



Universidad Nacional
de Entre Ríos

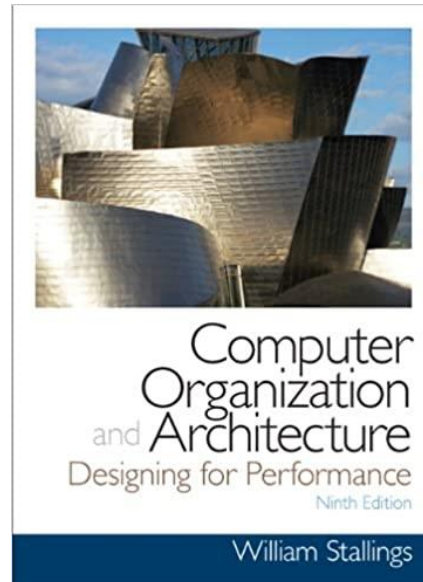
Tecnicatura universitaria en desarrollo web

Jerarquía de memoria

Semana 6 – Arquitectura de computadoras

Esta presentación esta basada en el libro de:

- ❑ William Stallings, Computer Organization and Architecture, 9th Edition, 2017.
 - Arquitectura acumulador: THE VON NEUMANN MACHINE: 2.1 / A BRIEF HISTORY OF COMPUTERS, pag 17.
 - Operandos de una instrucciones: CHAPTER 12 / INSTRUCTION SETS: CHARACTERISTICS AND FUNCTIONS, Number of Addresses, pag 410.



Archivos presentación y ejemplos se alojan en:

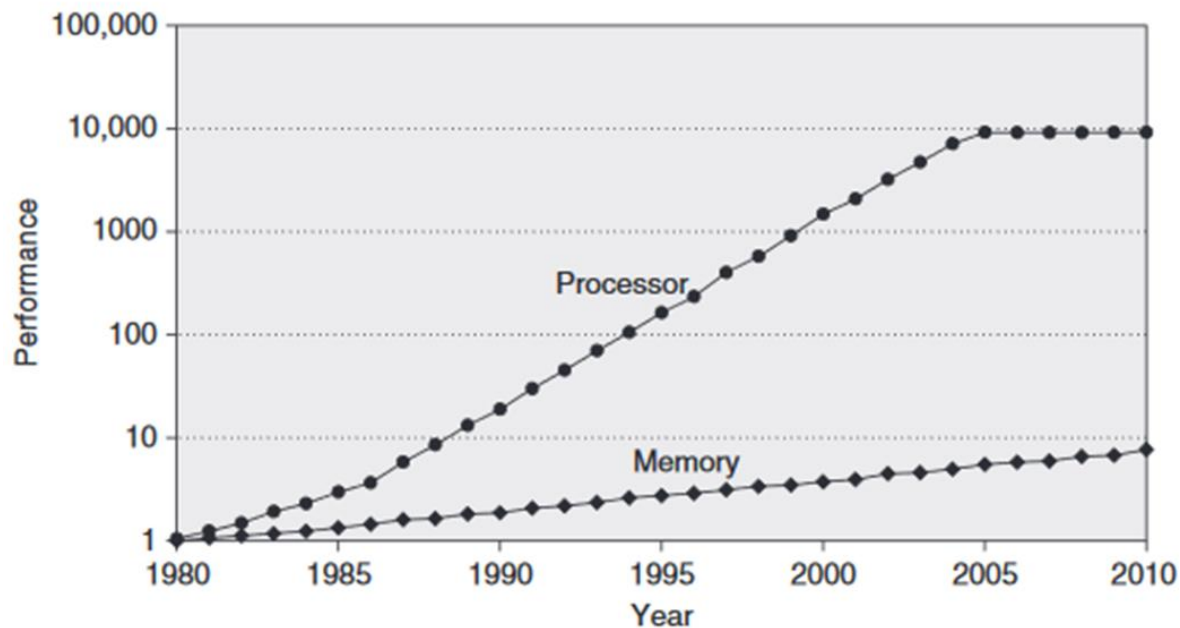


<https://github.com/ruiz-jose/tudw-arq.git>

Jerarquía de memoria

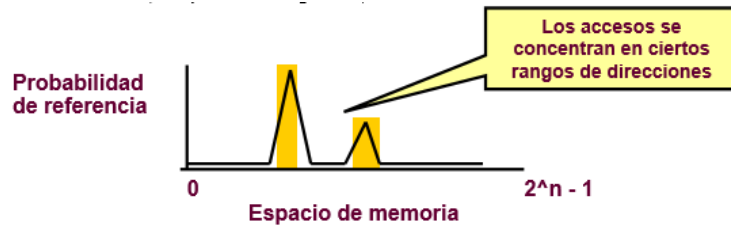
- Diferencia rendimiento CPU-RAM
- Principio de localidad
- Pirámide de jerarquía de memoria
- DRAM

