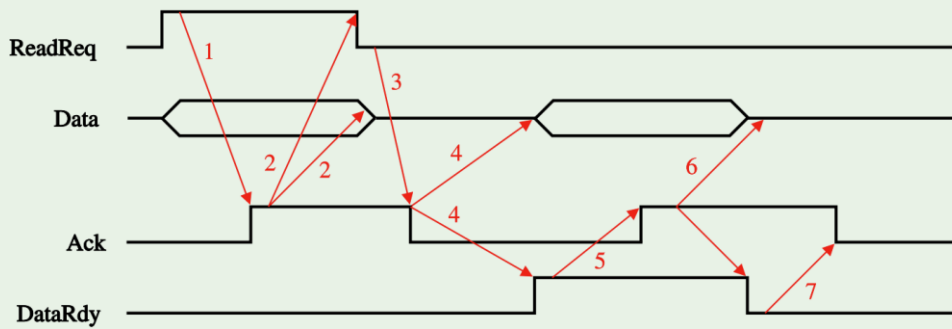


Ejemplo de transacción: lectura asíncrona



5.- Sistema con un sistema de memoria y bus que soporta acceso a bloques de entre 4 y 16 palabras de 32 bits. Bus síncrono de 64 bits a 200 MHz, donde una transferencia de 64 bits como envío de la dir. de memoria requieren 1 ciclo. Se necesitan 2 ciclos de reloj entre 2 operaciones de bus. Las 4 primeras palabras se leen en 200 ns; cada grupo adicional de 4 palabras se lee en 20 ns. Puede haber solapamiento con la lectura de las 4 palabras siguientes. Calcular latencia y ancho de banda para lectura de 256 palabras en transferencias de bloques de 4 y 16 palabras. Halla también el número de transacciones/segundo en cada caso.

Tciclo = $1/200 \text{ MHz} = 5 \text{ ns}$

-> Bloques de 4 palabras = $(256/4) = 64$ transacciones. Ciclos requeridos: 1 (envío de dirección) + 40 ($200\text{ns}/5\text{ns}=40$ ciclos para lectura) + 2 (enviar 2 y 2 palabras (tam.palabra=32bits)) + 2 (ciclos de hueco) = 45 ciclos.

TIEMPO TOTAL DE TRANSFERENCIA = 45 ciclos * 64 transacciones = 2880 ciclos * 5 ns = 14400 ns.

TRANSACCIONES/SEGUNDO = $64/14400 = 4,44 \text{ MTransacciones/s.}$

ANCHO DE BANDA = $256 \text{ (palabras)} * 4 \text{ (B)} / 14400 = 71,11 \text{ MB/s.}$

-> Bloques de 16 palabras = $(256/16) = 16$ transacciones. Ciclos requeridos: 1 (envío de dirección), 40 ($200\text{ns}/5\text{ns}=40$ ciclos para lectura), 2 (enviar 2 y 2 palabras pq. Tam palabra = 32 bits), 2 (ciclos de hueco para leer siguiente grupo). Como hay 4 grupos de 4 palabras = $1 + 40 + 4 * (2 + 2) = 57$ ciclos.

TIEMPO TOTAL DE TRANSFERENCIA = 57 ciclos * 16 transacciones = 912 ciclos * 5 ns = 4560 ns.

TRANSACCIONES/SEGUNDO = $16/4560 = 3,51 \text{ MTransacciones/s.}$

ANCHO DE BANDA = $256 \text{ (palabras)} * 4 \text{ (Bytes (tam.palabra))} / 4560 = 224,56 \text{ MB/s}$

6.- Mismo ejercicio con bloques de 16 bloques, pero suponiendo que las 4 primeras palabras requieren 150 ns y cada 4 palabras adicionales requieren 30 ns.

