

## Exercícios sobre Gerenciamento de Memória

01. Qual a função da MMU na gerência de memória?

R: É o hardware que traduz o endereço virtual em físico e mapear o endereço nas tabelas.

02. O que é o TLB?

R: Um dispositivo de hardware que é integrado na MMU

03. Explique a importância do TLB no desempenho da MMU.

R: O TLB ajuda MMU a ser mais rápido na busca na tabela de página.

04. Qual o tipo de fragmentação apresentado pelos métodos de gerência de memória baseados em partições fixas e por partições variáveis?

R: Nas partições fixas tem a fragmentação interna onde cada programa fica em uma partição que normalmente são maiores que o necessário, partições variáveis é a fragmentação externa que deixa espaços pequenos quando o programa vai terminando não deixando memória suficiente para outro programa.

05. Qual a diferença entre endereço físico e virtual?

R: Endereço virtual é gerado pela CPU e o físico é da memória RAM. Endereços virtuais são transformados em endereços físicos no momento de execução dos processos.

06. Qual a diferença entre monoprogramação e multiprogramação?

R: Multiprogramação é quando a memória está fazendo vários processos e monoprogramação é quando está em um único processo por vez.

07. Explique o que é a atividade de swapping e diga como ela pode prejudicar o desempenho do sistema computacional.

R: Faz uma transferência temporária entre a memória principal e a secundária só que pode demorar muito e consome muito a CPU.

08. Quando é mais interessante utilizarmos o algoritmo best-fit? E o worst-fit?

R: O best-fit, quando se deseja um processo mais rápido, e o worst-fit quando deseja criar o maior espaço possível.

09. Explique os algoritmos :

1. Considere um sistema de memória constituído por uma cache de 64K palavras de 1 octeto e uma memória principal de 8M palavras de 1 octeto. Considere ainda que a

cache tem blocos de 8 palavras e utiliza mapeamento directo.

a) Indique, justificando, o número de bits que se utiliza para endereçar a cache e a memória principal (isto é, a largura do barramento de endereços).

R: 16, 23

É dito que a cache tem 64K palavras de 1 octeto. Também é dito que a memória principal tem palavras de 1 octeto (8 bits). O comprimento de uma palavra é pois 8 bits (1 octeto = 1 byte).

Cache tem blocos de 8 palavras, como cada palavra = 1 octeto, cada bloco (ou linha) da cache usa 8 octetos. Capacidade da cache é  $64K = 2^6K = 2^{16}$ .

$2^{16} / 8 = 2^{16} / 2^3 = 2^{13}$  blocos (ou linhas). Uso 13 linhas de índice (para endereçar blocos) e mais 3 linhas de deslocamento (endereçar palavra dentro de um bloco). Na prática o número de linhas de endereço que entram na cache é de 16.

A largura do barramento de endereços é 23 porque  $2^{23}=8M$  (são necessárias 23 linhas para endereçar as 8M posições de memória).

[Se a largura de palavra fosse 16 bits (2 octetos) a capacidade da memória era de 4M palavras. Logo eram necessários 22 bits para endereçar os 4M endereços ( $2^{22}=4M$ ).]

É possível existir cache e memória principal com números de octetos por palavra diferente?

R: Isso não é expectável. Não é que seja impossível (por ex., podia ter um processador de 16 bits, uma cache de 16 bits e uma mem. primária de 8 bits; é claro que isto complicaria os acessos à mem. primária, obrigando a dois acessos por cada palavra a ler ou escrever...). Se nada for referido, assumir que tudo (processador, cache, mem. primária) tem o mesmo comprimento de palavra.

b) Indique, justificando, a largura dos campos de índice, etiqueta e deslocamento tal como interpretado por esta cache.

R: 13, 7, 3

A largura do campo de deslocamento é 3 porque  $2^3=8$  (número de palavras do bloco).

O índice será  $16-3=13$  porque  $2^{16}=64K$  e três bits são usados para o deslocamento (ver também texto da questão anterior).

A etiqueta será  $23-16=7$ .

? Como é constituída a cache?

É constituída por 8K blocos ( $2^{13}$ ), cada bloco com 8 palavras (uma palavra = 8 bits).

Na prática cada bloco (ou linha) da cache contém 8 palavras (cada com 1 octeto) e mais um campo de etiqueta com 7 bits.

NOTA: Quando se menciona a capacidade de uma cache, se nada for dito em contrário, não se inclui o espaço necessário às etiquetas nem aos bits de controlo; essa capacidade apenas se refere ao espaço destinado aos dados (palavras de memória do sistema).

? E se se tratasse de uma cache de mapeamento associativo de 2 vias?

Esta seria formada por 2 vias, cada uma com 4K blocos.

(64K total, 2 vias implica cada uma com 32K; considerando blocos de 8 palavras, temos cada via/módulo com 4K linhas ou blocos)

O campo de índice passaria a ter 1 bit a menos (12) e o campo etiqueta ganharia esse bit (passaria a 8 bits).

