

Synaptic axons and dendrits of HPC: memory controllers.

Axones y dendritas sinápticos del HPC: controladores de memoria.

María Daniela Lizarazo Sandoval
Escuela de Ingeniería de Sistemas e Informática
Universidad Industrial de Santander
Bucaramanga, Colombia
danimari98@hotmail.com

Lizeth Katherine Rey Olarte
Escuela de Ingeniería de Sistemas e Informática
Universidad Industrial de Santander
Bucaramanga, Colombia
katherey97@gmail.com

John Sebastian Martinez Cifuentes
Escuela de Ingeniería de Sistemas e Informática
Universidad Industrial de Santander
Bucaramanga, Colombia
sebas1997.sm@gmail.com

Carlos Arley Santos Sarmiento
Escuela de Ingeniería de Sistemas e Informática
Universidad Industrial de Santander
Bucaramanga, Colombia
carlos.santos.04.20@gmail.com

Abstract

With the boom of supercomputation, multiprocessing, and the big volume of information that all the above process and at the same time generates, it has become necessary to modify the way in which data and processes are stocked and organized in the RAM, just like a pile of LEGOS to build an architectonic masterpiece. This work aims to explore the challenges that different types of architectures of memory controllers (DRAM, flash memory, SDRAM) have, in order to satisfy the demands of high performance applications and masive storage, and the implications of it.

Keywords — memory controllers, RAM, applications, high performance computing

Resumen

Con el auge de la supercomputación, los multiprocesadores, y los grandes volúmenes de información que todo lo anterior procesa y a su vez genera, se ha vuelto necesario modificar la forma en que se almacenan y organizan datos y procesos en la RAM, es como si en un cerebro se replanteará la forma en que los axones y dendritas permiten la sinápsis entre neuronas. Este trabajo busca explorar los retos que las distintas arquitecturas de controladores de memoria (DRAM, flash memory, SDRAM) tienen para satisfacer las demandas de las aplicaciones de alto rendimiento y almacenamiento masivo, y las implicaciones de ello.

Palabras claves — controladores de memoria, RAM, aplicaciones, computación de alto rendimiento.

I. INTRODUCCIÓN

El concepto de memoria en la computación se asemeja a la de un cerebro humano ya que es el espacio de una computadora que se usa para el almacenamiento de datos y las instrucciones de procesamiento. Se dice que funciona como el cerebro humano ya que al igual que él posee una gran cantidad de celdas interconectadas (lo que en nuestro cerebro llamaríamos neuronas) que permiten su funcionamiento. Gracias a estas interconexiones la memoria es capaz de procesar información que viene tanto del interior como del exterior. Dicho de otra manera, la memoria es el conjunto de bits que almacena caracteres de manera temporal o permanente. El computador funciona con varios tipos de memoria: ROM, RAM, caché interna, caché externa, de video, etc.

Ahora bien, el controlador de memoria es un bloque lógico que realiza lecturas y/o escrituras desde una memoria basada en la tecnología de memoria. Qué tipo de controladores se usan y la forma en que se organizan puede variar, y depende altamente de los requerimientos del sistema. En la comparación con nuestro cerebro, los controladores de memoria vendrían a ser el medio por el cual se logra que dos neuronas procesen información, es decir, los axones y dendritas de las neruonas, siendo la sinápsis la memoria. Sin embargo, tal y como ocurrió con la evolución del cerebro humano, la forma en que dicho proceso fue cambiando a medida que el medio exterior le representaba retos o requerimientos al ser humano, y lo mismo

ocurre actualmente con los controladores de memoria.

LA SINAPSIS

Las neuronas cumplen la importante función de transmitir señales a cada célula, y la sinapsis es precisamente la vía empleada para ello.

