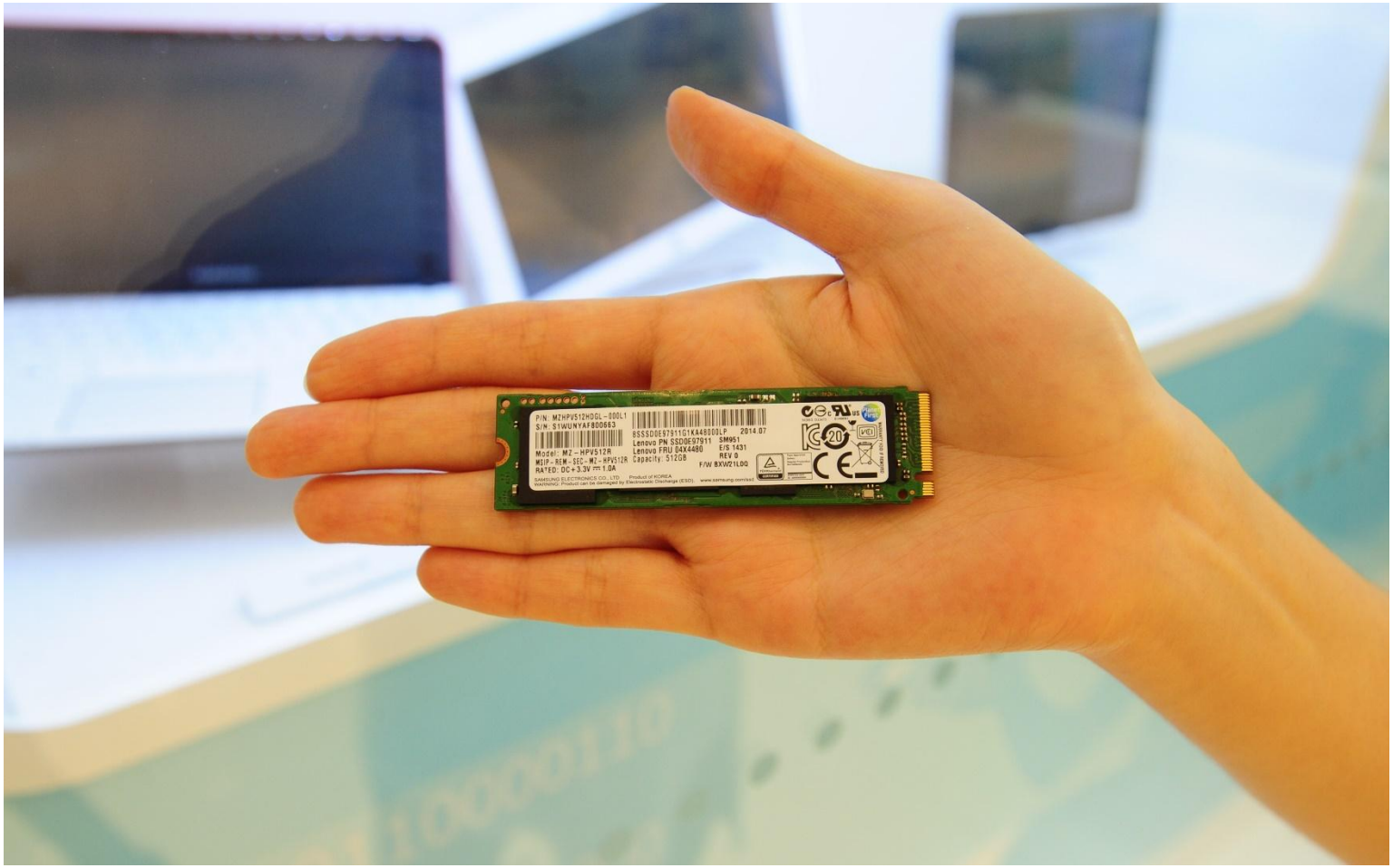
Tecnologias de Armazenamento

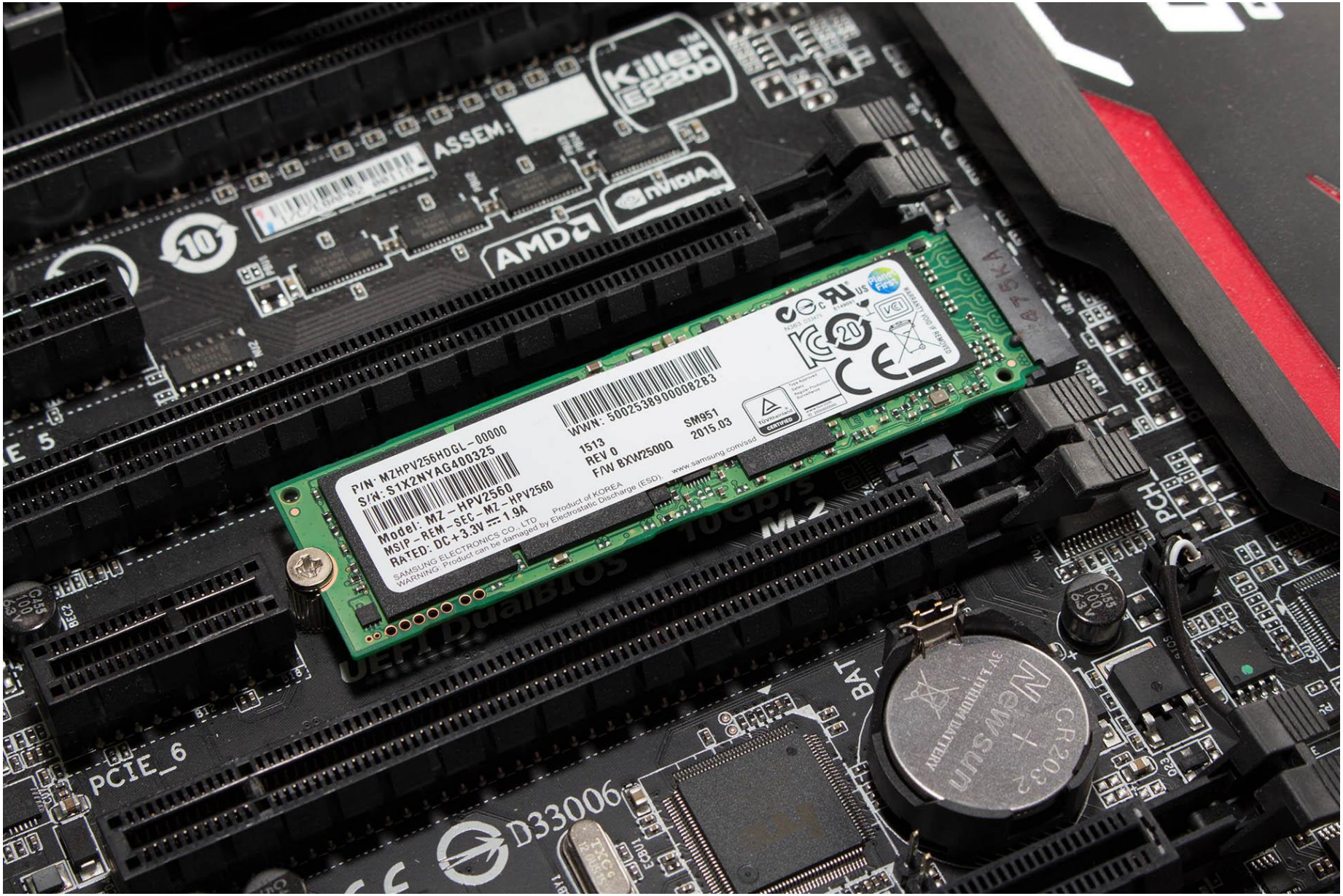
Tipos e Dispositivos

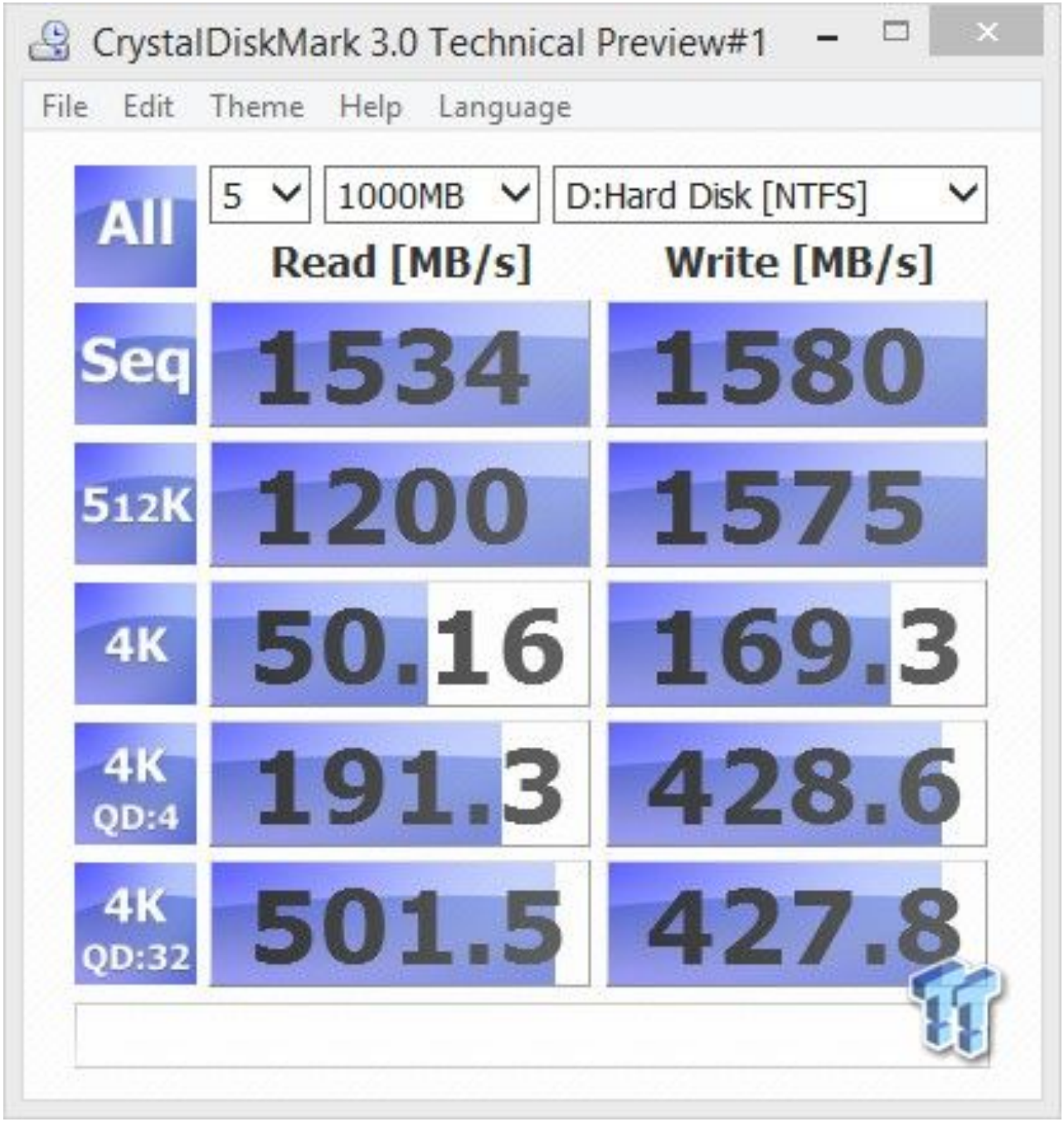
M.2 PCIe

- Implementação da interface PCIex (4 lanes) para dispositivos compactos, como SSDs
- Conhecido anteriormente como NGFF (Next Generation Form Factor)
- Utiliza o padrão SATA Express com AHCI (Advanced Host Controller Interface)
- Substitui padrão mSATA
- Banta teórica máxima 16Gb/s (2GB/s)

