

Capítulo 5

Entrada/Saída

- 5.1 Princípios do hardware de E/S
- 5.2 Princípios do software de E/S
- 5.3 Camadas do software de E/S
- 5.4 Discos

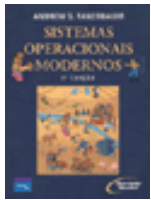
Princípios do Hardware de E/S



Dispositivo	Taxa de dados
Teclado	10 bytes/s
Mouse	100 bytes/s
Modem 56 K	7 KB/s
Canal telefônico	8 KB/s
Linhas ISDN dual	16 KB/s
Impressora a laser	100 KB/s
Scanner	400 KB/s
Ethernet clássica	1,25 MB/s
USB (<i>universal serial bus</i> — barramento serial universal)	1,5 MB/s
Câmara de vídeo digital	4 MB/s
Disco IDE	5 MB/s
CD-ROM 40x	6 MB/s
Ethernet rápida	12,5 MB/s
Barramento ISA	16,7 MB/s
Disco EIDE (ATA-2)	16,7 MB/s
FireWire (IEEE 1394)	50 MB/s
Monitor XGA	60 MB/s
Rede SONET OC-12	78 MB/s
Disco SCSI Ultra 2	80 MB/s
Ethernet Gigabit	125 MB/s
Dispositivo de Fita Ultrium	320 MB/s
Barramento PCI	528 MB/s
Barramento da Sun Gigaplane XB	20 GB/s

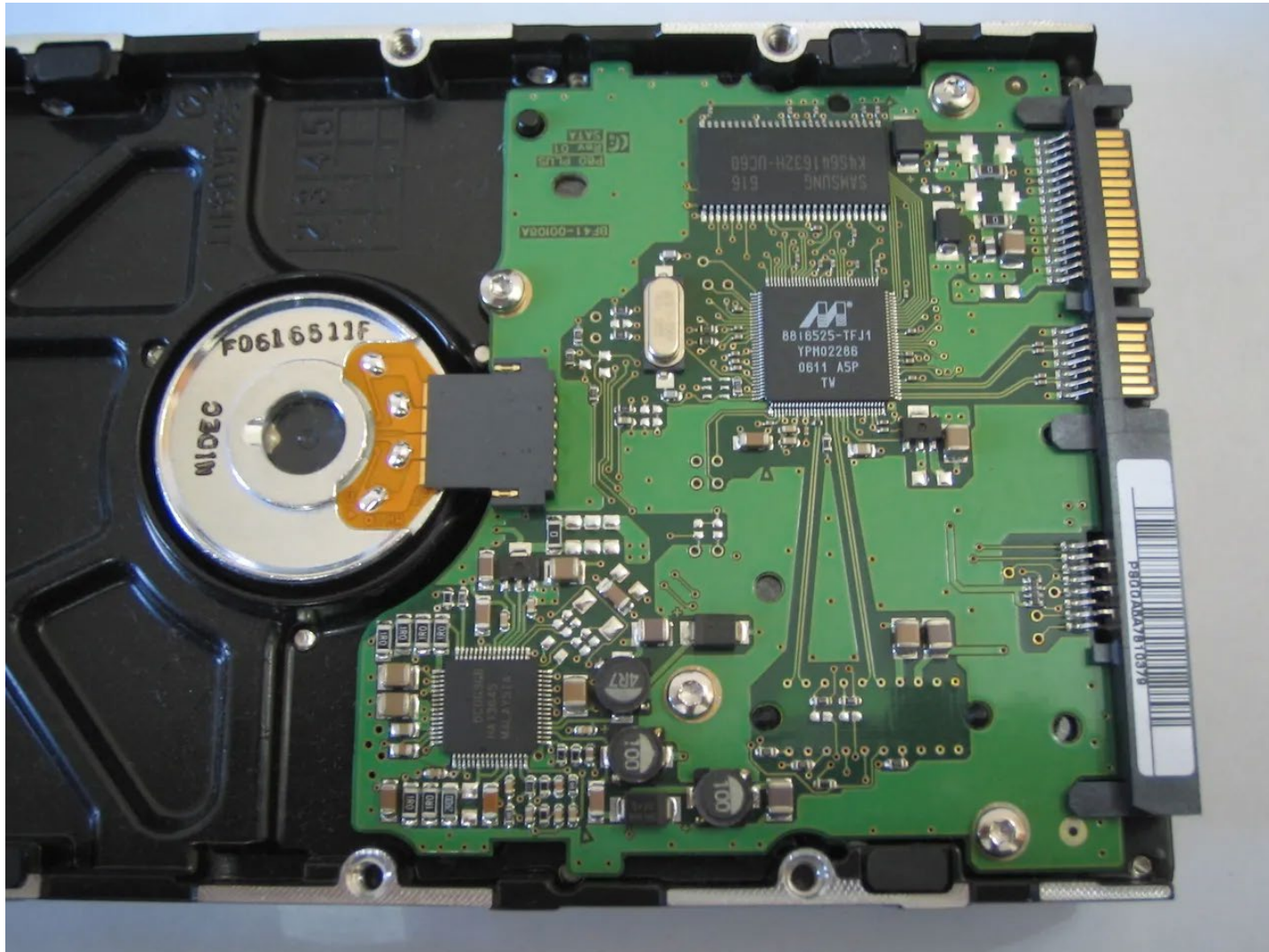
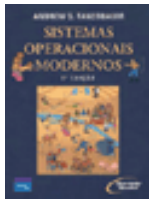
Taxas de dados típicas de dispositivos, redes e barramentos

Controladores de Dispositivos

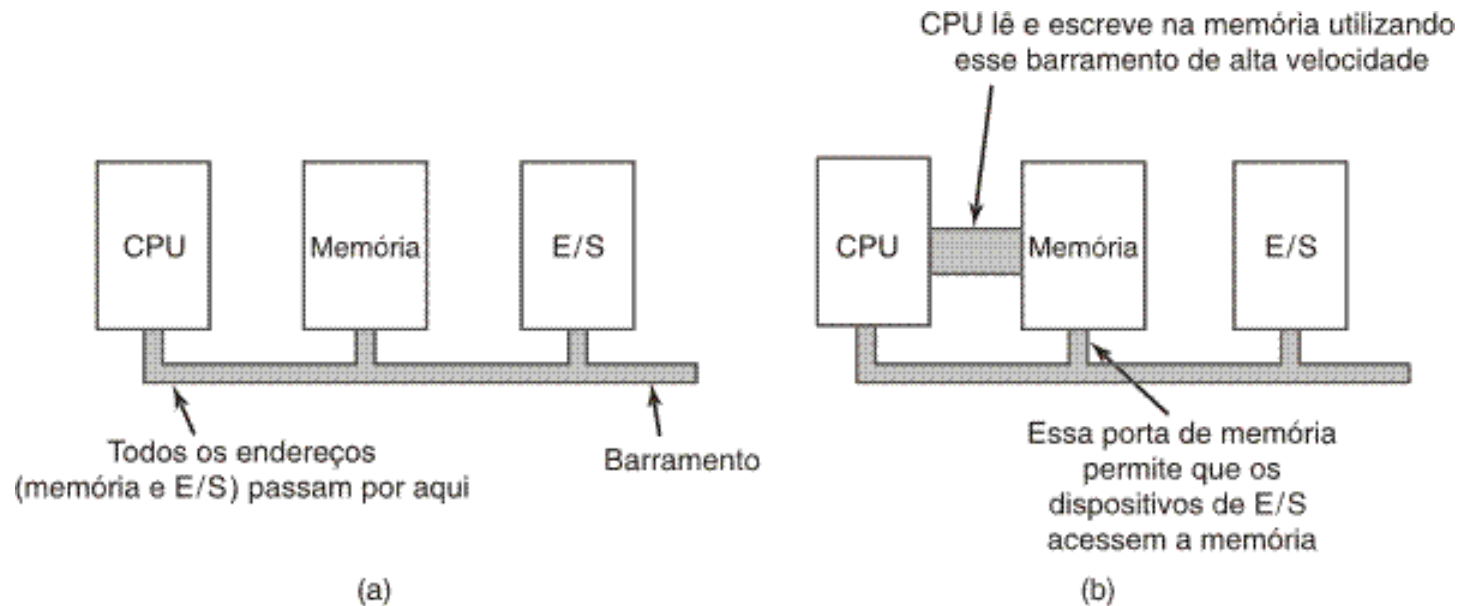
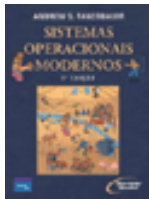


- Componentes de dispositivos de E/S
 - mecânico
 - eletrônico
- O componente eletrônico é o controlador do dispositivo
 - pode ser capaz de tratar múltiplos dispositivos
- Tarefas do controlador
 - converter fluxo serial de bits em bloco de bytes
 - executar toda correção de erro necessária
 - tornar o bloco disponível para ser copiado para a memória principal

Controladores de Dispositivos



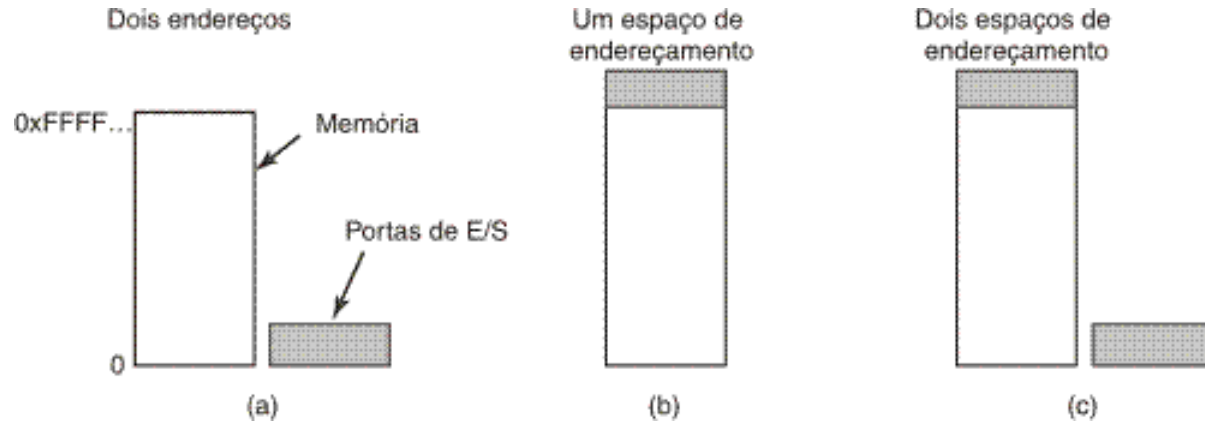
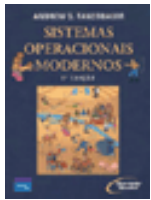
E/S Mapeada na Memória



(a) Arquitetura com barramento único

(b) Arquitetura com barramento dual

E/S Mapeada na Memória



- a) Espaços de memória e E/S separados
- b) E/S mapeada na memória
- c) Híbrido

E/S Mapeada na Memória



- Associação de endereços aos múltiplos dispositivos
- O SO só pode saber o status de cada dispositivo a partir dos registradores e memória
- Duas abordagens de endereços pelo modulo de E/S:
 - **Mapeado em memória:** O modulo de E/S opera dentro do espaço de endereçamento de memória, utilizando um conjunto de endereços reservados (Registradores são tratados como posições de memória)
 - **Mapeado em E/S ou E/S separado:** Existe um espaço de endereçamento independente para os dispositivos de E/S. Uso de instruções especiais de E/S.