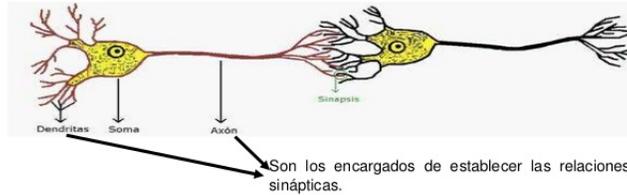
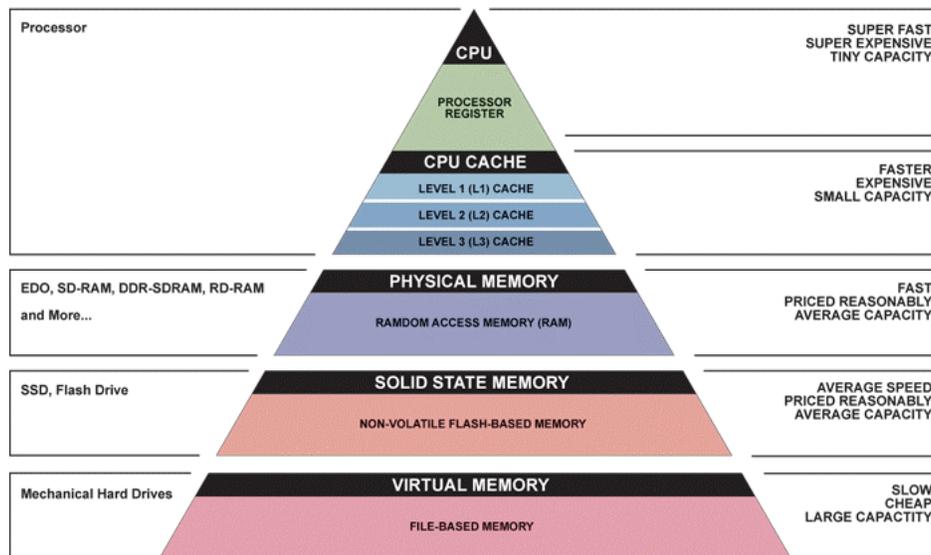


Figura 1. Proceso de sinápsis.

En la redacción posterior ahondaremos tres puntos importantes para comprender a cabalidad el tema de controladores de memoria: estableceremos una breve definición y una descripción de su funcionamiento, la importancia que tiene en la actualidad y la importancia en las aplicaciones. En la sección 2 se establece un recuento al estado del arte, posteriormente se dan definiciones relevantes al controlador de memoria y se explica cómo es su funcionamiento tomando en cuenta las operaciones que debe realizar un ordenador. En la sección 3 se establecera su importancia y se clasificarán algunos de los tipos de arquitecturas de memoria, especialmente para computación de alto rendimiento, que se encuentran presentes en el ordenador y la funcionalidad de los controladores que poseen. En la sección 4 hablaremos de la importancia en la capa de aplicación y por último se estableceran conclusiones en base a todo lo anterior.



▲ Simplified Computer Memory Hierarchy
Illustration: Ryan J. Leng

Figura 2. Clasificación de la memoria.

II. ESTADO DEL ARTE

Actualmente, es común que los controladores de memoria se alojen en la memoria RAM, pero esto no fue siempre así. Hubo un tiempo en el que se implantaron en el North Bridge de la placa base debido a que el bus que une al procesador con el North Bridge, solía ser el que mayor banda ancha tenía, evitando el tan famoso cuello de botella entre el controlador y el procesador. La integración de estos controladores surgen por primera vez en los ordenadores de escritorio, un poco después del

año 2000, con la llegada de los procesadores AMD ATHLON 64. Sin embargo, Intel siguió por más de 5 años implantando los controladores de memoria por fuera del procesador, hasta la llegada del Intel Nehalem, en el año 2008. [2] Como