-> Bloques de 16 palabras = 16 transacciones. Ciclos requeridos: 1 (envío de dirección) + 30 ($150\text{ns}/5\text{ns}=30$ ciclos para lectura) + 2 (enviar 2 y 2 palabras pq. Tam.palabra = 32 bits) + 4 (ciclos de hueco para siguiente grupo de 4 palabras [$(2 + 4) * 5 \text{ ns} = 30 \text{ ns}$]) + 2 (ciclos de hueco entre esta transacción y la siguiente [$(2 + 2 + 2 \text{ (del envío de la siguiente transacción)}) * 5 \text{ ns} = 30 \text{ ns}$])).

El cálculo resultaría = $1 + 30 + 3 * (2 + 4) + 2 + 2 = 53$ ciclos

TIEMPO TOTAL DE TRANSFERENCIA = 53 ciclos * 16 transacciones = 848 ciclos * 5 ns = 4240 ns

TRANSACCIONES/SEGUNDO (...) ANCHO DE BANDA DEL BUS (...)

7.- Sea un sistema con sistema de memoria y bus para acceso a bloques de 32 palabras de 32 bits cada una, bus síncrono de 64 bits a 1 GHz, en el que una transferencia de 64 bits y envío de dir. a memoria = 1 ciclo. t. acceso a memoria para cada 2 palabras es de 1 ns. Puede haber solapamiento con la lectura de las siguientes palabras. Calcular latencia y ancho de banda para lectura de 256 palabras.

Tciclo = $1 / 1\text{GHz} = 1 \text{ ns}$

-> Bloques de 32 palabras = $(256/32) = 8$ transacciones. Ciclos requeridos: 1 (envío de dirección) + 1 ($1\text{ns}/1\text{ns} = 1$ ciclo para lectura de 2 palabras) + 1 ciclo (envío de grupo de 2 palabras, $32 * 2 = 64$ bits)

Cada grupo restante de 2 palabras en el bloque de 32 palabras necesita solo el último paso, porque se solapa la lectura de un grupo con el envío del grupo anterior): $1 + 16 * 1 + 1 = 18$ ciclos por transacción

TIEMPO TOTAL DE TRANSFERENCIA = 18 ciclos * 8 transacciones = 144 ciclos * 1 ns = 144 ns

ANCHO DE BANDA DEL BUS = $256 * 4(\text{Bytes (tam.palabra)}) / 144\text{ns} = 7,11 \text{ GB/s}$

8.-Comparamos un bus síncrono y otro asíncrono. El bus síncrono tiene un t. ciclo de 50 ns, y cada transmisión dura 1 ciclo. El bus asíncrono requiere 40 ns para cada paso del protocolo de *handshaking*. El ancho de datos en ambos buses es de 32 bits. Suponer que cada lectura consume 200 ns. Calcular ancho de banda. ¿Cuál es más rápido?

-> Bus síncrono = 50 ns (petición a memoria de la dirección) + 200 ns (servicio de la palabra) + 50 ns (envío de la palabra) = 300 ns. Como el tam.palabra es de 32 bits, ANCHO DE BANDA = 4 Bytes, 4 Bytes / 300ns = 13,3 MB / s (+ rápido y mejor ancho de banda).

-> Bus asíncrono = Cada paso del protocolo requiere 40 ns. Pasos = petición de una dirección de memoria, reconocimiento y lectura de la posición de memoria y envío de la palabra de memoria. Cada uno de ellos requiere un conjunto de señales y reconocimientos por parte de los elementos involucrados, de modo que el tiempo total es: $40 \text{ ns} + \max(3 * 40\text{ns}, 200 \text{ ns}) + 3 * 40 \text{ ns} = \underline{360 \text{ ns}}$. ANCHO DE BANDA 4B / 360ns = 11,1 MB / s.