Princípios do Software de E/S

Objetivos do Software de E/S



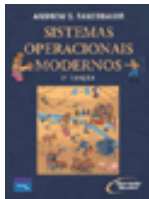
- Independência de dispositivo
 - Programas podem acessar qualquer dispositivo de E/S sem especificar previamente qual (disquete, disco rígido ou CD-ROM)
- Nomeação uniforme
 - Nome de um arquivo ou dispositivo pode ser uma cadeia de caracteres ou um número inteiro que é independente do dispositivo
- Tratamento de erro
 - Trata o mais próximo possível do hardware

Objetivos do Software de E/S

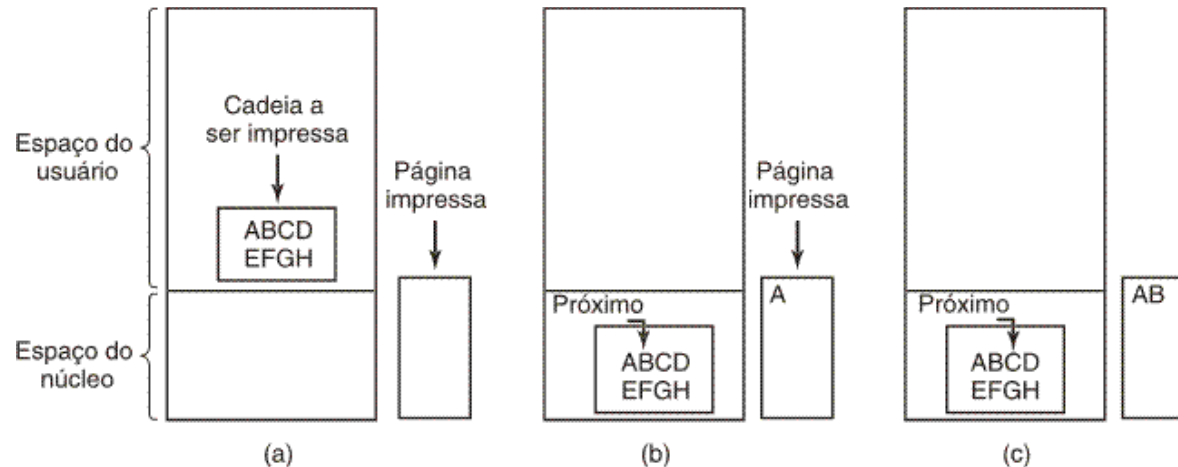


- Softwares Independente de E/S
 - Transferência de Dados:
 - Síncrona (Bloqueante): Requer bloqueio até que os dados estejam prontos para transferência
 - Assíncrona (Não-Bloqueante): Transferências acionadas por interrupções (Mais comuns)
 - Tipos de Dispositivos
 - Compartilháveis: Podem ser utilizados por vários usuários ao mesmo tempo. Exemplo: Disco Rígido
 - Dedicados: Podem ser utilizados por apenas um usuário de cada vez. Exemplo: Unidade de fita

E/S Programada



- E/S Programada:
 - Passos para a impressão de uma cadeia de caracteres (Laço até que toda a cadeia tenha sido impressa)
 - Impressoras modernas requerem uma ou mais páginas a serem impressas ao invés de caracteres



Passos da impressão de uma cadeia de caracteres

E/S Programada

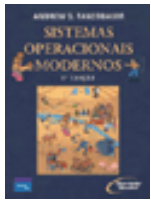


- E/S Programada:
 - Desvantagens:
 - CPU é ocupada o tempo todo até que a E/S seja feita
 - CPU continuamente verifica se o dispositivo está pronto para aceitar outro caractere → Espera ocupada
 - CPU deveria estar ocupada com trabalhos mais “nobres”



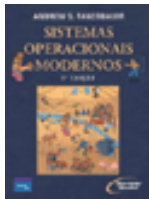
Recibo de Pagamento de Salário					
NOME DA EMPRESA LTDA CNPJ: 00.000.000/001-36			JANEIRO/2011		
Código	Nome do Funcionário	CSO	Emp.	Local	Depto. Secor. Seção
9011	LAURO ALBUQUERQUE				ANALISTA FINANCEIRO PLENO
Cód.	Descrição	Referência	Vencimentos	Descontos	
101	SALARIO	30 d	3.300,00		
973	INSS	11,00%		363,00	
987	IRFF S. SALARIO	7,50%		247,50	
			Total de Vencimentos	3300,00	
			Total de Descontos	610,50	
			Valor Líquido	2.689,50	
Salário Base	Sal. Contr. INSS	Base Cálcl. FGTS	FGTS do Mês	Base Cálcl. IRRF	Faixa IRRF
2.500,00	2.500,00	2.500,00	200,00	2225,00	02

E/S Orientada à Interrupção

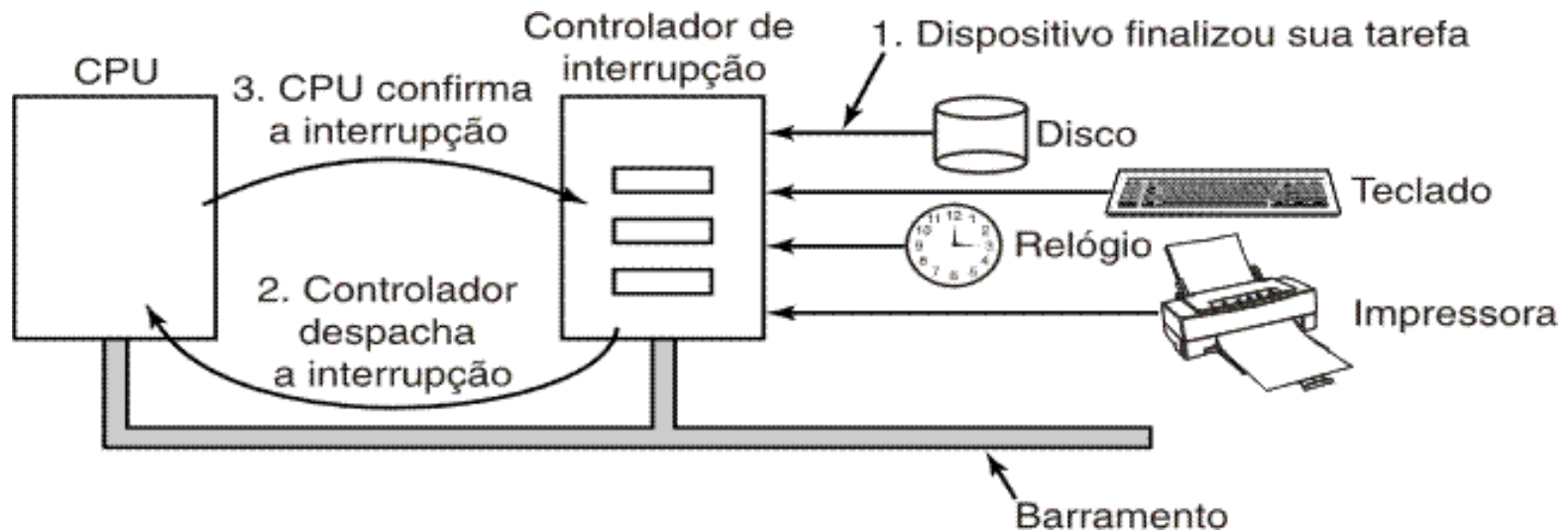


- Na E/S via interrupção:
 - O mecanismo de interrupções é utilizado para superar o problema de espera da CPU por operações nos periféricos
 - A interrupção permite que uma unidade ganhe a atenção imediata de outra, de forma que a primeira possa finalizar sua tarefa
- A CPU:
 - Envia um comando para o modulo de E/S e passa a executar outra tarefa
 - Quando a operação for concluída, o modulo de E/S interrompe a CPU
 - A CPU executa a troca de dados, liberando o modulo de E/S e retomando o processamento anterior

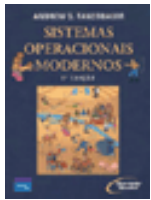
Interrupções Revisitadas



- Como ocorre uma interrupção. Conexões entre dispositivos e controlador de interrupção usam linhas de interrupção no barramento em vez de fios dedicados

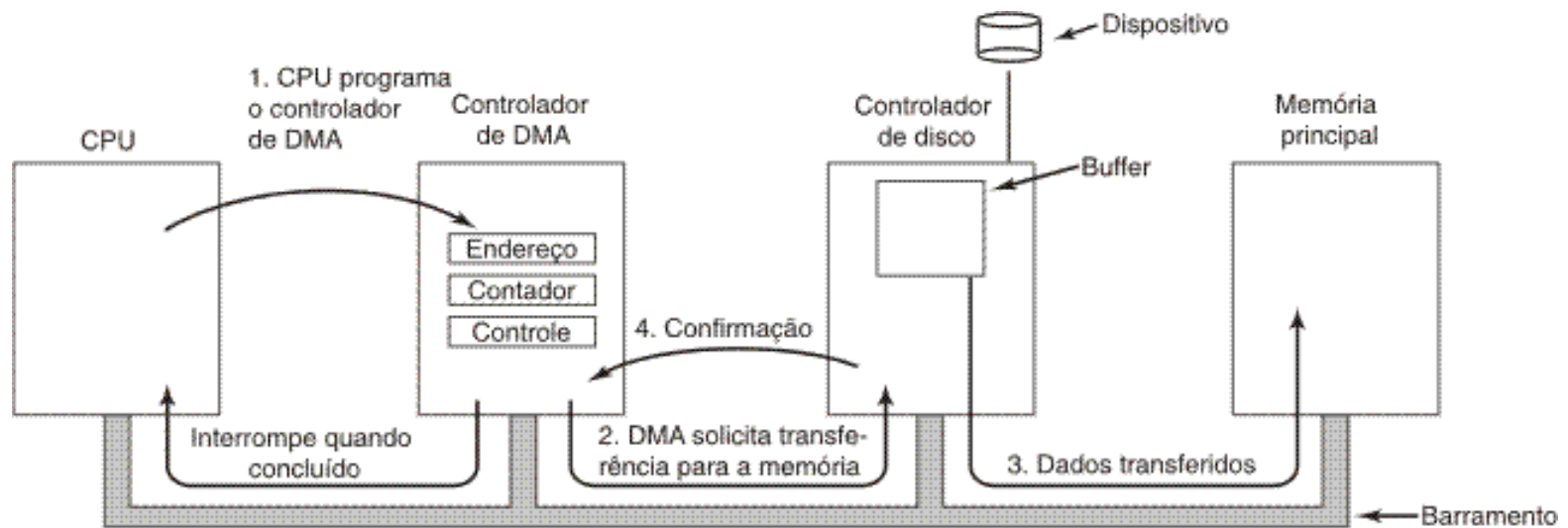
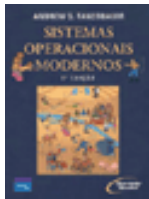


E/S Orientada à Interrupção



- E/S orientada à interrupção:
 - No caso da impressão, a impressora pode armazenar dados a serem impressos, no seu buffer ou não armazenar os caracteres
 - Em impressoras modernas, o documento inteiro ou algumas páginas podem ser enviados para o buffer da impressora
 - Quando a impressora está pronta para receber outros caracteres, gera uma interrupção
 - Processo é bloqueado

Acesso Direto à Memória (DMA)



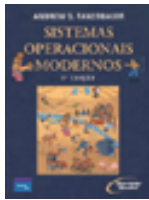
Operação de uma transferência com DMA

E/S Usando DMA



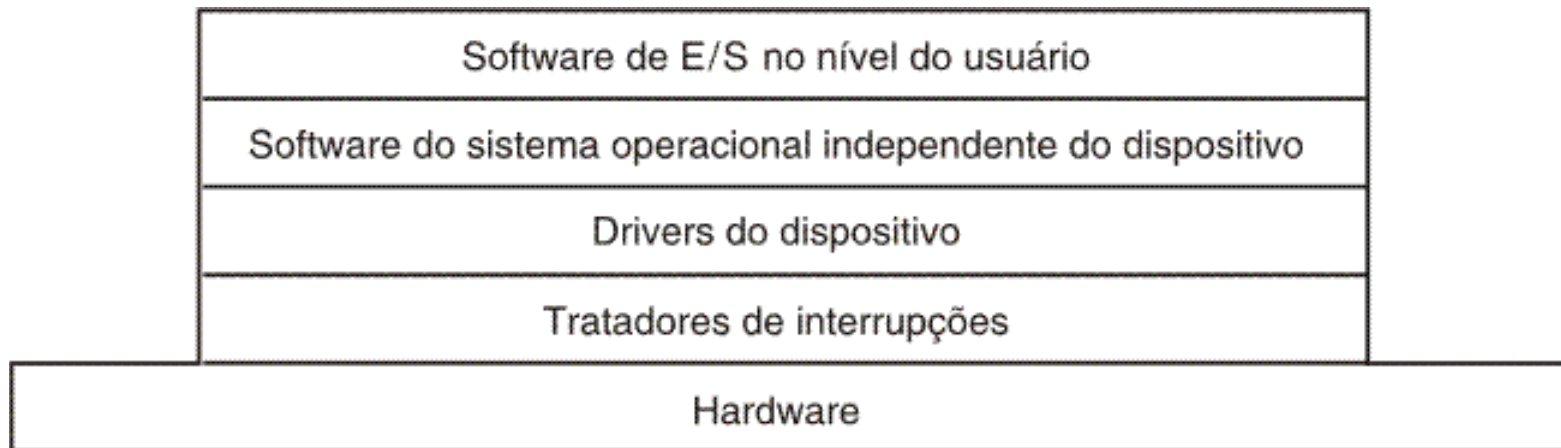
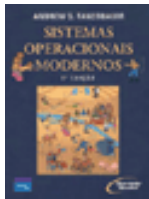
- Inconvenientes das técnicas anteriores:
 - Limitam a capacidade de transferência da CPU, entre o módulo de E/S e a Memória Principal
 - Uso de mais de uma instrução
 - CPU fica ocupada no gerenciamento
 - Grande quantidade de dados compromete desempenho
- Solução: Permitir acesso direto à memória
 - Uso de única interrupção para efetuar a transferência de um bloco de dados entre o periférico e a memória principal
 - CPU tem envolvimento mínimo no gerenciamento

E/S Usando DMA



- Necessidade de módulo adicional
 - Controlador de DMA
- Operação do Controlador de DMA:
 - CPU envia comando (Leitura ou Escrita) para o Controlador de DMA
 - CPU continua seu trabalho
 - Controlador de DMA, para acessar a memória, “rouba” ciclos da CPU, causando pequenos atrasos
 - Ao final da operação, o Controlador de DMA aciona a interrupção para sinalizar o término da operação
 - A CPU pode executar rotina de tratamento da interrupção, processando dados lidos ou produzindo novos dados para serem escritos