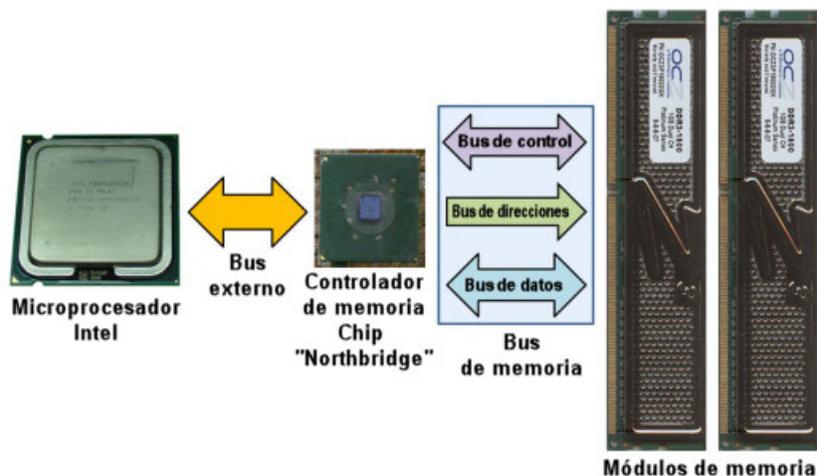


Figura 3. Funcionamiento de controlador de memoria Intel.

todo proceso de intervención, existen las ventajas y desventajas acerca de la instalación de los controladores de memoria en el procesador. Para empezar, la latencia de acceso a la memoria RAM se reduce de manera considerable; lo cual quiere decir que ni el procesador ni el North Bridge tienen que esperar al otro la información que se envíen, obtenida por la memoria. Ahora, el inconveniente del controlador de memoria integrado más importante es que el procesador sea dependiente de un solo tipo de memoria. El hecho de integrarlo una actualización del tipo de memoria que va a emplear el procesador supone rediseñar el propio controlador interno, la distribución de los pines y, en la mayoría de casos, el propio socket. A menos que el controlador del procesador se diseñe desde el principio para ser compatible con dos tipos de memoria, algo que sucedió con los procesadores Skylake, que eran compatibles con RAM DDR3 y DDR4.

El imparable progreso de hacer todo lo electrónico más pequeño ha hecho posible que muchos de estos componentes (procesador, memoria, periféricos) sean integrados en un solo paquete, llamado SoC (sistema en un chip). En HPC, es común el uso de los MPSoC, un sistema en chip multiprocesador que es un sistema en chip (SoC) que incluye múltiples microprocesadores. Dichas arquitecturas, también han traído consigo retos a la hora de implementarlos con controladores de memoria.

III. CONTROLADORES DE MEMORIA

A. Definiciones

La memoria RAM es la memoria principal de cualquier dispositivo donde almacenamos datos y programas de manera volátil, es decir, no permanentes, por ello, al momento de no haber fuente de energía en el dispositivo, esta deja de funcionar, perdiéndose toda la información alojada en ella. La importancia de la RAM es vital en un dispositivo, ya que todo programa pasa por esta para ejecutarse. Cualquier persona con conocimientos básicos en cuanto a ordenadores sabe perfectamente lo que conlleva tener una gran cantidad de memoria RAM: mayor procesamiento de datos, y por ende, los trabajos se realizarán a mayor velocidad. [1]

Su manera de actuar no es compleja, ya que su sistema accede a los datos almacenados de manera no secuencial, por eso al momento de ejecución de varios programas a la vez en un dispositivo es donde se hace más participe. Todos estos procesos conllevan acciones tomadas por diferentes chips dentro de la RAM. Los controladores de memoria actúan en la gestión de operaciones de lectura y escritura en la memoria del sistema, como también mantener la memoria activa al momento en el que se le esté suministrando corriente eléctrica. De esta manera, la memoria RAM se vuelve una solución más rápida comparada a otros tipos de almacenamiento como pueden ser los discos duros.

B. Funcionamiento

Como ya se mencionó antes, uno de los factores vitales es el suministro de la cantidad adecuada de flujo de energía. Una función principal de los controladores de memoria es cubrir esta necesidad refrescando la RAM a un ritmo constante mientras el ordenador o dispositivo está encendido.

