



Organización Computacional

Jerarquía de la memoria

Diego Ernesto Vázquez Briceño

M. en C. Jorge J. Pedrozo Romero

Licenciatura en Desarrollo en Software



Restricciones en el diseño de la memoria de una computadora

Costo: El costo es una de las restricciones más significativas en el diseño de la memoria de una computadora. El costo total de la memoria afecta directamente el costo final del sistema informático. Diferentes tipos de memoria tienen costos variables, por lo que el diseñador debe equilibrar el costo con los requisitos de rendimiento y capacidad del sistema.

Rendimiento: El rendimiento es otra restricción crítica en el diseño de la memoria. El rendimiento de la memoria afecta directamente la velocidad y la capacidad de respuesta del sistema. Para garantizar un rendimiento óptimo, el diseñador debe considerar varios aspectos, como la velocidad de acceso, la latencia, el ancho de banda y la capacidad de transferencia de datos.

Capacidad: La capacidad de la memoria es otra restricción importante. La capacidad total de la memoria determina la cantidad de datos que el sistema puede almacenar y procesar simultáneamente. Los requisitos de capacidad pueden variar según el tipo de aplicación y el entorno de uso.



Jerarquía de la memoria

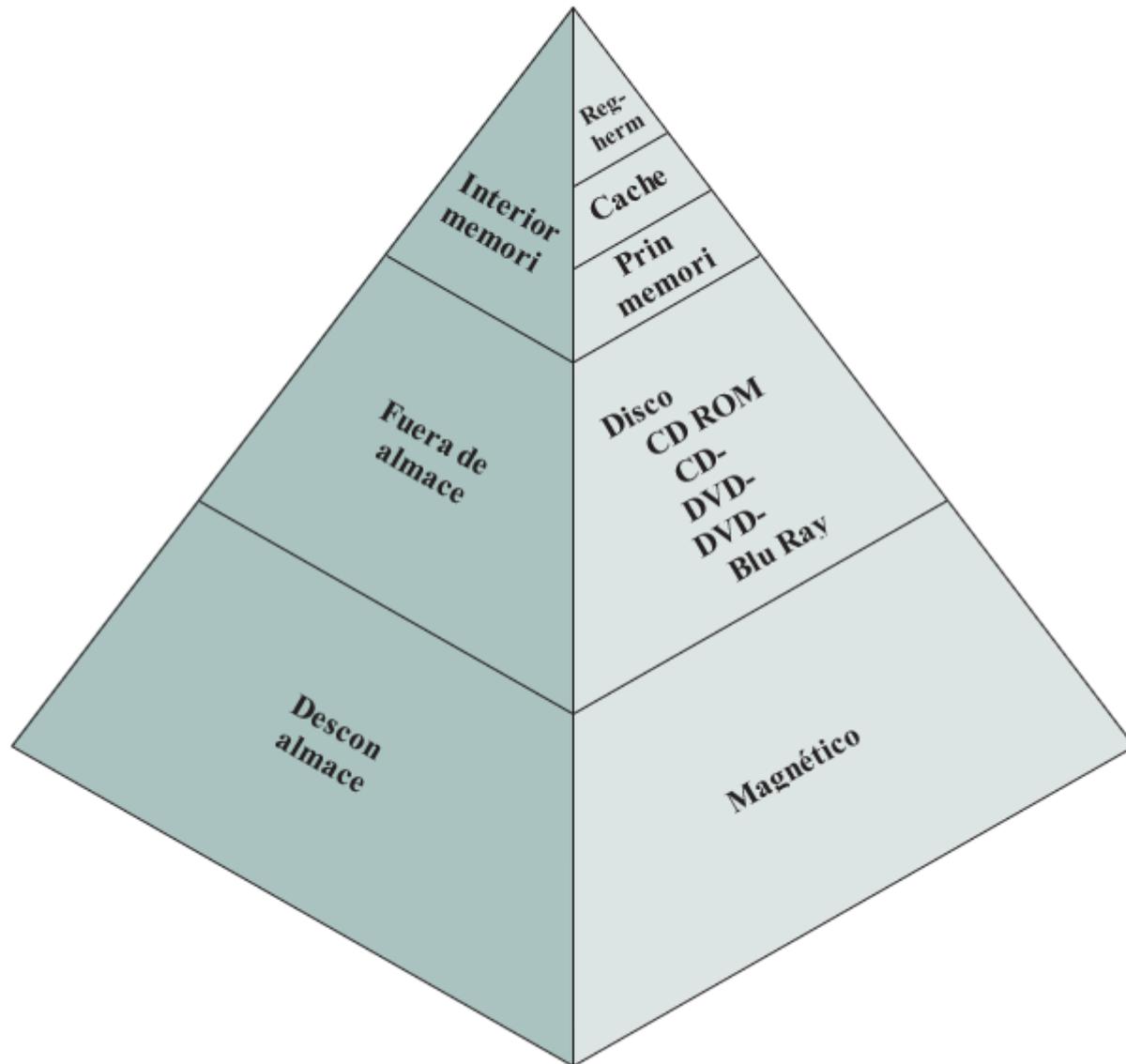


Figura 4.1 La jerarquía de la memoria



Localidad de referencia

La localidad de referencia es un principio importante en la arquitectura de computadoras que describe el patrón de acceso a la memoria por parte de un programa o proceso.

Se refiere a la tendencia de los programas a acceder a las mismas ubicaciones de memoria repetidamente durante un período de tiempo limitado.

Existen dos tipos: temporal y espacial.

