



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Estudio de nuevas tecnologías para memorias no volátiles emergentes

Trabajo Fin de Grado

Grado en Ingeniería Informática

Autor: Iván Aspas Coronado

Tutor: Daniel Gil Tomás

Curso 2018-2019

Resumen

Las tecnologías de memoria que constituyen el mercado actual están llegando a su límite de escalabilidad, y pronto la ley de Moore ya no podría aplicarse a estas tecnologías en su estado actual. La creciente demanda de más capacidad en el mismo espacio en Smartphones, cámaras digitales, reproductores de música, ordenadores, etc. y la ampliación del mercado de la memoria a la automoción y a la ropa inteligente, entre otras, hace más evidente la necesidad de avances en el campo del almacenamiento de datos. Diversas tecnologías desarrolladas a lo largo de las últimas décadas prometen combinar la velocidad de la SRAM, la densidad de almacenamiento de la DRAM o superior, y la no volatilidad de los datos que presenta la tecnología Flash, a un precio no muy alto. Una memoria con estas características sería capaz de revolucionar el mercado, cambiando por completo la jerarquía de memoria tal y como la conocemos ahora, y provocando más repercusión que la memoria Flash en la última década. Este trabajo se centra en analizar las diversas tecnologías de memoria no volátil que han surgido en los últimos años, especialmente las que poseen productos en el mercado actual, así como realizar una comparación entre éstas y las tecnologías clásicas. Para ello se ha efectuado un exhaustivo estudio bibliográfico a partir de publicaciones punteras, que se ha complementado con un análisis mediante algunas plataformas de simulación relevantes .

Palabras clave: Tecnologías de memories no volátiles emergentes, magnetic random-access memory (MRAM), spin-transfer torque random-access memory (STT-RAM), ferroelectric random-access memory (FeRAM), phase-change memory (PCM), resistive random-access memory(RRAM).

Abstract

The memory technologies that make up the current market are reaching their limit of scalability, and soon Moore's law could no longer apply to these technologies in their current state. The growing demand for storing more data in the same space in Smartphones, digital cameras, music players, computers, etc. and the expansion of the memory market to the automotive industry and smart clothing, among others, makes the need for advances in the field of data storage more evident. Several technologies developed over the last few decades promise to combine the speed of the SRAM, the storage density of the DRAM or higher, and the non-volatility of the Flash technology, at less cost. A memory with these characteristics would be able to revolutionize the market, completely changing the memory hierarchy as we know it now, and causing more impact than Flash memory in the last decade. This work focuses on analyzing the various non-volatile memory technologies that have emerged in recent years, especially those that have products in the current market, as well as making a comparison between them and classical technologies. For this, an exhaustive bibliographic study has been carried out based on leading publications, which has been complemented with an analysis by means of some relevant simulation platforms.



Keywords: Emerging nonvolatile memory technologies, magnetic random-access memory (MRAM), spin-transfer torque random-access memory (STT-RAM), ferroelectric random-access memory (FeRAM), phase-change memory (PCM), resistive random-access memory (RRAM).

Resum

Les tecnologies de memòria que constitueixen el mercat actual estan arribant al seu límit d'escalabilitat, i prompte la llei de Moore ja no podria aplicar-se a estes tecnologies en el seu estat actual. La creixent demanda de més capacitat en el mateix espai en Smartphones, càmeres digitals, reproductors de música, ordinadors, etc. i l'ampliació del mercat de la memòria a l'automoció i a la roba intel·ligent, entre altres, fa més evident la necessitat d'avanços en el camp de l'emmagatzemament de dades. Diverses tecnologies desenrotllades al llarg de les últimes dècades prometen combinar la velocitat de la SRAM, la densitat d'emmagatzemament de la DRAM o superior, i la no volatilitat de les dades que presenta la tecnologia Flash, a un preu inferior. Una memòria amb estes característiques seria capaç de revolucionar el mercat, canviant per complet la jerarquia de memòria tal com la coneixem ara, i provocant més repercussió que la memòria Flash en l'última dècada. Este treball se centra a analitzar les diverses tecnologies de memòria no volàtil que han sorgit en els últims anys, especialment les que posseïxen productes en el mercat actual, així com realitzar una comparació entre estes i les tecnologies clàssiques. Per a això s'ha efectuat un exhaustiu estudi bibliogràfic a partir de publicacions punteres, que s'ha complementat amb una anàlisi per mitjà d'algunes plataformes de simulació rellevants.

Paraules clau: Tecnologies de memories no volàtils emergents, magnetic random-access memory (MRAM) , spin-transfer torque random-access memory (STT-RAM) , ferroelectric random-access memory (FeRAM) , phase-change memory (PCM) , resistive random-access memory (RRAM).

Tabla de contenidos

1.	Introducción	10
1.1.	Motivación	10
1.2.	Problemática.....	10
2.	Jerarquía de memoria actual	12
2.1.	Estructura de la jerarquía de memoria.....	12
2.1.1.	Registros del procesador	13
2.1.2.	Memoria caché.....	13
2.1.2.1.	Memoria caché interna	14
2.1.2.2.	Memoria caché externa	14
2.1.3.	Memoria RAM.....	15
2.1.4.	Disco duro HDD	15
2.1.5.	Disco duro SSD.....	16
2.2.	Tecnologías empleadas actualmente	17
2.2.1.	Tecnología SRAM.....	17
2.2.2.	Tecnología DRAM	17
2.2.3.	Tecnología Flash	18
3.	Tecnologías de memoria emergentes	21
3.1.	MRAM.....	22
3.1.1.	STT-RAM	23
3.2.	FeRAM	25
3.3.	PCRAM	26
3.4.	RRAM	28
3.5.	Comparativa	30
3.5.1.	Tamaño de celda	30
3.5.2.	Densidad.....	32
3.5.3.	Latencia de lectura.....	34
3.5.4.	Latencia de escritura	35
3.5.5.	Consumo	36
3.5.6.	Conclusiones.....	37
3.6.	Arquitectura de chip	37
3.6.1.	1T1R Array.....	37
3.6.2.	X-Point Array	39



3.6.3.	3D X-Point Array	41
3.7.	Tecnologías futuras	42
3.7.1.	Memorias basadas en polímeros	42
3.7.2.	NRAM	43
3.7.3.	Memoria Millipede.....	44
3.7.4.	Memoria molecular	44
3.7.5.	Memoria MNW.....	45
3.7.6.	Memoria QD.....	46
4.	Situación de las NVM en el mercado.....	47
4.1.	Posibles aplicaciones de las NVM	48
4.1.1.	Reemplazo de memorias existentes.....	48
4.1.2.	Storage Class Memory	48
4.1.3.	Soluciones NVM de bajo consumo	49
4.1.4.	Informática inspirada en el cerebro humano	49
4.1.5.	Hardware de seguridad	50
4.1.5.1.	Generador de números aleatorios	50
4.1.5.2.	Función física irreproducible	51
4.1.6.	Hardware industrial	51
4.1.7.	Automoción	52
4.1.8.	Industria aeroespacial	52
4.2.	Dispositivos NVM existentes en el mercado.....	52
4.2.1.	Everspin 1-16Mb MRAM	52
4.2.2.	Avalanche Technology 1-32Mb MRAM.....	53
4.2.3.	Avalanche Technology 8-64Mb eMRAM	53
4.2.4.	Everspin 256Mb STT-RAM.....	54
4.2.5.	Avalanche Technology 1-32Mb STT-RAM.....	54
4.2.6.	Intel Optane 16-64GB 3D XPoint PCRAM	55
4.2.7.	Panasonic MN101L Series eRRAM.....	56
4.2.8.	Crossbar 256KB- 16MB eRRAM	56
4.2.9.	Crossbar 3D 1GB-1TB eRRAM	57
5.	Modelado de NVM	58
5.1.	NanoHUB	58
5.1.1.	Nanomaterial Mechanics Explorer.....	59
5.1.2.	Memristor Simulation Tool	59
5.1.3.	Magnetic Tunnel Junction Lab	59
5.1.4.	Subcircuit Generator for Ferromagnetic Nanomaterials	60

5.1.5.	Conclusiones.....	60
5.2.	NVSim.....	60
5.2.1.	Framework de NVSim	61
5.2.1.1.	Modelado del dispositivo.....	61
5.2.1.2.	Organización de los datos	61
5.2.1.3.	Tipos de banco de memoria	62
5.2.2.	Modelado del área	63
5.2.2.1.	Estimación del área de la celda	63
5.2.2.2.	Estimación del área de la circuitería externa.....	63
5.2.3.	Modelado de la latencia y el consumo	63
5.2.3.1.	Estimación del área de la circuitería externa.....	63
5.2.4.	Modelos de detección de datos	64
5.2.5.	Validación de los resultados de NVSim.....	64
5.3.	Modelado de NVM	65
6.	Conclusiones	72
6.1.	Trabajos futuros	72
7.	Referencias.....	73



Tabla de figuras

1. Jerarquía de memoria del computador.....	13
2. Estructura interna de un disco duro HDD.....	16
3. Estructura básica de una celda SRAM.....	17
4. Estructura básica de una celda DRAM.....	18
5. Estructura básica de una celda Flash NOR. Obtenida de [28].....	19
6. Estructura básica de una celda Flash NAND. Obtenida de [28].....	20
7. Clasificación de las tecnologías NVM. Obtenida de [1].....	21
8. Estructura básica de una celda MRAM. Obtenida de [1].....	23
9. Estructura básica de una celda STT-RAM. Obtenida de [1].....	24
10. Estructura básica de una celda FeRAM. Obtenida de [1].....	25
11. Estructura básica de una celda PCRAM. Obtenida de [3].....	27
12. Esquema básico del cambio entre los dos estados de la celda RRAM. Obtenida de [3].....	29
13. Estructura básica de una celda 1T1R-RRAM. Obtenida de [1].....	29
14. Comparativa entre tamaños de celda (Expresado en F2) . Obtenida de [17].....	31
15. Comparativa entre tamaños de celda medios (incluyendo circuitería de control). Obtenida de [17].....	31
16. Comparativa entre tamaños de celda y capacidad del chip. El tamaño de la burbuja indica el tamaño de la tecnología de fabricación de los semiconductores. Obtenida de [16].....	32
17. Gráfica que refleja la capacidad de los chips en el año que se ha realizado. El tamaño de la burbuja indica el tamaño de la tecnología de fabricación de los semiconductores. Obtenida de [16].....	33
18. Densidad promedio. Obtenida de [17].....	34
19. Tabla comparativa de la latencia de lectura entre NVM. Obtenida de [17].....	34
20. Latencia de escritura. Obtenida de [17].....	35
21. Comparativa entre el tiempo de escritura, el tamaño de celda, la capacidad del chip y la durabilidad de este. Un tamaño de burbuja mayor indica un mayor tamaño de celda. Las bandas de colores en los ejes indican el rango en el que se encuentra cada tecnología. Obtenida de [16].....	35
22. Comparativa entre la corriente necesaria para hacer una lectura y el año de fabricación. Un tamaño de burbuja mayor indica un mayor tamaño de celda. Obtenida de [16].....	36
23. Tabla resumen indicando las características generales de cada NVM. Obtenida de [1,2,3,4,9,18].....	37
24. Esquema del 1T1R Array: (a) Operación SET, (b) Operación RESET, (c) Operación READ. Obtenida de [18].....	38
25. Esquema del X-Point Array: (a) Operación SET en el esquema V/2, (b) Operación SET en el esquema V/3. Obtenida de [18].....	39
26. Esquema de los dos tipos de memoria 3D X-Point. Obtenida de [18].....	41
27. Estructura básica de una celda usando tecnología de polímeros. Obtenida de [1].....	42
28. Esquema básico de los dos estados de una celda NRAM. Obtenida de [1].....	43

29. Esquema básico de un cabezal en contacto con el material en el que los datos son almacenados. Obtenida de [1].....	44
30. Esquema básico de la tecnología de memoria molecular. Obtenida de [1].....	45
31. Esquema básico de la tecnología MNW. Obtenida de [1].....	46
32. Esquema básico de la tecnología Quantum Dot. Obtenida de [1].....	46
33. Comparativa entre los tiempos de acceso de la jerarquía de memoria actual y de las principales NVM emergentes, mostrando el hueco en el que las NVM se podrían ubicar en el futuro. Obtenida de [16].....	47
34. Memoria Intel Optane. Obtenida de [34].....	55
35. Microprocesador Panasonic MN101L Series. Versión de 64 pines. Obtenida de [35].....	56
36. Organización interna del modelo generado en NVSim. Obtenida de [38].....	62
37. Caché SRAM de 512KB.....	65
38. Caché SRAM de 8MB.....	65
39. Caché STT-RAM de 512KB.....	66
40. Caché STT-RAM de 8MB.....	66
41. RAM DRAM de 256MB.....	67
42. RAM DRAM de 1GB.....	67
43. RAM NAND Flash de 256MB.....	67
44. RAM NAND Flash de 1GB.....	67
45. RAM PCRAM de 256MB.....	68
46. RAM PCRAM de 1GB.....	68
47. RAM RRAM de 256MB.....	68
48. RAM RRAM de 1GB.....	68
49. Tabla resumen cualitativa de las prestaciones obtenidas en las simulaciones...	70



1. Introducción

Este capítulo está dedicado a enumerar las motivaciones y objetivos por los cuales he realizado este trabajo, enfocado al estudio de las nuevas tecnologías para memorias no volátiles que están surgiendo y desarrollándose en la actualidad.

1.1. Motivación

El mundo de la tecnología avanza a pasos de gigante, y resulta muy interesante ver hacia dónde se dirige el mundo de la informática, y en particular qué tecnologías de memoria van a aparecer en el mercado en un futuro cercano, y cómo éstas van a afectar a la jerarquía de memoria que está implantada en el mercado actual. Otro tema interesante es estudiar qué empresas están apostando por desarrollar este tipo de tecnologías y qué tipo de productos están disponibles para su compra en la actualidad.

En este trabajo también se realizaran simulaciones para dar una visión más práctica y poder comparar así las prestaciones entre las diversas tecnologías, ya que el resto del trabajo se centra más en aspectos bibliográfico-teóricos sobre el funcionamiento de las tecnologías de memoria no volátil emergentes, a nivel de celda y de arquitectura interna.

1.2. Problemática

En la actualidad, la demanda de un incremento en las prestaciones de los productos electrónicos está creciendo, y por ello la necesidad de incrementar las prestaciones de la memoria también aumenta. En su estado actual, un dispositivo de memoria Flash convencional funciona almacenando los datos en un transistor que contiene una puerta flotante (Floating Gate o FG) y éste está separado del circuito de control por una fina capa de óxido. Dicha capa, para asegurar la retención de los datos durante al menos 10 años, no puede ser de espesor inferior a 8nm. Esta necesidad limita la escalabilidad de la memoria Flash, ya que reducir más el tamaño de esta capa puede reducir la retención de los datos del dispositivo [18], debido a pérdida de la carga eléctrica de la celda

En los últimos años se han propuesto diversas estructuras alternativas para evitar los problemas de escalamiento en los dispositivos Flash, en un esfuerzo por continuar la ley de Moore y superar las limitaciones de los dispositivos basados en MOS. Alternativamente, también se han propuesto otros conceptos de almacenamiento (la mayoría basados en un elemento capaz de alternar entre dos estados de resistencia eléctrica diferentes) que intentan conseguir la llamada “memoria universal”, que poseería las ventajas de las tres tecnologías principales que componen la jerarquía de memoria actual: Alta velocidad, alta densidad, bajo consumo y no volatilidad. Un dispositivo con tales características sería el Santo Grial de la memoria, aunque en la

actualidad, aún no se ha podido desarrollar. Muchos dispositivos de memoria no volátil que proponen cambios en la forma en la que se almacena la información están siendo investigados con el propósito de encontrar una memoria que posea estas tres características. Ya que estos dispositivos utilizan tecnologías que van más allá de la celda basada en carga eléctrica tradicional, aún son necesarias más labores de I+D para explorar las posibilidades de estos dispositivos.

Actualmente, las tecnologías emergentes más prometedoras son STT-MRAM, FeRAM, PCM y RRAM. Cada una de estas tecnologías será descrita y analizada en las secciones siguientes. En cuanto a la producción comercial, todas ellas se encuentran o en fase de prototipado o ya disponen de productos comerciales, pero aún se limitan a aplicaciones de nicho en comparación con DRAM y Flash, ya que para que estas tecnologías puedan ser adoptadas por la industria en una escala mayor las prestaciones que deben conseguir (escalabilidad, velocidad, consumo, precio) deben ser superiores a los de las tecnologías de memoria existentes, y por el momento ese no es el caso.

Otra posible entrada al mercado de éstas sería mediante una tecnología híbrida, que combine los puntos fuertes de las nuevas tecnologías con las ya implantadas en el mercado, dando paso a mejoras en los procesos de desarrollo de las nuevas memorias, y como resultado se produzca una reducción de los costes de fabricación de éstas, haciendo más viable su entrada como dispositivo independiente.

2. Jerarquía de memoria actual

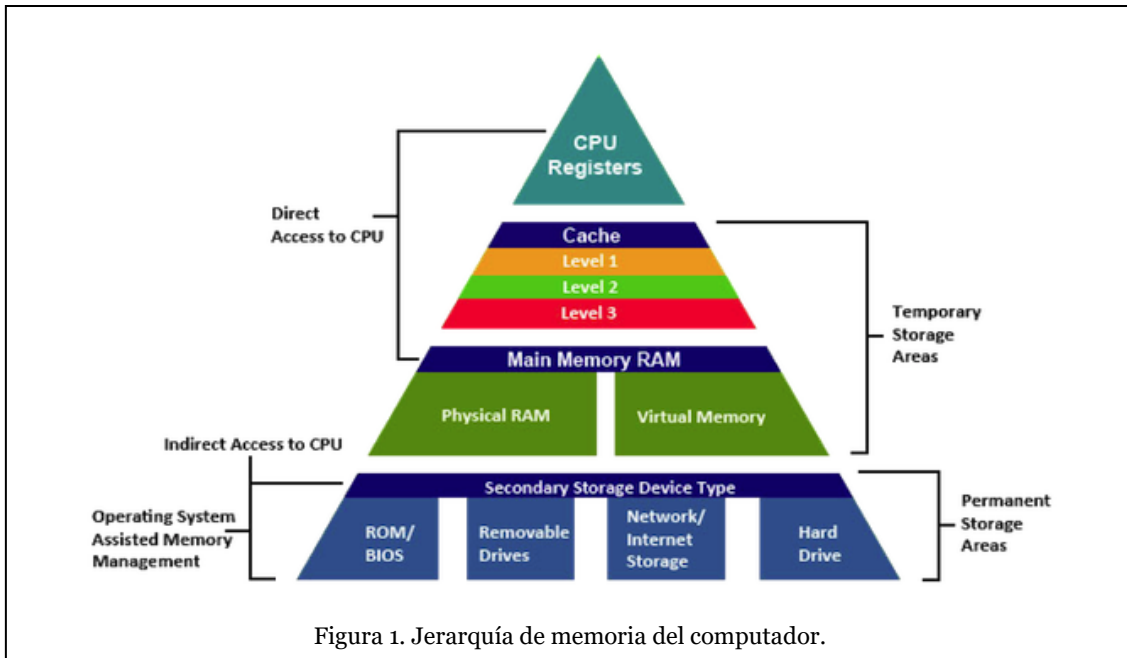
Un sistema de memoria está formado por las unidades donde se almacenan tanto el conjunto de instrucciones del sistema como los datos. Este tiene dos cometidos, el primero es trabajar con información, es decir, debe almacenar las instrucciones y datos con los que trabaja el procesador en un momento dado, así como resultados de operaciones, resultados intermedios, direcciones de memoria, etc. El segundo cometido es el de almacenar la información, ya sea de forma temporal o permanente. Un ejemplo de esto sería el almacenamiento de programas o datos listos para su uso en un futuro.

Desde un punto de vista tecnológico, una memoria "perfecta" debería ser rápida, con una elevada vida útil, de una gran densidad, y no volátil, pero como, de momento, aun no existe una memoria que reúna semejantes características, se optó por establecer una jerarquía de memoria, con la cual se pueda contar con capacidad suficiente para las necesidades del usuario, se pueda obtener los datos almacenados lo antes posible, se obtenga un ancho de banda elevado y el coste por bit sea lo más reducido posible.

2.1. Estructura de la jerarquía de memoria

Para conseguir estos cometidos, tradicionalmente, un ordenador tiene a su disposición diversas estructuras de memoria, que se podrían agrupar en dos categorías principales. Por un lado estaría la memoria interna, caracterizada por su volatilidad y una alta velocidad, y constituida por los registros del procesador, la memoria caché (SRAM), y la memoria principal (DRAM), y por otro lado tendríamos la memoria externa, caracterizada por su no volatilidad y su capacidad superior, y compuesta por los discos de estado sólido o SSD, las memorias Flash USB, los dispositivos magnéticos (HDD), y los dispositivos ópticos[27].

En la siguiente figura se detalla como estaría organizada la jerarquía de memoria de un computador, en la cual cuanto más cerca del ápice de la pirámide se encuentre una memoria, mayor será su velocidad, su ancho de banda, y su precio, mientras que las memorias en los estratos inferiores aumentarían en capacidad y en densidad de almacenamiento.



2.1.1. Registros del procesador

Los registros son, básicamente, elementos de memoria de rápido acceso que se encuentran en el procesador. Constituyen uno de los espacios de trabajo del procesador y se utilizan como espacio de almacenamiento temporal. Se implementan usando elementos de memoria RAM estática (SRAM)[21]. Resultan imprescindibles para ejecutar instrucciones en el procesador.

El conjunto de registros y su organización cambia de un procesador a otro. Estos registros pueden o no ser visibles para el programador de aplicaciones, y están divididos en seis grupos, cada uno con un fin específico[20]. Estos grupos son: registro de segmento, registro apuntador de instrucciones, registro apuntador, registro de propósito general, registro índice y registro de bandera

2.1.2. Memoria caché

La memoria caché es una memoria temporal, generalmente oculta y automática para el usuario, que proporciona acceso rápido a los datos de uso frecuente. La utilización de este tipo de memorias se basa en dos suposiciones que generalmente resultan ciertas: Los ordenadores tienden a utilizar las mismas instrucciones, y en menor medida, los mismos datos repetidamente, y que la información que necesita el procesador después de una operación se suele encontrar adyacente o muy cercana a la información que se está usando en esta, ya sea en memoria principal o en disco.



Su funcionamiento se basa en que al cargar una información en la memoria principal (sean instrucciones o datos), ésta se carga en zonas adyacentes de la memoria. Un controlador especial situado dentro del subsistema de caché determinará que posiciones de la memoria RAM convencional pueden ser utilizadas con más frecuencia por la aplicación que está ejecutándose en ese momento y traslada la información almacenada en ellas a la memoria caché. La siguiente vez que el microprocesador necesite acceder a la memoria RAM convencional existirá una gran probabilidad de que la información que necesita encontrar se encuentre en las direcciones adyacentes a las ya utilizadas[22]. Como estas direcciones ya están almacenadas en la memoria caché, el tiempo de acceso a la información disminuye en gran medida.

Desde el punto de vista del hardware, existen dos tipos de memoria caché: interna y externa.

2.1.2.1. Memoria caché interna

Denominada también caché primaria, caché de nivel 1 o simplemente caché L1, en realidad está dividida en dos, cada una con una misión específica: Una para datos y otra para instrucciones. Están incluidas en el procesador junto con su circuitería de control, lo que significa tres cosas: comparativamente es muy cara; extremadamente rápida, y limitada en tamaño. Como puede suponerse, su velocidad de acceso es comparable a la de los registros, es decir, centenas de veces más rápida que la RAM.

2.1.2.2. Memoria caché externa

Se conoce también como caché secundaria, caché de nivel 2 o caché L2. Es una memoria de acceso rápido incluida en la placa base, que dispone de su propio bus y controlador independiente que intercepta las llamadas a memoria antes que sean enviadas a la RAM. La caché externa típica es un banco SRAM ("Static Random Access Memory"). Esta memoria es considerablemente más rápida que la DRAM ("Dynamic Random Access Memory") convencional, aunque también mucho más cara. También se puede incluir esta caché en el procesador.

Actualmente también existe un tercer nivel de caché denominado L3. Es más rápida que la memoria principal (RAM), pero más lenta que L2. En esta memoria se agiliza el acceso a datos e instrucciones que no fueron localizadas en L1 o L2. Es generalmente de un tamaño mayor y ayuda a que el sistema guarde gran cantidad de información agilizando las tareas del procesador.

2.1.3. Memoria RAM

La memoria RAM está conformada por un conjunto de módulos de memoria, con un cuerpo compuesto de fibra de vidrio. Esta memoria utiliza la tecnología DRAM y sus sucesoras, SDRAM, DDR SDRAM, DDR2 SDRAM, DDR3 SDRAM y DDR4 SDRAM[24], ya que fabricar la capacidad necesaria para este módulo de memoria con tecnología SRAM sería demasiado caro. Es considerada como uno de los componentes más importantes a la hora de realizar cualquier equipo de cómputo, ya que se encarga principalmente de almacenar y ejecutar todas las instrucciones asignadas por la unidad central de procesamiento (procesador o CPU). Adicionalmente, la memoria RAM también almacena y realiza las ordenes de otros componentes en conjunto con la tarjeta madre, tal es el caso de la tarjeta gráfica, el disco duro e incluso las mismas aplicaciones que se encuentren instaladas en el equipo[23]. Los chips de memoria DRAM se agrupan en circuitos impresos que constituyen módulos de memoria y que son usados para ser insertados directamente en una computadora. Estos módulos están estandarizados y han evolucionado en paralelo a los diferentes tipos de DRAMs. En cada iteración se busca un incremento en la frecuencia de reloj y en la tasa de transferencia de datos, junto con una reducción del consumo frente a su iteración anterior.

2.1.4. Disco duro HDD

Un disco duro o disco rígido (en inglés hard disk drive) es un dispositivo no volátil, que conserva la información aun con la pérdida de energía, que emplea un sistema de grabación magnética digital. Dentro de la carcasa hay una serie de platos metálicos apilados girando a gran velocidad. Sobre los platos se sitúan los cabezales encargados de leer o escribir los impulsos magnéticos[25]. Hay distintos estándares para comunicar un disco duro con la computadora, siendo SATA el más común actualmente. Este tipo de unidades de memoria son las más delicadas de la lista, ya que tiene en su interior muchas partes móviles, y en parte también debido a la distancia extremadamente pequeña entre los cabezales y la superficie del disco, cualquier contaminación de los cabezales de lectura/escritura o las fuentes puede dar lugar a un accidente en los cabezales, un fallo del disco en el que el cabezal raya la superficie de la fuente, a menudo moliendo la fina película magnética y causando la pérdida de datos. Estos accidentes pueden ser causados por un fallo electrónico, un repentino corte en el suministro eléctrico o por golpes, entre otras razones. Pese a esto, sigue siendo la memoria no volátil más usada como almacenamiento masivo interno del ordenador, ya que son las más baratas y las que poseen mayor capacidad de almacenamiento hasta la fecha.



