Como surgiu o RAID

Quando interpretado literalmente, isso significa armazenar informações em uma variedade de unidades de disco rígido de baixo custo (HDDs).
A tecnologia foi inicialmente proposta por um grupo de estudo do professor. Patterson na Universidade da Califórnia. Quando apresentaram seu trabalho, o “I" no RAID representava “Inexpensive“(barato). Isso porque eles estavam procurando maneiras de usar os HDDs baratos que estavam no mercado, em seu estudo, em vez dos dispositivos de disco proprietários e muito caros, normalmente usados naquele momento, para melhorar a velocidade e a confiabilidade da unidade de disco.

O que é RAID?

RAID é um acrônimo para "Redundant Array of Inexpensive Disks".

Quando interpretado literalmente, isso significa armazenar informações em uma variedade de unidades de disco rígido de baixo custo (HDDs).

É geralmente considerado como "Tecnologia que combina um grande número de HDDs de baixo custo em um único HDD". RAID é o uso de vários discos para gerenciar dados de HDD usando uma variedade de técnicas diferentes. Estes são tipicamente divididos em 6 a 7 níveis; RAID 0, RAID 1, RAID 2, RAID 3, RAID 4, RAID 5, RAID6. Todas elas diferem em termos de implantação de dados e do tipo de redundância oferecida.

A tecnologia RAID não apenas evita a perda e a falha de dados, mas também melhora o desempenho dos negócios.

RAID

Basicamente existem três formas de se gerar um RAID, através de :

- Software;
- Hardware;
- Software e Hardware.

RAID de Software

O RAID de software pode ser implementado por meio de recursos que combinam vários dispositivos de disco conectados diretamente a um computador host (normalmente por meio de uma interface SCSI) e os considera como um único dispositivo de memória lógica. Este recurso introduzido com os sistemas operacionais Windows NT / 2000 é comumente usado por ser fácil e barato de implementar.

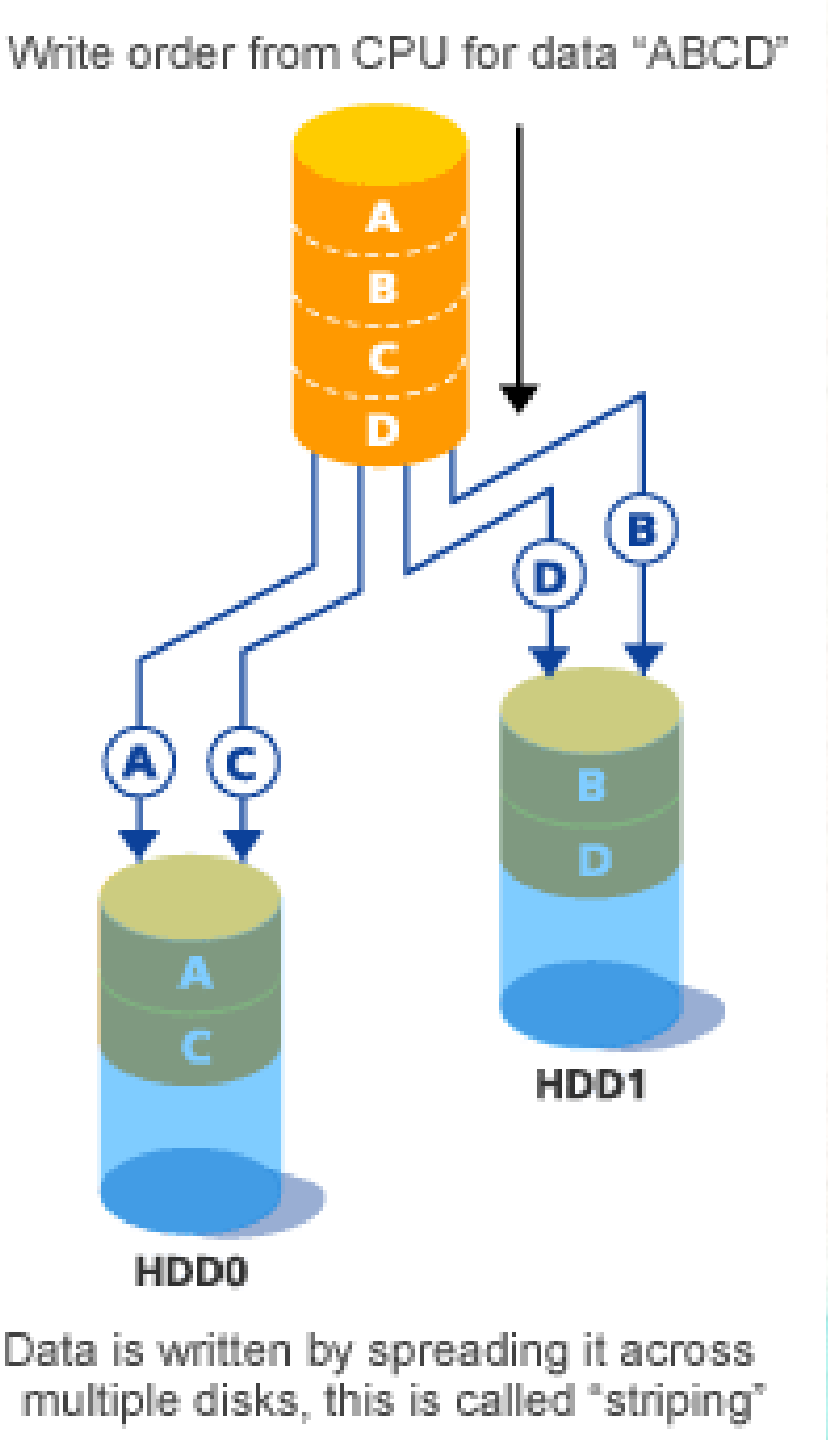
RAID de Hardware

Com Hardware RAID, um componente de controle, independente da CPU do host, implementa o RAID. Os dois métodos mais populares de Hardware RAID são via conexão de barramento PCI para o computador host usando um cartão, ou integrados com a unidade de disco e conectados ao computador host via Fibre-Channel ou SCSI.

O RAID de hardware é, de longe, o método mais comum em sistemas de servidores completos, pois não sobrecarrega o processamento do servidor.

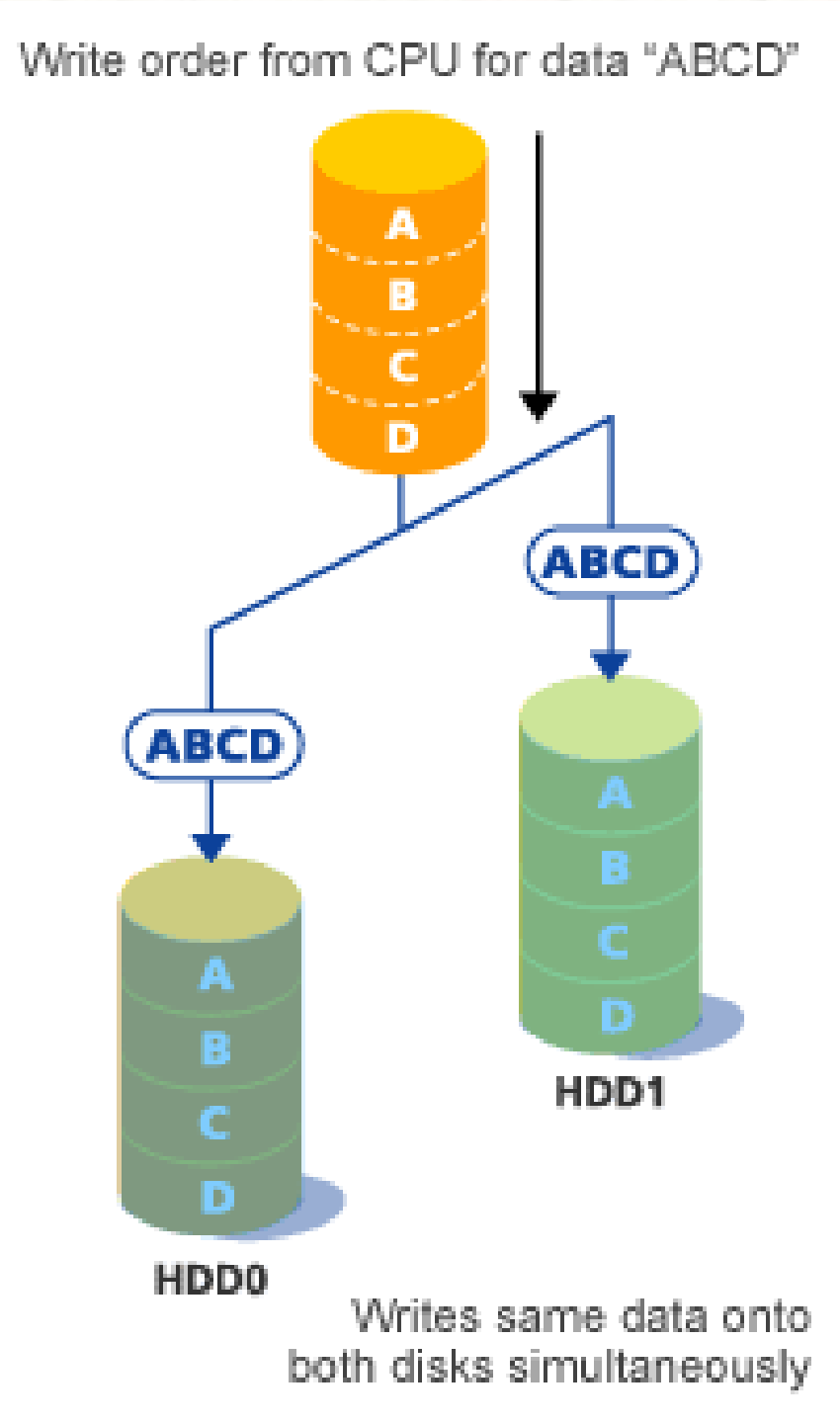
RAID 0

O RAID 0 divide os dados em unidades de bloco e os grava de maneira dispersa em vários discos. Como os dados são colocados em todos os discos, também são chamados de "striping". Este processo permite um desempenho de alto nível, pois o acesso paralelo aos dados em diferentes discos melhora a velocidade de recuperação. No entanto, nenhum recurso de recuperação é fornecido se ocorrer uma falha no disco. Se um disco falhar, ele afeta tanto as leituras quanto as gravações, e à medida que mais discos são adicionados à matriz, maior a possibilidade de ocorrer uma falha no disco.



