

Cuestiones conceptuales de memoria cache

1. ¿Cuál es el principal objetivo de la jerarquía de memoria?
2. ¿Cómo se explica que siendo la memoria cache tan sumamente pequeña en comparación a la MP se logre que más del 99% de las referencias a memoria se localicen en la cache?
3. ¿Cómo puede el programador sacar ventaja del conocimiento de cómo se encuentra organizada la cache en el sistema?
4. ¿Diferencia la CPU las direcciones de memoria cache de las de memoria principal? En caso afirmativo, indíquese cómo.
5. ¿Cómo se direcciona/indexa la memoria cache?
6. ¿Cómo es posible identificar el bloque de MP que se encuentra almacenado en un momento dado en una cierta línea de cache?
7. ¿Cuántos comparadores se necesitan en una cache de 1024 líneas con correspondencia directa? ¿Cuántos se necesitarían en la misma cache si la correspondencia fuese de 4-way?
8. ¿Qué tecnología de memoria RAM es la empleada en las cache, muy especialmente en los niveles superiores?
9. ¿Qué ventajas e inconvenientes tiene una cache completamente asociativa? ¿Cuándo estaría recomendado su empleo?
10. ¿Qué es más probable que se encuentre, una cache totalmente asociativa de 512B o una de 32KB?
11. Para un mismo tamaño de memoria cache ¿cuál consumirá más, una cache directa o una asociativa? ¿Por qué?
12. Si la cache explota el principio de localidad espacial ¿por qué no se hacen bloques de tamaño mayor?
13. Se dispone de una cache asociativa 4-way de tamaño 32KB, la cual se estructura en 256 conjuntos ¿Qué bits de la dirección definen el campo conjunto?
14. ¿Cómo se determina el total de bloques de MP que podrían llegar a mapearse a un mismo conjunto de memoria cache a lo largo del tiempo?
15. ¿Del resultado de qué operación aritmética se deriva el conjunto en el que se mapea un cierto bloque de MP?
16. ¿Del resultado de qué operación aritmética se deriva la etiqueta de un cierto bloque de MP?
17. ¿Del resultado de qué operación aritmética se deriva el número de bloque de MP?
18. En una cache completamente asociativa de tamaño 1KB, con tamaño de línea de 16B ¿cómo se determina la línea en la que se almacenaría el bloque de MP 0x1000202?
19. En el mismo escenario anterior, si se emplease, en su lugar, una correspondencia asociativa de 4-way ¿en qué líneas de la cache se podría almacenar el bloque de MP 0x1000202? **Nota:** Las líneas de MC se enumeran en el rango [0, 1, 2, ..., Num_total_bloques_MC-1]
20. ¿Qué ventaja ofrece el aumento del grado de asociatividad en la cache?
21. Suponiendo que en el conjunto 4 de una cache que dispone de 16 conjuntos se encuentra almacenado un bloque de MP cuya etiqueta es 0x3F ¿sabrías calcular de qué bloque se trata?
22. En la línea 4 de una cache completamente asociativa se halla almacenado un bloque de MP cuya etiqueta es 0x1000 ¿sabrías calcular de qué bloque se trata?
23. ¿Resulta válido un valor del número de vías que no sea potencia de 2?

24. Suponiendo que no varía ni el tamaño de la cache ni el número de vías ¿cómo se vería afectado el tamaño de la memoria de control si se duplicara el tamaño de bloque?
25. Suponiendo que el tamaño de la cache permanece constante, razonar que puede suceder para que el tamaño del campo etiqueta disminuya en un bit.
26. Partiendo del formato de dirección que se muestra en la figura y asumiendo un tamaño constante para la memoria cache, indíquese de qué forma variarían el tamaño de bloque y el número de vías para las cinco formatos de dirección resultantes de modificar el formato inicial de la dirección

	Etiqueta	Conjunto	Desplaz.
Dir. MP (inicial):	(23b)	(5b)	(4b)
Δ Bloque Δ N° Vías			
Caso 1:	(23b)	(4b)	(5b)
Caso 2:	(24b)	(4b)	(4b)
Caso 3:	(24b)	(3b)	(5b)
Caso 4:	(22b)	(5b)	(5b)
Caso 5:	(22b)	(7b)	(3b)