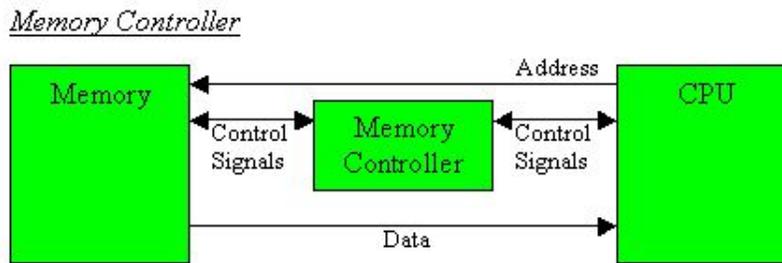


Figura 4. Diagrama de bloque básico de un controlador de memoria.

Para entender un poco mejor el funcionamiento, podemos pensar en que un ordenador es una casa y los circuitos comentados son las direcciones de calle. Para querer enviar un correo a alguna otra casa, el dispositivo o ordenador debe saber cuál es la dirección para enviar el paquete. El controlador de memoria es un intermediario en este proceso, ya que se encarga de que la información a enviar sea correcta a la localización adecuada.

IV. IMPORTANCIA DE LOS CONTROLADORES

Como se decía, las memorias son indispensables en cualquier sistema computacional, y el rendimiento de dichos dispositivos es crítico en MPSoCs (Sistemas multiprocesadores en chip) tanto en dominios de tiempo real como en alto-rendimiento (HPC). Los sistemas actuales tienen jerarquías de memoria complejas con distintos tipos de memorias volátiles y no volátiles, como DRAM y flash. Por los controladores de memoria en los sistemas tienen la tarea de manejar tales dispositivos de manera que se provea suficiente capacidad y rendimiento en tiempo real a los clientes de memoria de manera confiable, al mismo tiempo que se limita el costo y consumo de energía. [3] Para ello, los controladores de memoria tienen arquitecturas avanzadas, reglas y algoritmos que continúan evolucionando a medida que se presentan retos computacionales. Para poder ahondar en ello, se presentan los objetivos principales de cualquier controlador.

A. Objetivo de los controladores de memoria

Comúnmente cuando hablamos de controladores de memoria hacemos referencia al controlador de memoria principal (DRAM).

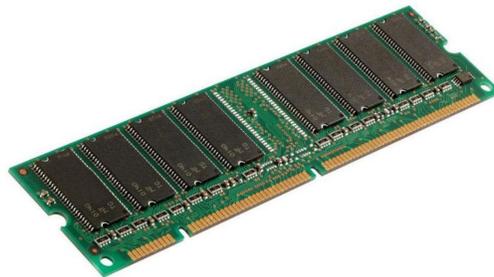


Figura 5. Memoria DRAM.

Al contener la lógica necesaria para leer y escribir en la DRAM (tipo de memoria semiconductor de acceso aleatorio que almacena cada bit de datos en un pequeño capacitor dentro de un circuito), si la DRAM no es actualizada de manera constante perderá los datos que se hayan guardado en ella, esto se debe a que los condensadores pierden su carga en fracción de segundos, en esto radica la importancia de que un controlador de memoria esté presente en el sistema.

El controlador de la memoria DRAM se encarga de traducir las direcciones que provienen de la lectura o escritura de la CPU en una dirección DRAM (las direcciones son dadas por datos de fila, columna, tamaño, etc.). Estos datos se dan a través de multiplexores y demultiplexores los cuales los utilizan para seleccionar la ubicación de memoria correcta y devolver los datos, esto se hace con la intención de que al pasar los datos de nuevo por el multiplexor estos se consoliden y reduzcan el ancho de bus requerido en la operación.

El ancho de bus de los controladores de memoria está constituido por el número de líneas paralelas disponibles para comunicarse con la celda de memoria. Los anchos de bus de los controladores de memoria varían desde 8 bits hasta 512 bits [5].