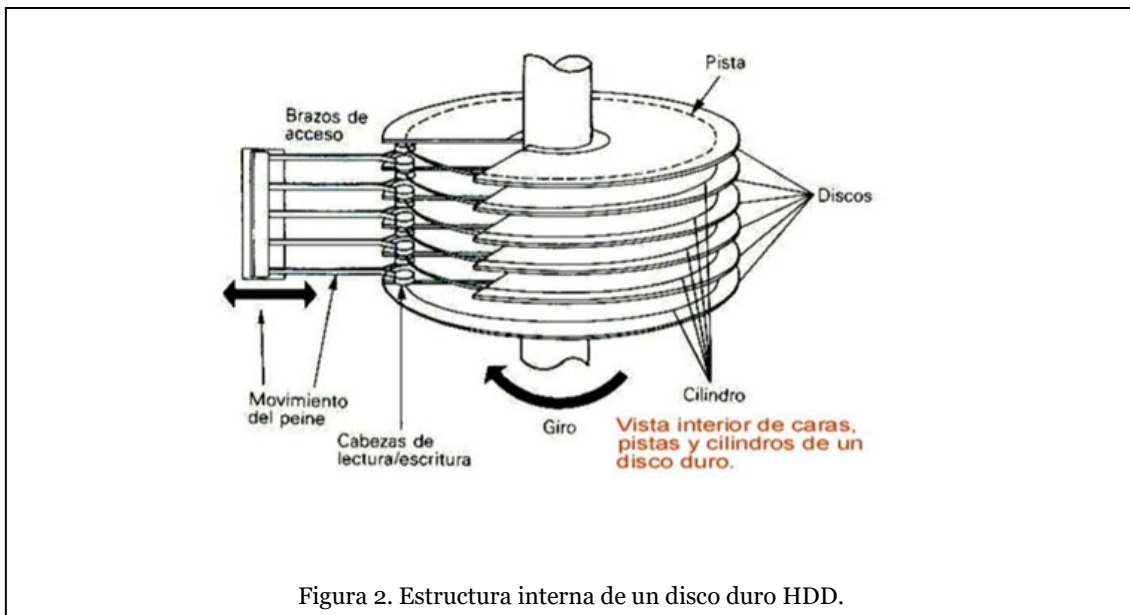
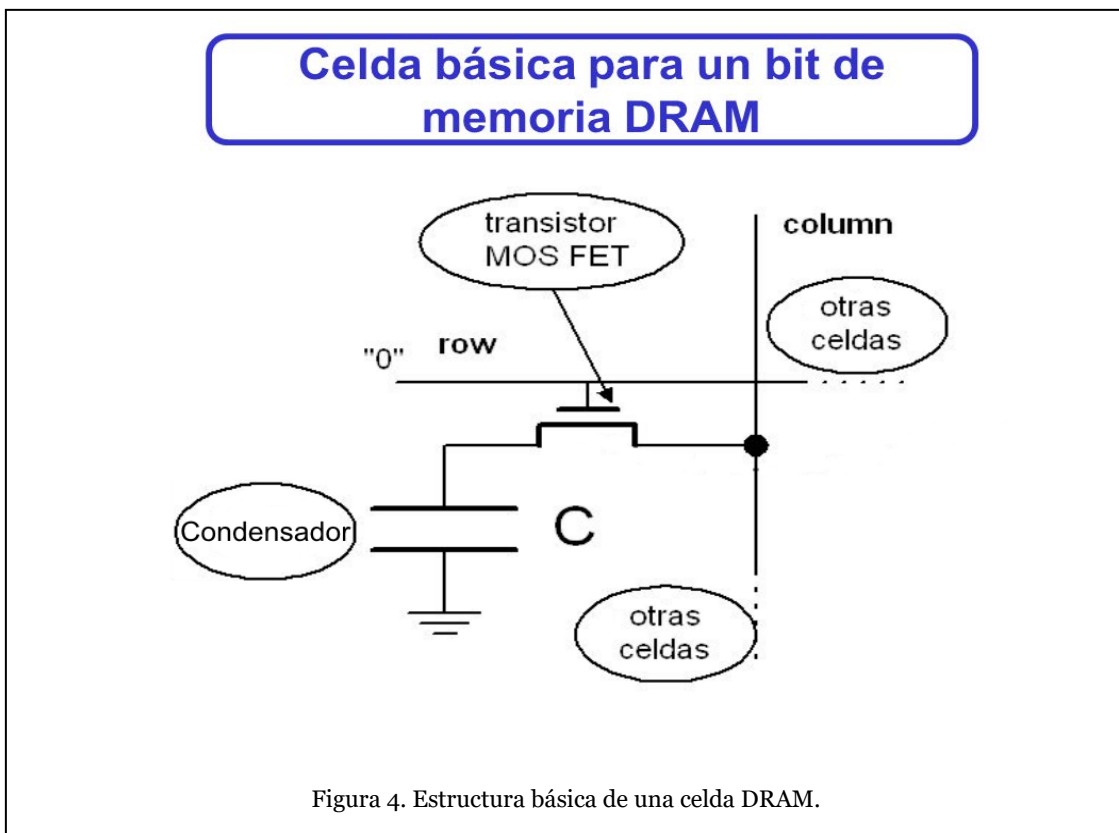


Figura 2. Estructura interna de un disco duro HDD.

2.1.5. Disco duro SSD

El disco duro SSD o unidad de estado sólido es un dispositivo de almacenamiento de datos que utilizan chips de memoria no volátil para almacenar los datos. También hacen uso de la misma interfaz SATA que los discos duros, por lo que son fácilmente intercambiables con los discos HDD. La tecnología más común con la que se fabrican este tipo de dispositivos es Flash, en concreto Flash tipo NAND. Dado que todo el proceso de lectura y escritura se realiza mediante impulsos eléctricos, los discos SSD no cuentan con piezas móviles en su interior[27]. Esto hace que los tiempos de acceso a los datos de un SSD resulten inapreciables, y esto constituye una gran ventaja respecto a unidades HDD, y hace a las unidades SSD mucho más rápidas que estas. Además, al no tener piezas móviles se minimiza la probabilidad de fallo mecánico y los hace más resistentes a los daños por pequeños impactos que pudieran sufrir al estar montados en un portátil. No obstante, la durabilidad no es el fuerte de las unidades SSD ya que las celdas de memoria de los discos SSD soportan un determinado número de operaciones de escritura y borrado y aunque puedes tardar años en llegar a ese límite, una vez superado ese número de operaciones se corre el riesgo de que el almacenamiento SSD comience a dar fallos de escritura o lectura, corrompiendo así los datos almacenados en este.

progresivamente, necesitando de un circuito dinámico de refresco que, cada cierto período, revisa dicha carga y la repone en un ciclo de refresco. Este ciclo puede interferir con los ciclos de acceso a memoria, por lo que se intenta que este periodo constituya el menor tiempo posible, sin afectar a la integridad de los datos, ya que, como el resto de memorias RAM, es volátil, es decir, si se interrumpe la alimentación eléctrica, la información almacenada se pierde. Cada bit en una DRAM consiste en un transistor de efecto de campo y un condensador[27]. El principio de funcionamiento básico, es sencillo: una carga se almacena en el condensador significando un 1 y sin carga un 0. El transistor funciona como un interruptor que conecta y desconecta al condensador. La principal ventaja de esto es la posibilidad de construir memorias con una gran densidad y que todavía funcionen a una velocidad alta, a un precio más reducido que un modulo SRAM de la misma capacidad.



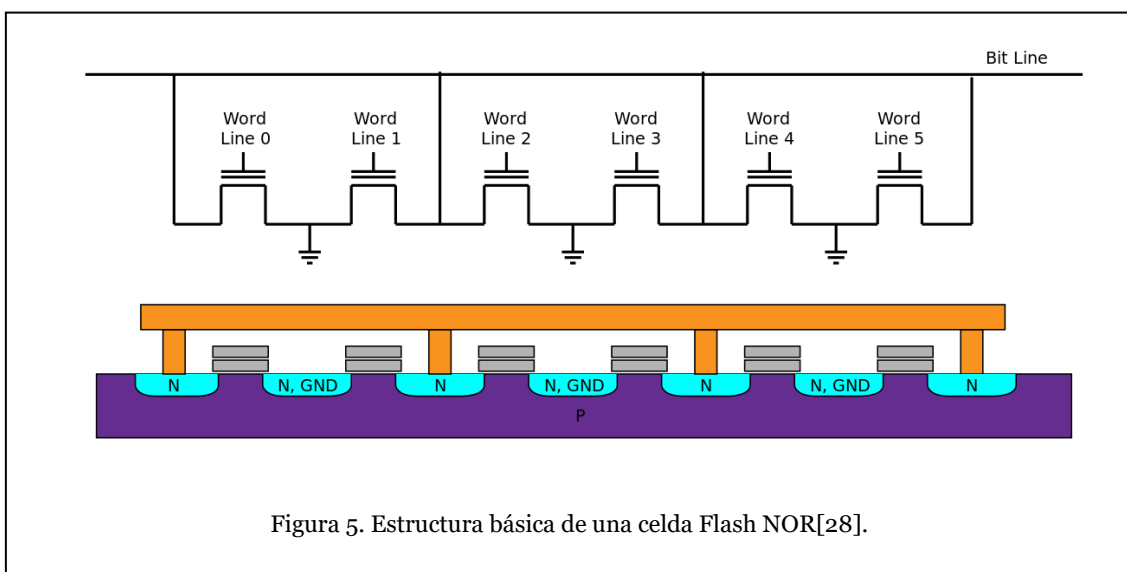
2.2.3. Tecnología Flash

Flash es una tecnología generalmente utilizada en tarjetas de memoria, discos duros SSD y otros dispositivos de almacenamiento. Es considerada uno de los mejores inventos en el campo de la electrónica desde la invención del transistor, ya que a día de hoy es una tecnología utilizada por todo el mundo, habiendo un sinnúmero de dispositivos cuya memoria secundaria se compone únicamente de una unidad Flash. Son un tipo de memoria EEPROM, que contienen una matriz de filas y columnas con celdas que tienen dos transistores en cada intersección. Tradicionalmente sólo almacenan un bit de información, aunque las más actuales pueden almacenar más de un bit por celda según el número de electrones que almacenan. Ambos transistores

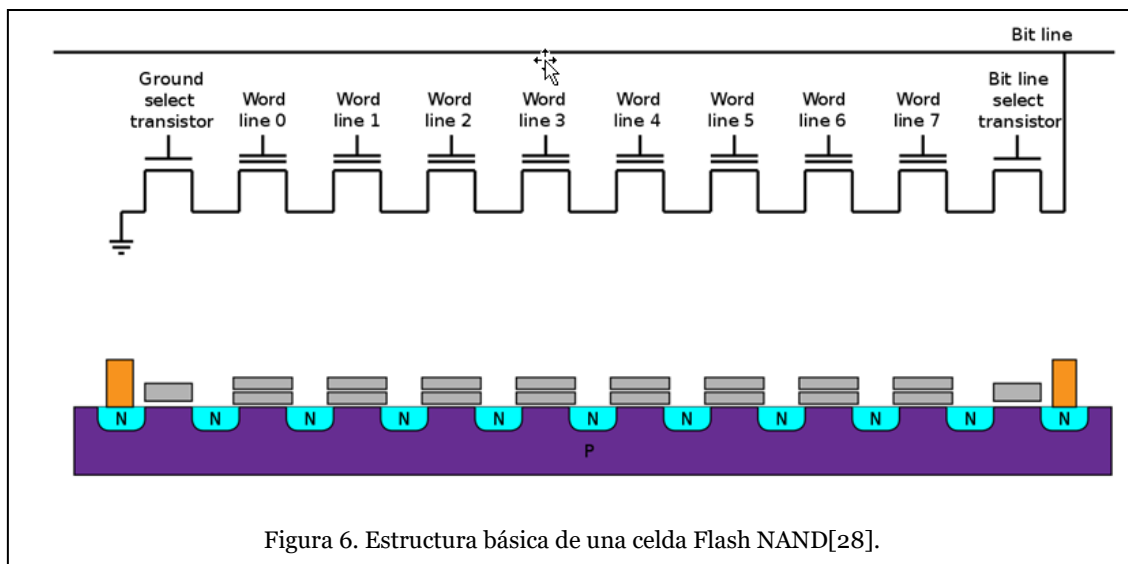
están separados por una capa de óxido. Uno de los transistores es denominado Floating Gate o FG. El FG está conectado a la fila a través del otro transistor, denominado Control Gate o CG. Cuando esta conexión se establece, el valor 0 es establecido, mientras que el valor es 1 cuando ambos transistores no están unidos. Para escribir los bits en las celdas se aplica una descarga eléctrica que transfiere o no, dependiendo del estado de los transistores. Si no se transfiere a través de los transistores equivaldría a un 1, mientras que si la corriente se transfiere equivaldría a un 0. El valor de la FG se mantiene en el estado en el que ha sido programado, incluso cuando se desconecta de la corriente, lo que hace que esta memoria sea no volátil.

Existen dos tipos principales de memorias Flash, tipo NOR y tipo NAND. Su nombre viene dado por la estructura usada para realizar las conexiones internas entre celdas de memoria. En el caso de las NOR, las celdas están interconectadas en paralelo a las líneas de bit, lo que permite a las celdas ser leídas y programadas individualmente, mientras que las celdas NAND están interconectadas en serie, lo que consume menos espacio, pudiendo incrementar la densidad de celdas[28].

El tipo NOR permite una lectura y escritura más lenta que NAND, pero archiva muy rápido las rutas de acceso aleatorias. Esto hace que NOR sea más adecuado para la ejecución y almacenamiento de comandos, mientras que NAND es más indicado para el almacenamiento masivo de datos.



En cuanto a la arquitectura, NAND puede almacenar más datos en un espacio de silicio más pequeño, lo que ahorra el coste por bit. En el pasado, cuando el almacenamiento de datos era más bajo, NOR tuvo mayor influencia en el mercado. Hoy, con el gran incremento de la necesidad de guardar más datos, el consumo de la electrónica y el negocio de los dispositivos, NAND ha superado de lejos a NOR.



La demanda de un incremento en las prestaciones en los dispositivos de memoria Flash ha resultado en el descubrimiento de una de sus mayores limitaciones. Esta es el grosor mínimo de la capa de óxido que separa los transistores. Dicha capa, para asegurar la retención de los datos durante al menos 10 años, no puede ser inferior a 8nm. Reducir más el tamaño de esta capa puede reducir la retención de los datos del dispositivo de 10 años a unos pocos segundos, debido a la que la celda ya no es capaz de retener la carga correctamente a partir de ese punto. Encontrar esta limitación ha hecho más evidente la necesidad de invertir en encontrar formas de almacenar la información distintas de el almacenamiento basado en la retención de una carga eléctrica, como por ejemplo el uso de un material que posea dos estados de resistencia eléctrica distintos. La investigación y el desarrollo de estas nuevas formas de almacenamiento pueden dar paso a dispositivos de memoria con mejores prestaciones que los actuales, con capacidad de escalar por debajo de los límites de Flash, por un precio y con una dificultad de fabricación menor.

3. Tecnologías de memoria emergentes

La jerarquía de memoria actual permanece inamovible desde hace décadas, y cuanto más tiempo pasa las limitaciones de las diversas tecnologías que la componen se hacen más evidentes con cada iteración de estas. Durante décadas estas tecnologías han ido siendo reducidas a un tamaño menor para obtener una mayor densidad a un menor coste y poder reducir así el consumo, y gradualmente se ha ido acercando al límite físico al que este tipo de memorias es posible escalar. Por esto, nuevos tipos de memorias basados en otras tecnologías para almacenar los datos han ido cogiendo importancia a lo largo de los últimos años. Para considerar un nuevo tipo de memoria como posible competencia de los sistemas actuales de memoria, este debería ser escalable, tener un consumo reducido, una velocidad alta de operación, una estructura simple, una alta durabilidad y que sea capaz de retener datos durante periodos largos y sin necesidad de corriente eléctrica.

Nuevas tecnologías están siendo investigadas y conducen al futuro como posibles alternativas a las memorias existentes en futuros sistemas informáticos. Tecnologías emergentes de memoria no volátil. tales como memoria RAM magnética (MRAM), Spin Transfer Torque Memory (STT-RAM), Memoria RAM ferroeléctrica (FeRAM), Phase Change Memory (PCM) y Resistive RAM Memory (RRAM), combinan la velocidad de la memoria SRAM, la densidad de la memoria DRAM, y la no volatilidad de la memoria Flash, por lo que se vuelven muy atractivas para futuras jerarquías de memoria, algunas de ellas ya se encuentran incluso en fases de producción comercial, aunque en el estado actual aun no poseen un mercado remotamente comparable al de SRAM, DRAM o Flash. Un esquema completo con la clasificación de las nuevas tecnologías de memoria no volátil se encuentra en la figura siguiente. En ella las clasificadas como emerging son las que tienen mayor proyección a corto plazo, mientras que las clasificadas como new son propuestas a más largo plazo.

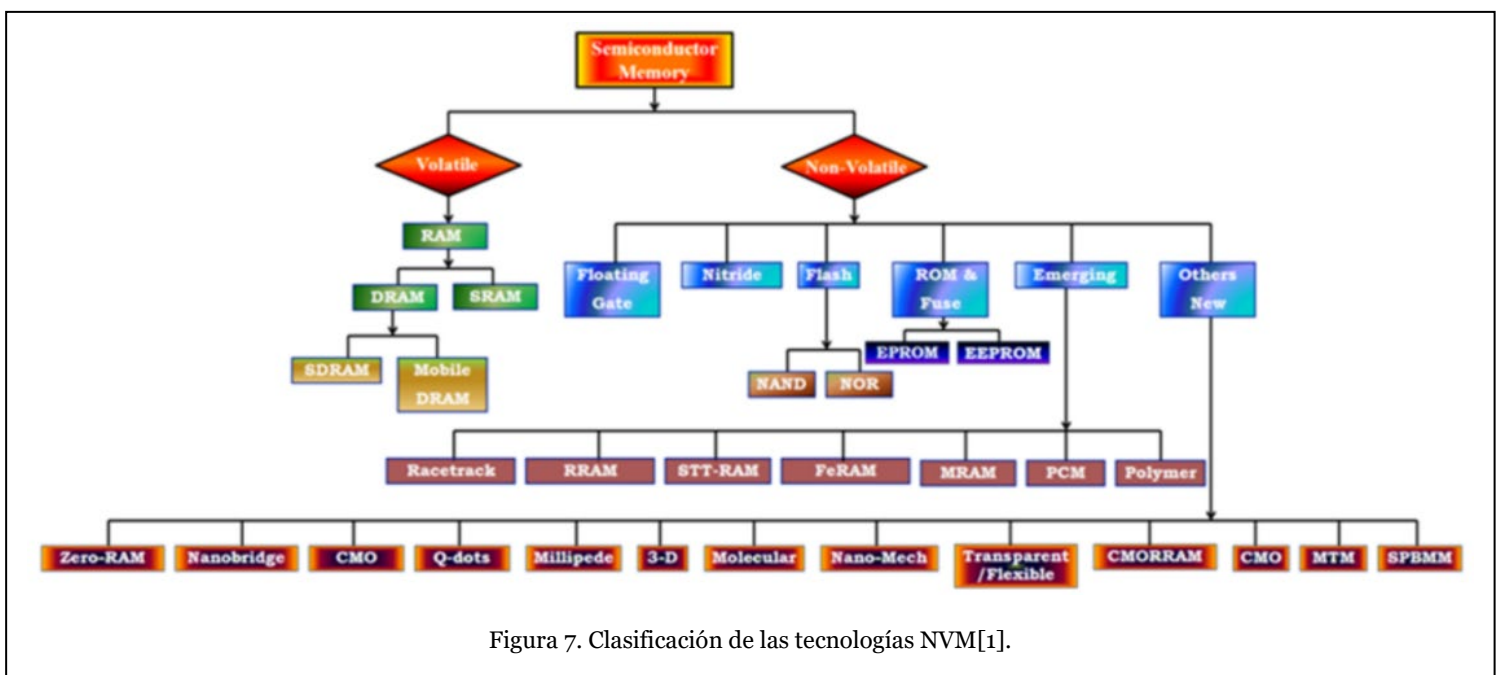


Figura 7. Clasificación de las tecnologías NVM[1].



En este apartado, además de intentar explicar el funcionamiento de las tecnologías emergentes de memoria no volátil también se explicará las nuevas tecnologías aun en fases experimentales que me parezcan más atractivas, a fin de ofrecer una visión lo más amplia posible del conjunto de tecnologías de NVM (Non Volatile Memory) que pueden cambiar la jerarquía de memoria actualmente establecida.

Ademas, tambien se compararan las tecnologias NVM entre ellas y con las tecnologias presentes en la jerarquia de memoria asentada en la actualidad, con el fin de ubicar las diversas NVM emergentes en el nicho de mercado que pretenden ocupar. Los criterios para esta comparación serán los siguientes:

1. **Estado de madurez de la tecnología:** Si esta tecnología se encuentra ya en algún producto a la venta o se encuentra en fases de desarrollo.
2. **Tamaño de celda**
3. **Densidad**
4. **Latencia de lectura**
5. **Latencia de escritura**
6. **Durabilidad de la celda**
7. **Consumo:** Consumo por bit accedido. Se indicaran los consumos de lectura y de escritura/borrado de bit si es necesario.

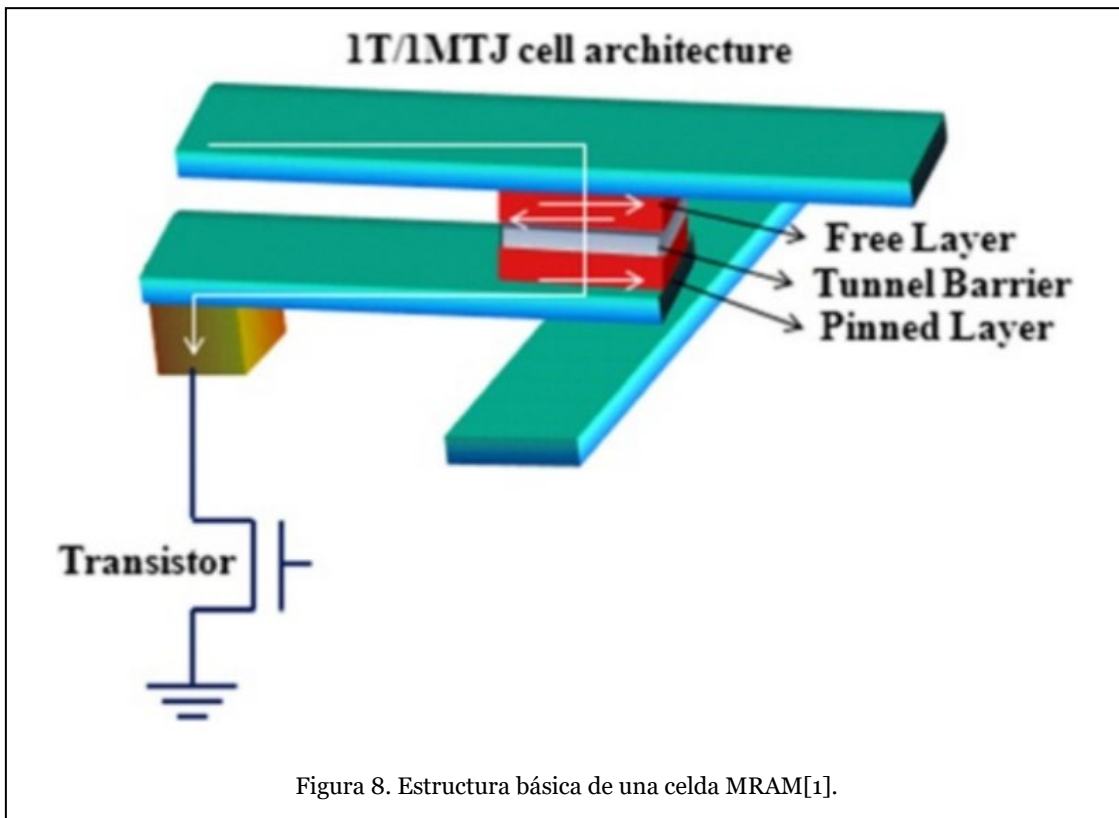
3.1. MRAM

La Memoria RAM Magnética o MRAM se encuentra en desarrollo desde la década de los 90. Esta tecnología utiliza dominios magnéticos en vez de eléctricas para almacenar la información. En términos simples, una celda que utilice esta tecnología contendría dos elementos de almacenamiento magnético, uno llamado fijo con polaridad fija, y otro denominado libre, capaz de cambiar entre una polaridad paralela y una opuesta al fijo. La polaridad del libre se puede cambiar aplicando un campo magnético sobre este. Estos dos componentes están separados por una capa aislante como se muestra en la figura 8, en la que actualmente se usa Oxido de Magnesio (MgO) [6]. Si la polaridad magnética de ambos elementos es paralela, debido a un fenómeno llamado efecto túnel magnético (Magnetic Tunnel Junction o MTJ), al establecer una corriente eléctrica por la celda esta dejaría pasar los electrones con facilidad, mientras que si la polaridad de ambos esta invertida, la resistencia de la celda seria alta.

En principio, a parte de su reducido consumo, con esta tecnología se pueden lograr velocidades de acceso a los datos cercanos al de la SRAM, con una densidad comparable a la de la DRAM, haciéndola una candidata interesante a la hora de reemplazar estas tecnologías en el futuro. Sin embargo, las principales desventajas de esta serian su elevado coste en comparación con las memorias actuales(del orden de 1000 veces más caro que un tamaño equivalente de memoria Flash), y su limitación de tamaño, ya que un tamaño muy pequeño a la hora de escribir los datos en las celdas

podría afectar a los datos escritos en las celdas cercanas, debido a los campos magnéticos creados por éstas [7]. Por esto se cree que esta tecnología nunca va a llegar a un tamaño de nodo por debajo de los 65nm.

En los últimos años, se han descubierto diversos mecanismos para realizar la escritura de datos en este tipo de memorias gracias a las corrientes eléctricas polarizadas (electrones con espín alineado). Esto ha dado paso a que esta tecnología pase a un segundo plano, ya que tecnologías sucesora de esta poseen menos limitaciones a un precio más reducido, haciéndola mucho más interesante.



Varias compañías como Samsung, IBM, Hitachi y Toshiba, y TSMC están desarrollando activamente variantes de tecnologías de chips MRAM. En vista del consumo de energía y la velocidad, MRAM compite favorablemente que otras memorias existentes como DRAM y Flash, con un tiempo de acceso de unos pocos nanosegundos, aunque sus limitaciones en la escritura (hacen falta cargas elevadas para crear el campo magnético en el *free layer*) y en la densidad hacen que esta tecnología se limite a aplicaciones muy especializadas.

3.1.1. STT-RAM

STT-RAM(Spin-Transfer Torque Random Access Memory o STT-MRAM es una tecnología de memoria magnética que utiliza la base establecida por la memoria MRAM para conseguir una NVM escalable, con reducidos tiempos de lectura y escritura, reducido tamaño de celda, y compatibilidad con tecnologías existentes como la DRAM y SRAM [5]. Esta tecnología utiliza el mismo modo de almacenar los bits de información

que MRAM, cada celda posee dos elementos magnéticos separados por una lámina aislante y dependiendo de si la polaridad de los dos campos magnéticos es paralela o antiparalela se obtienen los dos estados de almacenaje, gracias al MTJ. La novedad de esta tecnología se encuentra en la forma de cambiar la polarización del elemento libre del sistema, en este caso se utilizan fuerzas magnéticas denominadas spin transfer torque o STT, provocadas por corrientes *spin polarized* en elemento fijo. Esto permite tamaños de celda más pequeños, así como consumos menores en las escrituras.

La estructura de celda más prometedora por el momento consiste en el elemento de memoria descrito anteriormente conectado en serie a un transistor NMOS. Este transistor es controlado por la línea de palabra o WL (Word Line) y permite acceder a la información almacenada en la MTJ. Las operaciones de lectura y escritura se harían de la siguiente forma [10]:

- **Lectura:** Cuando el transistor de acceso es activado y una corriente eléctrica pasa entre la línea de bit y la fuente. Esta corriente debe tener un voltaje negativo y pequeño, para no ocasionar una escritura no deseada en la celda. Esta corriente pasa por un amplificador de señal y es comparada con una corriente de referencia para verificar el estado de la celda. Si la corriente es pequeña (resistencia de MTJ grande), se considera un "0". Si la corriente es grande, se considera un "1" almacenado.
- **Escritura:** Esta operación es realizada cuando una corriente con un voltaje suficiente pasa entre la línea de bit y la fuente. Este voltaje viene dado por el material por el que están compuestos los elementos magnéticos, entre otros factores. Cuando un voltaje positivo es aplicado, se escribirá un 0, en cambio si el voltaje es negativo se escribirá un 1.

