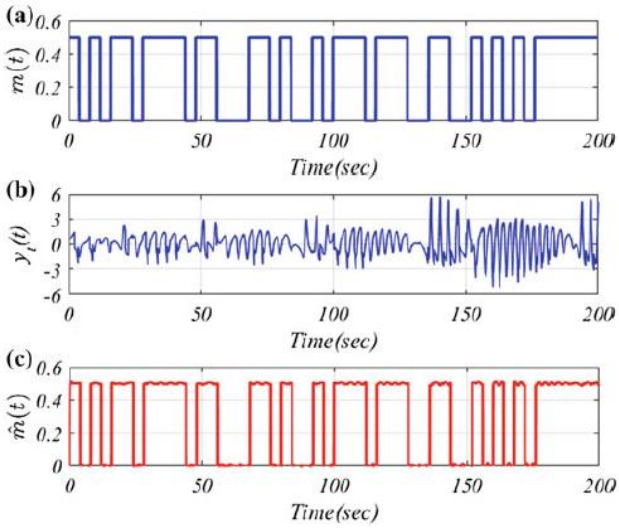


Ilustración 52: mensaje digital sin ruido. Rahma y Muneam. 2019

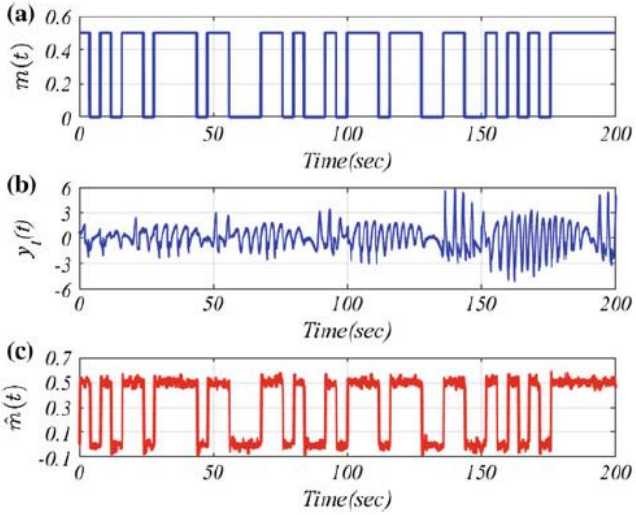


Ilustración 53: mensaje digital con ruido. Rahma y Muneam. 2019

Es arriesgado aventurarse a evaluar la calidad del mensaje desencriptado en el caso de señal analógica. Un análisis de Fourier, y aplicar la señal a un transductor que permita escuchar el mensaje, serían las mejores herramientas para hacerlo.

Sin embargo, en el caso de señal digital, la calidad y similitud entre el mensaje original y el obtenido tras el proceso de encriptado-transmión-desencriptado se ve a simple vista. Como era de esperar, cuando se añadió ruido a la señal, este se ve reflejado también en el mensaje obtenido. Pero al tratarse de información digital, no resultaría difícil conseguir una señal limpia como la original con un pequeño procesamiento posterior. Si el 1 lógico se ha asignado a una tensión de 0.5 voltios, y el 0 lógico a un valor de 0 voltios, se podría programar una función que convirtiera la señal en una binaria pura estableciendo un máximo en las oscilaciones del mensaje provocadas por el ruido:

$$f\left(\widehat{m}(t)\right) = 0.5\ V\ \ para\ \widehat{m}(t) \geq 0.4\ V$$

$$f\left(\widehat{m}(t)\right) = 0\ V\ \ para\ \widehat{m}(t) \leq 0.1\ V$$

Cuando el valor de la estimación de m(t) se encuentre entre 0.2 y 0.3 voltios, tan solo tendremos que indicar al programa que espere un pequeño tiempo, ya que sabemos que debe tratarse de un valor estacionario justo en el salto entre los dos valores lógicos posibles. Si la señal se mantiene en ese intervalo, sabremos que ha ocurrido un error en la transmisión o en el desencriptado.

5.2 Análisis y construcción de redes neuronales

Los sistemas de computación neuromórficos se consideran potenciales sustitutos de los sistemas tradicionales basados en la arquitectura de von Neumann para ciertas aplicaciones. El gigantesco aumento del volumen de datos disponibles para analizar y procesar en casi cualquier aplicación científica, tecnológica y empresarial que ha llegado con la industria 4.0 induce una necesidad de procesar estos datos en tiempo real y con la mayor eficacia de potencia posible. En este sentido, los sistemas de tipo von Neumann suelen presentar limitaciones debido a la gran diferencia de calidad y velocidad de funcionamiento entre los procesadores y las memorias que emplean. Los sistemas neuromórficos ofrecen la posibilidad de eliminar esa barrera ya que estos sistemas no diferencian entre zonas de procesamiento y zonas de almacenamiento. Esto es una de las muchas similitudes con los cerebros de los mamíferos a las que los adjetivos "neuromórfico" o "neuronal" hacen referencia. Dichas estructuras biológicas inspiran el paradigma de la arquitectura de estos nuevos sistemas de computación, los cuales están aún en los primeros compases de su existencia, quedando aún un larguísimo camino de resolución de obstáculos, mejoras, y optimización para poder llegar a ser una opción viable de hardware para implementar las tecnologías informáticas que articulan el mundo.

Es importante señalar que este campo incluye temas tan complejos y novedosos como son la inteligencia artificial, el aprendizaje automático y el diseño de circuitos analógicos neuromórficos; campos para los cuales es necesaria una formación muy especializada y que por tanto no tratamos de abarcar en este trabajo. Sin embargo, podemos ilustrar con algunos ejemplos y de forma general el peso que tienen los memristores en el desarrollo de estas tecnologías. En (Hassan, Liu, Yang, Li y Chen, 2019) encontramos un novedoso método de implementación de aprendizaje automático en una red neuronal basada en una estructura *crossbar* de memristores. Este método es una simplificación del llamado "backpropagation algorithm" que consigue ser más efectivo y más corto en cuanto a pasos en el método.