2. Considere um processador com um espaço de endereçamento físico de 64K palavras, com palavras de 2 octetos, e que usa uma cache de mapeamento directo que interpreta os bits de endereço da seguinte forma:

15-14-13-12-11-10-9-8-7-6-5-4-3-2-1-0

|-----Etiqueta-----|---Índice---| -Bloco

(3 bits - bloco, 6 bits - índice, 7 bits - etiqueta)

a) Qual é a capacidade para dados desta cache, em octetos?

R:  $2^{(3+6)} = 2^9 = 512$  palavras de 2 octetos; logo capacidade = 1K octeto ( $2^{10}$ ).

b) Mantendo a capacidade da cache, mas tornando-a uma cache com duas vias de associatividade, quais são as alterações em termos da forma como o endereço é interpretado?

R:

Passando a duas vias, cada via gastará metade da capacidade ( $512 / 2 = 256 = 2^8$ ).

Logo o índice reduzirá um bit (Desloc.: 3 bits, Índice: 5 bits, Etiqueta: 8 bits)

=====

II. 1. Considere um computador com 16 bits de endereço, em que a palavra de dados é representada por 16 bits, e uma cache em mapeamento directo, com capacidade de 1k octetos e bloco de 4 palavras.

a) Indique, justificando, quantos bits de índice tem a cache.

R: 7

Bloco = 4 ( $2^2$ ) => 2 bits de deslocamento

Cada bloco usa  $4 * 16$  bits = 8 octetos (4 palavras de 16 bits cada)

Se a cache tem 1K octeto ( $2^{10}$ ), então  $2^{10} / 8 = 2^7$  blocos, logo índice tem 7 bits

E a etiqueta será 16 bits de endereço menos 7 menos 2, ou seja, 7.

7, 7, 2

=====

IV a) Considere um processador com um barramento de endereços de 24 bits, no qual se pretende instalar um circuito de memória RAM com 8k, a partir do endereço 16k.

O circuito combinatório que descodifica esta RAM (ou seja, o circuito que gera o chip-select) tem 13 entradas.

R: Falso;  $8K = 2^3 * K = 2^{13}$ ,  $24 - 13 = 11$ ,  $11 + \text{mem\_enable} = 12$

24 bits endereço [ $2^{24} = 2^4 M = 16M$  ( $1K = 2^{10}$   $1M = 1K * 1K = 2^{20}$ )

$8K = 2^3 \cdot K = 2^{13}$  : 13 linhas de endereço da RAM

$24 - 13 = 11$  linhas de endereço a usar na descodificação da mem. (geração do chip select) (chip enable=chip select)

Se às 11 linhas identificadas atrás adicionarmos o sinal mem\_enable obtemos 12 linhas.

As 13 linhas menos significativas vão directamente às entradas de endereço da memória e usam-se as restantes para gerar o chip select (esta é a abordagem comum).

O facto de a memória ser colocada no endereço 16K não levanta qualquer problema pois é um múltiplo da capacidade da memória. Logo é um caso típico simples.

[Desafio: colocar em 14K era mais complicado... como se fazia?

Pista: Antes organizava a memória em blocos de 8K (capacidade da memória; se existissem várias memórias usaria a de menor capacidade; as outras de maior capacidade usariam 2 ou mais blocos...). Se 14K, o máximo divisor comum é 2K pelo que teria de organizar a memória em blocos de 2K e a RAM iria ocupar 4 blocos.

$2K = 2^{11} \Rightarrow$  uso onze linhas dentro de um bloco; para distinguir blocos uso as linhas de endereço a seguir ( $24 - 11 = 13$ ) (13 linhas mais significativas do endereço do sistema). O sinal de chip select seria obtido através do OR dos sinais que seleccionam cada um dos 4 blocos ocupados pela RAM]

? Neste caso, teríamos de ter em conta o memory enable? Se o circuito for implementado com um multiplexer, o memory enable será utilizado para activar (ou desactivar) o MUX ... Contabiliza-se o memory enable?

- 1) Diferencie memória interna de memória externa e apresenta exemplos de cada uma:  
**Resposta:** memória interna são aquelas acessadas directamente pelo processador, como memória principal, memória cache e registradores. Memória externa são aquelas que dependem de uma operação de I/O, como discos magnéticos, DVD, pen drives, entre outras.
- 2) Diferencie capacidade de memória e tamanho físico de memória.

**Resposta:** tamanho físico de memória é a dimensão do chip de memória, ou seja, seu tamanho físico em milímetros quadrados. Capacidade de memória é a capacidade de armazenamento em bytes, ou seja, sua densidade.

3) :

**Resposta:** acesso aleatório é possível através de uso de endereços de memória, onde o processador indica o endereço de memória e o controlador lê ou grava dados nesse endereço diretamente, não precisando acessar os demais endereços.

4) Como se obtém a taxa de transferência de um barramento?

**Resposta:** através da fórmula:  $TX = Fr.Qb$  onde TX = taxa de transferência, Fr é a frequência em Hertz e Qb é a quantidade de bits do barramento.

5) Comente os três elementos conflitantes do projeto de memórias?

