

Sistemas Operacionais

Entrada e Saída

Drivers e Hardware de Discos

Norton Trevisan Roman
Marcelo Morandini
Jó Ueyama

Apostila baseada nos trabalhos de Kalinka Castelo Branco, Antônio Carlos Sementille, Luciana A. F. Martimiano e nas transparências fornecidas no site de compra do livro "Sistemas Operacionais Modernos"

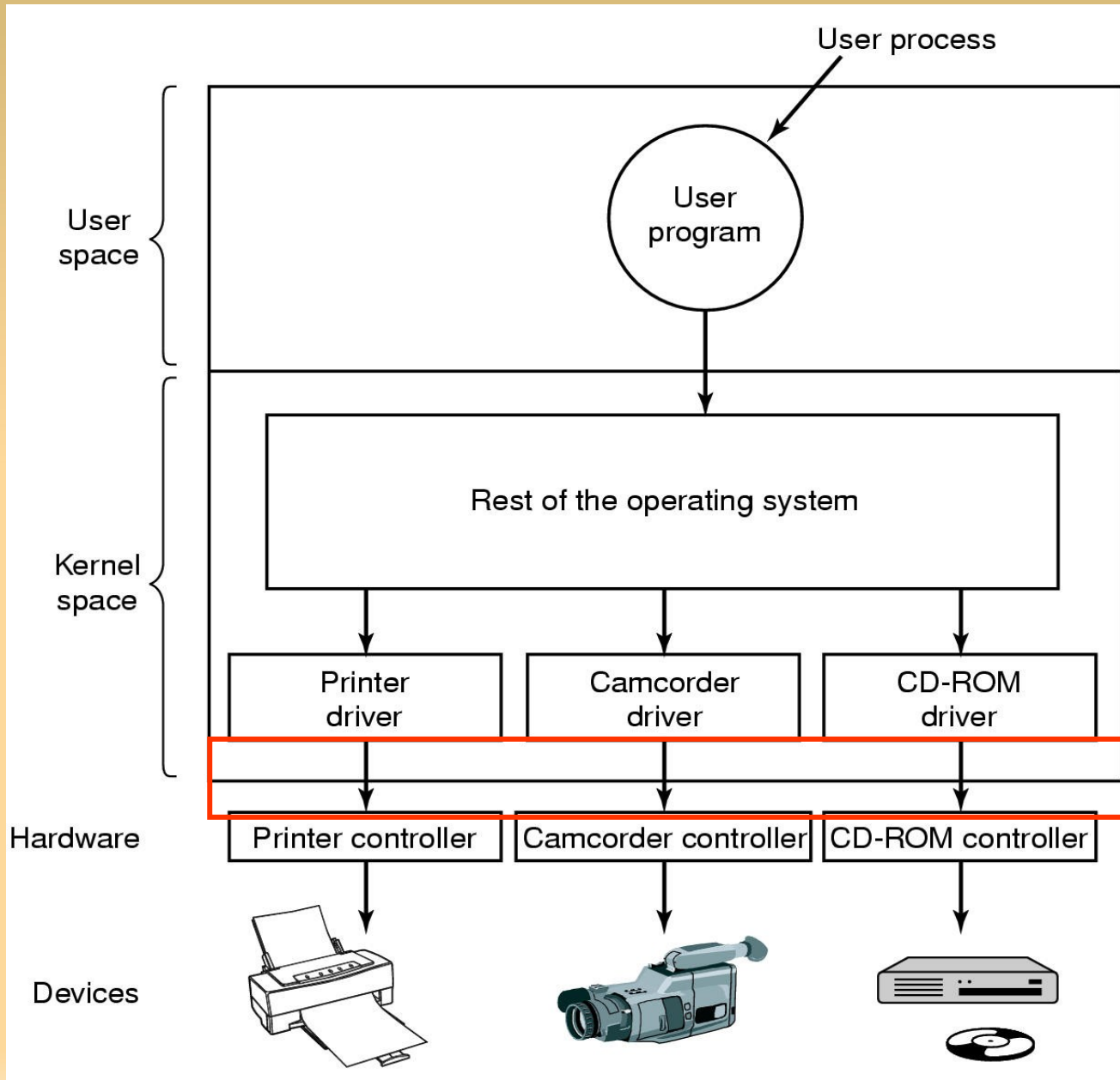
Drivers

- Cada controlador de dispositivo possui registradores
 - Usados para comunicação com a CPU
 - Vimos isso
- O número de registradores, e sua função, varia conforme o hardware
 - Cada dispositivo de E/S precisa de algum código específico para controlá-lo → o driver de dispositivo
 - Contêm todo o código dependente do dispositivo;
 - Dispositivos diferentes possuem drivers diferentes;
 - Classes de dispositivos podem ter o mesmo driver;

Drivers

- São geralmente escritos pelo fabricante do dispositivo
 - SOs diferentes precisam de drivers diferentes
- Fazem parte do kernel do SO
 - Permitindo acesso aos registradores do controlador de dispositivo
 - Controlam o funcionamento dos dispositivos por meio de seqüência de comandos escritos/lidos nos/dos registradores da controladora;
 - Problema;
 - Drivers defeituosos podem causar problemas no kernel do SO;

Drivers



Comunicação feita por barramento;

Drivers

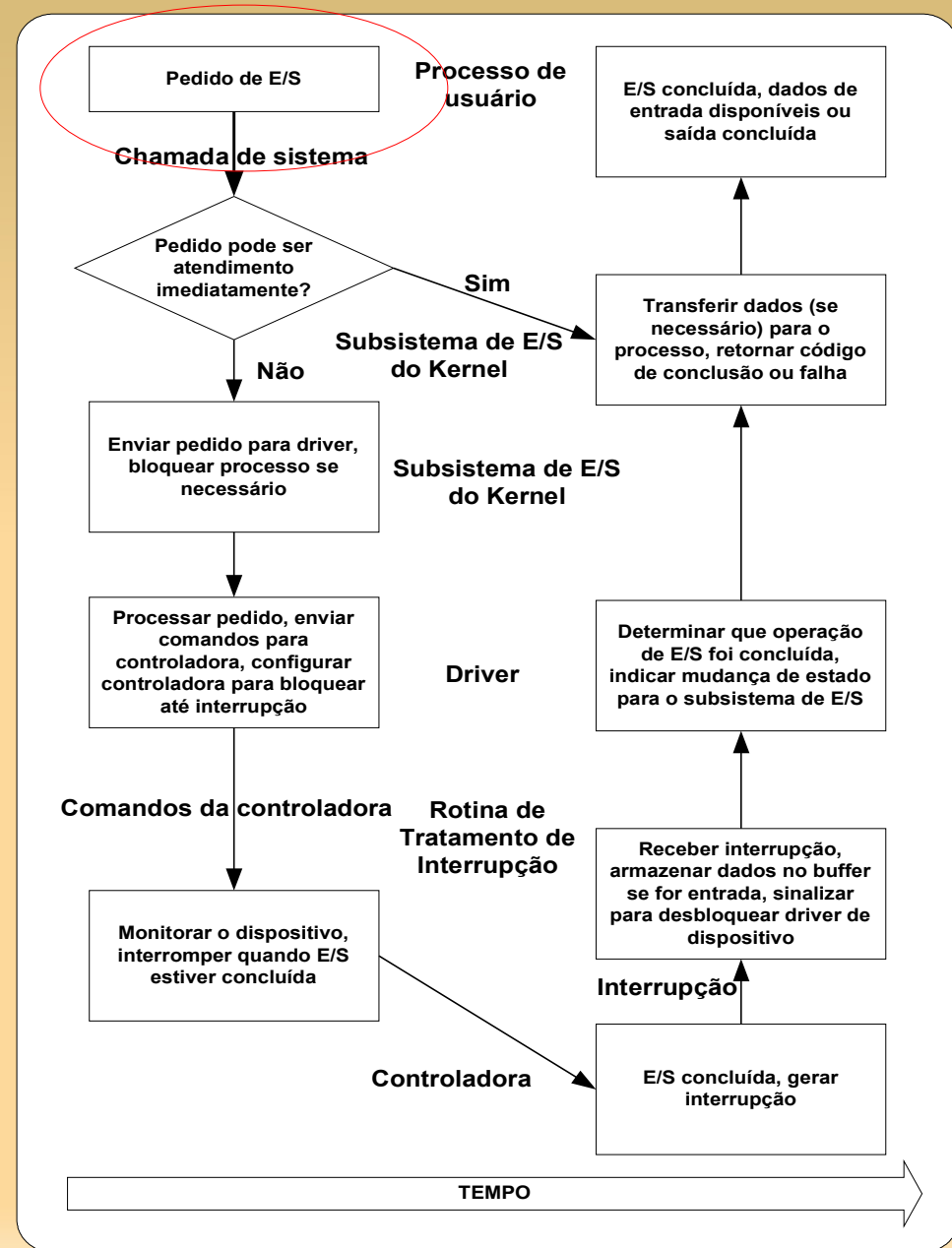
- A parte do SO:
 - Definir um modelo do que um driver deve fazer e como deve interagir com o resto do SO
 - Em geral, definem uma interface padrão para:
 - Drivers de dispositivos de bloco
 - Drivers de dispositivos de caracteres
 - Cada interface contém os procedimentos que o resto do SO pode chamar para fazer para usar o driver
 - Ex; ler um bloco, escrever um caractere
 - Carregar os drivers dinamicamente, durante a execução

Drivers

- Funções
 - Aceitar pedidos de leitura/escrita do software independente de dispositivos e cuidar que sejam executadas
 - Inicializar o dispositivo, se necessário
 - Gerenciar as necessidades energéticas do dispositivo
 - Criar um log de eventos