Camadas do Software de E/S



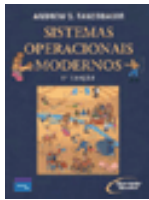
Camadas do sistema de software de E/S

Tratadores de Interrupção



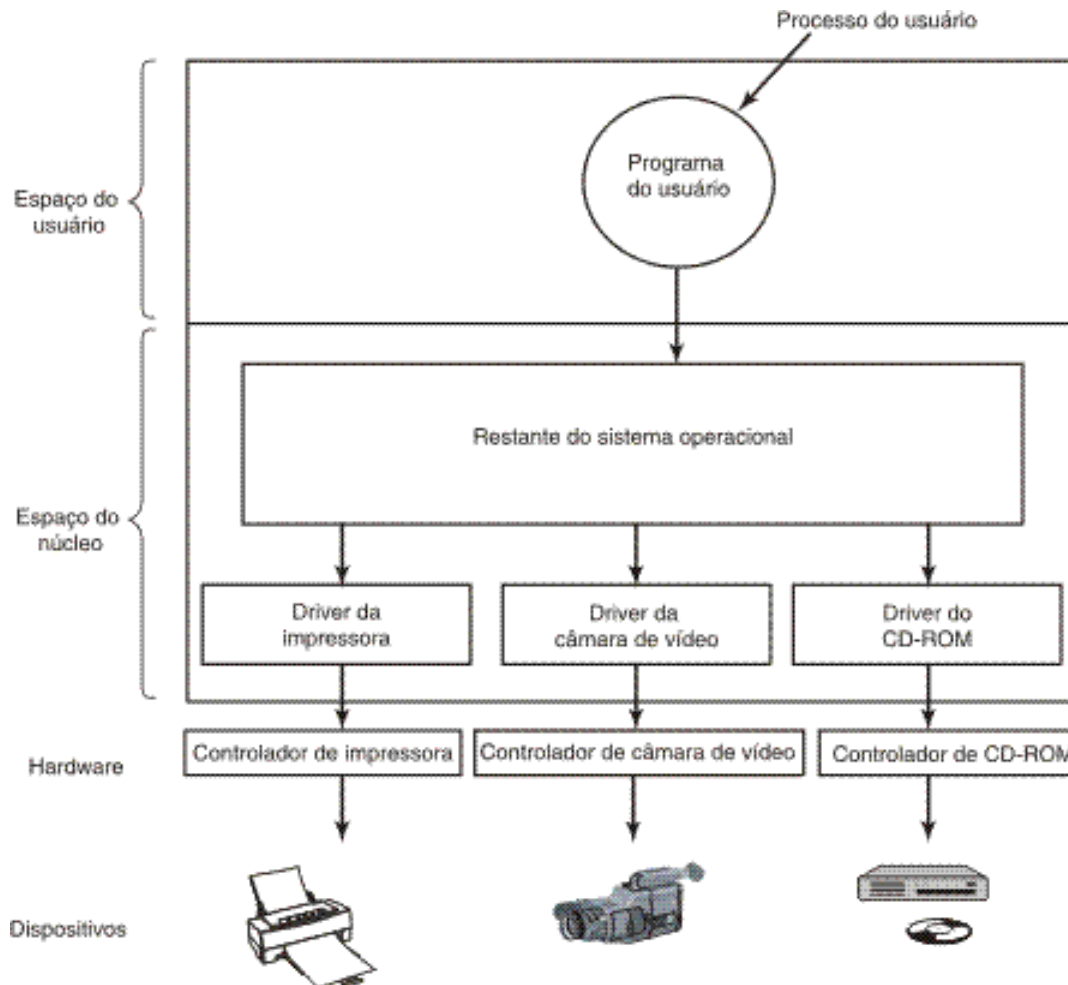
- As interrupções devem ser escondidas o máximo possível
 - Uma forma de fazer isso é bloqueando o driver que iniciou uma operação de E/S até que uma interrupção notifique que a E/S foi completada
- Rotina de tratamento de interrupção cumpre sua tarefa
 - Então desbloqueia o driver que a chamou

Tratadores de Interrupção



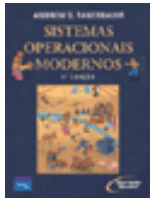
- Passos que devem ser executados em software depois da interrupção ter sido concluída:
 1. Salva registradores que ainda não foram salvos pelo hardware de interrupção
 2. Estabelece contexto para rotina de tratamento de interrupção
 3. Estabelece uma pilha para a rotina de tratamento de interrupção
 4. Sinaliza o controlador de interrupção, reabilita as interrupções
 5. Copia os registradores de onde eles foram salvos
 6. Executa rotina de tratamento de interrupção
 7. Escolhe o próximo processo a executar
 8. Estabelece o contexto da MMU para o próximo processo a executar
 9. Carrega os registradores do novo processo
 10. Começa a executar o novo processo

Drivers dos Dispositivos



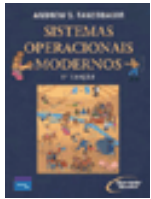
- Posição lógica dos drivers dos dispositivos
- A comunicação entre os drivers e os controladores de dispositivos é feita por meio do barramento

Drivers dos Dispositivos



- Drivers:
 - São gerenciados pelo kernel do SO
 - Contêm todo o código dependente do dispositivo
 - Controlam o funcionamento dos dispositivos por meio de sequência de comandos escritos/lidos nos/dos registradores da controladora
 - Dispositivos diferentes possuem Drivers diferentes:
 - Classes de dispositivos podem ter o mesmo Driver
 - São dinamicamente carregados
 - Drivers defeituosos podem causar danos ao kernel do SO

Software de E/S Independente de Dispositivo



Interface uniforme para os drivers dos dispositivos

Armazenamento em buffer

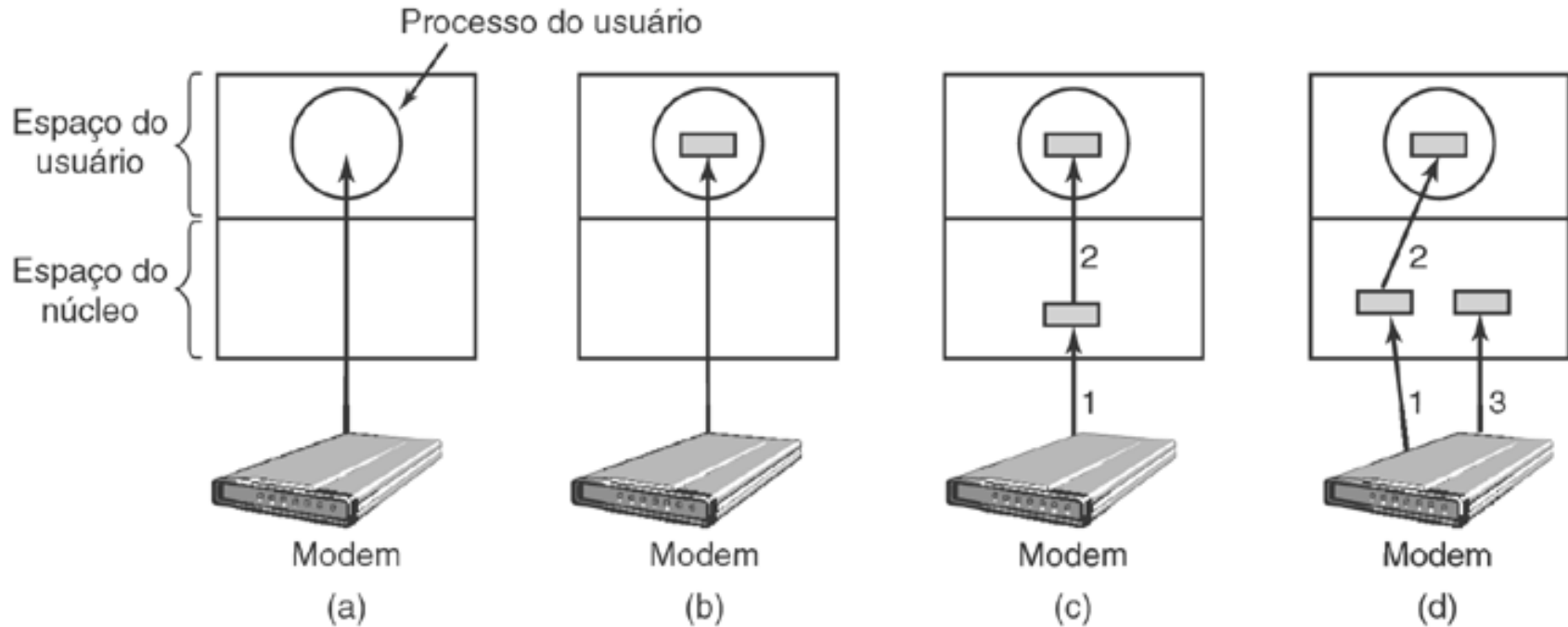
Relatório dos erros

Alocação e liberação de dispositivos dedicados

Fornecimento de tamanho de bloco independente
de dispositivo

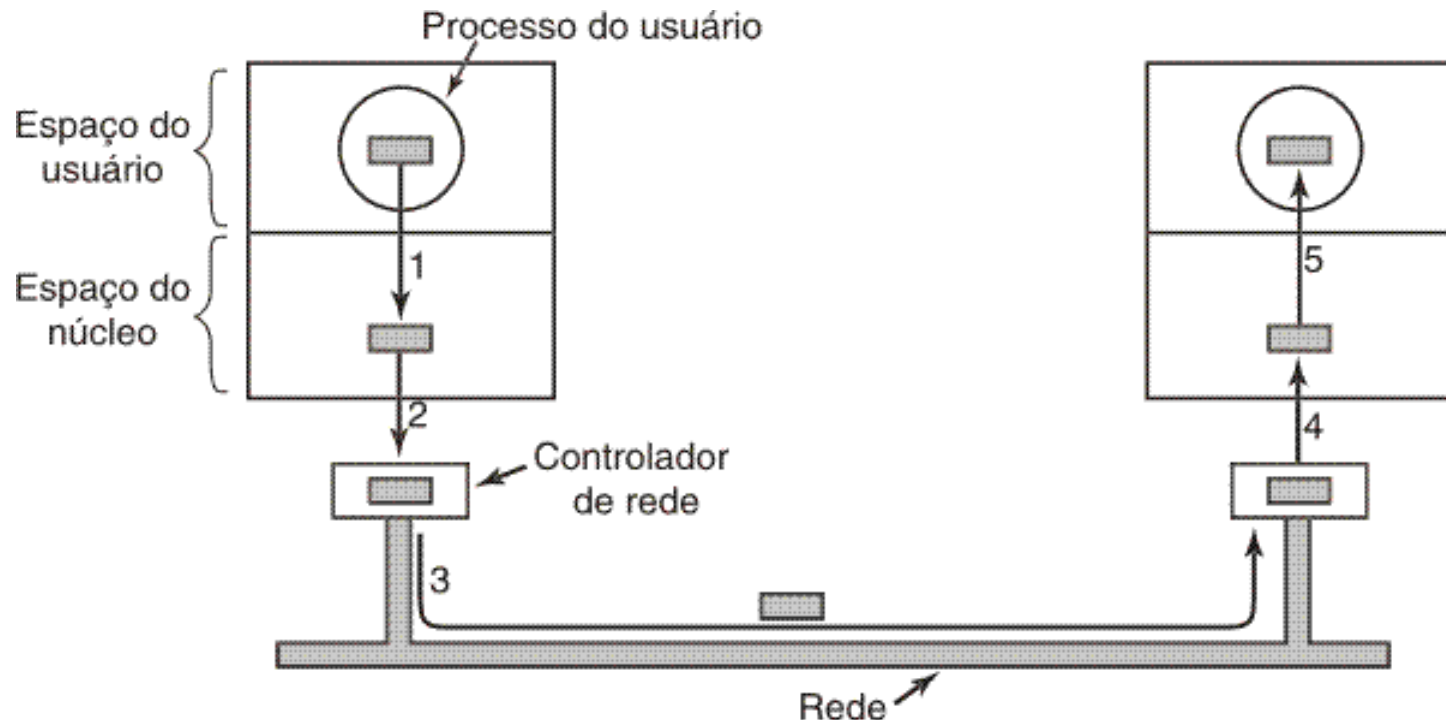
Funções do software de E/S independente de dispositivo