**Resposta:** os três elementos conflitantes de um projeto de memória são: capacidade, velocidade e custo. As memórias mais rápidas são mais caras, maiores fisicamente e são acessadas mais rapidamente, sendo utilizadas para cache e registradores. Memória principal é mais lenta, menor fisicamente e mais lentas.

6) Explique a hierarquia de memória:

**Resposta:** o subsistema de memória é dividido em vários níveis em uma estrutura hierárquica onde quanto mais próximo do processador, maiores fisicamente e mais rápidas são as memórias, porém são mais caras e menos densas. As memórias mais distantes do processador são menores fisicamente, mais densas, mais baratas, porém possuem um custo de por bit muito baixo. Devido as questões conflitantes de projeto de memórias, o computador possui esse subsistema de memória que visa equiparar as velocidades de operação entre memória e processador. Assim, para melhorar o tempo de acesso a memória principal, são utilizadas memórias mais rápidas intermediárias ao processador.

7) Exponha os elementos de memórias em um nível alto na hierarquia e em um nível abaixo na hierarquia:

**Resposta:** quanto mais baixo no nível hierárquico, mais lenta é a memória, maior é a capacidade de armazenamento, menor o tamanho físico da memória, menor o custo. Quanto mais alto na hierarquia, mais rápida são as memórias, menos densas, maior o tamanho físico e maior o custo.

8) Diferencie memória volátil e não volátil e de exemplos:

**Resposta:** memória volátil são aquelas em que os dados se perdem quando o computador é desligado, ou seja, são temporários e quando o fornecimento de eletricidade é cortado os dados são apagados, como a memória principal, cache e registradores. Memória não volátil são aquelas que os dados são permanentes, ou seja, mesmo sem alimentação elétrica os dados não se perdem, como HD, DVD, Pen drives, Fitas magnéticas, etc.

9) Comente sobre memórias RAM estáticas e RAM dinâmicas e onde são empregadas:

**Resposta:** as memórias RAM estáticas são feitas de transistores, que são maiores fisicamente, porém menos densas. Enquanto houver fornecimento de eletricidade seu

estado permanece inalterado, não necessitando de *refresh*. As memórias RAM dinâmicas são confeccionadas com capacitores, que por serem menores, tornam as memórias fisicamente maiores ou seja, mais densas. Capacitores perdem carga e precisam ser realimentados, assim necessitam de um circuito de *refresh*, que realimentará a memória regenerando os bits nela gravados. Isso torna a memória mais lenta, porém são mais baratas.

Memória RAM estáticas são utilizadas para registradores e caches, ao passo que as memórias RAM dinâmicas são empregadas em memória principal.

- 10) O que são e quais são as memórias de tecnologia de semicondutor somente leitura e principalmente leitura?

**Resposta:** As memórias somente leitura são as que são gravadas apenas uma vez, na fábrica ou por meio de aparelho em laboratório e podem ser ROM e a PROM. Já as memórias principalmente leitura são memórias predominantemente leitura, mas ocasionalmente podem ser regravadas poucas vezes. Como exemplo, podemos citar EPROM, EEPROM e memória Flash.

- 11) O elemento básico da organização da memória é a célula. Comente:

**Resposta:** a memória é organizada de forma que os bits possam ser acessados para leitura e escrita. Dessa forma, os bits devem armazenar duas informações possíveis: 0 ou 1. Essas duas informações deverão ser gravadas em compartimentos endereçáveis chamados de células, podendo conter apenas um bit ou vários bits, dependendo da arquitetura. As células possuem geralmente três terminais, onde um serve para selecionar a célula, outro para indicar uma leitura ou gravação e outro para inserir ou ler o dado.

As células são endereçadas por meio de linhas e colunas.

- 12) Por que usar memórias cache?

**Resposta:** como existe uma diferença entre as velocidades do processador e da memória principal, o usa-se memória cache para melhorar o desempenho.

- 13) Quais informações contém nas memórias cache?

**Resposta:** o conteúdo das memórias cache sempre serão cópias de conteúdo da memória principal.

- 14) Comente sobre acertos e erros em memórias cache?

**Resposta:** quando o processador encontra a informação na memória cache, não necessitando um acesso a memória principal, ocorre um *hit*. Quando a informação não está na memória cache, ocorre um *miss*, devendo ser buscada na memória principal.

- 15) Comente as duas formas de organização memórias cache?

**Resposta:** a organização de memória cache pode ser interposta ou direta. Na interposta, a informação sempre passará pela cache antes de ser disponibilizada para o processador. Na direta, o processador receberá o dado que não está na memória cache, mas uma cópia deste ficará para a memória cache.

- 16) Diferencia localidade temporal de localidade espacial nas memórias cache?

**Resposta:** as memórias cache funcionam baseadas em duas premissas: localidade temporal e espacial. Quando uma

região de memória é constantemente acessada essa região tende a ficar no cache por mais tempo, explorando a localidade temporal. Quando uma região de memória é acessada, provavelmente regiões próximas tendem a ser acessadas já que os programas são sequenciais, assim as memórias caches exploram o principio da localidade espacial.