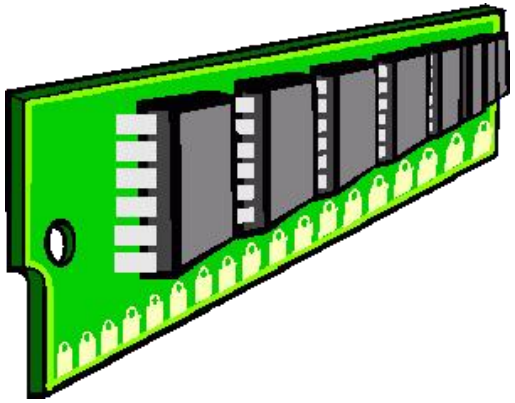
INSTITUTO FEDERAL
Santa Catarina



Gerenciamento de Memória

Professor: Alberto Felipe Friderichs Barros

INSTITUTO FEDERAL
Santa Catarina

Quem faz?

- SO Gerencia a memória e inclui a Memória Virtual.

O que faz?

- Usa técnicas para tornar a memória mais eficiente.

Por que faz?

- Busca do ideal para a memória.
- Resolver problemas: Concorrência, Proteção, Desempenho.

INSTITUTO FEDERAL
Santa Catarina

Introdução

Programas **precisam ser trazidos para a memória para serem executados**. É neste momento que programas transformam-se em processos. Para promover melhorias de desempenho, **o S.O deve manter vários processos na memória; isto é, compartilhar a memória**.

INSTITUTO FEDERAL
Santa Catarina

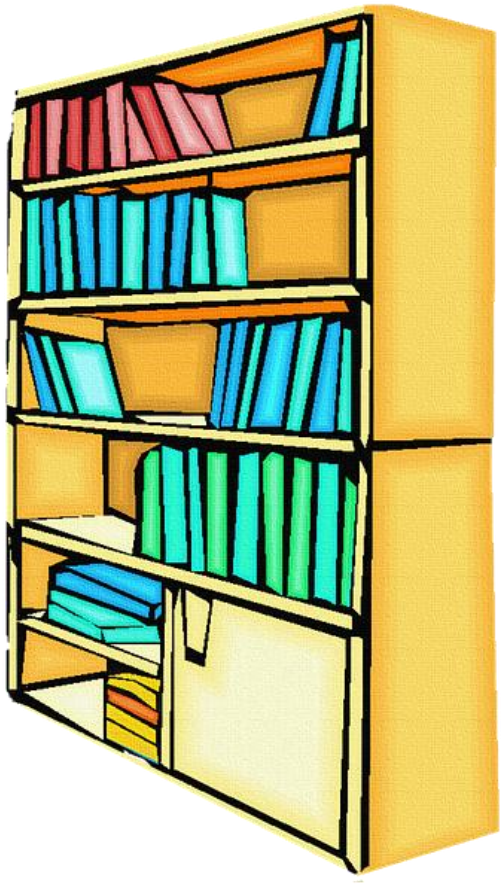
Memória

Todo computador é dotado de uma **quantidade limitada de memória**, que pode variar de máquina para máquina, a qual se constitui de um conjunto de circuitos capazes de armazenar valores: dados, programas a serem executados (alto nível) ou bits e bytes, estados (baixo nível)

INSTITUTO FEDERAL
Santa Catarina

Analogia

Armazenar →



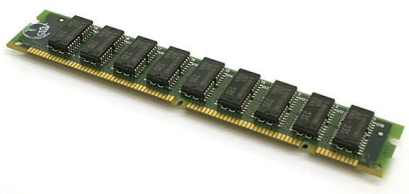
→ Recuperar

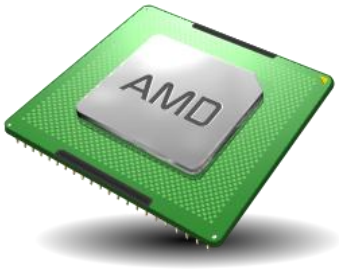
INSTITUTO FEDERAL
Santa Catarina

Idealmente programadores querem uma memória que seja:

- Grande
- Rápida
- Não volátil
- De baixo custo

* Infelizmente a tecnologia atual não comporta tais memórias.

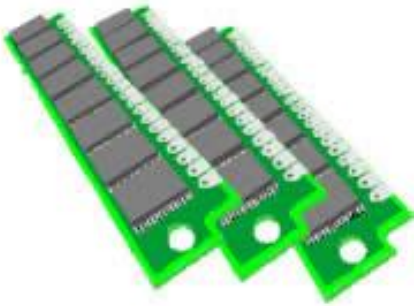




INSTITUTO FEDERAL
Santa Catarina

Introdução

Em um sistema de computação não é possível construir e utilizar **apenas um tipo de memória**. Para certas atividades, por exemplo, é fundamental que a transferência de informações seja a mais rápida possível. Outras atividades é preferido que os dados sejam armazenados por períodos mais longos.



INSTITUTO FEDERAL
Santa Catarina

Características

- Tempo de Acesso;
- Capacidade;
- Volatilidade;
- Tecnologia de Fabricação;
- Custo.

INSTITUTO FEDERAL
Santa Catarina

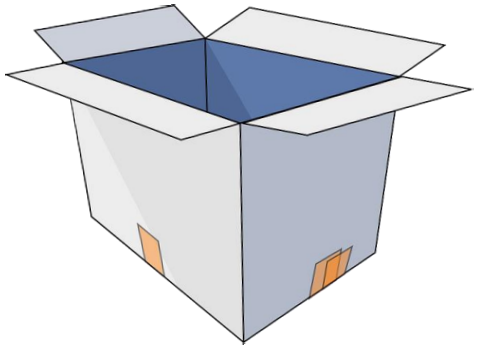
Tempo de Acesso

- Indica o tempo necessário para registrar uma informação, latência, nº de clock.
- Dispositivos eletromecânicos (discos, fitas, ..) tempo de acesso varia conforme a distância física entre dois acessos consecutivos - **acesso sequencial.**
- Memórias eletrônicas-igual, independentemente da distância física entre o local de um acesso e o local do próximo acesso - **acesso aleatório (direto).**

INSTITUTO FEDERAL
Santa Catarina

Capacidade

- Quantidade de informação que pode ser armazenada em uma memória;
- Unidade de medida mais comum é o byte, dependendo do tamanho da memória, isto é, de sua capacidade, indica-se o valor numérico total de elementos de forma simplificada, através da inclusão de K (kilo), M (mega), G (giga) ou T (tera).



INSTITUTO FEDERAL
Santa Catarina

Volatilidade

As memórias podem ser do tipo **volátil ou não volátil**.

- **Memória Volátil:** perde a informação armazenada quando a energia é desligada. Ex: Registradores, cache, memória principal.
- **Memória Não Volátil:** retêm a informação armazenada quando a energia é desligada. Ex: disco rígido, Fitas, pendrive, Cd-rom.

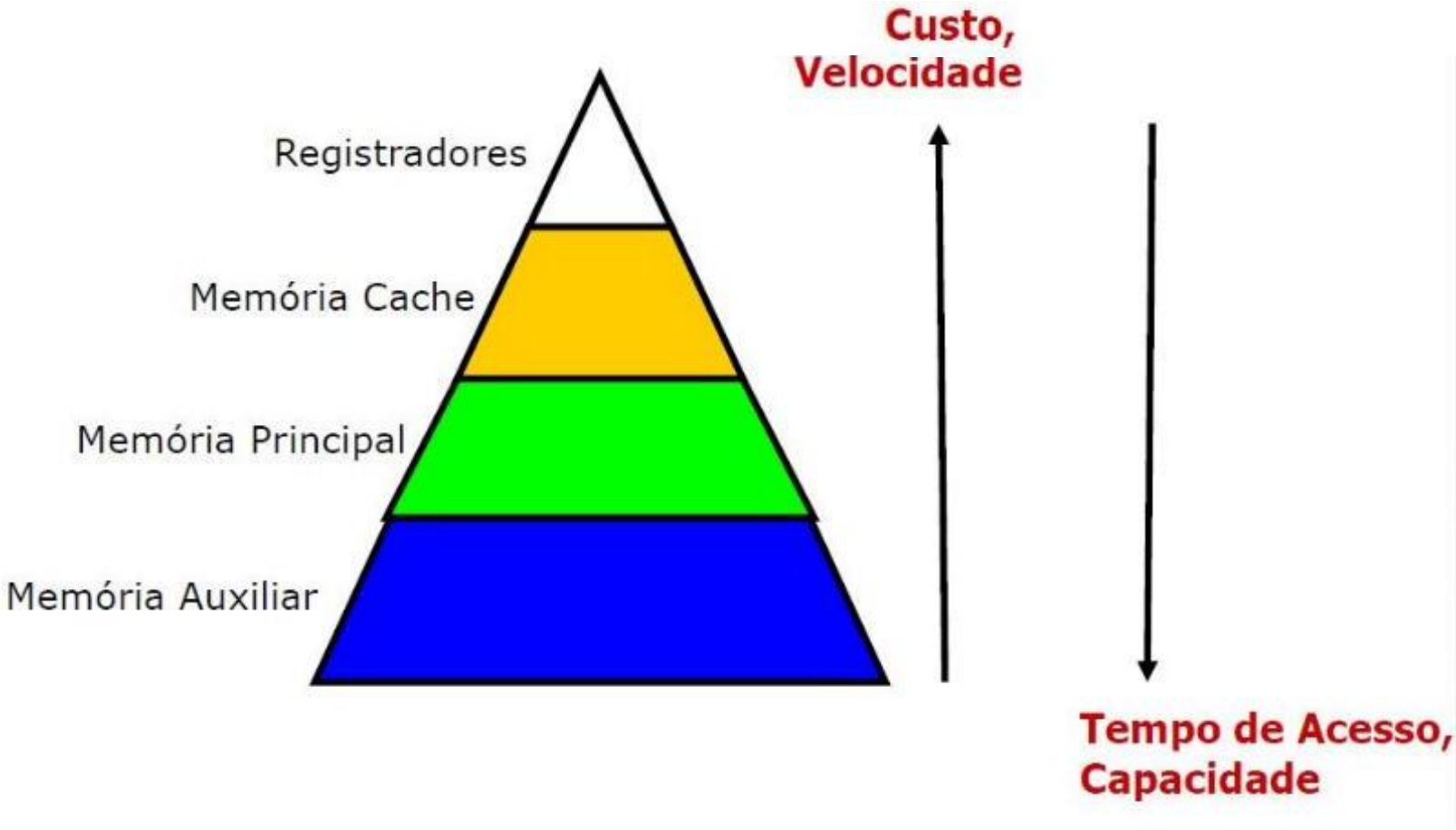
INSTITUTO FEDERAL
Santa Catarina

Tecnologias de Fabricação

- Semicondutores
- Magnético
- Óptico

INSTITUTO FEDERAL
Santa Catarina

Hierarquia

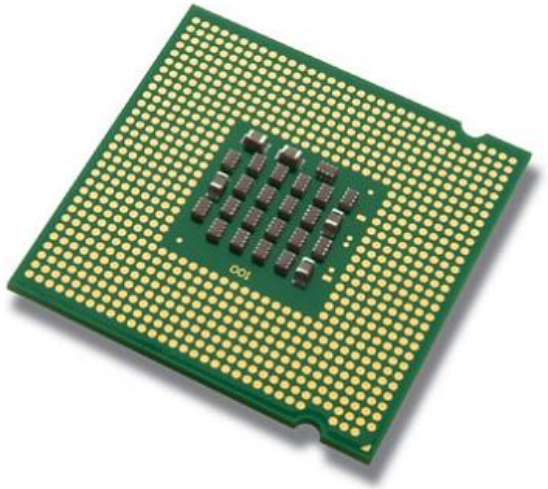


INSTITUTO FEDERAL
Santa Catarina

Registradores

A memória mais veloz e mais cara do sistema, são internos a CPU e possuem baixa capacidade de armazenamento.