Software de E/S Independente de Dispositivo



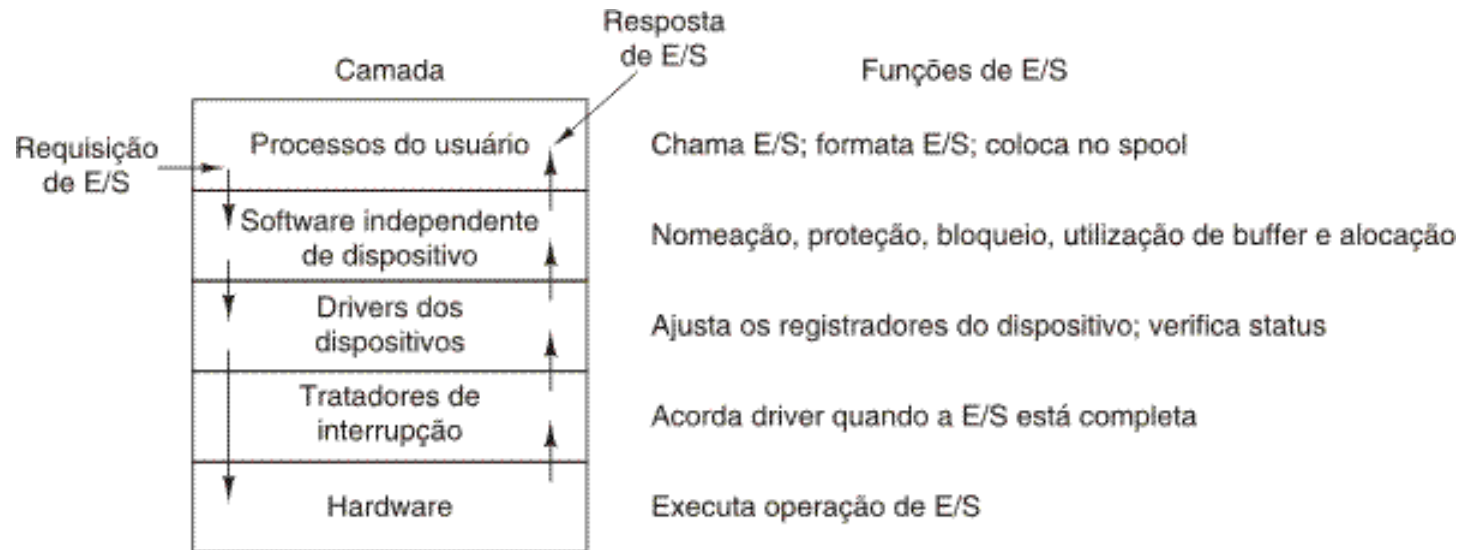
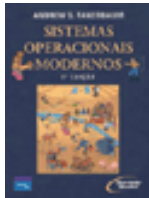
- a) Entrada sem utilização de buffer
- b) Utilização de buffer no espaço do usuário
- c) Utilização de buffer no núcleo seguido de cópia para o espaço do usuário
- d) Utilização de buffer duplo no núcleo

Software de E/S Independente de Dispositivo



A operação em rede pode envolver muitas cópias de um pacote

Software de E/S no Espaço do Usuário



Camadas do sistema de E/S e as principais funções de cada camada

Discos

Hardware do Disco



Parâmetro	Disco flexível IBM 360 KB	Disco rígido WD 18300
Número de cilindros	40	10 601
Trilhas por cilindro	2	12
Setores por trilha	9	281 (avg)
Setores por disco	720	35 742 000
Bytes por setor	512	512
Capacidade do disco	360 KB	18,3 GB
Tempo de posicionamento (cilindros adjacentes)	6 ms	0,8 ms
Tempo de posicionamento (caso médio)	77 ms	6,9 ms
Tempo de rotação	200 ms	8,33 ms
Tempo de pára/inicia do motor	250 ms	20 s
Tempo de transferência para um setor	22 ms	17 μ s

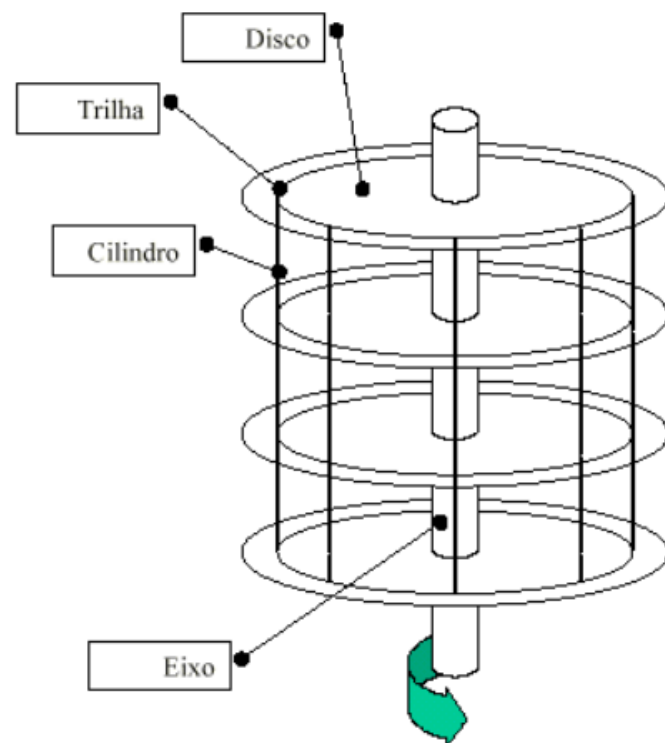
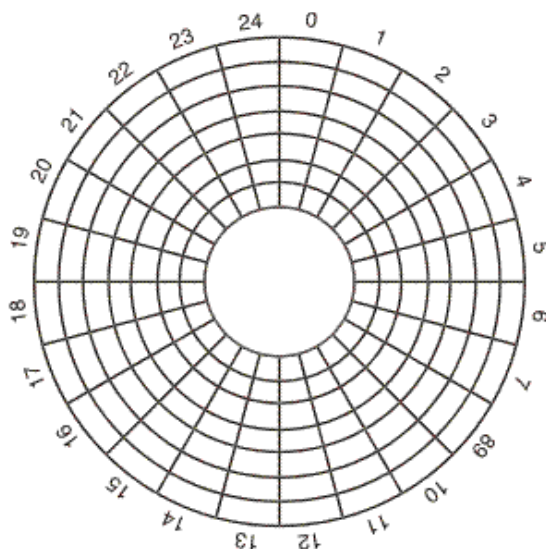
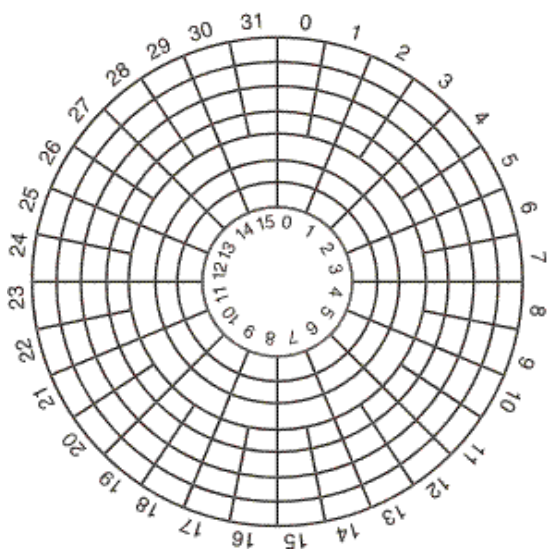
Parâmetros de disco para o disco flexível original do IBM PC e o disco rígido da Western Digital WD 18300

Discos Hardware do Disco



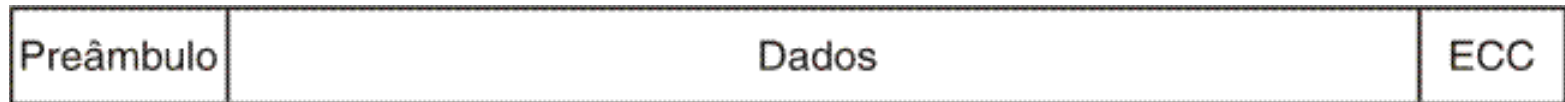
Especificações	4 TB	3 TB	2 TB	1 TB	500 GB
Número do modelo	WD4001FAEX	WD3001FAEX	WD2002FAEX	WD1002FAEX	WD5003AZEX
Interface	SATA 6 Gb/s	SATA 6 Gb/s	SATA 6 Gb/s	SATA 6 Gb/s	SATA 6 Gb/s
Capacidade formatada ¹	4.000.787 MB	3.000.592 MB	2.000.398 MB	1.000.204 MB	500.107 MB
Setores de usuários por unidade	7.814.037.168	5.860.533.168	3.907.029.168	1.953.525.169	976.773.168
Fator de forma	3,5 polegadas (6,3 cm)	3,5 polegadas (6,3 cm)	3,5 polegadas (6,3 cm)	3,5 polegadas (6,3 cm)	3,5 polegadas (6,3 cm)
Formato Avançado (FA)	Sim	Sim	Sim	Sim	Sim
Cumprir as normas RoHS ²	Sim	Sim	Sim	Sim	Sim
Desempenho					
Taxa de transferência de dados (máx)					
Buffer para host	6 Gb/s	6 Gb/s	6 Gb/s	6 Gb/s	6 Gb/s
Host para/desde drive (mantido)	154 MB/s	154 MB/s	138 MB/s	126 MB/s	150 MB/s
Cachê (MB)	64	64	64	64	64
Velocidade de rotação (RPM)	7200	7200	7200	7200	7200
Confiabilidade/Integridade dos dados					
Ciclos de carga/descarga ³	300.000	300.000	300.000	300.000	300.000
Erros de leitura irreversíveis por bits lidos	<1 em 10 ¹⁴	<1 em 10 ¹⁴	<1 em 10 ¹⁴	<1 em 10 ¹⁴	<1 em 10 ¹⁴
Garantia limitada (anos) ⁴	5	5	5	5	5
Gerenciamento de energia					
Requisitos médios de energia (W)					
Leitura/gravação	10,4	10,4	10,7	6,8	6,8
Ocioso	8,1	8,1	8,2	6,1	6,1
Standby/Descanso	1,2	1,2	1,3	0,7	0,8
Especificações ambientais⁵					
Temperatura (°C)					
Operacional	0 a 60	0 a 60	0 a 60	0 a 60	0 a 60
Não-operacional	-40 a 70	-40 a 70	-40 a 70	-40 a 70	-40 a 70
Choque (Gs)					
Operacional (2ms, gravação)	30	30	30	30	30
Operacional (2ms, leitura)	65	65	65	65	65
Não operacional (2ms)	300	300	300	300	350
Acústica (dBA) ⁶					
Ocioso	29	29	29	28	29
Pesquisa (média)	34	34	34	33	30
Dimensões físicas					
Altura (pol./mm, máx)	1,028/25,4	1,028/25,4	1,028/25,4	1,028/25,4	1,028/25,4
Comprimento (pol./mm, máx)	5,787/147	5,787/147	5,787/147	5,787/147	5,787/147
Largura (pol./mm, ± 0,01 pol.)	4/101,6	4/101,6	4/101,6	4/101,6	4/101,6
Peso (lb./kg, ± 10%)	1,72/0,78	1,72/0,78	1,66/0,75	1,52/0,69	0,97/0,44

Hardware do Disco



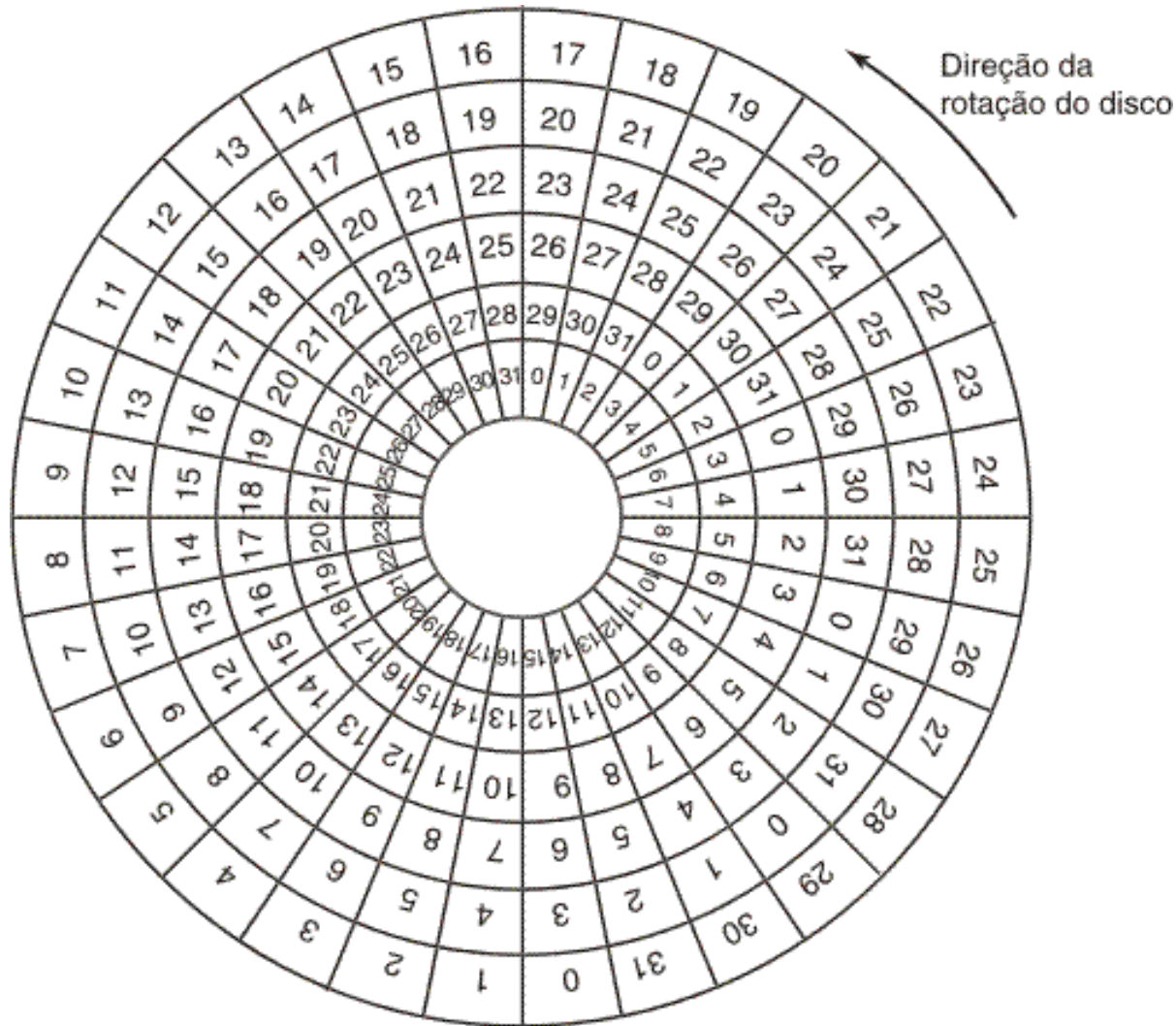
- Geometria física de um disco com duas zonas
- Uma possível geometria virtual para esse disco