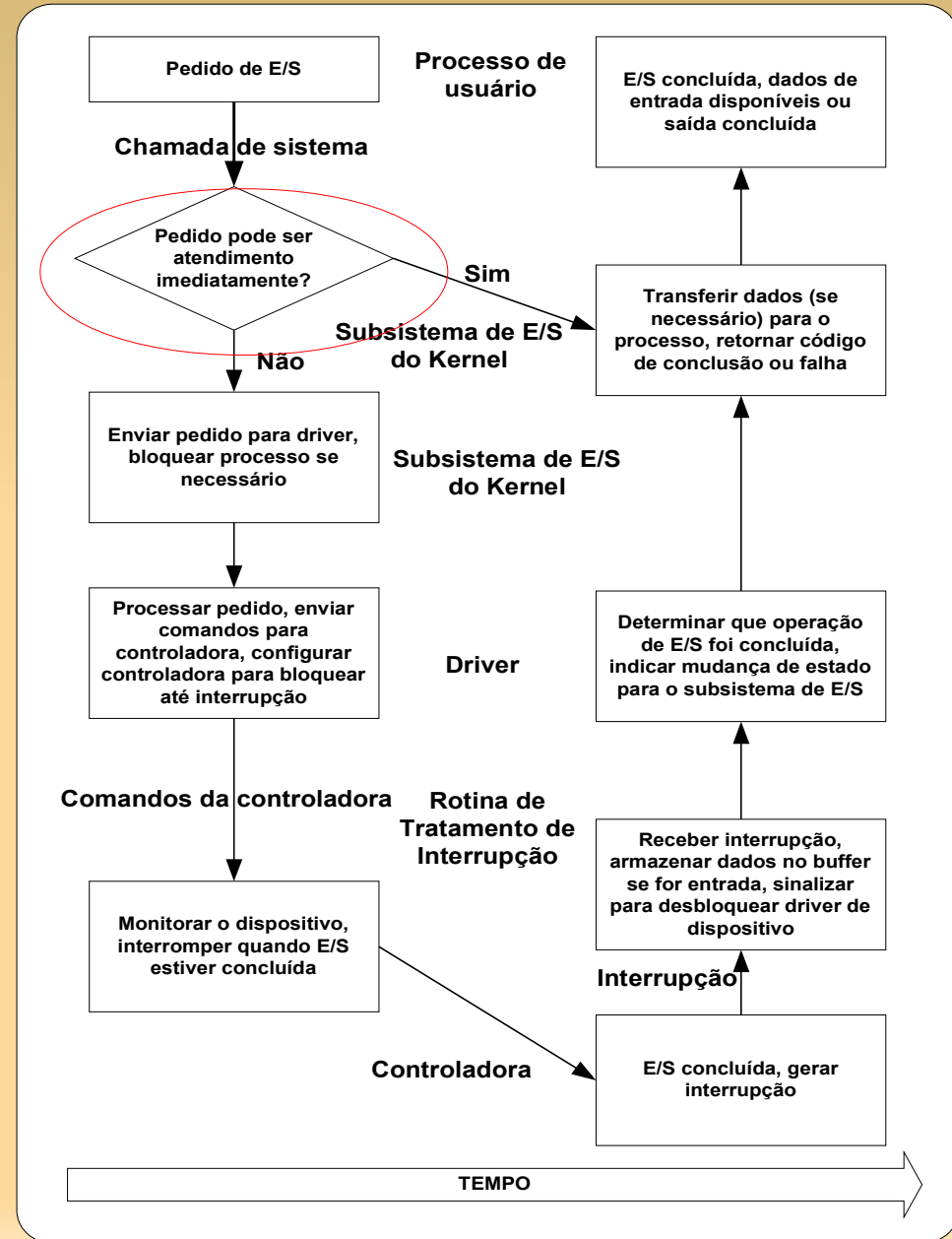
Drivers

- Funcionamento:
 - Um processo emite uma chamada de sistema bloqueante (ex: read) para um arquivo que já está aberto (open);



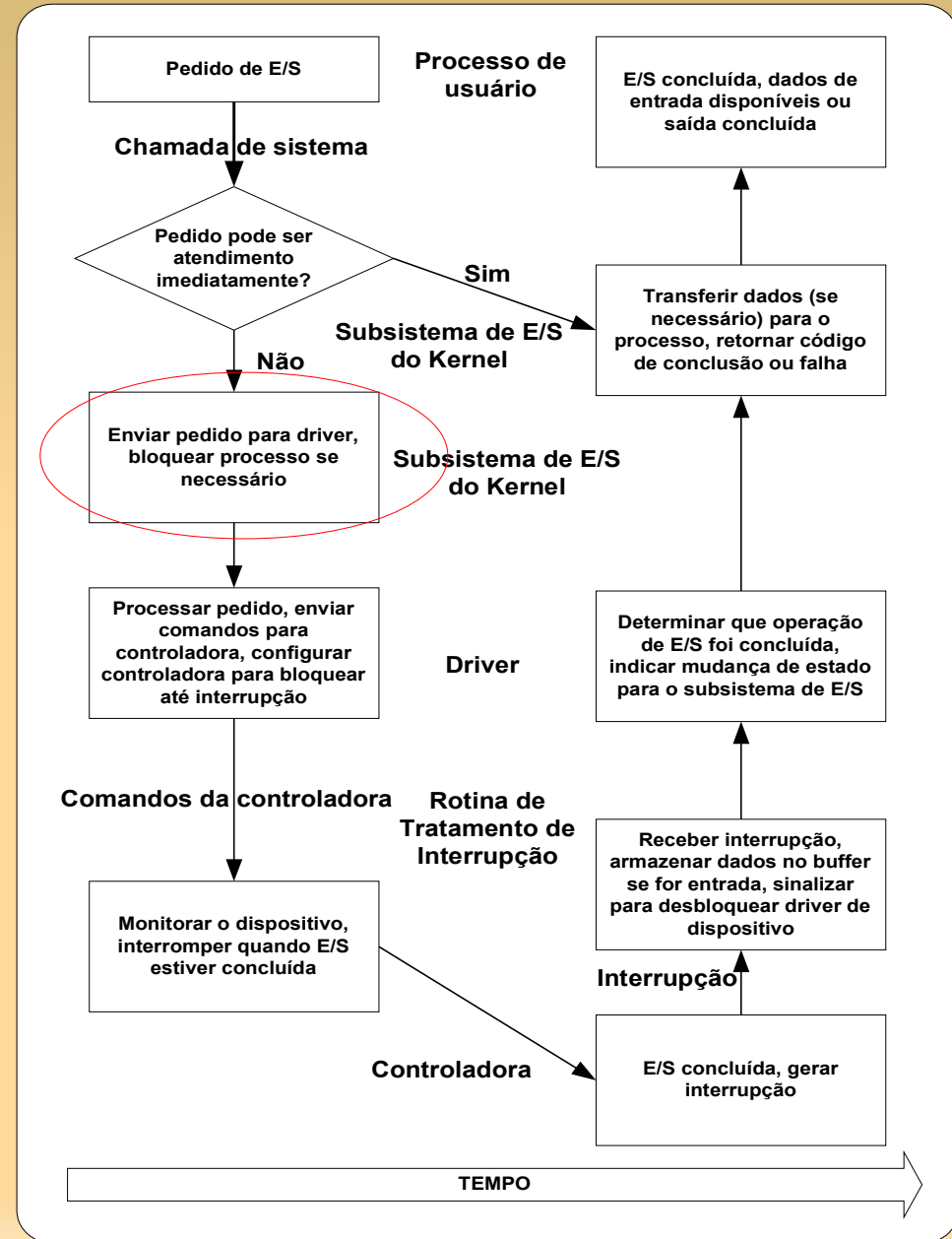
Drivers

- Funcionamento:
 - O código da chamada de sistema verifica os parâmetros. Se os parâmetros estiverem corretos e o arquivo já estiver no buffer (cache), os dados retornam ao processo e a E/S é concluída;



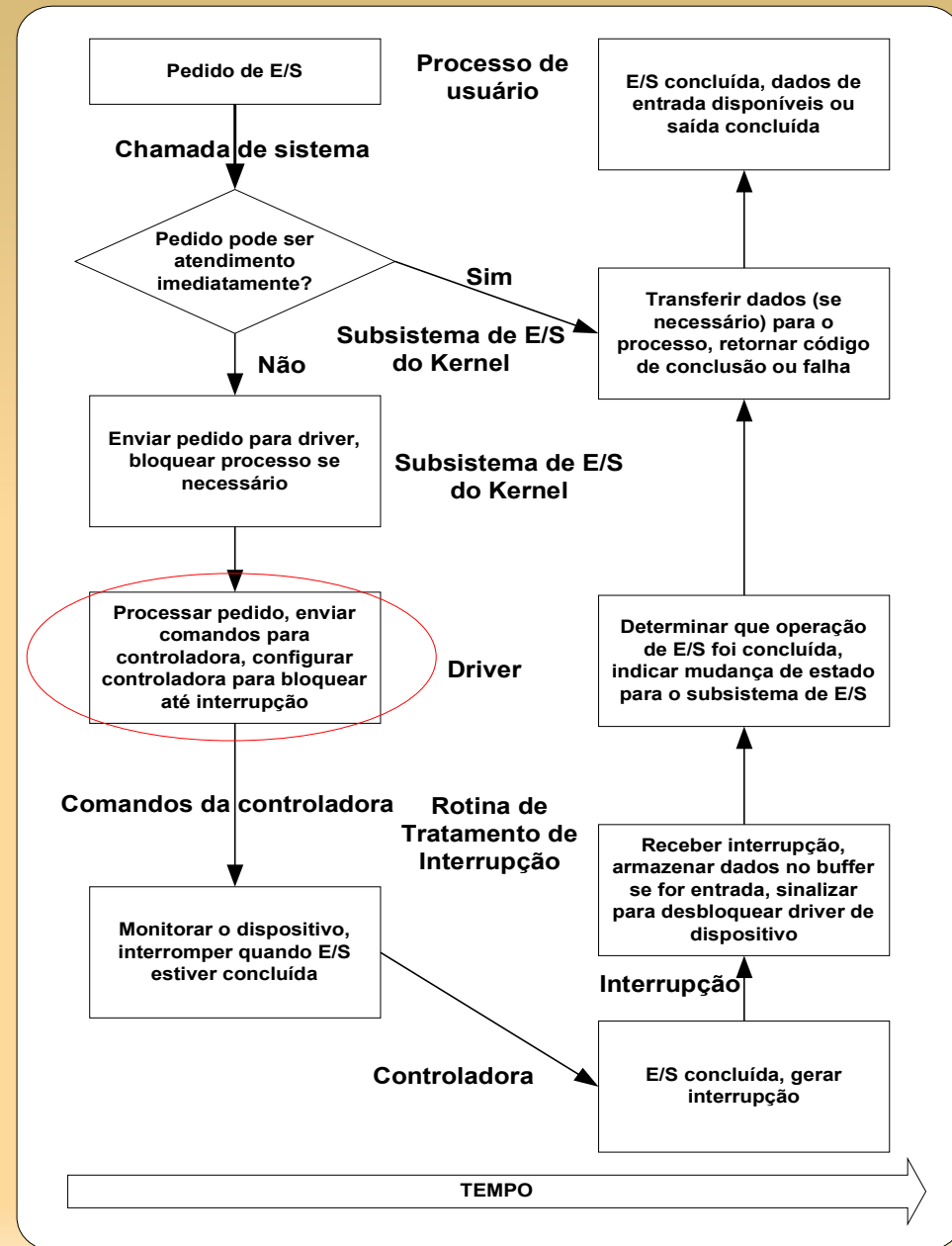
Drivers

- Funcionamento:
 - Se os parâmetros estiverem corretos, mas o arquivo não estiver no buffer, a E/S precisa ser realizada;
 - E/S é escalonada;
 - Subsistema envia pedido para o driver;
 - Se os parâmetros estiverem incorretos, um erro é retornado