9.- Computador con dir. de memoria y palabras de 64 bits. Jerarquía caché de 2 niveles (1º nivel: Tasa de fallo = 10 % / t. acierto = 5 ns; 2º nivel: Tasa de fallo = 40% / t. acierto = 8ns. T. medio de acceso a todo el sistema = 9 ns. Ambos niveles con tam. de línea = 8 palabras). En la caché de 2º nivel se conecta mediante un bus síncrono de 64 bits a la M.P., en el que tanto una transferencia de 64 bits como el envío de la dirección de memoria requieren 1 ciclo de reloj, y el sis. de memoria y bus soportan accesos a bloques de 8 palabras de 64 bits. Se necesitan 2 ciclos de reloj entre 2 operaciones de bus. La memoria tarda lo equivalente a 20 ciclos en acceder a las primeras 4 palabras del bloque; y luego cada grupo adicional de 4 palabras se lee en 13 ciclos. Las transferencias pueden solaparse con lectura. Calcular:

a) Frecuencia del bus síncrono que comunica la caché de 2º nivel con la M.P., ancho de banda y latencia para la lectura, desde la caché de 2º nivel, de 128 palabras de memoria, en el caso de transferencias de bloques de 8 palabras.

t. medio acceso = 9 ns = $5 + 0,1 * (8 + 0,4 * PF) \rightarrow PF = 80 \text{ ns}$. Tenemos 1 ciclo (enviar dir.) + 20 ciclos (acceder pal 0-3) + $\max(4 \text{ ciclos de envío pal 0-3}, 13 \text{ ciclos acceso pal 4-7}) + 4 \text{ ciclos envío pal 4-7} + 2 \text{ ciclos de espera} = 40 \text{ ciclos}$. 40 ciclos tardan 80 ns (PF) entonces T ciclo = $80 / 40 = 2 \text{ ns}$ y por tanto la frecuencia es su inversa, $1/0,002 = 500\text{MHz}$. Si 1 transacción son 40 ciclos y hay 16 transacciones (128 palabras / 8 líneas) la latencia es de $16 * 40 \text{ ciclos} = 640 \text{ ciclos}$, que son 1280 ns.

ANCHO DE BANDA = $128 * 8 (\text{Bytes/palabra}) / 1280 * 10^{-9} = 800 \text{ MB/s}$.

b) Supongamos que este bus soporta transacciones de bloques de 4 palabras y que la penalización por fallo de la caché de segundo nivel es de 54 ns cuando trabaja con líneas de 4 palabras. ¿Tendría sentido reemplazar este bus síncrono de memoria por un bus asíncrono donde cada paso del protocolo *handshaking* requiere de 30 ns en transferencias de bloques de 4 palabras?

Más allá de saber que al necesitar 7 pasos el protocolo *handshaking*, una transacción por el bus asíncrono tardará un mínimo de $7 * 30 = 210 \text{ ns}$, independientemente de lo que tarde una transferencia de datos por el bus o el acceso a una dirección de memoria. Por lo tanto, el bus síncrono, que tarda 54 ns será más rápido en cualquier caso.

10.- Tenemos un sistema con M.V. con páginas de 2 KB que conecta la M.P con el almacenamiento secundario mediante un bus síncrono de 128 bits a 100 MHz. Las transferencias de las páginas están gestionadas por una DMA a través del bus en bloques del tamaño de la anchura del bus, necesitando 280 ns para la lectura del 1º bloque y 100 ns para el resto en una operación de lectura de una página. Tanto el envío de 128 bits a través del bus como el de la dirección al controlador del almacenamiento secundario requieren 1 ciclo. Las transferencias de datos leídos más recientemente pueden solaparse con la lectura de los siguientes. Sabiendo que la DMA realiza la lectura de una página en una única transacción, calcula:

a) Latencia para la lectura de una página.