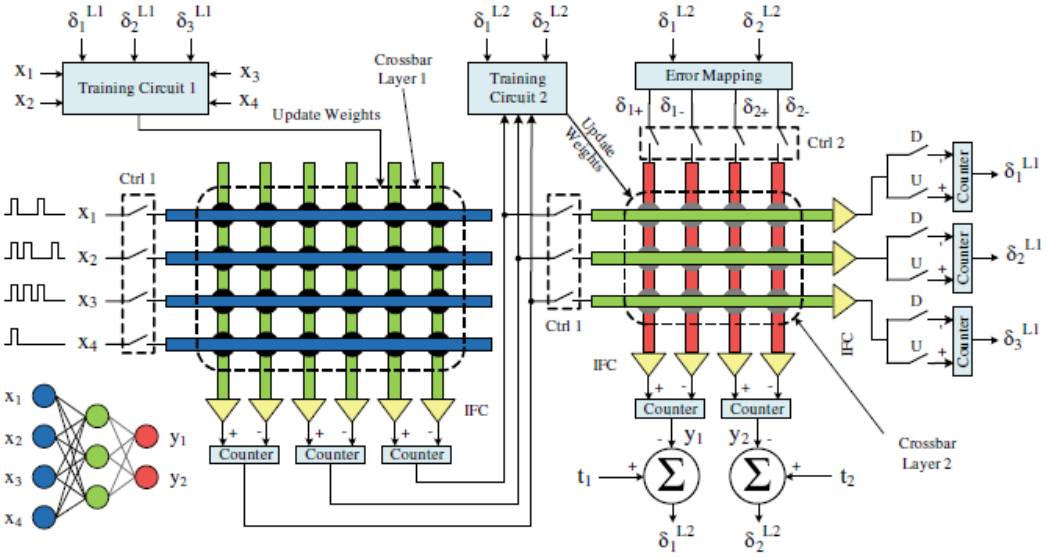


Ilustración 54: red neuronal basada en memristores. Hassan, Liu, et al. 2019

Aunque lo importante del trabajo es el nuevo enfoque al aplicar el método de aprendizaje a la red, nos centramos en el carácter ilustrativo de la filosofía general de la construcción de redes neuronales memristivas: El comportamiento de la conductancia del memristor y los niveles de tensión a la que esto ocurre resultan ser casi idénticas al comportamiento de la conductancia en las sinapsis neuronales biológicas. Así, cada memristor en el array representa una conexión entre neuronas. Más concretamente, cada memristor y su conductancia (resistencia) representan el llamado peso sináptico de la conexión entre dos de las neuronas de la red. En este caso, es una red sencilla con cuatro neuronas de entrada, dos neuronas de salida y una única capa oculta o de implementación lógica que consta de tres neuronas. Así, los dos arrays representan las dos familias de conexiones sinápticas entre la capa oculta y las capas de entrada/salida. El proceso de aprendizaje o entrenamiento consiste en la parametrización y actualización de los valores de peso sináptico, es decir, los valores de conductancia de los memristores, los cuales terminan generando una distribución en la probabilidad de establecer una conexión antes que otras. Esta distribución de pesos sinápticos es precisamente la lógica de la red, que genera las salidas en función de las señales de entrada.

Cuanto mejor sea el proceso de entrenamiento y menor sea la variabilidad en las señales de entrada, mejor será la respuesta obtenida en las salidas. Las aplicaciones de estas redes y métodos de aprendizaje van desde la creación de inteligencias artificiales y softwares que se autodesarrollan, hasta el reconocimiento de patrones digitales para visión artificial y ciberseguridad.

Destacamos también el dispositivo llamado coloquialmente "brain-on-a-chip" que muy recientemente han diseñado y fabricado en el MIT, uno de los centros mundiales de desarrollo tecnológico, y cuyos primeros resultados fueron publicados en la revista "Nature nanotechnology" (Yeon, et al., 2020).

En dicho trabajo los autores han construido una red neuromórfica con la típica estructura de *crossbar* apilados, y con memristores fabricados con diferentes aleaciones de plata, cobre y silicio. Emplearon este dispositivo tanto para guardar información, como para procesarla. En concreto, imágenes en forma de matriz tanto en escala de grises (1 byte por pixel) como de tipo RGB (3 bytes por pixel). El resultado de cargar o guardar imágenes en las redes construidas se presenta en la siguiente ilustración:

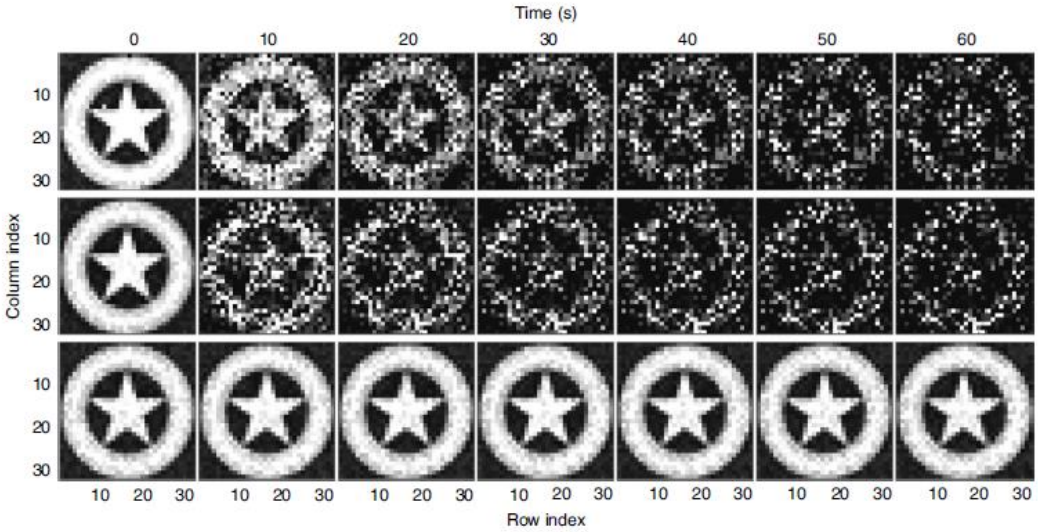


Ilustración 55: escritura y retención de datos en red neuronal memristiva. Yeon, et al. 2020

El mejor, o más bien, el único caso de retención de información efectiva corresponde a la red construida con memristores de plata y cobre (tercera fila de la ilustración 55). En cuanto al procesamiento de datos, la red se entrena para aplicar diferentes filtros de tipo matriz de convolución o kernel a otra imagen almacenada también en la red. Los resultados se muestran en las ilustraciones 56 y 57:

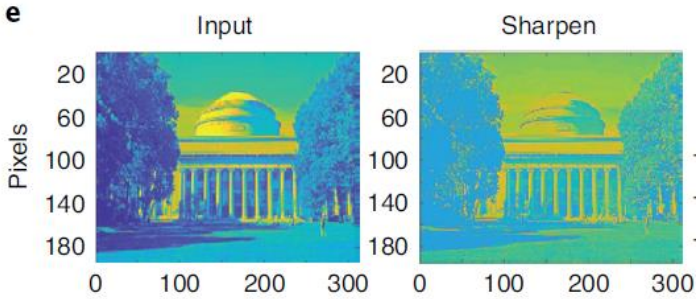


Ilustración 56: procesamiento de imágenes en red neuronal memristiva 1. Yeong, et al. 2020

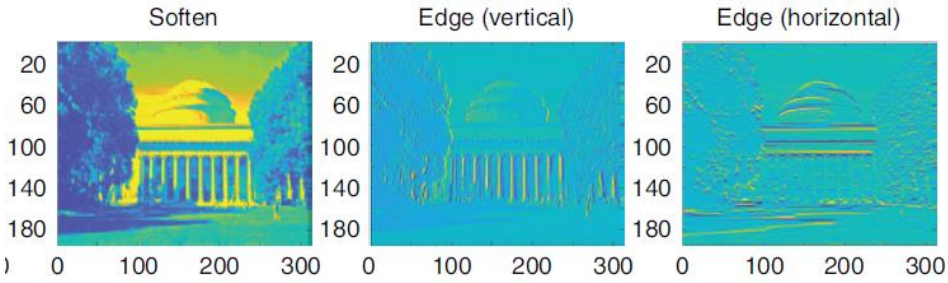


Ilustración 57: procesamiento de imágenes en red neuronal memristiva 2. Yeong, et al. 2020

Estos resultados son muy prometedores y abren un nuevo y gran margen de desarrollo de dispositivos neuromórficos cada vez más eficaces, pequeños, y multifuncionales. También alimentan la posibilidad de la implementación efectiva y relativamente sencilla de inteligencia artificial y de tareas computacionales que hoy en día son solo posibles en grandes y caros superordenadores cuyo software requiere mucho tiempo y esfuerzo para desarrollarse y actualizarse.