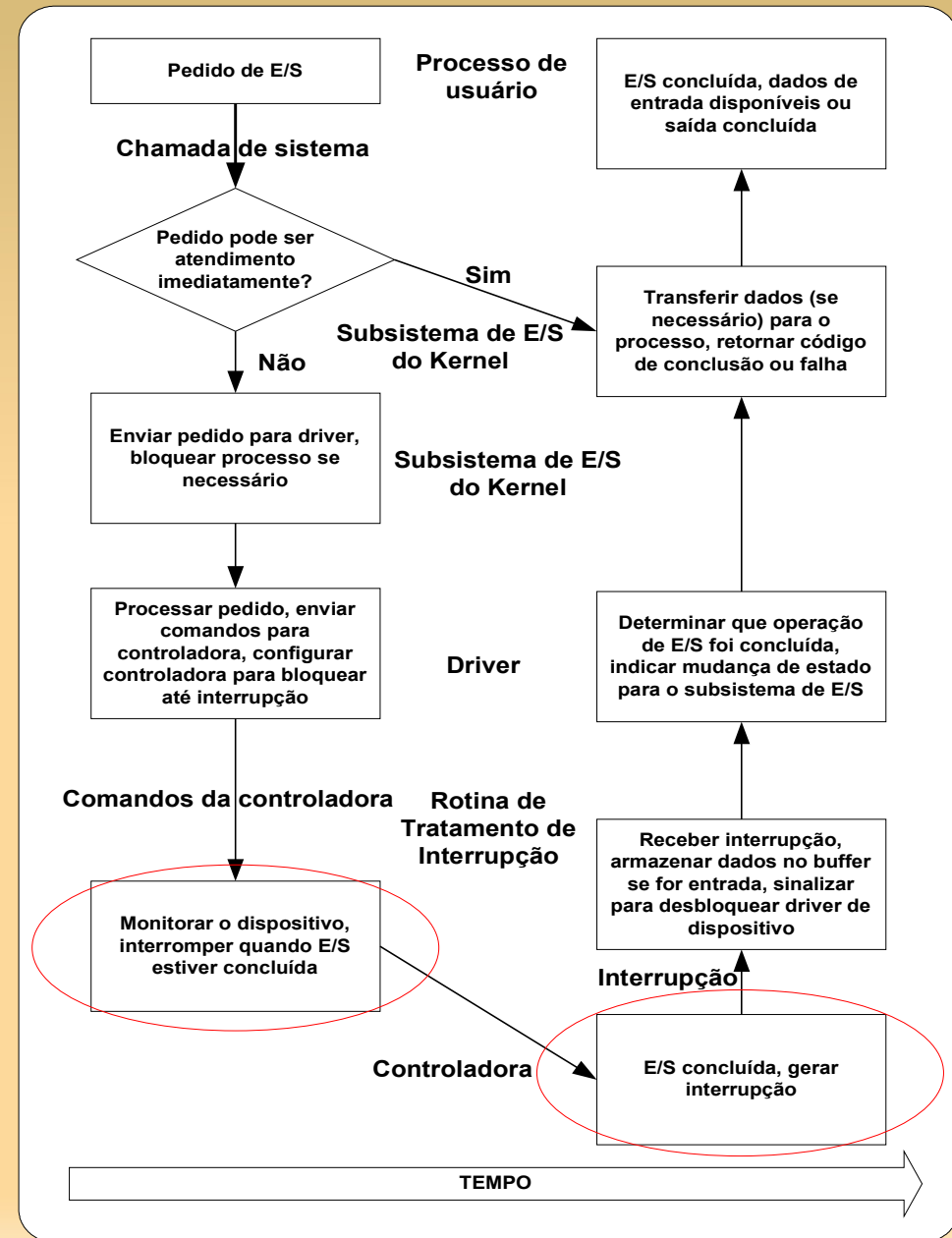
Drivers

- Funcionamento:
 - O Driver aloca espaço de buffer, escalona E/S e envia comando para a controladora do dispositivo escrevendo nos seus registradores de controle;
 - Driver pode usar a DMA;



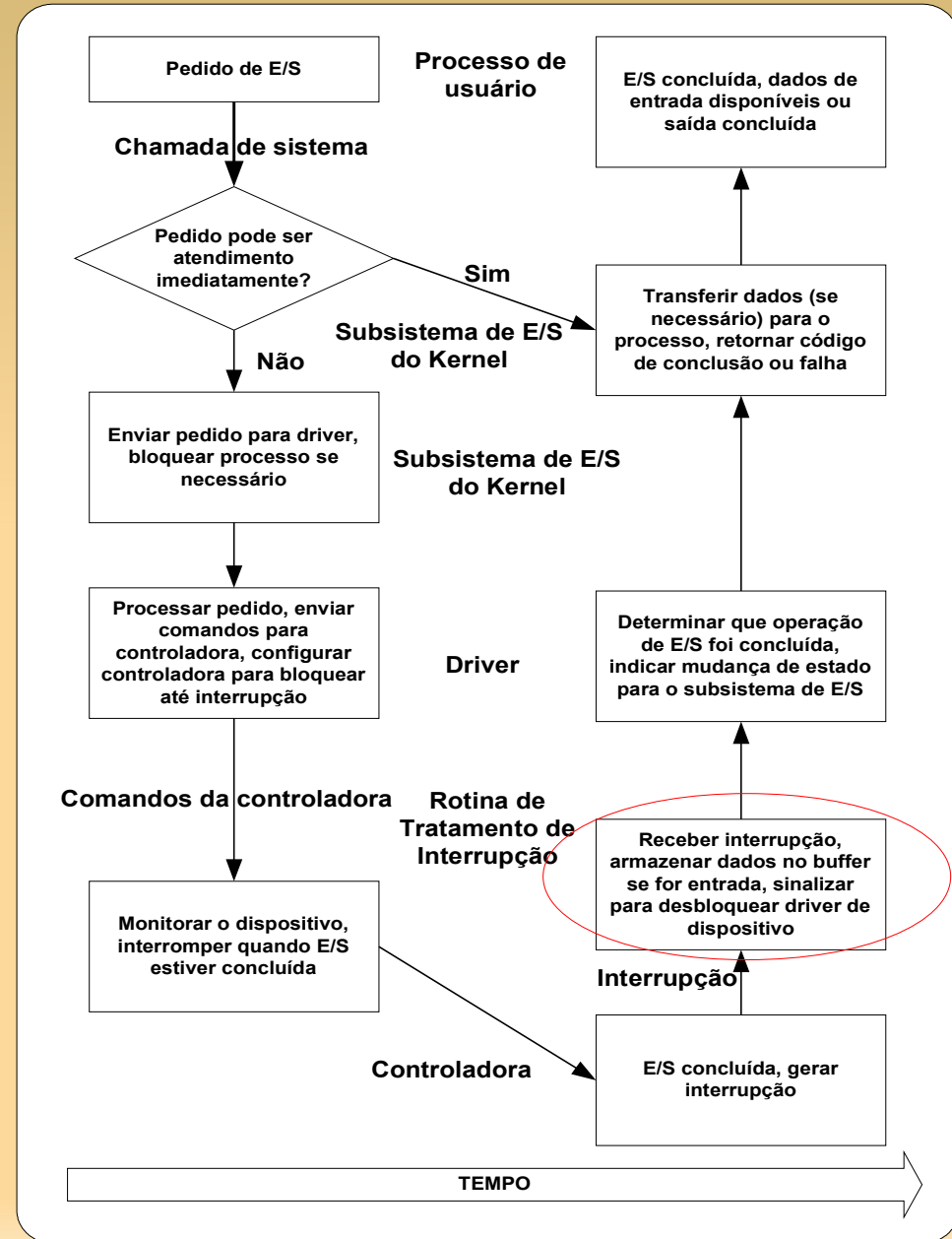
Drivers

- Funcionamento:
 - A controladora do dispositivo opera o hardware, ou seja, o dispositivo propriamente dito;
 - Após a conclusão da E/S, uma interrupção é gerada;



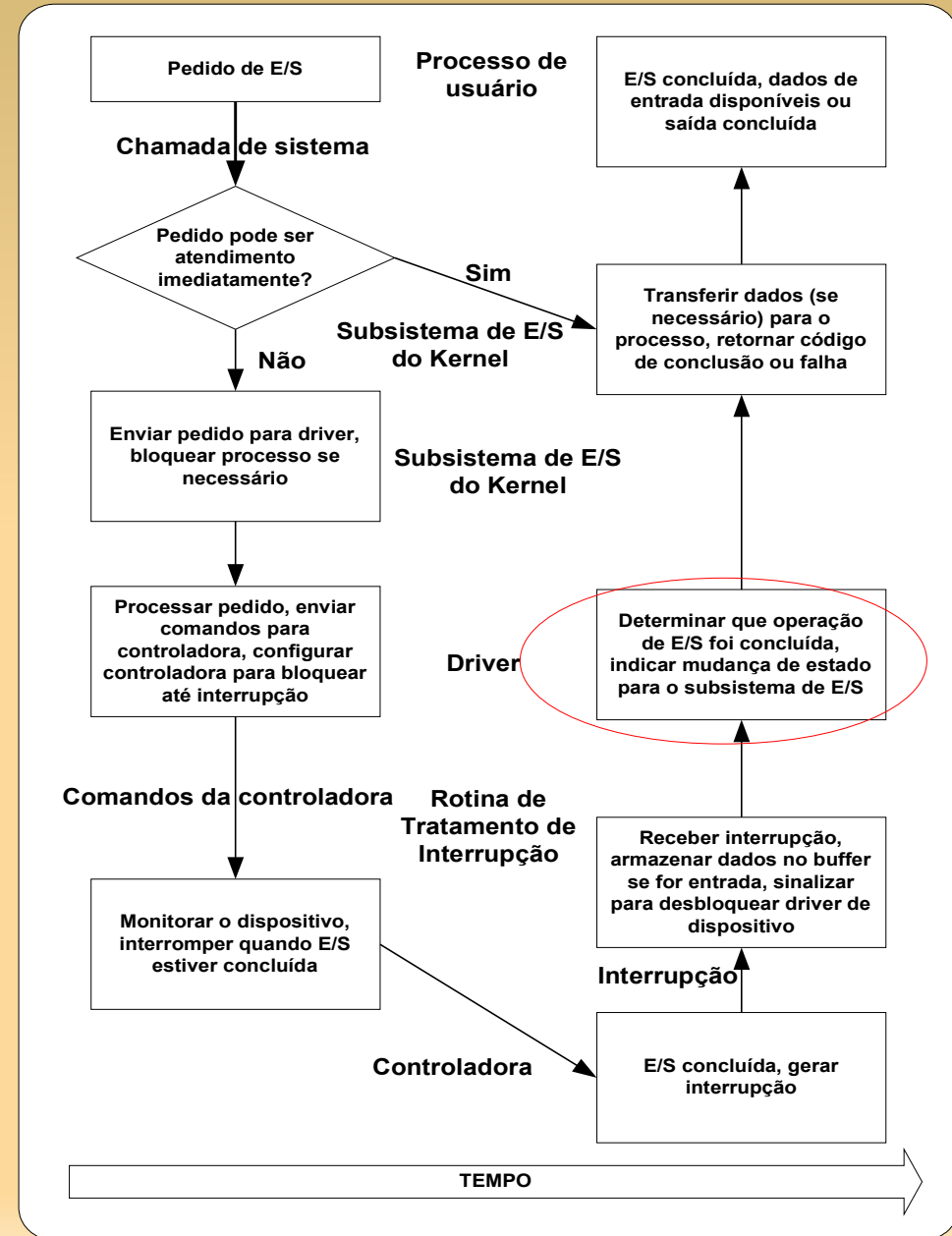
Drivers

- Funcionamento:
 - A rotina de tratamento de interrupções apropriada recebe a interrupção via vetor de interrupção, armazena os dados, sinaliza o driver e retorna da interrupção;