RAID 1

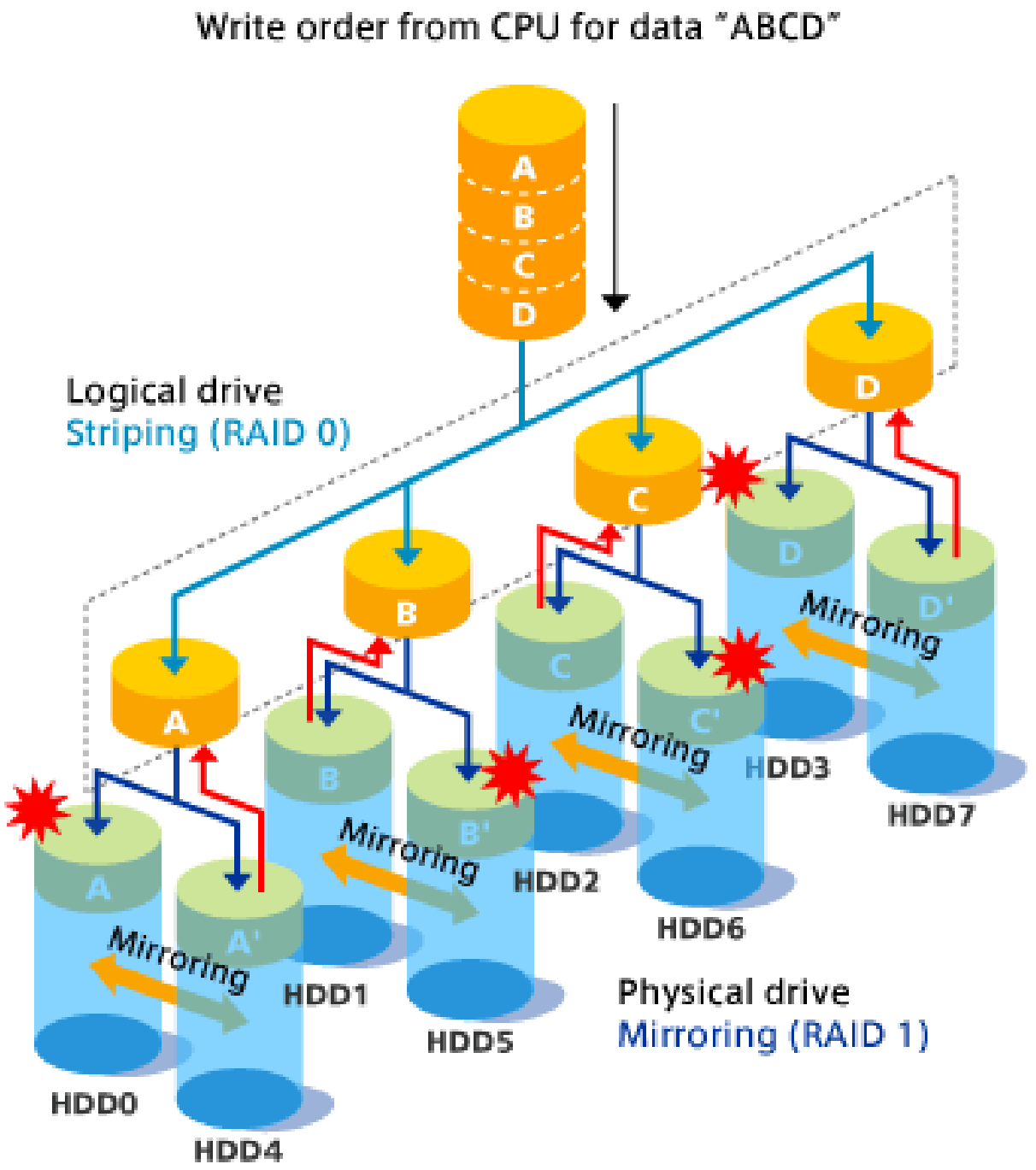
Esse nível é chamado de "espelhamento", pois copia os dados em duas unidades de disco simultaneamente.
Embora não haja aprimoramento nas velocidades de acesso, a duplicação automática dos dados significa que há pouca probabilidade de perda de dados ou tempo de inatividade do sistema. O RAID 1 fornece tolerância a falhas. Se um disco falhar, o outro assumirá automaticamente e a operação contínua será mantida. Não há melhoria no desempenho do custo de armazenamento, pois a duplicação de todos os dados significa que apenas metade da capacidade total do disco é capaz de armazenar.



RAID 1 + 0

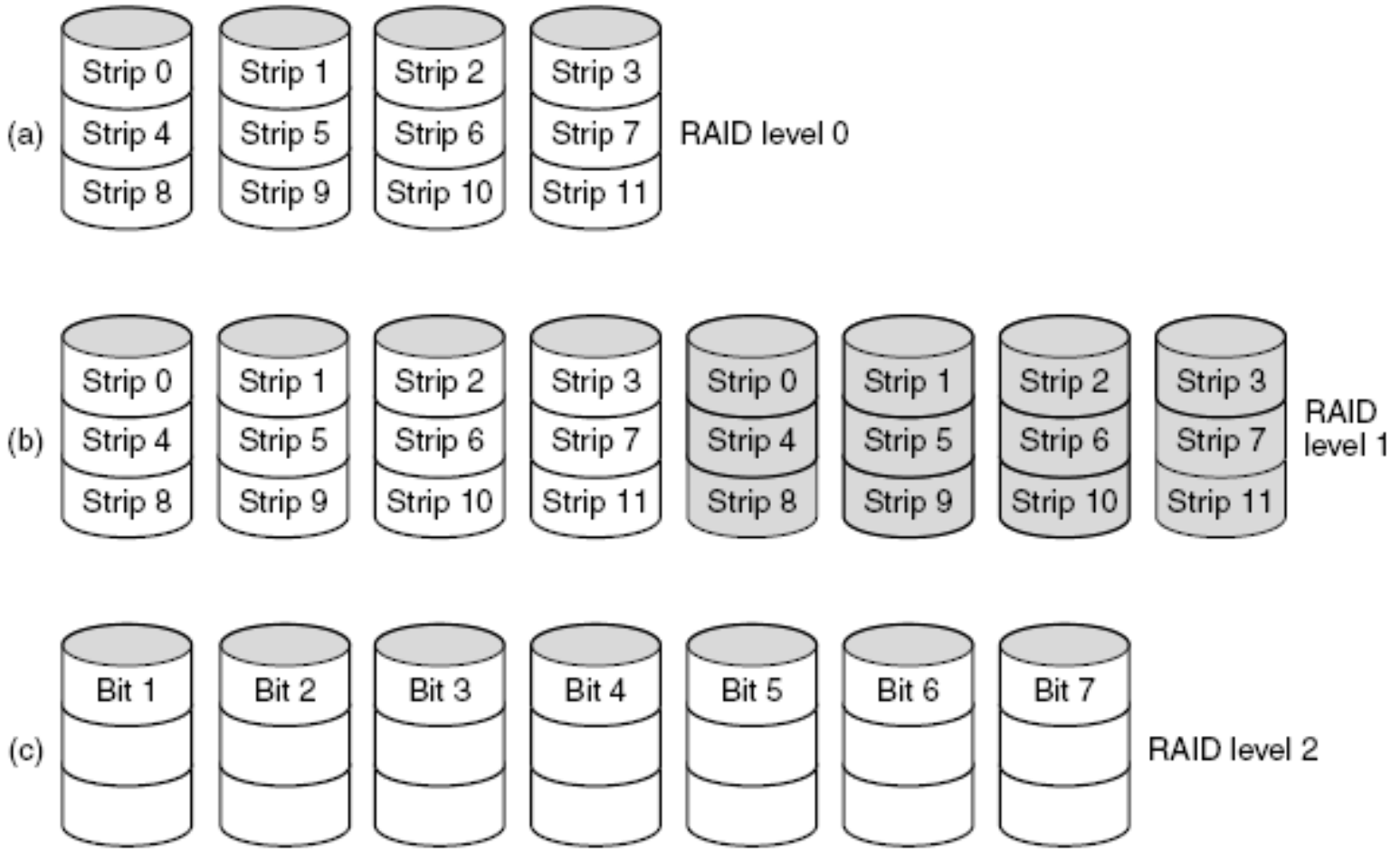
RAID 1 + 0 combina RAID 0 e RAID 1.

Ao configurar ambas as tecnologias em um único array, a duplicação de dados e a velocidade de acesso aprimorada podem ser fornecidas. Embora essa combinação torne a instalação mais cara em comparação com outras tecnologias, a confiabilidade e o alto desempenho de E/S podem ser garantidos. Isso ocorre porque uma única falha no disco não impede a distribuição para outros discos.