5.3 Memorias no volátiles

5.3.1 Contexto y conceptos generales

Tanto a nivel usuario como industrial, la memoria es un elemento fundamental en cualquier dispositivo electrónico o informático. Podemos hablar en general de 5 tipos de memoria, aunque hay diferentes clasificaciones si tomamos como parámetro la capacidad máxima, o la velocidad de acceso.

-Memoria RAM: almacenamiento temporal de información, tanto datos como instrucciones. Es una memoria de lectura y escritura. Sus siglas vienen de *Random Access Memory*, es por tanto una memoria que no necesita seguir un orden concreto para acceder a uno de sus espacios o bytes de memoria, lo cual es una ventaja a la hora de minimizar el tiempo de acceso a la información que almacena. Es una memoria volátil, es decir, los datos se pierden si se deja de alimentar el circuito de memoria (apagar el dispositivo, corte inesperado en la corriente…)

-Memoria ROM: siglas de *Read Only Memory,* memoria de solo lectura. Es decir, los datos que almacene en general no se pueden modificar. Está incorporada en la placa base de los ordenadores y contiene instrucciones necesarias para arrancar y ejecutar el *Basic Input Output System* (o BIOS) del sistema. Las memorias ROM son de tipo secuencial, por tanto los datos que almacena solo se pueden leer en un orden concreto y definido.

-Memoria PROM: programable de solo lectura. Son memorias digitales basadas en fusibles, por lo que la información es invariante una vez se ha escrito.

-Memoria caché: una memoria de apoyo para el funcionamiento del computador. Se encuentra a medio camino en cuanto a velocidad y capacidad entre los registros del microprocesador y el resto de memorias de la máquina.

-Memoria virtual: semejante a la caché. Sirve de apoyo al sistema operativo. Su velocidad de acceso es muy baja, pero sigue presente en sistemas Windows y Linux.

-Memoria flash: también llamadas EEPROM. Permiten ser escritas o leídas en varias posiciones durante la misma operación. Las memorias extraíbles USB o microSD son memorias flash.

Aunque la anterior clasificación es importante, tiene un enfoque más informático que ingenieril. Veamos como se dividen las memorias si nos basamos en la tecnología empleada, ya que esto nos dará una visión más técnica desde el punto de vista del hardware:

-Memoria de semiconductor: emplean circuitos integrados basados en transistores. Actualmente la memoria principal de las máquinas consiste en memoria de semiconductor volátil tipo RAM. También existen memorias de semiconductor no volátiles, como las memorias flash.

-Memoria magnética: emplean patrones de magnetización para almacenar la información sobre una capa magnetizada. Son no volátiles y de acceso secuencial. De este tipo son los discos duros.

-Memoria de disco óptico: basados en la reflexión de luz de diodos láser. Son no volátiles y de acceso secuencial. Los DVD, CD-CDROM y Blue-Ray son memorias de disco óptico.

-Memoria de disco magneto-óptico: la información se almacena codificando el estado magnético de una superficie, pero se lee ópticamente. Son no volátiles y de acceso secuencial.

-Nuevos métodos propuestos: nuevas tecnologías que generan nuevos tipos de memoria con características mejoradas se desarrollan actualmente, como la memoria de cambio de fase, la memoria holográfica o la molecular. El objetivo de los desarrolladores e investigadores es conseguir mejorar la capacidad, velocidad, durabilidad y coste de fabricación para poder competir con las anteriores, las cuales están muy desarrolladas y llevan muchas décadas en el mercado.

5.3.2 Funcionamiento básico. El problema de los caminos furtivos

La clave sobre la que se construyen estas nuevas memorias se encuentra en la capacidad de variar la resistencia de un memristor tan solo controlando el tiempo que se le aplica una pequeña tensión, cuya polaridad determina si aumentamos o disminuimos dicha resistencia. Un bit se podrá así almacenar en un memristor haciendo corresponder un valor de resistencia máxima Roff con un 0 lógico, y un valor de

resistencia mínima Ron con un 1 lógico. Estos dos valores extremos de resistencia se podrá "leer" aplicando al memristor una tensión por debajo de su tensión umbral, es decir, una tensión tan pequeña que no sea capaz de desplazar el umbral entre las dos zonas con diferente conductividad (concentración de vacantes), y que por tanto no varíe la resistencia que previamente se ha "escrito".

Se construye así un array o matriz de almacenamiento de bits, con una estructura de *crossbar* que puede convertirse en una célula de almacenamiento apilando dichos arrays uno encima de otro.

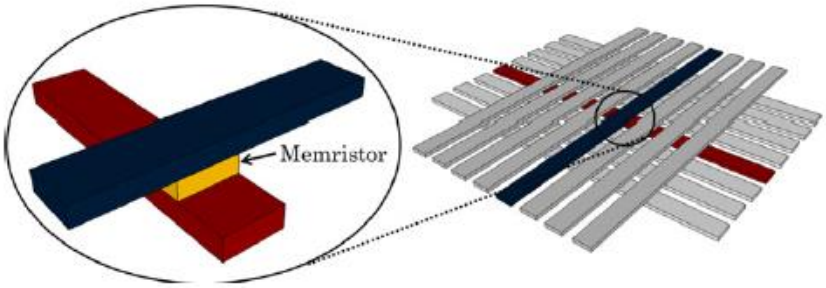


Ilustración 58: crossbar de memoria con memristores. Zidan, Fahmy, et al. 2012

Estas memorias poseen, a priori, unas propiedades de tamaño y densidad excelentes en comparación con otras tecnologías de memoria Flash. Pero además, la posibilidad de comprimir información es muy interesante: supongamos que somos capaces de controlar la tensión de escritura y su tiempo de aplicación con la suficiente precisión. Si además la fabricación del memristor es óptima en el sentido de minimizar impurezas, maximizar la homogeneidad de propiedades y en general eliminar perturbaciones que afecten al comportamiento según el modelo matemático adecuado de su variable intrínseca, entonces no hay ninguna razón para limitarnos a usar los estados Ron y Roff para almacenar información. Si, por ejemplo, tomamos otros dos valores de resistencia intermedios como valores válidos de estado de memoria, entonces tendremos la capacidad de almacenar cuatro estados lógicos en una sola célula o unidad básica. Esto equivale a que en una sola unidad básica podemos almacenar el equivalente a dos bits.

Evidentemente, estos valores serán en realidad rangos de valores de resistencia válidos, que se traducirán en rangos de tensión de lectura que se interpretarán siempre con el mismo valor digital. En la bibliografía existente se proponen siempre unos rangos intermedios o *buffers* que ayuden a separar lo suficiente los distintos rangos lógicos de manera que no haya posibilidad de error o de superposición al leer analógicamente la unidad.

Sin embargo, un rápido vistazo a la arquitectura *crossbar* nos descubrirá uno de los principales problemas que se han tenido que solventar para poder construir y comercializar memorias basadas en memristores: los llamados *sneak paths* o caminos furtivos. Cuando queremos escribir o leer en una unidad (un memristor) de la matriz, tendremos que aplicar la tensión adecuada entre los conductores en los que se encuentra dicha unidad (fila y columna de la matriz). Sin embargo, al no haber ninguna ruta concreta definida, la corriente podrá circular por varios caminos además del deseado, y como ya hemos visto en los modelos matemáticos es la corriente y el tiempo que se aplica la que controla la variación de la memristancia. Por tanto, si tenemos otros caminos por los que fluye corriente, no se cumplirá el modelo matemático que conocemos y que representa un único memristor atravesado por la corriente que pide a la fuente de tensión aplicada.