Diferencia en rendimiento CPU-RAM (Gap:brecha)



Principio de localidad:

- ❖ El rango de direcciones de memoria al que accede un programa es relativamente pequeño para periodos de tiempo cortos.
- ❖ La información de las direcciones más frecuentadas pueden estar en una memoria pequeña y rápida (memoria caché), y el resto de datos en una grande y barata.
- ❖ Lo más referenciado tenerlo más cerca de la CPU.



¿Cómo se cumple el principio de localidad?

○ Localidad temporal

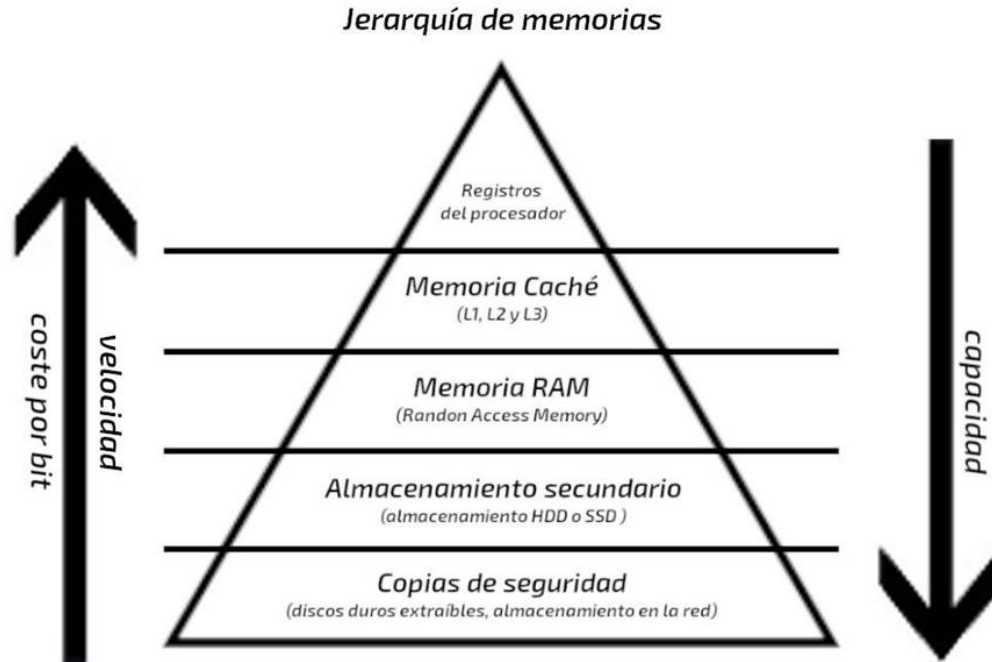
Si se accede a una posición de memoria, frecuentemente se vuelve a acceder a la misma posición de memoria

- Código: bucles, recursividad, funciones frecuentemente llamadas.
- Datos: datos globales

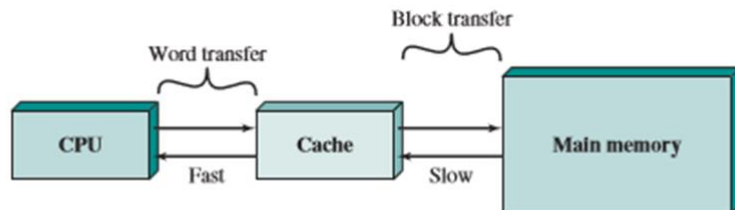
○ Localidad espacial

Si se accede a una posición de memoria, frecuentemente también se accede a posiciones cercanas