B. Mayor seguridad con más niveles de traducción.

A pesar que son controladores experimentales existen algunos que permiten un segundo nivel de traducción para las direcciones de memoria o incluso hay algunos Intel que ya tienen implementada una codificación de la memoria la cual tendría como función convertir los datos del usuario que se encuentran guardados en la memoria principal en unos patrones pseudoaleatorios, estos sistemas de seguridad evitan cualquier tipo de obtención, análisis, recuperación de datos o información privada, además de que previenen casos en los que la información sea desconstruida con la intención de revelar su estructura, arquitectura o diseño para extraer información que pueda terminar siendo lucrativa. A finales del 2010 no se había logrado que se cumplieran estas funciones y lo máximo que hacía la codificación de memoria era resolver problemas eléctricos relacionados con la DRAM.

V. VARIANTES PARA LOS CONTROLADORES DE MEMORIA.

Anteriormente se dijo que cuando se hablaba de los controladores de memoria normalmente se refería a los controladores de la memoria DRAM, sin embargo, existen un par de controladores de memoria para las de tipo DDR, SDRAM, doble canal, completamente almacenada y memoria flash. Para el ámbito de HPC resultan relevantes la DDR, SDRAM y la memoria flash: [4]

1) *Memoria SDRAM*: La memoria sincrónica de acceso aleatorio (SDRAM) es igual a la DRAM con la excepción de que la DRAM es asincrónica, es decir, se mantiene sincronizada con el reloj de la computadora, lo que permite una mayor eficiencia para guardar y acceder a información en comparación con la memoria asincrónica DRAM.



Figura 6. Memoria SDRAM.

2) *Memoria DDR*: En comparación con las SDRAM las cuales soportan una sola operación de memoria a la vez, estas memorias soportan dos operaciones de memoria por ciclo de reloj proporcionando así el doble de desempeño, dándole así el nombre de “Velocidad doble de datos”.

Nombre estándar	Velocidad del reloj	Tiempo entre señales	Velocidad del reloj de E/S
DDR 200	100 MHz	10 ns	100 MHz
DDR 266	133 MHz	7,5 ns	133 MHz
DDR 300	150 MHz	7 ns	150 MHz
DDR 333	166 MHz	6 ns	166 MHz
DDR 366	183 MHz	5,5 ns	183 MHz
DDR 400	200 MHz	5 ns	200 MHz
DDR 433	216 MHz	4,6 ns	216 MHz
DDR 466	233 MHz	4,2 ns	233 MHz
DDR 500	250 MHz	4 ns	250 MHz
DDR 533	266 MHz	3,7 ns	266 MHz

Figura 7. Frecuencias de trabajo para DDR.

Los controladores de memoria de doble velocidad de datos o DDR son utilizados para controlar la SDRAM DDR en la cual la información se transfiere en el flujo de altas y bajas del reloj de memoria de la computadora, permitiendo que se transfieran casi el doble de datos al ser de doble velocidad. [6]

3) *Memoria flash*: Siendo una nueva tecnología de almacenamiento de datos no volátil que los preserva, aunque el dispositivo sea desconectado de la fuente de electricidad, se ha convertido en uno de los medios de almacenamiento favoritos. Esta clase de dispositivos no es solamente utilizada para la lectura y escritura de datos, también permite ser usado en el formateo del equipo ya que puede almacenar los datos del sistema para que sean previamente instalados [6].



Figura 8. Memorias flash.

Con el avance de los chips flash semi-conductores, la alta densidad y los bajos precios de celdas multinivel (MLC), estos han reemplazado el papel de los chips de celda de un solo nivel en muchas aplicaciones y en el mercado de los dispositivos que usan memorias flash. Sin embargo, los MLC tienen menos rendimiento de escritura/lectura, mayor error de bit, y menos duración. Por lo que el crecimiento de su capacidad y su competencia con las SCL representan retos para su aplicación en computación de alto rendimiento.