El bus de 128 bits indica que se leen 16 bytes de cada vez. Así, el nº de lecturas para transferir una página sería: $NL = 2 \text{ KB} / 2^4 \text{ B} = 2^{11} / 2^4 = 2^7 = 128$. La frecuencia del bus es 100 MHz, con lo cual el período es:

$T = 1 / 100 * 10^6 = 10^{-8} \text{ s} = 10 \text{ ns}$. Ciclos necesarios para la transferencia de una página requiere: 1 ciclo (envío de dirección), 28 ciclos ($280\text{ns}/10\text{ns} = 28$ ciclos para leer primeros 16 bytes.), 1270 ciclos ($127 * 100/10 = 1270$ ciclos para leer cada uno de los 127 grupos de 16 bytes restantes), 1 ciclo (transferir los últimos 16 bytes).

Cálculo: $1 + 28 + 1270 + 1 = 1300$ ciclos, siendo la LATENCIA: $1300 * 10 = 13000 \text{ ns}$.

b) Ancho de banda del bus para las operaciones de lectura de página desde el almacenamiento secundario.

$\text{ANCHO DE BANDA} = 2\text{KB} / 13 * 10^{-6} = 0,154 * 10^9 = 154\text{MB/s}$

11.- Supongamos que tenemos un sistema con las siguientes características:

- Un tándem bus-memoria que soporta operaciones de un nº arbitrario de palabras de 32 bits. El bus es de 32 bits y tiene una frecuencia de 500 MHz. Las direcciones de memoria son también de 32 bits.
- El t. acceso a memoria para las 4 primeras palabras de un bloque es de 50 ns. Para bloques mayores, cada grupo adicional de 4 palabras se obtiene en 10 ns.
- Las transferencias por el bus y los accesos a memoria pueden solaparse. Se supone que el bus está disponible antes de cada acceso.
- Las lecturas y escrituras en memoria consumen la misma cantidad de tiempo.

El sistema incorpora, además, una memoria caché con política de post-escritura, de la que se han obtenido las siguientes medidas de rendimiento:

- Cuando se produce un fallo caché, el 40 % de las veces es necesario realizar la operación de post-escritura en M.P., mientras que en el resto de las ocasiones solo la lectura es necesaria.
- Una tasa de fallos caché de 0,05 fallos por instruc.. para un tamaño de línea de 4 palabras.
- Una tasa de fallos caché de 0,03 fallos por instruc.. para un tamaño de línea de 16 palabras.

Suponiendo que el procesador se queda esperando durante la duración completa de un fallo caché (incluyendo el tiempo de post-escritura, si es necesario), ¿cuántos ciclos por instrucción se gastan en la gestión de los fallos caché para los dos tamaños de línea considerados (4 y 16 palabras)?

Comenzamos pensando en cómo calcular la penalización media que produce un fallo caché, que supuestamente son siempre en lectura de una línea. Además de esta lectura, en la caché de post-escritura de este ejercicio, el 40% de las veces que hay un fallo se va a producir además la escritura de la línea reemplazada. Por tanto:

Penalización media que produce un fallo caché = tiempo de lectura + $0,4 * \text{tiempo de escritura} = 1,4 * \text{tiempo de lectura}$ (como el enunciado nos dice que lecturas y escrituras consumen la misma cantidad de tiempo...)

Obtenemos ahora la latencia para los dos tamaños de línea que se nos piden:

$T_{\text{clock}} = 1/500\text{MHz} = 2\text{ns}$

$T_{\text{wait}} = 50 \text{ ns} = 25 \text{ ciclos}$:

-> Lectura de 4 palabras: 1 (inicio de la operación (dirección: 32 bits)) + 25 (T_{wait}) + 4 (leer 4 palabras de 32 bits con un bus de 32 bits) = 30 ciclos.

-> Lectura de 16 palabras: 1 (inicio op.) + 25(T_{wait}) + 4(lectura) + 1(espera para 5 ciclos) + 4(") + 1(") + 4(") + 1(") + 4(") = 45 ciclos.

