

[illegible]

a) Supongamos que tenemos dos niveles de memoria, M1 y M2. La memoria M1 es la más cercana al procesador y contiene 1000 bytes con un tiempo de acceso promedio de 0.1 microseg. La memoria M2 contiene 100000 bytes, tiene un acceso promedio de 1 microseg (Podemos asumir que la M2 es la memoria Principal)
Se solicita:
Calcular el tiempo de acceso promedio al sistema de memoria (teniendo en cuenta estos dos niveles memorias) considerando un HIT cache que varía en estos dos valores 0.15; y 0.85. ¿Cuál de estos dos valores de HIT cache Ud. considera que es el más deseable desde el punto de vista del rendimiento?

Grafique la función de HIT ratio para estos valores. Explique a se puede atribuir su

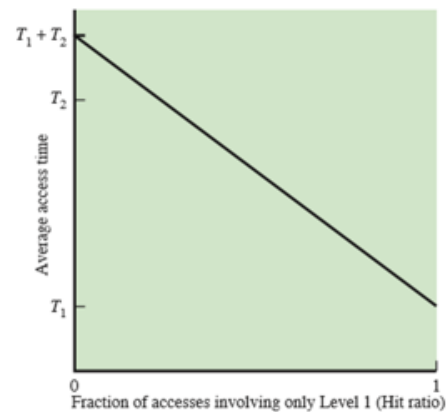
T1 es el tiempo de acceso a la memoria 1 y T2 el tiempo de acceso a la memoria 2

Para altos porcentajes de acceso al nivel 1, el promedio de tiempo de acceso es mas cercano al nivel 1 que al nivel 2.

Supongamos que el 95% de los accesos a la memoria se encuentran en el cache (H=0.95). Entonces el tiempo promedio de acceso a una palabra se puede expresar como:

$(0.95) * (0.01 \text{ microseg}) + (0.05) * (0.01+0.1) = 0.0095+0.0055=0.015$ microseg.

El resultado es cercano al tiempo de acceso a la memoria mas rápida. (0.015 es mas cerca de 0.01 que de 0.1)



T1	0,1 ms		
T2	1 ms		
	(0,15) * 0,1 mas	0,85 (0,1+1)	1,1
Tiempo promedio accesos H	0,015	0,935	0,95 ms
	0,85	0,15	
Tiempo promedio accesos H	0,085	0,165	0,25 ms
	0,95	0,05	
	0,095	0,055	0,15
	0,98	0,02	
	0,098	0,022	0,12

T1+T2 1,1 ms

x

1 ms T2

x

0,1 ms T1

x

x

15

85

100

Así, si el módulo de memoria RAM se encuentra en una placa madre con un bus que funciona a 400 Mhz (400 millones de ciclos por segundo) y en cada ciclo del reloj se realiza una transferencia de datos; y el bus es de 64 bits, o sea que se pueden transferir 64 bits simultáneamente por las pistas del bus en cada ciclo; si multiplicamos $400.000.000 \times 64 = 25.600.000.000$ bits por segundo, y como cada byte está compuesto por 8 bits, si dividimos $25.600.000.000 / 8 = 3.200.000.000$; esto significa que se pueden transferir teóricamente 3200 Megabytes por segundo. Obviamente estos números pueden variar por distintos motivos.

		400 Mhz	400.000.000	ciclos por segundo					
	bus de 64 líneas			400000000 * 64 líneas del bus de datos (cada línea transporta un bit en cada ciclo)					
	Tasa de Transferencia	25.600.000.000	bits por segundos						
		8 its = 1 byte							
	Tasa de Transferencia	25.600.000.000 dividido 8	es igual a	3.200.000.000 bytes por seg				3.200 Mbytes por seg	