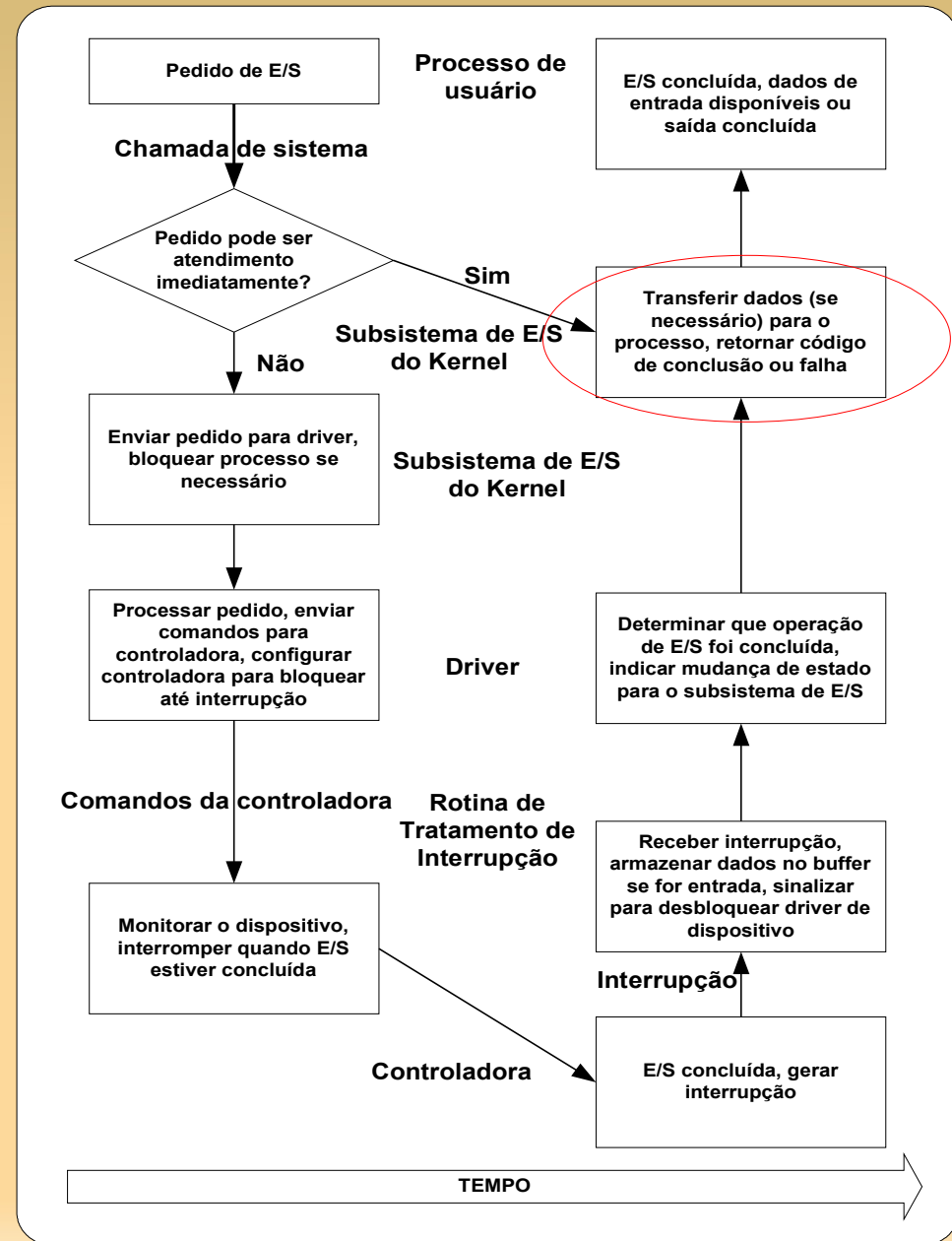
Drivers

- Funcionamento:
 - O Driver recebe o sinal, determina qual pedido de E/S foi concluído, determina o status e sinaliza que o pedido está concluído;



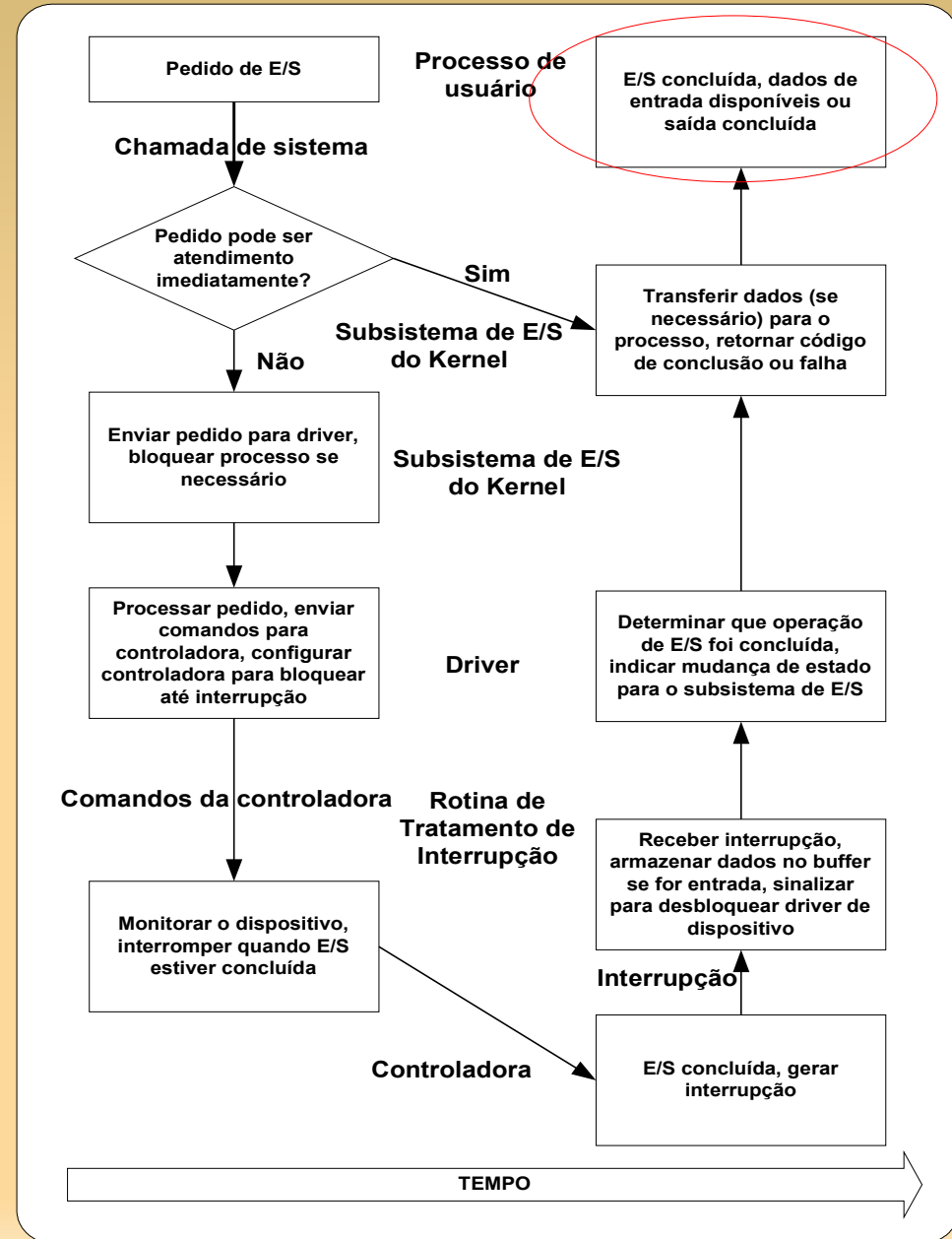
Drivers

- Funcionamento:
 - O Kernel transfere dados ou códigos de retorno para o espaço de endereçamento do processo que requisitou a E/S e move o processo da fila de bloqueados para a fila de prontos;

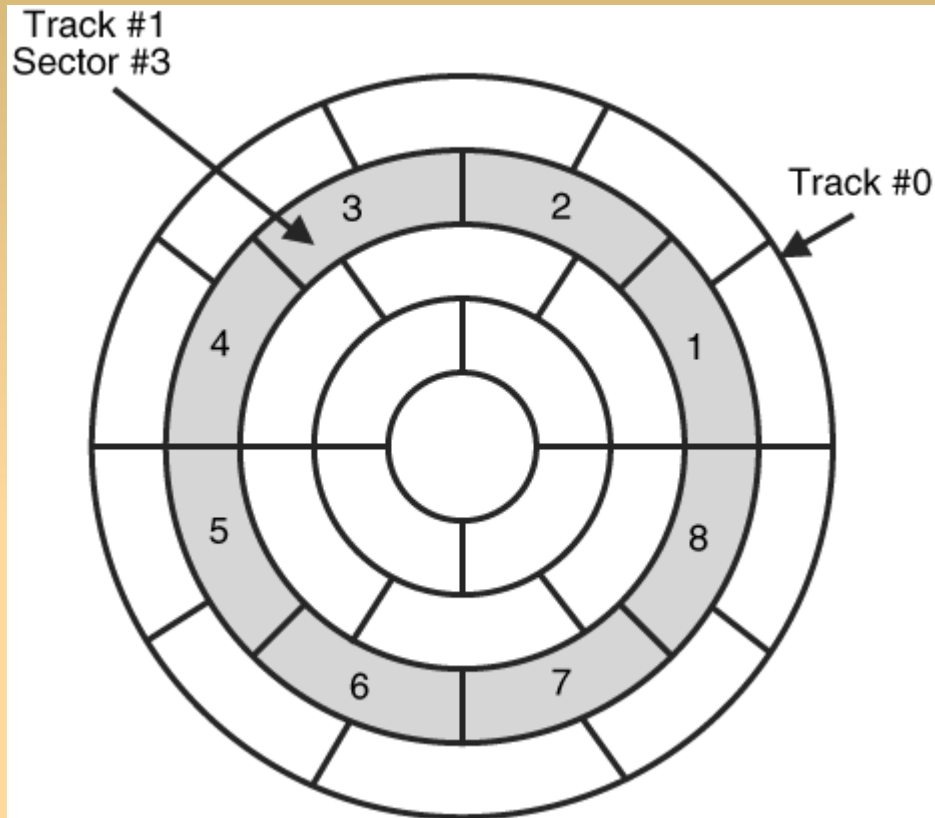


Drivers

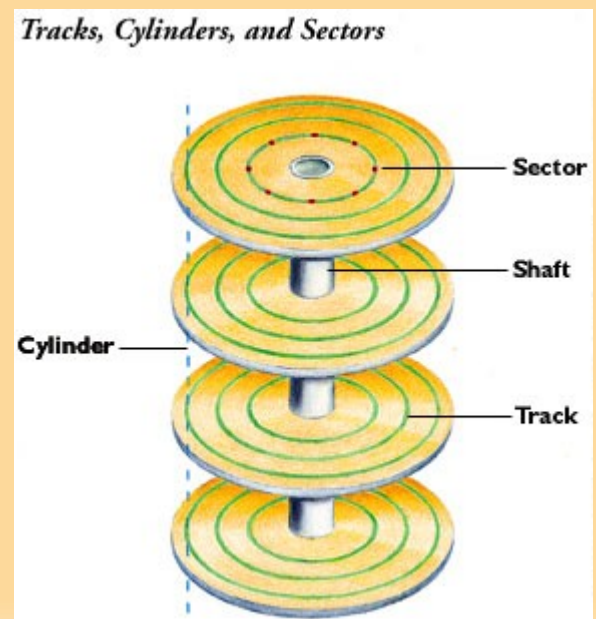
- Funcionamento:
 - Quando o escalonador escalona o processo para a CPU, ele retoma a execução na conclusão da chamada ao sistema.