Tendremos un circuito equivalente tal que la resistencia del memristor en el que queremos leer o escribir tiene en realidad una resistencia en paralelo equivalente a las resistencias de los caminos posibles y no deseados que puede tomar la corriente. Esa resistencia equivalente dependerá por tanto del estado de memoria almacenado en los memristores que están en esos caminos furtivos, haciendo por tanto casi imposible hacer funcionar correctamente esta arquitectura. Los caminos furtivos deben suprimirse o eliminarse de alguna forma.

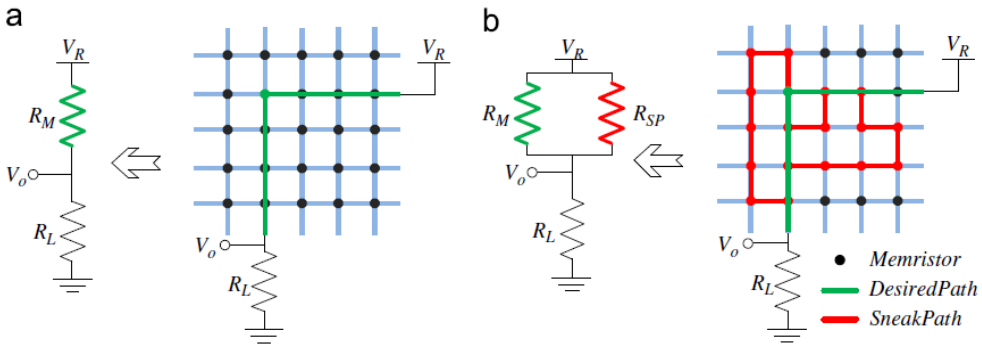


Ilustración 59: funcionamiento ideal (a) y efecto de los caminos furtivos (b). Zidan, Fahmy, et al. 2012

Hay varias propuestas de solución a los caminos furtivos en la bibliografía disponible, tales como encaminamiento con diodos, o el uso de memristores de tres terminales o memistores. Nos centraremos en el siguiente punto en dos soluciones concretas que nos parecen las más completas y eficientes.

5.3.3 Arquitecturas propuestas con solución a los caminos furtivos.

Primero analizaremos un modelo más general de memoria basada en un *crossbar* de memristores de N niveles o estados de memoria propuesto por Merkel, Nagpal y otros, en la publicación "Proceedings of international joint conference on neural networks" (2011). Posteriormente veremos una arquitectura más desarrollada que, siguiendo las líneas del anterior trabajo citado, presenta resultados más concretos y prácticos. Dicha arquitectura fue presentada en la prestigiosa revista "IEEE Transactions on nanotechnology" con la autoría de Yilmaz y Mazumder (2017).

En el primer modelo, los autores proponen una original ecuación para evaluar y comparar la calidad de una RRAM (non-volatile memristor-based resistive random access memory) en función de distintos parámetros:

$$\gamma = N * \frac{t_{ret}}{t_{Wavg}} * \left(\frac{2^{\beta}}{2N} + S_{NM}S_B\right) \quad [41]$$

"t ret" es el tiempo de retención de la memoria, que en el caso de los memristores está en el orden de las décadas.

"t Wavg" es el tiempo medio de escritura.

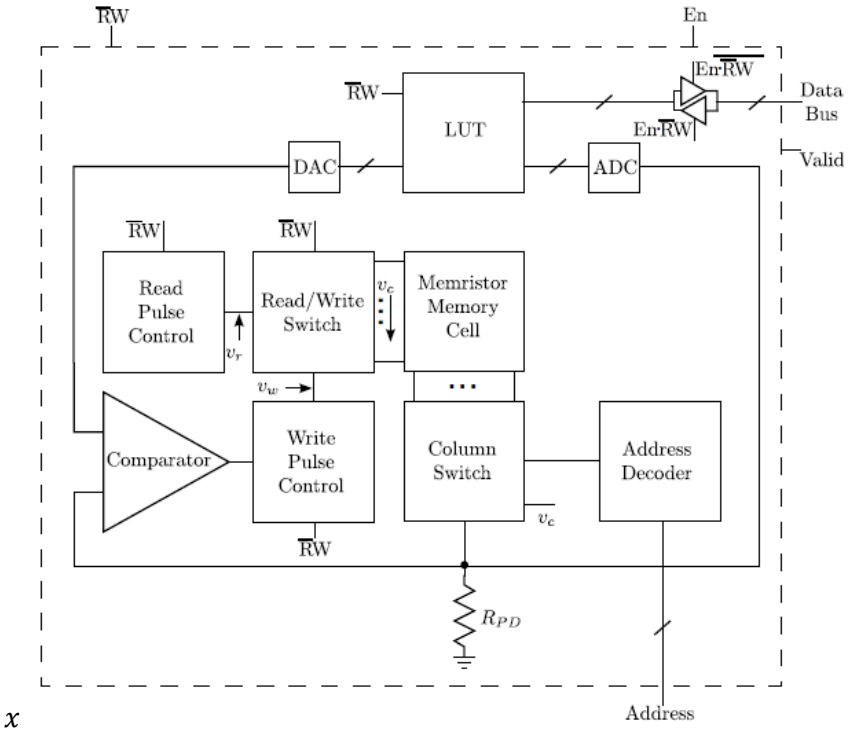
"N" es el número de estados lógicos de memoria asociados a un rango de resistencia del memristor.

"S nm" es el margen de ruido, la anchura por encima y por debajo del valor central de dicho valor.

"S b" es el ancho de buffer, la zona de separación entre los distintos rangos de valores de memoria y que corresponde con valores no válidos de memoria.

"β" es el número de bits del convertidor analógico-digital que convierte el valor de tensión correspondiente al estado de la memoria a un valor lógico.

Recordemos como último paso preliminar las ecuaciones que muestran de manera sencilla y geométrica (posición del umbral entre comportamiento rectificador y comportamiento de conducción con histéresis) la relación entre tensión, corriente y resistencia en el memristor, las ecuaciones 36 y 37, ya que son las mismas que emplean los autores para desarrollar este modelo de célula de memoria que mostramos a continuación:



x

Ilustración 60: arquitectura de RRAM multinivel. Yilmaz y Mazumder. 2017

Como ya hemos explicado, una mayor densidad de memoria se puede conseguir aumentando el número de niveles N, con la inevitable consecuencia de que se disminuye así la fiabilidad de la memoria al reducir los márgenes de ruido y de separación entre los niveles de memoria. Sin embargo en este modelo suponemos que se ha buscado el equilibrio adecuado para la aplicación concreta, centrándonos en la circuitería y esquema de bloques de control para la memoria.

El núcleo de esta célula de memoria es una *crossbar* de nanomemristores de N niveles de memoria, situados en los cruces entre M filas y C columnas. Los caminos furtivos se evitan de la siguiente manera:

Cuando se quiere leer o escribir en un memristor de una columna i-ésima, se conecta dicha columna a masa, mientras que al resto de columnas se aplica la tensión de lectura o escritura. Por otro lado, en todas las filas se aplica dicha tensión de lectura o escritura, por tanto los únicos memristores que tendrán aplicada una diferencia de potencial serán los de la columna en masa, y por tanto solo se escribirá o leerá en dichos memristores. Este método tiene un problema evidente, y es que en realidad estaremos leyendo la resistencia equivalente a todos los memristores en paralelo que haya a lo largo de la columna seleccionada. Por lo tanto será necesario un trabajo previo de definición de estados de memoria considerando la combinación de M memristores para representar de manera inequívoca y fiable N valores de memoria asociados a sus resistencias equivalentes.