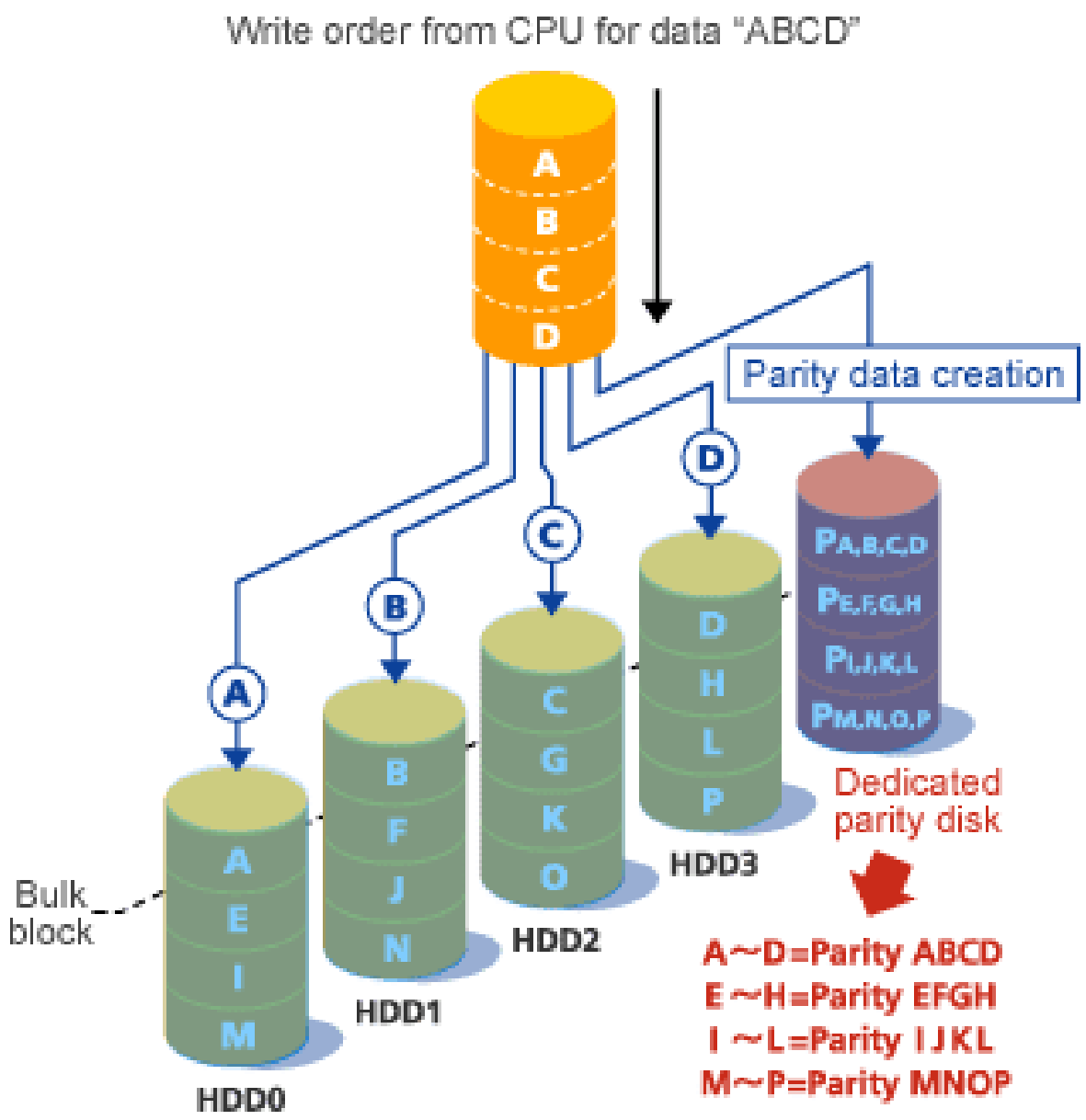
RAID 2

O RAID 2 emprega o uso de Códigos de Correção de Erros (ECC - Error Correction Codes), também chamados de Hamming Codes (em homenagem a Richard Hamming, da Bell Labs, que os inventou). Eles fornecem a capacidade de procurar e corrigir erros nos dados. Os dados são divididos em unidades de bits ou bytes e mantidos em várias unidades de dados dedicadas. Na prática, no entanto, o RAID 2 é pouco usado, pois é inferior a outros níveis de RAID em termos de custo e



RAID 3

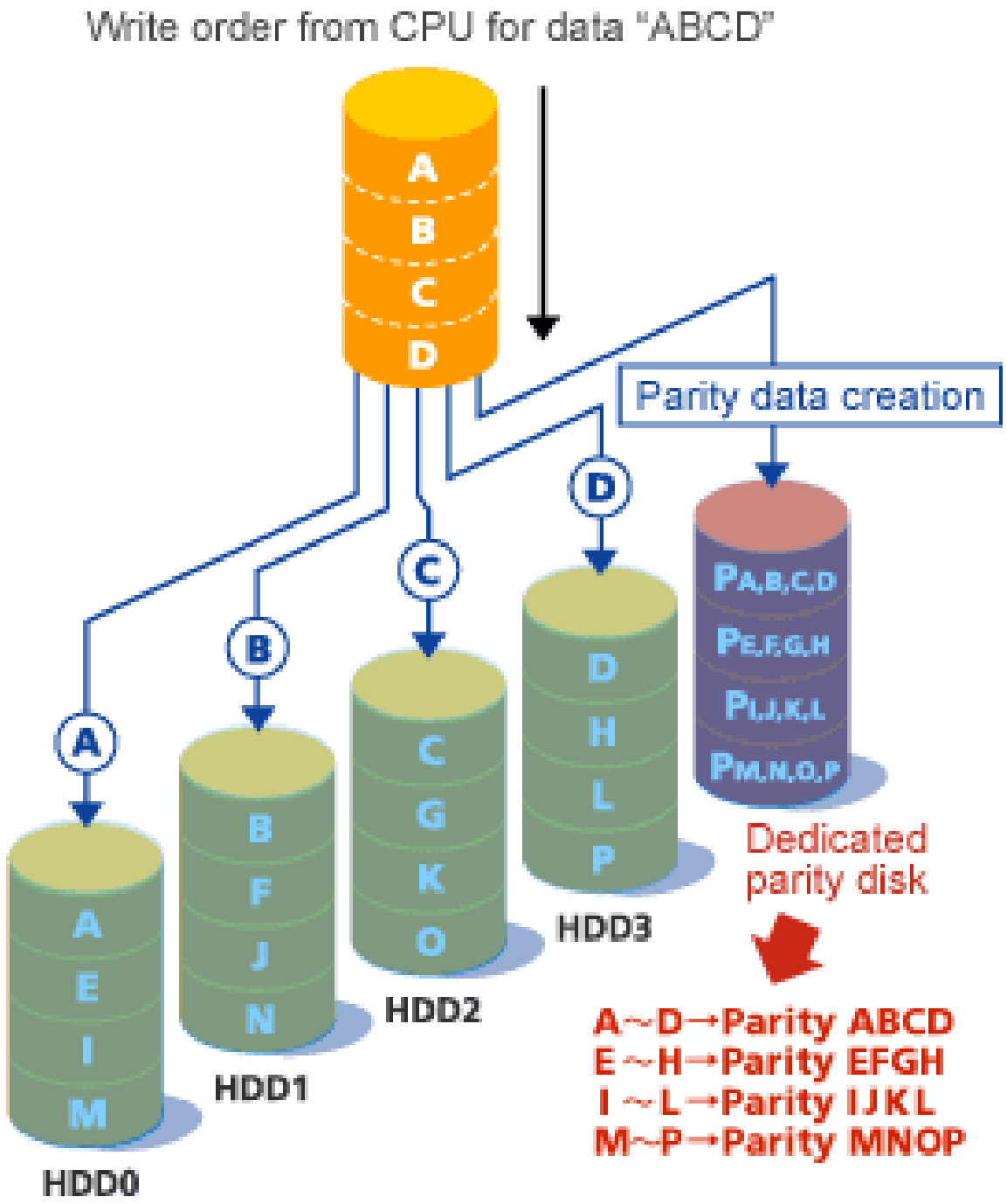
Com o RAID 3, os dados são divididos em unidades de bits ou bytes e gravados em várias unidades de disco de dados dedicadas. As informações de paridade são criadas para cada seção de dados separada e gravadas em uma unidade de paridade dedicada. Todas as unidades de disco podem ser acessadas em paralelo o tempo todo e os dados podem ser transferidos em massa, garantindo a transferência de dados em alta velocidade



Parallel bulk access across the striping width
("ABCD" in the figure above)
enables high-speed data transfer

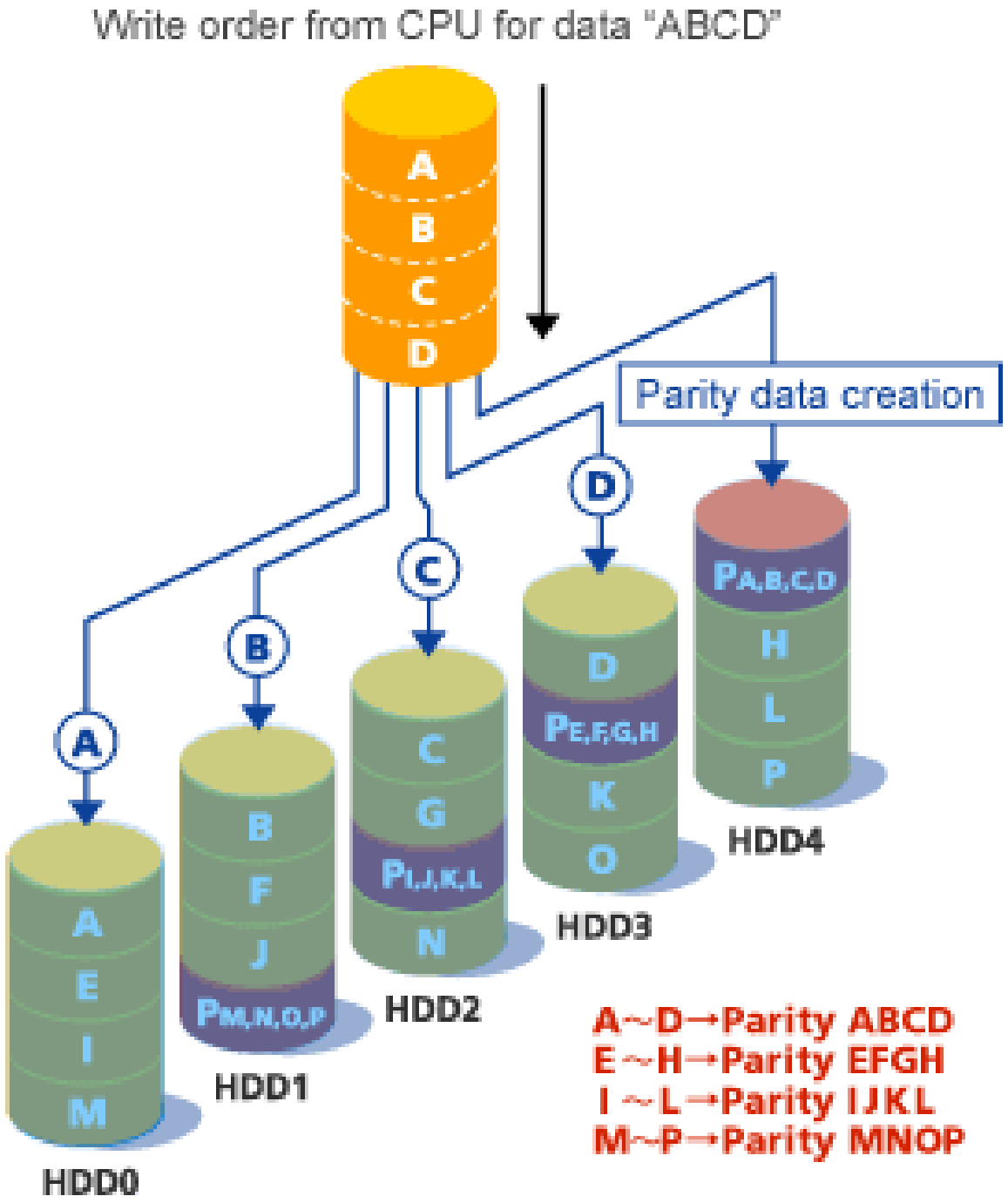
RAID 4

O RAID 4 apresenta recriação de dados por meio de uma combinação de striping de RAID 0 e o uso de um disco de paridade dedicado.
Os dados são divididos em blocos de unidades e mantidos em unidades de disco de dados dedicadas, enquanto os dados de paridade são mantidos em um único disco de paridade dedicado. Ao atualizar, é necessário pré-ler os dados existentes e de paridade e gravar os dados de paridade atualizados quando a atualização for concluída. Esse processo é chamado de "penalidade de gravação". Geralmente, é impossível implementar esse processo em ambientes de negócios, pois o disco de paridade dedicado torna-se um gargalo durante surtos de tráfego e o desempenho sofre.



