



Universidade Federal de Pelotas

Instituto de Física e Matemática

Departamento de Informática

Bacharelado em Ciência da Computação

Arquitetura e Organização de Computadores II

Aula 18

4. Memória virtual: manipulação de faltas de página e de faltas na TLB. Uma estrutura comum para hierarquias de memória.

Prof. José Luís Güntzel

guntzel@ufpel.edu.br

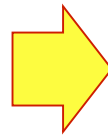
www.ufpel.edu.br/~guntzel/AOC2/AOC2.html

4. Hierarquia de Memória: integração de níveis

► Manipulação de Faltas de Página e de Faltas na TLB

- Uma falta no acesso à TLB ocorre quando nenhuma de suas entradas é igual ao endereço virtual gerado pelo processador
- Uma falta no acesso à TLB pode ter um dos seguintes desdobramentos:

A página requisitada está presente na memória



Basta criar uma nova entrada para ela na TLB

A página requisitada não está presente na memória

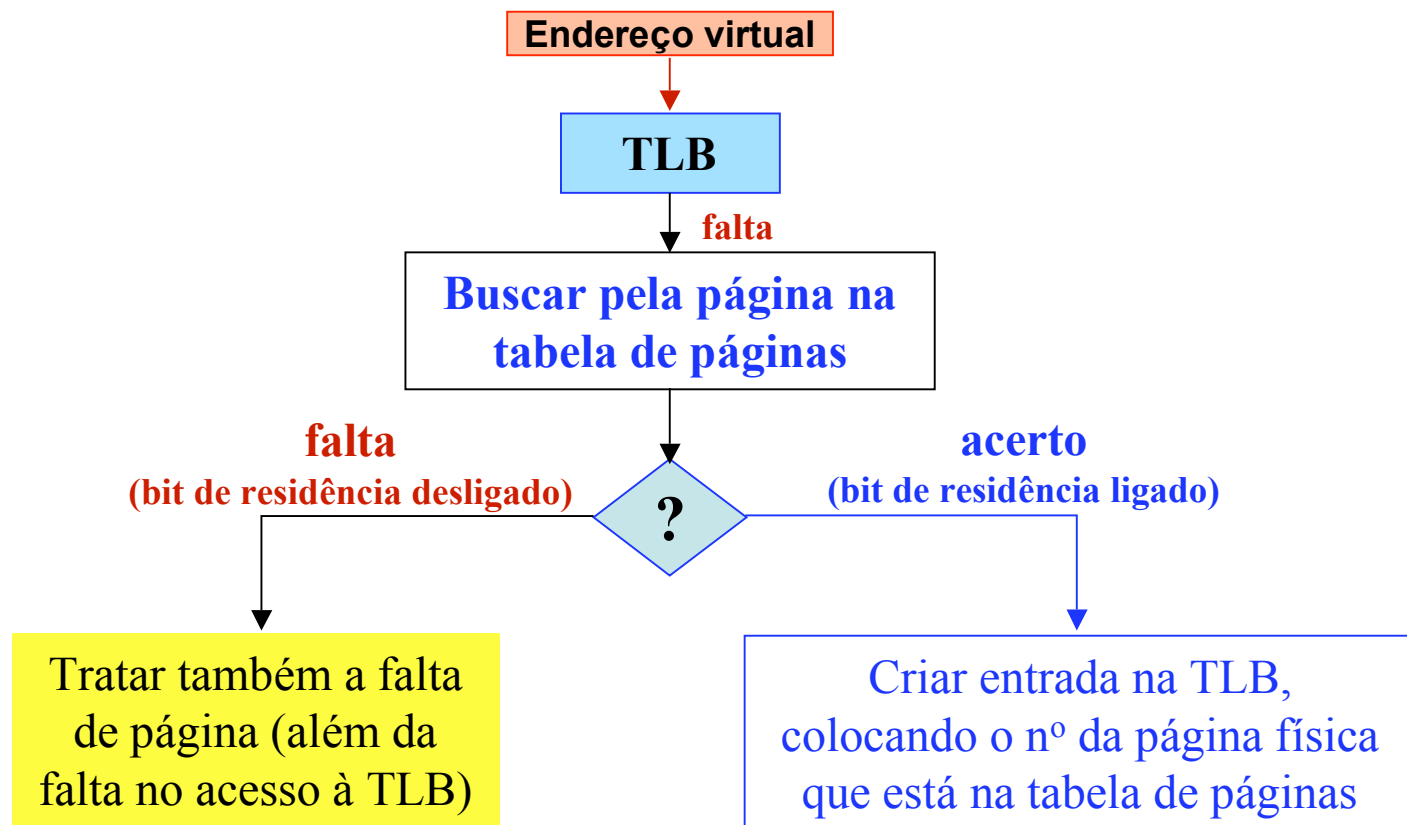


É necessário transferir o controle para o S.O., que vai tratar a falta de página

Como saber qual das duas situações ocorreu?

4. Hierarquia de Memória: integração de níveis

► Manipulação de Faltas de Página e de Faltas na TLB Como saber qual das duas situações ocorreu?



4. Hierarquia de Memória: integração de níveis

► Manipulação de Faltas de Página e de Faltas na TLB

Tratamento de falta na TLB pode ser por software ou por hardware:

- Seu tratamento requer uma pequena seqüência de operações para copiar a entrada da tabela de páginas da memória principal para a TLB

O Tratamento de faltas de página requer uso do mecanismo de exceção:

- Para interromper o processo ativo
- Transfere o controle para o S.O.
- Falta de página geralmente é reconhecida durante o ciclo de relógio usado para acessar a memória
- PC é salvo em um registrador especial chamado **EPC** (a fim de poder retornar a execução a partir da próxima instrução)