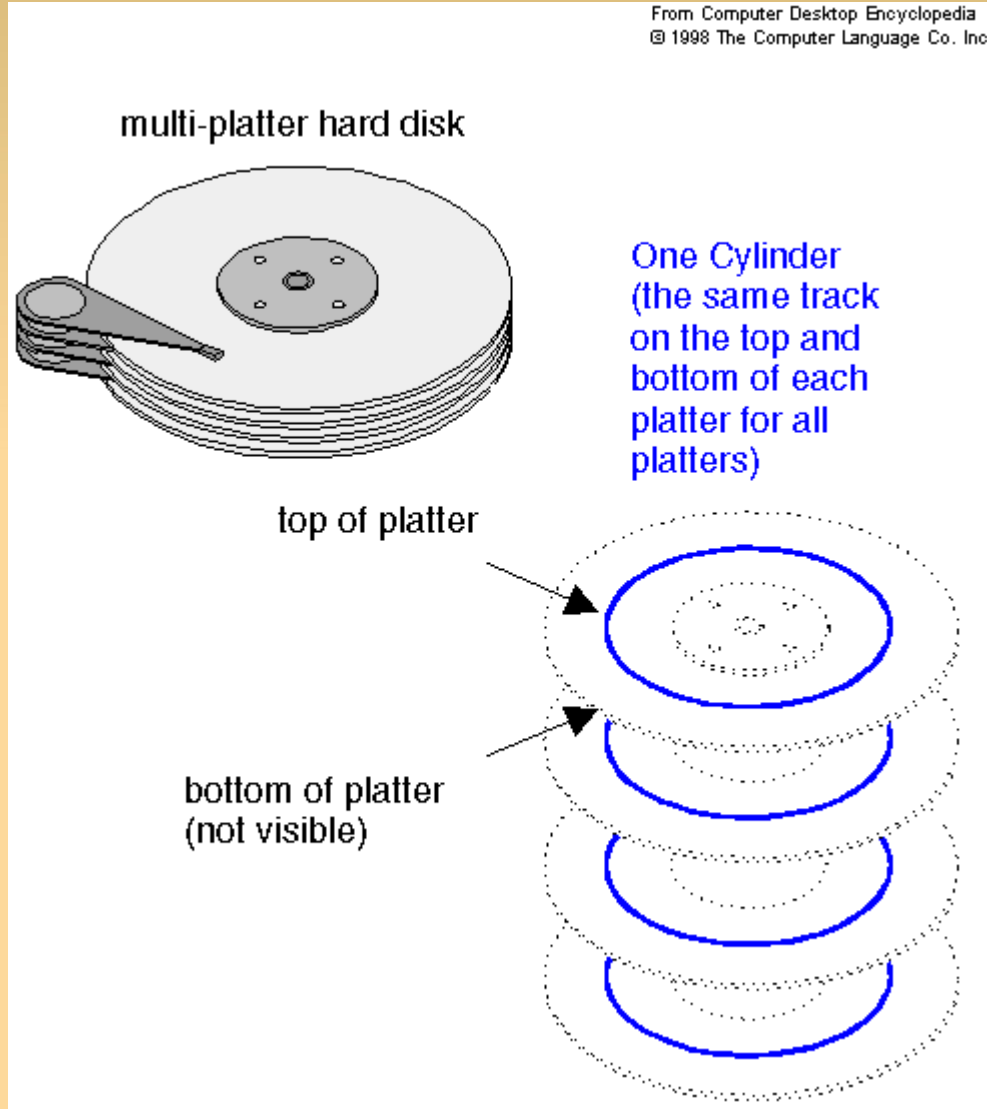
Hardware de Disco



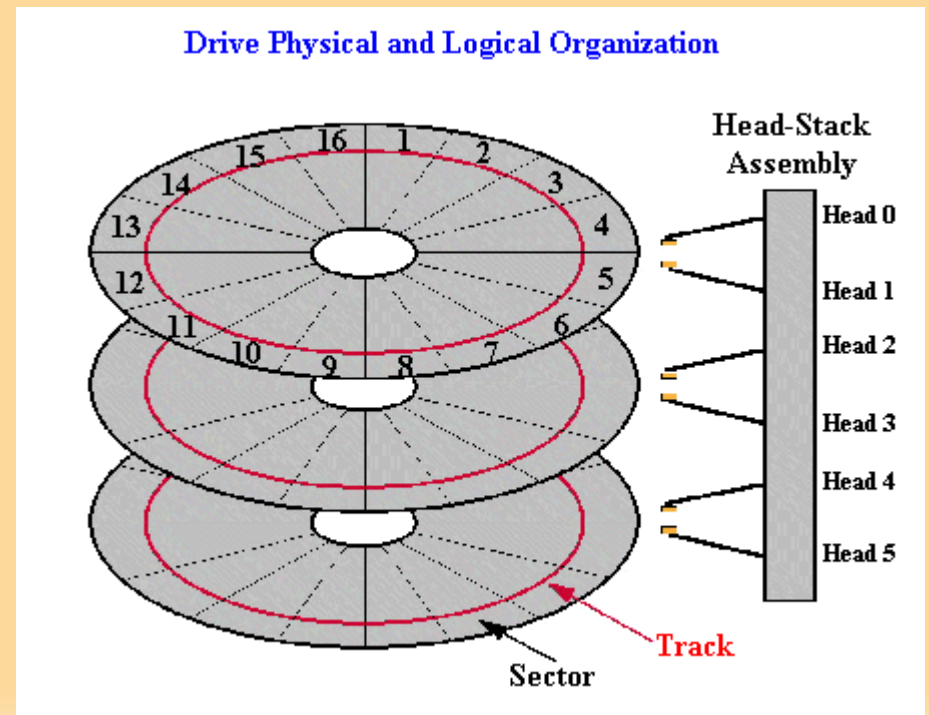
- * Cada superfície é dividida em **trilhas**;
- * Cada trilha é dividida em **setores** ou **blocos** (512 bytes a 32K);
- * Um conjunto de trilhas (com a mesma distância do eixo central) formam um **cilindro** (versão 3D da trilha)



Hardware de Disco

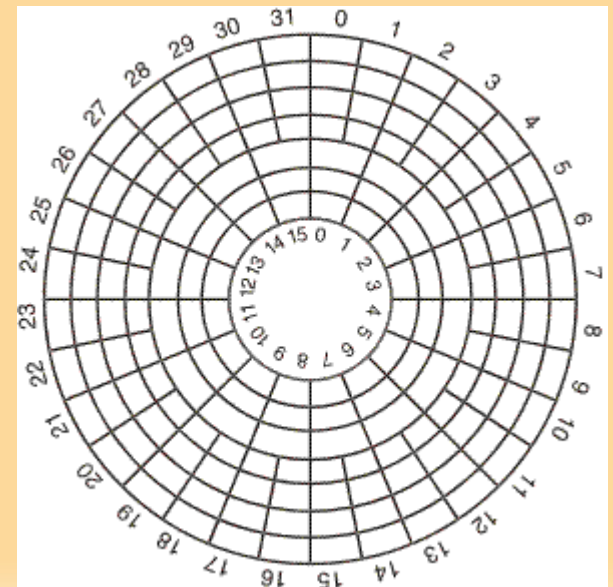
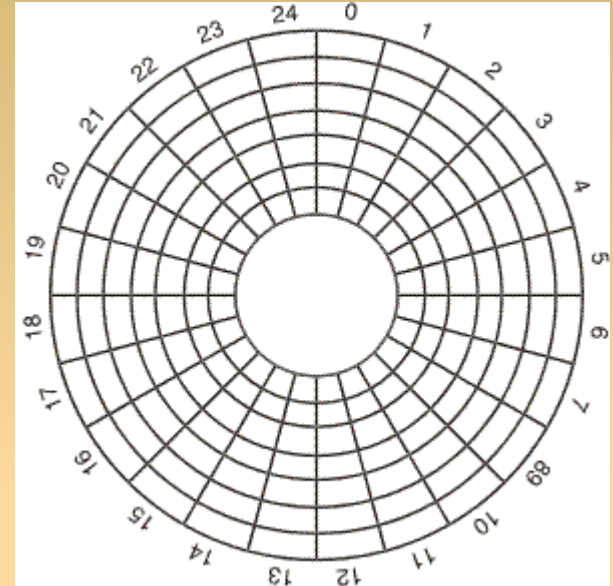


- * Cabeças de leitura e gravação;
- * Tamanho do disco:
 n° cabeças (faces) x n° cilindros
(trilhas) x n° setores x
tamanho_setor;



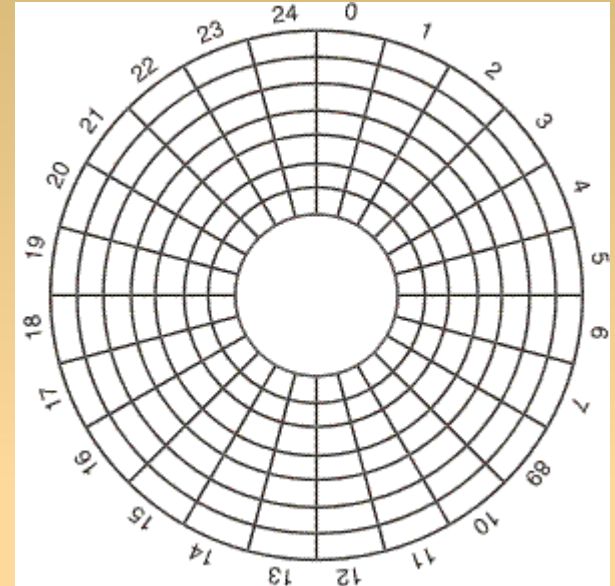
Hardware de Disco

- Geometria:
 - A geometria especificada (usada pelo driver) pode diferir da real
 - Em discos antigos, o número de setores por trilha era o mesmo para todos os cilindros
 - Discos modernos são divididos em zonas
 - Mais setores nas externas que nas internas



Hardware de Disco

- Geometria:
 - Apresentam uma geometria virtual ao SO
 - Escondem os detalhes de quantos setores há em cada trilha
 - O software age como se houvessem x cilindros, y cabeças e z setores por trilha.
 - O controlador do dispositivo mapeia um pedido de (x,y,z) para o cilindro, cabeça e setor reais
 - Discos modernos possuem endereçamento de bloco lógico:
 - Setores são numerados consecutivamente, iniciando no 0

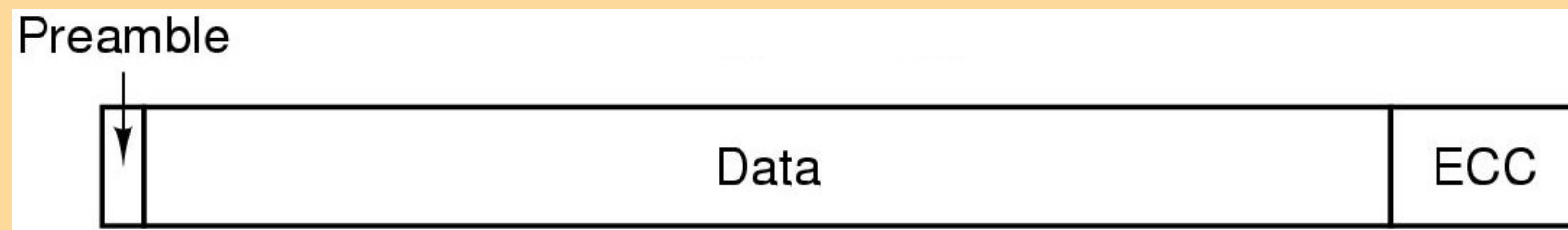


Hardware de Disco

- Formatação:

- Antes que possa ser usado, recebe uma formatação de baixo nível

- Trilhas concêntricas, contendo um número de setores
- Há um espaço pequeno entre os setores
- O setor é formatado da seguinte maneira:



Permite ao hardware reconhecer o início do setor. Contém também o número do setor e cilindro

Contém informação redundante que pode ser usada para recuperação de erros de leitura (ex: quando alguns bits dos dados estão inúteis)

Hardware de Disco

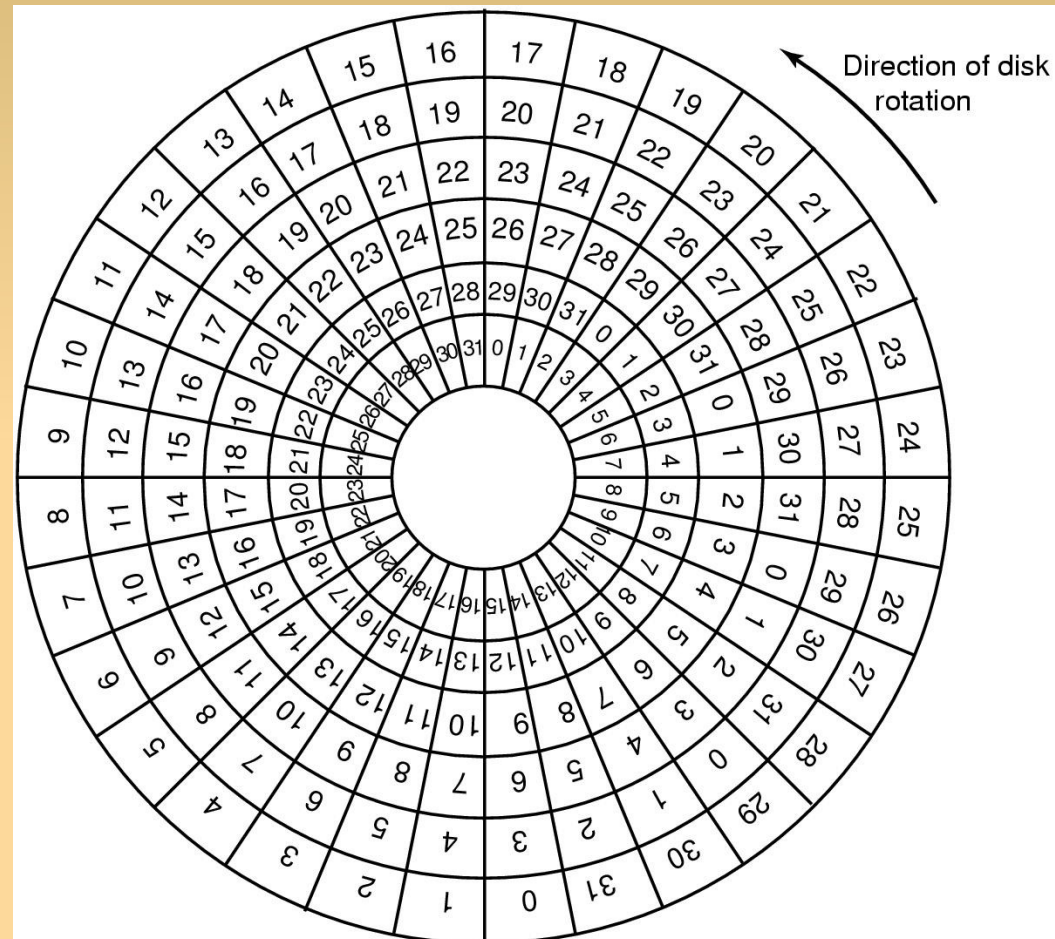
- Formatação:

- A posição do setor 0 de cada trilha é deslocada em relação à trilha anterior

- Torção cilíndrica (cylinder skew)

- Aumenta o desempenho

- Se o que deve ser lido for além do limite da trilha, não é preciso fazer nova busca para a trilha seguinte. Basta mover a cabeça e manter o disco rodando, que se chegará ao setor 0 da trilha seguinte.

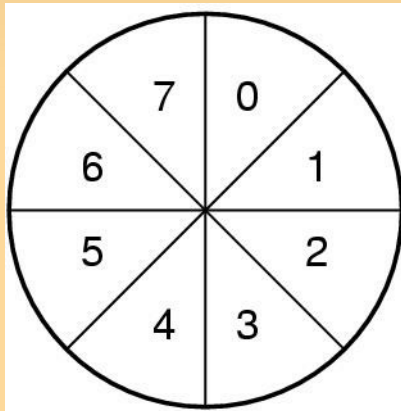


Hardware de Disco

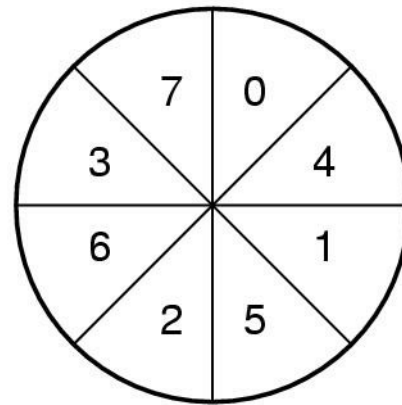
- Formatação:
 - Ao ser lido, o conteúdo do disco é transferido a um buffer
 - Quando fica cheio, o buffer é transferido à memória
 - O que toma tempo
 - Entre duas transferências do buffer à memória, pode-se passar da posição do dado no disco
 - deve-se esperar nova rotação
 - Solução: entrelaçamento
 - Dá algum tempo para que o buffer seja transferido

Hardware de Disco

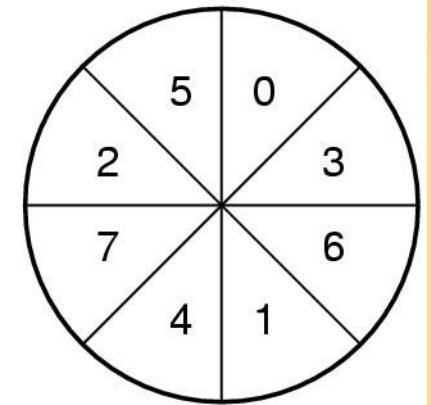
- Formatação:
 - Entrelaçamento



Sem entrelaçamento



Entrelaçamento simples



Entrelaçamento duplo

Hardware de Disco

- Formatação:
 - Particionamento
 - Executado após a formatação de baixo nível
 - Setor 0 contém o master boot record (MBR), com
 - Código (programa) de boot
 - Tabela de partições, com o setor de início e tamanho de cada partição
 - Normalmente, com espaço para 4 partições
 - Uma delas é marcada como ativa na tabela (para que se possa iniciar o computador a partir do HD)

Hardware de Disco

- Formatação:
 - Formatação de alto nível
 - Último passo, feito em cada partição separadamente
 - Define
 - Bloco de boot
 - Lista ou bitmap de blocos livres no disco
 - Diretório raiz (localização)
 - Sistema de arquivos
 - Altera a tabela de partição
 - Dizendo o SO que é usado na partição

Hardware de Disco

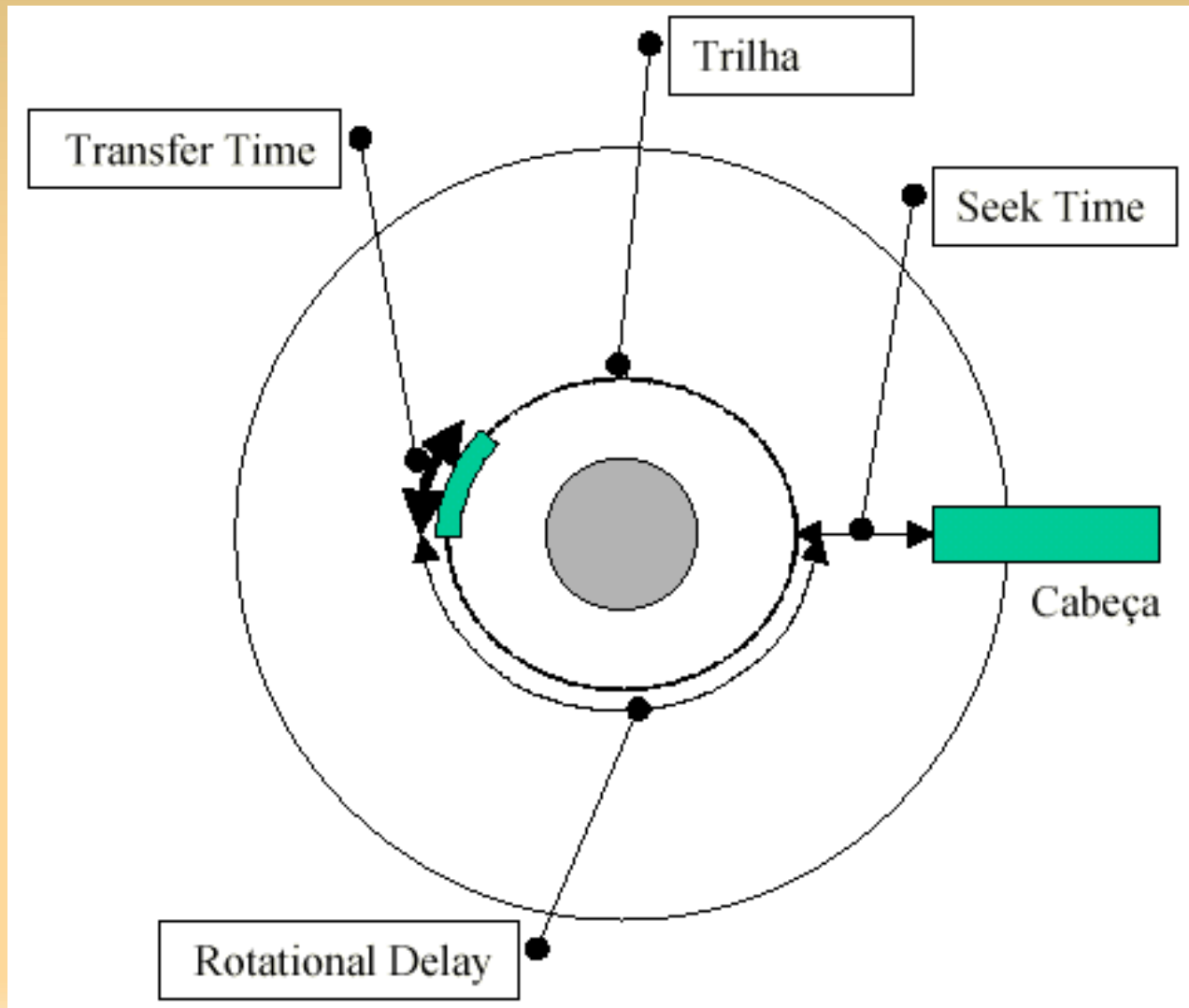
- Boot:
 - Ao ser ligado o computador, a bios lê o MBR
 - Verifica então que partição está ativa
 - Olhando a tabela de partições
 - Aqui cabe escolha, no caso de múltiplos SO
 - Lê então o setor de boot daquela partição e o roda
 - Este setor contém um programa que carrega um outro maior – bootstrap loader
 - O bootstrap loader varre o sistema de arquivos para encontrar o kernel do SO, carregando-o e executando-o