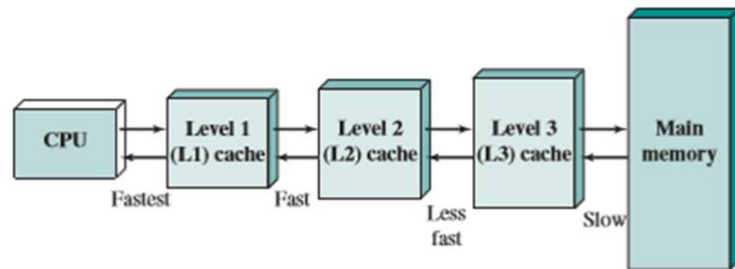
- Código: código secuencial
- Datos: vectores, matrices



Cache: pequeña componente de almacenamiento interno que transparentemente almacena datos que recientemente han sido accedidos.

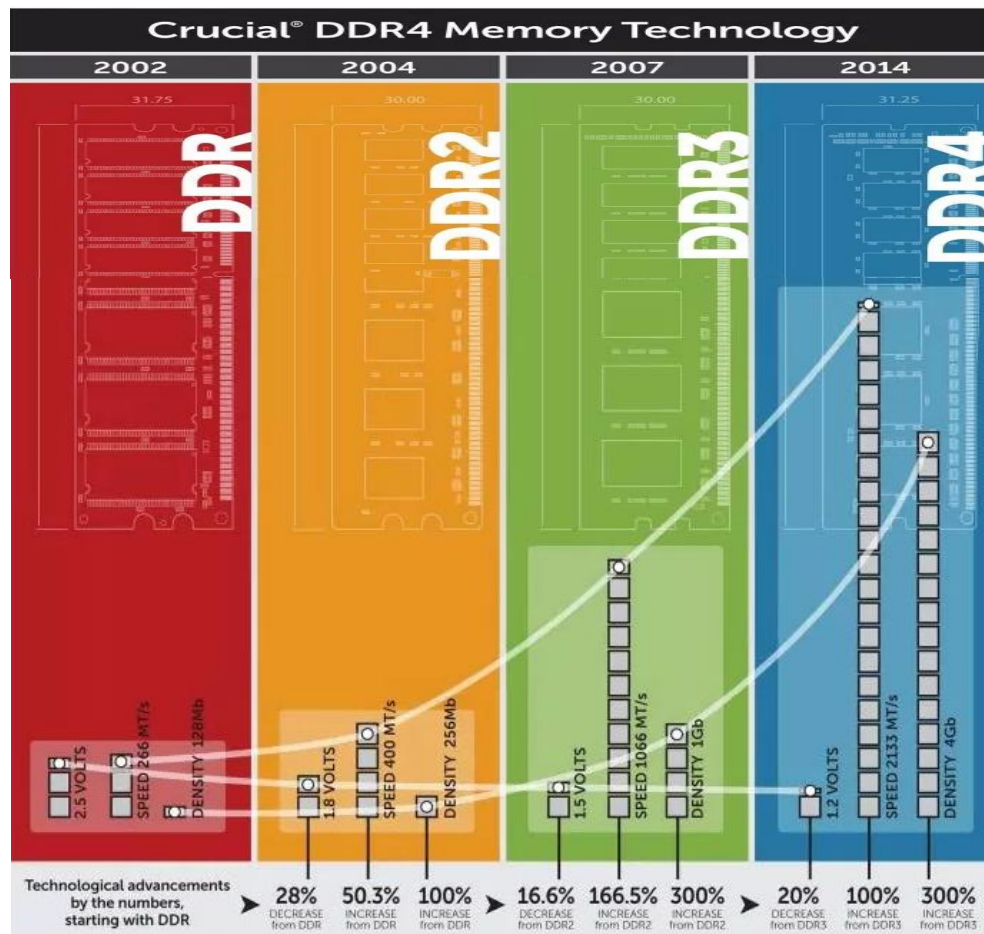


(a) Single cache



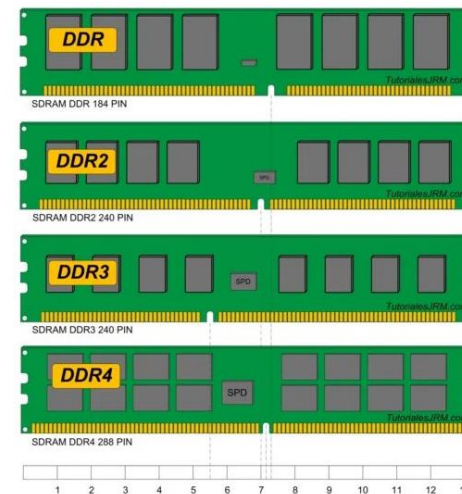
(b) Three-level cache organization

Figure 4.3 Cache and Main Memory



Evolución DDR:

- ❖ Reduce el consumo en volts.
- ❖ Aumenta el speed en millones de transferencias por segundos (MT/s).
- ❖ Aumenta la cantidad de bytes sobre la misma superficie (densidad).