Ex: SM951 SSD


P/N: MZHPV512HDGL-000L1
S/N: S1WUNYAF800563
Model: MZ-HPV512R
MSIP-REM-SEC-MZ-HPV512R
RATED: DC+3.3V ⎓1.0A
Lenovo PN SSD0E97911
Lenovo FRU 04X4480
Capacity: 512GB
SM951
E/S 1431
REV 0
F/W 8XW2L0Q


P/N: MZHPV256HDGL-00000
S/N: S1X2NYAG400325
Model: MZ-HPV2560
MSIP-REM-SEC-MZ-HPV2560
RATED: DC+3.3V ⎓ 1.9A
SAMSUNG ELECTRONICS CO., LTD
Product of KOREA
WARNING: Product can be damaged by Electrostatic Discharge (ESD)
www.samsung.com/ssd
WWN: 5002538900008283
1513
REV 0
F/W BXW2500Q
SM951
2015.03
Killer E2200
ASSEM:
AMD
PCIE_6
M.2
BAT


CrystalDiskMark 3.0 Technical Preview#1
File Edit Theme Help Language
All
5
1000MB
D:Hard Disk [NTFS]
Read [MB/s]
Write [MB/s]
Seq
1534
1580
512K
1200
1575
4K
50.16
169.3
4K QD:4
191.3
428.6
4K QD:32
501.5
427.8


PATA
SATA

SATA e PATA

Name	Raw bandwidth	Transfer speed
M.2	16 Gbit/s	2000 MB/s
eSATA	6 Gbit/s	600 MB/s
eSATAp	3 Gbit/s	300 MB/s
SATA revision 3.2	16 Gbit/s	1.97 GB/s
SATA revision 3.0	6 Gbit/s	600 MB/s
SATA revision 2.0	3 Gbit/s	300 MB/s

High-Capacity Hard Drives

- Drives Acima de 2TB
- Com hard drives iguais ou acima de 2.2TB, a industria teve que lidar com limitações de capacidade de endereçamento introduzidas no projeto original do PC, principalmente causadas pelo uso de definições de 32-bits para tamanhos de partição e LBA (logical block addresses) resultando numa máxima capacidade endereçável de of 2.199TB.

Seagate 10TB helium drive

Seagate reaches 1Tb per square inch, hard drive to reach 60TB capacity

Seagate says that it has reached the milestone of storage density that offers 1 terabit (1 trillion bits) per square inch, using Heat-Assisted Magnetic recording technology that promises a 60TB hard drive within the next decade.

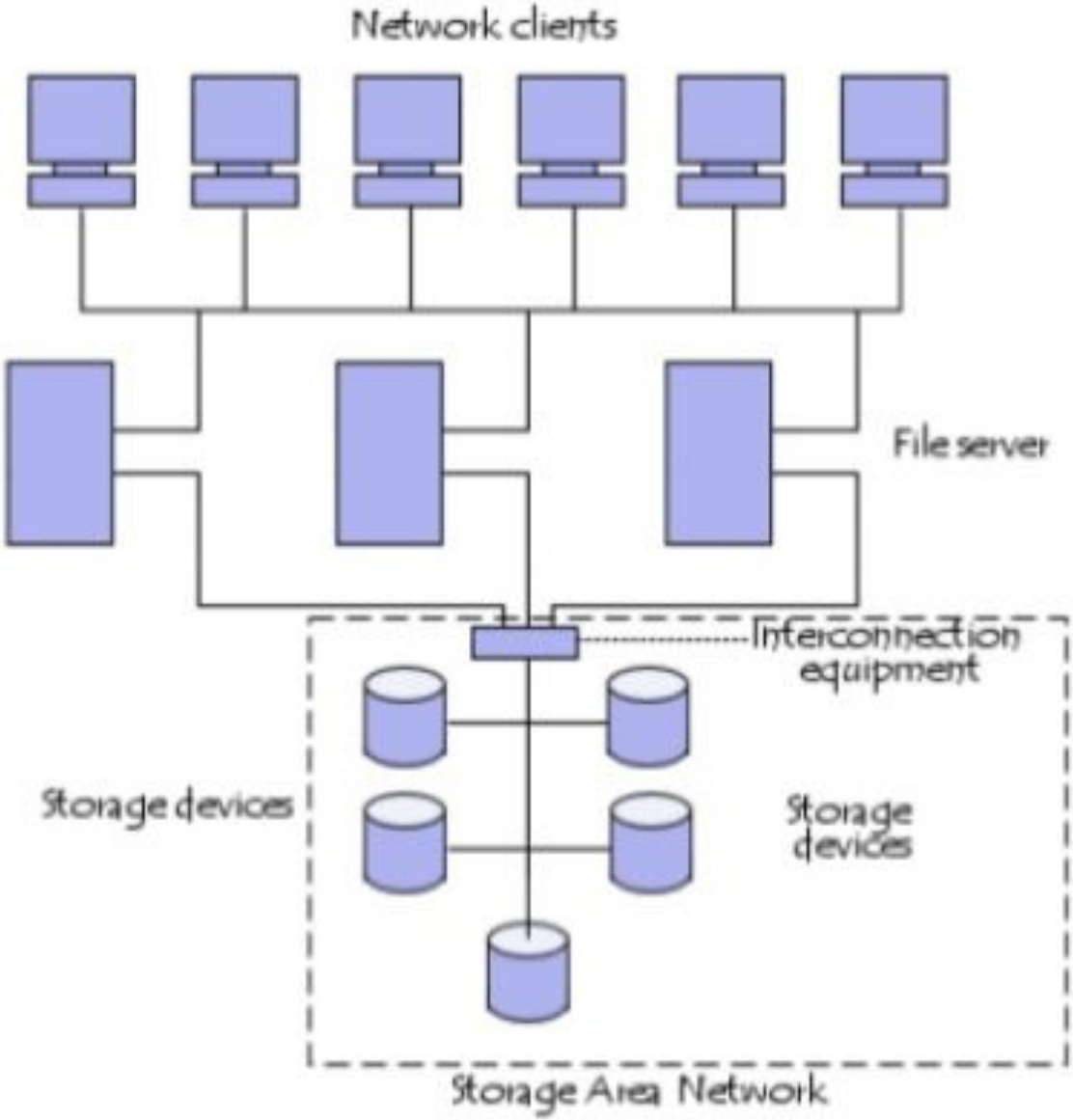
by Dong Ngo @riceandstirfry / March 19, 2012 12:39 PM PDT