- Tempo de acesso: 1 a 5 ns.
- Capacidade: 8 a 64 bits.
- Volatilidade: Volátil.
- Tecnologia: Semicondutores.
- Custo: Muito elevado.

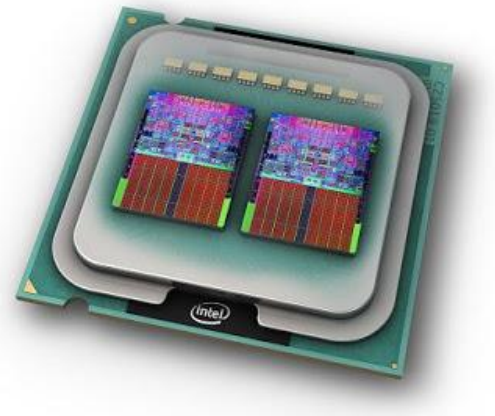


INSTITUTO FEDERAL
Santa Catarina

Cache

Memória intermediária utilizada para diminuir a latência entre o processador e a memória principal subdivide-se em: cache L1, L2 e L3; Atualmente existem no mercado memórias cache por volta de 16 MB

- Tempo de acesso: 5 a 7 ns.
- Capacidade: 16K a 16 MB.
- Volatilidade: Volátil.
- Tecnologia: SRAM.
- Custo: Elevado.



INSTITUTO FEDERAL
Santa Catarina

Memória Principal

É o dispositivo no qual armazena os dados que estão sendo executados, para que o processador busque as instruções a serem processadas.

- Tempo de acesso: 50 a 80 ns.
- Capacidade: 8 Gbytes.
- Volatilidade: Volátil.
- Tecnologia: DRAM.
- Custo: baixo.

INSTITUTO FEDERAL
Santa Catarina

Memória Secundária

É a memória mais barata, com mais espaço e comum nos computadores. São as mais lentas unidades de armazenamento de um sistema computacional. Exemplos: CD, DVD, Disco Rígido, Pendrives, Fitas, CD-ROM, etc.

- Tempo de acesso: 120 a 300 ms.
- Capacidade: 2 TB.
- Volatilidade: Não Volátil.
- Tecnologia: Ótico, Magnético, etc...
- Custo: baixo.

INSTITUTO FEDERAL
Santa Catarina

Hardware Básico

A memória principal e os registradores embutidos dentro do próprio processador são o único espaço de armazenamento de uso geral que a CPU pode acessar diretamente. Portanto, quaisquer instruções em execução e quaisquer dados que estiverem sendo usados pelas instruções devem estar na RAM. Se os dados não estiverem na memória, devem ser transferidos para lá antes que a CPU possa operar sobre eles.

INSTITUTO FEDERAL
Santa Catarina

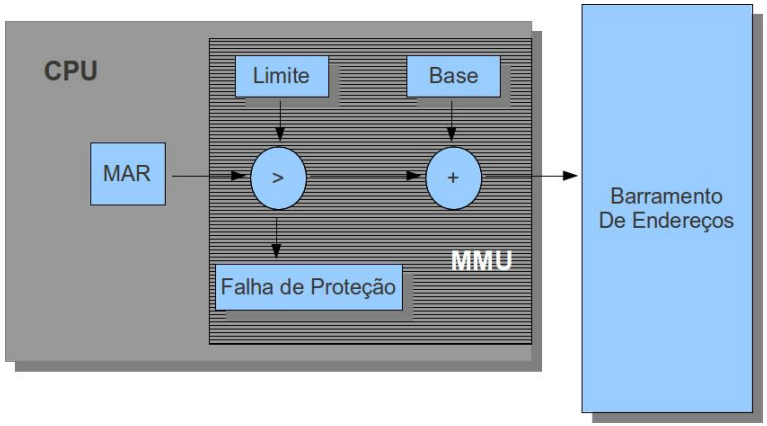
Hardware Básico

Os registradores que estão embutidos na CPU são geralmente acessáveis dentro de um ciclo do relógio da CPU. O mesmo não pode ser dito da memória principal que é acessada por uma transação no bus da memória. Para completar um acesso à memória podem ser necessários muitos ciclos do relógio da CPU. Nesses casos, o processador normalmente precisa ser interrompido, já que ele não tem os dados requeridos para completar a instrução que está executando. Essa situação é intolerável por causa da frequência de acessos à memória. A solução é adicionar uma memória rápida entre a CPU e a memória principal, geralmente no chip da CPU para acesso rápido, **a cache.**

INSTITUTO FEDERAL
Santa Catarina

Hardware Básico

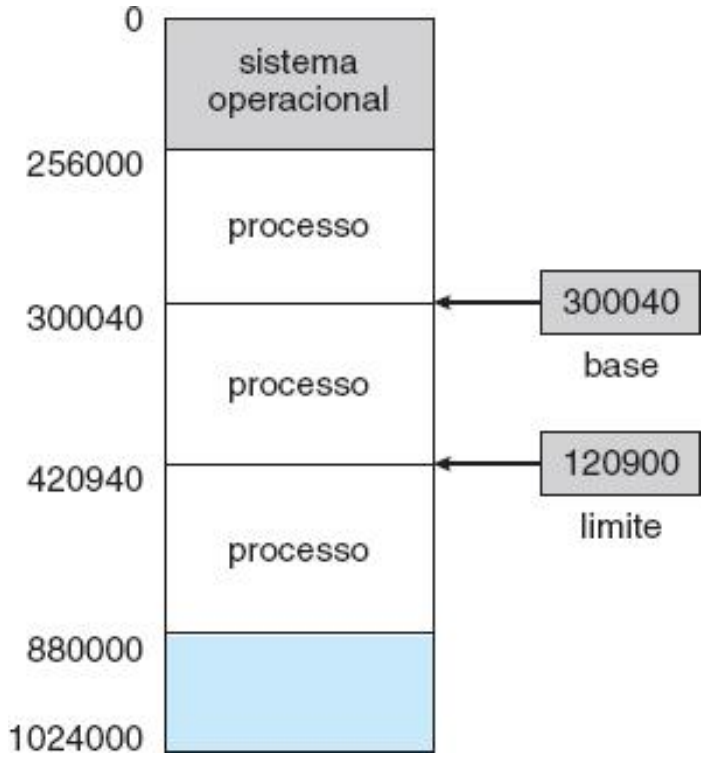
Além disso o S.O deve assegurar a operação correta. Para a operação apropriada do sistema, **proteger o sistema operacional** contra o acesso de processos de usuário. Em sistemas multiusuários, deve-se adicionalmente proteger os processos de usuário uns dos outros. Essa proteção deve ser fornecida pelo hardware porque o sistema operacional não costuma intervir entre a CPU e seus acessos à memória (em razão do comprometimento do desempenho).



INSTITUTO FEDERAL
Santa Catarina

Registradores

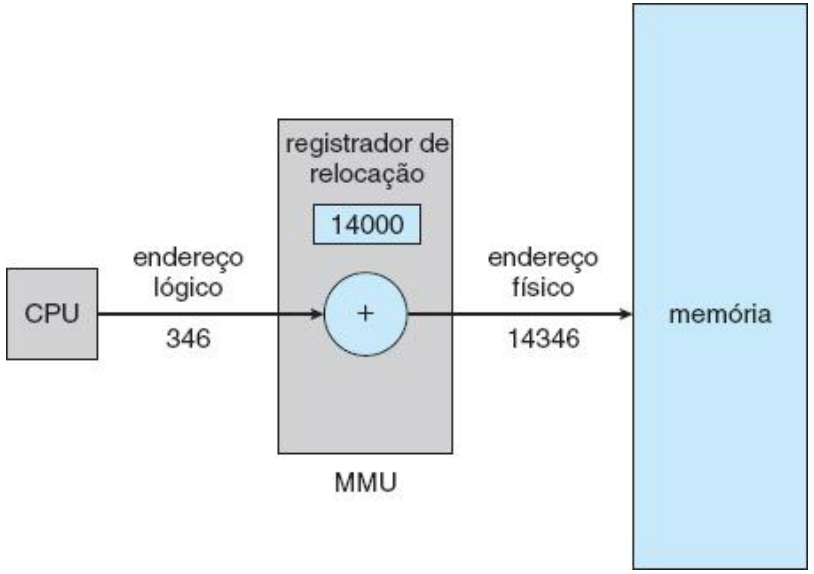
É possível fornecer essa proteção usando dois registradores, usualmente um registrador base e um registrador limite, O **registrador base** contém o menor endereço físico de memória legal; o **registrador limite** especifica o tamanho do intervalo.



INSTITUTO FEDERAL
Santa Catarina

Tipos de Endereços

Um endereço gerado pela CPU é comumente referenciado como **endereço lógico**, enquanto um endereço visto pela unidade de memória, isto é, aquele que é carregado no registrador de endereços da memória, costuma ser referenciado como **endereço físico**.

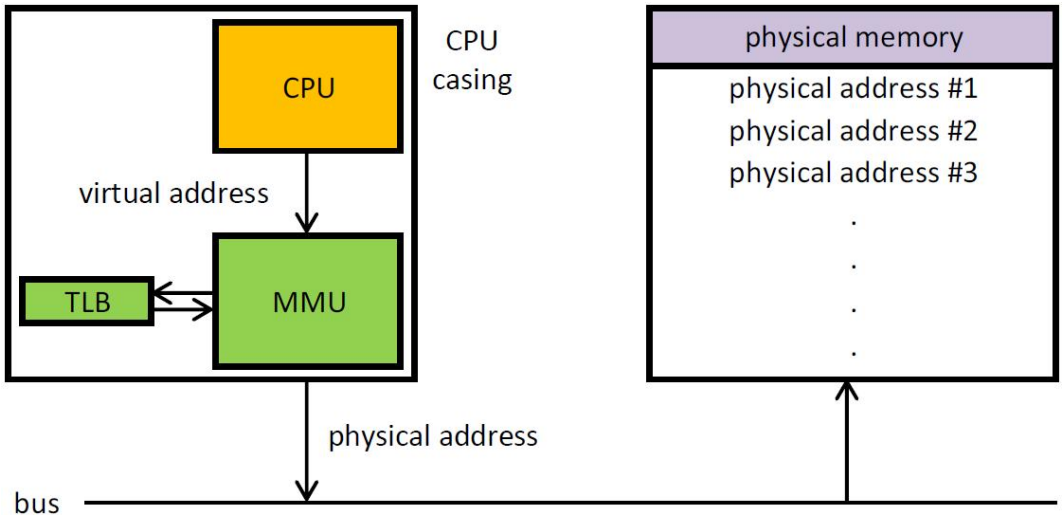


INSTITUTO FEDERAL
Santa Catarina

MMU

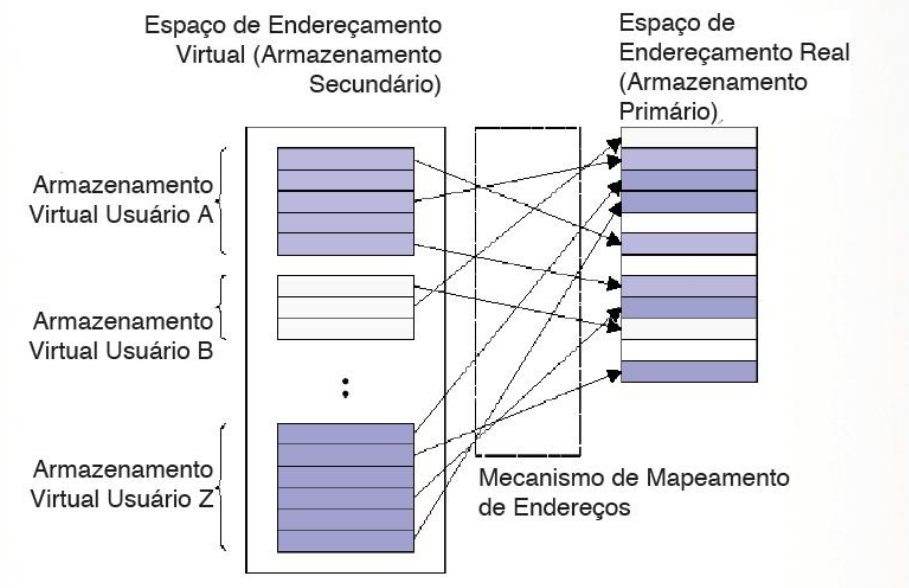
Memory Management Unit – componente da CPU que transforma os endereços virtuais em endereços físicos.