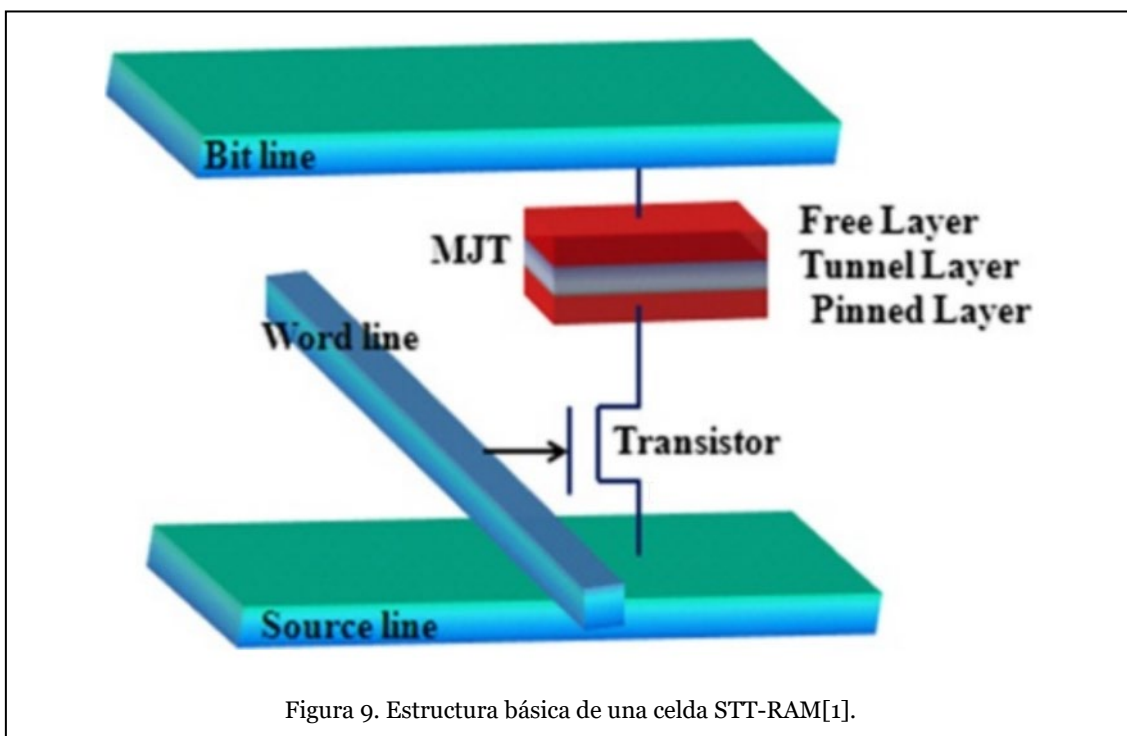


Figura 9. Estructura básica de una celda STT-RAM[1].

Por esto, esta tecnología resulta una opción más atractiva para el futuro de la memoria MRAM, pudiendo ser utilizada en las siguientes generaciones de dispositivos

semiconductores tales como microprocesadores, microcontroladores y SoC (Software on Chip).

Para que esta tecnología pueda considerarse una opción atractiva en el mercado actual aún deben resolverse ciertos aspectos clave en su funcionamiento tal y como está en la actualidad. Se debe conseguir un consumo bajo a la hora de escribir datos, y una alta estabilidad térmica para su correcto funcionamiento. Además, la diferencia en resistencia eléctrica entre los dos estados por el momento es relativamente pequeña, lo que puede resultar en problemas a la hora de determinar el estado de la celda. Otro aspecto a tener en cuenta sería que cualquier campo magnético externo relativamente fuerte podría cambiar los valores de las celdas, ocasionando la pérdida de los datos [4].

La tendencia actual para mejorar esta tecnología se basa en cambiar la polaridad de los dos elementos magnéticos para que sean perpendiculares el uno con el otro, en vez de paralelos. Este cambio reduce la corriente necesaria, mejora la escalabilidad y aumenta el tiempo de retención de los datos.

Actualmente, el tamaño de celda STT-RAM más pequeño que se ha conseguido es $0,26 \mu\text{m}^2$, lo que la hace 150 veces más grande que la celda NAND más pequeña hoy en día. En cuanto a tiempos de lectura y escritura STT-RAM se coloca como la más prometedora al implementar celdas con doble MTJ, consiguiendo tiempos menores que 10ns para ambas, pero aumentando considerablemente el tamaño de celda. La durabilidad de las celdas sería superior a 10^{16} ciclos, mientras que el consumo requerido por STT-RAM se reduce cada vez que el tamaño de celda disminuye, lo que hace que por el momento consiga consumos 90% inferiores al consumo de una memoria SRAM convencional.

Diversas compañías como Everspin, Hitachi, IBM, Qualcomm, Samsung y Toshiba están investigando formas de trabajar con esta tecnología y desarrollando prototipos, algunas incluso poseen dispositivos en el mercado, como Everspin y Qualcomm.

3.2. FeRAM

Memoria RAM Ferroeléctrica o FeRAM es una nueva tecnología de memoria no volátil que utiliza una composición de celda parecida a la de la memoria DRAM, pero usando un material ferroeléctrico en vez de uno dieléctrico entre las placas del condensador. Cuando se aplica un campo eléctrico a materiales dieléctricos, el material se polariza, y normalmente perdería la polarización una vez eliminado el campo eléctrico, pero con materiales ferroeléctricos, al

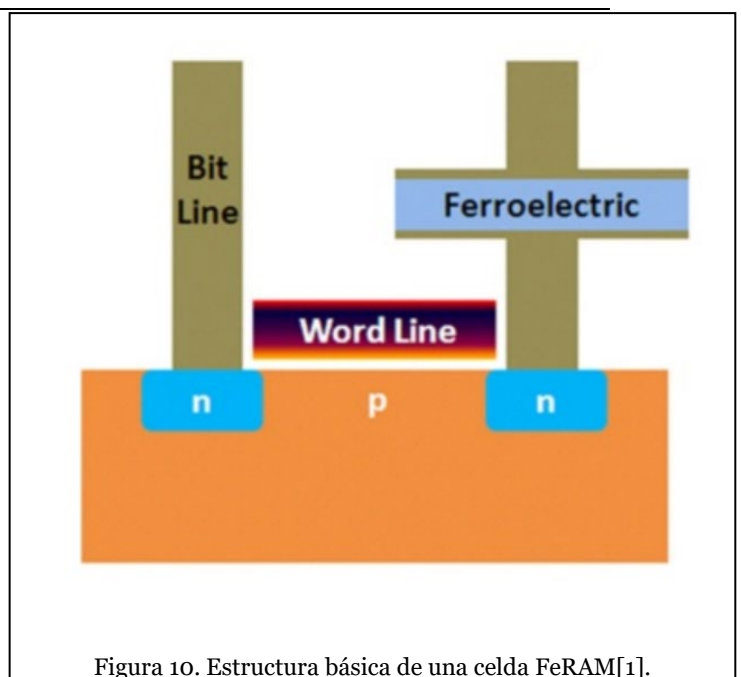


Figura 10. Estructura básica de una celda FeRAM[1].

aplicar el campo eléctrico exhiben histéresis y al eliminar dicho campo la polarización se mantiene, pudiendo almacenar así la información de forma indefinida, sin recurrir a corrientes de refresco como la DRAM. Pese al nombre de esta tecnología el material que se utiliza en esta no es el hierro, materiales como el titanato de circonato de plomo (PZT) son usados en la actualidad como material ferroeléctrico, aunque otros materiales están siendo considerados para este rol.

Una de las desventajas de esta tecnología es que el ciclo de lectura es destructivo, es decir, al leer un bloque de datos estos son eliminados de la memoria, y es necesario volver a escribirlos si se desea mantenerlos. Esto se da porque el método de lectura es escribir un bit en cada celda [1]. Si el estado de la celda cambia, un pulso débil será detectado indicando que la celda contenía un 0 lógico. Esta desventaja no supone una reducción de la vida útil del dispositivo, ya que esta tecnología soportaría un gran número de ciclos de lectura. Otra desventaja sería que los materiales ferroeléctricos empleados para fabricar las celdas de esta tecnología pueden ser degradados por los procesos de manufactura utilizados en la industria actual.

La tecnología FeRAM en la actualidad ha alcanzado velocidades mayores que la Flash, aunque aún por debajo de la DRAM, por lo que se espera que reemplace a esta en diversos dispositivos como PDA, teléfonos móviles, sistemas de seguridad y dispositivos wireless, y los dispositivos que se han sacado al mercado en los últimos años poseen un rendimiento comparable si no superior a dispositivos Flash con características similares.

Actualmente, FeRAM ha alcanzado tamaños de celda en torno a los 130nm^2 , y posee una durabilidad superior a 10^{15} ciclos de lectura/escritura. En cuanto a tiempos de lectura y escritura, FeRAM alcanza los 20ns y 50ns como velocidades máximas, respectivamente, aunque el tiempo de lectura puede ser más lento.

Esta tecnología resulta muy interesante para aplicaciones que requieran voltajes reducidos para un funcionamiento óptimo o aplicaciones que requieran sobrescribir los datos de forma frecuente, como por ejemplo monitorización medioambiental, seguimiento inalámbrico, dispositivos de registro de datos, seguimiento logístico, entre otras. Empresas como Texas Instruments o Fujitsu poseen dispositivos FeRAM en el mercado en la actualidad

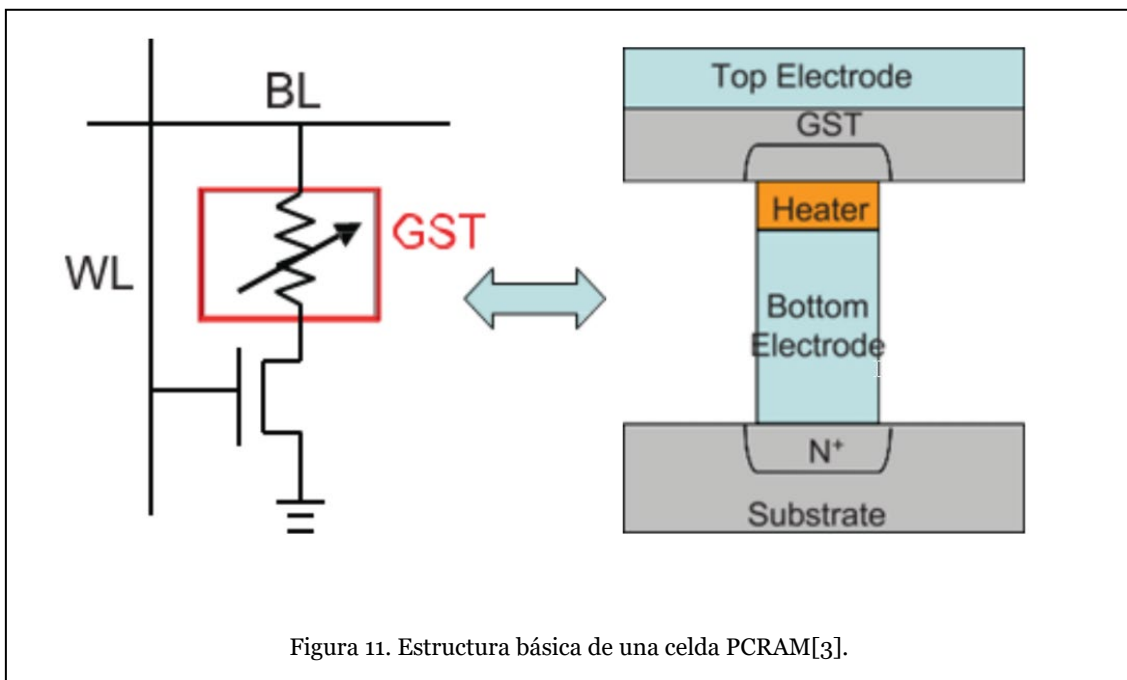
3.3. PCRAM

También conocida como chalcogenide RAM, es un tipo de memoria RAM no volátil, basada en un material denominado calcogenuro que puede encontrarse en dos estados diferentes. El material más prometedor por el momento es el GST, una aleación compuesta por germanio, antimonio y telurio. Estos materiales también son utilizados en la capa que almacena los datos en los discos CD-RW y DVD-RW. La estructura de este material puede cambiar rápidamente entre cristalina y amorfa a una escala microscópica. La estructura cristalina posee muy poca resistencia a la electricidad, mientras que esta resistencia aumenta considerablemente cuando se encuentra en su estructura amorfa, y este cambio en la resistencia hace posible la representación de datos según la resistencia en un punto concreto de material.

Esta tecnología tiene el potencial para proveer almacenamiento de alta densidad, con una tasa de transferencia de datos muy alta por un precio reducido. Además para escribir en una celda no es necesario borrar los datos, y las celdas tendrían una durabilidad superior a la de la memoria flash.

A parte del calcogenuro, en una estructura básica de celda, PCRAM tiene un transistor NMOS en serie con este, tal y como muestra la figura 11. Las operaciones de lectura y escritura se realizan de la siguiente forma [11]:

- **Lectura:** Para leer los datos almacenados en el GST, es aplicado un pequeño voltaje, y dependiendo del estado en el que se encuentra el material se leería un voltaje distinto en la BL.
- **Escritura:** En esta tecnología son necesarias dos operaciones de escritura:
 - **SET:** En esta operación se realiza el cambio de amorfo a cristalino del calcogenuro. Esta transición se realiza enviando una corriente que calienta el material hasta su temperatura de cristalización durante un periodo de tiempo determinado por el tipo de material.
 - **RESET:** Esta operación realiza el cambio del calcogenuro de cristalino a amorfo. Esta transición se realiza enviando una corriente eléctrica mayor que calienta el material sobrepasando el punto de fusión del material, y una vez se vuelve a enfriar, el material se solidifica en estado amorfo.



El principal problema que tiene que afrontar esta tecnología es el elevado consumo necesario para borrar las celdas, problema que se espera resolver al reducir el tamaño de estas.

Para intentar afrontar las limitaciones inherentes a su tecnología, en la actualidad el esquema de celda ha evolucionado a un esquema 3D XPoint, que será

descrito en el siguiente apartado. Dicho esquema permite desempeños al nivel de la DRAM con un coste más bajo que esta, siendo una alternativa más atractiva y viable para un sistema de almacenamiento que combine una velocidad similar a la DRAM con la persistencia de los datos. Además, se está estudiando la viabilidad de nuevos materiales que posean mejores características que el GST para este tipo de aplicaciones. Entre ellos se encuentra el $\text{Cr}_2\text{Ge}_2\text{Te}_6$, que se dice que puede reducir el consumo para el cambio de fase un 90%.

En la actualidad, las celdas de memoria PCRAM tienen un tamaño de $0.0069 \mu\text{m}^2$, lo que las hace tan solo 5 veces más grandes que la celda Flash más pequeña [12]. En cuanto a tiempos de lectura y escritura, PCRAM tarda menos de 20ns en realizar una lectura, mientras que tarda mínimo 20/30 ns, aunque el tiempo de escritura aumenta conforme el tamaño de celda se reduce, haciendo que la operación de escritura en un dispositivo 8Gbit PCRAM sea de 150ns. La durabilidad de PCRAM es superior a las 10^{12} operaciones, y esta tecnología necesita en torno a 9-10V para realizar una operación de RESET, mientras que la operación SET requeriría 7-8 V.

Diversos prototipos que utilizan PCRAM han sido desarrollados por empresas como IBM, Infineon, Macronix y Samsung, entre otras. Además, la producción de dispositivos PCRAM por empresas como Intel, Samsung y STMicroelectronics ha sido anunciada en los últimos años.

3.4. RRAM

RRAM o RAM resistiva es un tipo de NVM con un funcionamiento similar a PCRAM, utilizando el cambio en la resistencia de un material dieléctrico de estado sólido. Esta tecnología resulta interesante ya que es compatible con los procesos de manufactura de semiconductores actuales. En concreto, la RRAM basada en dispositivos memristor es una de las tecnologías con más potencial, debido a su coste reducido y a la simplicidad de la celda. Esta consiste en un transistor NMOS conectado en serie a un biestable MIM (metal-insulator-metal) de dos terminales, el cual puede alternar entre dos estados de conductividad distintos, inducidos en este con la aplicación de distintos voltajes a través de sus terminales [13].

En comparación con otros sistemas de almacenamiento explicados en este documento, esta tecnología utiliza cambios en la estructura cristalina del material situado en el centro del biestable (generalmente un óxido metálico), en concreto mediante la redistribución de las vacantes de oxígeno del material, tal y como se muestra en la Fig. 12. Un estado de alta resistencia (HRS) equivaldría a un 0, mientras que un estado de baja resistencia (LRS) equivaldría a un 1.

La configuración de celda de la figura 13 es conocida como 1T1R-RRAM (1Transistor-1Resistencia-RRAM). La estructura 1T1R es muy compacta y puede permitir apilar verticalmente las capas de memoria, lo cual la hace ideal para dispositivos de almacenamiento masivo [15]. Aunque también se ha observado que al operar matrices RRAM muy grandes es necesario utilizar un dispositivo “selector”(un diodo, por ejemplo) como elemento aislante, lo que limita su escalabilidad por el momento.

Entre todas las candidatas, se considera que la RRAM es la más prometedora, ya que funciona más rápido que la PCRAM y tiene una estructura de celda más simple y más pequeña que las memorias magnéticas, haciendo posible densidades muy altas.

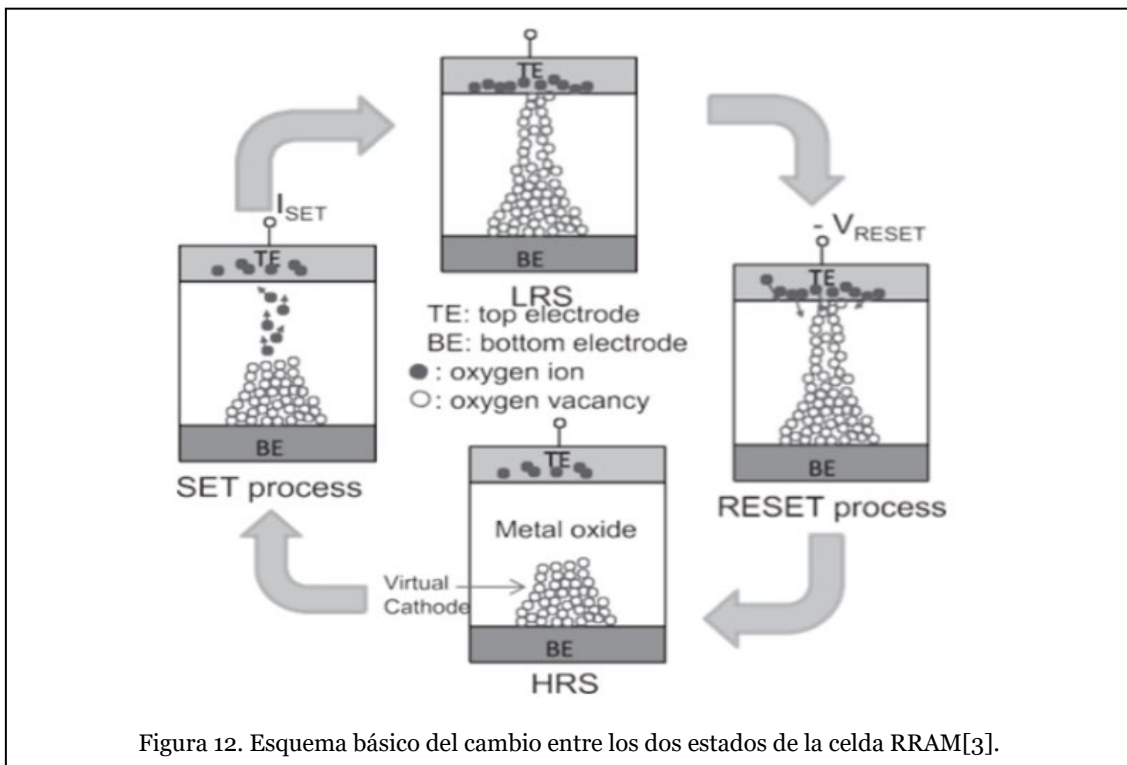


Figura 12. Esquema básico del cambio entre los dos estados de la celda RRAM[3].

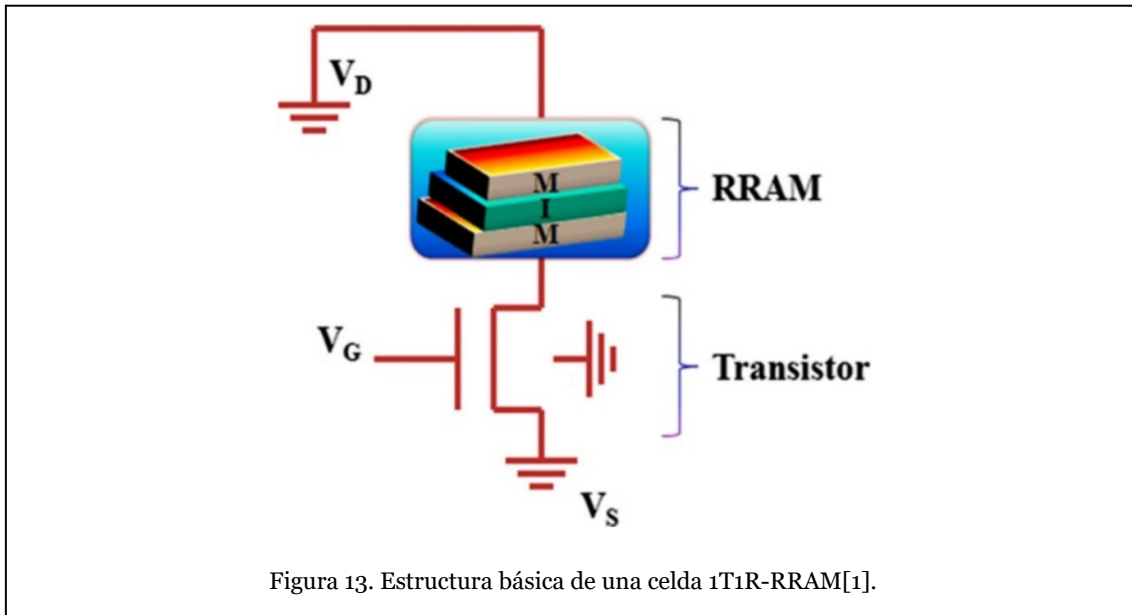


Figura 13. Estructura básica de una celda 1T1R-RRAM[1].

Aunque es un elemento de memoria más prometedor, los problemas críticos para el futuro desarrollo de dispositivos de RRAM son la retención de datos y la resistencia de la memoria, y se requiere de más esfuerzos de investigación así como estudios estadísticos de confiabilidad y mantenibilidad de esta tecnología para que esta

tecnología alcance todo su potencial. También se ha observado en las pruebas y prototipos que la interfaz usada para acceder y ordenar el espacio de memoria desempeña un papel muy importante en el rendimiento de la RRAM.

El desarrollo de ReRAM de alta densidad se ha visto limitado por la falta de un buen dispositivo selector. Sin embargo, los avances recientes en la memoria 3D XPoint parecen resolver este cuello de botella y ReRAM podría progresar rápidamente en un futuro cercano si se logra resolver problemas como los bits erráticos.

En la actualidad, las celdas de memoria RRAM tienen un tamaño de $0.0076 \mu\text{m}^2$, lo que las hace tan solo 6 veces más grandes que la celda Flash más pequeña. En cuanto a tiempos de lectura y escritura, RRAM tarda menos de $10 \mu\text{s}$ en realizar una lectura, mientras que tarda en torno a 10 ns en realizar una escritura, aunque, como sucede con PCRAM, el tiempo de escritura aumenta conforme el tamaño de celda se reduce, haciendo que la operación de lectura en un dispositivo 32 Gbit RRAM sea de 23 ns . La durabilidad de RRAM es superior a las 10^6 operaciones, y esta tecnología necesita menos de 3 V para funcionar de manera óptima.

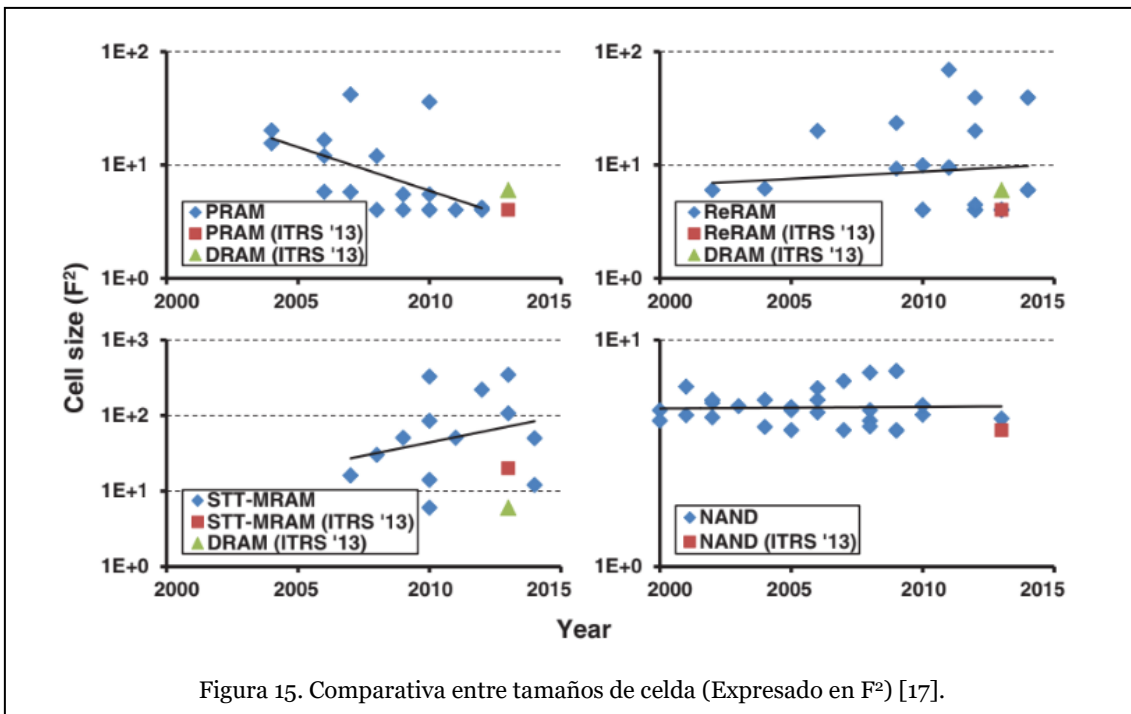
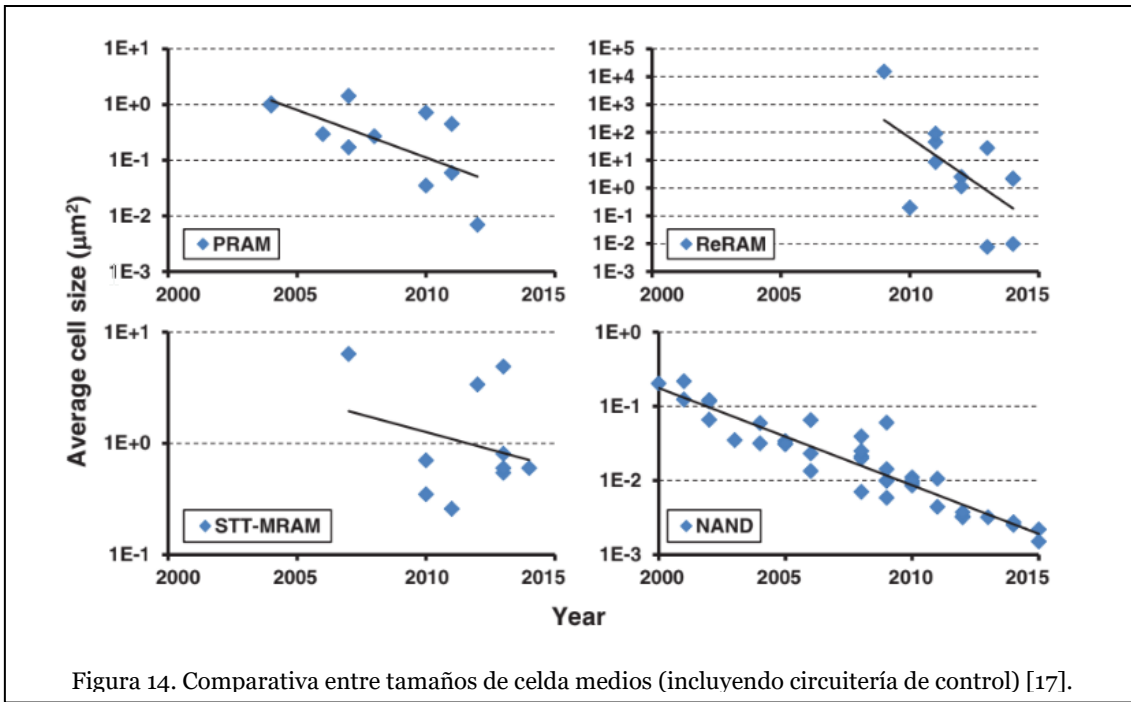
Diversos prototipos que utilizan RRAM han sido desarrollados por empresas como 4DS Memory, Adesto Technologies, Crossbar, Fujitsu, Sandisk y Panasonic. Además, la producción de dispositivos PCRAM por empresas como Sandisk, 4DS Memory y Panasonic ha sido anunciada en los últimos años.

