

Infraestrutura de Hardware- Engenharia da Computação

Lista – Capítulo 5 - Hierarquia de Memória

1. [1.0] Memória caches são importantes para fornecer uma hierarquia de memória de alto desempenho para processadores. Abaixo está uma lista de endereços de memória que estão endereçando palavras:

Endereços: 0x03, 0xb4, 0x2b, 0x02, 0xbf, 0x58, 0xbe, 0x0e, 0xb5, 0x2c, 0xba, 0xfd

 - a. Para cada um desses endereços, identifique o valor dos campos de tag e de índice para uma cache de mapeamento direto com 32 blocos de uma palavra. Também liste se cada referência é um hit ou uma miss, assumindo que a cache está inicialmente vazia.
 - b. Repita a questão anterior considerando blocos de quatro palavras e um tamanho total de oito slots. Também liste se cada referência é um hit ou uma miss, assumindo que a cache está inicialmente vazia.
 - c. Considerando os endereços de memória informados anteriormente devemos escolher a melhor configuração para uma cache de mapeamento direto que pode armazenar 16 palavras. Para as opções abaixo, qual a melhor configuração? Justifique sua resposta.
 - C1: a cache possui blocos de 1 palavra,
 - C2: a cache possui blocos de 2 palavras, e
 - C3: a cache possui blocos de 4 palavras
2. [1.5] Por convenção, uma cache é nomeada de acordo com a quantidade de dados que contém (ou seja, uma cache de 4 KiB pode conter 4 KiB de dados). No entanto, as caches também armazenam informações como tags e bits de validade. Supondo que as caches são endereçadas por byte, e que os endereços e palavras acessadas são de 32 bits.
 - a. Calcule o número total de bits necessários para implementar uma cache de 64 KiB com blocos de duas palavras.
 - b. Calcule o número total de bits necessários para implementar uma cache de 128 KiB com blocos de 32 palavras. Quanto maior é esta cache do que a cache de 64 KiB descrita na letra a.? (Observe que, mudando o tamanho do bloco, nós dobramos a quantidade de dados por slot sem dobrar o tamanho total da cache).
 - c. Explique por que a cache de 128 KiB, apesar da maior quantidade de dados, pode ter um desempenho mais lento do que a primeira cache.
3. [2.0] Para o projeto de uma cache de mapeamento direto com um endereço de 64 bits, os seguintes campos do endereço são usados para acessar a cache.
 - TAG: bits 63 – 11
 - Index: Bits 10-5
 - Offset: bits 4-0
 - a. Qual é o tamanho do bloco da cache (em palavras)?
 - b. Quantos blocos a cache tem?
 - c. Qual é a razão entre bits totais necessários para a implementação da cache sobre os bits de armazenamento de dados?

Na utilização da cache descrita acima os seguintes endereços foram acessados (decimal):
0,4,16,132,232,160,1024,30,140,3100,180,2180

- d. Para cada endereço acessado, liste (1) sua TAG, índice e deslocamento, (2) se foi um hit ou uma miss, e (3) que slots foram substituídos (se houver).
 - e. Qual é a razão de hits?
 - f. Mostre o estado final do cache, a cada entrada válida representada como <índice, tag, dados>. Por exemplo: <0,3, Mem(0xC00) – Mem[0xC1F]
4. [1.5] Em geral, o tempo de acesso ao cache é proporcional à capacidade. Suponha que os acessos de memória principais tenham uma latência de 70 ns e que 36% de todas as instruções acessem a memória dos dados. Abaixo tem-se os dados das caches L1 de cada um dos dois processadores, P1 e P2.
- P1: Cache L1: 2KiB, L1 miss rate: 8.0%, L1 Hit time: 0,66ns
P2: Cache L1: 4KiB, L1 miss rate: 6.0%, L1 Hit time: 0,90ns
- a. Supondo que o hit time de L1 determine os tempos de ciclo para P1 e P2, quais são suas respectivas frequências do clock?
 - b. Qual o tempo médio de acesso à memória para P1 e P2 (em ciclos)?
 - c. Assumindo uma CPI base de 1.0 sem qualquer atraso na memória, qual é a CPI total para P1 e P2? Qual processador é mais rápido?