Formatação de Disco



Um setor do disco

Formatação de Disco



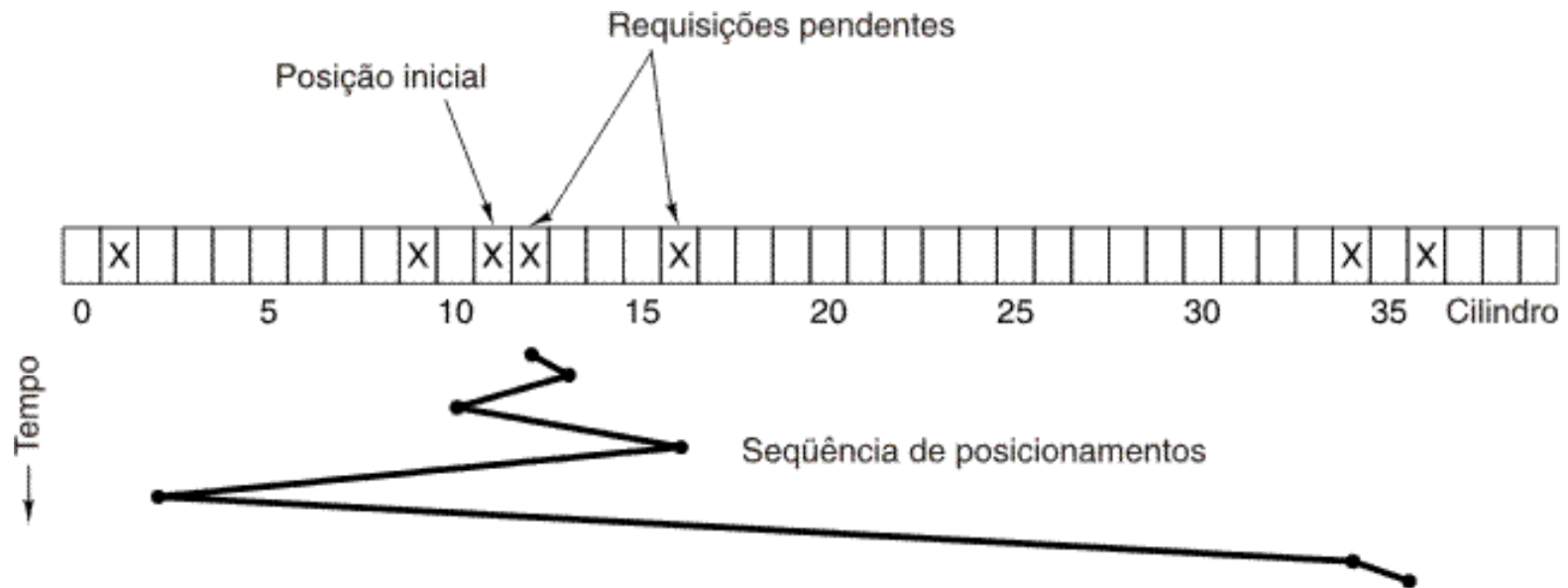
Uma ilustração da torção cilíndrica

Algoritmos de Escalonamento de Braço de Disco



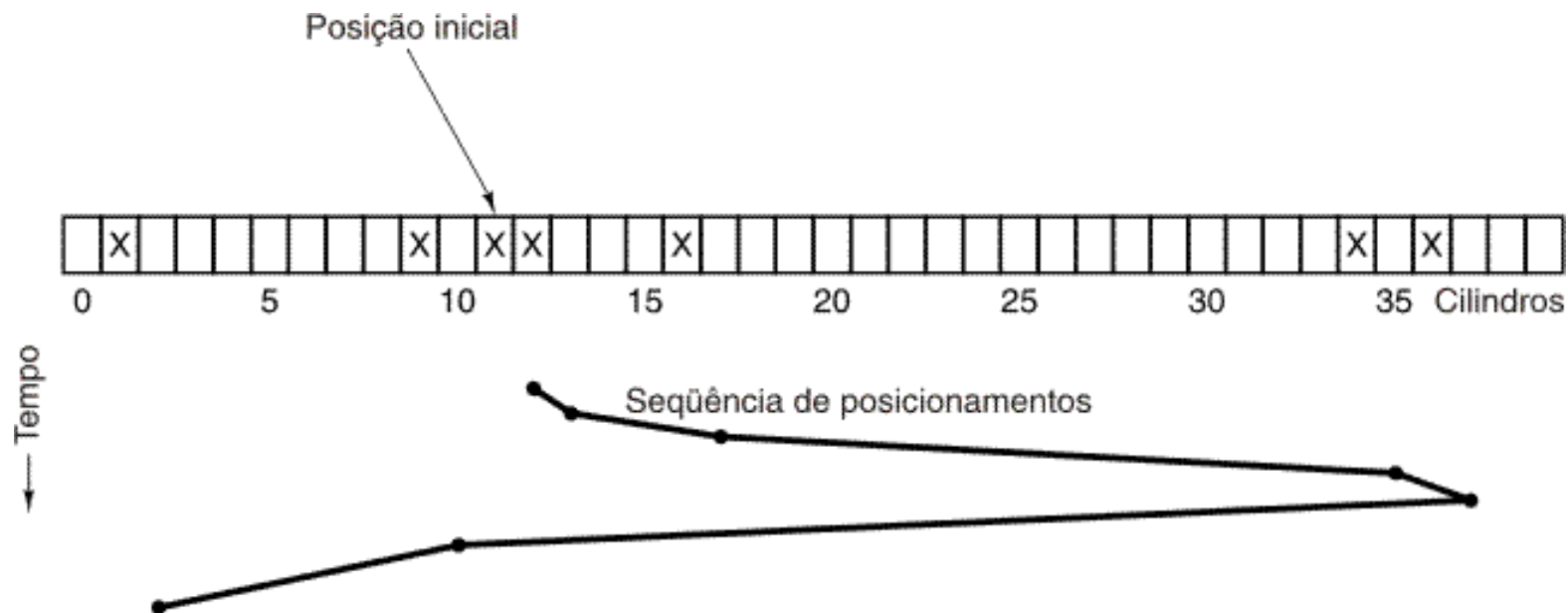
- Tempo necessário para ler ou escrever um bloco de disco é determinado por 3 fatores
 1. tempo de posicionamento
 2. atraso de rotação
 3. tempo de transferência do dado
- Tempo de posicionamento domina
- Checagem de erro é feita por controladores

Algoritmos de Escalonamento de Braço de Disco



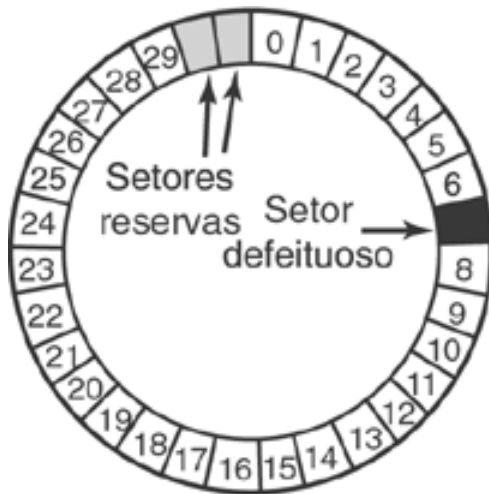
Algoritmo de escalonamento de disco *Posicionamento Mais Curto Primeiro (SSF)*

Algoritmos de Escalonamento de Braço de Disco

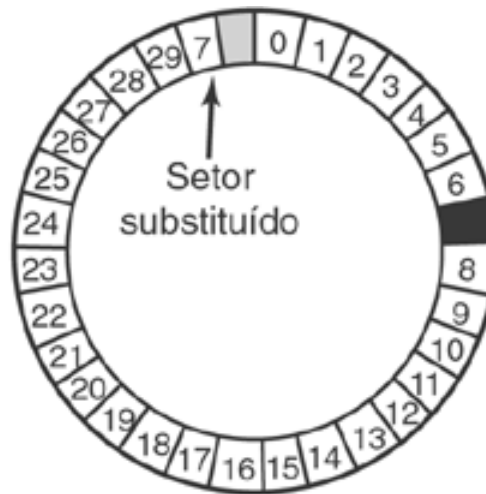


O algoritmo do elevador para o escalonamento das requisições do disco

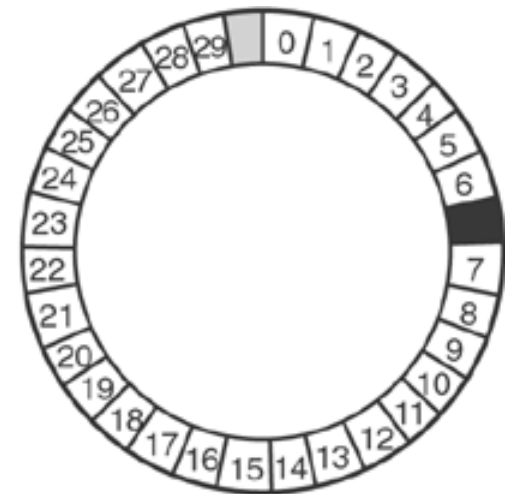
Tratamento de Erro



(a)



(b)



(c)

- a) Uma trilha de disco com um setor defeituoso
- b) Substituindo um setor reserva por um setor defeituoso
- c) Deslocando todos os setores para pular o setor defeituoso

Hardware do Disco



- Embora presente com quaisquer tipos de discos, a técnica RAID foi utilizada, inicialmente em discos SCSI
- **RAID** = **R**edundant **A**rray of **I**ndependent **D**isks
- Anteriormente, RAID significava **R**edundant **A**rray of **I**nexpensive **D**isks
- Há vários tipos de RAID, cada qual com sua finalidade

NAS (Network Attached Storage)

- Servidor NAS ou Network Attached Storage, com tradução livre do inglês "servidor de armazenamento conectado em rede" é uma unidade de armazenamento exclusiva para guardar e compartilhar arquivos de diversos computadores através de uma rede local.
- Também conhecidos como storages NAS, esses equipamentos podem ser encontrados montados em gabinetes externos ou em armários tipo rack, são compostos por um ou mais discos rígidos e possuem uma ou mais portas de comunicação para conexão em redes locais LAN.
- Geralmente utilizado para organizar, centralizar e fazer backup de dados, um servidor NAS pode ainda executar aplicações sofisticadas, como ser alvo virtualizado para diversos hosts numa grande rede corporativa, servir de servidor de nuvem de datacenters ou disponibilizar vários bancos de dados para centenas de usuários.

Hardware do Disco



Storage Synology NAS DiskStation Annapurna Labs Alpine AL-314 Quad Core 1.7GHz 4GB DDR3L

Tipo de RAID Suportado:

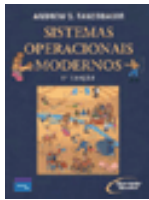
- - RAID híbrido Synology
- - Basic
- - JBOD
- - RAID 0
- - RAID 1
- - RAID 5
- - RAID 6
- - RAID 10



WD My Cloud Expert Series 56TB EX4100 4-Bay NAS Server (4 x 14TB)

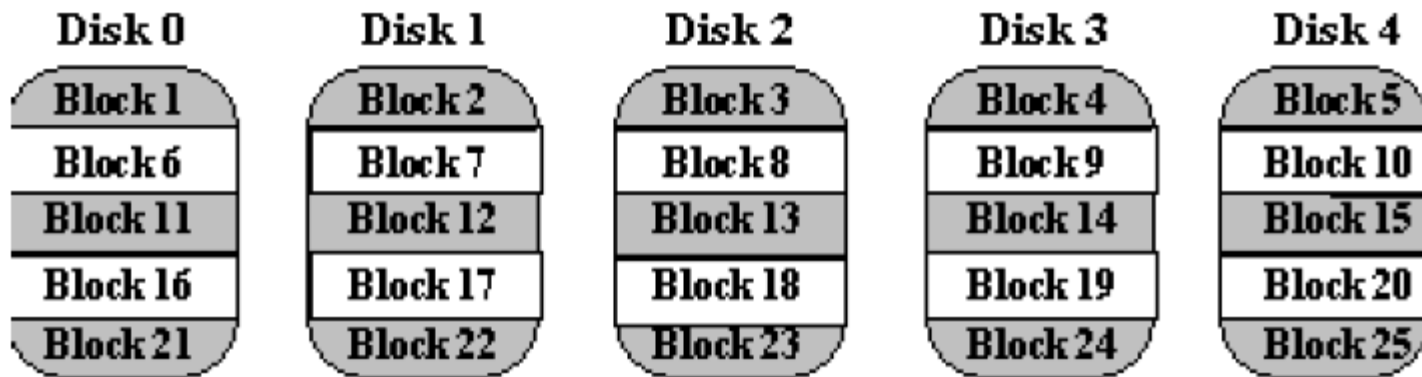
Multiple RAID and backup options, including RAID 0, 1, 5, 10, hot spare, JBOD, and SPAN help to protect all your valuable media, photos, and important files.