3.5. Comparativa

Este apartado está dedicado a realizar una comparativa entre las tecnologías de memoria no volátil emergentes más prometedoras descritas en los puntos anteriores y las tecnologías SRAM, MRAM y Flash. En concreto, voy a centrarme en las tecnologías STT-RAM, PCRAM y RRAM, ya que son las tecnologías que la comunidad científica piensa que son más prometedoras, y por consecuencia tienen más presencia en estudios sobre tecnologías de memoria no volátil emergentes [17,16].

3.5.1. Tamaño de celda

En las dos siguientes graficas (Fig. 14 y 15) puede apreciarse la evolución del tamaño de celda de las NVM frente a la tecnología NAND Flash. En ellas se ve como PCRAM y RRAM alcanzan tamaños de celda parecidos en los últimos años (con circuitería de control y sin circuitería de control), pero aún no consiguen reducir el tamaño por debajo de una celda NAND Flash, mientras que STT-RAM se mantiene un orden de magnitud por encima de la competencia si se tiene en cuenta los circuitos de control, mientras que logra tamaños de celda muy similares sin ellos.



En la última grafica (Fig. 16) se compara el tamaño de celda, la capacidad del chip y el tamaño de la tecnología de fabricación de los semiconductores, además muestra el dispositivo selector que ha sido utilizado en la fabricación del chip. En ella podemos apreciar que RRAM consigue más densidad en el chip utilizando el mismo tamaño de celda y la misma tecnología de fabricación de semiconductores que PCRAM, y que el tamaño de STT-RAM no consigue reducirse más allá de $10F^2$. Cabe destacar que en la mayoría de chips que tienen un tamaño de celda inferior a $10F^2$ un circuito selector MOSFET ha resultado inviable y se han recurrido a otras técnicas de selección como transistores verticales o celdas auto-seleccionables.

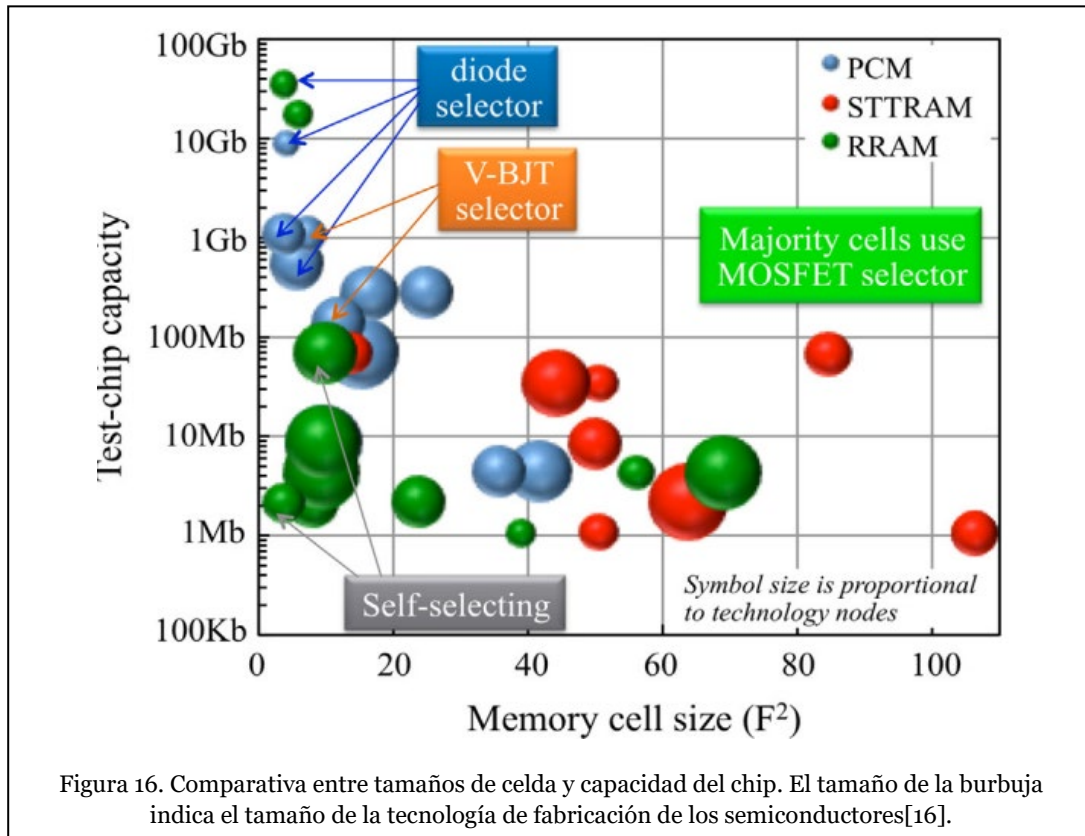


Figura 16. Comparativa entre tamaños de celda y capacidad del chip. El tamaño de la burbuja indica el tamaño de la tecnología de fabricación de los semiconductores[16].

3.5.2. Densidad

En la siguiente tabla (fig. 17) muestra una comparativa entre la densidad media de cada tecnología y el año en el que se fabricó. En ella podemos apreciar como RRAM y PCRAM, pese a conseguir tamaños de celda parecidos a NAND Flash, aún se encuentran un orden de magnitud por debajo en cuanto a densidad de bit en el chip. Mientras tanto, STT-MRAM se encuentra 2 órdenes de magnitud por debajo de sus competidores.

La Fig. 18 muestra los prototipos de chip desarrollados desde 2003 a 2015, y los compara por capacidad del chip y por el tamaño de la tecnología de fabricación de los semiconductores que lo componen. Además, también muestra la empresa fabricante del chip. En la tabla se puede apreciar como PCRAM fue la primera tecnología en ser prototipada, y ha variado de tamaño desde Mb a Gb. En ella también se aprecia como la capacidad de STT-RAM es menor que el resto, debido a que sus aplicaciones no necesitan almacenamiento de alta densidad. RRAM es la tecnología que muestra más variabilidad, desde tamaño de celda hasta tamaño de la tecnología CMOS empleada, ya que las diferentes empresas que la desarrollan tienen en cuenta distintas aplicaciones para esta memoria.

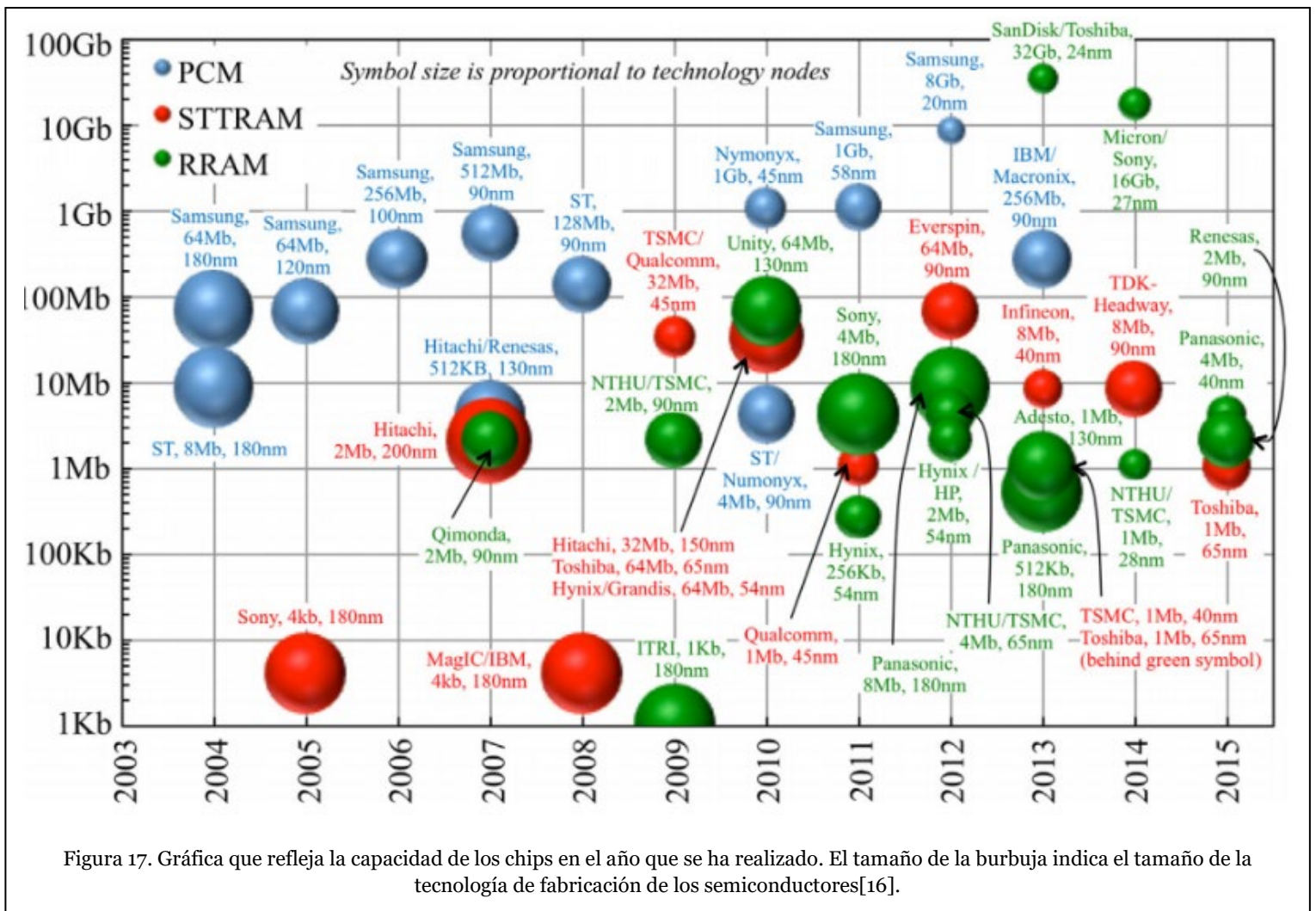


Figura 17. Gráfica que refleja la capacidad de los chips en el año que se ha realizado. El tamaño de la burbuja indica el tamaño de la tecnología de fabricación de los semiconductores[16].



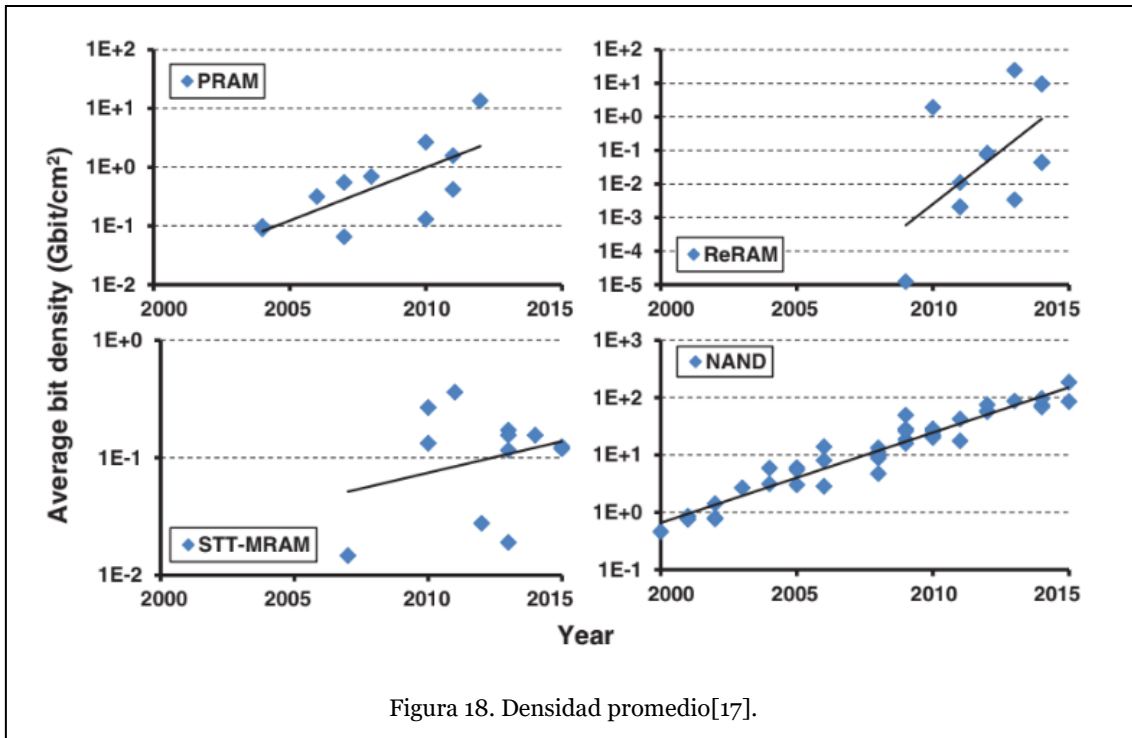


Figura 18. Densidad promedio[17].

3.5.3. Latencia de lectura

Los tiempos de lectura de la Fig. 19 muestran el tiempo requerido para leer una celda, pero no incluyen la latencia ocasionada por la circuitería de control que las opera. En estas tablas se pueden apreciar como los tiempos de lectura de PCRAM y RRAM tienden a aumentar a medida que el tamaño de celda disminuye y la densidad aumenta. En el caso de STT-MRAM los tiempos de lectura se han visto reducidos por el uso de celdas con doble MTJ. En este aspecto todas las memorias que aparecen en las gráficas muestran latencias de lectura inferiores a NAND Flash, y en el caso de STT-MRAM, su latencia de lectura es inferior incluso a la de DRAM.

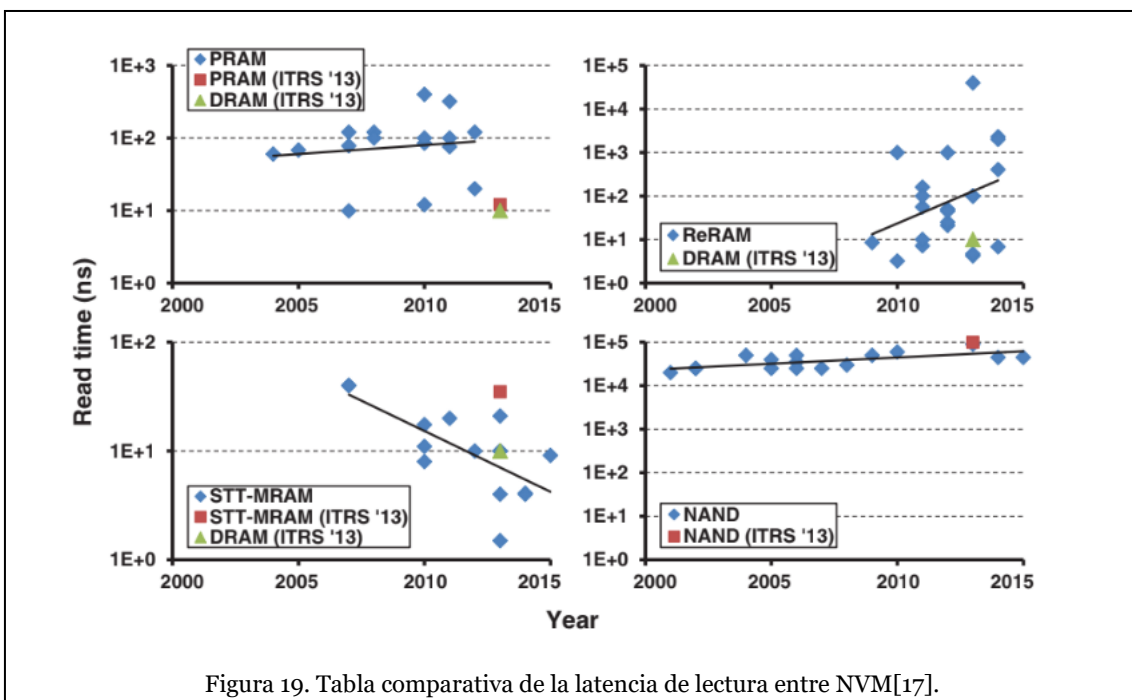


Figura 19. Tabla comparativa de la latencia de lectura entre NVM[17].

3.5.4. Latencia de escritura

En el caso de la latencia de escritura sucede lo mismo que con la de lectura. Pese a ser más densa, RRAM tiene peores tiempos de escritura de media que PCRAM. Los tiempos de RRAM suelen ser 10 veces más lentos que PCRAM. STT-RAM tiene muy buenos tiempos de escritura pero su capacidad de chip es muy baja. Las tres tecnologías se sitúan por encima de la NAND Flash, y solo STT-RAM logra tiempos parecidos a la DRAM.

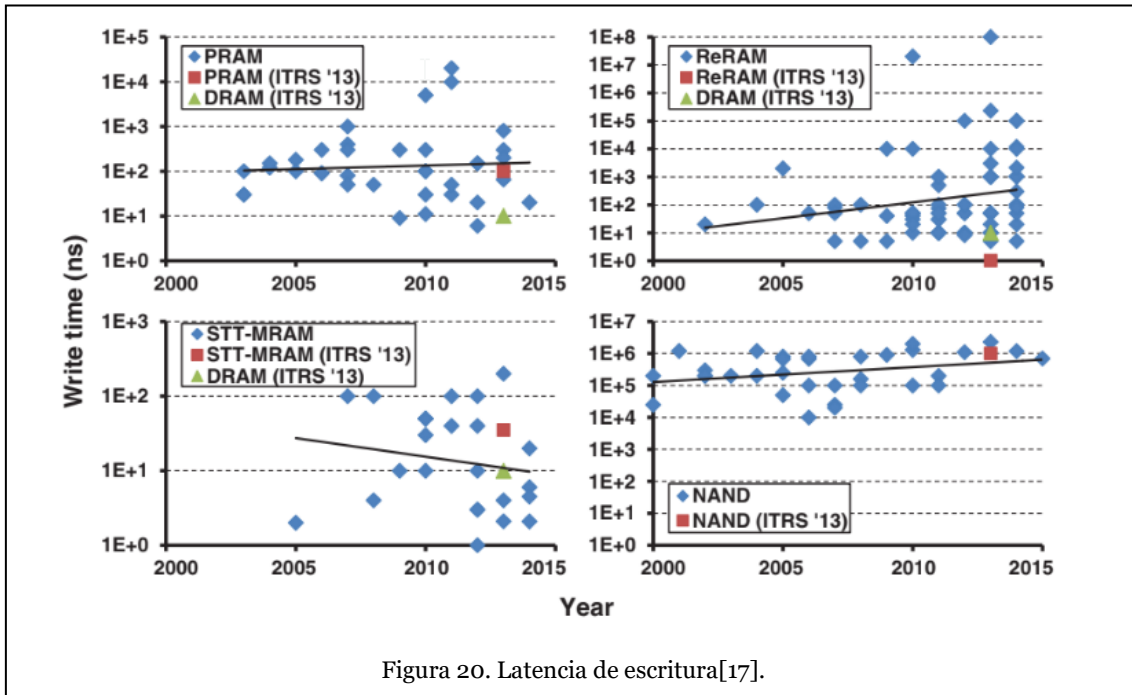


Figura 20. Latencia de escritura[17].

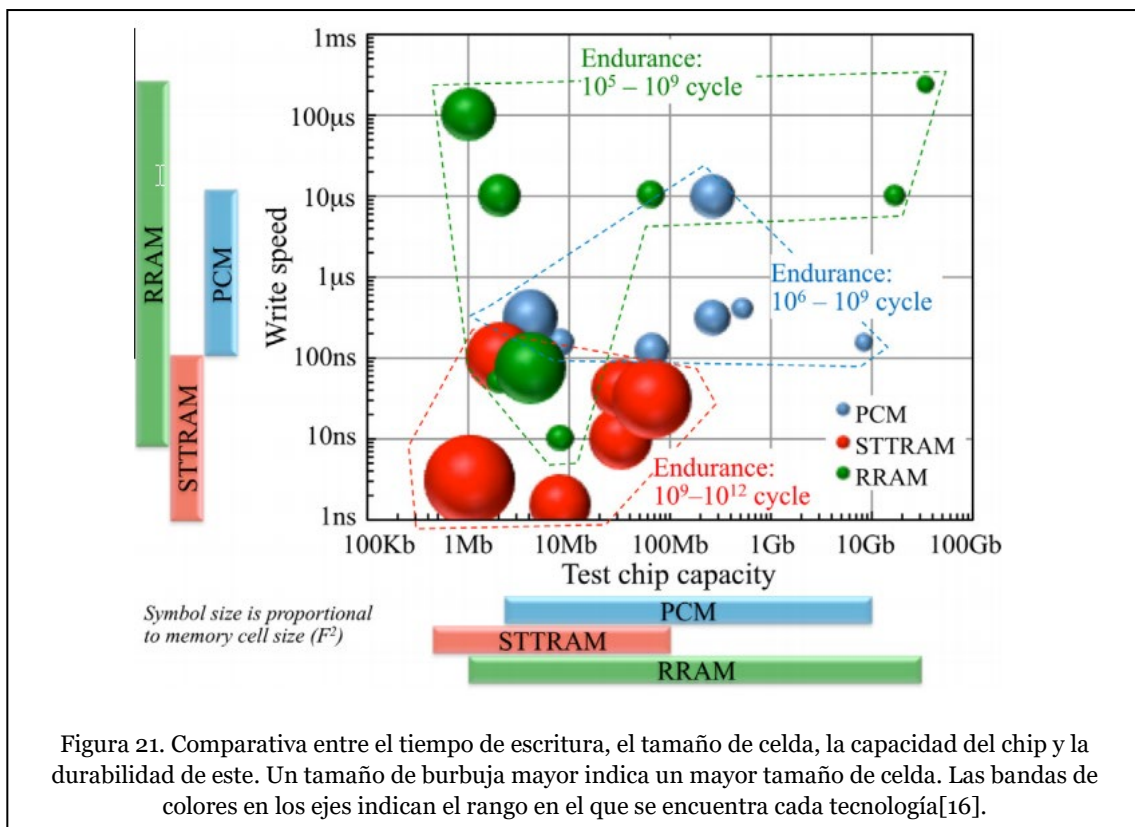


Figura 21. Comparativa entre el tiempo de escritura, el tamaño de celda, la capacidad del chip y la durabilidad de este. Un tamaño de burbuja mayor indica un mayor tamaño de celda. Las bandas de colores en los ejes indican el rango en el que se encuentra cada tecnología[16].



La Fig. 21 refuerza lo que se muestra en la gráfica anterior de una forma más clara. En esta se puede apreciar cómo se logran mejores tiempos con tamaños de celda mayores y menor capacidad en el dispositivo, aunque no siempre es así. En el caso de STT-RAM, para conseguir semejantes tiempos es necesaria una corriente de conmutación alta, y por lo tanto, un tamaño de celda mayor. La mayor variabilidad en el rango de tiempos de lectura de RRAM demuestra que todavía es una tecnología inmadura.

3.5.5. Consumo

Un consumo reducido es un factor clave en los dispositivos de memoria actuales. Los dispositivos de memoria no volátil no tienen un consumo en reposo ya que se pueden mantener apagados sin provocar pérdida de datos. Su consumo es determinado por la corriente y el voltaje necesarios para que se produzca una escritura. El voltaje de las NVM emergentes oscila entre los 0.5 y los 5 V. El factor clave para realizar la comparación es la corriente necesaria para la escritura. En la Fig. 22 se puede observar una tabla que relaciona la corriente de escritura con el año en el que fue desarrollado el chip y su tamaño de celda. En ella podemos observar como la tendencia de todas las tecnologías es reducir su consumo. En ella también se puede observar como RRAM consigue mejores consumos con un tamaño de celda más pequeño.

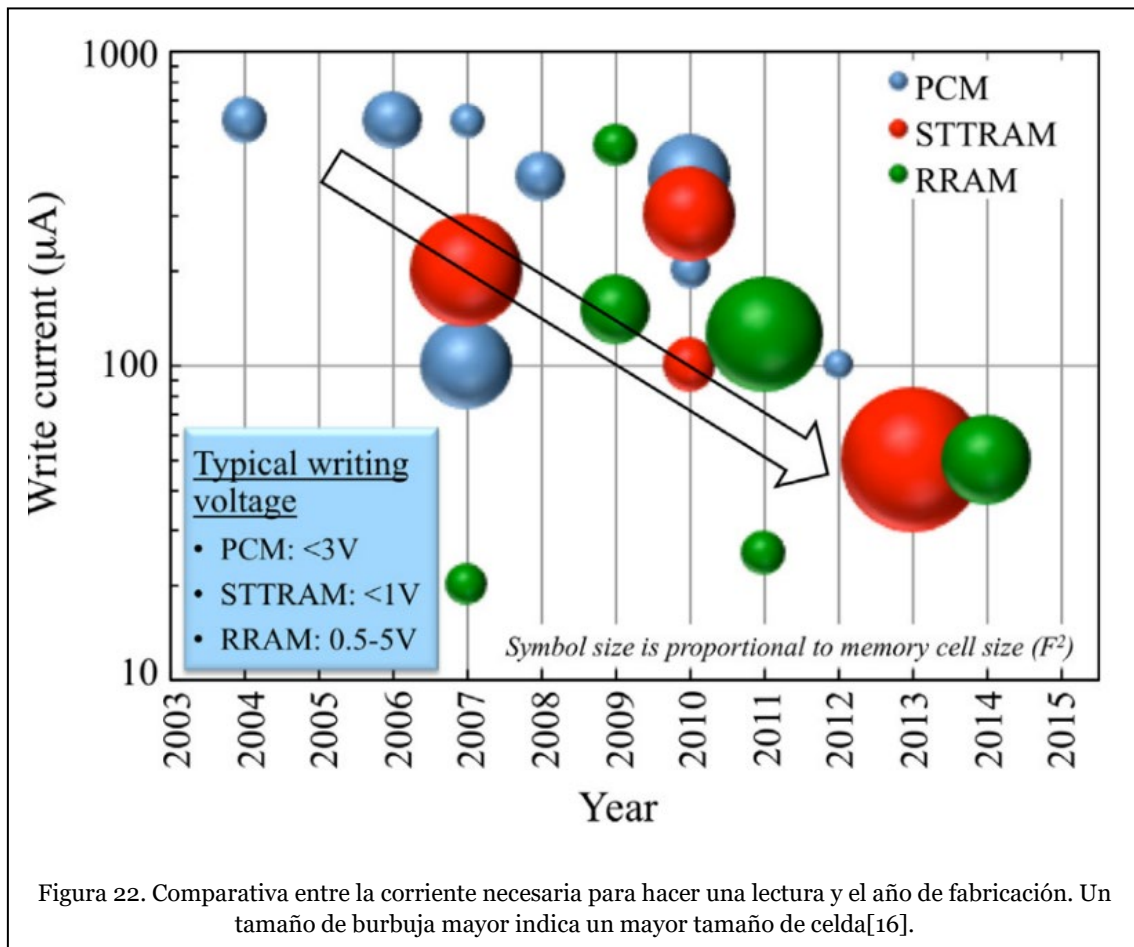


Figura 22. Comparativa entre la corriente necesaria para hacer una lectura y el año de fabricación. Un tamaño de burbuja mayor indica un mayor tamaño de celda[16].