Storage Area Network (SAN)

Características do Padrão

Fibre Channel

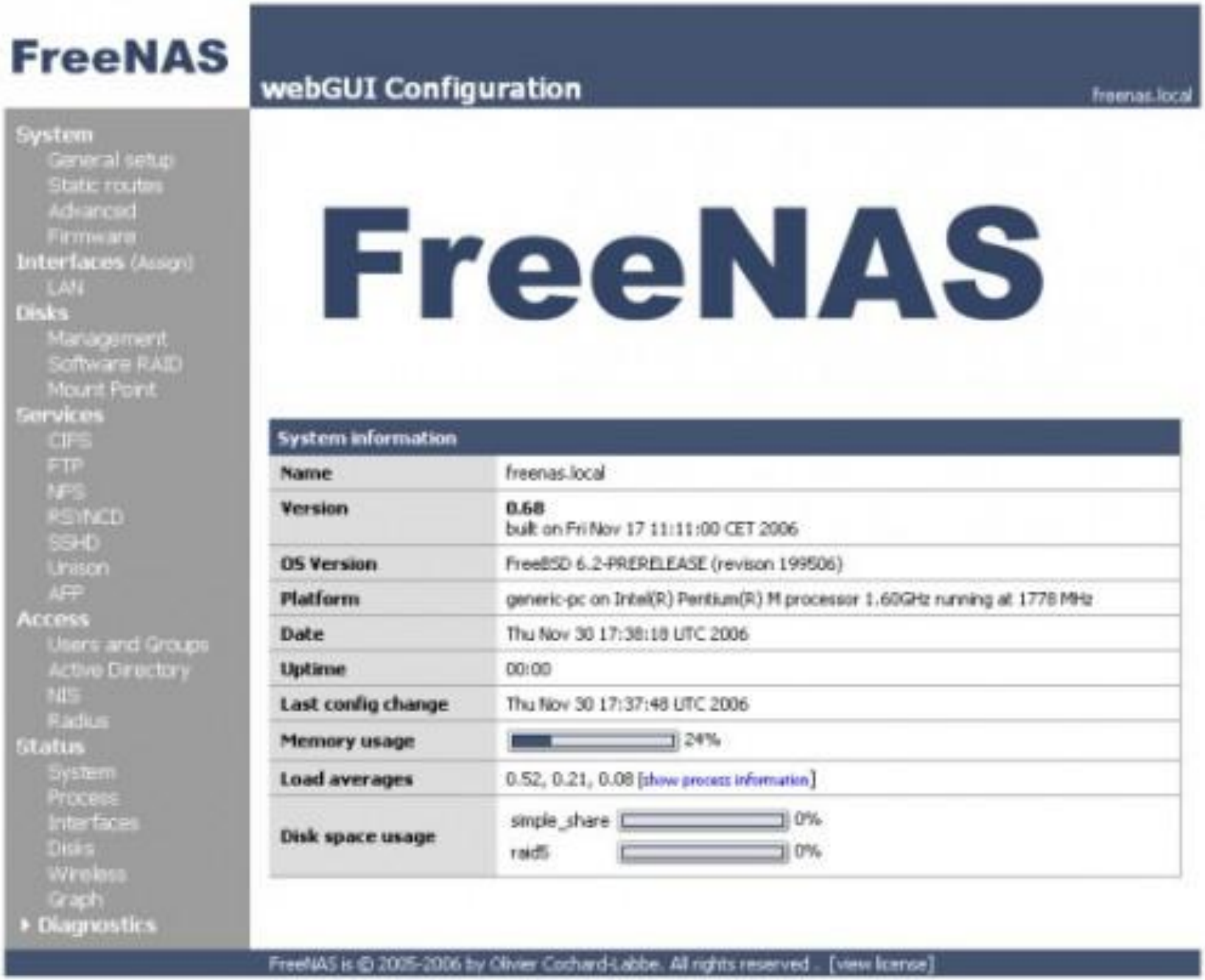
- 10km
- 10Gb/s



NAS

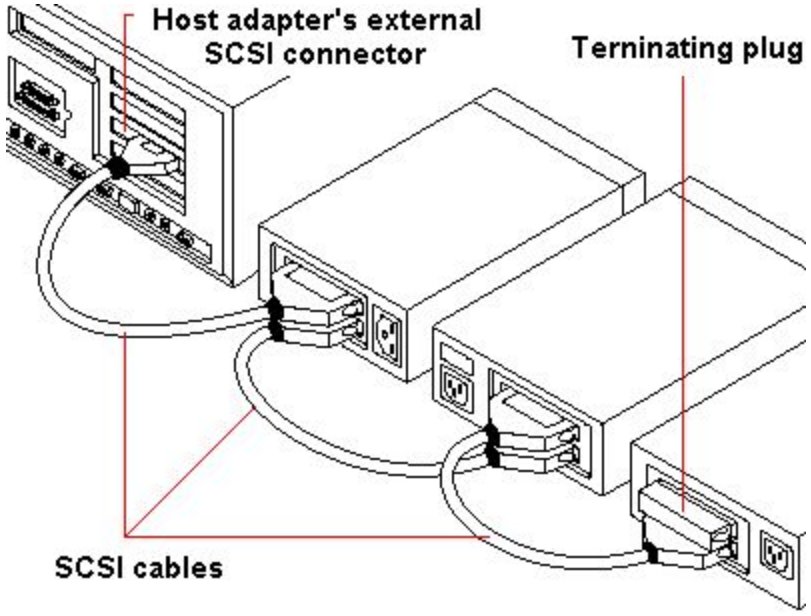
- Network Attached Storage
  - Servidor de arquivos
  - Sistema operacional próprio
  - Baseado em LAN
  - Sistema de arquivos NFS e CIFS (common internet)

NAS

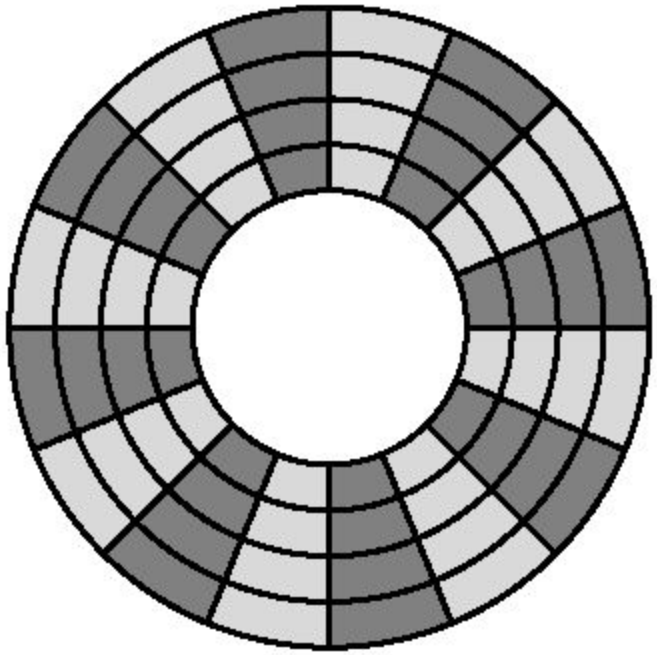

FreeNAS
webGUI Configuration
freenas.local
System
General setup
Static routes
Advanced
Firmware
Interfaces (Assign)
LAN
Disks
Management
Software RAID
Mount Point
Services
CIFS
FTP
NFS
RSYNCD
SSHD
Unison
APP
Access
Users and Groups
Active Directory
NIS
Radius
Status
System
Process
Interfaces
Disks
Wireless
Graph
Diagnostics