4. Hierarquia de Memória: integração de níveis

► Observações Finais - 1

A gerência da memória virtual é um grande desafio devido o alto custo das faltas de página

Várias técnicas são usadas para resolver este problema:

1. As páginas devem ser grandes o suficiente para tirar proveito da localidade espacial (reduzindo a taxa de faltas)
2. O mapeamento de um endereço virtual em um endereço físico usando a tabela de páginas é feito de maneira totalmente associativa
3. O S.O. usa técnicas especiais (e.g., LRU e bit de referência) para determinar qual das páginas deve ser substituída

4. Hierarquia de Memória: integração de níveis

► Observações Finais - 2

As escritas no disco também são muito caras. A memória virtual:

- Usa o esquema *write-back*
- Controla se uma página não foi modificada (para evitar a escrita de páginas que não foram modificadas)

Mecanismo de memória virtual

- Realiza a tradução do endereço virtual para o espaço de endereçamento físico (usado no acesso à memória principal)
- Permite o compartilhamento protegido da memória principal entre diversos processos
- Simplifica a alocação da memória

A TLB funciona como uma cache para as traduções da tabela de páginas

4. Hierarquia de Memória: integração de níveis

► Uma Estrutura Comum para Hierarquias de Memória

Características	Valores típicos para caches	Valores típicos para memórias paginadas	Valores típicos para TLB
Tamanho total em blocos	1.000-100.000	2.000-250.000	32-4.000
Tamanho total em KB	8-8000	8.000-8.000.000	0,25-32
Tamanho do bloco em bytes	16-256	4.000-64.000	4-32
Penalidade por falta (ciclos de relógio)	10-100	1.000.000-10.000.000	10-100
Taxa de faltas	0,1%-10%	0,00001%-0,0001%	0,01%-2%

4. Hierarquia de Memória: integração de níveis

► Onde se Pode Colocar um Bloco?