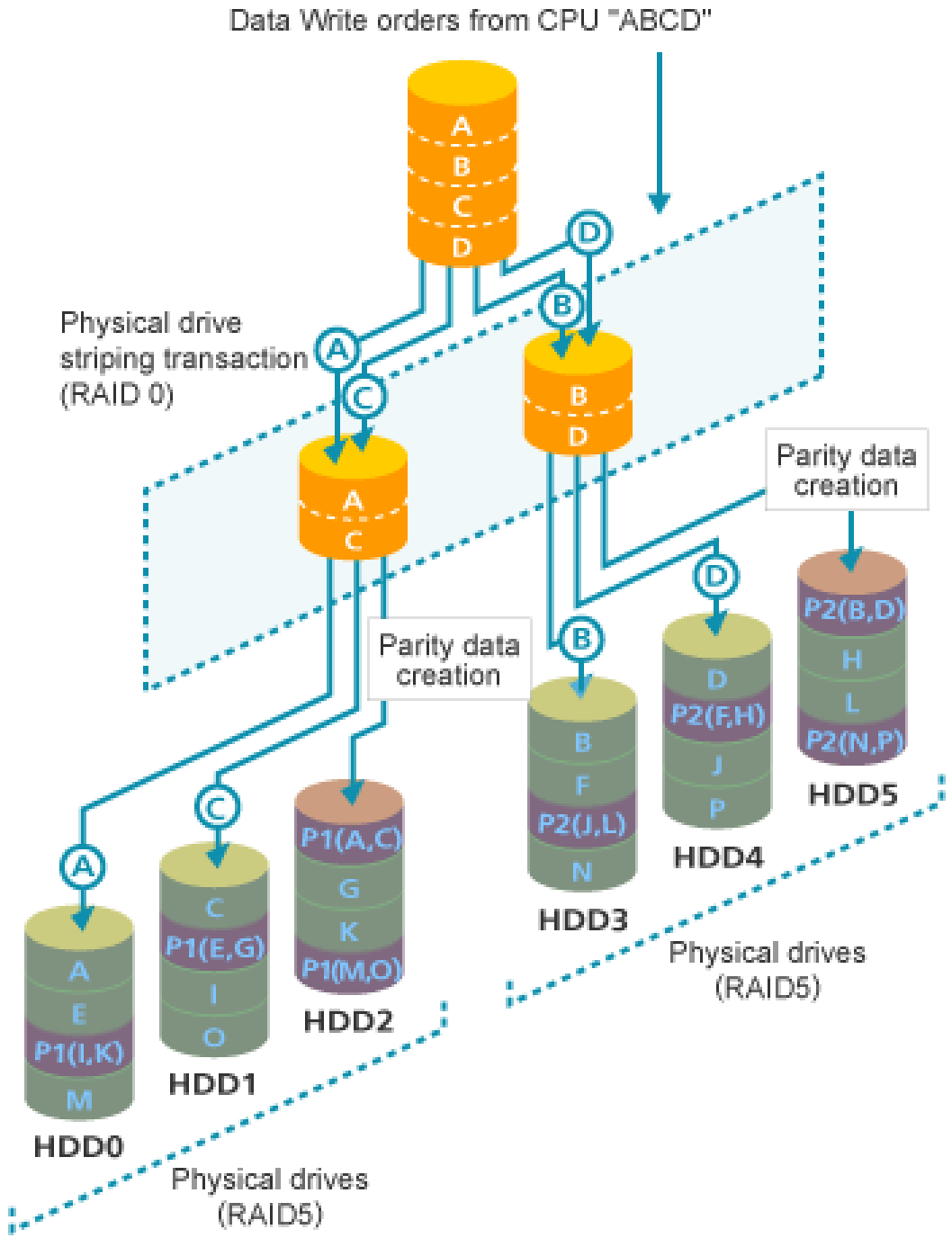
RAID 5

O RAID 5 é a tecnologia RAID mais popular atualmente em uso. Ele usa uma técnica que evita a concentração de E/S em um disco de paridade dedicado, que ocorre com o RAID 4.
Embora o RAID 5 divida os dados e crie informações de paridade semelhantes ao RAID 4, ao contrário do RAID 4, os dados de paridade são gravados separadamente em vários discos. A Penalidade de Gravação ainda ocorre porque os dados existentes devem ser pré-lidos antes que os dados de atualização e paridade sejam atualizados após a gravação dos dados. No entanto, o RAID 5 permite que vários pedidos de gravação sejam implementados simultaneamente porque os dados de paridade atualizados são dispersos pelos vários discos. Esse recurso garante maior desempenho em comparação com o RAID 4.



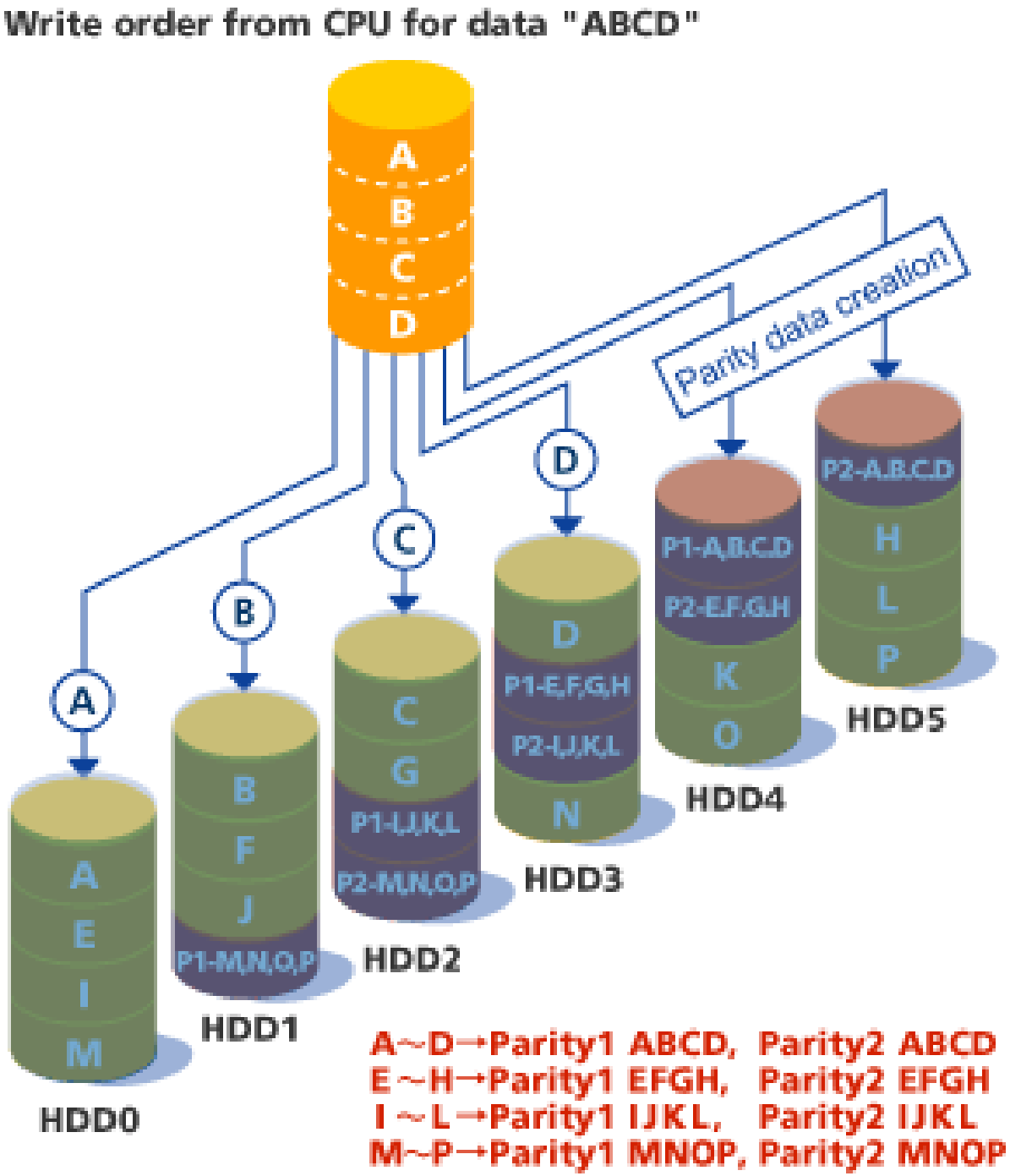
RAID 5 + 0

O RAID5 + 0 distribui dados entre grupos de multiplos RAID5 usando um método RAID0 front-end.
Esse striping de múltiplos RAID5 permite que um disco por grupo seja salvo no caso de falha no disco. Isso proporciona maior confiabilidade em sistemas de configuração de grande capacidade em comparação com um único grupo RAID5. Além disso, a reconstrução de transações, que com RAID 5 e RAID 6 leva um tempo cada vez maior à medida que aumenta a capacidade de disco, pode ser executada muito mais rapidamente com RAID5 + 0, pois a quantidade de dados em cada grupo RAID é menor.