- O programa manipula endereços lógicos, ele nunca vê o endereço físico.



Gerenciador de Memória

Ao sistema operacional é destinada a função de coordenar e gerenciar a utilização dessas memórias de forma eficiente. Este serviço é implementado pelo sistema operacional através do gerenciador de memória



INSTITUTO FEDERAL
Santa Catarina

Gerenciador de Memória

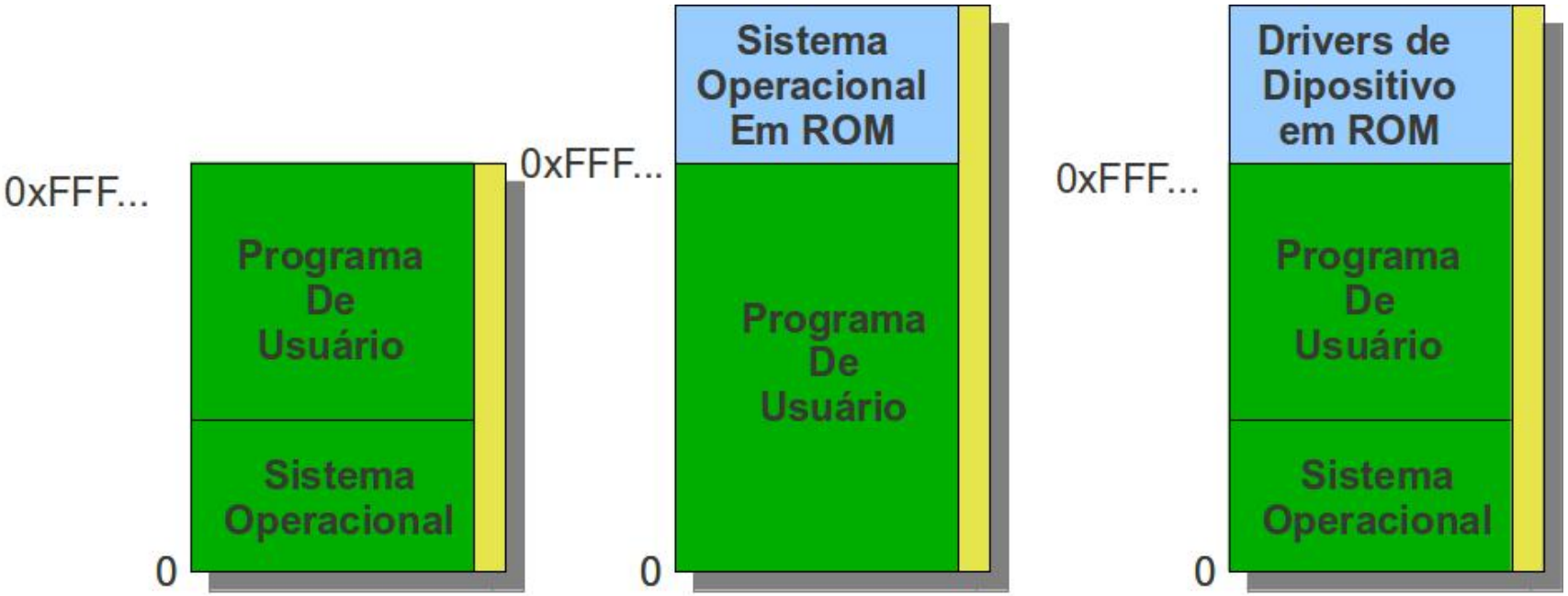
Módulo do Sistema operacional que cuida da administração da memória.

- **Gerenciar a hierarquia de Memória:**
- Gerenciar os espaços livres e ocupados (partições);
- Alocar e localizar processos/dados na memória (leitura e gravação).

- Controlar as partes da memória da memória que estão em uso, e as que não, para:
- Alocar memória aos processos, quando estes precisarem;
- Liberar memória quando um processo termina.
- Tratar do problema de swapping.

Monoprogramação

1) A Memória é dividida em duas áreas: do Sistema Operacional e do Usuário;
2) Sem proteção, pois um usuário pode acessar a área do Sistema Operacional;
3) Uso ineficiente da memória.



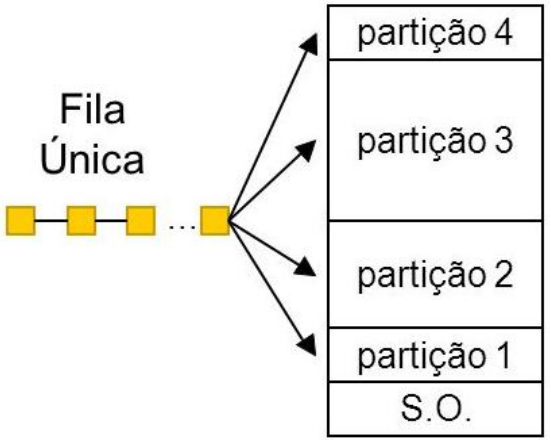
a) Mainframes b) handhelds c) Primeiros PCs

Multiprogramação

1) Como fazer para armazenar n processos na memória?

Com a multiprogramação começou a existir a necessidade do uso da memória por vários usuários simultaneamente

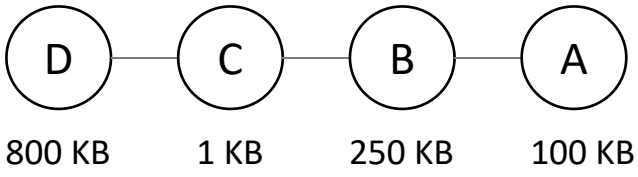
- Ocupação mais eficiente do processador;
- Divida a memória em partições de tamanho fixo;
- O tamanho de cada partição era estabelecido na inicialização do sistema;

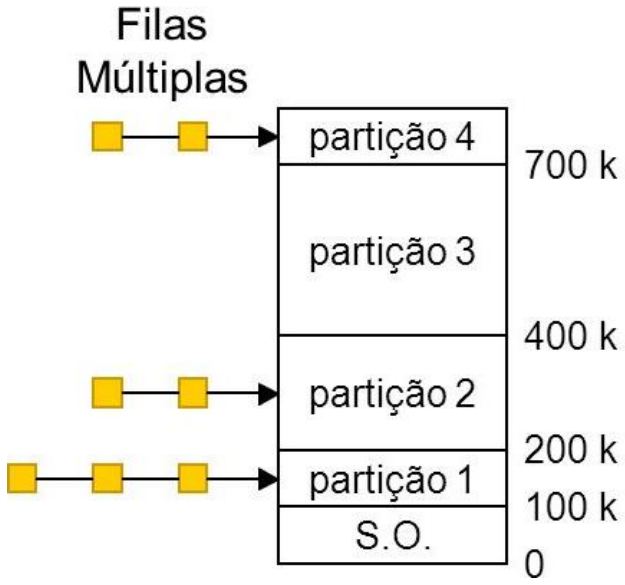


Multiprogramação

- Para alteração do particionamento, era necessário uma nova inicialização com uma nova configuração.
- O espaço que sobrar não será alocado.
- Sistema gerencia uma tabela de partições.

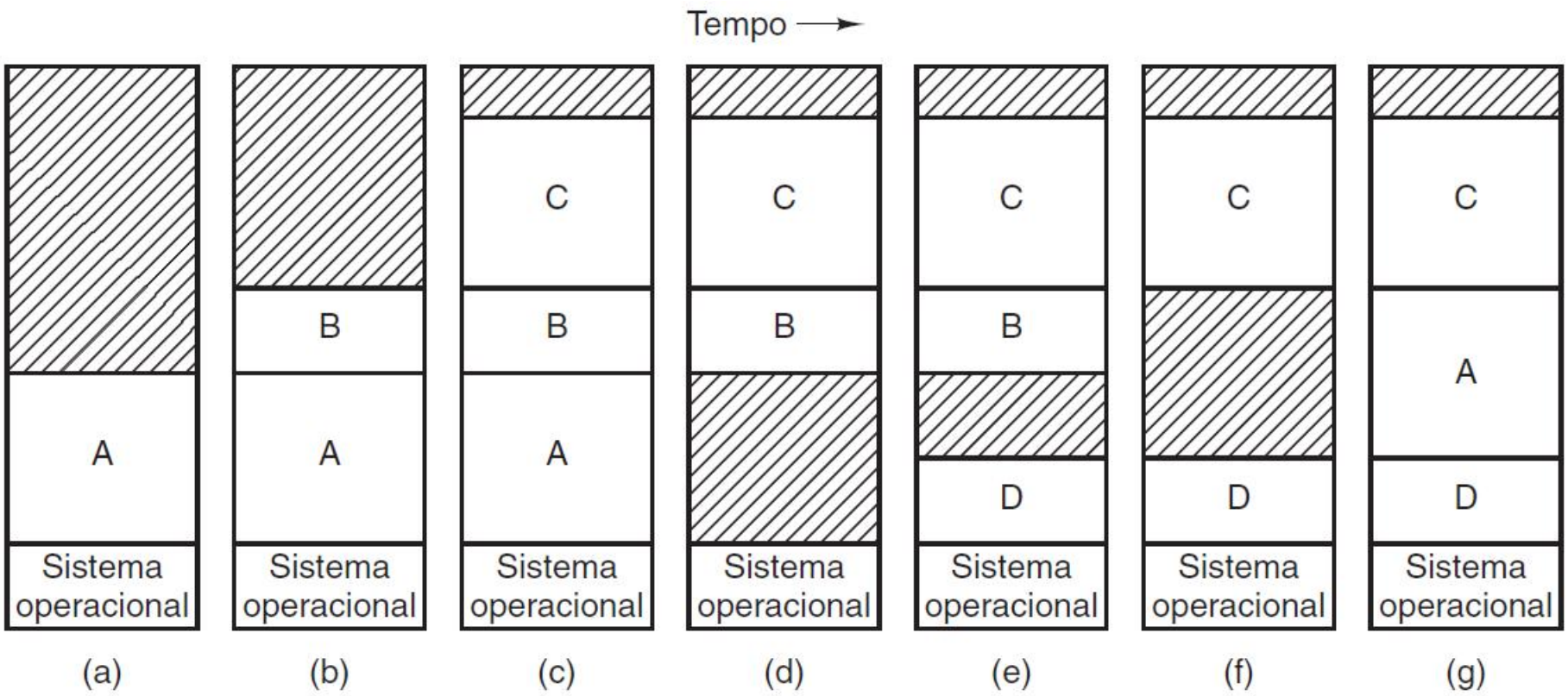
Partição	Tamanho
1	100 KB
2	200 KB
3	300 KB
4	100 KB