Lectura:

La señal de control RW habilita la escritura en 0. La dirección de memoria se introduce desde el bus de datos, determinando así qué columna se quiere leer. El bloque "Column switch" conectará dicha columna a masa a través de la resistencia Rpd, que se empleará como divisor de tensión para determinar el estado de memoria. Este bloque aplicará al resto de columnas la misma tensión Vc que se aplicará a las filas durante el mismo tiempo. El bloque "Read/Write switch" aplica esta tensión Vc a todas las filas de la memoria, por lo que como ya explicamos solo habrá una diferencia de tensión neta en la columna seleccionada. Esta tensión es lo bastante pequeña y se aplica durante el tiempo adecuado como para asegurar que no se modifique el estado de los memristores de la columna seleccionada. Así, la tensión entre los extremos de la resistencia Rpd es directamente proporcional al valor de la resistencia equivalente entre los memristores de la columna, que es precisamente el estado de memoria almacenado que queremos detectar. Dicha tensión se lleva a un convertidor ADC donde se convierte en la señal digital que podrá ser empleada por la aplicación que usa esta memoria como información útil, siempre y cuando el bloque LUT confirme que ese valor digital corresponde con un estado válido de memoria, en cuyo caso el bit de validación se pone a 1. Si los memristores en paralelo presentan un valor de resistencia fuera de rangos (es decir, en la zona de buffer) entonces el bit de validación valdrá 0 y la LUT no transmite el valor digital obtenido de la conversión.

Escritura:

El algoritmo de escritura es muy similar al de lectura. La señal RW en este caso vale 1, y la tensión que se aplica ahora a las filas es Vw. Lo más interesante es el método para asegurar que ha habido éxito en la escritura: el bloque LUT transmite el valor digital que se introduce para escribir en la memoria a un convertidor digital-analógico. Esta señal analógica de tensión se lleva a un comparador junto con la tensión que se lee simultáneamente mientras se escribe (divisor de tensión Rp). Cuando el comparador detecta que ambas tensiones son iguales, se interpreta que la escritura se ha completado correctamente y se detiene la aplicación de Vw.

Veamos a continuación el segundo modelo presentado al comienzo de esta sección, para analizar las mejoras y detalles de implementación que lo hacen más interesante y realizable:

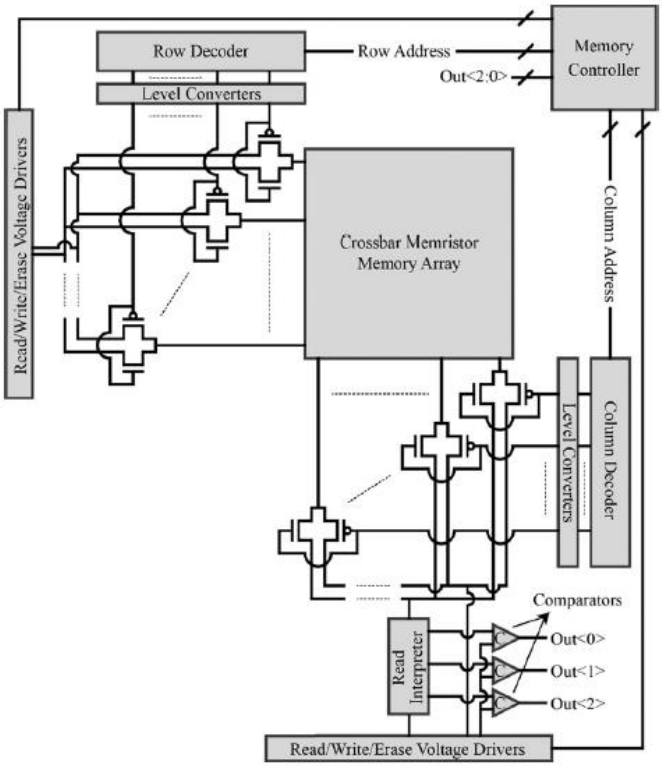


Ilustración 61: arquitectura de RRAM multinivel. Yilmaz y Mazumder. 2017

En este caso, la selección del elemento de memoria es exacta, y tanto la tensión de lectura/escritura como la conexión al divisor de tensión desde el que se lee el estado de memoria se aplican a la fila y a la columna deseada respectivamente, consiguiendo así acceder al memristor adecuado de forma aislada. Esta selección en filas y columnas se consigue de una forma relativamente sencilla con dos transistores con puerta común por cada entrada al array. Esto significa que, por ejemplo, para un array de memoria "plano" de M unidades básicas (M memristores, $\sqrt{M}$ filas y $\sqrt{M}$ columnas) se necesitarán 4*$\sqrt{M}$ transistores. Así, además de eliminar los caminos furtivos de corriente se trabaja con la información en cada memristor por separado, por lo que la correspondencia entre valor de resistencia y estado de memoria vuelve a ser individual, sencilla e intuitiva, y solo varia en cuanto a número de niveles de memoria o cantidad de bits que queramos almacenar en cada memristor. Aunque podemos considerarlas despreciables al analizar el funcionamiento general, a la hora de implementar esta arquitectura tendremos que controlar las corrientes de corte y la potencia disipada por los transistores y asegurarnos de que no influyan negativamente.

El bloque "Memory controller" genera las direcciones de memoria que controlan los estados de los transistores, y las señales que controlan los drivers de voltaje de lectura/escritura de filas y columnas. Además, recibe la información que genera el bloque "Read interpreter" para informar al usuario o al bloque ante el que responde sobre el éxito o fracaso de la operación de escritura, o el valor leído en el caso de operación de lectura. Analizamos por último el funcionamiento de este bloque, el interpretador de lectura:

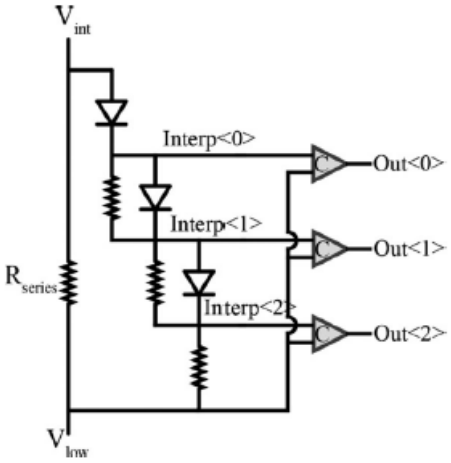


Ilustración 62: bloque "read interpreter". Yilmaz y Mazumder. 2017

Para interpretar correctamente este bloque, hay que destacar que los autores han desarrollado este modelo para una unidad básica de memoria que consiste en un memristor en el que se definen cuatro valores de resistencia con un cierto margen de ruido (rangos de resistencia) que corresponden a cuatro estados de memoria útiles, es decir que cada unidad básica almacenará la información equivalente a 2 bits. El bloque interpretador es básicamente un divisor de tensión cuyos voltajes entre los extremos son el de la salida de los transistores "abiertos" (en saturación) que depende directamente del valor de memoria de la unidad que se ha seleccionado, y el proporcionado por el driver de lectura/escritura llamado Vlow. Esta diferencia de potencial se divide en 3 con 3 resistencias y un diodo en paralelo con cada una, y se lleva a 3 comparadores en los que la segunda entrada es siempre la tensión inferior proporcionada por el driver, que será siempre constante. Por tanto, con una correspondencia que podemos ajustar con los valores y polaridades de las tensiones de lectura, escritura y la Vlow, cada vez que uno de los comparadores cambie de estado significará que una de las tensiones intermedias del divisor de tensión ha superado o caído respecto a Vlow, y que nos encontramos en un nivel de memoria de los cuatro posibles. Cuando nos encontremos en el estado de memoria de menor resistencia, la tensión Vint será máxima y no conmutará ninguno de los comparadores. Así, las tres señales Out<0,1,2> permiten al controlador conocer el estado de memoria leído, así como indicar que una escritura se ha producido correctamente.