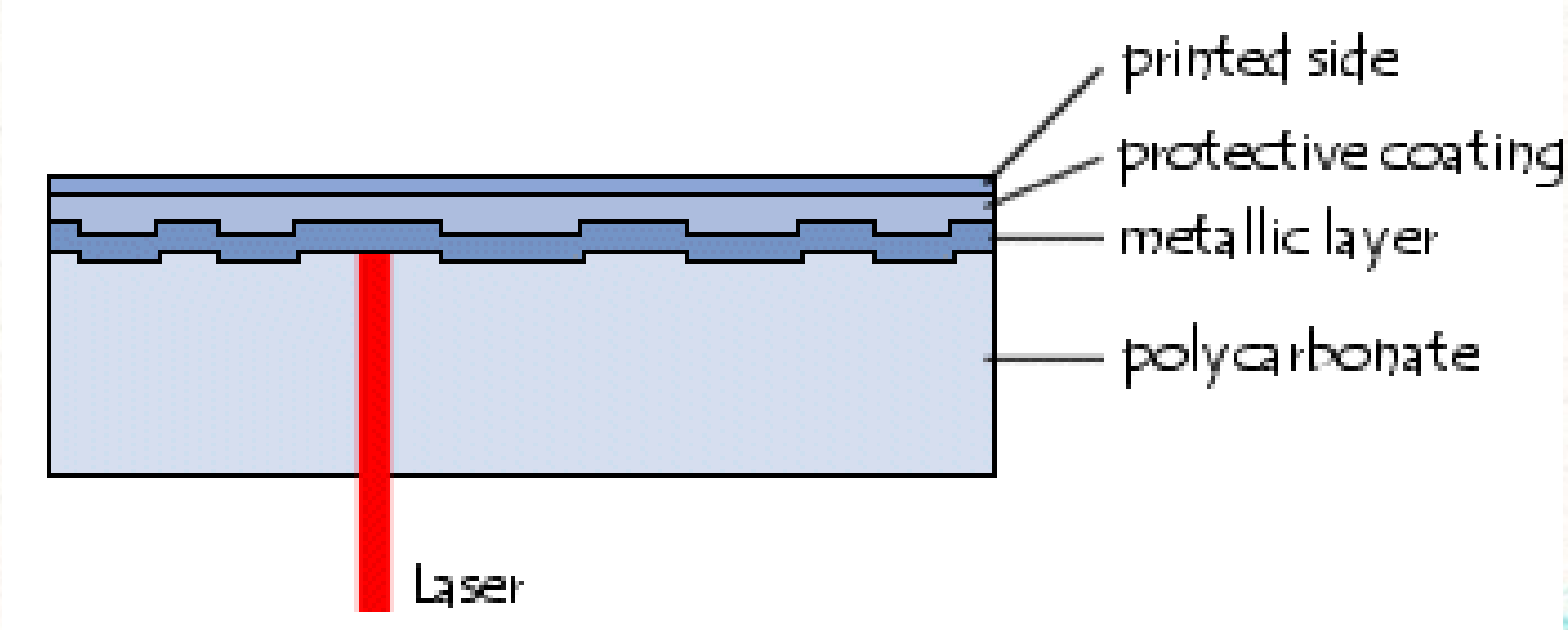
RAID 6

O RAID 6 implementa dois registros de paridade em diferentes unidades de disco (paridade dupla), permitindo a recuperação de duas falhas simultâneas de unidade de disco no mesmo grupo de RAID.
O RAID 6, em que as atualizações de paridade são alocadas separadamente em vários discos, bem como o RAID 5, podem implementar vários pedidos de gravação ao mesmo tempo. Esse recurso garante maior desempenho quando comparado à tecnologia RAID 4.



CD-ROMs

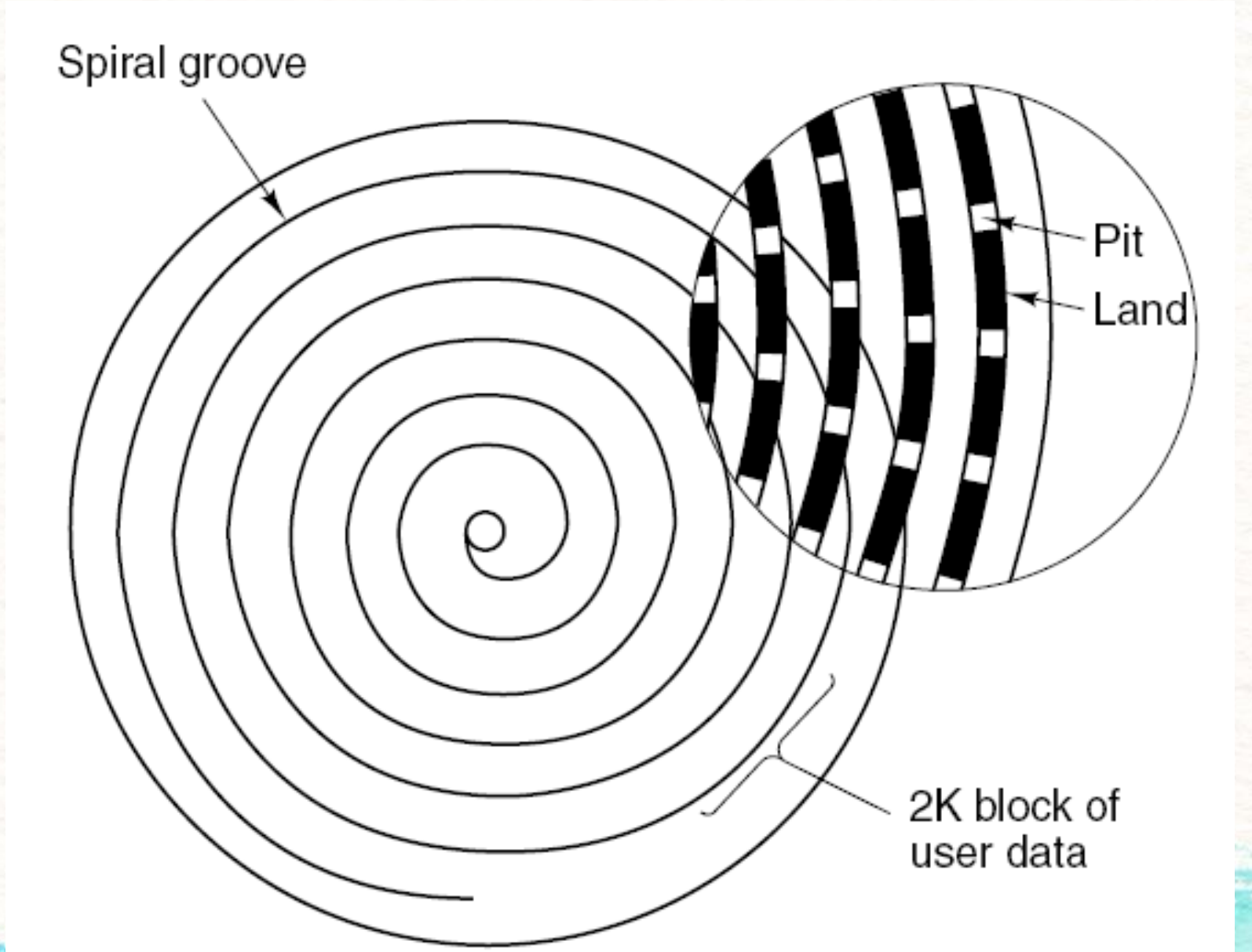
O CD é composto por um substrato de plástico (policarbonato) e uma fina película metálica refletora (ouro de 24 quilates ou liga de prata). A camada refletora é coberta por uma laca anti-UV acrílica que cria um filme protetor para os dados. Por último, uma camada adicional pode ser acrescentada para obter uma face superior impressa:



CD-ROMs

A camada refletora possui pequenos alvéolos. Assim, quando o laser atravessa o substrato de policarbonato, a luz se reflete na camada refletora, exceto quando o laser passa num alvéolo, é o que permite codificar a informação. Esta informação é armazenada nas 22.188 pistas gravadas em espirais (na realidade, trata-se apenas de uma pista concêntrica).

CD-ROMs



CD-ROMs

O laser do Compact Disc não é uma unidade de alta. Ele é um laser de baixa voltagem, estado sólido com rpotenciaadiação de um cristal semicondutor tão pequeno como quanto a cabeça de um alfinete. A radiação é monocromática, infra-vermelha. No sentido exato da palavra o feixe é invisivel e inofensivo, porque ele é concentrado somente no ponto focal. De qualquer forma a potência requerida para o funcionamento do leitor óptico é de apenas alguns miliwatts e só é ativado quando a gaveta esta fechada.

A trajetória da luz: À partir do laser, o feixe passa através do prisma semi-refletor (dois prismas juntos) até atingir a superfície refletora do disco, na sua trajetória de retorno, o feixe não consegue passar os 2 prismas novamente, mas apenas 1 deles, e na superfície de separação se reflete, indo incidir a luz no fotodiodo com a informação do disco.

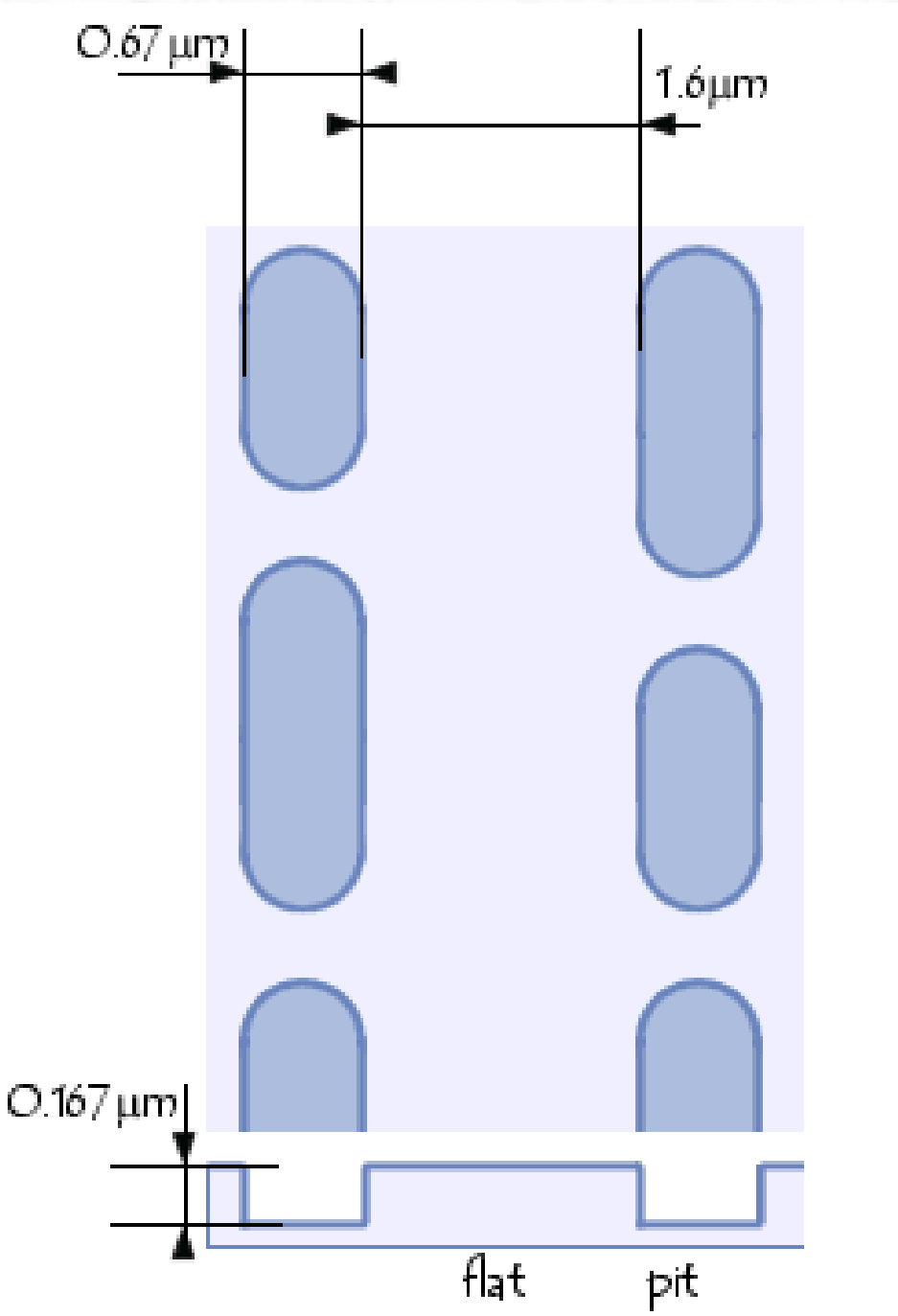
Codificação no CD-ROM

Na verdade, a pista física é constituída de alvéolos de profundidade de 0,168 μm, amplitude de 0,67μm e comprimento variável. As pistas físicas têm, entre elas, uma distância de cerca de 1,6 μm. Chamamos de cova (pit) o fundo do alvéolo e terra (land) os espaços entre eles, como mostrado no desenho.