INSTITUTO FEDERAL
Santa Catarina

Partições Variadas



INSTITUTO FEDERAL
Santa Catarina

Tipos de Memória Particionada

1. Partições Fixas:
- Tamanho e numero de partições são estáticas;
- Tendem a desperdiçar memória;
- Mais simples, gera fragmentação interna

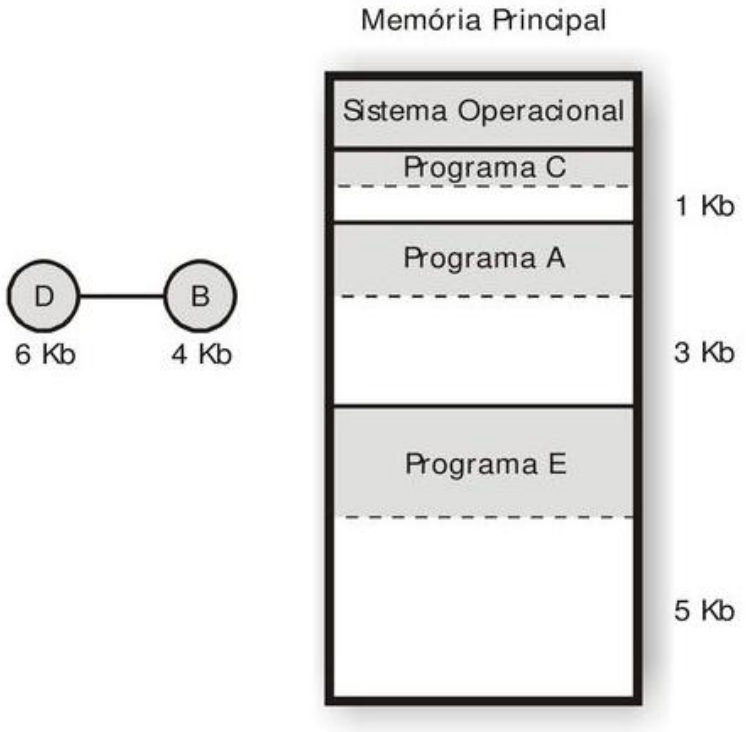
2. Partições Variáveis:
- Tamanho e numero de partições dinâmicas;
- Otimiza a utilização da memória, mas complica a alocação e liberação;
- Partições alocadas dinamicamente, gera fragmentação externa.

3. Segmentação:
- Programa é dividido em segmentos e fica espalhado pela memória. otimiza espaços

4. Paginação:
- Utiliza espaço não-contiguo, memória física dividida em frames e lógica em páginas;
- Processos preenchem páginas diminui fragmentação externa e otimiza alocação e liberação.

1. Partições Fixas

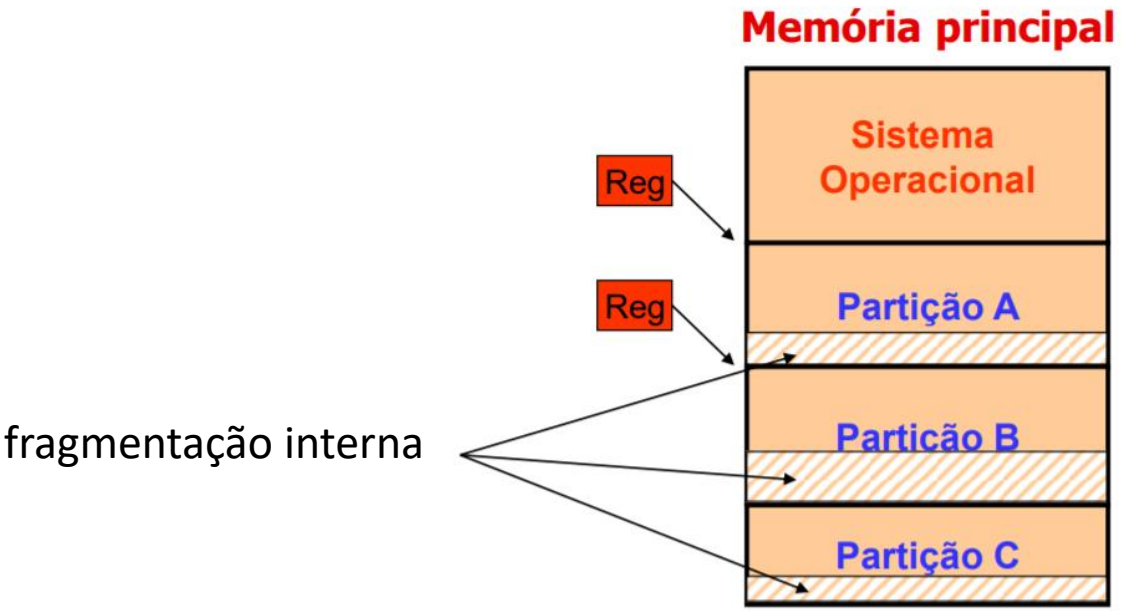
- Esquema mais simples de alocação de memória: dividir a memória em várias partições de tamanho fixo, de mesmo tamanho ou não. Um processo, não importando quão pequeno seja, ocupa uma partição inteira;
- Problema: **fragmentação interna.**
- Usado pelo IBM OS/360.



INSTITUTO FEDERAL
Santa Catarina

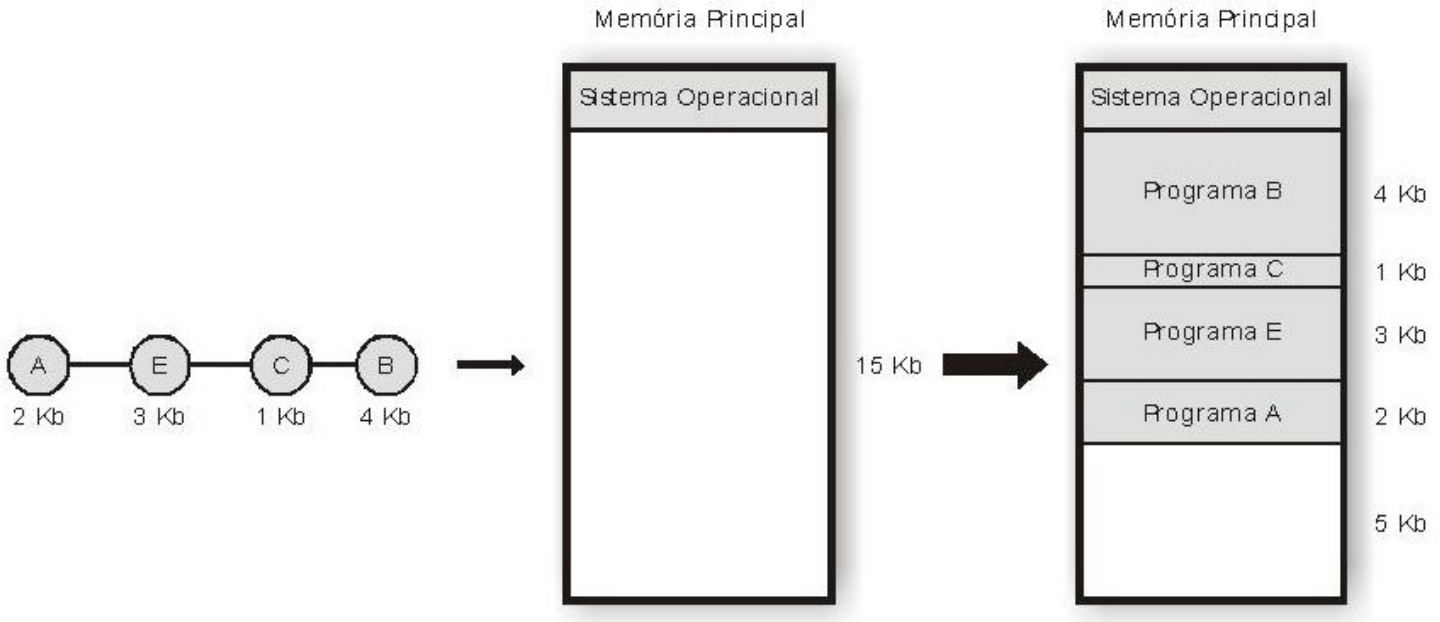
Fragmentação Interna

A fragmentação interna ocorre quando processos não ocupam totalmente o espaço das partições, ou seja, uma parte da memória é perdida devido a memória alocada a um processo ser maior que a solicitada, causando uma diferença de tamanho, e esta diferença ficará inutilizável entre os blocos de memória até que o processo termine.



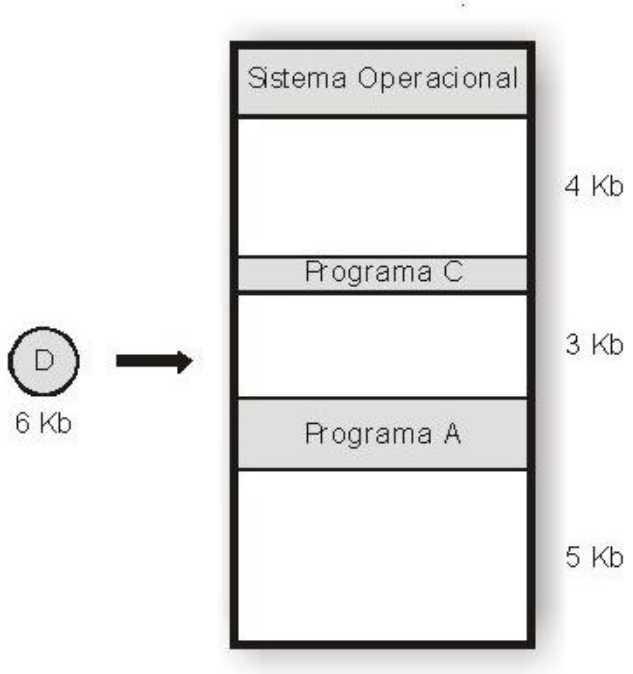
2. Partições Variáveis

- Neste esquema o tamanho da partição é definido de acordo com a demanda do processo (cada processo utiliza apenas o espaço de memória necessário);
- O Sistema gerencia uma tabela indicando partes livres (brechas) e ocupadas;
- Problema: **fragmentação externa.**