Este funcionamiento se aprecia en estas figuras, que reflejan la operación de escritura de un estado de memoria '11' que corresponde por mero convenio al valor de máxima resistencia del memristor:

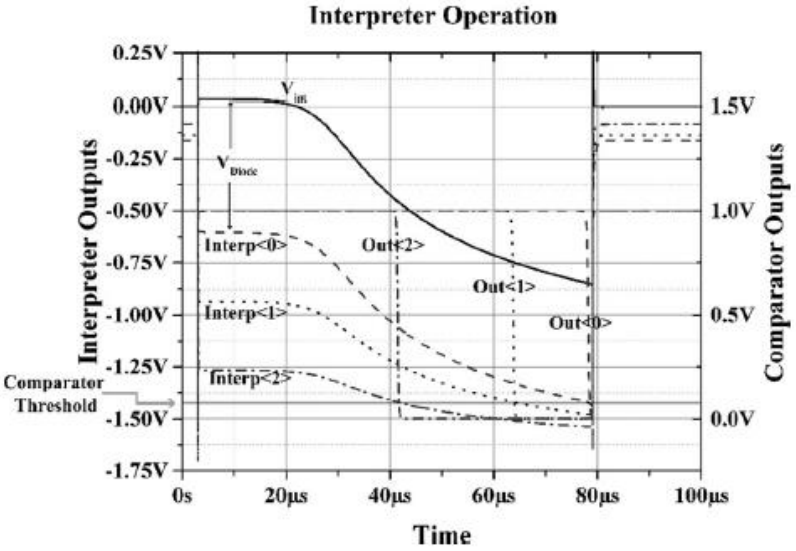


Ilustración 63: salidas del bloque interpretador. Yilmaz y Mazumder. 2017

La tensión umbral de los comparadores es Vlow. El valor de Vwrite y la tensión umbral del primer diodo hacen que las tensiones Interp<i> sean siempre negativas pero mayores que Vlow hasta que Vint comienza a caer por el aumento de resistencia del memristor. Entonces el divisor de tensión y los diodos hacen que los comparadores se activen en los instantes en los que se alcanza la tensión Vint que corresponde a un valor de resistencia en el memristor asignado a un estado de memoria.

Vemos la evolución del estado del memristor junto con la tensión Vint que se alcanza para estos cuatro valores de memoria asignados en este modelo:

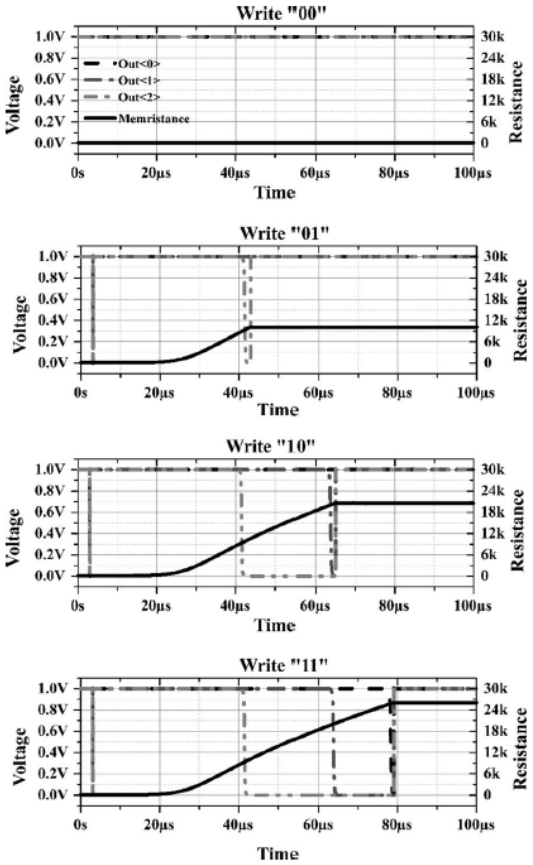


Ilustración 64: distintas operaciones de escritura. Yilmaz y Mazumder. 2017

Se ha elegido aplicar siempre la misma tensión de escritura, de manera que los diferentes estados se alcanzan a medida que aumenta el tiempo que se aplica dicha tensión. Estos tiempos que observamos (el máximo corresponde obviamente con el de máxima resistencia partiendo de un estado de mínima) también nos dan una idea del orden de los tiempos de escritura en esta arquitectura.

Para comprobar la fiabilidad y la desviación de los márgenes de ruido en cada nivel de memoria de los memristores, además de la homogeneidad de las propiedades de éstos, los autores escriben el mismo valor en todo el array procediendo después representar la distribución de valores de resistencia de todos los memristores, obteniendo los siguientes resultados:

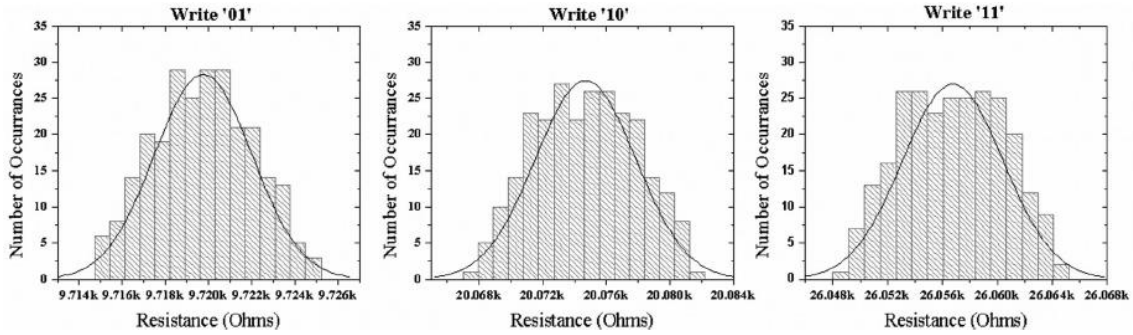


Ilustración 65: distribuciones de resistencia para cada estado. Yilmaz y Mazumder. 2017

Vemos que las distribuciones son normales y con márgenes bastante pequeños y del orden de unas pocas decenas de ohmios. Esto abre la puerta a posibles aumentos de número de estados o de niveles almacenados en cada memristor, es decir aumento de la densidad de memoria de esta arquitectura.

6 Problemas inherentes al memristor de HP

Como hemos visto en este trabajo, existe una gran cantidad de bibliografía disponible sobre aplicaciones y prototipos prácticos de memristores. Estos dispositivos se han estudiado y empleado (o estudiado su empleabilidad) en las mejores universidades y centros tecnológicos del mundo, y sus bases matemáticas y teóricas son fuertes. Cabe preguntarse entonces: ¿por qué los memristores aún no se han comercializado en gran escala?