FreeNAS
System information
Name
freenas.local
Version
0.68
built on Fri Nov 17 11:11:00 CET 2006
OS Version
FreeBSD 6.2-PRERELEASE (revison 199506)
Platform
generic-pc on Intel(R) Pentium(R) M processor 1.60GHz running at 1778 MHz
Date
Thu Nov 30 17:38:18 UTC 2006
Uptime
00:00
Last config change
Thu Nov 30 17:37:48 UTC 2006
Memory usage
24%
Load averages
0.52, 0.21, 0.08 [show process information]
Disk space usage
simple_share
0%
raid5
0%
FreeNAS is © 2005-2006 by Olivier Cochard-Labbe. All rights reserved. [view license]

SCSI (Barramento)

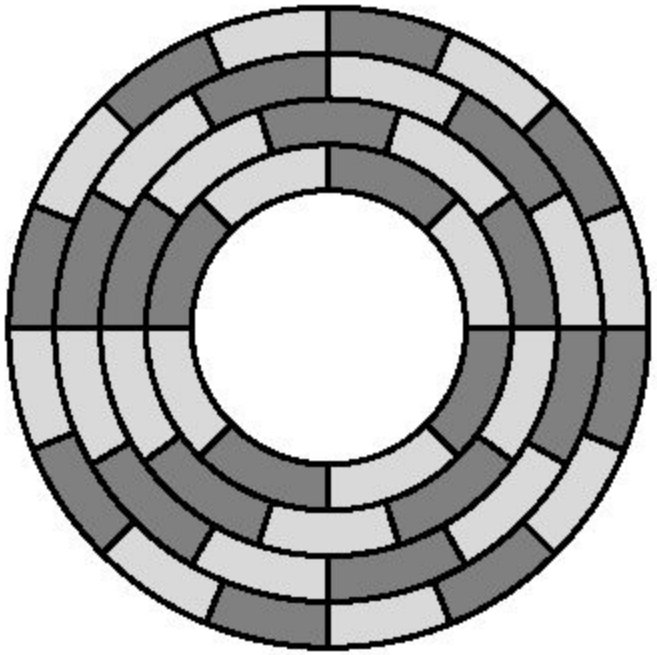
- Antigo padrão para HDDs de alto desempenho
- Interconexão estilo “Daisy Chain”
- High Bandwidth/Throughput