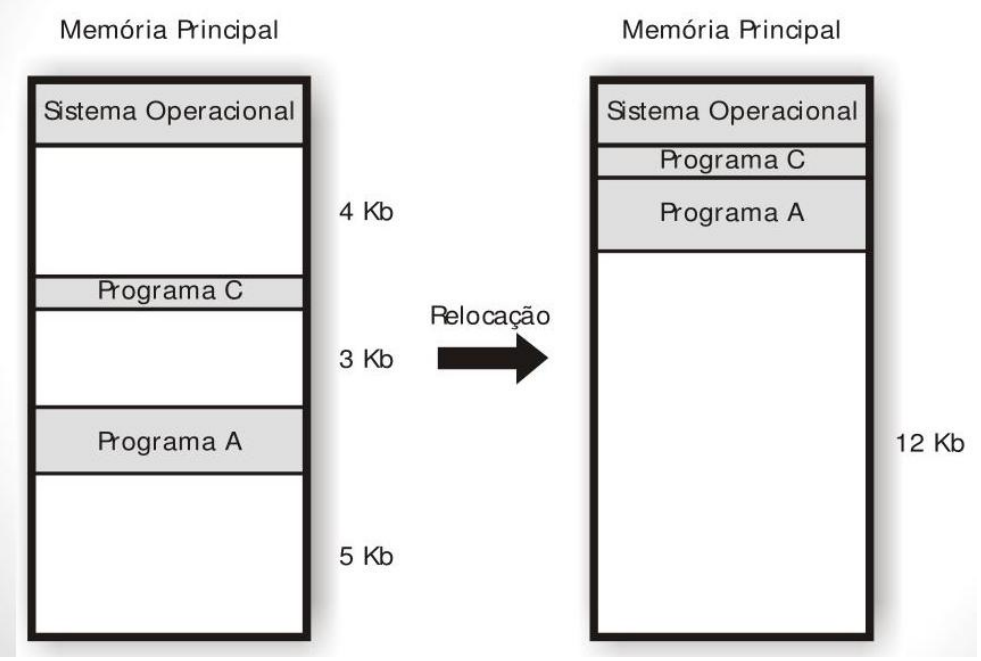
Fragmentação Externa

Término do processo deixa espaços que podem ser insuficientes para novos processos. Uma solução é a desfragmentação, **compactação**, **que é reorganizar a memória livre**. No entanto, a compactação nem sempre é possível, somente se a realocação for dinâmica feita em tempo de execução, mas existe um custo.



Realocação Dinâmica

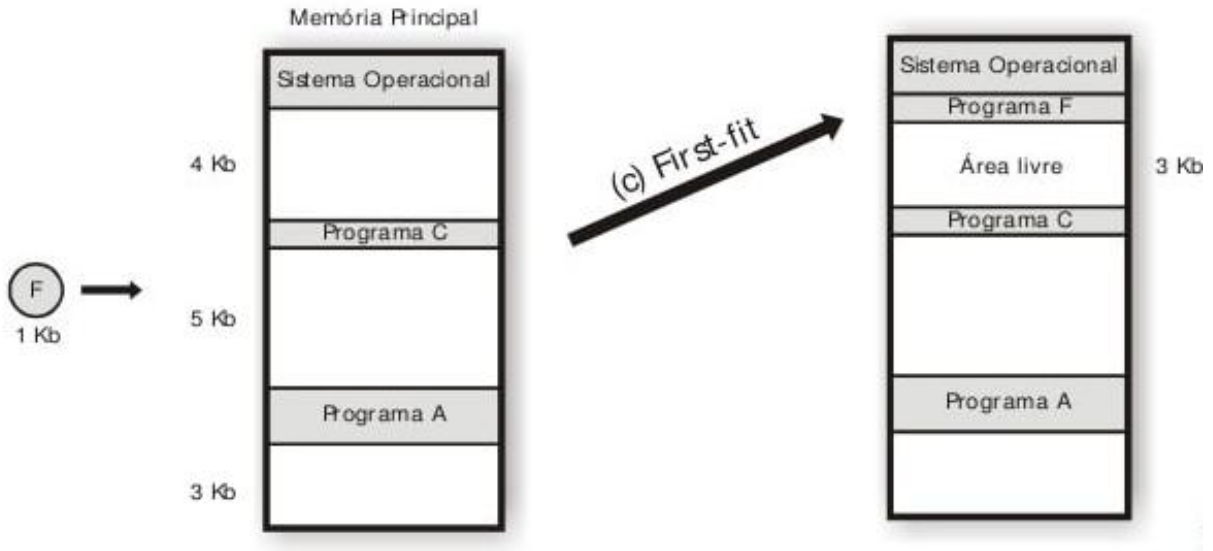
- Reunião dos espaços contíguos.
- Movimentação dos programas pela memória principal;
- Resolve o problema da fragmentação;
- Consome recursos do sistema tais como tempo de processamento, uso de discos, etc.



Estratégias de Alocação

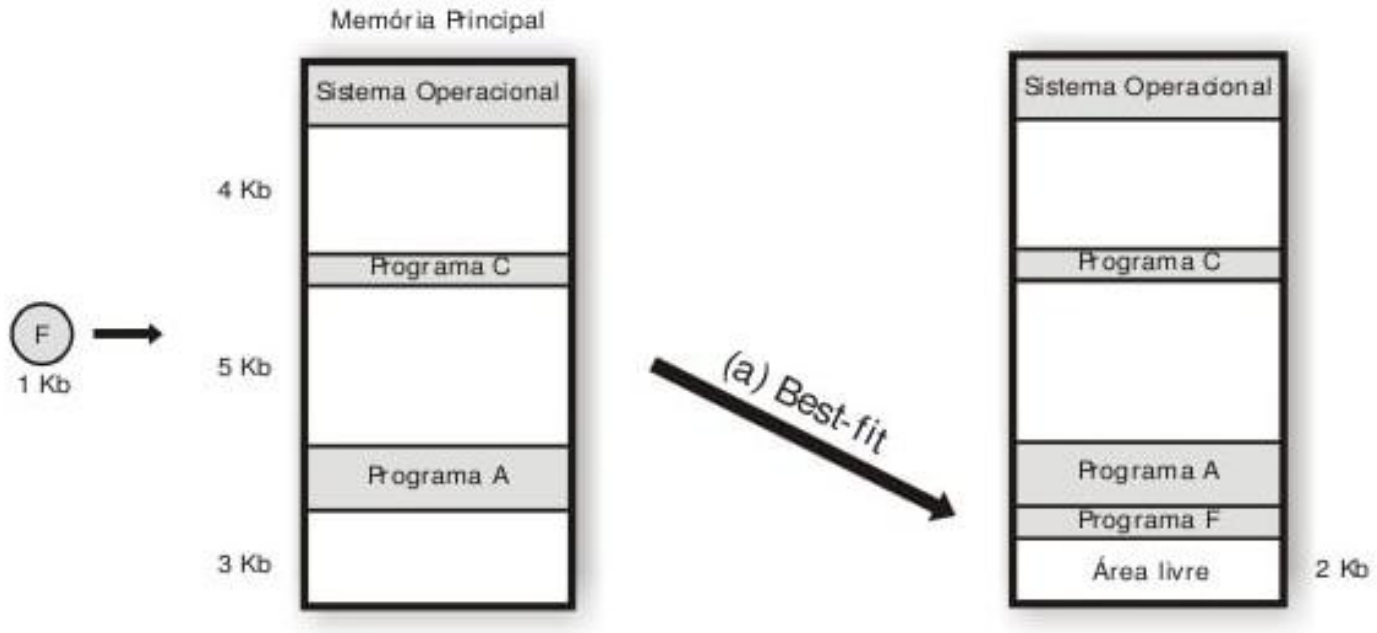
Sempre que um novo processo é criado uma lista de espaços é percorrida e será usada uma lacuna maior ou igual ao tamanho do processo em questão. Basicamente, há 3 estratégias de escolha para determinar em qual área livre um programa será carregado para execução: First-Fit, Best-Fit, Worst-Fit.

First-Fit: Primeiro-apto, Aloca o primeiro bloco grande o suficiente para carregar o programa (Não é preciso percorrer toda a lista de brechas livres), estratégia mais rápida, mas gera fragmentação.



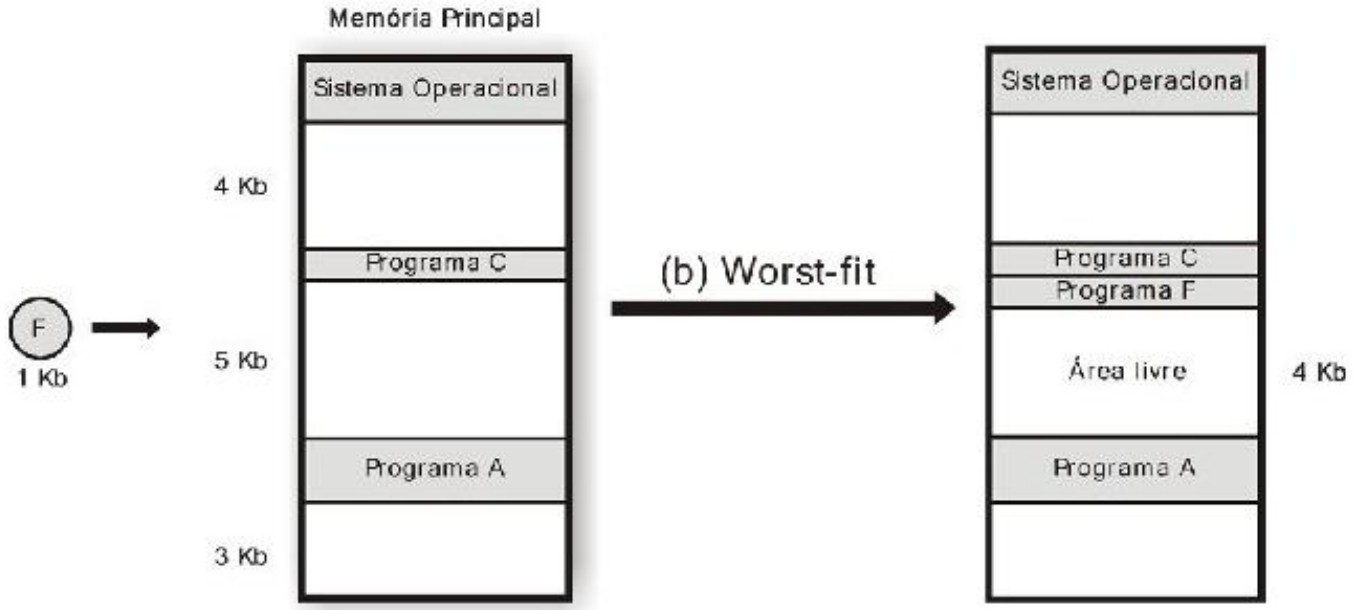
Estratégias de Alocação

Best-Fit: Melhor-apto, Melhor partição, aloca o **menor** bloco grande o suficiente; é preciso procurar na lista inteira, tempo de busca grande. Gera fragmentação.



Estratégias de Alocação

<u>Worst-Fit</u>: Pior-apto, Aloca o **Maior** bloco; é preciso procurar na lista inteira. tempo de busca grande, minimiza a fragmentação externa, não apresenta bons resultados.



INSTITUTO FEDERAL
Santa Catarina

Estratégias de Alocação

- O primeiro-apto (First-Fit) e o o mais-apto (Best-Fit) são melhores em **redução de tempo e utilização de memória**.
- O primeiro-apto (First-Fit) geralmente é mais rápido.
- A estratégia do (First-Fit) e do (Best-Fit) para alocação de memória sofrem de fragmentação externa.
- O problema de fragmentação é grave, podendo resultar em até um terço da memória ficar inutilizável!
- Solução possível: permitir que o espaço de endereçamento lógico dos processos seja **não contíguo**. Duas técnicas são: a segmentação e a paginação.