3.5.6. Conclusiones

En este apartado se ha intentado reflejar las diferencias entre las tecnologías NVM emergentes más prometedoras mediante tablas y gráficas. En ellas se puede observar cómo, pese a ser la tecnología más inmadura de las tres que se están comparando, RRAM resulta ser la más prometedora. Con los mejores resultados en cuanto a densidad, consumo, capacidad, y tamaño de celda, resulta una opción muy atractiva para integrar a toda clase de productos presentes en nuestro día a día. Además, si se consigue desarrollar un dispositivo selector óptimo para esta tecnología se podrían reducir los tiempos de lectura y escritura manteniendo un tamaño de celda reducido, y quizás también se logre aumentar su durabilidad y reducir la variabilidad sin realizar cambios significativos a la tecnología en sí. La Fig. 23 es una tabla resumen que cuantifica aproximadamente las características generales de SRAM, DRAM y Flash así como las tecnologías NVM emergentes que se han definido en apartados anteriores, con el fin de facilitar la comparación entre estas.

	SRAM	DRAM	Flash	FeRAM	MRAM	STT-RAM	RRAM	PCRAM
Tam. celda	>100F ²	6 - 10F ²	4 - 10F ²	15 - 40F ²	25F ²	6 - 50F ²	4 - 12F ²	4 - 30F ²
Densidad chip	128Mb	8-16Gb	256-512Gb	128Mb	16-32Mb	2-64Mb	256Kb-32Gb	8Gb
Durabilidad	10 ¹⁶	10 ¹⁶	10 ⁴	10 ¹⁵	10 ¹⁵	10 ¹⁵	10 ⁶ - 10 ¹²	10 ⁹
T. Lectura	1 - 2ns	10ns	10µs	<10ns	<10ns	<10ns	<10ns	10ns
T. Escritura	1 - 2ns	10ns	100µs	10ns	<20ns	10ns	10ns	50ns
Energía por cada escritura	0.2pJ	2pJ	1000mJ	2pJ	120pJ	0.02pJ	2pJ	100pJ
Voltaje	<1V	<1V	>10V	2V	<1.5V	<1.5V	<5V	<3V
Volatilidad	SI	SI	NO	NO	NO	NO	NO	NO
Retencion	Volátil	Volátil	>10 años	>10 años	>10 años	>10 años	>10 años	>10 años

Figura 23. Tabla resumen indicando las características generales de cada NVM[1,2,3,4,9,18].

3.6. Arquitectura de chip

3.6.1. 1T1R Array

Es una de las arquitecturas más comunes para las NVM emergentes. En este diseño, cada celda está conectada en serie con un transistor que actúa como selector de la celda. La línea de palabra (WL) controla la puerta del transistor, por tanto, al ajustar el voltaje de WL se puede controlar la corriente de escritura que recibe la celda. El electrodo superior de la celda NVM se conecta a la línea de bit (BL), y el electrodo inferior se conecta al drenaje del transistor. La fuente del transistor es conectada a la

línea de fuente (SL). El área de celda típica del 1T1R array es $12 F^2$ (F es el tamaño de la función de litografía) si el ancho y el largo de la puerta del transistor es uno. El área mínima de la celda se puede reducir a $6 F^2$ si se realizan mejoras como compartir SL en dos columnas, entre otras [18]. El área de la celda aumentaría si un transmisor de tamaño mínimo no puede proporcionar suficiente corriente de escritura (en el caso de STT-MRAM y PCRAM, por ejemplo). en la Fig. 24 se detallan las operaciones de SET, RESET y READ:

- SET: para esta operación, se aplica voltaje sobre WL para encender el transistor de la celda seleccionada, y un voltaje de escritura es aplicado a la BL de la celda mientras SL está conectado a tierra.
- RESET: para la operación de reset se aplica voltaje sobre WL para activar el transistor de la celda seleccionada, y también se aplica voltaje de escritura sobre SL mientras Bl está conectado a tierra.
- READ: para realizar una lectura se aplica voltaje sobre la WL de la celda seleccionada y se aplica un voltaje de lectura sobre BL mientras SL está conectado a tierra. se amplifica la señal y después se compara con las señales de referencia para saber si la celda se encuentra en HRS o LRS.

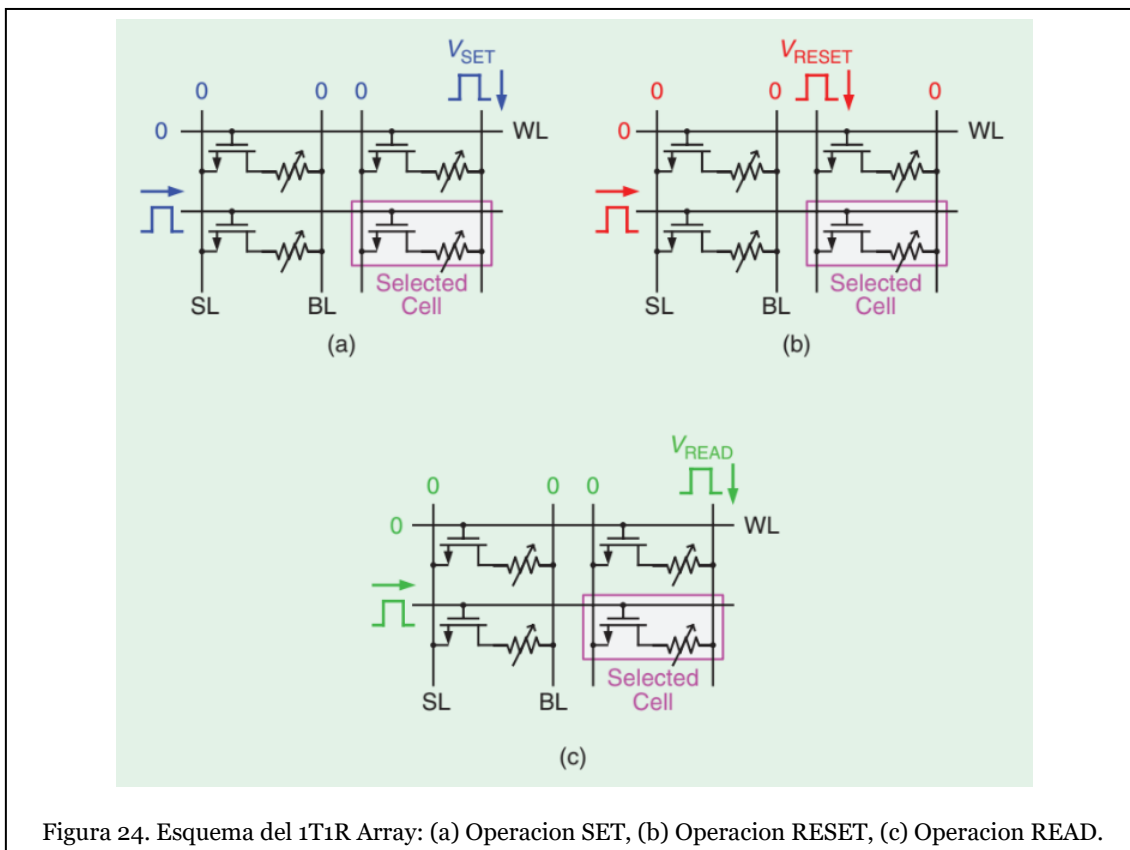


Figura 24. Esquema del 1T1R Array: (a) Operación SET, (b) Operación RESET, (c) Operación READ.

Debido a que los transistores están apagados para las celdas no seleccionadas, no hay problemas de interferencia o interferencia en la matriz 1T1R, y se puede acceder a cada celda de forma independiente y aleatoria. Se pueden escribir (o leer) varios bits en paralelo en (o desde) la misma fila activando múltiples columnas.

Esta arquitectura enfrenta desafíos de escalabilidad si la corriente de escritura de las NVM no puede escalar en consecuencia. Para proporcionar la corriente necesaria

para que la escritura de la celda se realice es necesario utilizar un transistor más grande que el de tamaño mínimo por el momento, por lo tanto es poco probable que se llegue a conseguir un tamaño de celda de $6F^2$ o inferior teniendo en cuenta la corriente de escritura necesaria actualmente. Además, reducir el voltaje de escritura a menos de 1 V es necesario también para aplicaciones embebidas.

3.6.2. X-Point Array

Esta arquitectura consiste en filas y columnas perpendiculares entre sí con celdas NVM intercaladas entre ellas, como se muestra en la Fig. Cross-point, en principio, puede lograr un área de $4F^2$, consiguiendo una mayor densidad de integración que la matriz $1T1R$. En esta arquitectura los selectores se agregan en serie con las celdas NVM para evitar interferencias entre celdas, lo que se conoce como arquitectura de un selector y una resistencia ($1S1R$). Cross-point puede admitir PCRAM y RRAM, pero no admite STT-MRAM debido a que este tipo de celda tiene una relación de activación / desactivación demasiado pequeña para esta arquitectura [1]. Para programar con éxito las celdas NVM, se pueden aplicar dos esquemas de escritura ($V/2$ y $V/3$).

En el esquema $V/2$, para la operación SET, a la WL de la celda seleccionada se le aplica el voltaje de escritura V_w y la BL se conecta a tierra. Para la operación RESET, las conexiones mencionadas anteriormente para WL y BL se invierten. Tanto en las operaciones de SET como de RESET todos los WL y BL no seleccionados están sesgados en $V_w/2$. Por lo tanto, solo la celda seleccionada ve un V_w completo, mientras que las celdas semiseleccionadas a lo largo del WL o BL seleccionado ven la mitad de V_w y todas las otras celdas no seleccionadas en la matriz ven un voltaje muy cercano a cero

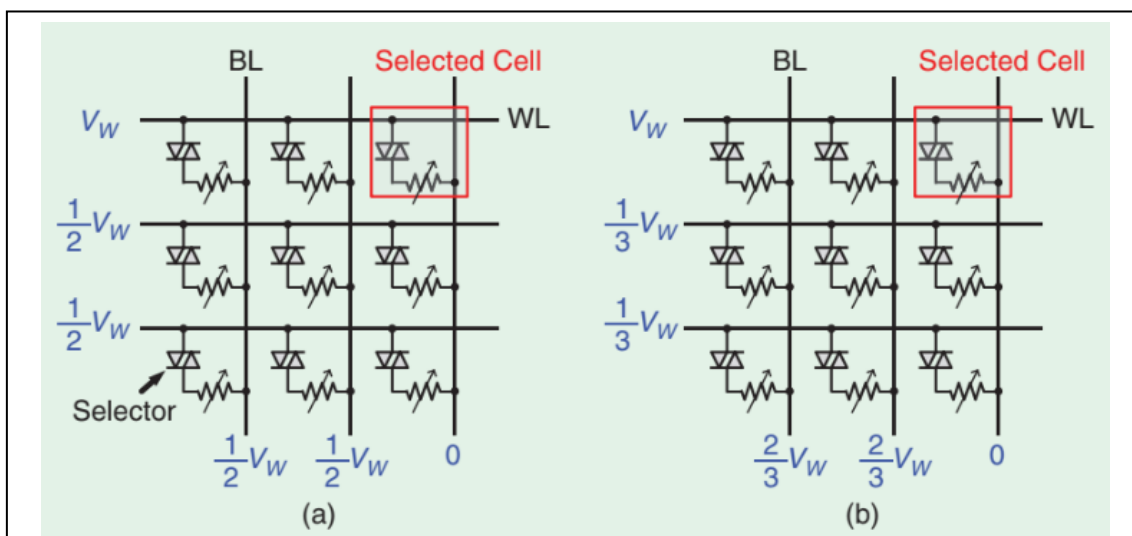


Figura 25. Esquema del X-Point Array: (a) Operación SET en el esquema $V/2$, (b) Operación SET en el esquema $V/3$.

En el esquema V / 3, para la operación SET, a la WL de la celda seleccionada se le aplica el voltaje de escritura V_w y la BL se conecta a tierra, igual que en el esquema V / 2. Lo mismo sucede en la operación RESET. La principal diferencia entre esquemas es que en el esquema V / 3, para la operación de SET, las WL y BL no seleccionadas están sesgadas a $1/3 V_w$ y $2/3 V_w$, respectivamente. Para la operación de RESET, las WL y BL no seleccionadas están sesgadas a $2/3 V_w$ y $1/3 V_w$, respectivamente. De esta manera, la celda seleccionada ve V_w , mientras que todas las demás celdas no seleccionadas en la matriz solo ven $1/3 V_w$.

Los pros y los contras de estos dos esquemas de escritura se pueden resumir de la siguiente manera: el esquema V / 2 generalmente tiene menos consumo de energía que el esquema V / 3. Esto se debe a que las celdas no seleccionadas (no a lo largo de WL y BL seleccionadas) en el esquema V / 2 ven idealmente voltaje cero, mientras que todas las celdas no seleccionadas en el esquema V / 3 ven $1/3 V_w$, consumiendo así energía durante el periodo de escritura. Por otro lado, el esquema V / 3 tiene mejor inmunidad a la perturbación de escritura que el esquema V / 2, ya que el voltaje máximo que ven las células no seleccionadas es $1/3 V_w$ en el esquema V / 3 mientras que es $1/2 V_w$ en el esquema V / 2.

Para realizar una lectura, se aplica todas las columnas el voltaje de lectura V_r , y solo la fila seleccionada está conectada a tierra. Por lo tanto, solo las celdas de la fila seleccionada ven un voltaje de lectura y todas las otras celdas no seleccionadas ven voltaje cero. Los dispositivos utilizados para realizar la amplificación y la detección de la señal se pueden clasificar en dos tipos: de voltaje y de corriente. En diseños prácticos, la elección entre detección de voltaje y detección de corriente depende del tamaño de la matriz y las características de la celda NVM. La conclusión general es que para una matriz con una BL larga o un LRS más alto, la detección de corriente proporciona un acceso más rápido.

La matriz de punto de cruce sufre de dos desafíos de diseño conocidos: 1) La caída de IR a lo largo del cable de interconexión y 2) El sneak path (cuando la corriente viaja por rutas no deseadas dentro de un circuito y sus interfaces externas) a través de las celdas no seleccionadas. El problema de la caída de IR se vuelve significativo cuando el ancho del cable WL y BL escala a un régimen por debajo de 50 nm donde la resistividad de interconexión aumenta drásticamente. Para garantizar una operación de escritura exitosa, el voltaje de escritura proporcionado por el controlador debe aumentarse sobre el voltaje de conmutación real de la celda NVM para compensar la caída de IR. Sin embargo, el voltaje de escritura no se puede aumentar demasiado porque $1/2 V_w$ (en el esquema V / 2) no debería alterar la resistencia NVM para las celdas cercanas al controlador.

El problema del sneak path está asociado con el problema de la caída de IR. Las celdas semiseleccionadas a lo largo de las líneas WL y BL seleccionadas forman los sneak paths durante la operación de escritura. Las rutas de fuga contribuyen con corriente adicional a la caída de IR y degradan aún más el margen de escritura. Mientras tanto, el sneak path aumenta la corriente de escritura (por lo tanto, la potencia de escritura) que proporcionan los controladores del cross-point array. Aumentar la resistencia LRS (o reducir de manera equivalente la corriente de escritura) y aumentar la no linealidad de la celda NVM (con la ayuda del selector) son las soluciones más viables para estos problemas.

3.6.3. 3D X-Point Array

La memoria 3D XPoint o 3D cross-point no es una NVM emergente per se, es más un esquema de organización 3D para otras tecnologías, pero considero que merece ser mencionada ya que tiene el potencial para que otras NVM explicadas anteriormente puedan superar algunas de sus limitaciones y consigan aumentar drásticamente su viabilidad. Se basa en una matriz de acceso a datos de cuadrícula cruzada apilable (cross-point memory array). El diseño básico de la celda consiste en un diodo en posición vertical en serie con un elemento de memoria (RRAM o PCRAM, principalmente). Esta configuración en vertical permite una reducción del tamaño del chip utilizado, obteniendo así mejoras en el retardo de conexión en la enrutación de bloques, así como permitiendo densidades mucho mayores y escalabilidad 3D[1].

Existen dos enfoques de integración 3D utilizando el esquema de cuadrícula cruzada. El primero consiste en apilar la matriz de cuadrícula cruzada convencional por capas. El segundo se basa en una estructura de pilares compuestos de celdas de memoria y electrodos conectados al plano de palabra, como se muestra en la Fig. 30 (b)[18].

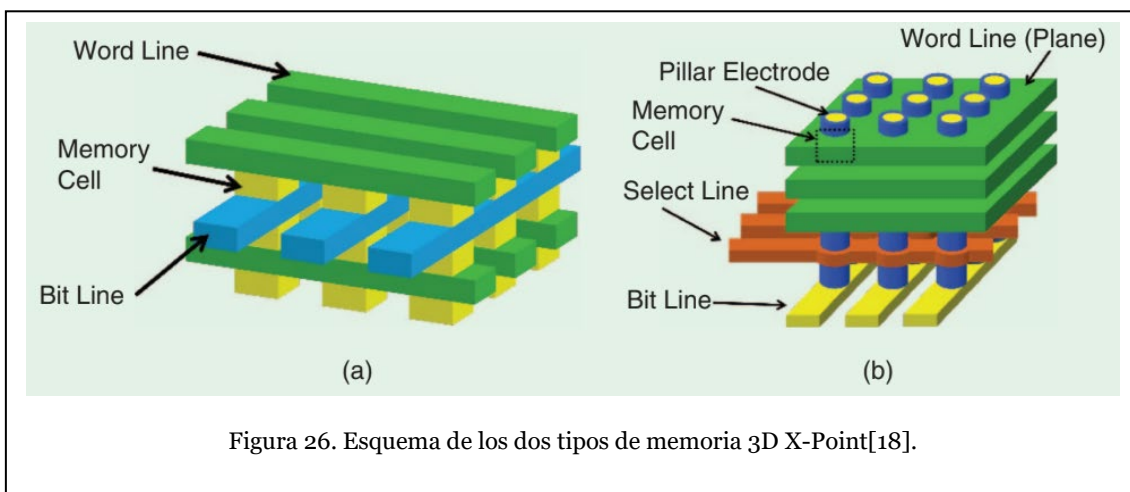


Figura 26. Esquema de los dos tipos de memoria 3D X-Point[18].

El primer método resulta más costoso ya que su fabricación es más compleja además de requerir más pasos en su producción que el segundo. El segundo enfoque es relativamente más sencillo que el primero, haciendo que este sea más prometedor a la hora de reducir el coste de fabricación.

Para que esta tecnología sea viable, aun es necesario estudiar la escalabilidad de los componentes para conseguir tamaños inferiores a los 20 nm, así como diseñar un circuito periférico para realizar la operación de lectura lo suficientemente compacto.

3.7. Tecnologías futuras

En este apartado se encuentran tecnologías NVM con potencial, pero, ya sea porque la tecnología es muy nueva y aún necesita más investigación y desarrollo, o bien por qué ha sido eclipsada por tecnologías más atractivas y ha perdido recursos, se encuentran en fases más tempranas de desarrollo que las tecnologías mencionadas anteriormente.

3.7.1. Memorias basadas en polímeros

Caracterizada por un proceso de fabricación simple y por el uso de materiales fáciles de manejar, la memoria basada en polímeros consiste en una capa compuesta de moléculas y / o nano partículas en una matriz de polímero orgánico, situada entre un conjunto de electrodos, tal y como se muestra en la figura. Los materiales poliméricos tienen una estructura simple, capacidad de lectura y escritura libre, mejor escalabilidad, capacidad de apilamiento 3D, potencial de bajo costo y gran capacidad de almacenamiento de datos.

Esta tecnología, al igual que otras en esta lista, almacena los datos en función de los cambios en la conductividad en una celda al aplicar un voltaje sobre esta. La capa compuesta por el polímero orgánico se crea mediante la técnica de recubrimiento por centrifugación sol-gel. El resto de materiales que constituyen la memoria son disueltos en un disolvente, para poder revestir el sustrato con estos [1]. Una vez se ha evaporado el solvente, una fina película del material se deposita en el fondo del material. Este proceso se repite para los electrodos superiores. La conductividad de la capa orgánica se cambia aplicando un voltaje a través de la celda, Cuando esta se vuelve conductora, los electrones se pueden introducir y extraer de ella.

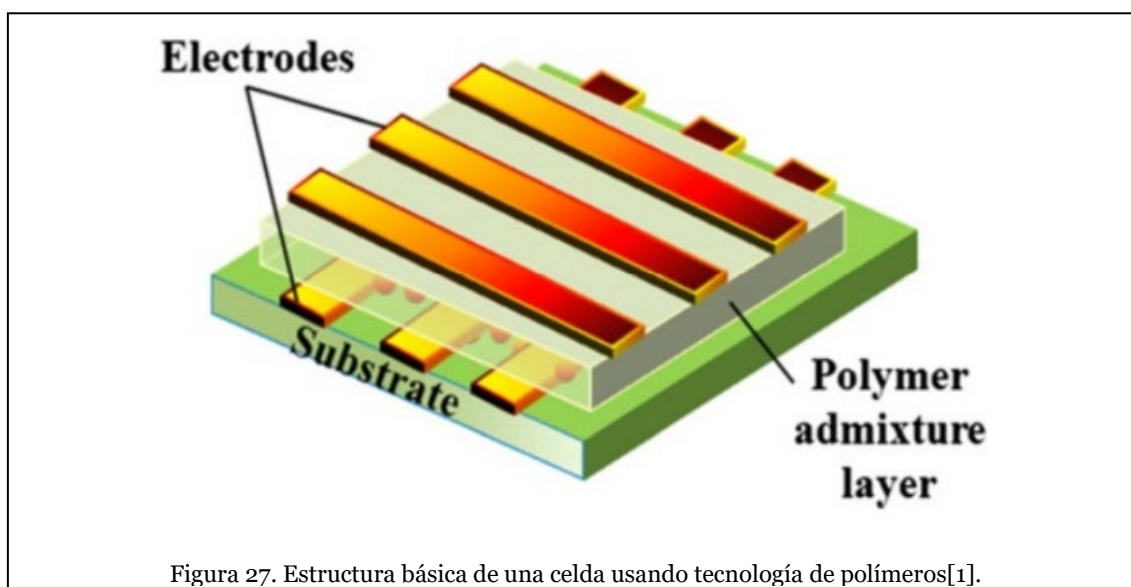


Figura 27. Estructura básica de una celda usando tecnología de polímeros[1].

Pero convertir la memoria de polímeros en un producto comercial no sería fácil. Las tecnologías de memoria compiten no solo por la capacidad de almacenamiento sino también por la velocidad, el consumo de energía y la fiabilidad. La dificultad está en cumplir con todos los requisitos de los chips de memoria de silicio actuales.

3.7.2. NRAM

NRAM es una memoria basada en nanotubos de carbono (CNT), que funciona según un principio nanomecánico, en lugar de un cambio en las propiedades del material. NRAM utiliza los nanotubos de carbono para determinar si la celda contiene 0 o 1, y este estado está determinado por la posición física del CNT: arriba con alta resistencia, o abajo y conectado a tierra. Se espera que la NRAM sea más rápida y más densa que la DRAM y también más escalable, ya que se espera que sea capaz de manejar celdas de bits de 5 nm cuando la fabricación de CMOS avance a ese nivel. La estructura de la celda consiste en una serie de electrodos en el fondo de la celda. Los CNT se depositan encima de está dejando un pequeño espacio que actúa como aislante, y encima de estos se colocan los contactos superiores y sus patrones de conexión. Cuando el CNT se encuentra separado del electrodo la celda se considera apagada (0), ya que no hay una ruta entre los electrodos superior e inferior [1]. Sin embargo, cuando se aplica un voltaje suficiente el CNT es atraído hacia el electrodo inferior, y se mantiene en esa posición gracias a las fuerzas de van der Waals. Esto hace que los electrodos estén conectados y la celda pasaría a estar encendida (1). Para apagar una celda basta con aplicar un voltaje con la polaridad opuesta al inicial, y el CNT será repelido del electrodo inferior.

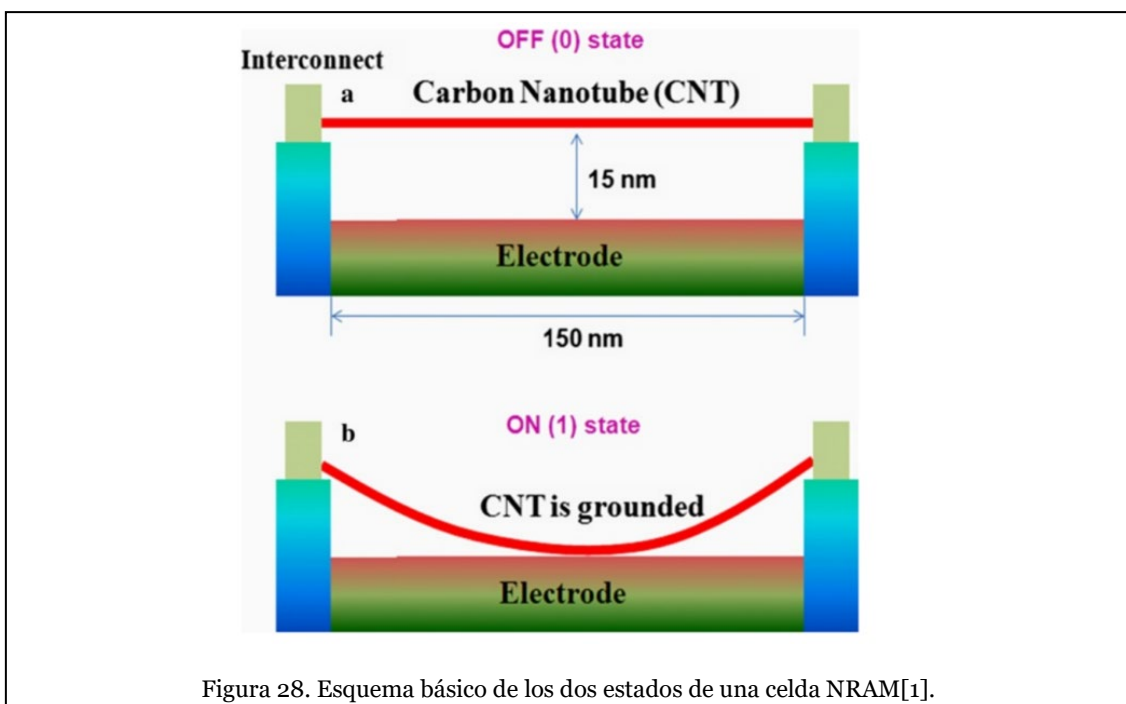


Figura 28. Esquema básico de los dos estados de una celda NRAM[1].

NRAM ofrece la posibilidad de una arquitectura de celda simple, que podría operar a velocidades mucho más altas que el Flash convencional, manteniendo un consumo reducido. Sin embargo, NRAM sufre problemas de fabricación que son inherentes a los dispositivos basados en nanotubos de carbono. Los problemas incluyen el costo y la complejidad de fabricación de los CNT, la inconsistencia en la colocación de los CNT en las posiciones deseadas y las dificultades para eliminar los nanotubos de las posiciones no deseadas del sustrato.

3.7.3. Memoria Millipede

Millipede es una NVM que almacena los datos en una lámina delgada de polímero mediante orificios nanoscópicos. Se podría considerar la tecnología sucesora de las tarjetas perforadas utilizadas a finales del siglo XIX. Esta tecnología proporciona una manera simple de almacenar datos binarios, pudiendo almacenar cientos de gigabytes de datos por centímetro cuadrado. Sin embargo, el polímero vuelve a su forma pre-perforada con el tiempo, perdiendo así los datos almacenados. Esta tecnología utiliza miles de pequeños puntos afilados (de ahí el nombre) para perforar agujeros en una película plástica delgada, donde cada uno de los agujeros de 10 nm representa un solo bit.