EJEMPLO 4.1 Supongamos que el procesador tiene acceso a dos niveles de memoria. El nivel 1 contiene 1000 palabras y tiene un tiempo de acceso de 0,01 metros; El nivel 2 contiene 100.000 palabras y tiene un tiempo de acceso de 0,1. metros. Supongamos que si una palabra a la que se accede está en el nivel 1, entonces el procesador accede a ella directamente. Si está en el nivel 2, la palabra se transfiere primero al nivel 1 y luego el procesador accede a ella. Para simplificar, ignoramos el tiempo necesario para que el procesador determine si la palabra está en el nivel 1 o en el nivel 2. La figura 4.2 muestra la forma general de la curva que cubre esta situación. La figura muestra el tiempo medio de acceso a una memoria de dos niveles en función de la tasa de aciertos, h , donde h se define como la fracción de todos los accesos a la memoria que se encuentran en la memoria más rápida (por ejemplo, el caché), t_1 es el tiempo de acceso al nivel 1, y t_2 es el tiempo de acceso al nivel 2.¹ Como se puede observar, para porcentajes elevados de acceso al nivel 1, el tiempo medio total de acceso es mucho más cercano al del nivel 1 que al del nivel 2.

En nuestro ejemplo, supongamos que el 95% de los accesos a la memoria se encuentran en el nivel 1. Entonces el tiempo promedio para acceder a una palabra se puede expresar como

$$(0,95)(0,01 \text{ metros}) + (0,05)(0,01 \text{ metros} + 0.1 \text{ metros}) = 0.0095 + 0.0055 = 0,015 \text{ metros}$$

En consecuencia, es posible organizar los datos a lo largo de la jerarquía de manera que el porcentaje de accesos a cada nivel sucesivamente inferior sea sustancialmente menor que el del nivel superior. Consideremos el ejemplo de dos niveles ya presentado. Deja el nivel 2