27. ¿Qué factor determina los fallos de inicio en una cache?
28. Para reducir el número de fallos de capacidad bastaría con aumentar el tamaño de la cache ¿por qué entonces no se aumenta el tamaño de L1 lo suficiente como para minimizar la tasa de fallos?
29. ¿Cómo se pueden reducir o incluso eliminar los fallos por conflicto de la cache?
30. Para un mismo tamaño total de cache L1 ¿Cuáles son las ventajas de emplear cache dual (I+D) frente a emplear cache unificada (I y D)?
31. Se dispone de una cache dual con tasas de acierto $H_I=0,994$ y $H_D=0,89$. Sabiendo que el 75% de los accesos a memoria corresponden a instrucciones ¿cuál sería la tasa de aciertos H resultante del combinado de ambas memorias?
32. En el caso anterior, si el tiempo de acceso a ambas cache es de 1 ns y el tiempo de acceso a la memoria principal es de 100ns ¿cuál sería el tiempo medio de acceso a memoria?
33. Sabiendo que el 75% de los accesos se realizan a instrucciones y que el tiempo de acceso a memoria principal es de 100ns ¿cuál de las siguientes configuraciones de L1 obtendría un menor tiempo medio de acceso a memoria?:
- Cache dual: 32KB L1I ($H_{L1I}=99,4\%$) + 32KB L1D ($H_{L1D}=89\%$); $T_{L1I}=T_{L1D}=1\text{ ns}$
 - Cache unificada: 64KB L1 ($H_{L1}=97,8\%$); $T_{L1}=2,5\text{ ns}$
34. En el caso anterior, si el porcentaje de acceso a instrucciones fuese del 65% ¿Qué configuración proporcionaría el menor tiempo medio de acceso a memoria?

35. ¿Sería correcto tomar la tasa de aciertos cómo el factor determinante de las prestaciones de la cache (menor tiempo medio de acceso a memoria)?
36. Se dispone de una cache L1 de tiempo de acceso $T_{L1} = 1\text{ns}$ y tasa de aciertos del 96%, la cual proporciona un tiempo medio de acceso a memoria de 4,96ns. Se desea reducir dicho tiempo medio de acceso por debajo de los 2ns, para lo cual se emplea un segundo nivel de cache (L2) cuya tasa de aciertos es del 99% ¿Qué tiempo de acceso, como máximo, debería tener la cache L2?
37. Como es sabido, tanto la tasa de aciertos como el tiempo de acceso a la cache son factores determinantes del tiempo medio de acceso a memoria. Sin embargo, a menudo ambos factores son contrapuestos, en el sentido de que la mejora de uno comporta el deterioro del otro. Con esta premisa, suponiendo la existencia de un sistema de cache con dos niveles L1-L2 ¿cuál de los dos factores debería priorizarse en el diseño de la cache L1? ¿cuál de los dos factores debería priorizarse en el diseño de la cache L2? Razonar la respuesta
38. ¿Para qué se utiliza un algoritmo de reemplazo y cuándo se requiere su empleo?
39. Algunos algoritmos de reemplazo, como el LRU, requieren de contadores ¿Cómo se determina el tamaño de dichos contadores? ¿Cuántos contadores harían falta en una cache MC{64KB,8,64B}?
40. ¿Cuáles son las ventajas e inconvenientes del algoritmo LRU? ¿Cómo se aplica realmente en los chips comerciales?
41. ¿Qué ventajas y/o inconvenientes presenta una política de escritura write-through (escritura directa) respecto a otra write-back (escritura posterior)?
42. ¿En qué niveles de la jerarquía de memoria sería más propio el empleo de write-through frente a write-back? ¿Por qué?
43. Razonar por qué la política de escritura write-no allocate (no ubicación en escritura) puede tener sus ventajas en determinados contextos
44. ¿Para qué sirven los buffers de escritura asociados a las cache? ¿Con qué política de escritura se hacen obligatorios? ¿Qué ocurre si el buffer se llena?
45. Supóngase un escenario en el que las operaciones de escritura se encadenan una tras otra abarcando bloque completos de datos (p.e., con objeto de inicializar extensas áreas de datos -algo habitual en S.O.) ¿Qué resultaría más ventajoso en este caso, una política de write allocate u otra de write-no allocate? Justificar la respuesta
46. ¿Se puede beneficiar la política de escritura write-back de la existencia de buffers de escritura?
47. Sea una cache cuya geometría está definida por los siguientes parámetros: MC{64KB,1,64B}. Asumiendo política de escritura write-back, indíquese qué porcentaje del total de bits ocupados por la cache corresponderían a la memoria de control
48. Si la anterior configuración de cache se modificase aumentando a 16 el número de vías y empleando algoritmo de reemplazo LRU ¿en qué porcentaje variaría el total de bits destinados a la memoria de control?
49. Sin variar el tamaño de la cache ¿de qué maneras se podría reducir el tamaño de la memoria de control asociada?
50. ¿Cómo se puede borrar el contenido de la memoria cache?