Una memoria flash usualmente consiste en un controlador, ROM para el almacenamiento del firmware, y una RAM para almacenar la traducción de direcciones y manejo de información. Puede consistir en uno o más chips de flash, y se puede clasificar en dos tipos [7]:

- SSD: Diseñada para reemplazar hard drives, consiste en varios chips (8). Son distribuidas con canales múltiples, y cada canal puede ser accedido independientemente para mejorar el grado de paralelismo cuando se accede a chips flash.
- Flash card: Como las memorias USB y las tarjetas SD, es usada para almacenar datos de un usuario y se compone de 1 o 2 chips. Requieren un costo bajo por unidad, incluyen un controlador con poder de cómputo limitado, tamaño pequeño de RAM, y bajo costo de chips flash MLC.

Ambos tipos tienen la característica de capacidad de crecimiento rápida para competir con el rendimiento de los diseños de manejo de flash, en términos de escalabilidad espacial. A su vez, los dispositivos que suelen hacer uso de memorias flash tienen frecuencia de actualización muy alta, por lo que la confiabilidad que ofrezcan los controladores de memoria es fundamental para cualquier aplicación.

VI. DESDE LAS APLICACIONES

Los diseños de computadores cuentan con una característica fundamental para su correcto funcionamiento: la abstracción. Una de las ventajas que esto ofrece es que permite que las capas no se alteren completamente entre sí en caso de que una de ellas presente una falla, sin embargo, que exista una abstracción entre capas o niveles no quiere decir que la importancia de cada una de ellas se haga invisible a medida que se asciende. Un claro ejemplo de ello son los controladores de memorias y las aplicaciones.

Las aplicaciones tienen diferentes requerimientos en tiempo real, dependiendo de la naturaleza del procesamiento y el grado de “pipelining” en los cores. Estos requerimientos se reflejan en los requerimientos de sus correspondientes clientes de memoria. Dichos requerimientos se puede clasificar en cuatro clases distintas, dependiendo de la criticidad de la aplicación.

- Requerimientos firmes en tiempo real (FRT): Para aplicaciones, como un software de radio definido, en donde la existencia de fallas al satisfacer los requerimientos de sus clientes de memoria es altamente indeseable y puede resultar en fallas para cumplir con cierto estándar, o incluso violar el correcto funcionamiento de MPSoC.

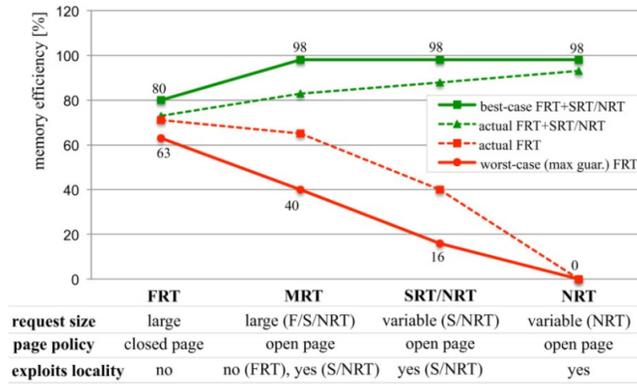


Figura 9. Mejor y peor caso de eficiencia de memoria para los distintos tipos de controladores en DRAM.

- Requerimientos suaves en tiempo real (SRT): Para aplicaciones, como los decodificadores de media, para las cuales pasarse un límite suave resulta en degradación de la calidad de la aplicación, como causar defectos visuales en videos decodificados. Aunque esto puede resultar molesto para el usuario, puede ser aceptable desde que no se repita mucho.
- No requerimientos en tiempo real (NRT): Como una interfaz gráfica de usuario, dado que sus clientes de memoria no han definido bien sus requerimientos, pero de igual manera debe ser lo suficientemente rápida para que la aplicación tenga una valoración positiva por el usuario.
- Sistemas de criticidad de tiempo mixto: Los clientes de memoria tienen una mezcla de requerimientos FRT, SRT, y NRT, lo cual representa un reto para el diseño de controladores de memoria que se ajuste a dichas combinaciones.