Hardware de Disco

- Drivers de Disco:
 - Fatores que influenciam tempo para leitura/escrita no disco:
 - Tempo de acesso (seek) → tempo para o movimento do braço até o cilindro;
 - Atraso rotacional (latência) → Tempo necessário para o cabeçote se posicionar no setor de escrita/leitura;
 - Tempo da transferência dos dados;
 - $T_{\text{acesso}} = T_{\text{seek}} + T_{\text{latência}} + T_{\text{transferência}}$

Hardware de Disco

$$t_{\text{access}} = t_{\text{seek}} + t_{\text{rotationaldelay}} + t_{\text{transfer}}$$



Hardware de Disco

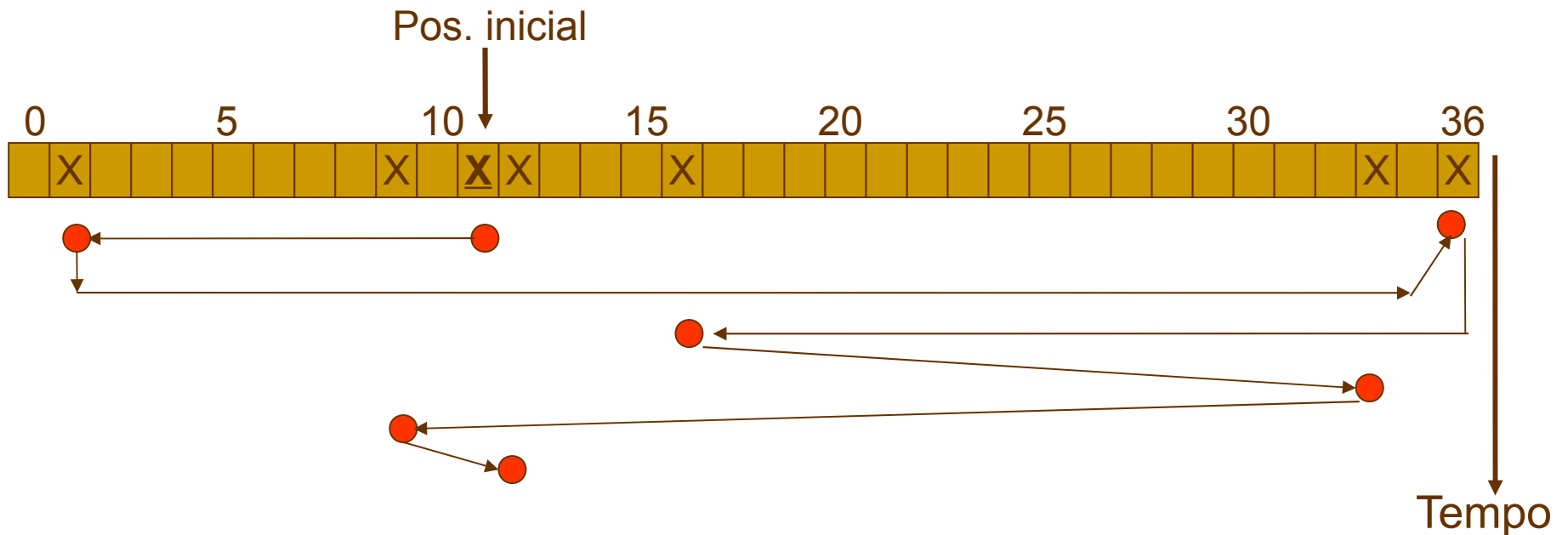
- Drivers de Disco – Escalonamento do braço:
 - Para muitos discos, o tempo de acesso domina
 - Bom lugar para reduções
 - Quando o disco está muito carregado, é provável que, durante uma busca, outras requisições sejam geradas por outros processos
 - O driver mantém uma tabela de requisições pendentes, indexada pelo número do cilindro
 - Com todas as requisições pendentes em uma lista ligada
 - Cada entrada da tabela tem a lista de requisições para seu cilindro correspondente

Hardware de Disco

- Drivers de Disco – Escalonamento do braço:
 - Algoritmos
 - First-Come, First-Served (FCFS)
 - O driver aceita uma requisição por vez, e as executa nessa ordem
 - Pouco pode ser feito para otimização
 - Ex:
 - Disco com 37 cilindros;
 - Atualmente lendo bloco no cilindro 11;
 - Surgem requisições para os cilindros 1,36,16,34,9,12, nesta ordem
 - As requisições pendentes são colocadas na tabela
 - Mantidas em lista ligada – uma para cada cilindro requisitado

Hardware de Disco

Disco com 37 cilindros;
Lendo bloco no cilindro 11;
Requisições: 1,36,16,34,9,12, nesta ordem



FCFS → atendimento: 1,36,16,34,9,12;

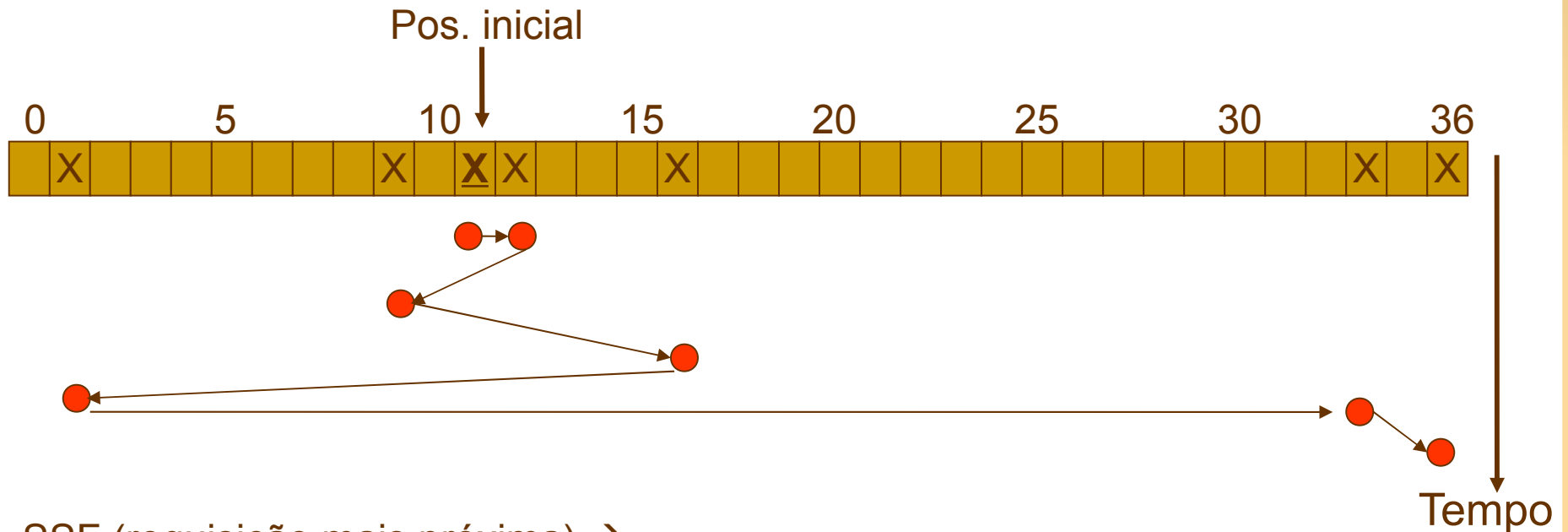
movimentos do braço (número de cilindros): 10,35,20,18,25,3 = 111;

Hardware de Disco

- Drivers de Disco – Escalonamento do braço:
 - Algoritmos
 - Shortest Seek First (SSF)
 - Sempre atenda em seguida a requisição mais próxima
 - Minimiza o tempo de busca

Hardware de Disco

Disco com 37 cilindros;
Lendo bloco no cilindro 11;
Requisições: 1,36,16,34,9,12, nesta ordem



SSF (requisição mais próxima) →
atendimento: 12,9,16,1,34,36;
movimentos do braço (número de cilindros): 1,3,7,15,33,2 = 61;

Hardware de Disco

- Drivers de Disco – Escalonamento do braço:
 - Algoritmos
 - Shortest Seek First (SSF)
 - Problema;
 - Se mais requisições forem chegando, a cabeça tenderá a não se mover muito de sua posição original
 - Se o disco estiver carregado, tenderá a ficar no meio a maior parte do tempo
 - Requisições nos extremos do disco demorarão a ser atendidas

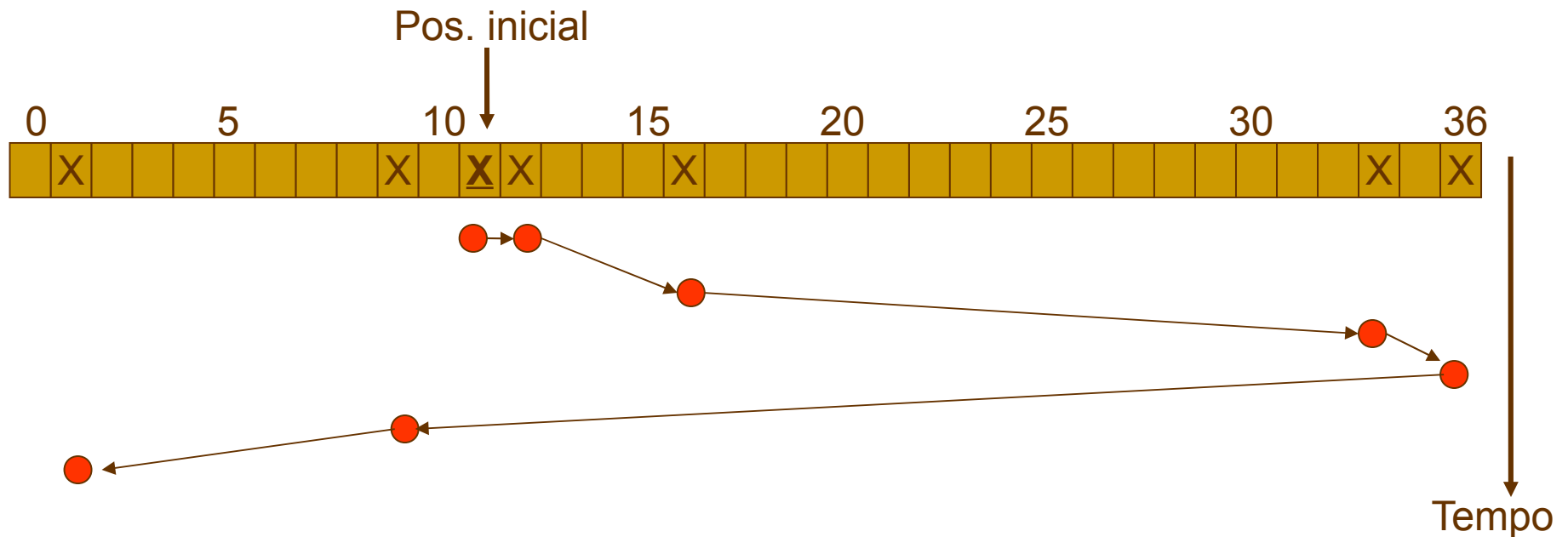
Hardware de Disco

- Drivers de Disco – Escalonamento do braço:
 - Algoritmos
 - Elevador
 - O problema de escalonar os andares de um elevador, em um edifício alto é semelhante ao braço do disco
 - Requisições chegam continuamente e aleatoriamente
 - Muitos elevadores tentam conciliar eficiência e justiça
 - Continuam se movendo na mesma direção até não haver mais requisições pendentes naquela direção
 - Então trocam de direção
 - No disco, o driver deve manter 1 bit a direção (up ou down)
 - Quando uma requisição termina, o driver verifica o bit
 - Se for up, o braço é movido à próxima requisição mais alta
 - Se não houver requisições pendentes nessa direção, o bit é feito down, e o braço se move à próxima requisição mais baixa