Esta pregunta cae fuera del ámbito puramente académico y es muy complicado responderla sin otras herramientas que la especulación. Sin embargo, encontramos interesante exponer el análisis por un lado intuitivo y por otro basado en la experiencia que publicó el reputado tecnólogo, empresario e inventor estadounidense Alex Nugent. El propósito no es otro que trascender el campo académico y científico para introducirnos en un aspecto igual de importante del desarrollo tecnológico que no se estudia ni se aprende: el factor humano y empresarial. Estos son sin duda elementos fundamentes en cualquier actividad, y la investigación no es una excepción. Es indispensable que cualquier científico, ingeniero o desarrollador conozca lo mejor posible el estado real del mercado de su campo de investigación, las empresas o agencias que lo financian, y la disponibilidad del trabajo previo de otros colegas de su campo para poder desarrollar su trabajo de forma óptima. El estado de la tecnología de los memristores es un ejemplo perfecto de estas variables no técnicas en el desarrollo de la electrónica.

Alex Nugent, licenciado en Física por la universidad de Whitman (Washington, USA) es el CEO de la empresa "Knowm" y un prolífico investigador de aplicaciones de nanotecnología, aprendizaje automático y redes neuronales, campos sobre los cuales ha publicado varios artículos y en los que es titular de 40 patentes. En la página web de su empresa publica con bastante regularidad artículos que, sin ser técnicos, abordan temas tecnológicos de actualidad siempre relacionados con sus propios campos de trabajo. Analizaremos las ideas principales que expone en (Nugent, 2019) sobre la ausencia de disponibilidad comercial de memristores y los problemas que identifica como causas de esta ausencia.

Knowm, la empresa de Nugent, desarrolló y fabricó el "Memristor Discovery USB platform", el primer dispositivo con memristores disponible en el mercado y que podemos ver en la ilustración 66. Se trata de una sencilla placa de PCB con los drivers y conectores para poder interactuar y experimentar con los 16 memristores de tipo SDC que lleva integrados.

Ilustración 66: memristor discovery USB platform. Sitio web de Knowm

Nugent explica como pensó que el siguiente paso más lógico era comparar el funcionamiento de 'sus' memristores de tecnología SDC o "self-directed channel" (los cuales analizaremos en el siguiente punto) con los memristores de óxido de titanio o de hafnio de HP. Parece ser que le resultó imposible durante meses conseguir ningún prototipo aislado o discreto, y cita textualmente el testimonio de la inventora de los memristores SDC diciendo que ella misma intentó durante años lo mismo, sin resultados. Lo único que podía obtener de los desarrolladores eran memristores ya integrados en placas de circuito diseñadas y construidas para un estudio concreto. Esto, incluso para un profano, llama mucho la atención, ya que no sería lógico ni científica ni económicamente pasar directamente de modelos de simulación a obleas o circuitos integrados, sin estudiar primero el rendimiento de los dispositivos de forma aislada.

La explicación que Nugent deduce es que los memristores basados en óxido de metales tienen un tiempo de vida útil suficientemente largo para ser fabricados, testeados y publicar los resultados de dichos tests, pero no lo bastante como para ser comercializados. La proliferación de la bibliografía y el trabajo del departamento de márketing de una empresa tan grande como HP generaron una falsa esperanza al no poner el foco en los inconvenientes de esta tecnología, sino en las interesantes aplicaciones potenciales. La posición de la empresa de no querer reconocer antes de tiempo que su tecnología es un callejón sin salida es no solo comprensible si no también esperable, y en cuanto a la gran cantidad de bibliografía académica publicada el señor Nugent resume su explicación en el dicho "*Publish or perish*", que se traduce como "Publica o perece". Lo que intenta decir es que es muy normal (estadísticamente y comprensivamente hablando) que un grupo de investigación que ha trabajado durante meses en una aplicación o proyecto con cierta tecnología y que han demostrado que dicha aplicación es funcional no se arriesguen a esperar un tiempo no específico para simplemente comprobar que el dispositivo no se ha oxidado y por tanto estropeado antes de publicar su trabajo, ya que durante ese tiempo su investigación ha podido quedar obsoleta. Nugent pone el ejemplo de la oxidación precisamente porque apunta a la oxidación como causa más probable del corto periodo de vida de los memristores de HP, y por supuesto esto haría actualmente inviable la implementación de estos dispositivos en aplicaciones orientadas a su venta al público.

Por último, no podemos obviar la polémica existente en cuanto a la coherencia de la descripción física y del modelo de funcionamiento del memristor que dan los científicos de HP. Importantes expertos en el campo de la materia condensada señalan imposibilidades físicas en la descripción del memristor, como en (Meuffels y Soni, 2012) donde se expone que las ecuaciones que describen la dinámica de un memristor con la propiedad de no-volatilidad en su resistencia y controlado por corriente violan el principio de Landauer, el cual dicta que en cualquier proceso de manipulación de información almacenada se cumple el segundo principio de la termodinámica, y establece una energía mínima de energía que se disipa en forma de calor de manera inevitable en cualquier cambio de estado de bits, ya sea escritura o borrado.

7 Alternativas funcionales: el memristor "self-directed channel "

Analizaremos los memristores de canal autodirigido o "self-directed channel" (SDC) al ser estos los únicos memristores actualmente disponibles en el mercado y por tanto representar la principal esperanza que esta tecnología tiene de salir adelante, puesto que las principales aplicaciones que se barajan para ellos son claramente de orientación comercial e industrial (hardware para clasificación y aprendizaje profundo, y memorias no volátiles)

En (Campbell, 2017) vemos la descripción estructural y funcional del memristor de canal autodirigido. La estructura que se esquematiza en la ilustración 67 consiste en los dos electrodos (capa gris) entre los cuales encontramos las capas del dispositivo como tal.

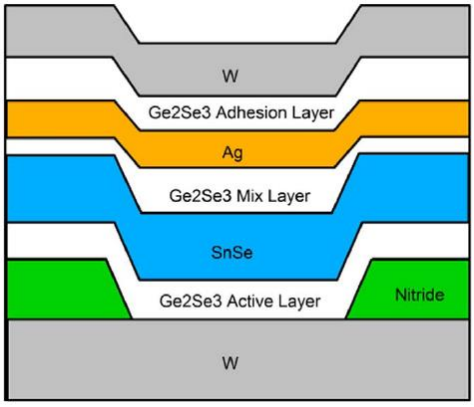


Ilustración 67: estructura del memristor SDC. Campbell (2017)

Entre otras propiedades, la autora destaca como gran ventaja de este memristor que se puede fabricar con una deposición de las capas mostradas muy sencilla, rápida y barata, usando tan solo una máquina de pulverización ya que no es necesario controlar el espesor de las capas para el correcto funcionamiento del dispositivo (los márgenes de error son de hasta 500 ángstroms). Además, el memristor SDC puede funcionar en régimen alto de temperatura, a 150ºC y picos de más de 250ºC. Al no necesitar recocido térmico ni procesos similares y fabricarse de una manera sencilla por deposición, se trata de un dispositivo rápido y barato de construir.

Tras su fabricación, el memristor SDC presenta una impedancia del orden de los MΩ. Cuando aplicamos un potencial positivo a los electrodos (en la ilustración 67 el positivo es el superior, y el negativo el inferior), este potencial genera iones de estaño que pasan de la capa SnSe a la capa activa de Ge2Se3. Estos iones generan la aparición de pares de electrones auto-atrapados localizados alrededor de los pares de Ge-Ge de la capa Ge2-Se3, facilitando la sustitución de átomos de germanio por átomos (iones) de plata. Dichas sustituciones forman los canal de conducción por desplazamientos de iones Ag+ que incrementan la conductividad entre los electrodos. Aplicando una tensión eléctrica negativa, los iones de plata se separan y vuelven a su capa de origen, disminuyendo así la conductividad.