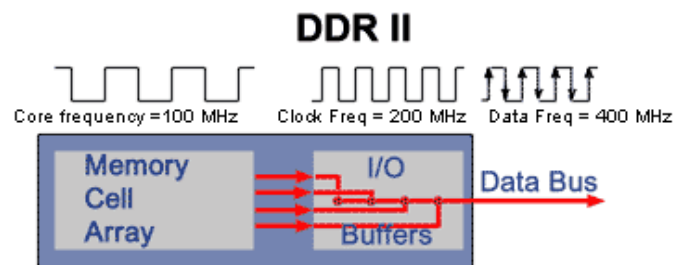
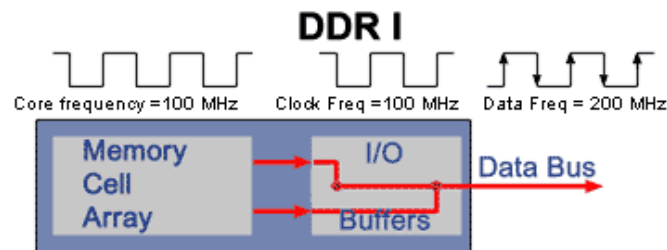
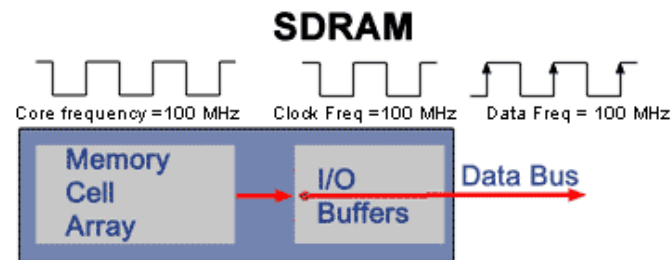


Fuente: <https://tuorlalesjrm.com/>

Evolución DDR:

Por lo general las memoria tienen un núcleo de celdas que funcionan a 100 MHz y utilizan un bus de 64 bits.

- ❖ En una memoria DRAM single, la celda de memoria proporciona 1 bit cada 10 ns al búfer de E/S que, a su vez, lo envía a través del bus a una frecuencia de 100 millones de bits por segundo (100 MHz). Por lo que obtiene un ancho de banda de $(100 \times 64) / 8 = 800$ MB/s.
- ❖ En la DDR I, ahora devuelve 2 bits a cada ciclo, por lo que esto se conoce como captación previa de 2 bits. Sin embargo, para que el búfer de E/S continuara funcionando a 100 MHz, se empleó la idea de usar ambos extremos de la señal de reloj en lugar de los estados binarios. En resumen, mientras permanece en 100 MHz, DDR permite transferir 200 millones de bits por segundo, es decir, un ancho de banda de $((100 \times 2) \times 64) / 8 = 1600$ MB/s.
- ❖ **Limitación:** la tecnología limitó la frecuencia de las celdas de memoria a alrededor de 166 MHz. Al alcanzar los 200 MHz, las matrices de memoria comenzaron a sufrir. ¿Solución? ¡Vuelva a 100 MHz, por supuesto! Pero esta vez, para mantener el flujo, fue necesario duplicar la frecuencia de los búferes de E/S.
- ❖ En la DDR II, las celdas proporcionan 4 bits por ciclo al búfer de E/S que ahora funciona a 200 MHz. Por lo tanto, la captación previa cambia de 2 bits a 4 bits. Al final, en el bus de datos, obtienes lo mismo que con DDR (uso de los flancos ascendentes y descendentes de la señal), pero al doble de frecuencia. Entonces, la tasa de datos para una frecuencia base de 100 MHz es $((100 \times 4) \times 64) / 8 = 3200$ MB/s.



Preguntas?