Disk Layout Methods



(a) Constant angular velocity



(b) Constant linear velocity

RAID

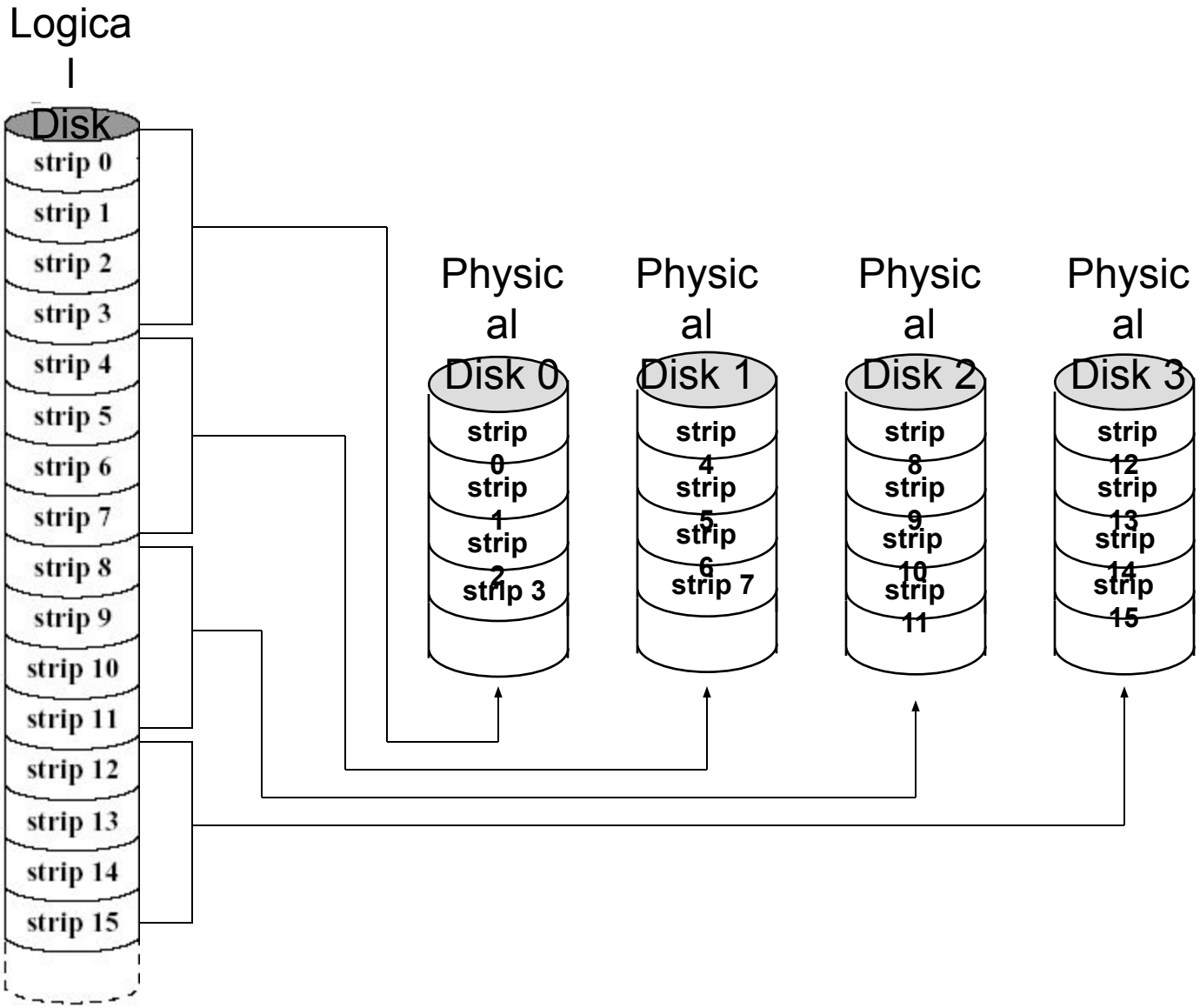
Redundant Array of Independent (Inexpensive) Disks

- Conjunto de discos tradados como uma unidade lógica
- Dados distribuidos pelos discos do array
- Uso opcional de redundância e/ou paridade, permitindo recuperação de dados no caso de falhas
- 6 padrões inicialmente propostos
- Padrões 2 e 4 não são comercializados, só incluídos para facilitar o entendimento