Por tanto, las penalizaciones por fallo para las 2 configuraciones quedan:

Penalización media que produce un fallo cache (4 palabras/línea): $1,4 * 30 = 42$ ciclos

Penalización media que produce un fallo cache (16 palabras/liena): $1,4 * 45 = 63$ ciclos

Ahora ya podemos obtener los ciclos por instrucción que se gastan en la gestión de fallos cache en ambas configuraciones:

Cache de 4 palabras/línea: $0,05$ (fallos/inst.) * $42 = 2,1$ ciclos

Cache de 16 palabras/línea: $0,03$ (fallos/inst.) * $63 = 1,89$ ciclos

12.- Sea un computador dotado de un sistema de jerarquía de memoria con memoria caché y memoria principal. El ancho de palabra en ambas es de 32 bits, y las direcciones de memoria también son de 32 bits.

a) La memoria caché tiene un tamaño de 16K palabras, es asociativa por conjuntos de 4 vías, cada bloque tiene 16 palabras, tiene una tasa de fallos de 4% y el t. cierto consume un ciclo del reloj.

Indicar cuales son los campos en los que se descompone una dirección de memoria de cara a la caché, cuantos bits se necesitan para cada uno de ellos y para qué se usan dichos campos.

Como cada bloque tiene 16 palabras * 4B = 64B, el desplazamiento de bloque será de 6 bits.

Al tratarse de una memoria asociativa por conjuntos de 4 vías, si cada bloque tiene 16 palabras, un conjunto tendrá 4 vías/conjunto * 16 palabras/vía = 64 palabras/conjunto.

Como la cache tiene 16K palabras, el numero de conjuntos será 16K palabras / 64 palabras/conjunto = 2^8 conjuntos, así que el índice de conjuntos ocupa 8 bits.

Para la etiqueta quedan entonces $32 - (6+8) = 18$ bits

ETIQUETA = 18 BITS / ÍNDICE = 8 BITS / DESPLAZAMIENTO = 6 BITS

b) El bus que comunica a la M.P con la caché es de 2 palabras de ancho. Por otra parte, cada vez que se quiere transferir un bloque de 16 palabras de memoria principal a caché: Procesador la 1ª dirección del bloque, lo que ocupa 1 ciclo; La memoria tarda 2 ciclos en devolver las 2 primeras palabras del bloque; El procesador envía a la memoria la siguiente dirección al mismo tiempo que transmite; Una vez la memoria ha recibido la siguiente dirección, tarda otros 2 ciclos en enviar al procesador las dos palabras siguientes, repitiendo estos dos pasos hasta completar el envío del bloque completo y si se requiere la transferencia de un nuevo bloque, es preciso dejar un hueco de 2 ciclos antes de iniciarla. Indicar cuántos ciclos consume un fallo de caché (cuánto tiempo se tarda en enviar bloque completo desde M.P. a caché).

Dado que el bloque ocupa 16 palabras y que cada envío de M.P. a caché consta de 2 palabras, el proceso será el siguiente: 1 (enviar dir1) + 2(Devolver pal1-2 que concatena con enviar dir2) + 2(Devolver pal3-4 que concatena con enviar dir3) + ... + 2(Devolver pal13-14 que concatena con enviar dir8) + 2 (devolver pal 15-16).

Por tanto, si además tenemos en cuenta el hueco de 2 ciclos para una nueva petición de bloque, en un fallo de caché consumiremos $1 * 2 * 8 + 2 = 19$ ciclos.

c) Si un ciclo de reloj dura 2 ns, indicar el t. medio de acceso a caché medido en segundos.

Un acierto consume 1 ciclo. Por tanto, según la fórmula del tiempo medio de acceso a memoria tenemos que:

$T_{\text{acceso}} = t_{\text{cierto}} + T_{\text{fallos}} * P_{\text{fallo}} = 1 + 0,04 * 19 = 1,76$ ciclos.

Si el ciclo de reloj dura $2 * 10^{-9}$ s entonces el tiempo medio de acceso a memoria es de $3,52 * 10^{-9}$ s.