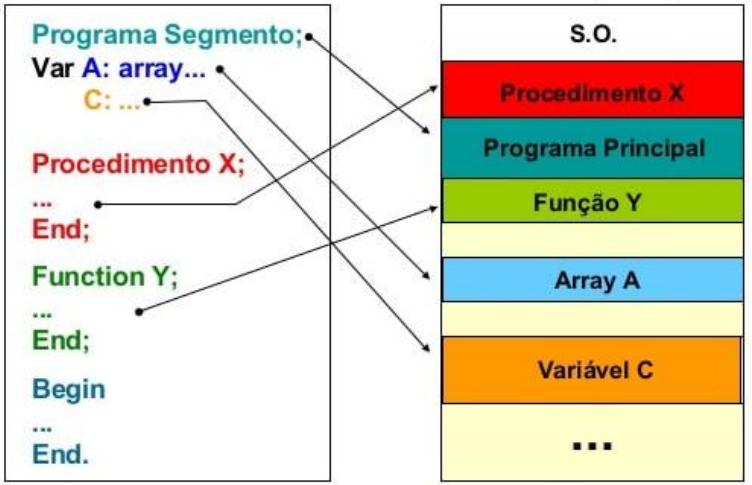
INSTITUTO FEDERAL
Santa Catarina

3. Segmentação

- Esquema de gerenciamento de memória baseado na visão do programador. Um programa é uma coleção de segmentos de tamanho variável como:

  - Pilha
  - Variável local, global
  - Código do programa
  - Estrutura de dados
  - Sub-rotina
  - etc.

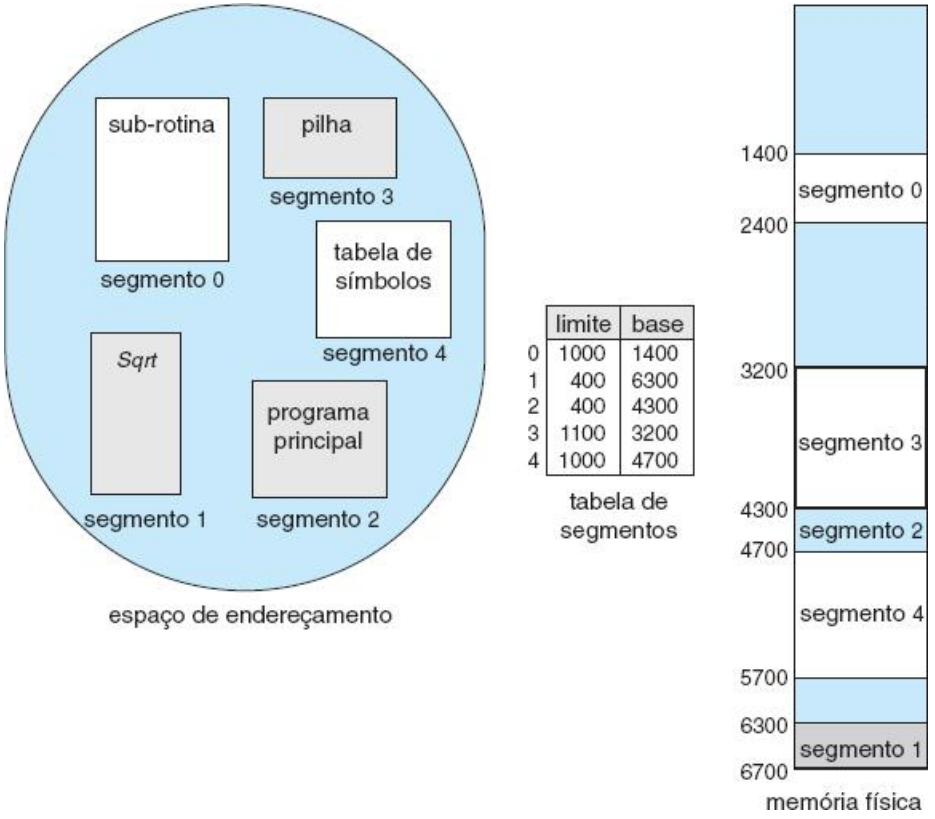
- O programador fala sobre “a pilha”, e “o programa principal” sem se preocupar com os endereços que esses elementos ocupam na memória. (pedaços do programa espalhados pela memória, segmentos ocupam páginas)



INSTITUTO FEDERAL
Santa Catarina

3. Segmentação

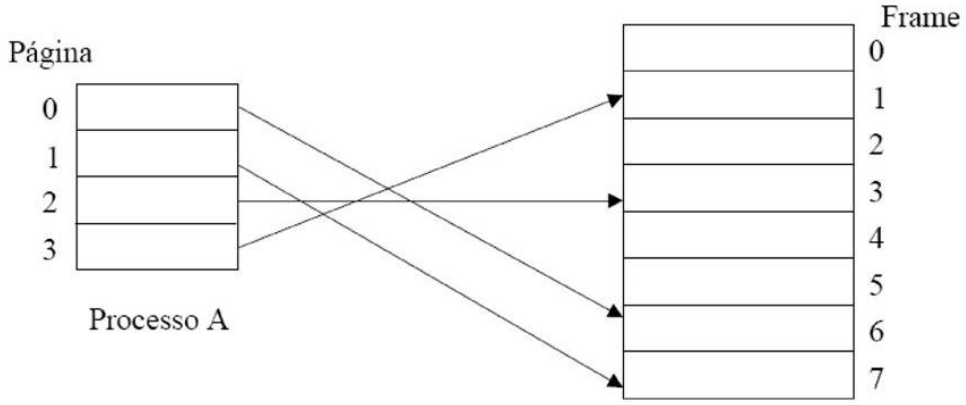
- Cada segmento tem um nome e um tamanho. O programador, então, especifica cada endereço com dois valores: um nome de segmento e um deslocamento.
- O Sistema faz o mapeamento por meio de uma **tabela de segmentos**. Cada entrada na tabela de segmentos tem uma **base** e um **limite de segmento**. A base do segmento contém o endereço físico inicial em que o segmento reside na memória, e o limite do segmento especifica o tamanho do segmento



INSTITUTO FEDERAL
Santa Catarina

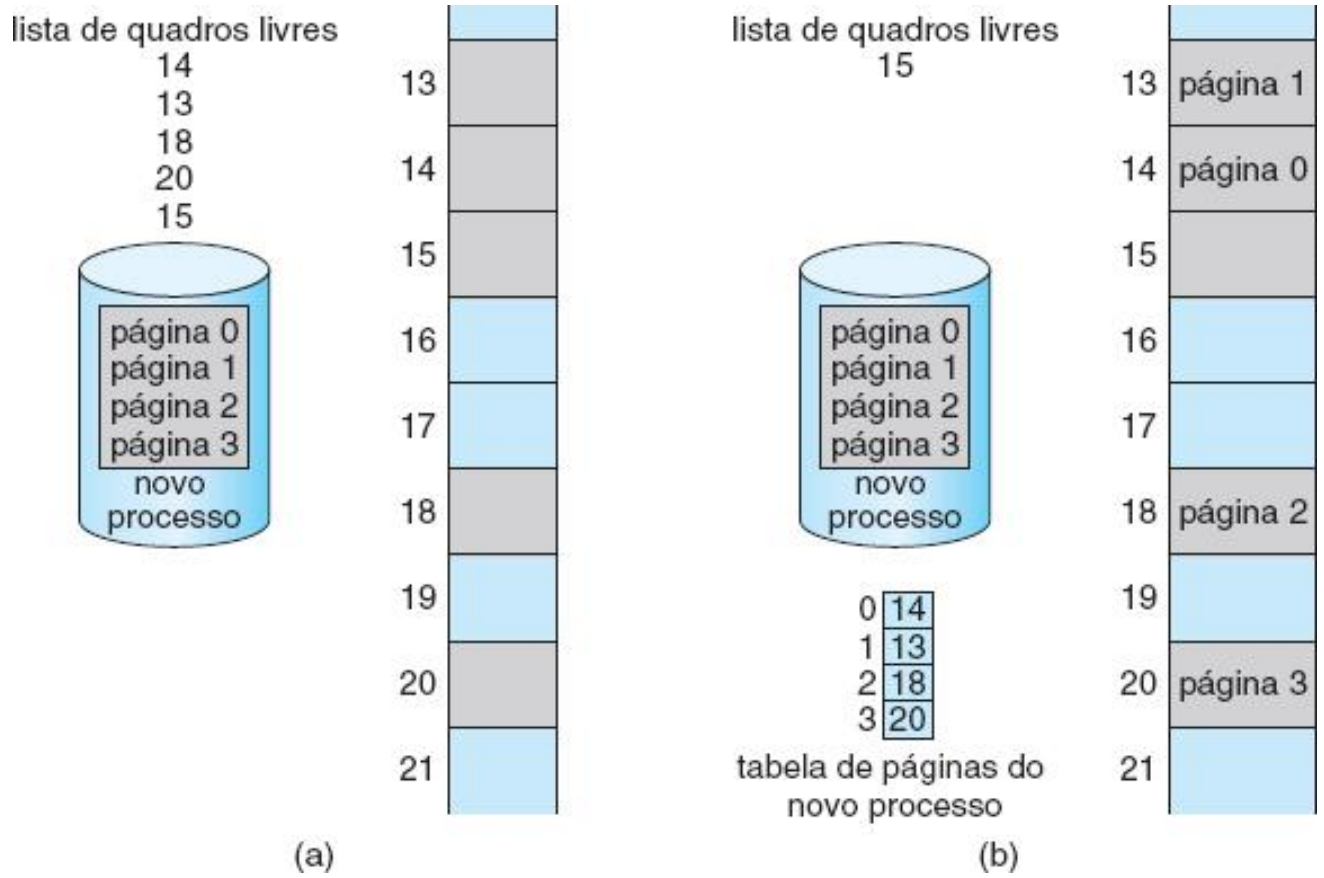
4. Paginação

- Alternativamente, o espaço de alocação pode ser não-contiguo, o processo é alocado na memória física onde houver blocos livres.
- A memória física é dividida em blocos de tamanho fixo chamados quadros (frames)
- A memória lógica é dividida em blocos de mesmo tamanho chamado páginas (pages)
  - Normalmente potência de 2, entre 512 bytes e 16MB por página.
- É usada uma tabela de páginas para traduzir endereços lógicos em físicos.
- Evita a fragmentação externa e a necessidade de compactação, mas pode ter alguma fragmentação interna.