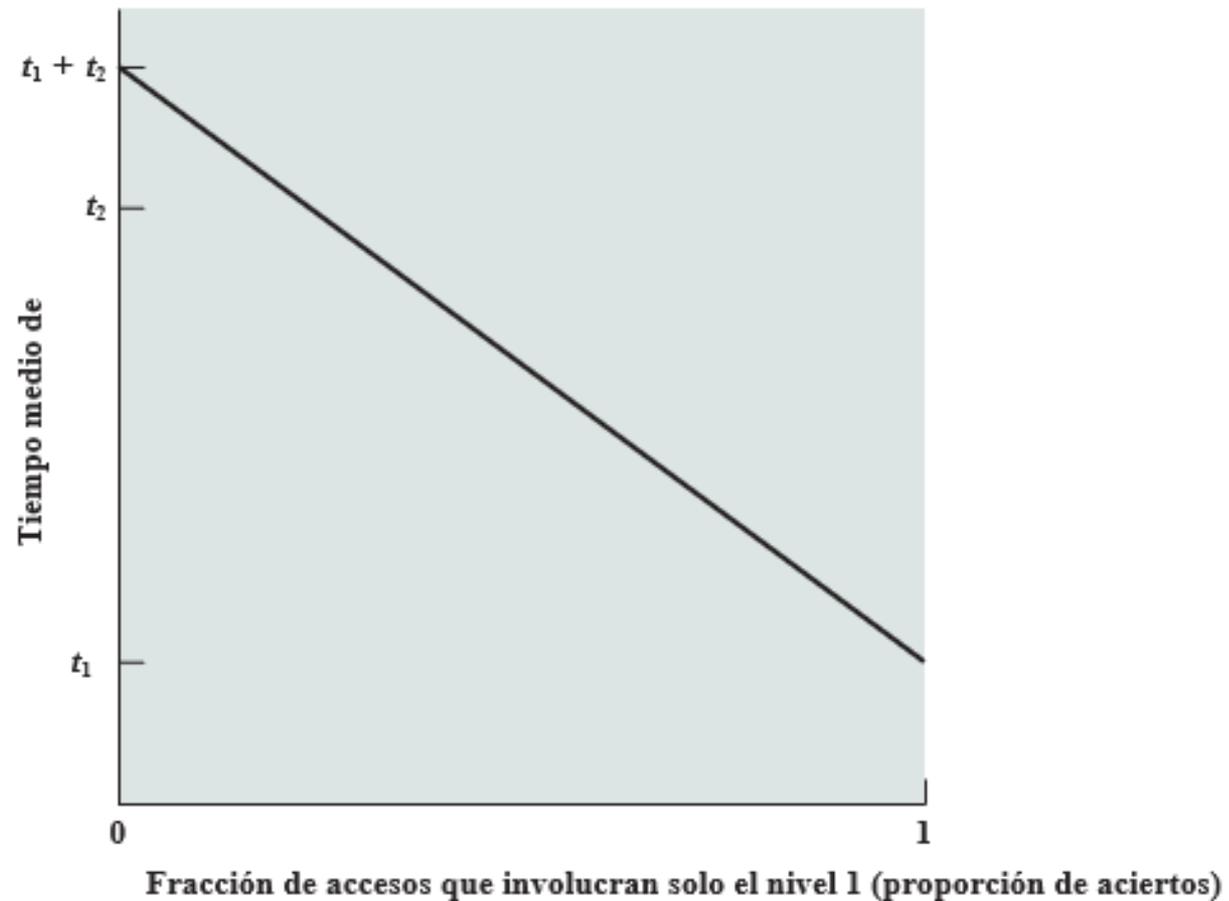
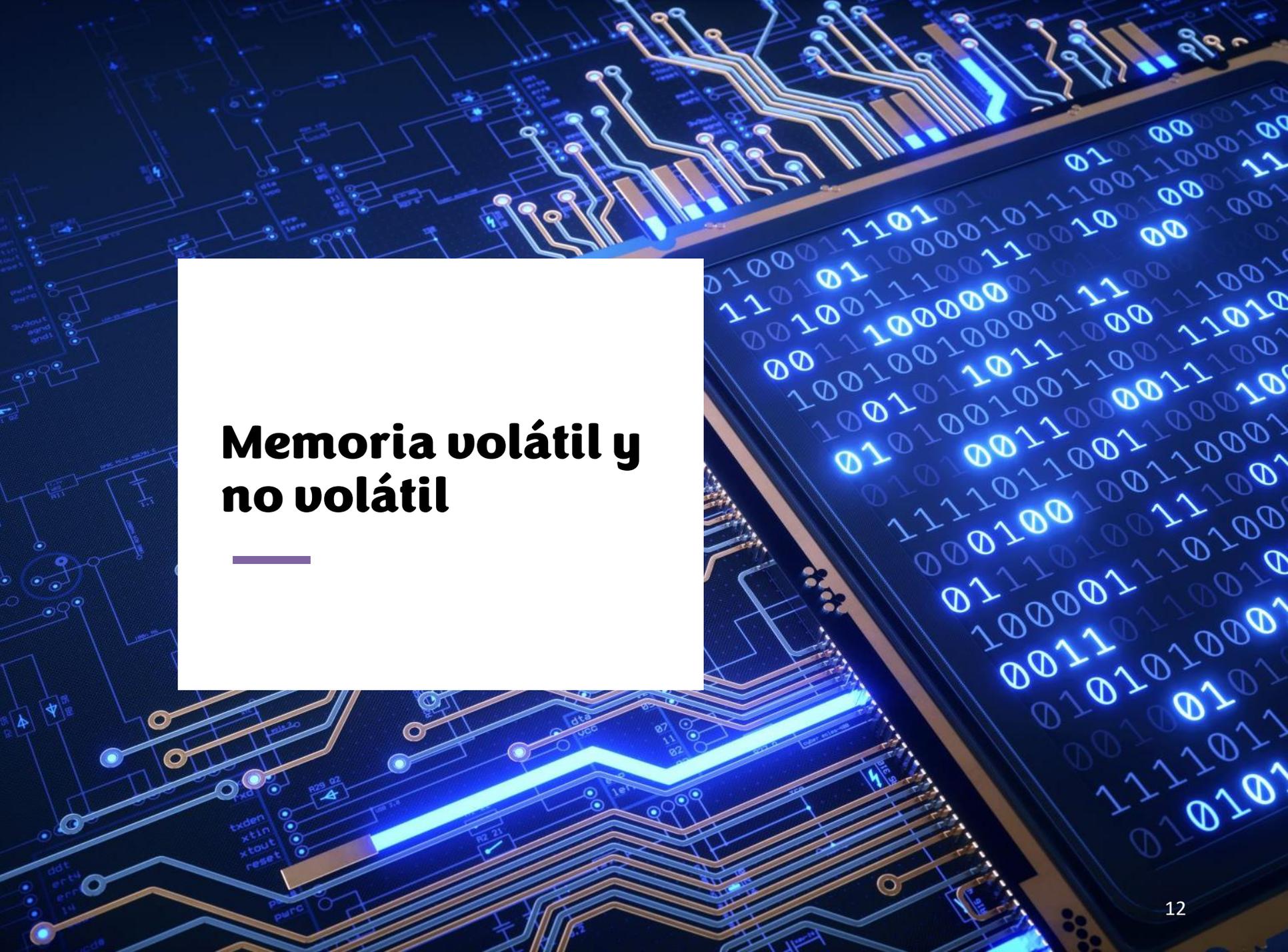


Figura 4.2 Rendimiento de los Accesos que involucran únicamente el Nivel 1 (proporción de aciertos)



Memoria volátil y no volátil

Memoria Volátil:

La memoria volátil es aquella que pierde su contenido cuando se interrumpe el suministro de energía eléctrica. Es decir, los datos almacenados en la memoria volátil se borran cuando se apaga el dispositivo o se pierde la energía eléctrica.

Memoria No Volátil:

La memoria no volátil es aquella que retiene su contenido incluso cuando se interrumpe el suministro de energía eléctrica. Es decir, los datos almacenados en la memoria no volátil permanecen intactos incluso después de apagar el dispositivo.



El cache de disco

El "caché de disco" es una técnica de almacenamiento que implica el uso de una porción de memoria volátil de alta velocidad para almacenar temporalmente datos que se han accedido recientemente desde un disco duro. Esta memoria caché actúa como un intermediario entre la unidad de disco y el sistema operativo o la aplicación que accede a los datos. Su propósito principal es mejorar el rendimiento del sistema al reducir los tiempos de acceso a los datos.