O laser utilizado para ler o CD tem comprimento de onda de 780 nm no ar. Como o índice refrator do policarbonato é de 1,55, o comprimento de onda do laser no policarbonato equivale a 780/1.55 = 503 nm = 0,5 μm.
Levando em conta que a profundidade do alvéolo corresponde a um quarto do comprimento de onda do feixe do laser, a onda de luz refletida por uma cova se move de volta à metade do comprimento (125% de longitude para chegar ao disco e o mesmo para voltar) da onda refletida no plano.

Desta maneira, toda vez que o laser alcança o nível de um alvéolo com buracos, a onda e seu reflexo são defasados de metade da onda se anulando entre si (interferências destrutivas), de modo que tudo se passa como se nenhuma luz tivesse refletido. A passagem de um buraco para um plano provoca queda de sinal, representando um bit.

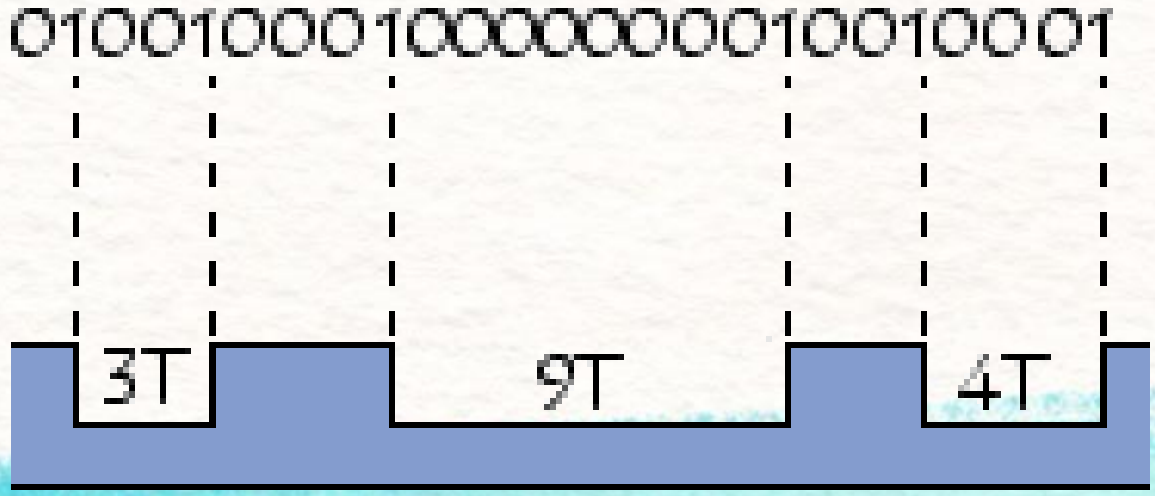
Codificação no CD-ROM



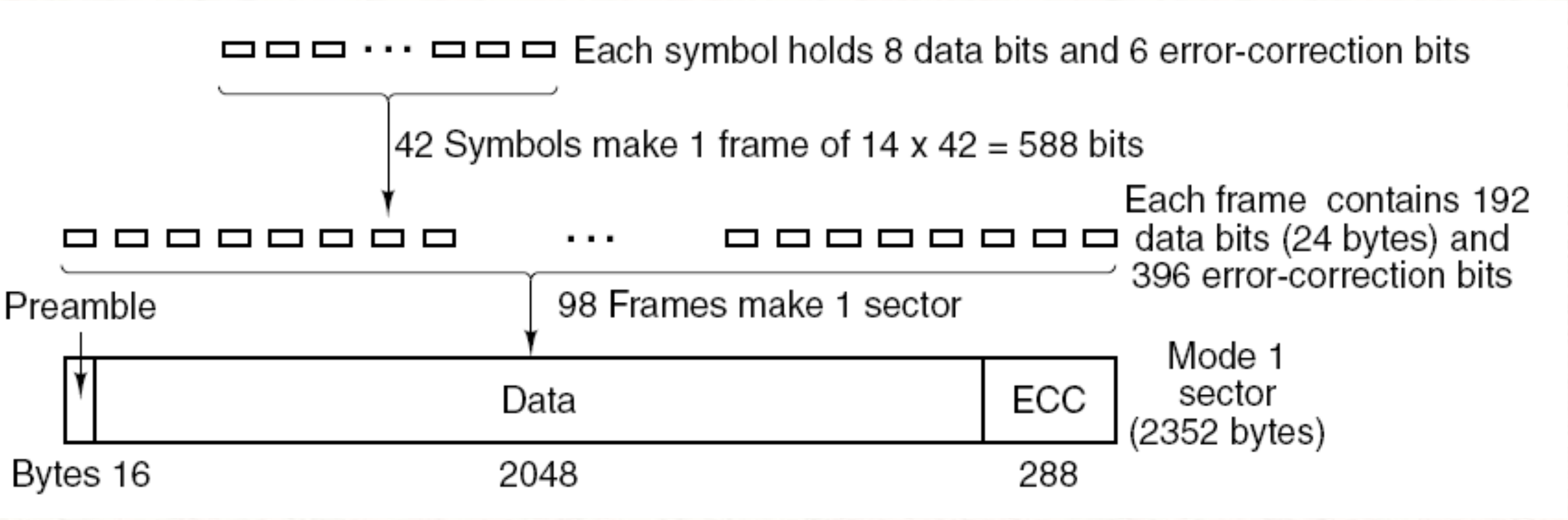
Codificação no CD-ROM

É o comprimento do alvéolo que permite armazenar a informação. O tamanho de um bit em um CD foi padronizado e corresponde à distância percorrida pelo feixe luminoso em 231,4 nanossegundos, ou seja, 0,278 μm e a velocidade padrão mínima de 1,2 m/s.

De acordo com o padrão EFM (Eight-to-Fourteen Modulation), utilizado para o armazenamento da informação em um CD, deve ter, no mínimo, dois bits em 0 entre dois bits 1 consecutivos e não pode não ter mais de 10 bits consecutivos em zero entre dois bits 1 para evitar os erros. É por isso que o comprimento de um alvéolo corresponde, no mínimo, ao comprimento necessário para armazenar o valor 001 (3T, ou seja 0,833 μm) e, no máximo, ao comprimento que corresponde ao valor 00000000001 (11T, ou seja, 3,054 μm)