Por otro lado, el mercado ha demostrado un deseo insaciable por dispositivos que puedan producir consumir y comunicar cantidades y calidades cada vez mas altas de contenido media. Para ello se usan dispositivos MPSoC cuyo diseño implica un gran reto para sus arquitecturas multi-core heterogéneas: optimizar costo y rendimiento de forma agresiva del sistema fuera de chip DRAM, porque el ancho de banda deseado crece más rápidamente que la tecnología DRAM subyacente y los consumidores demandan precios bajos. [7] La TV de alta resolución, el procesamiento de imágenes y la realidad aumentada son ejemplos típicos de aplicaciones que requieren un incremento drástico de ancho de banda para la memoria. Como solución a los retos que esto representa, se ha propuesto integración 3D, entre ellos el más avanzado actualmente es el esquema memoria DRAM en procesador con una integración vertical con datos amplios interconectado a través de TSV (Through Silicon Vias), pues provee esquemas de conexión e interfaces basados en un bus de data amplio entre el MPSoC y la memoria. [8] [9]

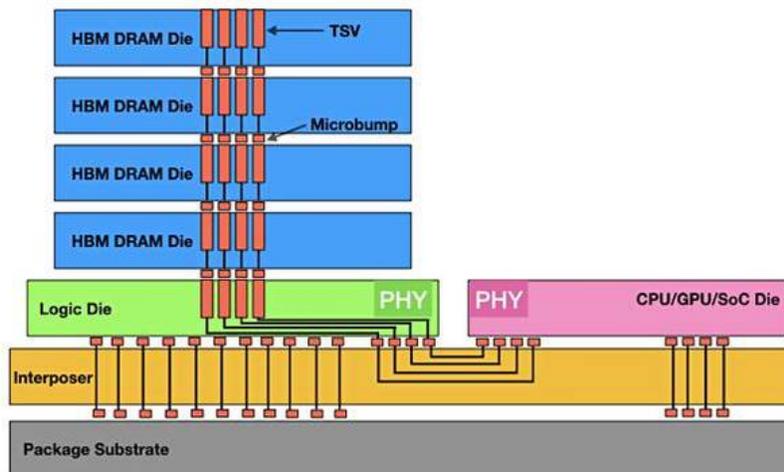


Figura 10. Memoria de alto ancho de banda "high bandwidth memory".

Debido a flexibilidad creciente, rendimiento, y restricciones de consumo de energía, las arquitecturas multi-core enfrentan muchos problemas. Otro reto, es la latencia de comunicación entre sistemas de memoria CPU y GPU al unirlos, para que puedan compartir una misma dirección de espacio de memoria. Dado que las tareas asignadas a los cores del CPU y/o GPU tienen distintas demandas de ancho de banda (bw), se necesita un sistema de memoria de 2 niveles. Para lo cual existen propuestas como Region-Aware Memory Controller (RAMON), un sistema configurable de memoria donde distintas regiones de direcciones de espacios están disponibles para dedicarse a diferentes números de controladores de memoria, iniciativa que tuvo como resultados la mejora del Bw por un factor de 9 veces para regiones de CPU, 14.1 veces para regiones GPU, y 4.5 veces para regiones heterogéneas combinadas. [10]

Plataformas virtuales como SystemC Simulations para SoCs, que es una de las tecnologías de vanguardia, se está usando en la industria para el desarrollo temprano software con poca disponibilidad de hardware. Esto requiere de un trabajo en conjunto de memorias flash y controladores de memoria en la arquitectura del SoC desde el boot-up hasta sus funciones regulares. [11]