Hardware do Disco



RAID 0

- Voltado para melhoria do desempenho
- Dados escritos em seções sequenciais do disco
- Vários dispositivos acessados de uma só vez
- Não é voltado para tolerância à falhas

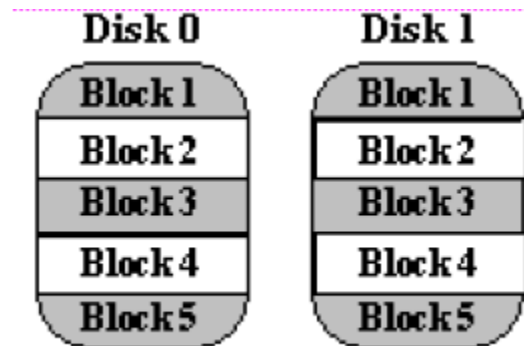


Hardware do Disco



RAID 1

- Dados são escritos em um ou mais discos
- Também chamado de espelhamento
- Redundância prove tolerância a falhas
- Desempenho ruim se comparado ao RAID 0

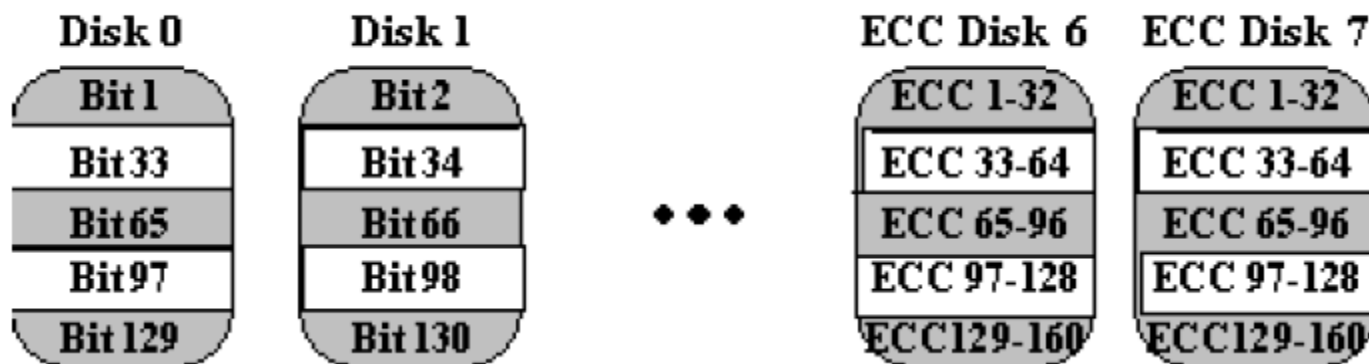


Hardware do Disco



RAID 2

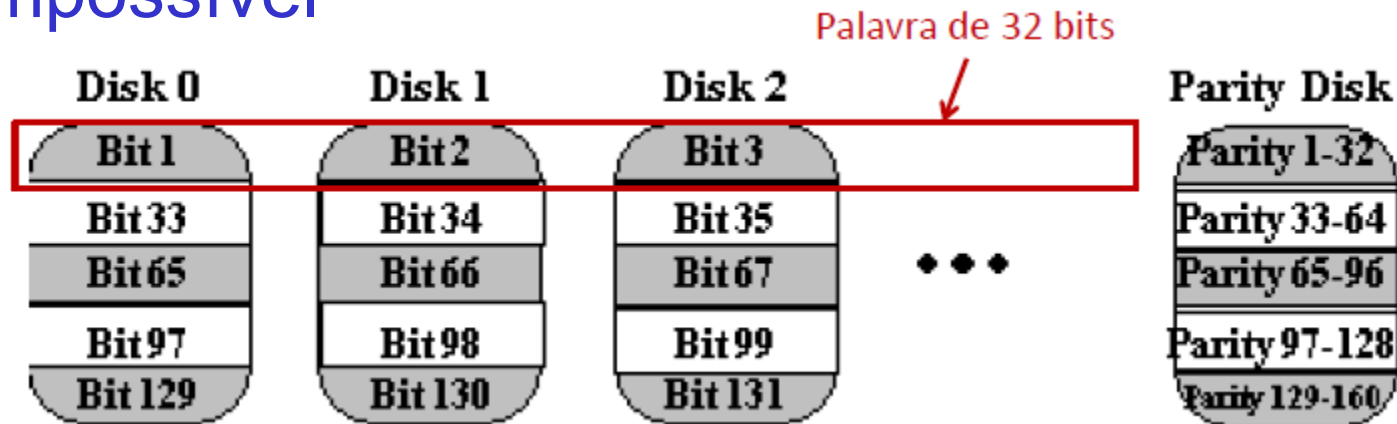
- Dados são escritos em seções sequenciais dos discos, no nível de bit
- Discos extras contêm Códigos de Hamming, no nível de bit, para detecção e correção de erros
- São necessários vários discos de ECC



Hardware do Disco

RAID 3

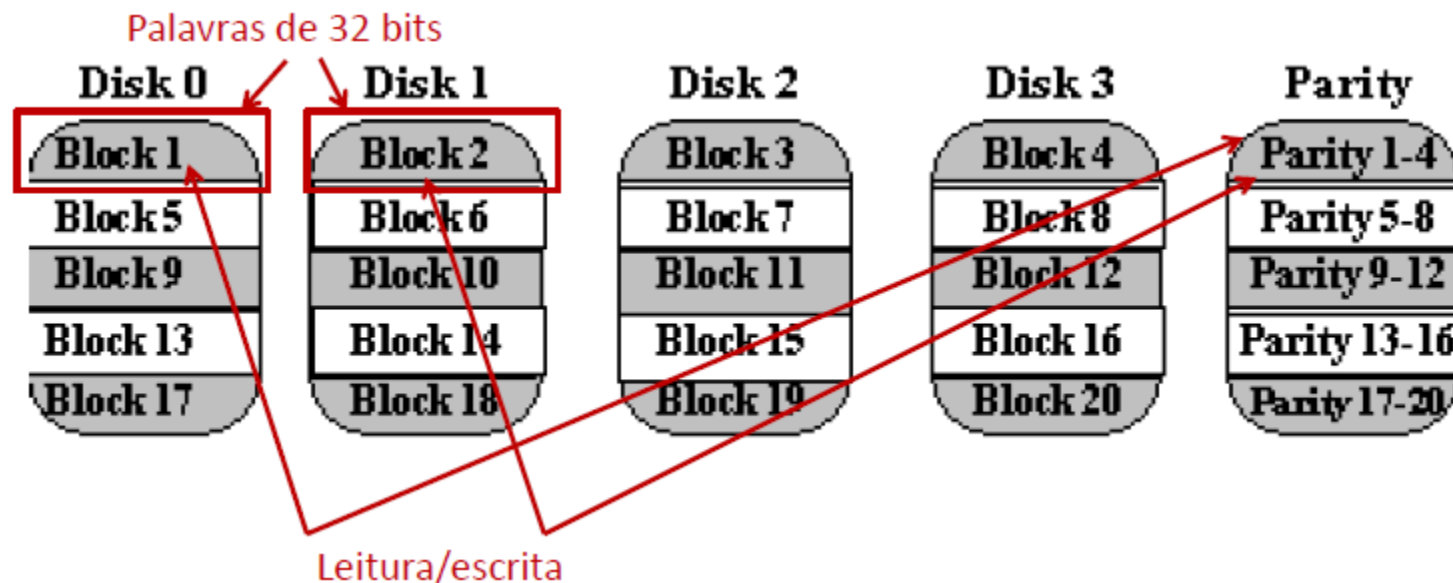
- Dados são escritos em seções sequenciais dos discos, no nível de bit
- Apenas um disco extra com bit de paridade
- Atendimento a solicitações simultâneas (leitura/escrita de duas ou mais palavras) é impossível



Hardware do Disco

RAID 4

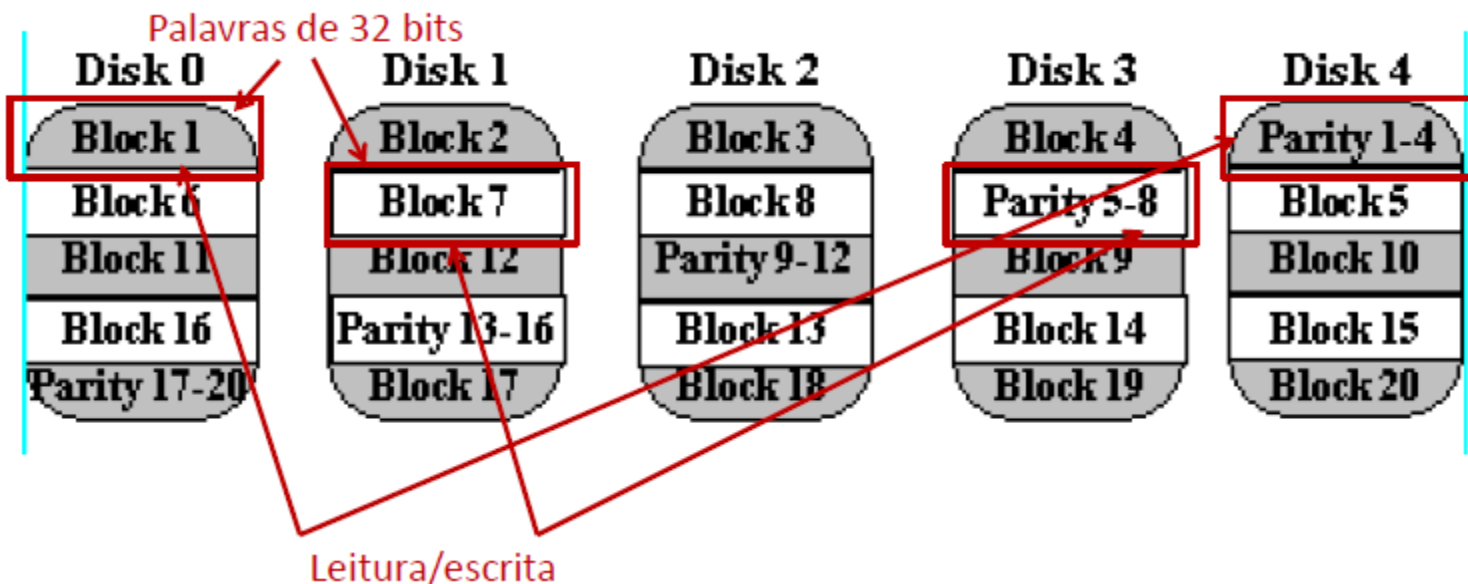
- Similar ao RAID 3, mas os discos são organizados em nível de bloco
- Disco de paridade é um gargalo para atendimento de solicitações simultâneas



Hardware do Disco

RAID 5

- Bits de paridade são distribuídos de maneira uniforme por todos os discos
- Diminui a ocorrência de gargalos no atendimento a solicitações simultâneas

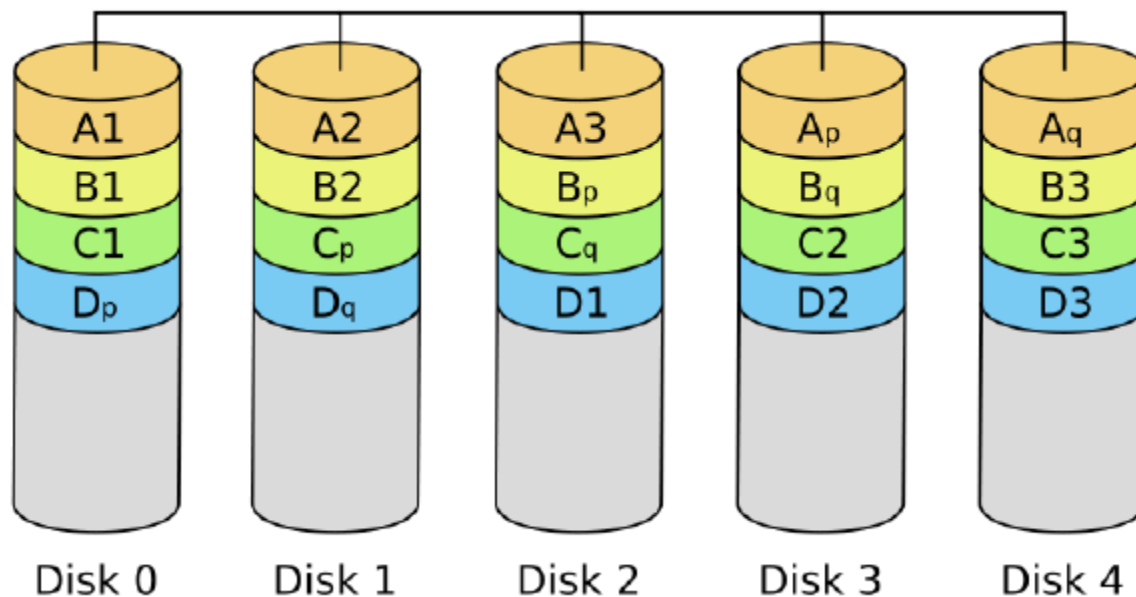


Hardware do Disco

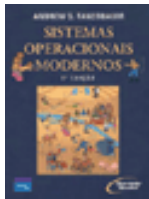


RAID 6

- Provê bloco de paridade adicional
- Usa códigos Reed-Solomon para proteção de erros em dois discos