A colocação de um bloco no nível superior da hierarquia pode usar um dos seguintes esquemas:

- Mapeamento direto
- Mapeamento associativo por conjunto (caso genérico)
- Mapeamento totalmente associativo

Esquema	Número de conjuntos	Blocos por conjunto
Mapeamento direto	Número de blocos na cache	1
Associativo por conjunto	$\frac{\text{Número de blocos na cache}}{\text{associatividade}}$	Associatividade (tipicamente de 2 a 8)
Totalmente associativo	1	Número de blocos na cache

4. Hierarquia de Memória: integração de níveis

► Como Encontrar um Bloco?

Depende do esquema adotado para colocação dos blocos

Associatividade	Método de localização	Número de comparações necessárias
Mapeamento direto	Índice	1
Associativo por conjunto	Indexar o conjunto, pesquisar entre os elementos	Grau de associatividade
Totalmente associativo	Pesquisar todas as entradas da cache	Tamanho da cache
	Tabela separada para busca	0

4. Hierarquia de Memória: integração de níveis

► Como Encontrar um Bloco?

Nos sistemas de memória virtual, a tabela de páginas é usada para indexar a memória principal

Fatores que motivam o mapeamento totalmente associativo nas tabelas de páginas:

1. Alto custo das faltas de página
2. Permite que o software utilize algoritmos bastante sofisticados para substituição de página
3. Pode ser facilmente indexado, sem necessidade de hardware extra
4. Quanto maior o tamanho da página, menor o overhead relativo representado pelo tamanho da tabela de páginas

4. Hierarquia de Memória: integração de níveis

► Como Encontrar um Bloco?

Caches e TLBs normalmente usam esquema associativo por conjunto:

- Acesso combina indexação e busca em um conjunto pequeno de elementos

Porém, alguns sistemas mais recentes têm usado caches com mapeamento direto devido ao pequeno tempo de acesso e simplicidade

4. Hierarquia de Memória: integração de níveis

► Qual dos Blocos Deve ser Substituído?

Totalmente associativa: todos os blocos são candidatos à substituição

Conjunto associativa: escolher um dos blocos pertencentes a um conjunto

Mapeamento direto: não há escolha!

Estratégias para a substituição de blocos:

- **Aleatória:** os blocos candidatos à substituição são escolhidos ao acaso, possivelmente contando com algum auxílio de hardware
- **Bloco usado há mais tempo (LRU):** o bloco substituído é aquele usado há mais tempo

4. Hierarquia de Memória: integração de níveis

► O Que Acontece em uma Escrita?

Write-through: a informação é escrita tanto na cache quanto no bloco da memória no nível inferior da hierarquia (mem principal, no caso da cache)

Write-Back/copy-back: a informação é escrita somente no bloco da cache. O bloco modificado é escrito no nível inferior da hierarquia somente no momento de sua substituição

4. Hierarquia de Memória: integração de níveis

► O Que Acontece em uma Escrita?

Vantagens do *Write-back*

- ❑ As palavras podem ser escritas individualmente pelo processador na velocidade da cache (ao invés de sê-lo na velocidade da memória principal)
- ❑ Escritas múltiplas dentro de um bloco podem ser feitas com uma única operação de escrita no componente da hierarquia inferior
- ❑ Quando os blocos são escritos de volta, o sistema pode fazer uso de uma banda passante extremamente alta para efetivar a transferência (já que todo o bloco é escrito)

4. Hierarquia de Memória: integração de níveis

► O Que Acontece em uma Escrita?

Vantagens do *Write-through*

- As faltas são mais simples e mais baratas de tratar (nunca há necessidade de escrever todo o bloco no componente de memória da hierarquia inferior)
- É mais simples de implementar que a *write-back* (que exige um *buffer* de escrita)

4. Hierarquia de Memória: integração de níveis

► O Que Acontece em uma Escrita?

Nos sistemas de Memória Virtual

- ❑ Na prática, somente write-back é usado (devido a longa latência nas escritas no componente de memória da hierarquia inferior - HD)
- ❑ O desempenho dos processadores vem melhorando a uma taxa bem maior do que a melhora no acesso à memória principal baseada em DRAMs
- ❑ Cada vez mais caches estão usando (ou virão a usar) a estratégia *write-back*