CONCLUSIONES

Todo proceso evolutivo requiere de adaptaciones si lo que se busca es prevalecer, los seres humanos somos un ejemplo de ello, nuestro cuerpo se adaptó, dejamos órganos detrás y mejoramos otros, nuestro cerebro se volvió cada vez más complejo, nuestras neuronas y la comunicación entre ellas permitieron que construyéramos una civilización en donde hoy día tenemos máquinas que tratan de simular dicho comportamiento. Máquinas computacionales que han demostrado que tienen todo lo necesario para seguir evolucionando y adaptándose a medida que se requiera. La memoria, siendo una de las partes más importantes en la arquitectura de cualquier sistema computacional, debe ir a la par con dicha evolución, no solo en crecimiento de capacidad sino en estructuración en búsqueda de eficiencia, dado que si bien una de las observaciones de la Ley de Moore es que los procesadores se hacen cada vez más pequeños, la relación con la cantidad de implicaciones para el diseño de las arquitecturas de los controladores de memoria es inversamente proporcional, pues se deben romper paradigmas y encontrar nuevas formas de comunicar el procesador con la memoria que sean eficientes.

Lo anterior, se hace evidente en la capa del usuario: las aplicaciones, y al mismo tiempo estas mismas son generadoras de limitaciones para los controladores. Pues no basta con que sean eficientes sino que también sean de bajo costo y consumo de energía. No obstante, toda esta serie de encrucijadas para los controladores de memoria lleva a un estado constante de actualización de sí mismo que resulta beneficioso no solo para lo que se pide, sino que también para la arquitectura en general del computador, pues en el proceso de re-estructurar la memoria se pueden encontrar formas más óptimas de organizar el hardware entero, y finalmente de ello se tratan los procesos evolutivos: de encontrar en los desafíos oportunidades de mejora completa.

REFERENCIAS

- [1] Patterson, David, et al. Computer Organization and Design. Editorial MK. Chapter 6, pp. 568-580.
- [2] Usera, Juan. Por qué se emplea un controlador de memoria integrado en el procesador. Julio 2018. Recuperado de: <https://hardzone.es/2018/07/01/controlador-memoria-integrado-procesador/>
- [3] Quora. Forum: What is a memory controller. (2017). Recuperado de: <https://www.quora.com/What-is-memory-controller> [Revisado 2019].
- [4] InCo Fing. Arquitectura de computadores: Memorias. (2019). Recuperado de: <https://www.fing.edu.uy/tecnoinf/mvd/cursos/arqcomp/material/teo/arq-teo09.pdf> [Revisado 2019].
- [5] Uruguay OC. Una Explicación Rápida de lo que es la Memoria DRAM y SRAM, Que es Cache y más. (2018). Recuperado de: <https://uruguayoc.com/2018/10/17/una-explicacion-rapida-de-lo-que-es-la-memoria-dram-sram-y-que-es-cache-o-buffer/> [Revisado 2019].
- [6] Mora, Eduardo. 2003. Las Memorias de un Computador - Monografias.com. Retrieved from <https://www.monografias.com/trabajos16/memorias/memorias.shtml> [Revisado 2019].
- [7] Akesson, B. et al. Memory Controllers for High-Performance and Real-Time MPSoCs.
- [8] L. Cadix et al. Integration and frequency dependent electrical modeling of Through Silicon Vias (TSV) for high density 3DICs. In Proc. IITC, 2010.
- [9] P. Franzon. Creating 3D-Specific Systems-Architecture, Design,CAD. In Proc. 2010.
- [10] RAMON: Region-Aware Memory Controller. (2018). IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Very Large Scale Integration (VLSI) Systems, IEEE Transactions on, IEEE Trans. VLSI Syst, (4), 697. <https://bibliotecavirtual.uis.edu.co:2236/10.1109/TVLSI.2018.2789520>
- [11] Efficient implementation of memory controllers and memories and virtual platform. (2014). 2014 International Conference on Communication and Signal Processing, Communications and Signal Processing (ICCSP), 2014 International Conference On, 1645. <https://bibliotecavirtual.uis.edu.co:2236/10.1109/ICCSP.2014.6950127>