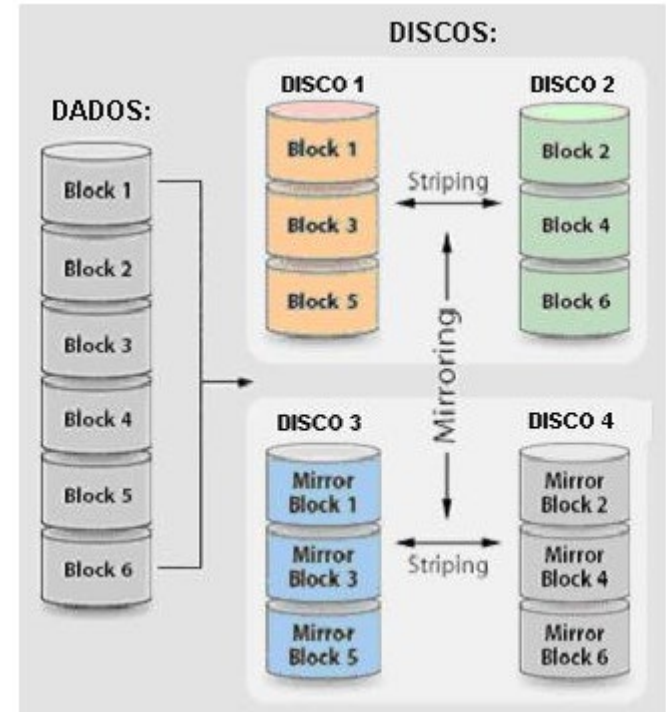


Hardware do Disco



RAID 0 + 1

- Os dados são espelhados através de grupos de discos segmentados
 - Dados primeiro segmentados e depois espelhados
 - No exemplo com 4 discos, se faz a gravação segmentada nos discos 1 e 2 para se fazer o espelho nos discos 3 e 4, na sequência
 - Se perder um disco em um dos lados o outro assume, tornando-se, praticamente, um RAID 0
 - Fragilidade no tempo entre gravação dos segmentos e espelhamento

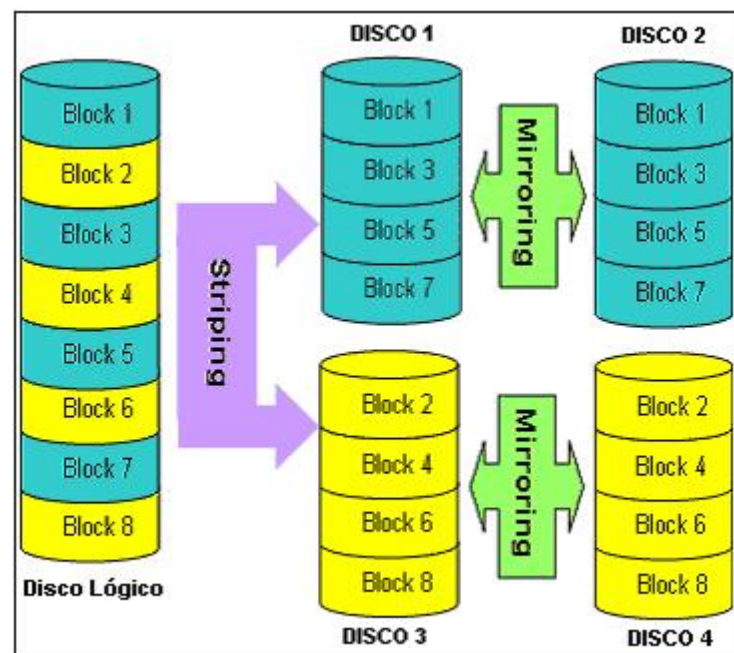


Hardware do Disco



RAID 1 + 0

- Os dados são segmentados através de grupos de discos espelhados
 - Dados primeiro espelhados e depois segmentados
 - No exemplo com 4 discos, se faz a gravação do disco 1 e o espelho no disco 3, para se fazer o mesmo entre os discos 2 e 4
 - Falhas nas gravações dos discos 1 e 3 mantém a estrutura intacta
 - Mais seguro que o RAID 0 + 1



JBOD – Just a Bunch of Disks

- Embora seja relacionado ao RAID, o JBOD não é um modo RAID, mas também é bastante usado, sobretudo em servidores de arquivos. No JBOD (Just a Bunch Of Disks) os HDs disponíveis são simplesmente concatenados e passam a ser vistos pelo sistema como um único disco, com a capacidade de todos somada. Os arquivos são simplesmente espalhados pelos discos, com cada um armazenando parte dos arquivos (neste caso arquivos completos, e não fragmentos como no caso do RAID 0).
- No JBOD não existe qualquer ganho de desempenho, nem de confiabilidade. Caso um dos HDs apresente defeito, os arquivos armazenados nele são perdidos, mas os arquivos armazenados nos demais continuam intactos. Na verdade, o único ganho é o de praticidade, com a possibilidade de usar vários discos para formar um único volume de grande capacidade, ao invés de ter que espalhar os arquivos e pastas entre os vários HDs.
- Ao contrário dos outros modos RAID, não existe nenhum problema em combinar HDs com capacidades e desempenho variados num sistema JBOD. Cada HD pode dar sua parcela de contribuição, independentemente de sua capacidade.

Hardware do Disco

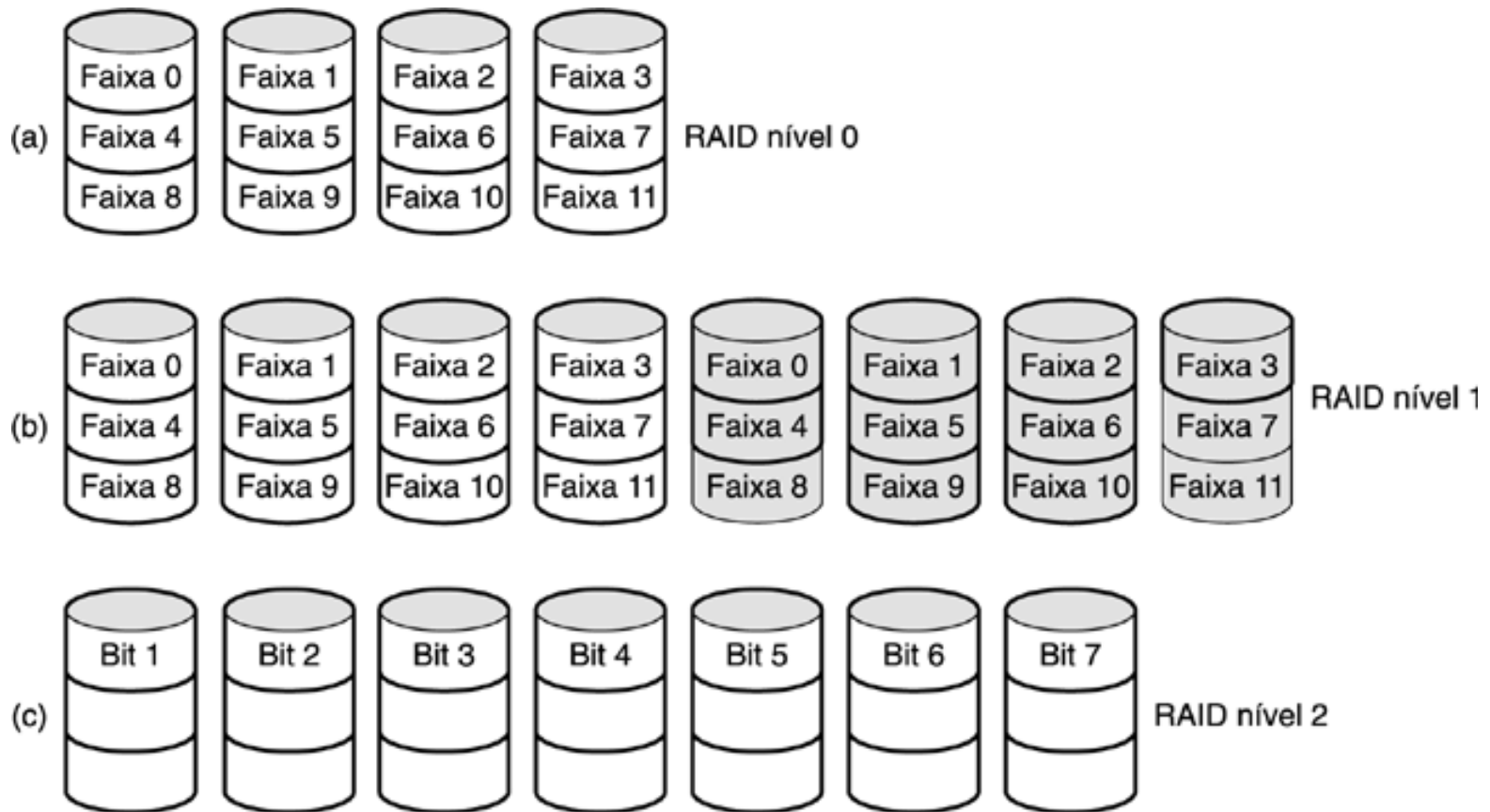
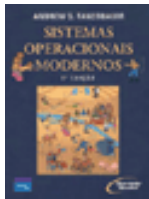


RAID

It is a technique that combines multiple disk drives into a logical unit (RAID set) and provides protection, performance, or both.

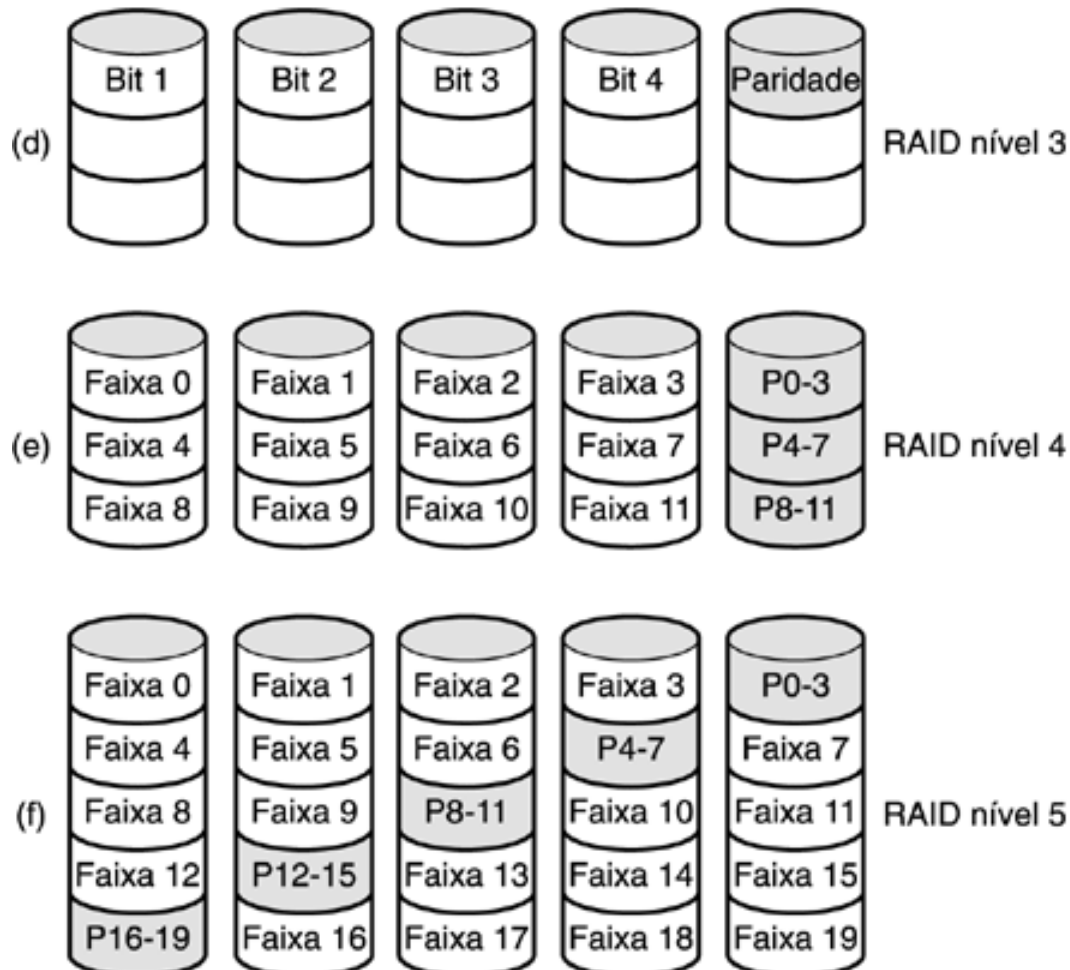
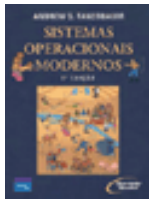
RAID level	Min disks	Available storage capacity (%)	Read performance	Write performance	Write penalty	Protection
1	2	50	Better than single disk	Slower than single disk, because every write must be committed to all disks	Moderate	Mirror
1+0	4	50	Good	Good	Moderate	Mirror
3	3	$[(n-1)/n]*100$	Fair for random reads and good for sequential reads	Poor to fair for small random writes fair for large, sequential writes	High	Parity (Supports single disk failure)
5	3	$[(n-1)/n]*100$	Good for random and sequential reads	Fair for random and sequential writes	High	Parity (Supports single disk failure)
6	4	$[(n-2)/n]*100$	Good for random and sequential reads	Poor to fair for random and sequential writes	Very High	Parity (Supports two disk failures)