Considere agora a adição de uma cache L2 ao processador P1. Use as capacidades da cache L1 e o hit time mencionados anteriormente. Para a cache L1 temos os seguintes dados: L2 Size: 1 MiB - L2 Miss Rate: 95% - L2 Hit Time 5.62 ns

- d. Calcule o AMAT para P1 com a adição da cache L2. O AMAT é melhor ou pior com a cache L2?
 - e. Assumindo uma CPI base de 1.0, qual é o CPI total para P1 com a adição da cache L2?
5. [1.5] O uso de vários níveis de caches é uma técnica importante para superar o tamanho limitado da cache de primeiro nível em detrimento de sua velocidade. Considere um processador com os seguintes parâmetros:
- a. CPI base (sem stalls de memória): 1.5
 - b. Velocidade processador: 2GHz
 - c. Tempo de acesso a memória principal: 100ns
 - d. taxa falta L1 por instrução: 8% **
 - e. Tempo de acesso da Cache L2 – Mapeamento direto: 15 ciclos
 - f. Taxa de Falta com cache L2 – mapeamento direto: 3,2%
 - g. Tempo de acesso da Cache L2 – 8-way associativa: 32 ciclos
 - h. Taxa de falta com cache L2- 8-way associativa: 1.6%

**Taxa de falta de cache de primeiro nível é por instrução. Suponha que o número total de faltas na cache L1 (instrução e dados combinados) equivale a 8% do número de instruções. Calcular o CPI para o processador na tabela usando:

- a. apenas uma cache de primeiro nível,

- b. uma cache de segundo nível de mapeamento direto e
- c. uma cache de segundo nível 8 way set-associativa.
- d. Como esses números mudam se o tempo de acesso principal à memória for o dobro do tempo? Observe até que ponto uma cache L2 pode esconder os efeitos de uma memória lenta.

6. [1.5] A memória virtual usa uma tabela de páginas para rastrear o mapeamento de endereços virtuais para endereços físicos. Este exercício mostra como essa tabela deve ser atualizada à medida que os endereços são acessados. Os seguintes dados constituem um fluxo de endereços virtuais. Assuma páginas de 4 KiB, uma TLB totalmente associativa de quatro entradas e a política de substituição LRU. Se as páginas devem ser trazidas a partir do disco, incrementar o próximo maior número de página.

Decimal	4653	2195	13900	34619	48838	12624	49193
---------	------	------	-------	-------	-------	-------	-------

TLB

Valid	Tag	Número página física	Tempo último acesso
1	0xb	12	4
1	0x7	4	1
1	0x3	6	3
0	0x4	9	7

TABELA DE PÁGINAS

Índice	Valid	Página Física ou Disco
0	1	5
1	0	Disco
2	0	Disco
3	1	6
4	1	9
5	1	11
6	0	Disco
7	1	4
8	0	Disco
9	0	Disco
A	1	3
B	1	12
C	0	Disco

- a. Para cada acesso mostrado acima, liste se o acesso é um hit ou erro no TLB; se o acesso é um hit ou erro na tabela de página, se o acesso é uma falta de página e o estado atualizado do TLB.
- b. Repita a questão da letra a. anterior, mas desta vez use páginas de 16 KiB. Quais seriam algumas das vantagens de ter um tamanho de página maior? Quais são algumas das desvantagens?
- c. Repita a questão da letra a., mas desta vez use uma TLB 2-way associativa.

7. [1.0] Qual o propósito de se usar código corretor de erros em um sistema hierárquico de memória e como funciona o código de Hamming com distância 3?
8. (extra – 1,5) Considere o programa abaixo que acessa os elementos de um vetor que tem os seguintes valores

```
start:
    addi x6, x0, 64
    lw x7, 104(x0)
    lw x7, 116(x0)
    lw x7, 132(x0)
    lw x7, 120(x0)
    lw x7, 180(x0)
    lw x7, 168(x0)
    lw x7, 176(x0)
    lw x7, 324(x0)
    lw x7, 136(x0)
loop:  lw x7, 144(x0)
    lw x7, 116(x0)
    lw x7, 272(x0)
    lw x7, 120(x0)
    lw x7, 124(x0)
    lw x7, 136(x0)
    lw x7, 168(x0)
    addi x6, x6, -1
    addi x0, x0, 0
    addi x0, x0, 0
    bne x6, x0, loop
    addi x0, x0, 0
    addi x0, x0, 0
    addi x0, x0, 0
    halt
vetor: .word 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41,
42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56
```

Simule o programa no CompSim faça um relato comparativo nos casos enumerados abaixo, considerando as seguintes características da memória cache: Número de faltas e se existe localidade espacial e temporal.

- Caso 1: Compare os resultados obtidos com uma cache de 16 palavras versus uma cache de 32 palavras com os mesmos tamanhos de bloco e graus de associatividade (a seu critério). Qual a justificativa para os resultados obtidos na taxa de faltas para as duas configurações analisadas?
- Caso 2: Compare os resultados obtidos com uma cache de 16 palavras com três tipos de associatividade: Mapeamento direto, Associativa por conjunto 2-way e Completamente associativo. Qual a justificativa para os resultados obtidos na taxa de faltas para as configurações analisadas?
- Caso 3: Compare uma cache de 32 palavras com blocos de 2 palavras versus uma cache de 32 palavras com blocos de 4 palavras, ambas de mapeamento direto. Qual a

justificativa para os resultados obtidos na taxa de faltas para as duas configurações analisadas?