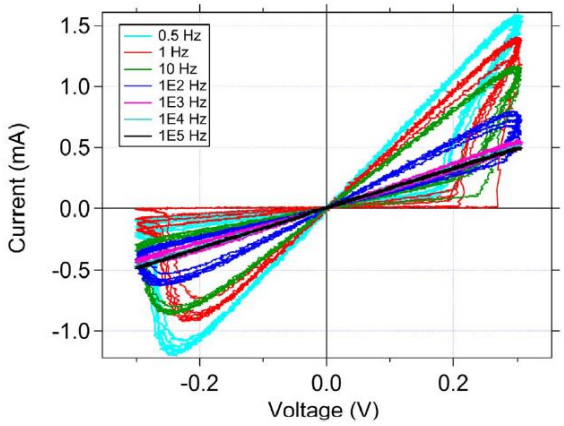


Ilustración 52: PHL del memristor SDC para distintas frecuencias. Campbell (2017)

Conclusiones

Como nos habíamos propuesto en los objetivos hemos descrito en el presente trabajo la fundamentación teórica de la existencia del memristor, su modelización matemática y simulación, y hemos analizado en detalle algunas de sus aplicaciones más importantes y prometedores.

Como principal conclusión resolvemos que esta tecnología puede ser un pilar fundamental en la evolución a corto y medio plazo de la industria 4.0 al servir de base para nuevas formas de criptoseguridad y para las redes neuronales, con todo lo que estas suponen: aprendizaje profundo y automático, inteligencia artificial, problemas dinámicos de gran tamaño, y reconocimiento de patrones. Sin embargo, los riesgos o fracasos de ciertos dispositivos pueden atrasar gravemente la llegada de los memristores al mercado.

El memristor de canal autodirigido es el candidato más prometedor ya que se ha demostrado ser funcional, sencillo de fabricar y además aunque no hemos encontrado estudios específicos sobre su vida útil, aún no se han reportado problemas graves en ese sentido, y sus fabricantes ya lo han puesto a disposición del consumidor esperando una nueva inyección de capital y una liberación de prototipos por parte de HP para poder continuar explorando y/o mejorando su empleo en las diferentes aplicaciones descritas.

Bibliografía

Adhikari, S. P., Kim, H., Sah, M., y Chua, L. (2019). Three fingerprints of memristor. En L. Chua, R. S. Williams, S. P. Adhikari, H. Kim, V. A. Gaurav Gandhi, M. P. Sah, et al., *Handbook of memristor networks.* Springer.

Ashbee, K. H., y Smallman, R. E. (1963). The plastic deformation of titanium dioxide single crystals. *Proceedings of the royal society*, 195-205.

Biolek, D., y Biolek, Z. (2014). Fourth fundamental circuit element: SPICE modeling and simulation. En R. S. Williams, L. Chua, M. D. Pickett, Z. Biolek, D. Biolek, A. Ascoli, . . . D. Sacchetto, *Memristor and memristive systems.* Springer.

Campbell, K. A. (2017). Self-directed channel memristor for high temperature operation. *Microelectronics journal* , 10-14.

Chua, L. (1971). Memristor - the missing circuit element. *Transactions on circuit theory*, 507-519.

Chua, L. (2019). If it´s pinched it´s a memristor. En L. Chua, R. S. Williams, S. P. Adhikari, H. Kim, V. A. Gaurav Gandhi, M. P. Sah, . et al., *Handbook of memristor networks* (págs. 15-88). Springer.

Chua, L. (2019a). Everything you wish to know about memristors but are afraid to ask. En L. Chua, R. S. Williams, S. P. Adhikari, H. Kim, V. A. Gaurav Gandhi, M. P. Sah, et al., *Handbook of memristor networks.* Springer.

Chua, L. (2019b). If it´s pinched it´s a memristor. En L. Chua, R. S. Williams, S. P. Adhikari, H. Kim, V. A. Gaurav Gandhi, M. P. Sah, et al., *Handbook of memristor networks* (págs. 15-88). Springer.

Chua, L. (2019c). The fourth element. En L. Chua, R. S. Williams, S. P. Adhikari, H. Kim, V. A. Gaurav Gandhi, M. P. Sah, et al., *Handbook of memristor networks* (págs. 1-14). Springer.

Farb, N. (1965). High temperature deformation of rutile. *Journal of applied physics*, 1746-1751.

Farb, N., Johnson, O., y Gibbs, P. (1965). High temperature deformation of rutile. *Journal of applied physics*, 1746-1751.

Gardos, M. N. (1988). The effect of anion vacances on the tribological properties of rutile. *Tribology transactions*, 427-436.

Hassan, A. M., Liu, C., Yang, C., Li, H., y Chen, Y. (2019). Designing neuromorphic computing systems with memristor devices. En L. Chua, & A. Adamatzky, *Handbook of memristor networks* (págs. 469-494). Springer.

Jiang, H., Belkin, D., Savel´ev, S., Lin, S., Wang, Z., Li, Y., Wu, Q. (2017). A novel true random number generator based on a stochastic diffusive memristor. *Nature communications*.

Lu, Y., Hirohashi, M., y Sato, K. (2006). Thermoelectric properties of non-stoichiometric titanium dioxide fabricated by reduction treatment using carbon powder. *Materials Transactions*, 1449-1452.

Merkel, C. E., Nagpal, N., Mandalapu, S., y Kudithipudi, D. (2011). Reconfigurable N-level memristor memory design. *Proceedings of international joint conference on neural networks*, 3042-3047.

Meufels, P. , y Soni, R. (2012). [Online]: Fundamental issues and problems in the realization of memristors. https://arxiv.org/abs/1207.7319

Nugent, A. (26-02-2019). *The oxide-memristor paradox.* https://knowm.org/oxide-memristors-have-shelf-life-problems

Rahma, F., y Muneam, S. (2019). *Memristive nonlinear electronic circuits: dynamics, synchronization and applications.* Springer.

Sahin, M. E., y Guler, H. (2017). TiO2 memristor modelling with Labview. *Turkish journal of science and technology*, 79-83.

Tang, H., Prasad, K., Sanjinés, R., Schmid, P. E., y Levy, F. (1994). Electrical and optical properties of TiO2 anatase thin films. *Journal of applied physics*, 2042-2047.

Valsa, J., Biolek, D., y Biolek, Z. (2010). An analogue model of the memristor. *International journal of numerical modeling: electronic networks, devices and fields*, 400-408.

Williams, R. S. (2014). How we found the missing memristor. En R. S. Williams, L. Chua, M. D. Pickett, Z. Biolek, D. Biolek, A. Ascoli, B. Linares-Barranco, *Memristor and memristive systems.* Springer.

Yang, J. J., Pickett, M. D., Li, X., Ohlberg, D. A., Stewart, D. R., y Williams, R. S. (2008). Memristive switching mechanism for metal/oxide/metal nanodevices. *Nature nanotechnology*, 429-433.

Yeon, H., Lin, P., Choi, C., Tan, S. H., Park, Y., Lee, D., Nie, Y. (2020). Alloying conducting channels for reliable neuromorphic computing. *Nature technology*.

Yilmaz, Y., y Mazumder, P. (2017). A drift-tolerant read/write scheme for multilevel memristor memory. *IEEE transactions on nanotechnology*, 1016-1027.

Zidan, M. A., Fahmy, H. A., Hussain, M. M., y Salama, K. N. (2012). Memristor-based memory: the sneak paths problem and solutions. *Microelectronics journal* , 176-183.