CD-ROMs



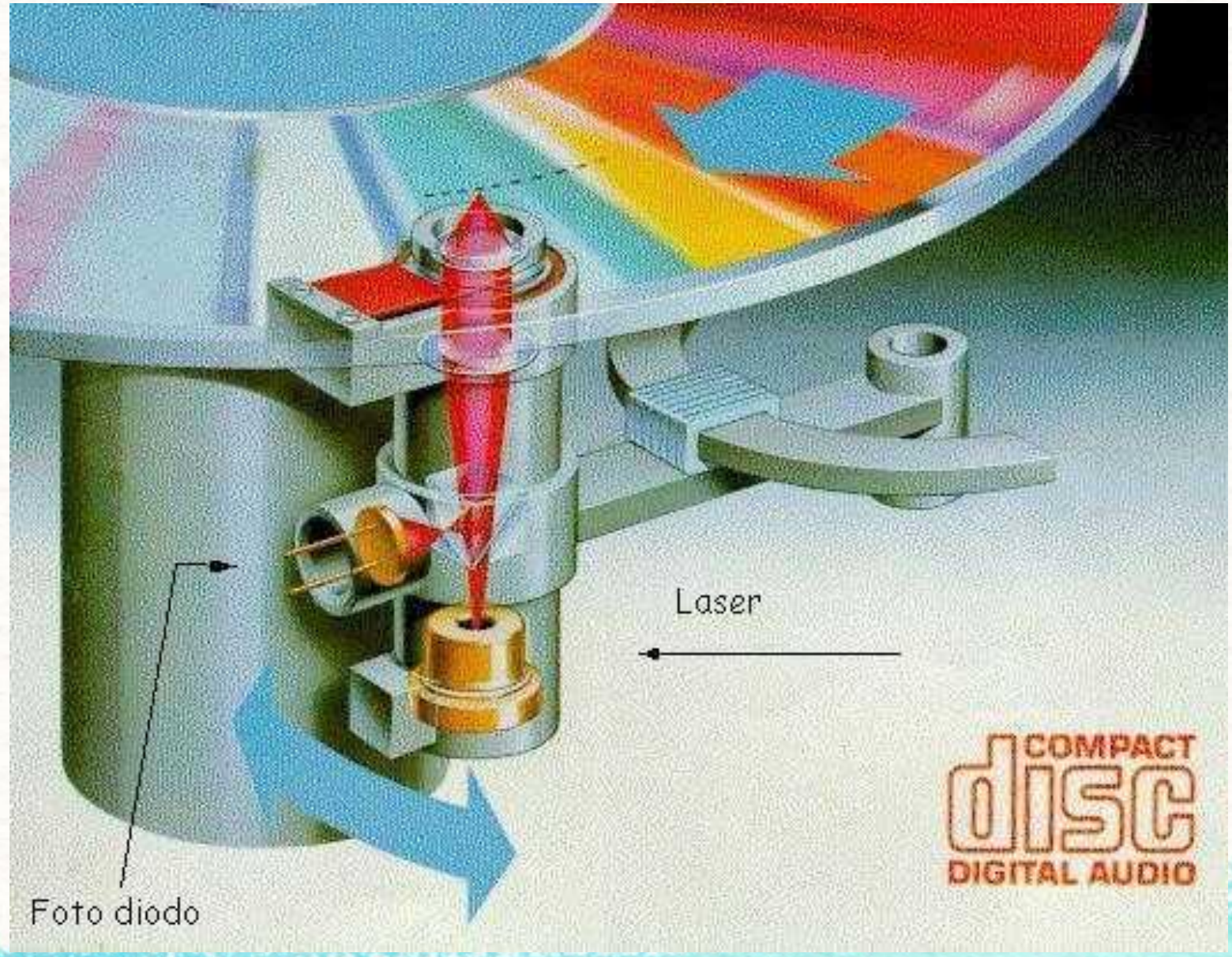
Layout lógico de um CD-ROM

Sistemas de arquivo do CD-ROMs

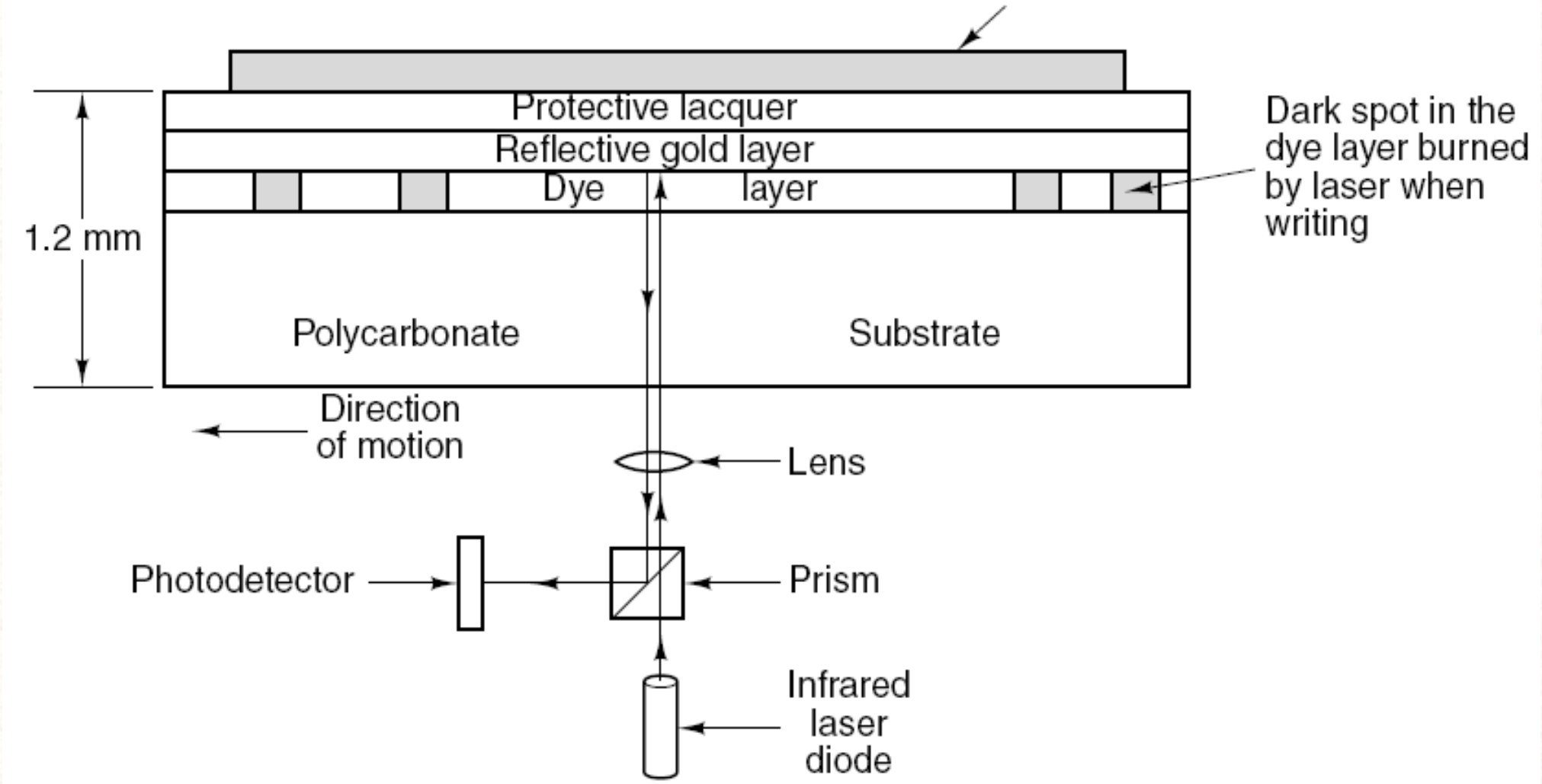
O padrão mais comum para sistemas de arquivos CD-ROM ó o ISO9660. Praticamente todos os CD-ROM atualmente no mercado são compatíveis com este padrão. Um dos objetivos desse padrão era tornar todos os CD-ROM legíveis em todos os computadores, independentemente da ordem de bytes e do sistema operacional utilizado.

Os CD-ROMs não possuem cilindros concêntricos como os discos magnéticos. Em vez disso, há uma única espiral contínua contendo os bits em uma sequência linear. Os bits ao longo da espiral são divididos em blocos lógicos (também chamados de setores lógicos) de 2352 bytes. Algumas delas são para preâmbulos, correção de erros e outras despesas indiretas. A porção de carga útil de cada bloco lógico é de 2048 bytes. Quando usado para música, os CDs têm leadins, leadouts e intervalos intertrack, mas eles não são usados para CD-ROMs de dados. Muitas vezes a posição de um bloco ao longo da espiral é indicada em minutos e segundos. Pode ser convertido em um número de bloco linear usando o fator de conversão de 1 seg = 75 blocos.

CD-ROMs



CD-ROMs Regraváveis



Seção transversal de um disco de CD-R e laser. Um CD-ROM de prata tem estrutura semelhante, porém sem camada de corante e com camada de alumínio sem cor ao invés de camada de ouro.

DVD

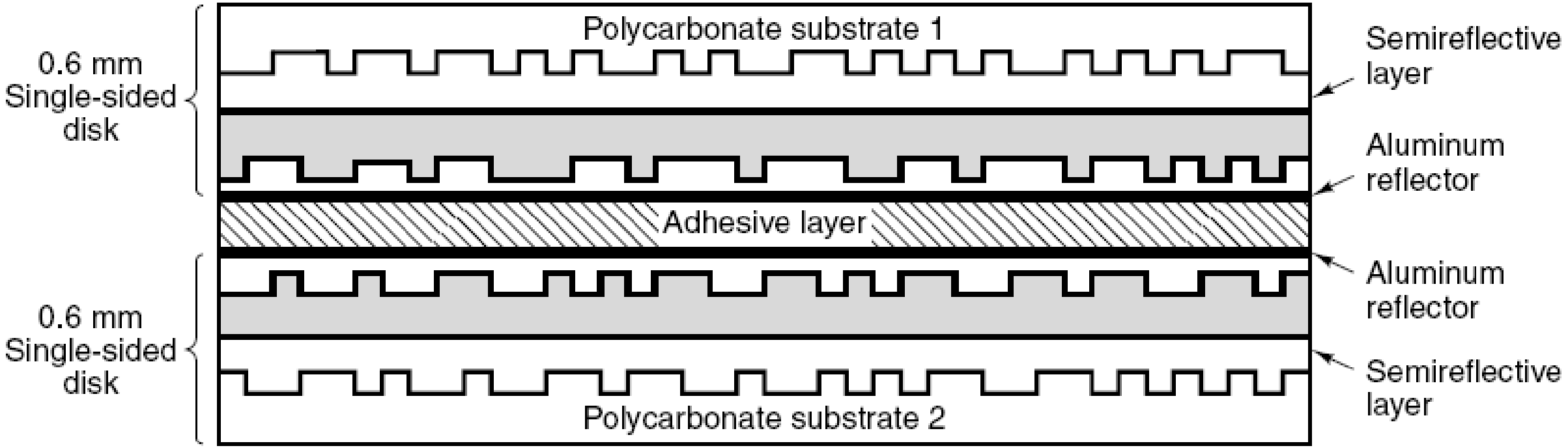
O DVD é uma melhoria do CD com:

- Fendas (pits) menores (0,4 microns versus 0,8 microns para CDs).
- Uma espiral mais apertada (0.74 microns entre as faixas versus 1.6 microns para CDs).
- Um laser vermelho (a 0,65 mícrons versus 0,78 mícrons para CDs).

Formatos de DVD

- Único lado, camada única (4,7 GB).
- Único lado, camada dupla (8,5 GB).
- Dois lados, camada única (9,4 GB).
- Dois lados, camada dupla (17 GB).

Estrutura do DVD



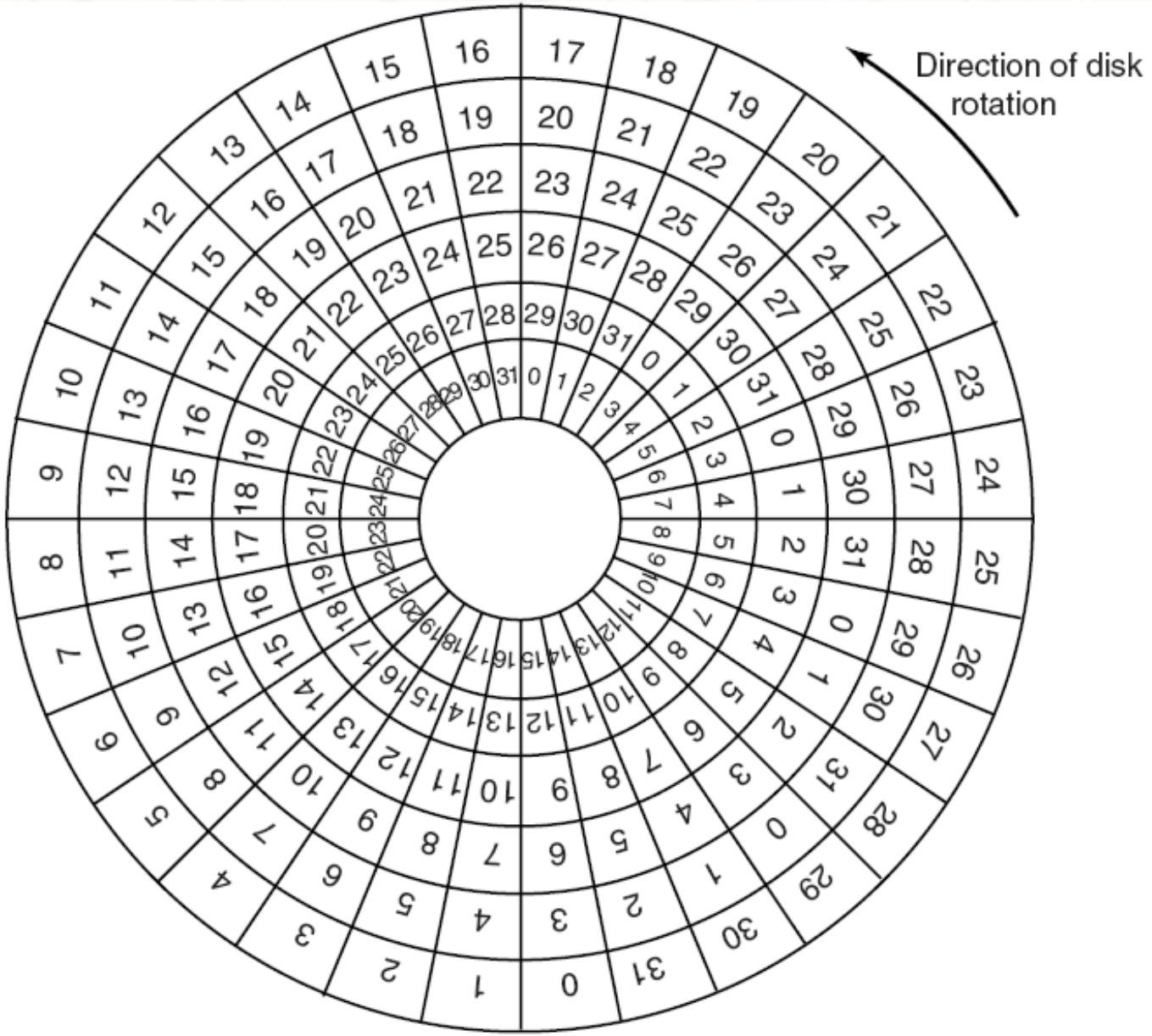
Um disco de dois lados com camadas duplas