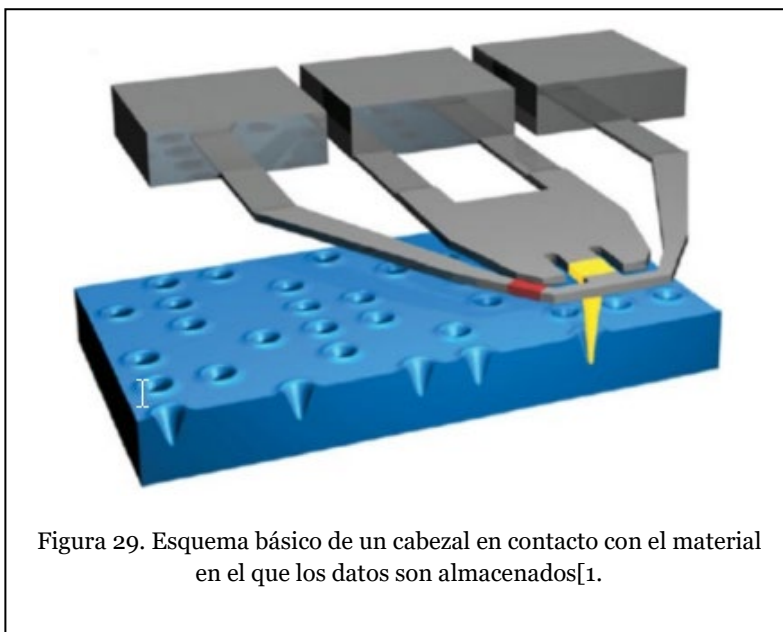


Figura 29. Esquema básico de un cabezal en contacto con el material en el que los datos son almacenados[1].

Los dispositivos que hagan uso de esta tecnología pueden ser fabricados con las técnicas de producción existentes, por lo que no serían costosos de hacer, y los últimos estudios han registrado que la tasa de pérdida de información durante una década estaría entre el 10% y el 20%, lo que la hace comparable a la memoria flash [1].

3.7.4. Memoria molecular

Una memoria molecular es una tecnología de memoria de almacenamiento de datos no volátil que utiliza moléculas como elemento de almacenamiento de datos, en

lugar de circuitos electrónicos o magnéticos. En una memoria molecular, una monocapa de moléculas se intercala entre dos matrices de electrodos. Las moléculas se empaquetan de una manera altamente ordenada, con un extremo de la molécula conectado eléctricamente al electrodo inferior y el otro extremo de la molécula conectada al electrodo superior, y este componente se describe como un interruptor molecular. Al aplicar un voltaje entre los electrodos, se altera la conductividad de las moléculas, lo que permite que los datos se almacenen en ellas [1]. Este proceso se puede revertir y los datos se pueden borrar aplicando un voltaje con polaridad opuesta al inicial.

Esta tecnología, pese a eliminar varias limitaciones relativas al rango de temperatura a la que puede operar y a restricciones de forma de esta, aún debe solucionar problemas relativos a la densidad y la potencia necesaria, así como de ampliar la diferencia en conductividad entre los dos estados de la celda, ya que actualmente solo es capaz de producir un cambio del 20% y una cifra tan baja puede dar problemas con la integridad de los datos.

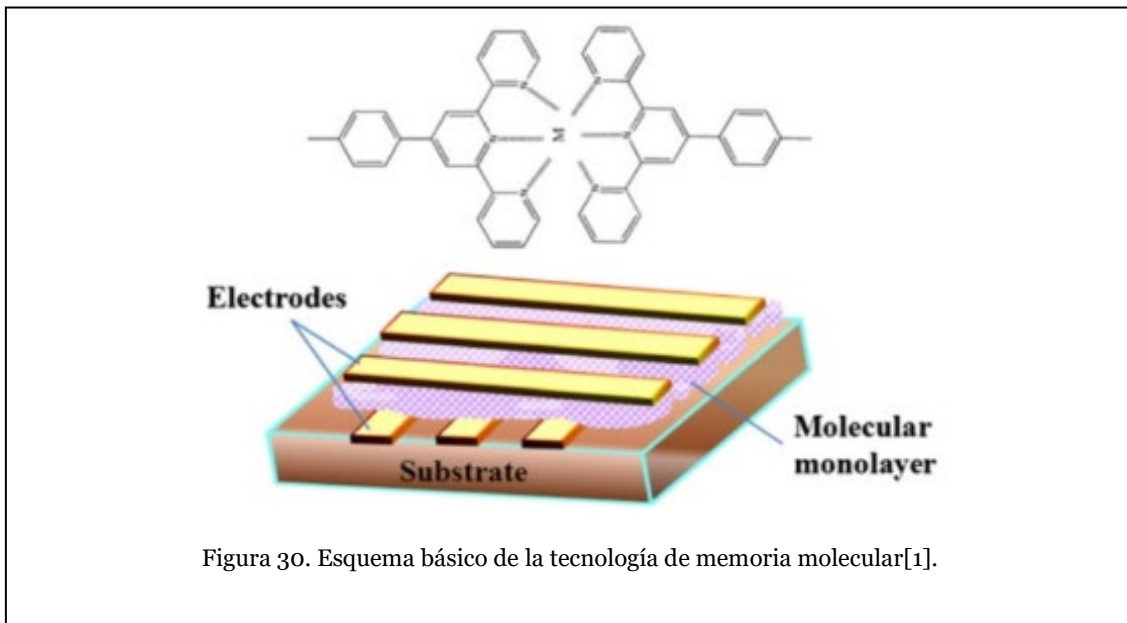


Figura 30. Esquema básico de la tecnología de memoria molecular[1].

3.7.5. Memoria MNW

La memoria MNW (Molecular NanoWire array) es fundamentalmente diferente de otras memorias de semiconductores, ya que el almacenamiento de información se logra a través de un transistor de nanocables que opera con moléculas redox activas en lugar de a través de la manipulación de pequeñas cantidades de carga. Por el momento, esta tecnología es relativamente lenta y carece de la capacidad de acceso aleatorio, aunque se está trabajando para superar estas limitaciones. Pese a esto, la tecnología de memoria basada en nanocables es un enfoque interesante a la hora de ensamblar dispositivos electrónicos / fotónicos en escalas ultra pequeñas debido a su tamaño sub-litográfico, su estructura monocristalina sin defectos y su geometría única. La celda de esta memoria consiste en un MNW que conecta dos electrodos, colocados sobre una fina capa aislante, seguida del sustrato. El canal de un transistor de nano cables está

recubierto con moléculas redo activas. Mediante el control del voltaje que actúa sobre el sustrato es posible cambiar el estado de oxidación y reducción de las moléculas activas, y midiendo la conductividad del MNW se pueden obtener los dos estados necesarios para almacenar la información.

Esta tecnología sería capaz de obtener una muy alta densidad, con un proceso de fabricación simple, y pudiendo ser fabricada en una estructura 3D [1]. Sin embargo está limitada por un reducido tiempo de retención y un bajo rendimiento en general del dispositivo en su estado actual. Sin embargo puede resultar un paso importante en la creación de computadoras moleculares en el futuro.

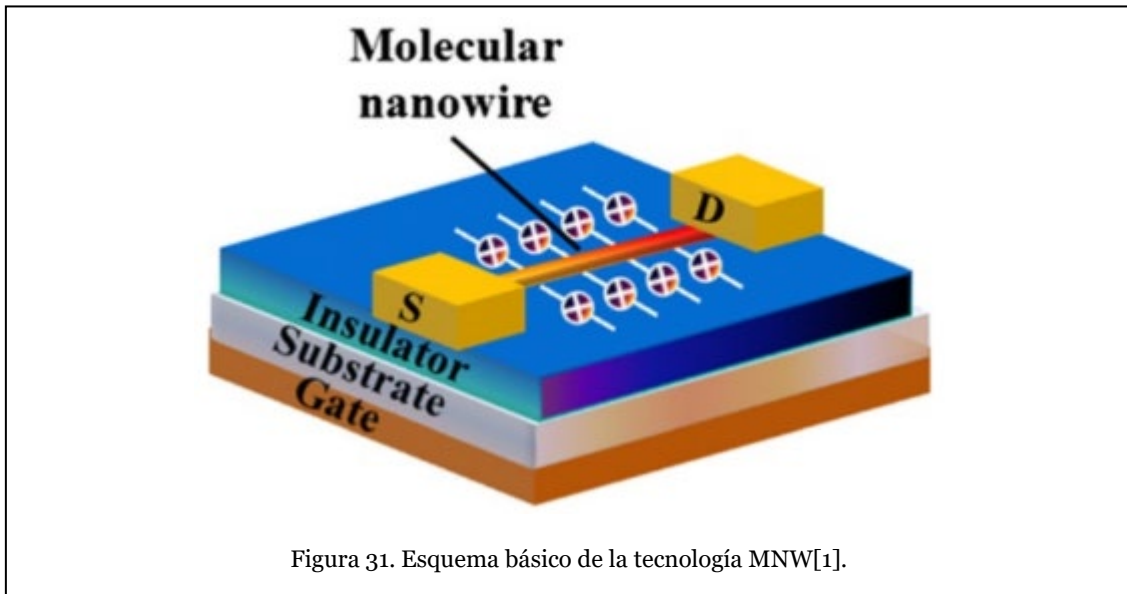


Figura 31. Esquema básico de la tecnología MNW[1].

3.7.6. Memoria QD

La memoria QD es un tipo de memoria basada en pequeñas islas de semiconductores, conocidos como puntos cuánticos o QD (Quantum Dots). Esta memoria podría proporcionar almacenamiento a largo plazo con velocidades de escritura casi tan rápidas como la DRAM. Los investigadores afirman que un gran conjunto de pequeñas islas, cada una de aproximadamente 15 nm de ancho, podría almacenar 1 terabyte de datos por pulgada cuadrada, siendo capaz de escribir datos en estos en menos de 10ns[14]. Además de una alta densidad y una excelente velocidad de lectura y escritura, esta memoria tiene un reducido voltaje de operación y es capaz de almacenar más de 1 bit en cada punto. Sin embargo, la incapacidad de producir puntos de igual tamaño así como la falta de un revestimiento aislante uniforme en estos es lo que mantiene esta tecnología aun en desarrollo.

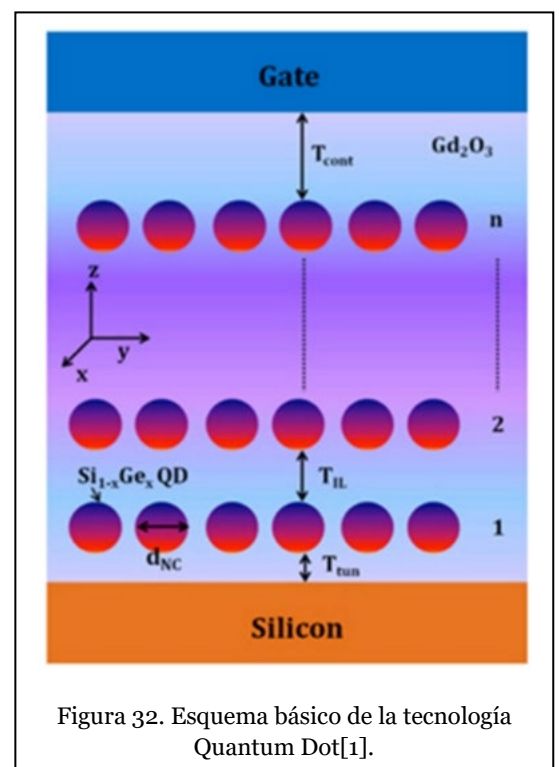
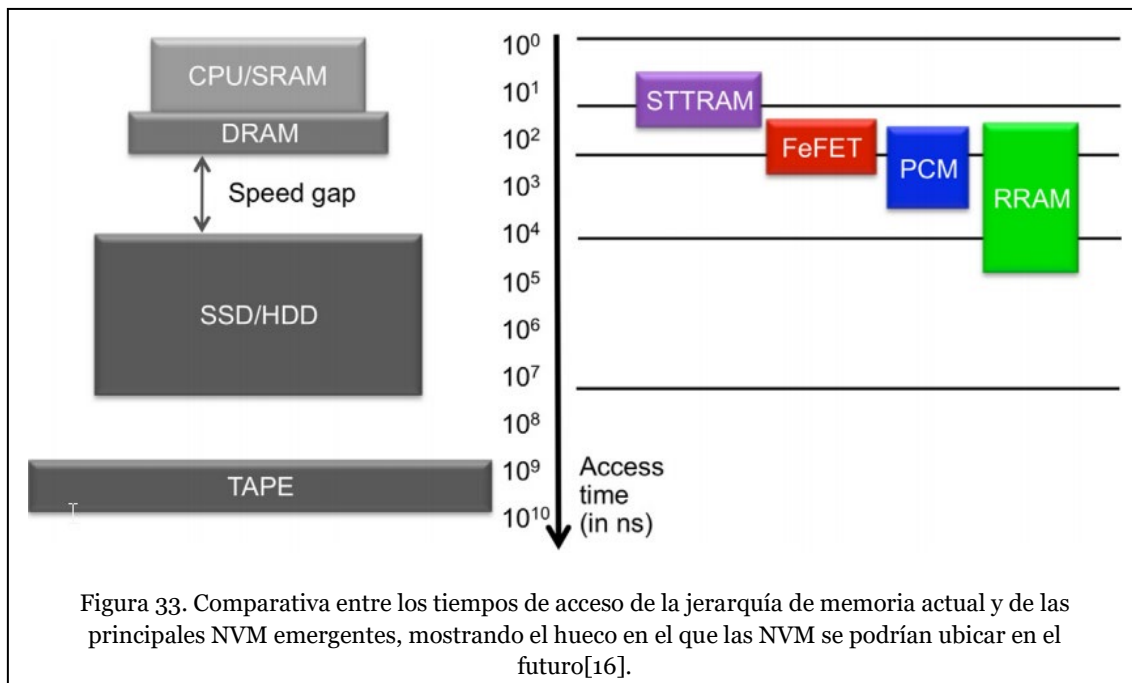


Figura 32. Esquema básico de la tecnología Quantum Dot[1].

4. Situación de las NVM en el mercado

Diferentes tecnologías NVM emergentes pueden tener aplicaciones distintas en la jerarquía de memoria debido a las características que poseen cada una. STT-MRAM resulta atractiva para reemplazar a SRAM y DRAM debido a sus tiempos de lectura ($<10\text{ns}$) y escritura y a su durabilidad ($>10^{15}$ ciclos). Sin embargo, su tamaño de celda actual y su coste, hacen que no pueda competir con SRAM y DRAM por el momento. Con el paso del tiempo, si se logra la reducción del tamaño de celda es muy probable que pueda reemplazarlas ya que no necesita un refresco constante de los datos. Por otra parte, PCRAM y RRAM tienen un tamaño de celda muy pequeño, y por consiguiente, una alta capacidad, lo que las hace muy apropiadas para servir como almacenamiento de datos. Es poco probable que logren sustituir a la memoria Flash como almacenamiento masivo, ya que ésta tiene un coste por bit muy reducido en comparación con RRAM y PCRAM, pero ambas NVM pueden ubicarse entre la memoria DRAM y Flash como Storage Class Memory o SCM, un concepto de memoria nuevo que combina la velocidad de la DRAM con la persistencia de los datos. Además, con el paso del tiempo y a medida que se realizan más estudios sobre las NVM, se descubren nuevas aplicaciones para estas que inicialmente no se tenían en cuenta. Las aplicaciones de las NVM emergentes y los productos que hay actualmente en el mercado son los temas que se van a tratar en este apartado. La figura 33 muestra el posible espacio en el que se podrían situar las NVM emergentes en el futuro.



4.1. Posibles aplicaciones de las NVM

4.1.1. Reemplazo de memorias existentes

Reemplazar las memorias existentes es la justificación por la que se empezó a estudiar nuevas tecnologías de almacenamiento de datos, debido a que durante los últimos años se ha ido haciendo más evidente que los sistemas de memoria actuales no van a poder satisfacer las prestaciones que se les va a demandar en un futuro no muy lejano. No obstante, estas demandas también se van a aplicar a cualquier producto que utilice una tecnología NVM emergente, lo que pone un listón muy elevado de entrada al mercado para una nueva tecnología. STT-RAM puede acercarse la durabilidad y la velocidad de la SRAM, sin embargo, si al principio se plantea como reemplazo de otra memoria con rendimiento más bajo (como reemplazo de la NOR Flash, por ejemplo), su entrada a la producción en masa podría ser mejor acogida [18]. También es atractiva para reemplazar a SRAM o eDRAM en el ultimo nivel de caché. Partiendo de ese punto, el perfeccionamiento de su proceso de fabricación y la reducción de costes podrían sacar a la luz su verdadero potencial. RRAM y PCRAM destacan por su escalabilidad y su pequeño tamaño de celda, lo que las hace idóneas para el almacenamiento masivo. Sin embargo, competir con NAND Flash en cuanto a densidad y coste por bit hacen que el reemplazo de esta sea muy improbable a corto plazo, mas ahora con el creciente interés por desarrollar arquitecturas 3D NAND Flash.

4.1.2. Storage Class Memory

Tal y como se muestra en la Fig. 31, los tiempos de acceso a los datos varían mucho de la caché a los discos duros HDD. La jerarquía de memoria se implementó para optimizar el tiempo de acceso a los datos sin elevar mucho el coste total del sistema. NVMs con capacidades de escalabilidad, con la velocidad y la durabilidad de la memoria RAM puede elevar mucho el rendimiento de un sistema. Pero la búsqueda de esta memoria universal está resultando un verdadero desafío. Sin embargo, existe un hueco en la jerarquía de memoria actual, situado entre la memoria principal y la memoria secundaria. Una memoria capaz de rellenar ese hueco tendría un gran impacto en el precio y el rendimiento de un sistema. Storage Class Memory o SCM posee velocidades comparables a la DRAM, a la vez que reduce el coste por bit al nivel de HDD, combinando los beneficios de la RAM y el almacenamiento masivo [6]. Para ello, SCM requiere una tecnología NVM fácilmente escalable, y con un proceso de fabricación sencillo, lo que hace PCRAM y RRAM unos candidatos excelentes para este nuevo tipo de memoria.

4.1.3. Soluciones NVM de bajo consumo

El consumo en reposo (estático, debido a corrientes de fuga) se ha convertido en una fuente importante del consumo de energía en muchos sistemas informáticos. Las NVM emergentes podrían eliminar la necesidad de entrar en modo de reposo de estos sistemas, ya que al apagarse retendrían la información almacenada. Los nodos de red inalámbricos y la red del Internet of Things (IoT) permanecen mucho tiempo en reposo, y solo se activan cuando reciben información. Estos sistemas, al usar SRAM como memoria caché, solo pueden desactivarse después de haber transferido los datos a una memoria no volátil. Para estos dispositivos la mejor solución para reducir su consumo sin afectar a su rendimiento sería sustituir la tecnología SRAM con la que operan por una tecnología NVM emergente que tenga un consumo reducido con la que se pueda diseñar un sistema que se encuentre normalmente apagado y que posea una caché no volátil. Esto, además de reducir el consumo, incrementaría drásticamente el rendimiento del sistema, y eliminaría el consumo en reposo del sistema. STT-RAM es un candidato muy prometedor para este cometido, gracias a sus elevadas velocidades de lectura y escritura de datos. Sin embargo debería mejorar su consumo dinámico para poder reemplazar la tecnología existente en este campo.

Por otra parte, la arquitectura de von Neumann genera un elevado consumo a la hora de realizar transferencias de datos. Las NVM emergentes son un recurso que permitiría nuevas arquitecturas donde la lógica y el almacenamiento estuvieran mejor integrados. Algunos dispositivos lógicos que han surgido en los últimos años incorporan funciones de almacenamiento. Esto reduce el movimiento de los datos a través del dispositivo, aumentando el rendimiento del sistema al reducir latencias y bajando el consumo del dispositivo [16]. Para que un dispositivo de estas características pueda ser posible el rendimiento de las NVM debe ser superior al actual, para que no se produzca un cuello de botella en el acceso a memoria de estos dispositivos.

4.1.4. Informática inspirada en el cerebro humano

En relación con los ordenadores de hoy en día, el cerebro humano consigue un rendimiento bastante alto consumiendo una cantidad muy baja de recursos. En las redes neuronales del cerebro, las sinapsis conectan las neuronas entre si y desempeñan una función clave en el proceso de aprendizaje. En los últimos años se ha demostrado que el comportamiento analógico de algunas NVM emergentes puede utilizarse para imitar los comportamientos sinápticos. Al ajustar el cambio en la resistencia de los elementos que componen estas NVM se pueden implementar los pesos sinápticos que se producen en el cerebro con el uso de los pulsos de conmutación de las NVM [16]. Aunque el uso de dispositivos de estado sólido para imitar las sinapsis no es nuevo, NVM emergentes como PCRAM y RRAM pueden lograr una densidad sináptica cercana a la de los cerebros (10^{10} cm⁻²), debido a su buena escalabilidad y baja potencia. Sin embargo, la demostración a nivel de matriz de redes neuronales y algoritmos todavía es limitada, y para que se puedan realizar pruebas a gran escala deben mejorarse la

estabilidad y la precisión de la modulación. El impacto de la imperfección de los dispositivos NVM en el rendimiento de la red sináptica es un tema de investigación importante si se quieren obtener mejores resultados en este ámbito.

4.1.5. Hardware de seguridad

Algunos comportamientos estocásticos de NVM emergentes son indeseables para las aplicaciones de memoria, pero podrían utilizarse como fuentes de entropía para aplicaciones de seguridad que incluyen variaciones verdaderamente aleatorias. Estos comportamientos incluyen la variabilidad de los parámetros del dispositivo NVM, ruido en la señal de lectura o RTN y conmutación probabilística (es decir, rendimiento de conmutación controlado por las condiciones de operación). Por ejemplo, la conmutación de RRAM es intrínsecamente estocástica con una gran variación en la resistencia de la celda y los voltajes de conmutación, además en RRAM la generación de RTN es muy común [16]. La probabilidad de cambio dependiente de la condición se ha caracterizado para RRAM y STTRAM. Estos comportamientos aleatorios de NVM emergentes pueden utilizarse para generar primitivas de seguridad de hardware[18].

4.1.5.1. Generador de números aleatorios

Principalmente existen 2 métodos para generar números aleatorios. El primer método utiliza algoritmos computacionales que pueden producir secuencias largas de resultados aparentemente aleatorios, que de hecho están completamente determinados por un valor inicial más corto, conocido como valor semilla o clave. Como resultado, toda la secuencia aparentemente aleatoria puede reproducirse si se conoce el valor semilla. Este tipo de generador de números aleatorios se suele denominar generador de números pseudoaleatorios.

El segundo método mide algún fenómeno físico que se espera que sea aleatorio. Este método incluye la medición del ruido atmosférico, el ruido térmico y otros fenómenos electromagnéticos y cuánticos externos. La conmutación probabilística de STTRAM y RRAM permite la programación de estos para que exista la misma probabilidad de que una celda genere en "0" o "1". Este comportamiento se puede utilizar para crear un TRNG (True Random Number Generator), que se ha demostrado experimentalmente con la prueba de la aleatoriedad real para las aplicaciones de seguridad [18]. Alternativamente, la fuerte RTN (Random Telegraph Noise, transiciones súbitas y escalonadas entre dos o más niveles discretos de voltaje o corriente, tan altos como varios cientos de microvoltios, en tiempos aleatorios e impredecibles.) en RRAM también se puede usar para generar números aleatorios en un circuito simple.

4.1.5.2. Función física irreproducible

Una función física irreproducible o PUF utiliza la aleatoriedad física para generar características de seguridad específica que no se puedan clonar y se pueda usar como "huella digital" para identificar o autenticar hardware. Las soluciones PUF existentes explotan principalmente la inevitable variación en la fabricación de circuitos integrados que es incontrolable, no se puede clonar y es única para cada circuito individual. La variabilidad intrínseca de la RRAM proporciona una fuente alternativa de aleatoriedad para la implementación de PUF. A diferencia de la variación de fabricación que se fija una vez que se fabrica un circuito, la variabilidad de la RRAM es potencialmente reconfigurable, y no depende de el proceso de fabricación.

Un PUF ideal debería generar respuestas con el mismo porcentaje de bits cero y uno. También debe asegurar que las diferentes respuestas de diferentes PUF para el mismo producto deben ser completamente diferentes y no correlacionadas. Finalmente, una PUF siempre debe generar la misma respuesta.

PUF también se ha implementado en otros NVM emergentes, incluidos PCM y STTRAM. Los PUF basados en NVM son mucho más pequeños que otros PUF (por ejemplo, SRAM PUF), por lo tanto, no solo son más adecuados para aplicaciones de seguridad ligeras, sino que también permiten una mayor longitud de bits para mejorar las soluciones de seguridad [16]. La reprogramación de NVM también puede ser útil a la hora de diseñar PUF con mejores prestaciones.

4.1.6. Hardware industrial

La mayoría de sistemas industriales son centralizados, donde los nodos envían al sistema central información para ser almacenada y analizada. Estos sistemas necesitan enlaces entre los nodos y el sistema central que sean síncronos, fiables y que tengan una latencia muy baja. Las *smart factories* de hoy en día utilizan un enfoque modular, al delegar la toma de decisiones a cada nodo, utilizando aprendizaje automático y inteligencia artificial. Con la toma de decisiones produciéndose en el nodo sus requisitos han cambiado, antes solo tenían que almacenar los datos de los sensores en su último estado y ahora necesitan almacenar los datos producidos por el aprendizaje automático para funcionar correctamente. Para ello es necesaria una memoria densa, rápida, con capacidad de alteración de nivel de bit/byte, una alta retención y un amplio rango de temperatura operativa. Tecnologías como RRAM y PCRAM pueden resultar idóneas para este cometido

4.1.7. Automoción

Los datos recogidos a bordo de un automóvil deben procesarse y almacenarse al instante, ya que cualquier pérdida de datos a velocidades de autopista podría resultar catastrófico. Para ello la memoria idónea como sistema de memoria central de un automóvil sería una memoria robusta, capaz de soportar vibraciones, golpes y cambios de temperatura, además de baja latencia y variabilidad de los datos. STT-RAM, PCRAM y RRAM están siendo estudiadas como reemplazo de las NOR y NAND Flash utilizadas en la actualidad

4.1.8. Industria aeroespacial

Las condiciones a las que debe funcionar una memoria en el espacio son extremadamente más duras que las presentes en la atmosfera terrestre. La radiación acelera el envejecimiento de las piezas y materiales semiconductores, lo que lleva a una degradación del rendimiento eléctrico o un fallo prematuro. Además de la radiación, los dispositivos que operan en el espacio son expuestos a temperaturas extremas que oscilan entre -55°C y 125°C . Para combatir estos desafíos, es necesaria una memoria altamente confiable, resistente a la radiación y con alta retención de datos. STT-RAM presenta unas características idóneas para este cometido, al ser resistente a la radiación y a los cambios de temperatura, y al tener una durabilidad tan elevada. Actualmente el satélite SpriteSat integra un chip STT-RAM de 4Mb [39,40]. Adicionalmente, RRAM también ha demostrado ser resistente a la radiación [15].

4.2. Dispositivos NVM existentes en el mercado

En este apartado se presentaran de manera breve diversos productos ya a la venta que hacen uso de las tecnologías NVM emergentes descritas en el apartado 3.

4.2.1. Everspin 1-16Mb MRAM

Este dispositivo (MR4A08B y MR4A16B) es un chip MRAM de 16,777,216 bits organizado como 2,097,152 palabras de 8 bits. Existen dos configuraciones distintas del dispositivo, una con un bloque de 16 Mb y otra con dos bloques de 8Mb. Ofrece un tiempo de lectura / escritura de 35 ns con resistencia muy elevada. El tiempo de retención de datos es superior a 20 años, y estos están protegidos en caso de pérdida de energía mediante un circuito de inhibición de baja tensión para evitar escrituras con voltaje fuera de las especificaciones. Es la solución de memoria ideal para aplicaciones

que deben almacenar y recuperar de forma permanente datos y programas críticos rápidamente. Proporciona almacenamiento de datos altamente confiable en un amplio rango de temperaturas. El producto se ofrece con opciones de rango de temperatura de funcionamiento comercial (0 a +70 ° C) e industrial (-40 a +85 ° C)[29].

Estos chips están diseñados para sustituir otros módulos de memoria en un circuito integrado, por ejemplo en el nuevo Controlador lógico programable de Koyo Electronics (DL205), utilizado en la ingeniería automática o automatización industrial para automatizar procesos electromecánicos, o el Schneider Electric Modicon M580 Automation Controller, un dispositivo industrial orientado al control automatizado del Industrial Internet of Things (IIOT) de una empresa.