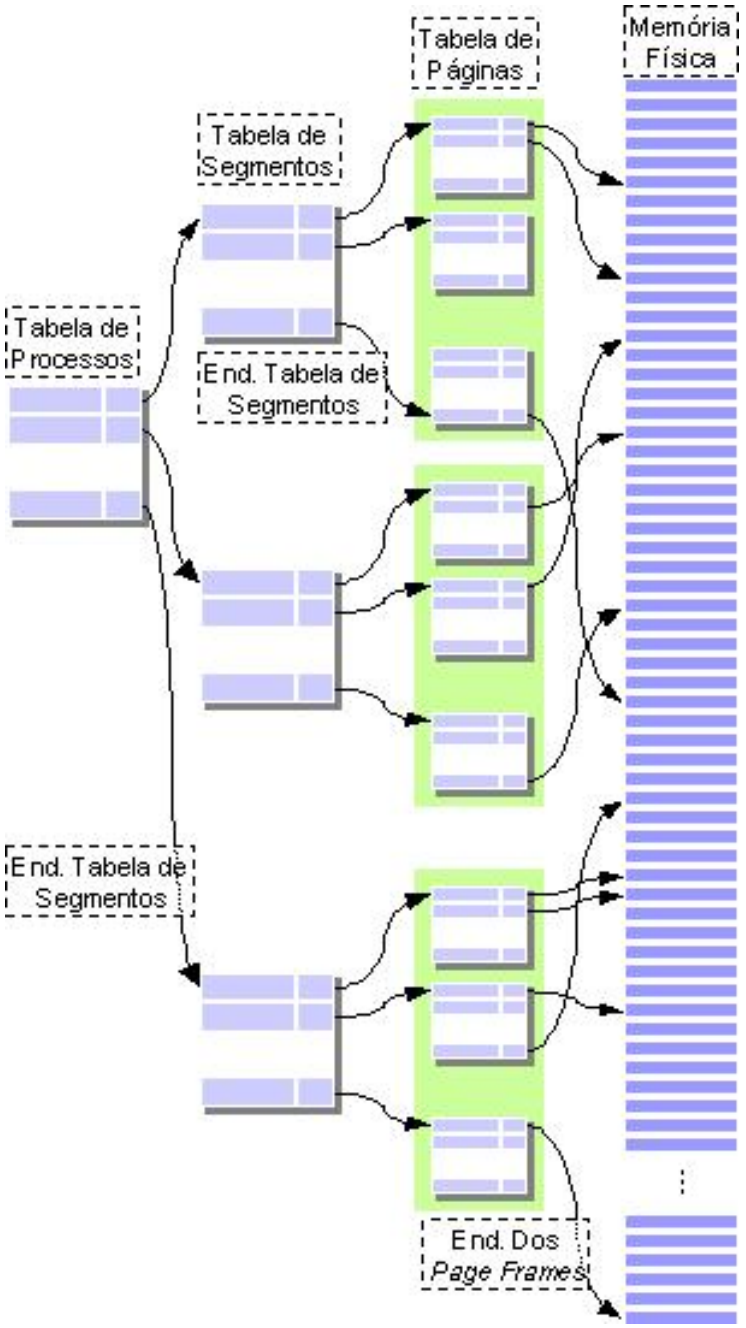
INSTITUTO FEDERAL
Santa Catarina

4. Paginação



INSTITUTO FEDERAL
Santa Catarina

S.O Modernos

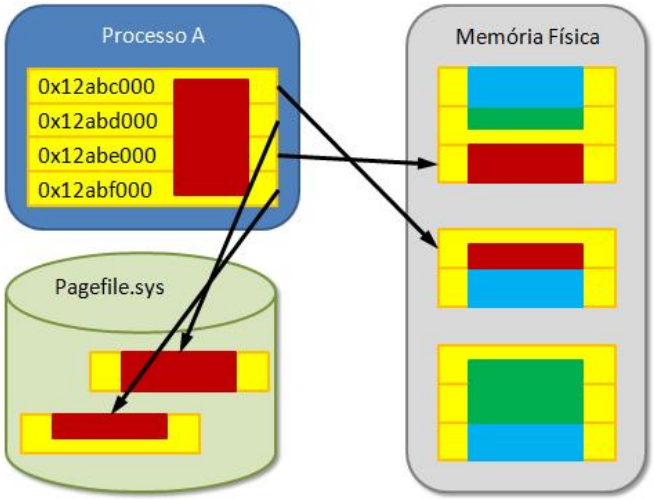


INSTITUTO FEDERAL
Santa Catarina

Memória Virtual

Técnica de gerência de memória que combina a memória principal (RAM) e secundária (Disco)

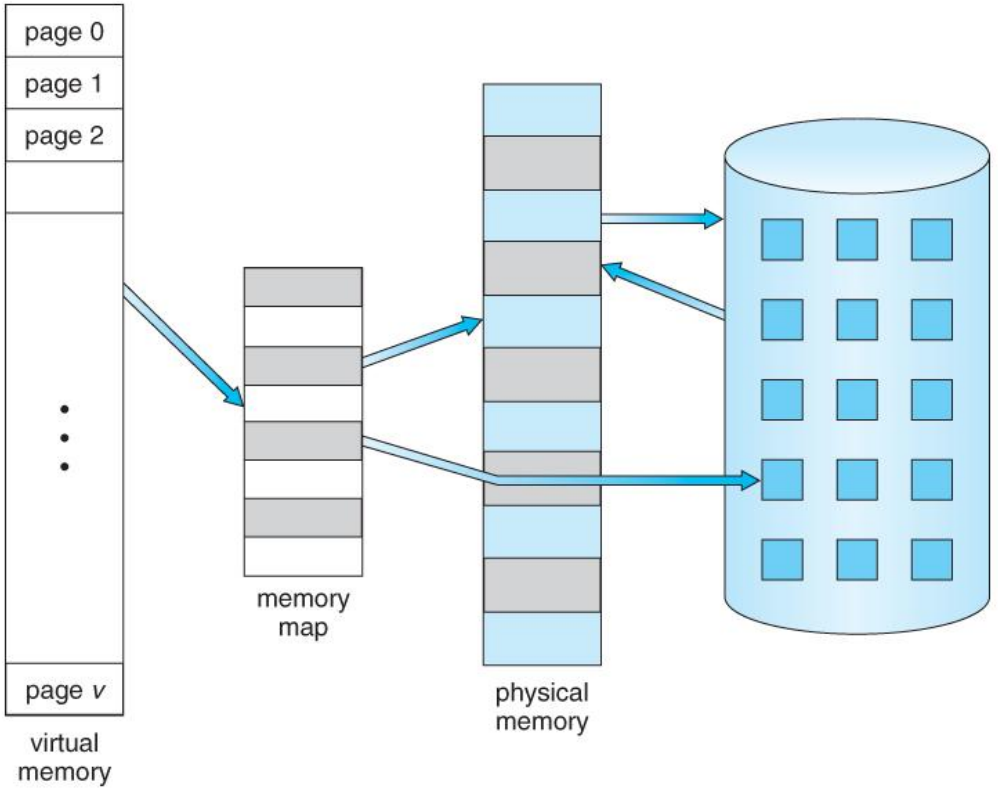
- Permite o compartilhamento seguro e eficiente da memória entre vários programas;
- Remove transtornos de programação de uma quantidade pequena e limitada de memória;
- Os programas podem ser maiores que a memória física disponível;
- Somente parte do programa pode estar na memória física para execução;
- Reduz problemas de fragmentação de memória;
- Evita colapsos do S.O entretanto deixa lento.



INSTITUTO FEDERAL
Santa Catarina

Memória Virtual

Se uma página solicitada não estiver carregada na memória física, ela deve ser trazida do disco rígido (e possivelmente outra página já utilizada será levada de volta ao disco rígido)



INSTITUTO FEDERAL
Santa Catarina

Memória Virtual

A Memória Virtual pode ser implementada por:

- Paginação sob demanda (mais comum)
- Segmentação sob demanda (mais complexo)

INSTITUTO FEDERAL
Santa Catarina

Paginação sob demanda

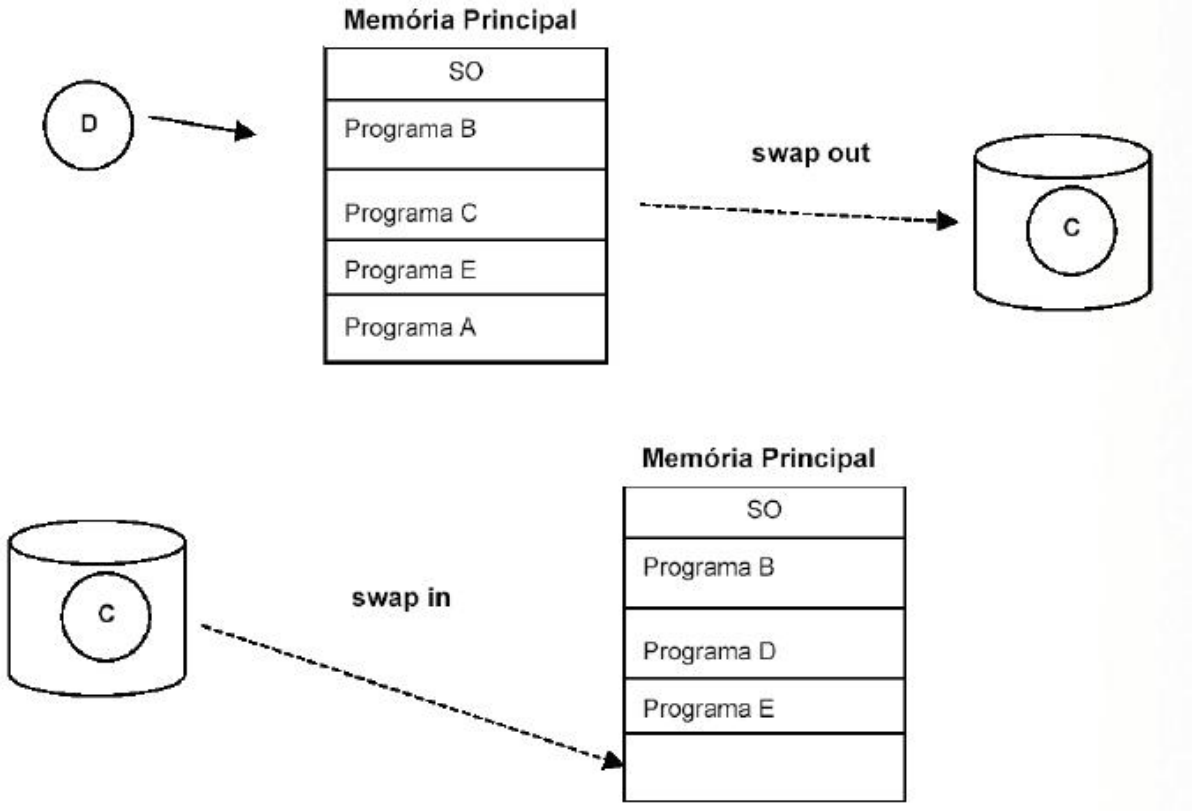
Um programa pode ser carregado totalmente na memória, mas muitas vezes não é preciso dele inteiro inicialmente para execução. Carrega uma página na memória apenas quando é necessária.

- Necessita menos E/S;
- Necessita menos memória física;
- Possibilita um maior número de processos na memória;
- Se uma página é necessária basta referenciá-la;
- Se uma página não está na memória basta carrega-la na memória.

