Componentes del computador.

Memoria.

La memoria en las computadoras es el conjunto de recursos que almacenan instrucciones y datos para que la CPU y otros dispositivos los puedan procesar. Desde el punto de vista del diseño de sistemas, la memoria no trabaja por si sola ya que esta incluye niveles con distintas latencias, anchos de banda y persistencia (registros, caché L1/L2/L3, memoria principal DRAM y almacenamiento secundario). Entender los niveles y sus características es importante para optimizar software y hardware, porque muchas optimizaciones de rendimiento resultan de ajustar la localidad de datos a la jerarquía de memoria. [1]

Los tipos de memoria usados en computadores se diferencian por capacidad, velocidad y persistencia. En la cima de todo están los registros que son los más rápidos y de menor capacidad, la caché que funciona en microsegundos o nanosegundos de latencia; en el medio está la memoria interna (RAM/DRAM) con mayor capacidad y mayor latencia; al final está la memoria externa (SSD, HDD, almacenamiento en red) que ofrece mucha capacidad y persistencia pero con latencias y anchos de banda peores. Además, aparecen tecnologías emergentes (NVM, PCM, ReRAM) y técnicas near-memory / processing-in-memory (PIM) que buscan reducir el coste de mover datos entre niveles. [2]

La jerarquía de memoria organiza esos niveles por latencia y capacidad para mejorar la velocidad de acceso, los niveles superiores (caché) ejecutan datos rápidamente y filtran solicitudes hacia niveles más lentos solo cuando es necesario. Modelos analíticos y de colas han demostrado cómo el número de niveles y sus parámetros (latencias, tasas de aciertos/fallos, ancho de banda compartido) afectan la respuesta del sistema y las variaciones en el servicio, esto es especialmente crítico en arquitecturas multicore donde la memoria es compartida y los efectos de contención aumentan la variabilidad del tiempo de respuesta. [3]

Memoria Caché, Interna y Externa.

La memoria caché tiene funciones y mecanismos específicos: mantener copias de bloques de memoria más utilizados, reducir latencia efectiva y disminuir la presión sobre la RAM/almacenamiento. Las políticas de reemplazo (LRU, LFU, políticas híbridas) y las estrategias de coherencia en caches compartidas influyen fuertemente en la tasa de aciertos y en su rendimiento. Trabajos recientes proponen caches in-memory con privación dinámica que cambian los tiempos de ejecución para minimizar la tasa de fallos, mostrando mejoras reales en entornos de almacenamiento y sistemas distribuidos. [4]

Al comparar memoria interna (RAM) y memoria externa (almacenamiento secundario) destacan diferencias funcionales: la RAM es volátil, optimizada para acceso aleatorio rápido y altas tasas de transferencia en acceso a datos activos; el almacenamiento externo es no volátil y está optimizado para capacidad y coste por cada byte. Sin embargo, nuevas arquitecturas de almacenamiento computacional y de memoria cercana estrechan la distancia entre ambos, moviendo parte del procesamiento hacia el almacenamiento o acercándose a la memoria para reducir movimiento de datos y mejorar latencias o consumo energético en aplicaciones intensivas en datos. [5]

La importancia de la caché en el rendimiento se cuantifica con métricas como las tasas de acierto y tiempo de acceso efectivo (EAT). Modelos y simuladores muestran que pequeñas mejoras en tasa de aciertos o en la política de reemplazo pueden dar reducciones significativas del tiempo promedio de acceso y carga sobre la RAM y el almacenamiento, especialmente en cargas concurrentes (multicore, DNNs). Por eso, la co-diseño de software (localidad de datos, optimizaciones de acceso) y hardware (tamaños de línea, niveles de caché, ancho de banda) es una práctica recurrente para mejorar el rendimiento y la latencia. [6]

Las tendencias emergentes modifican el panorama, ya sea en el procesamiento en memoria (PIM), computación en memoria y dispositivos con capacidades especiales integradas proponen reducir la transferencia de datos entre CPU y memoria, mejorar la eficiencia energética y aumentar el rendimiento para tareas específicas (p. ej. inferencia de redes neuronales u operaciones de reducción en grandes volúmenes de datos). Investigaciones en materiales y dispositivos (ferroeléctricos, memoria no volátil) y revisiones de arquitecturas PIM señalan tanto prometedoras ganancias de rendimiento como retos (programabilidad, coherencia, herramientas de verificación). [7]

Para investigación y medición práctica, los trabajos de conferencias sobre sistemas de memoria (actas MEMSYS) y artículos recientes sobre caches y dispositivos de almacenamiento computacional recomiendan usar simuladores detallados (modelado de canales, ranks, banks), benchmarks orientados a memoria y experimentos controlados con métricas de latencia, rendimiento y perfiles de localidad. Además, los estudios de revisión sistemática sobre computational storage recomiendan definir claramente el caso de uso y medir tanto el impacto en latencia como en consumo energético y coste por operación. [8]

Bibliografías

[1] B. Jacob, *The Memory System*. Cham: Springer International Publishing, 2009. doi: 10.1007/978-3-031-01724-7.

[2] K. Asifuzzaman, N. R. Miniskar, A. R. Young, F. Liu, and J. S. Vetter, “A survey on processing-in-memory techniques: Advances and challenges,” *Memories - Materials, Devices, Circuits and Systems*, vol. 4, p. 100022, Jul. 2023, doi: 10.1016/j.memori.2022.100022.

[3] A. M. Mohamed, N. Mubark, and S. Zagloul, “Performance aware shared memory hierarchy model for multicore processors,” *Sci Rep*, vol. 13, no. 1, p. 7313, May 2023, doi: 10.1038/s41598-023-34297-3.

[4] K. Shakiba and M. Stumm, “PaperCache: In-Memory Caching with Dynamic Eviction Policies,” in *Proceedings of the 17th ACM Workshop on Hot Topics in Storage and File Systems*, New York, NY, USA: ACM, Jul. 2025, pp. 107–113. doi: 10.1145/3736548.3737836.

[5] S. A. Shirke, N. Jayakumar, and S. Patil, “Design and performance analysis of modern computational storage devices: A systematic review,” *Expert Syst Appl*, vol. 250, p. 123570, Sep. 2024, doi: 10.1016/j.eswa.2024.123570.

[6] “Near-ideal in-memory sensing and computing devices using ferroelectrics,” *Nat Mater*, vol. 22, no. 12, pp. 1447–1448, Dec. 2023, doi: 10.1038/s41563-023-01692-0.

[7] D. Fakhry, M. Abdelsalam, M. W. El-Kharashi, and M. Safar, “A review on computational storage devices and near memory computing for high performance applications,” *Memories - Materials, Devices, Circuits and Systems*, vol. 4, p. 100051, Jul. 2023, doi: 10.1016/j.memori.2023.100051.

[8] “MEMSYS: Memory Systems,” in *The International Symposium on Memory Systems*, New York, NY, USA: ACM, 2021.