Formatação de um disco

Setores de um disco

Formatação de um disco

Sequência de um cilindro



Programação do Braço de Disco

Fatores de tempo de leitura / gravação

- Tempo de busca (o tempo para mover o braço para o cilindro adequado).
- Atraso de rotação (o tempo para o setor adequado girar sob a cabeça).
- Tempo real de transferência de dados.

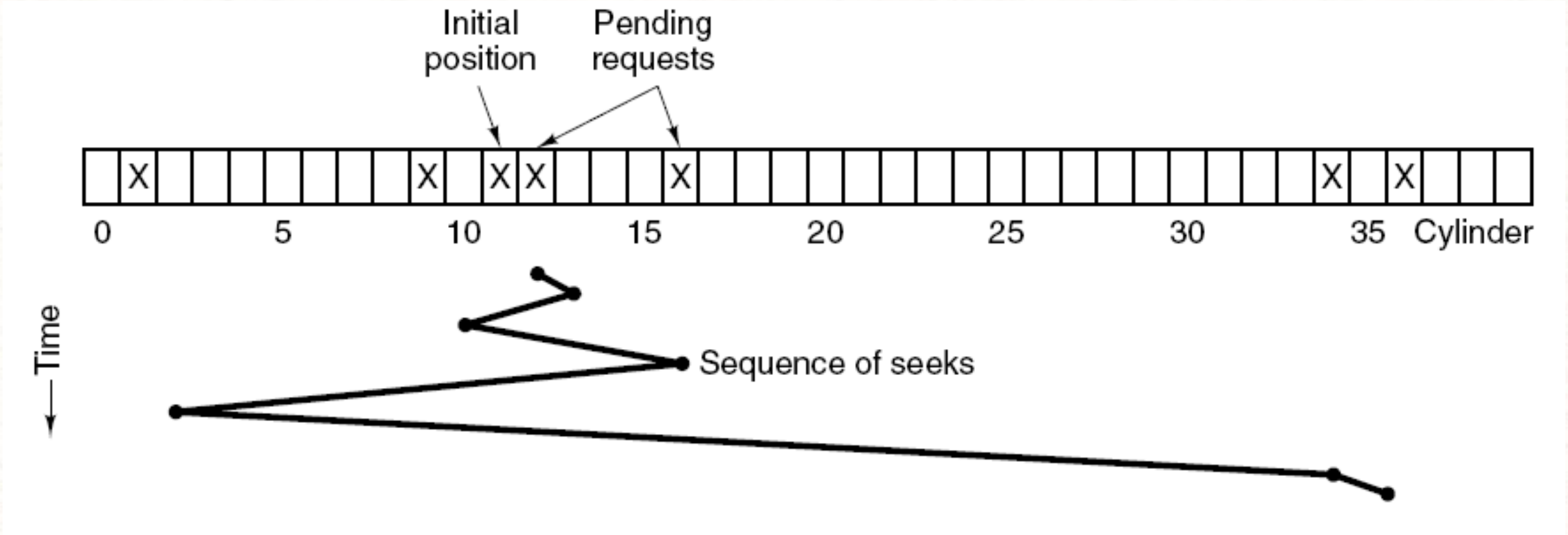
Para a maioria dos discos, o tempo de busca domina os outros dois tempos, portanto, reduzir o tempo médio de busca pode melhorar substancialmente o desempenho do sistema.

Programação SSF (short seek first)

Considere um disco imaginário com 40 cilindros. Um pedido vem para ler um bloco no cilindro 11. Enquanto o buscador de cilindros 11 está em andamento, novas solicitações chegam aos cilindros 1, 36, 16, 34, 9 e 12, nessa ordem. Eles são inseridos na tabela de solicitações pendentes, com uma lista de links separada para cada cilindro.

Quando a solicitação atual (para o cilindro 11) é concluída, o driver de disco pode escolher qual solicitação deve ser executada a seguir. Usando o FCFS (First-Come, First-Served), ele iria ao lado do cilindro 1, depois para o 36 e assim por diante. Esse algoritmo exigiria movimentos de braço de 10, 35, 20, 18, 25 e 3, respectivamente, para um total de 111 cilindros.

Programação do Braço de Disco



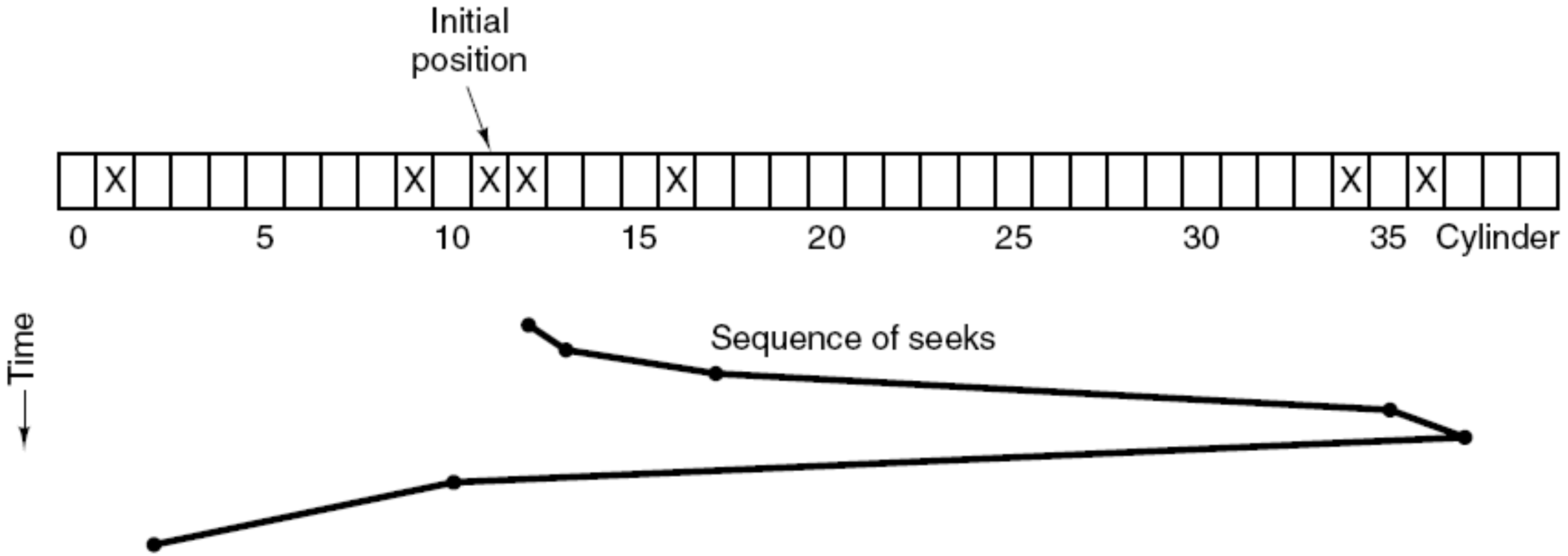
Shortest Seek First (SSF) Algoritmo de escalonamento de disco.

Programação do Elevador

O problema de programar um elevador em um prédio alto é semelhante ao de agendar um braço de disco. Solicitações entram continuamente chamando o elevador para andares (cilindros) aleatoriamente. O computador que opera o elevador pode facilmente rastrear a sequência em que os clientes pressionaram o botão de chamada e atendê-los usando FCFS ou SSF.

No entanto, a maioria dos elevadores usa um algoritmo diferente para conciliar as metas de eficiência e justiça que são conflitantes entre si. Eles continuam se movendo na mesma direção até que não haja mais pedidos pendentes nessa direção, então eles mudam de direção. Esse algoritmo, conhecido tanto no mundo do disco quanto no mundo dos elevadores como o algoritmo do elevador, requer que o software mantenha 1 bit: o bit de direção atual, UP ou DOWN. Quando uma solicitação termina, o driver do disco ou do elevador verifica o bit. Se estiver na posição UP, o braço ou a cabine será movido para a próxima solicitação pendente mais alta. Se nenhuma solicitação estiver pendente em posições mais altas, o bit de direção será invertido. Quando o bit é ajustado para DOWN, o movimento é para a próxima posição solicitada mais baixa, se houver. Se nenhum pedido estiver pendente, ele simplesmente para e aguarda.

Algoritmo do Elevador



O algoritmo do elevador para agendar solicitações de disco.

Referências Bibliográficas

- TANENBAUM, Andrew S., BOSS, Herbert. **Sistemas Operacionais Modernos,** Pearson - 4ª ed., 2016.
- SILBERSCHATZ, A., GALVIN, P.B., GAGNE, G. **Fundamentos de Sistemas Operacionais**, Ed. LTC, 8ª ed., 2011
- DEITEL, H.M.; DEITEL, P.J.; CHOFFNES, D.R. – **Sistemas Operacionais**. Prentice Hall, Tradução da 3ª ed., 2005
- DEITEL, H.M.; DEITEL, P.J. – **C How to Program**. Prentice Hall, Tradução da 3ª ed., 2001
- MIZRAHI, Victorine Viviane. **Treinamento em Linguagem C – Curso Completo módulos 1 e 2**, Ed. Person Education.
- http://www.fujitsu.com/global/products/computing/storage/eternus/glossary/raid/feature.html