INSTITUTO FEDERAL
Santa Catarina

Swapping

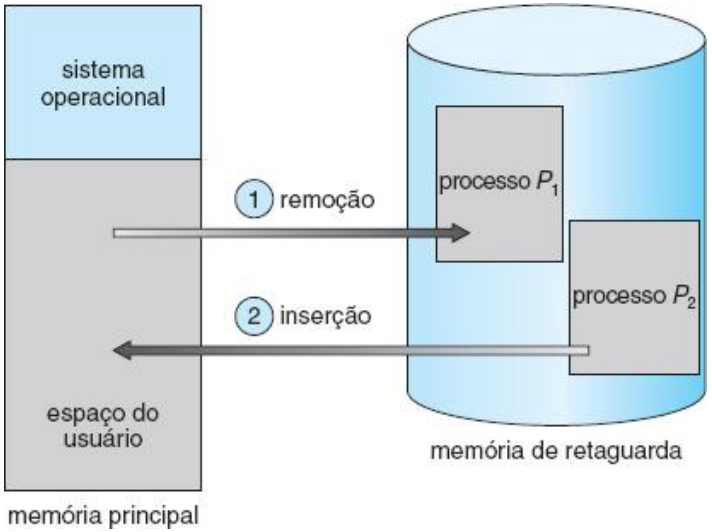
É o chaveamento de processos entre a memória e o disco. Introduzida para contornar o problema de insuficiência de memória principal. Técnica aplicada a gerência de memória para programas que esperam por memória livre para serem executados. **Swap-out:** da memória para o disco em uma área de “swap”; **Swap-in:** do disco para a memória.



INSTITUTO FEDERAL
Santa Catarina

Swapping

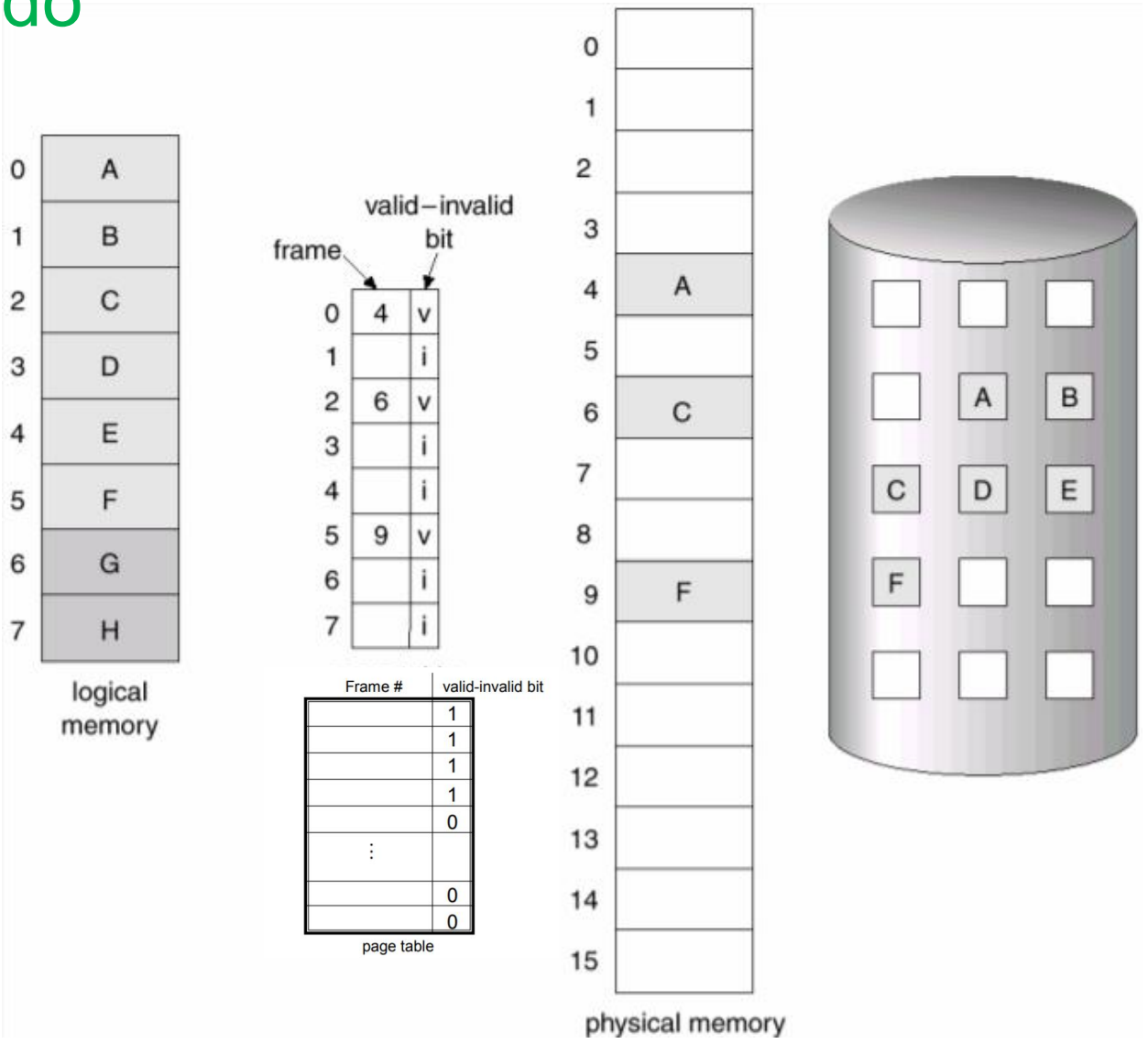
O tempo de mudança de contexto nesse sistema de permuta **é bem alto**. Para termos uma ideia desse tempo, suponha que o processo do usuário tenha um tamanho de 100 MB e a memória de retaguarda seja um disco rígido padrão com taxa de transferência de 50 MB por segundo. A transferência real do processo de 100 MB em uma das transferências (para dentro ou para fora) da memória principal leva: **100 MB / 50 MB por segundo = 2 segundos**



INSTITUTO FEDERAL
Santa Catarina

Bit Válido-Inválido

Bit inválido igual a 0 quando estiver sendo realizada uma tradução de endereços falta de página (page fault)



INSTITUTO FEDERAL
Santa Catarina

Substituição de Página

Quando não há quadros livres na memória é necessário substituir páginas que não estão sendo usadas e armazená-las no disco.

- Uma mesma página pode ser carregada/retirada da memória várias vezes durante a execução do processo.
- **Qual processo retirar?** Algoritmo de substituição de páginas.

Algoritimo First-in-First-Out (FIFO)

- Verifica o tempo em que a página foi trazida na memória, quando é necessária a substituição de página a mais antiga é utilizada.
- Mantém uma Fila com todas as páginas trazidas.
- Há situações que quanto mais quadro, pode gerar mais falhas (Anomalia de Belady)
- Por exemplo: 3 páginas de memória, string de referência: 1,2,3,4,1,2,5,1,2,3,4,5

1	**1**	**4**			**5**						9 falhas de página
2	**2**		**1**			**3**					
3	**3**			**2**			**4**				

INSTITUTO FEDERAL
Santa Catarina

Algoritimo Ótimo

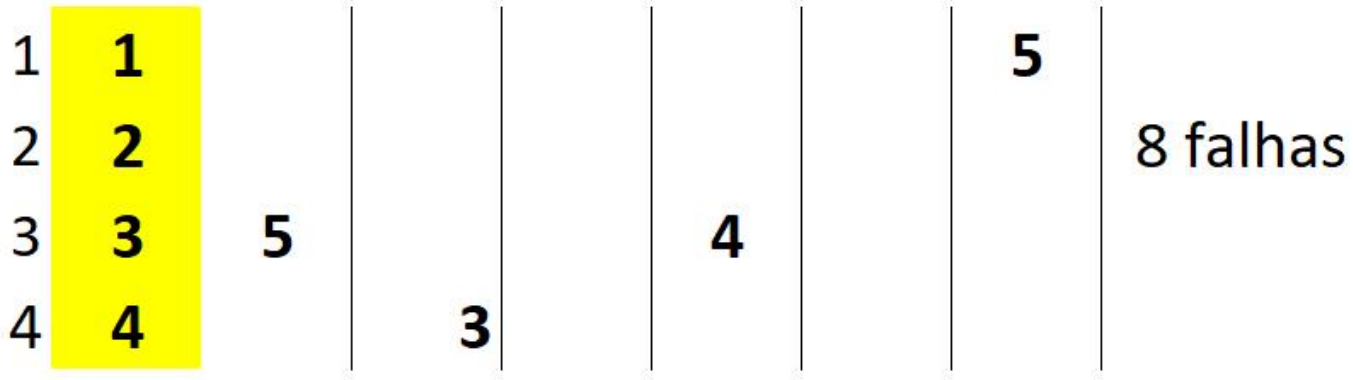
- Substituir a página que não será utilizada pelo período mais longo.
- Por exemplo: 3 páginas de memória, string de referência: 1,2,3,4,1,2,5,1,2,3,4,5
- Reduz o dobro de falhas em relação ao FIFO.

1	**1**		**4**		
2	**2**				6 falhas
3	**3**				
4	**4**	**5**			

INSTITUTO FEDERAL
Santa Catarina

Algoritimo Least Recently Used (LRU)

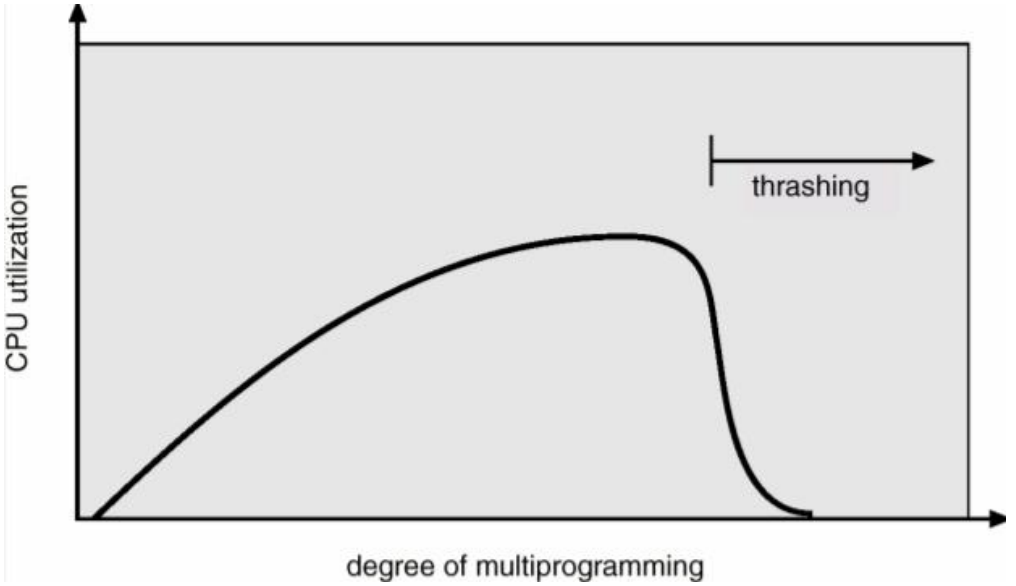
- Página menos usada recentemente
- Substituir a página que não foi utilizada pelo maior período de tempo passado.
- Por exemplo: 3 páginas de memória, string de referência: 1,2,3,4,1,2,5,1,2,3,4,5



Thrashing

Quando o processo gasta mais tempo paginando do que executando.

- Se um processo não tem quadros suficientes a taxa de falha é muito alta resultando baixa utilização da CPU (processos suspensos por paginação)
- Resulta em graves problemas de desempenho



INSTITUTO FEDERAL
Santa Catarina

Bibliografia

SILBERSCHATZ, A. Fundamentos de Sistemas Operacionais 9ª edição. Grupo GEN, 2015.

TANENBAUM, A. S. Sistemas Operacionais Modernos: 2ª edição, São Paulo, editora Prentice Hall, 2003..

MACHADO, F. B. & MAIA, L. P., Arquitetura de Sistemas Operacionais, 4 Edição, São Paulo, LTC, 2007

INSTITUTO FEDERAL
SANTA CATARINA
Câmpus Garopaba

Obrigado!