Hardware de Disco

Disco com 37 cilindros;
Lendo bloco no cilindro 11;
Requisições: 1,36,16,34,9,12, nesta ordem

Bit de direção corrente (*driver*):
Se *Up* → atende próxima requisição;
senão *Bit = Down*;
muda direção e atende requisição;



Elevator (requisições na mesma direção) →

atendimento: 12,16,34,36,9,1

movimentos do braço (número de cilindros): 1,4,18,2,27,8 = 60;

Hardware de Disco

- RAID
 - Originalmente: Redundant Array of Inexpensive Disks
 - Redefinidos pela indústria como Redundant Array of Independent Disks
 - Surge da do fato de que o desempenho da CPU aumenta muito mais rápido que o dos discos
 - Projetado para armazenar grandes quantidades de dados

Hardware de Disco

- RAID
 - Combina diversos discos rígidos em uma estrutura lógica:
 - Aumenta a confiabilidade, capacidade e o desempenho dos discos;
 - Recuperação de dados → redundância dos dados;
 - Armazenamento simultâneo em vários discos permite que os dados fiquem protegidos contra falha (não simultânea) dos discos;
 - Performance de acesso, já que a leitura da informação é simultânea nos vários dispositivos;

Hardware de Disco

- RAID

- Pode ser implementado por:

- Hardware (controladora):

- Instalação de uma placa RAID no servidor, o subsistema RAID é implementado totalmente em hardware;
 - Libera o processador para se dedicar exclusivamente a outras tarefas;
 - A segurança dos dados aumenta no caso de problemas devido à checagem da informação na placa RAID antes da gravação;

- Software (sistema operacional)

- Menor desempenho no acesso ao disco;
 - Oferece um menor custo e flexibilidade;
 - Sobrecarrega o processador com leitura/escrita nos discos;
 - Qualquer que seja, para o SO existe um único disco;

Hardware de Disco

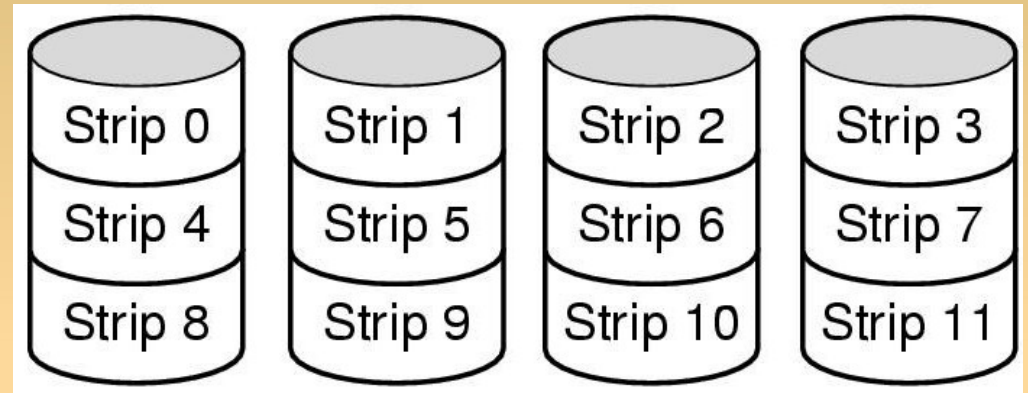
- RAID
 - Os dados são distribuídos pelos drives
 - Permite operações paralelas
 - Obtidos a partir de diferentes esquemas (chamados níveis)
 - A forma pela qual os dados são escritos e acessados define os níveis de RAID (até 9 níveis):
 - RAID 0:
 - Também conhecido como Stripping;
 - Vê o disco virtual simulado pelo RAID como dividido em tiras de k setores cada
 - Strips consecutivos escritos com round-robin

Hardware de Disco

■ RAID

■ RAID 0:

- Melhora desempenho das operações de E/S
 - Se o software manda ler um bloco de 4 tiras consecutivas, , iniciando em um limite de tira, o controlador do RAID quebrará esse comando em 4 – um para cada disco
- lê e escreve os dados de maneira paralela
- divide os discos em fatias
- Utilizam mesma controladora (controladora RAID);
- Aplicações multimídia (alta taxa de transferência);

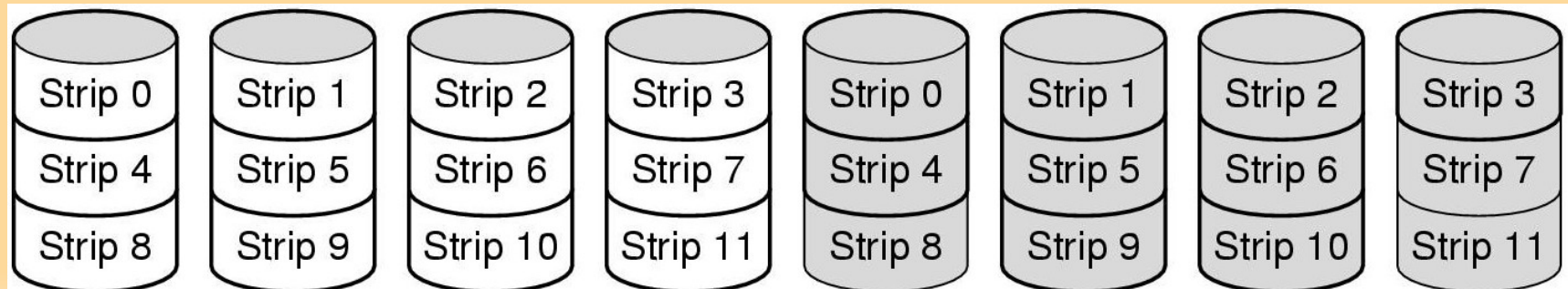


Hardware de Disco

- RAID

- RAID 1:

- Conhecido como espelhamento (mirroring);
 - Duplica todos os discos
 - 4 principais e 4 de reserva
 - Operações de escrita no disco primário são replicadas em um disco secundário;
 - Leitura pode ser feita de qualquer cópia → distribui a carga
 - Pode ter controladoras diferentes;



Hardware de Disco

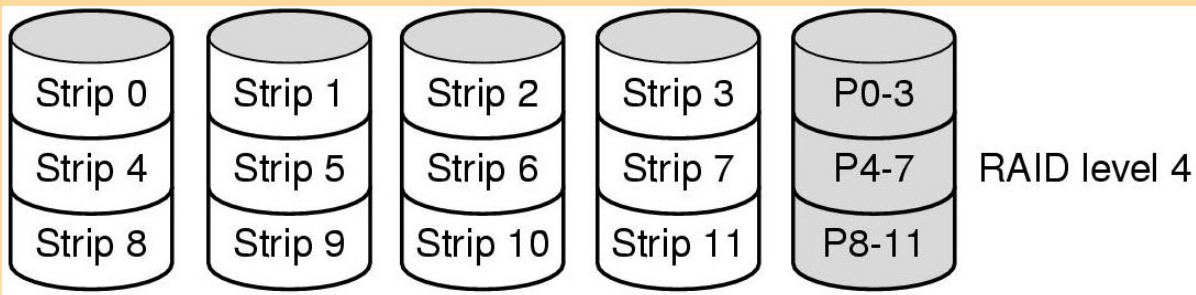
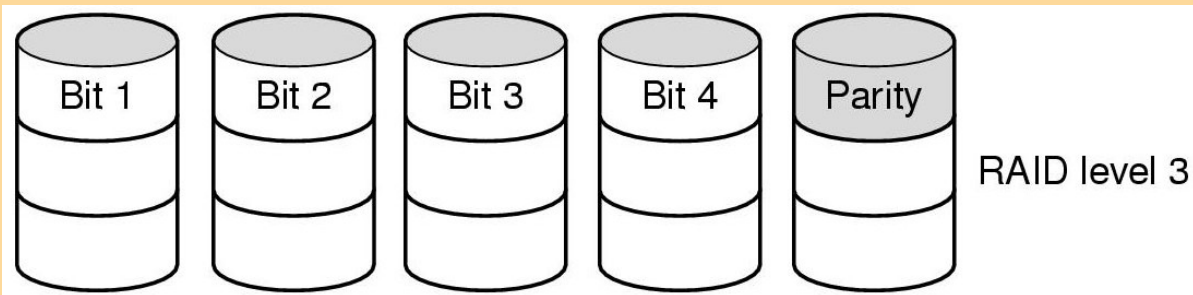
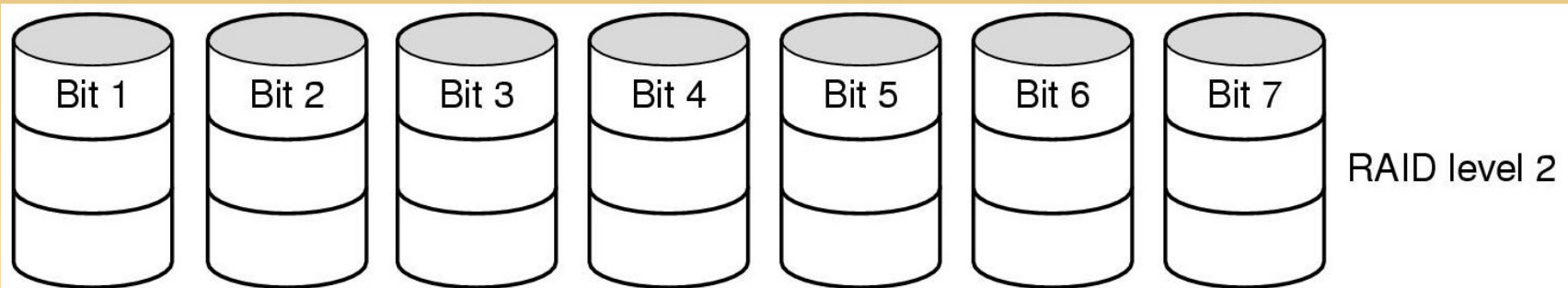
- RAID
 - RAID 1:
 - Excelente tolerância a falhas
 - Se um drive falhar, a cópia é usada
 - Recuperação consiste em instalar um novo disco e copiar do backup para ele
 - Desvantagem: espaço físico em dobro (alto custo);
 - Transações on-line (tolerância a falhas);
 - RAID 10:
 - Combinação dos RAID 1 e RAID 0;

Hardware de Disco

- RAID 2
 - os dados podem ser quebrados em bytes ou palavras
- RAID 3
 - versão simplificada do RAID 2, possui um disco separado apenas para os bits de paridade
- RAID 4
 - semelhante ao RAID 0
 - possui bits de paridade em disco separado para aumentar a confiabilidade

Hardware de Disco

- RAID



Hardware de Disco

- Raid

- RAID 5:

- Stripes;
 - Paridade XOR ECC distribuída - nível de bloco;
 - Paridade está distribuída nos discos;