El principal motivo técnico que hace que estos números ideales de velocidad de transferencia de datos no se den es la *latencia*. Cuando se transporta la información de la memoria RAM a su microprocesador, también hay que tener en cuenta el tiempo en que tarda en leer dicha información de la memoria. Por ejemplo cuando se quiere transmitir una serie de datos primero hay que ir a buscarlos y si la lectura de dicha información puede durar 4 ciclos del reloj del bus, ese tiempo deberá ser sumado a los valores de tiempo de transferencia; en otras palabras si el reloj realiza 400.000.000 ciclos por segundo y en cada ciclo transporta 64 bits, esto significa que requiere 0,0000000025 segundos (2,5 nanosegundos; o sea 2,5 mil millonésimas de segundo) para transportar cada grupo de 64 bits, ya que cada ciclo del reloj del bus tiene una duración de 2,5 nanosegundos. Pero supongamos que para llegar a leer los bits que luego serán transferidos, como se mencionó antes, se necesita un tiempo de 4 ciclos del reloj del bus; por lo que tardará $2,5 \times 4 = 10$ nanosegundos adicionales en leer los datos; más 2,5 para transferirlos, entonces el tiempo total será de 12,5 nanosegundos; o sea 10 nanosegundos más que si la tarea consistiera únicamente en transferir los datos de la memoria a su microprocesador.

400 Mhz	400.000.000	ciclos por segundo			0,0000000003				
	1 ciclo	dura	1/400000000	0,0000000025	segundos	2,5 nanosegundos			
	1 bit por linea	1 ciclo por bit	1 bit cada 2,5 nanosegundos						
	64 lineas		64 bits cada 2,5 nanosegundos						
A)	Tasa Transferencia	Transmitir 64 bits requieren 2,5 nanosegundos							
B)	Latencia de acceso a memoria medido en ciclos, supongamos 4 ciclos que es igual a 4 * 2,5 nanosegundods que es igual a 10 nanosegundos								
	Entonces tiempo de acceso a memoria								
	A+ B	igual	2,5 nanoseg + 10 nanoseg		igual a	12,5 nanoseg			

Especifique que significa la siguiente nomenclatura para las memorias DDR: 4 GB DDR3 – 2133 PC3-17000

La nomenclatura que se utiliza para definir los módulos de memoria de tipo DDR sigue un patrón establecido que describe sus características: *DDRx-yyyy PCx-zzzz*; donde x define la generación DDR, yyyy la frecuencia aparente o efectiva (en Megatransferencias por segundo) y zzzz la máxima tasa de transferencia de datos (en Megabytes) por segundo que se puede lograr entre un módulo de memoria y su controlador. Esta tasa depende de dos factores, ancho de bus de datos (siempre de 64 bits) y la frecuencia efectiva de trabajo, mientras que para calcularla se utiliza la siguiente fórmula:

Frecuencia DDR efectiva x 64 bits / 8 bits por cada byte

Veamos algunos ejemplos:

1 GB DDR-400 PC-3200: Indica que se trata de un módulo de 1 GB (Gigabyte) de tipo DDR, frecuencia aparente o efectiva de trabajo de 400 Mhz o para ser más correctos 400 Megatransferencias por segundo y una tasa de transferencia de datos máxima de 3200 MB/s.

4 GB DDR3-2133 PC3-17000: Indica que se trata de un módulo de 4 GB de tipo DDR3, frecuencia aparente o efectiva de trabajo de 2133 Mhz o para ser más correctos 2133 Megatransferencias/s y una tasa de transferencia de datos máxima de 17000 MB/s.

¿Cuál sería la tasa de transferencia de una memoria DDR3-2133 PC3-17000 en dual channel con un bus de datos que opera a 1067 MHz?

El Dual Channel es una tecnología con la que cuentan algunos controladores de memoria la cual les permite ensanchar el bus de datos que los une con los módulos de memoria de 64 bits a 128 bits. Si los valores de frecuencia de reloj, latencias, entre otros se mantienen iguales; teóricamente el monto de datos transmitidos por segundo debería duplicarse, ya que como vimos depende de la fórmula:

Frecuencia del reloj del bus x tasa de datos transferidos por ciclo (2 en tecnologías DDR) x bits transferidos por ciclo / 8 bits por byte = Bytes transmitidos por segundo

Por ejemplo:

Si se tiene una memoria **DDR3-2133 PC3-17000** en un bus de datos que opera a 1067 Mhz se puede lograr una tasa de transferencia de datos por segundo de:

$$1067 \times 2 \times 64 / 8 = 17072 \text{ MB / s}$$