Hardware do Disco



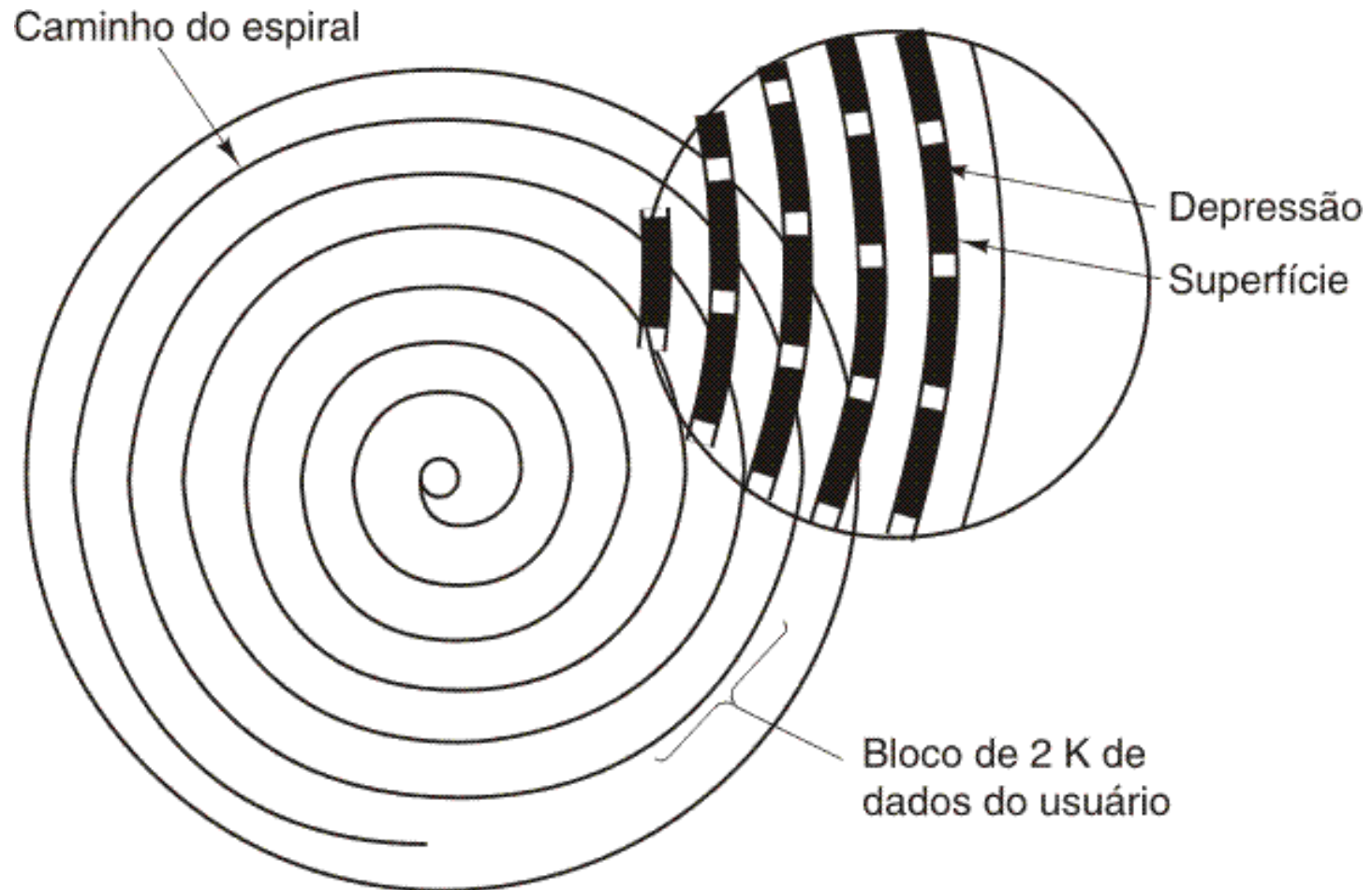
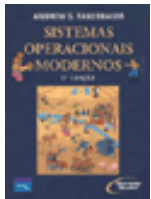
- RAIDs níveis 0 a 2
- Discos de segurança e de paridade são os sombreados

Hardware do Disco



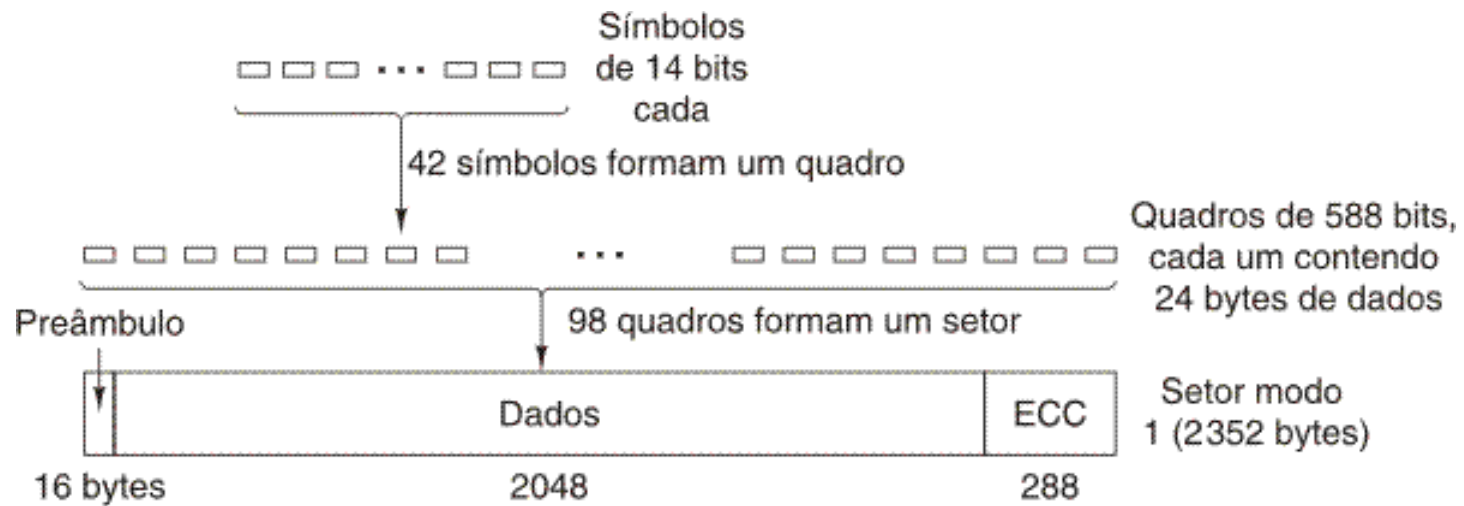
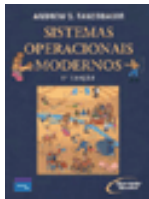
- RAIDs níveis 3 a 5
- Discos de segurança e de paridade são os sombreados

Discos Óticos



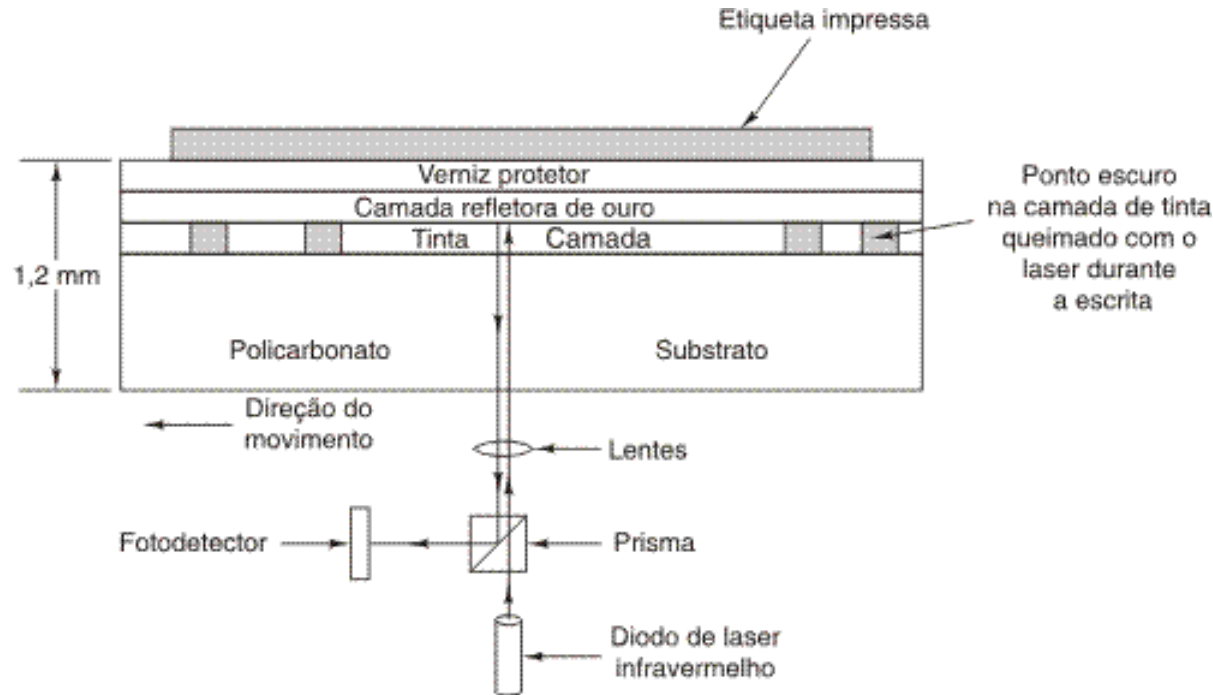
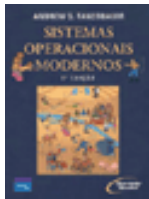
Estrutura de gravação de um CD ou CD-ROM

Discos Óticos



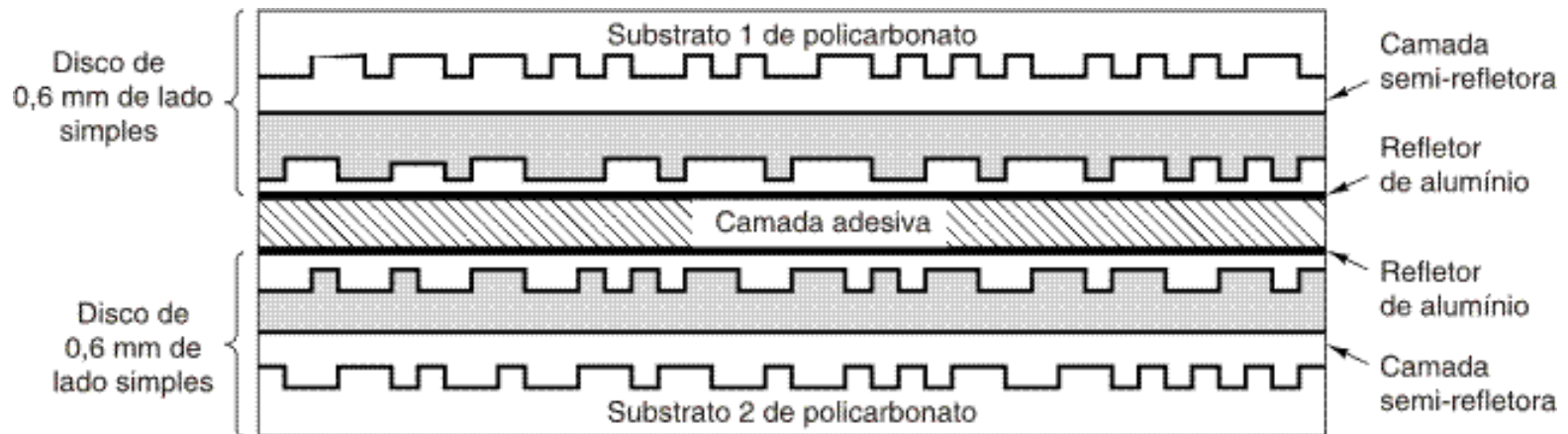
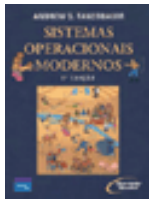
Esquema lógico dos dados em um CD-ROM

Discos Óticos



- Secção transversal de um disco CD-R e laser
 - sem escala
- CD-ROM prateado tem estrutura similar
 - sem camada de tinta
 - com camada de alumínio em vez de ouro

Discos Óticos



Disco DVD com lado duplo e camada dupla