Varios chips de la misma familia (MR2A08A y MR2A16A, entre otros) poseen la certificación AEC-Q100 GRADE 1, otorgado a los dispositivos después de superar una prueba de estrés que garantiza el correcto funcionamiento sin fallos en temperaturas desde -40° hasta 125°, y que otorga al chip el grado de producto de calidad automotriz. Estos chips han sido integrados en la unidad de control del motor R2M5 de las BMW Motorrad, que compiten en el Superbike World Championship. Además, un chip de similares características fue diseñado para su uso en el satélite SpriteSat, concebido para monitorizar las emisiones luminosas en la atmósfera superior.

4.2.2. Avalanche Technology 1-32Mb MRAM

Esta familia de dispositivos MRAM fabricada por Avalanche Technology (AS3001316, AS3004316, AS3008316, AS3016316) poseen 16 bloques de memoria que funcionan de forma paralela y asíncrona. Se fabrican en una capacidad que varía desde 1Mb a 16Mb. Ofrece tiempos de retención de datos superiores a los 10 años. Presenta baja latencia, bajo consumo de energía, resistencia infinita, alto rendimiento y buena escalabilidad. Está compuesta de celdas MRAM de 40 nm, y su rango de voltaje operativo es 2,7-3,6V. Además, Ofrece 3 opciones que pueden operar a temperaturas distintas: Comercial (0° a 70°), Industrial (-40° a 85°) e Industrial Plus (-40° a 105°) [32].

Estos chips están diseñados para sustituir otros módulos de memoria en un circuito integrado, por ejemplo en un Controlador lógico programable industrial. Esta familia de chips está disponible en tres encapsulados: 44 o 54 pin TSOP, y 48 ball FBGA

4.2.3. Avalanche Technology 8-64Mb eMRAM

La eMRAM (embedded MRAM) fabricada por Avalanche Technology es una memoria de acceso aleatorio magneto-resistiva (MRAM) que varía en densidad de 8Mbit a 64Mbit y está organizada en palabras de 32 bits. eMRAM es una verdadera memoria de acceso aleatorio, permitiendo que tanto las lecturas como las escrituras ocurran aleatoriamente en la memoria. eMRAM es una memoria no volátil altamente



confiable y rápida, y gracias a ello se está convirtiendo rápidamente en el reemplazo de eFlash como memoria incorporada en el chip (SoC).

El eMRAM es ideal para aplicaciones que deben almacenar y recuperar datos sin incurrir en grandes penalizaciones de latencia. Ofrece baja latencia, buena escalabilidad, bajo consumo de energía, resistencia muy elevada y un alto rendimiento. Compuesta de celdas MRAM de 28 nm, puede operar en temperaturas desde los -40° hasta los 125° . Además su rango de voltaje operativo oscila entre 0,8 y 2V, y ofrece una durabilidad de 10^9 ciclos [33].

4.2.4. Everspin 256Mb STT-RAM

La DDR3 STT-MRAM (Familia EMD3D256M) es una memoria de alta velocidad configurada internamente como una memoria RAM de ocho bancos. Es una memoria que ofrece no volatilidad y alta resistencia a velocidades DDR3. El dispositivo es capaz de funcionar con DDR3 a velocidades de hasta 1333 MT/s/Pin [30]. Está diseñado para cumplir con todas las características de DRAM DDR3. Con la tecnología STT-MRAM, no se requiere la refresco de la celda, lo que simplifica enormemente el diseño del sistema y reduce el consumo. El dispositivo utiliza un esquema de multiplexación RAS/CAS y funciona a 1.5V.

Posee una interfaz diseñada para transferir dos palabras de datos por ciclo de reloj en los pines de E/S. Las operaciones de lectura y escritura están orientadas a la ráfaga, comienzan en una ubicación seleccionada y continúan durante una ráfaga de ocho o una ráfaga "cortada" de cuatro en una secuencia programada. La operación comienza con el registro de un comando de activación, que es seguido por un comando READ o WRITE. Los bits de dirección registrados coincidentes con el comando de activación se utilizan para elegir el banco y la fila que se activarán. Existen dos configuraciones disponibles para este dispositivo, una con 8 bloques de memoria de 32Mb, y otra con 16 bloques de 16Mb. La primera consta de un tamaño de página de 512 bits, mientras que la segunda es de 1024 bits. El dispositivo cuenta con una durabilidad de 10^{10} ciclos, un tiempo de retención de los datos estimado de 3 meses, y el Bit Error Rate o BER es de 6.3×10^{-8} .

A partir de esta familia de chips Buffalo Memory ha estado trabajando en un SSD altamente confiable para el mercado industrial y planea lanzar un SSD con caché STT-MRAM con factor de forma de 2.5 pulgadas e interfaz SATA3 para proporcionar tolerancia a fallos de energía, mayor velocidad y mayor consistencia. El nuevo SSD se lanzará para las aplicaciones en las que se requiere la máxima fiabilidad.

4.2.5. Avalanche Technology 1-32Mb STT-RAM

Esta familia de chips (AS1001101, AS1004101, AS1008101, AS1016101, AS3001101, AS3004101, AS3008101, AS3016101, AS1001204, AS1004204, AS1008204, AS1016204, AS3001204, AS3004204, AS3008204, AS3016204) fabricada

por Avalanche Technology funcionan con la tecnología STT-MRAM. Están divididos en dos subfamilias, ASxxxx204 la forman los chips de altas prestaciones, caracterizados por sus dos modos de funcionamiento: SDR y DDR. Mientras tanto, ASxxxx101 está formada por los chips de bajo consumo. Utilizan una celda con un MTJ perpendicular, y el tamaño de esta es 40nm.

Existen 5 tamaños disponibles: 1Mb, 4Mb, 8Mb y 16Mb, y su rango de voltaje de operación es de 1,7V a 3,6V [31]. Aseguran una durabilidad de 10^{16} ciclos y una retención de los datos superior a los 20 años. Además se ofrecen 3 opciones que pueden operar a temperaturas distintas: Comercial (0° a 70°), Industrial (-40° a 85°) e Industrial Plus (-40° a 105°) para cada dispositivo.

Estos chips están diseñados para sustituir otros módulos de memoria en un circuito integrado, por ejemplo en un Controlador lógico programable industrial. Esta familia de chips está disponible en dos formatos: 8 pad WSON, 8 pin SOIC, y 24 ball FBGA.

4.2.6. Intel Optane 16-64GB 3D XPoint PCRAM

La memoria Intel Optane ofrece baja latencia y alto rendimiento con altas velocidades de lectura / escritura bajo cargas de trabajo exigentes. Este tipo de rendimiento hace que la memoria Intel Optane sea una unidad de sistema operativo ideal, que ofrece un arranque rápido tanto del sistema como de las aplicaciones, así como un entorno multitarea fluido [34].

Utiliza la tecnología PCRAM con una arquitectura 3D Xpoint. Dispone de 3 capacidades, 16, 32 y 64 GB, y su consumo cuando esta activo el dispositivo oscila entre los 2 y los 3,25 W. Su durabilidad ronda los 10^{15} ciclos y su rango de temperatura operativo es de 0 a 85° .

Esta memoria utiliza un puerto PCIe para conectarse a un PC, y es utilizada para manejar grandes cantidades de datos a altas velocidades, evitando así el cuello de botella que suele formarse al utilizar archivos ubicados en la memoria secundaria, ya sea NAND Flash o un HDD.

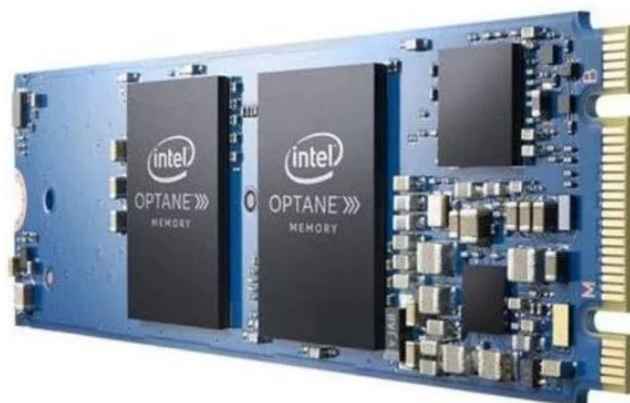


Figura 34. Memoria Intel Optane [34].

4.2.7. Panasonic MN101L Series eRRAM

Este chip fabricado por Panasonic es un microprocesador de 8 bits de baja potencia que lleva incorporado en él un módulo ROM de 64KB que utiliza la tecnología RRAM. Posee una frecuencia máxima de operación de 10MHz y está disponible con 48, 64 y 80 pines en el chip. El voltaje necesario para operar oscila entre 1,1 y 3,6V[35].

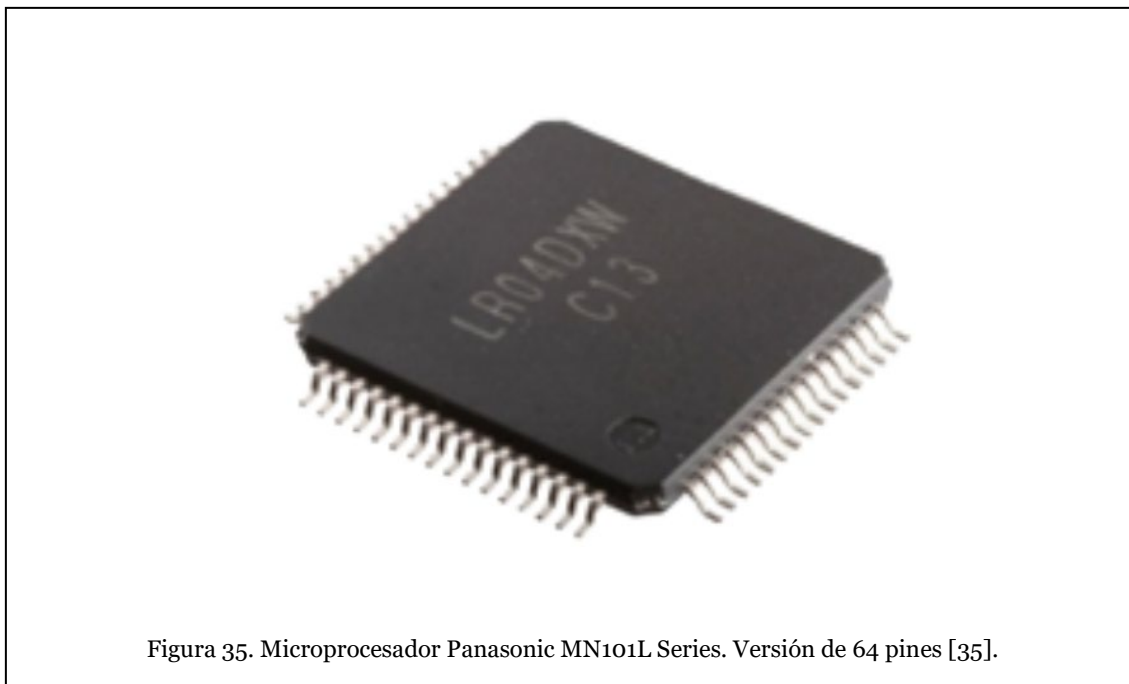


Figura 35. Microprocesador Panasonic MN101L Series. Versión de 64 pines [35].

Este microprocesador utiliza RRAM para obtener un menor coste, y a su vez aumentar la duración de la batería y la velocidad de operación del producto. Tales beneficios hacen que sea una MCU ideal para aplicaciones en los campos de la seguridad, la industria y la medicina, o como Electronic Shelf Labels, como Panasonic demuestra en un video corporativo presentando el producto[45].

4.2.8. Crossbar 256KB- 16MB eRRAM

Estos chips de memoria integrada son una opción ideal en aplicaciones de memoria no volátil incorporada, como Internet of Things (IoT), dispositivos portátiles, tabletas, teléfonos inteligentes, electrónica de consumo, inteligencia artificial, industrial, automotriz y médica. Comenzando a 40 nm y escalando por debajo de 10 nm, la serie P permite soluciones de memoria de almacenamiento de datos y ejecución de código de bajo consumo, de bajo costo, de baja latencia y de alto rendimiento. Las densidades admitidas varían de 256K Bytes a 16M Bytes, aunque también se puede elegir un tamaño personalizado [36].

Aseguran una durabilidad superior a 10^6 , una retención de más de 10 años y la posibilidad de operar en temperaturas desde los -40° hasta los 125° .

Utilizar RRAM en un microprocesador reduce el coste y el consumo del dispositivo, y a su vez aumenta la velocidad de operación del producto. Tales beneficios hacen que sea una MCU ideal para aplicaciones en los campos de la industria, la medicina, el IOT y la inteligencia artificial.

4.2.9. Crossbar 3D 1GB-1TB eRRAM

Estos chips de memoria integrada son perfectos para aplicaciones de memoria de alta densidad y baja latencia, como el almacenamiento de centros de datos, computación móvil, electrónica de consumo e inteligencia artificial. Ofrecen alta densidad, baja latencia, alto rendimiento y baja potencia en una solución de memoria no volátil. La integridad de los datos y las características operativas superan el flash 3D NAND actual. Las densidades admitidas son de 1GByte a 1 TeraByte, aunque también se puede personalizar el tamaño del chip [37].

Aseguran una durabilidad superior a 10^6 , una retención de más de 10 años y la posibilidad de operar en temperaturas desde los -40° hasta los 125° , además de poseer el doble de densidad que un chip 3D NAND.

A partir de sus chips de memoria Crossbar ofrece soluciones SSD NVMe y NVDIMM. La solución NVDIMM basada en ReRAM es una solución de memoria persistente no volátil que combina chips de memoria ReRAM para funcionar en las ranuras de memoria DRAM estándar. Las soluciones SSD NVMe basadas en ReRAM reducen la brecha de rendimiento entre el almacenamiento de datos y el procesamiento de la CPU, y sirven como el reemplazo perfecto de SSD basado en Flash con mayor resistencia, confiabilidad y bajo consumo de energía.

5. Modelado de NVM

Durante las últimas décadas se han desarrollado muchas herramientas de modelado para permitir la exploración de diseño a nivel de sistema para memoria caché y memoria basada en SRAM o DRAM. Por ejemplo, CACTI es una herramienta que se ha utilizado ampliamente en la comunidad de arquitectura de computadores para estimar el rendimiento, la energía y el área de cachés de SRAM y DRAM [46]. A este se le incorporaron algunas mejoras, como un modelo de energía para las SRAM, que se utilizó para desarrollar una organización óptima de la memoria caché, y también se le incorporó un modelo que estimaba la potencia estática. Además, CACTI también se ha ampliado para evaluar el rendimiento, la energía y el área de STTRAM, PCRAM, ReRAM y flash NAND. Sin embargo, como CACTI se diseñó originalmente para modelar una caché basada en SRAM, algunas de sus suposiciones no coinciden con las implementaciones reales del circuito NVM, y por lo tanto las prestaciones de la NVM modelada en CACTI y los chips NVM fabricados basándose en estos cálculos difieren bastante.

Por lo tanto, se requieren circuitos periféricos especializados para cada tecnología. Sin embargo, como pocas de estas tecnologías NVM están lo suficientemente maduras actualmente, solo se ha demostrado un número reducido de chips prototipo los cuales solo cubren una pequeña parte del espacio de diseño completo. Por lo tanto, las herramientas de modelado a nivel de circuito son necesarias para facilitar la investigación de NVM al predecir el rendimiento, el consumo de energía y el área del chip para reducir y optimizar la construcción de prototipos.

5.1. NanoHUB

NanoHUB.org [47] es la principal comunidad cibernética para la investigación, educación y colaboración en nanotecnología computacional. La Red de Nanotecnología Computacional (NCN), que opera nanoHUB, se estableció en 2002 y está financiada por la National Science Foundation (NSF) para apoyar la Iniciativa Nacional de Nanotecnología (NNI). NanoHUB.org avanza esa iniciativa y ahora atiende a más de 1.4 millones de visitantes anuales, incluidos investigadores, educadores, estudiantes y profesionales.

NanoHUB.org ofrece 320 herramientas de simulación disponibles a través de un navegador web. Las herramientas de simulación se ejecutan como *applets* en la ventana del navegador, y se ejecutan remotamente en los servidores de la Universidad de Purdue.

Entre todas las herramientas que dispone la plataforma, hay 4 que resultan interesantes y que entran dentro del ámbito de este trabajo:

5.1.1. Nanomaterial Mechanics Explorer

Esta herramienta permite a los usuarios explorar las propiedades de los materiales desde la escala atómica, incluida la propagación de microfisuras, la deformación plástica, la fusión y la transformación martensítica. La herramienta utiliza simulaciones de dinámica molecular con ejemplos preconstruidos y control total sobre los parámetros de simulación individuales. La herramienta ejecuta el código de dinámica molecular LAMMPS y se conecta a OpenKIM, una base de datos de modelos interatómicos, que permite la simulación de más de 30 elementos y 50 aleaciones de más de 180 modelos.

Para cada simulación, se genera una visualización atómica del sistema, así como la visualización específica de los átomos defectuosos (aquellos que no coinciden con la estructura cristalina inicial). Todas las demás salidas son elegidas por el usuario y son específicas de la simulación individual.

Entre las simulaciones disponibles se encuentra la transformación de fase. En este módulo, la fusión se estudia a nivel atómico. Un cristal sólido se calienta por encima de su temperatura de fusión a presión ambiente hasta que esté completamente líquido para así medir el cambio de estructura y el consumo de energía

5.1.2. Memristor Simulation Tool

Esta herramienta analiza el cambio en la resistencia de la memoria, de acuerdo a el voltaje y la corriente empleadas para realizar la simulación. Esta herramienta permite la modificación de diversos parámetros, tales como la frecuencia y la amplitud de el voltaje aplicado, el tamaño de celda, y los valores de R_{on} y R_{off} de la resistencia de la memoria. A partir de estos datos realiza la simulación y muestra los resultados mediante diversas graficas.

5.1.3. Magnetic Tunnel Junction Lab

Esta *applet* calcula la resistencia, Tunneling Magneto Resistance, Spin Torque, y las características de conmutación de un Magnetic Tunnel Junction (MTJ). Consiste en un simulador de transporte cuántico basado en la función de no equilibrio de Green (NEGF) que calcula las propiedades críticas de transporte de un MTJ, además de las resistencias paralelas y antiparalelas, Tunneling Magneto-Resistance (TMR) y Spin Transfer Torque (STT, tanto en el plano como fuera del plano), utilizando los parámetros de material y las dimensiones geométricas del dispositivo MTJ como datos de entrada.

5.1.4. Subcircuit Generator for Ferromagnetic Nanomaterials

Esta herramienta desarrolla modelos de circuitos compactos de la evolución temporal de la dinámicas de magnetización dentro de un material ferromagnético al emular el comportamiento matemático a partir de la ecuación Landau - Lifshitz - Gilbert (LLG). La herramienta también incorpora el efecto spin-transfer torque (STT) en un material ferromagnético.

La herramienta es útil para aplicaciones electrónicas ya que produce una lista SPICE que describe modelos de circuitos para la dinámica de conmutación de magnetización subyacente necesaria para simular materiales magnéticos a nanoescala en dispositivos spintronic.

5.1.5. Conclusiones

NanoHUB proporciona herramientas de simulación muy potentes, pero las que entran dentro del ámbito de este trabajo solo simulan ciertos aspectos que ocurren a nivel de celda de las NVM emergentes, que no dejan de ser interesantes, pero que serían más apropiadas para un trabajo en el campo de la física. Este trabajo busca un simulador capaz de simular NVM emergentes, así como DRAM, SRAM y Flash, a nivel de chip para poder comparar prestaciones.

5.2. NVSim

NVSim [38] es una herramienta de modelado a nivel de circuito para el cálculo del desempeño, la energía y el área. Al ser una extensión de CACTI, NVSim admite SRAM y DRAM, y además NAND flash, STTRAM, PCRAM y ReRAM.

Los principales objetivos del desarrollo de la herramienta NVSim son los siguientes:

1. Estimar el tiempo de acceso, la energía de acceso y el área de silicio de chips NVM con una organización específica antes de gastar recursos en producir un prototipo.
2. Encontrar la organización de chip optima y sus opciones de diseño que logren el mejor rendimiento, energía o área.
3. Encontrar las opciones de diseño optimas definiendo una o varias variables previamente.

A continuación se va a proporcionar una breve explicación de el funcionamiento interno de NVSim, para obtener documentación más detallada puede acceder a [38].

5.2.1. Framework de NVSim

El framework de NVSim es una variación del framework de CACTI, al que se le han añadido varias características nuevas, como modos de activación de datos más flexibles y diversas alternativas en la organización de bancos.

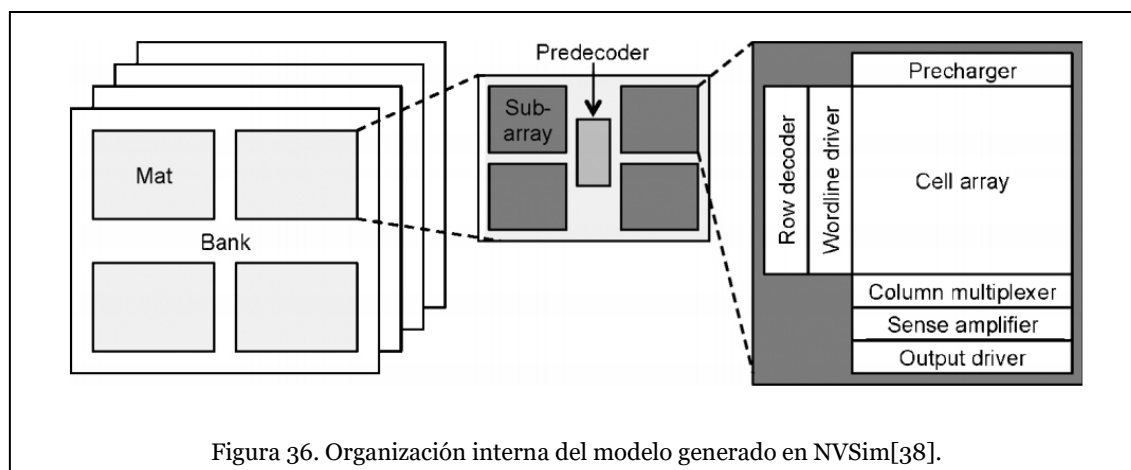
5.2.1.1. Modelado del dispositivo

NVSim utiliza los datos del último ITRS Report [7] y de la herramienta MASTAR [48] para obtener los parámetros necesarios para el modelado. Es capaz de modelar celdas desde 22 nm hasta 180 nm y admite tres tipos de transistor: Alto rendimiento (HP), Bajo consumo activo (LOP) y bajo consumo en reposo (LSTP).

5.2.1.2. Organización de los datos

Hay tres niveles de jerarquía en dicha organización, que son banco, *mat* y *subarray*. Básicamente, las descripciones de estos niveles son las siguientes:

- El banco es la estructura de nivel superior modelada en NVSim. Un chip NVM puede tener múltiples bancos. El banco es una unidad de memoria completamente funcional, y se puede operar de manera independiente. En cada banco, varios mats están conectados entre sí en forma de árbol H o de bus.
- El mat es el bloque de construcción del banco. Múltiples mats en un banco operan simultáneamente para cumplir una operación de memoria. Cada mat consta de varios subconjuntos y un predecodificador.
- El subarray es la estructura elemental modelada en NVSim. Cada subarray contiene circuitos periféricos que incluyen decodificadores de fila, multiplexores de columna y controladores de salida.



Normalmente, en el modelado en CACTI, los amplificadores están integrados en el nivel de subarray. Sin embargo, en el modelo NVSim, los amplificadores de detección pueden colocarse tanto en el nivel de subarray como en el nivel de mat.

5.2.1.3. Tipos de banco de memoria

Para cubrir todos los diseños de memoria posibles, existen cinco tipos de bancos de memoria en NVSim: uno para RAM, uno para CAM y tres para cachés asociados con diferentes maneras de acceso. Las funcionalidades de estos cinco tipos de bancos de memoria se enumeran a continuación.

- RAM: obtiene el contenido de los datos en la interfaz de E/S al dar la dirección de los datos.
- CAM: obtiene la dirección de los datos en la interfaz de E/S al dar el contenido de los datos si se produce una coincidencia.
- Caché con acceso normal: accede al array de datos de la caché y al array de etiquetas al mismo tiempo. El contenido de los datos se almacena temporalmente en cada mat; si hay un acierto, la señal de acierto de la memoria caché generada a partir del array de etiquetas se enruta a los mats adecuados y el contenido de la línea de caché deseada se envía a la interfaz de E/S.
- Caché con acceso secuencial: accede primero al array de etiquetas de la caché; si hay un acierto, entonces accede al array de datos de la caché con la dirección establecida y la información del acierto, y finalmente envía la línea de caché deseada a la interfaz de E/S.
- Caché con acceso rápido: accede al array de datos de la caché y al array de etiquetas simultáneamente. Lee todo el contenido del conjunto desde los mats a la interfaz de E/S, y genera de manera selectiva la línea de caché deseada si hay una señal de acierto de caché generada desde el array de etiquetas.

5.2.2. Modelado del área

5.2.2.1. Estimación del área de la celda

NVSim puede modelar tres tipos distintos de celda de memoria.

- MOS-Accessed Cell: Corresponde a la estructura 1-Transistor 1-Resistor (1T1R) usada por la mayoría de chips NVM, en la cual un transistor NMOS está conectado en serie con el elemento de memoria.
- Cross-Point Cell: Corresponde a la estructura 1-Diode 1-Resistor (1D1R) y a la 0-Transistor 1-Resistor (0T1R) utilizadas por bastantes chips NVM de alta densidad.
- NAND-String Cell: Esta estructura esta modelada específicamente para NAND Flash.

5.2.2.2. Estimación del área de la circuitería externa

Para estimar el área de cada componente del circuito periférico, se profundiza en el diseño lógico a nivel de puerta de forma similar a CACTI. Sin embargo, en NVSim, los transistores se dimensionan de forma más generalizada que CACTI. Se ofrecen tres opciones de tamaño de transistor en el modelo de área de NVSim: uno optimizado para la latencia, uno optimizado para el área, y uno equilibrado en área y latencia.

5.2.3. Modelado de la latencia y el consumo

Como herramienta de modelado analítico, NVSim utiliza análisis RC para temporización y potencia. En esta sección, describimos cómo se estiman las resistencias y capacitancias en NVSim y cómo se combinan para calcular el retraso y el consumo de energía.

5.2.3.1. Estimación del área de la circuitería externa

NVSim considera la resistencia del cable y la capacitancia del cable de las interconexiones, la R_{on} , la resistencia de conmutación, las capacitancias de puerta y drenador de los transistores, y la resistencia y capacitancia equivalentes de los elementos de almacenamiento de memoria. Los métodos para estimar las resistencias y capacitancias de cables y parásitos son modificados a partir de los métodos utilizados en CACTI. Las mejoras se basan en la última actualización de los modelos de transistores según el ITRS Report mas reciente [7].

Considerando que NVSim es una herramienta de estimación a nivel de sistema, solo modelamos el comportamiento de los elementos de almacenamiento y registramos las resistencias y capacidades equivalentes de los estados RESET y SET. La latencia general de acceso a la memoria y el consumo de energía se estiman combinando todos los valores de temporización y potencia de los componentes del circuito. NVSim sigue la misma metodología que CACTI con ligeras modificaciones.

5.2.4. Modelos de detección de datos

A diferencia de otros circuitos periféricos, el amplificador de detección es un diseño analógico en lugar de un diseño lógico. Por lo tanto, NVSim desarrolla un modelo de temporización aparte para los esquemas de detección de datos. Los diferentes esquemas de detección tienen su impacto en la compensación entre rendimiento, energía y área. NVSim considera tres tipos de esquemas de detección: current sensing, current-in voltage sensing, y voltage-divider sensing. En el esquema current sensing, el estado de la celda de memoria se lee midiendo la corriente resultante a través de la celda de memoria seleccionada cuando se aplica un voltaje de lectura. La corriente en la línea de bits se compara con la corriente de referencia generada por las celdas de referencia, la diferencia de corriente se amplifica y, finalmente, se convierten en señal digital. Current-in voltage sensing es un método de detección alternativo que funciona aplicando una corriente en la celda de memoria seleccionada y detectando el voltaje a través del amplificador. El esquema voltage-divider sensing se presenta al introducir una resistencia en serie con la celda de memoria. El valor de resistencia se selecciona para lograr el margen máximo de detección de lectura.

5.2.5. Validación de los resultados de NVSim

NVSim se basa en supuestos de diseños de celdas de memoria, reglas de diseño de circuitos y parámetros de fabricación CMOS. Tal y como se describe en las secciones anteriores, NVSim proporciona un conjunto de opciones en NVSim para ajustar los parámetros de diseño, como la organización de la memoria, el tipo de cable, el tipo de transistor, los esquemas de detección de datos, etc. Por lo tanto, NVSim es capaz de emular un chip de memoria real, y al comparar el resultado de la estimación de NVSim con los parámetros reales del chip de memoria se debería hallar la precisión de NVSim. NVSim está validado con chips NAND Flash y con prototipos industriales de STT-RAM, PCRAM y ReRAM en términos de área, latencia y energía. La información de las especificaciones de diseño de chips reales fue utilizada para establecer los parámetros de entrada requeridos por NVSim, como la capacidad, el tamaño de línea, el nodo tecnológico y la organización de la matriz. Luego, se compararon los números de estimación de rendimiento, energía y área generados por NVSim con los números reales reportados en los diseños de chip, y el margen de error obtenido en la comparación es inferior al 30%.

5.3. Modelado de NVM

En este apartado se comparan los resultados de los modelos de diversas tecnologías de memoria realizados en el simulador NVSim. Para ello se han realizado diversas simulaciones de cada tecnología, cada una con un objetivo de optimización distinto, para encontrar así la mejor opción con una capacidad y un tamaño de palabra específico. El modelo de SRAM y STT-RAM se ha simulado para una memoria caché de 512KB y 8MB de capacidad y un tamaño de palabra de 64 bits. El modelo de DRAM, NAND Flash, RRAM y PCRAM se basa en una memoria RAM con 256MB y 1GB de capacidad, con un tamaño de palabra de 64 bits. Hay que tener en cuenta en esta última que estas simulaciones son a nivel de chip y no de módulo de memoria RAM, por tanto no se disponen datos más específicos sobre el rendimiento de éstos en una computadora. El consumo estático aparece solo en los mejores casos por que ha sido simulado a posteriori.

Tipo optimización	Área (mm ²)	Latencia Hit/Miss/Write (ns)	Consumo por Hit/Miss/Write (nJ)	Consumo estático (mW)
lectura	0.757	0.481/ 0.245/ 0.165	0.046/ 0.046/ 0.046	394.157
Escritura	0.794	0.488/ 0.252/ 0.161	0.051/ 0.051/ 0.051	-
Consumo Lectura	0.600	10.674/ 8.187/ 7.655	0.009/ 0.009/ 0.006	-
Consumo Escritura	0.466	4.531/ 3.853/ 3.835	0.037/ 0.037/ 0.002	-
Área	0.457	25.736/ 3.992/ 16.140	0.117/ 0.117/ 0.006	-

Figura 37. Caché SRAM de 512KB.

Tipo optimización	Área (mm ²)	Latencia Hit/Miss/Write (ns)	Consumo por Hit/Miss/Write (nJ)	Consumo estático (W)
lectura	8.269	3.405/ 0.769/ 1.598	0.113/ 0.113/ 0.111	4.476
Escritura	9.100	3.570/ 0.783/ 1.557	0.130/ 0.130/ 0.129	-
Consumo Lectura	8.503	92.993/ 54.665/ 53.837	0.036/ 0.036/ 0.024	-
Consumo Escritura	6.970	32.779/ 23.166/ 23.148	0.527/ 0.527/ 0.007	-
Área	6.929	383.998/ 33.788/ 232.438	1.766/ 1.766/ 0.022	-

Figura 38. Caché SRAM de 8MB.

Tipo optimización	Área (mm ²)	Latencia Hit/Miss/Write (ns)	Consumo por Hit/Miss/Write (nJ)	Consumo estático (mW)
lectura	0.336	1.844/ 1.720/ 10.191	0.098/ 0.098/ 0.100	103.657
Escritura	0.453	1.861/ 1.720/ 10.182	0.119/ 0.119/ 0.122	-
Consumo Lectura	0.193	8.608/ 3.824/ 14.856	0.049/ 0.049/ 0.053	-
Consumo Escritura	0.285	8.176/ 6.580/ 15.060	0.189/ 0.189/ 0.053	-
Área	0.176	56.032/ 39.886/ 48.360	0.058/ 0.058/ 0.056	-

Figura 39. Caché STT-RAM de 512KB.

Tipo optimización	Área (mm ²)	Latencia Hit/Miss/Write (ns)	Consumo por Hit/Miss/Write (nJ)	Consumo estático (mW)
lectura	3.667	3.318/ 2.245/ 10.703	0.197/ 0.197/ 0.199	285.438
Escritura	4.046	3.378/ 2.305/ 10.703	0.216/ 0.216/ 0.219	-
Consumo Lectura	2.785	173.588/ 33.309/ 129.230	0.062/ 0.062/ 0.059	-
Consumo Escritura	2.668	173.752/ 33.309/ 129.248	0.064/ 0.064/ 0.058	-
Área	2.608	612.945/ 101.849/ 433.490	0.159/ 0.159/ 0.093	-

Figura 40. Caché STT-RAM de 8MB.

Para la memoria caché de 512KB el resultado con mejores prestaciones en general ha sido el modelo en el que el objetivo de optimización ha sido la latencia de lectura. Si comparamos este modelo con la caché STT-RAM obtenemos que la SRAM ocupa 2,25 veces más área que la STT-RAM. Además, las latencias en un hit, en un miss o en un write son 4, 7 y 60 veces menor, respectivamente. Por último, SRAM consume aproximadamente la mitad del consumo de STT-RAM. Para la cache de 8MB disminuye la distancia entre las dos memorias. SRAM sigue siendo un 44% más grande, sin embargo, la latencia de STT-RAM en un hit se equipara a la de SRAM, y las latencias en los miss y en los write pasan a 3 y 7 veces más lentas que SRAM, respectivamente. En cuanto a consumos dinamicos, SRAM consume un 40% menos que STT-RAM. En resumen, STT-RAM mejora el área y consumo estático pero empeora la latencia y el consumo dinámico por el momento.

Tipo optimización	Área (mm ²)	Lat. lectura(ns)	Lat. escritura (ns)	Consumo lectura(pJ)	Consumo escritura (pJ)	Consumo estático (W)
lectura	12.925	7.057	7.543	106.648	98.721	103.458
Escritura	15.047	7.645	7.069	114.397	110.529	-
Consumo Lectura	13.811	159.503	158.994	67.448	39.803	-
Consumo Escritura	10.742	193.783	193.783	3.866nJ	16.721	-
Área	10.472	2.602us	2.602us	15.052nJ	40.848	-

Figura 41. RAM DRAM de 256MB.

Tipo optimización	Área (mm ²)	Lat. lectura(ns)	Lat. escritura (ns)	Consumo lectura(pJ)	Consumo escritura (pJ)	Consumo estático (W)
lectura	46.644	14.814	14.216	222.319	191.885	373,152
Escritura	51.700	24.96	13.052	215.960	208.033	-
Consumo Lectura	54.851	624.815	622.779	136.324	81.018	-
Consumo Escritura	42.299	752.607	752.607	15.216nJ	33.325	-
Área	41.763	10.313us	10.313us	60.036nJ	81.605	-

Figura 42. RAM DRAM de 1GB.

Tipo optimización	Área (mm ²)	Lat. lectura(ns)	Lat. borrado/ programming	Consumo lectura (nJ)	Consumo(nJ) borrado(bloque)/ Programming(página)	Consumo estático (W)
lectura	33.793	54.374	1.250ms/ 200.015us	33.651	53.328/ 35.870	37.462
Escritura	24.020	403.435	1.250ms/ 200.006us	862.937	17.491/ 1.870uJ	-
Consumo Lectura	28.876	941.716	1.251ms/ 200.907us	7.536	41.301/ 11.974	-
Consumo Escritura	28.876	941.716	1.251ms/ 200.907us	7.536	41.301/ 11.974	-
Área	17.418	10.877us	1.254ms/ 203.642us	2.706uJ	62.475/ 6.700uJ	-

Figura 44. RAM NAND Flash de 1GB.

Tipo optimización	Área (mm ²)	Lat. lectura(ns)	Lat. borrado/ programming	Consumo lectura (nJ)	Consumo(nJ) borrado(bloque)/ Programming(página)	Consumo estático (W)
lectura	10.096	34.286	1.250ms/ 200.003us	18.564	47.147/ 20.792	14.111
Escritura	6.118	447.032	1.250ms/ 200.002us	849.481	17.345/ 1.866uJ	-
Consumo Lectura	7.763	256.536	1.250ms/ 200.227us	3.933	21.002/ 6.149	-
Consumo Escritura	7.763	256.536	1.250ms/ 200.227us	3.933	21.002/ 6.149	-
Área	4.386	4.487us	1.251ms/ 200.912us	678.555	15.633/ 1.677uJ	-

Figura 43. RAM NAND Flash de 256MB.

Tipo optimización	Área (mm ²)	Lat. lectura(ns)	Lat. SET/RESET(ns)	Consumo lectura(pJ)	Consumo SET/RESET(nJ)	Consumo estático (W)
lectura	12.593	7.518	44.363/ 154.363	113.967	6.585/ 6.583	12.956
Escritura	15.150	8.256	44.138/ 154.138	196.569	6.630/ 6.626	-
Consumo Lectura	14.020	86.538	123.318/ 233.318	102.700	6.571/ 6.559	-
Consumo Escritura	11.098	324.996	3663.538/ 473.538	187.164	6.560/ 6.535	-
Área	9.795	5.160us	5.198/ 5.308us	4.025nJ	6.910/ 6.706	-

Figura 45. RAM PCRAM de 256MB.

Tipo optimización	Área (mm ²)	Lat. lectura(ns)	Lat. SET/RESET(ns)	Consumo lectura(pJ)	Consumo SET/RESET(nJ)	Consumo estático (W)
lectura	47.227	24.852	53.002/ 163.002	234.289	6.696/6.694	48.044
Escritura	49.174	25.305	52.993/ 162.993	370.119	6.747/6.741	-
Consumo Lectura	48.716	42.889	71.422/ 181.422	210.838	6.679/6.673	-
Consumo Escritura	40.888	1.294us	1.332/ 1.442us	599.142	6.638/6.587	-
Área	39.077	20.627us	20.665/ 20.775us	4.039nJ	6.924/6.720	-

Figura 46. RAM PCRAM de 1GB.

Tipo optimización	Área (mm ²)	Lat. lectura(ns)	Lat. escritura(ns)	Consumo lectura(pJ)	Consumo escritura(pJ)	Consumo estático (W)
lectura	7.883	6.822	23.536	139.773	170.932	-
Escritura	10.347	7.195	23.173	198.094	229.167	-
Consumo Lectura	11.273	10.429	27.025	108.600	196.280	-
Consumo Escritura	9.596	6.920	23.193	113.070	140.302	26.657
Área	6.482	22.088	36.893	209.201	295.427	-

Figura 47. RAM RRAM de 256MB.

Tipo optimización	Área (mm ²)	Lat. lectura(ns)	Lat. escritura(ns)	Consumo lectura(pJ)	Consumo escritura(pJ)	Consumo estático (W)
lectura	31.530	17.690	28.970	267.779	298.938	69.597
Escritura	31.530	17.690	28.970	267.779	298.938	-
Consumo Lectura	31.530	17.691	28.971	201.466	232.626	-
Consumo Escritura	31.530	17.691	28.971	201.466	232.626	-
Área	25.927	48.248	49.972	330.629	416.855	-

Figura 48. RAM RRAM de 1GB.

En el caso de la RAM de 256MB el mejor resultado ha sido el modelo optimizado para la latencia de lectura en el caso de DRAM, Flash y PCRAM y el modelo

optimizado para el consumo de escritura en el caso de RRAM. En cuanto a tamaños RRAM es la que consigue un área menor, aunque no mucho más pequeña que Flash. En cuanto a latencia de lectura RRAM es la más rápida, siendo 5 veces más rápida que Flash, mientras que DRAM y PCRAM son solo 0.2 y 0.6ns más lentas, respectivamente. DRAM obtiene el mejor resultado en latencias de escritura, 24000 veces más rápida que Flash, entre 6 y 15 veces más rápida que PCRAM y solo 3 veces más rápida que RRAM. Por último, DRAM vuelve a destacar siendo la memoria con menores consumos. en la lectura obtiene resultados ligeramente menores que los de PCRAM y RRAM, y consigue consumos 160 veces menores que Flash. En la escritura RRAM consume un poco menos que el doble que DRAM, PCRAM consume 66 veces más que DRAM, y ésta consume 210 veces menos que Flash. Para la memoria RAM de 1GB el mejor resultado que han tenido las distintas tecnologías han sido los modelos en los que el objetivo de optimización ha sido la latencia de lectura. el que obtiene mejores resultados en área es RRAM seguido muy de cerca por NAND Flash. En cuanto a latencia de lectura es DRAM la que consigue el mejor resultado, siendo 1.2, 3.6 y 1.6 veces más rápida que DRAM, Flash y PCRAM, respectivamente. También es DRAM la memoria con mejor latencia de escritura, siendo 2 veces más rápida que RRAM, casi 14000 veces más rápida que Flash y aproximadamente 12 veces más rápida que PCRAM. En cuanto a consumos, en la lectura DRAM, PCRAM y RRAM consumen más o menos lo mismo, unas 140 veces menos que Flash, mientras que en la escritura DRAM vuelve a destacar consumiendo 1,5 veces menos que RRAM, 30 veces menos que PCRAM y 180 veces menos que Flash. En cuanto a consumos estáticos PCRAM y Flash tienen consumos muy similares, RRAM consume un poco menos del doble que éstas, y DRAM alcanza consumos estáticos muy altos, 7 veces más altos para la memoria de 256MB y 10 veces más altos para la memoria de 1GB

Los datos obtenidos en los modelos anteriores hacen ver que por el momento las NVM emergentes no alcanzan a ofrecer mejores prestaciones que las memorias ya establecidas. Por ejemplo STT-RAM mejora el área y el consumo estático de el chip pero es inferior a SRAM en latencias y consumos dinámicos. Por otro lado, RRAM y PCRAM mejoran en área y en consumo estático respecto a DRAM y en latencias y consumos dinámicos respecto a Flash, pero no alcanzan resultados lo suficientemente atractivos para sustituir a la combinación ambas. Sin embargo, teniendo en cuenta la inmadurez de las NVM emergentes, los resultados obtenidos en estas simulaciones demuestran el potencial de estos tres nuevos tipos de memoria. Con una inserción en el mercado gradual, ya sea mediante la combinación de diversas memorias creando memorias híbridas o mediante el desarrollo de nuevos productos, es muy probable que se encuentren circuitos selectores y arquitecturas de chip nuevas específicas para cada tipo de nueva tecnología, que abarate su coste y mejore sus prestaciones, o que simplemente la reducción de costes y el incremento en prestaciones venga dado por el perfeccionamiento de los procesos de manufacturación de estas. La siguiente tabla intenta resumir de forma cualitativa los resultados de las tablas anteriores [Fig. 49].

	SRAM	DRAM	Flash	STT-RAM	PCRAM	RRAM
Densidad	Baja	Alta	Muy Alta	Media	Alta	Muy Alta
Velocidad	Muy Alta	Alta	Muy Baja	Muy alta (lectura) Alta (escritura)	Media (lectura) Baja (escritura)	Alta
Consumo activo	Bajo	Muy Bajo	Muy Alto	Bajo	Muy Bajo (lectura) Alto (escritura)	Muy Bajo
Volátil	Si	Si	No	No	No	No

Fig. 49 Tabla resumen cualitativa de las prestaciones obtenidas en las simulaciones.

6. Conclusiones

Durante las últimas tres décadas y media de existencia, la familia de memorias de semiconductores se ha expandido enormemente y ha alcanzado densidades más altas, velocidades más altas, menor potencia, más funcionalidad y menores costos. En los últimos 40 a 50 años, Las tecnologías de memoria no volátil basadas en semiconductores (NVSM) han evolucionado del concepto de puerta flotante a muchas otras formas de almacenar la información. Desde 1990, las tecnologías NVSM han permitido el desarrollo de todos los sistemas electrónicos modernos y brindando beneficios sin precedentes a la humanidad. Aunque, con el paso del tiempo, se hacen evidentes algunas de las limitaciones dentro de cada tipo de memoria. A medida que la dimensión del dispositivo se reduce por debajo de los 10 nanómetros, NVSM enfrenta muchos desafíos de escalamiento graves, como la interfaz de las celdas vecinas, la reducción de la carga almacenada y el ruido aleatorio asociado a los semiconductores. Debido a estos problemas existe una importante actividad investigadora intentando resolverlos, y el campo de las memorias convencionales está cambiando mediante la introducción de las tecnologías no volátiles emergentes, distintas a las convencionales basadas en el almacenamiento de los datos a partir de una carga eléctrica. De éstas tecnologías se espera que complementen a las convencionales a corto plazo, ya que surgen de la necesidad de solucionar los problemas intrínsecos de escalabilidad de las tecnologías de memoria clásicas. Esta transición conlleva un importante cambio en la magnitud que representa la información binaria. El cambio de celdas basadas en el almacenamiento de carga eléctrica a celdas basadas en el cambio en la resistencia.

6.1. Trabajos futuros

En unos pocos años, la mayoría de tecnologías de las que se habla en este texto pueden experimentar cambios que generen la evolución de ésta, que puedan motivar estudios similares. Del mismo modo, nuevos productos estarán disponibles para el consumidor, y habrán surgido nuevas aplicaciones que puedan ser objeto de estudio.

También se podría expandir este trabajo en un estudio más profundo de los simuladores, así como simulaciones más completas y detalladas, que den una idea más concreta de las prestaciones que estos dispositivos son capaces de ofrecer, así como la inclusión de mas tecnologías NVM a las simulaciones, para ofrecer un espectro más amplio del panorama de la memoria.

7. Referencias

1. Jagan Singh Meena, Simon Min Sze, Umesh Chand, Tseung-Yuen Tseng: **Overview of emerging nonvolatile memory technologies**. *Nanoscale Research Letters* 2014. 9:526
2. Yuan Xie: **Modeling, Architecture, and Applications for Emerging Memory Technologies**. 0740-7475/11/; 2011 IEEE
3. O.D. Alao, J.V. Joshua, D.O. Kehinde, E.O. Ehinlafa, M.O. Agbaje, J.E.T Akinsola: **Emerging Memory Technologies**. *The International Journal Of Engineering And Science (IJES)*; 2016
4. Rodrigue Rizk, Dominick Rizk, Ashok Kumar, Magdy Bayoumi: **Demystifying Emerging Nonvolatile Memory Technologies: Understanding Advantages, Challenges, Trends, and Novel Applications**. The Center for Advanced Computer Studies University of Louisiana at Lafayette, Louisiana, USA; 2019
5. INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS 2.0: **BEYOND CMOS**; 2015
6. INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS 2.0: **MORE MOORE**; 2015
7. INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS 2.0: **EXECUTIVE REPORT**; 2015
8. Alexander Makarov, Viktor Sverdlov, Siegfried Selberherr: **Modeling Emerging Non-volatile Memories: Current Trends and Challenges**. 2012 International Conference on Solid State Devices and Materials Science
9. Sparsh Mittal, Jeffrey S. Vetter: **A Survey of Software Techniques for Using Non-Volatile Memories for Storage and Main Memory Systems**. DOI 10.1109/TPDS.2015.2442980, *IEEE Transactions on Parallel and Distributed Systems*
10. Tetsuo Endoh, Hiroki Koike, Shoji Ikeda, Takahiro Hanyu, Hideo Ohno: **An Overview of Nonvolatile Emerging Memories— Spintronics for Working Memories**. *IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS*;2016
11. Scott W. Fong, Christopher M. Neumann, and H.-S. Philip Wong: **Phase-Change Memory—Towards a Storage-Class Memory**. 0018-9383; 2017 IEEE.
12. H.-S. Philip Wong, Simone Raoux, SangBum Kim, Jiale Liang, P. Reifenberg, Bipin Rajendran, Mehdi Asheghi, Kenneth E. Goodson: **Phase Change Memory**. 0018-9219; 2010 IEEE
13. G. Molas, G. Piccolboni, M. Barci, B. Traore, J. Guy, G. Palma, E. Vianello, P. Blaise, J. M. Portal, M. Bocquet, A. Levisse, B. Giraud, J. P. Noel, M. Harrand, M. Bernard, A. Roule, B. De Salvo, L. Perniola: **Functionality and reliability of resistive RAM (RRAM) for non-volatile memory applications**. 978-1-4673-9478-9/16; 2016 IEEE
14. Supriya Karmakar: **Quantum dot gate non-volatile memory as single level cell (SLC), multi-level cell (MLC) and triple level cell (TLC)**. Farmingdale State College-SUNY, Farmingdale, NY, 11735



15. Shimeng Yu: **Resistive Random Access Memory (RRAM)**; March 2016
16. An Chen: **A review of emerging non-volatile memory (NVM) technologies and applications**. 0038-1101/ 2016
17. Kosuke Suzuki, Steven Swanson: **A Survey of Trends in Non-Volatile Memory Technologies: 2000–2014**. 978-1-4673-6933-6/15; 2015 IEEE
18. Shimeng Yu, Pai-Yu Chen: **Emerging Memory Technologies**. 10.1109/MSSC.2016.2546199
19. GIANLUCCA O. PUGLIA, AVELINO FRANCISCO ZORZO, CÉSAR A. F. DE ROSE, TACIANO D. PEREZ, DEJAN MILOJICIC : **Non-Volatile Memory File Systems: A Survey**. 10.1109/ACCESS.2019.2899463
20. Miquel Albert Orega, Gerard Enrique Manonellas: **El procesador**. PID_00177072
21. Oscar borja: **REGISTROS** [Online]. Available: <https://www.monografias.com/trabajos7/regi/regi.shtml>
22. José Manuel Mendías Cuadros: **Memoria caché**. Dpto. Arquitectura de Computadores y Automática Universidad Complutense de Madrid
23. **Memoria RAM** [Online]. Available: <https://sites.google.com/site/modulodefundamentosdehardware/5--memoria-ram>
24. **Memoria RAM** [Online]. Available: <https://sites.google.com/site/hardwarejoseangel/memoria-ram>
25. Roger Bedoya: **Disco Duro (fundamentos de Hardware)** [Online]. Available: <http://rogerbedoya.blogspot.com/2010/05/disco-duro-fundamentos-de-hardware.html>
26. CORPORACIÓN DE INGENIERÍA EN CIRCUITOS INTEGRADOS: **TECNOLOGÍA SRAM** [Online]. Available: http://www.ieec.uned.es/investigacion/Dipseil/PAC/archivos/Tecnologia_SRAM.pdf
27. **Tecnología de los sistemas de memoria**. Departamento de Arquitectura y tecnología de computadores. Universidad de Sevilla [Online]. Available: <http://icaro.eii.us.es/descargas/Tecnologia%20de%20Memorias%20AC%20curso%2004-05.pdf>
28. **Flash** [Online]. Available: https://es.wikipedia.org/wiki/Memoria_flash
29. Everspin Technologies: **MR4A08B and MR4A16B Families** [Online]. Available: <https://www.everspin.com/getdatasheet/MR4A16B>; <https://www.everspin.com/getdatasheet/MR4A08B>
30. Everspin Technologies: **EMD3D256M Family** [Online]. Available: <https://www.everspin.com/getdatasheet/EMD3D256M>
31. Avalanche Technology: **High Performance and Ultra-Low Power Families** [Online]. Available: <http://www.avalanche-technology.com/wp-content/uploads/2019/08/1Mb-16Mb-Serial-ULP-MRAM.pdf>; <http://www.avalanche-technology.com/wp-content/uploads/2019/08/1Mb-16Mb-Serial-HP-MRAM.pdf>
32. Avalanche Technolog: **Parallel Persistent SRAM families** [Online]. Available: <http://www.avalanche-technology.com/wp-content/uploads/2019/08/1Mb-16Mb-Parallel-x16-MRAM.pdf>

33. Avalanche Technology: **Embedded MRAM** [Online]. Available: <http://www.avalanche-technology.com/wp-content/uploads/2019/06/8Mb-64Mb-eMRAM.pdf>
34. Intel: **Intel Optane** [Online]. Available: <https://www.intel.com/content/www/us/en/products/memory-storage/optane-memory/optane-memory-m10-series/optane-memory-m10-16gb-m-2-80mm.html>;
<https://www.intel.com/content/www/us/en/products/memory-storage/optane-memory/optane-memory-m10-series/optane-memory-m10-64gb-m-2-80mm.html>
35. **Panasonic: 8-bit Low power Microcomputer MN101L Series** [Online]. Available: <https://na.industrial.panasonic.com/products/semiconductors/microcontrollers/lineup/8-bit-low-power-microcomputers-mn101l-series/series/70510/model/72639>
36. Crossbar: **Crossbar eRRAM** [Online]. Available: <https://www.crossbar-inc.com/products/p-series/>
37. Crossbar: **Crossbar 3D eRRAM** [Online]. Available: <https://www.crossbar-inc.com/products/t-series/>
38. Xiangyu Dong, Cong Xu, Yuan Xie, Norman P. Jouppi: **NVSim: A Circuit-Level Performance, Energy, and Area Model for Emerging Nonvolatile Memory**. IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, VOL. 31, NO. 7, JULY 2012
39. Herbert J. Kramer: **SpriteSat (Rising)** [Online]. Available: <https://directory.eoportal.org/web/eoportal/satellite-missions/s/spritesat>. , eoPortal Directory; Sharing Earth Observation Resources.
40. Kazuya YOSHIDA, Yukihiro TAKAHASHI, Yuji SAKAMOTO, Eriko UJIIE, Kei TAKIUCHI, Yasuhiro NAKAZATO, Tomoki SAWAKAMI, Takeshi SAKANOI, Yasumasa KASABA, Satoshi KONDO, Kozo YAMASHITA, Shinya UEDA, Takeshi TAKASHIMA, Kazuhiro NAKAZAWA, Takefumi MITANI, Teruaki ENOTO, Mitsuteru SATO, Umran INAN, Ivan LINSKOTT, Fredrik BRUHN, Yoshinari MASUMOTO: **SPRITE-SAT: a Micro Satellite for Scientific Observation of Transient Luminous Events and Terrestrial Gamma-ray Flashes**. Trans. JSASS Aerospace Tech. Japan Vol. 8, No. ists27, pp. Tm_7-Tm_12, 2010.
41. Nanomaterial Mechanics Explorer [Online]. Available: <https://nanohub.org/resources/nanomatmech>
42. Memristor Simulation Tool [Online]. Available: <https://nanohub.org/resources/memristor>
43. Magnetic Tunnel Junction Lab [Online]. Available: <https://nanohub.org/resources/mtjlab>
44. SPICE Subcircuit Generator for Ferromagnetic Nanometaterials [Online]. Available: <https://nanohub.org/resources/fmmacrospace>
45. Panasonic demonstrate battery-less NFC electronic shelf labels with his RRAM MCUs [Online]. Available: <https://www.rram-info.com/panasonic-demonstrate-battery-less-nfc-electronic-shelf-labels-its-reram-mcus>
46. CACTI [Online]. Available: <https://www.hpl.hp.com/research/cacti/>

47. nanoHUB [Online]. Available: <https://nanohub.org/about/presskit>
48. MASTAR tool [Online]. Available:
[https://ipfs.io/ipfs/QmXoyvizjW3WknFiJnKLwHCnL72vedxjQkDDP1mXWo6uco/wiki/MASTAR_\(software\).html](https://ipfs.io/ipfs/QmXoyvizjW3WknFiJnKLwHCnL72vedxjQkDDP1mXWo6uco/wiki/MASTAR_(software).html)