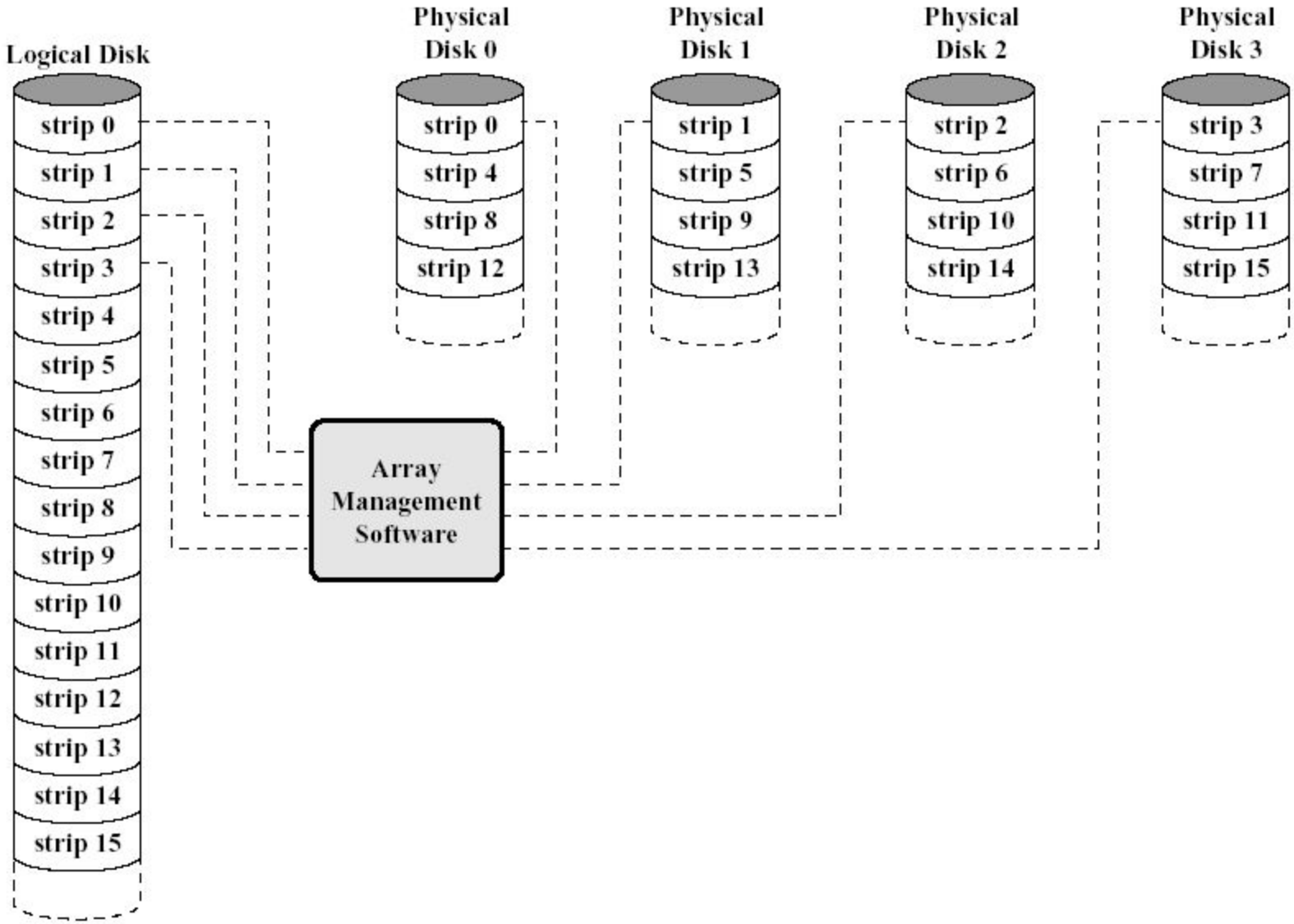
RAID 0 (Striping)

- No RAID 0 todos os HDs passam a ser acessados como se fossem um único drive. Ao serem gravados, os arquivos são fragmentados nos vários discos, permitindo que os fragmentos possam ser lidos e gravados simultaneamente, com cada HD realizando parte do trabalho. Isso permite melhorar brutalmente a taxa de leitura e de gravação e continuar usando 100% do espaço disponível nos HDs. O problema é que no RAID 0 não existe redundância. Os HDs armazenam fragmentos de arquivos, e não arquivos completos. Sem um dos HDs, a controladora não tem como reconstruir os arquivos e tudo é perdido. Isso faz com que o modo RAID 0 seja raramente usado em servidores.

Data mapping for a RAID 0 (Linear) Array



Data mapping for a RAID 0 (Striping) Array



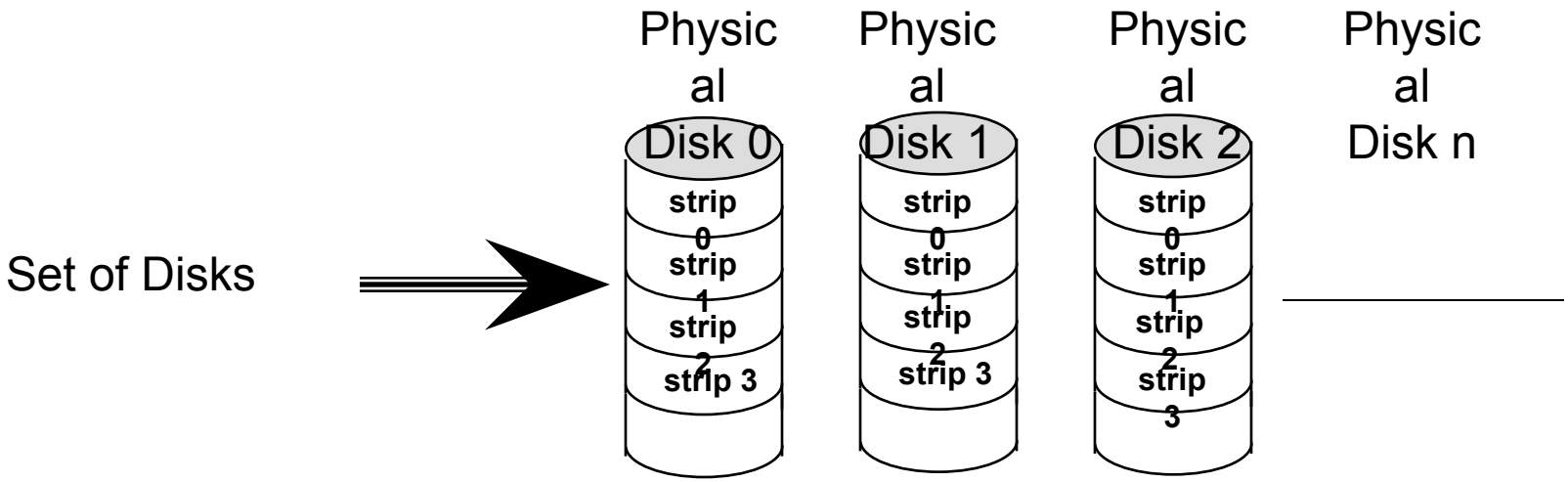
RAID 0:

- Sem redundância
- Desempenho depende dos padrões de requisição
  Altas taxas de acesso são atingidas quando:
  - Caminho de dados integral é rápido, inclui controladora, I/O BUS e acesso a memória (ex. Intel ICH11r)
  - A aplicação gera requisições de forma eficiente de forma a usar o array linearmente, acessando sempre que possível stripes consecutivos
- Requisições de I/O são tratadas em paralelo
- Basicamente uma sequencia de strips (blocos e setores), distribuídos intercaladamente pelos HDDs

RAID

- **RAID 1 (Mirroring)**: No RAID 1 são usados dois HDs (ou qualquer outro número par). O primeiro HD armazena dados e o segundo armazena um cópia exata do primeiro, atualizada em tempo real. Se o primeiro HD falha, a controladora automaticamente chaveia para o segundo HD, permitindo que o sistema continue funcionando. Em servidores é comum o uso de HDs com suporte a hot-swap, o que permite que o HD defeituoso seja substituído a quente, com o servidor ligado. A desvantagem em usar RAID 0 é que metade do espaço de armazenamento é sacrificado.

Raid 1 (mirrored)



RAID 1 não usa paridade, apenas espelha os dados

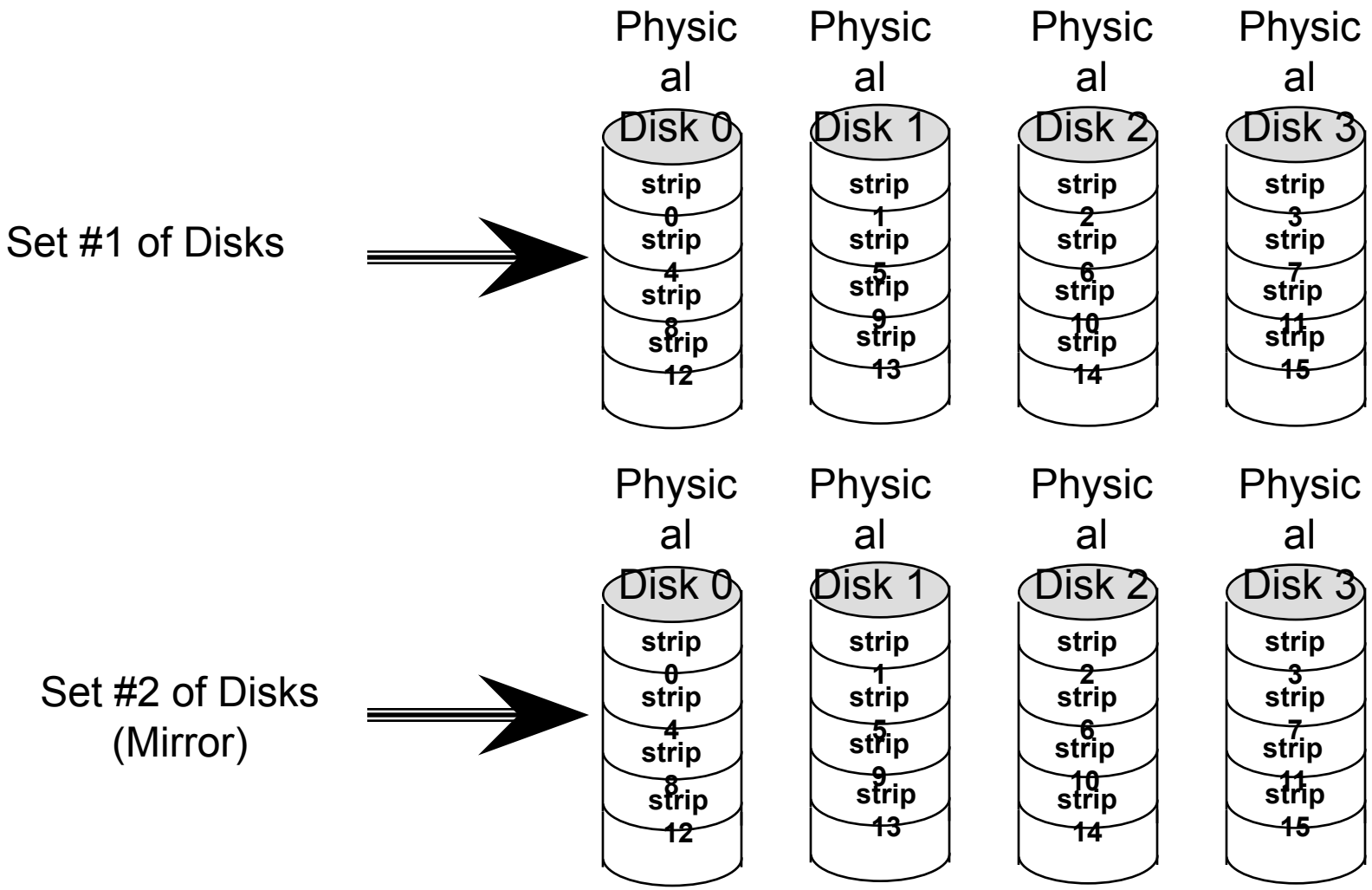
RAID 1

- Mais:
  - Aumenta segurança
  - Requisições podem ser atendidas por qualquer dos discos (minimum search time)
  - Gravações são executadas em paralelo sem penalidade (writing penalty)
  - Recuperação de erros é fácil, bastando copier o dado do disco correto
  - Pode ser usado em combinaçãocom RAID 0 para velocidade e segurança
- Menos:
  - Preço dobra
  - RAID 0+1 dobra o número de discos do RAID 0.

RAID 10

- **RAID 10 (Mirror/Strip)**: Este modo combina os modos 0 e 1 e pode ser usado com a partir de 4 HDs (ou outro número par). Metade dos HDs são usados em modo striping (RAID 0), enquanto a segunda metade armazena uma cópia dos dados dos primeiros, oferecendo redundância.

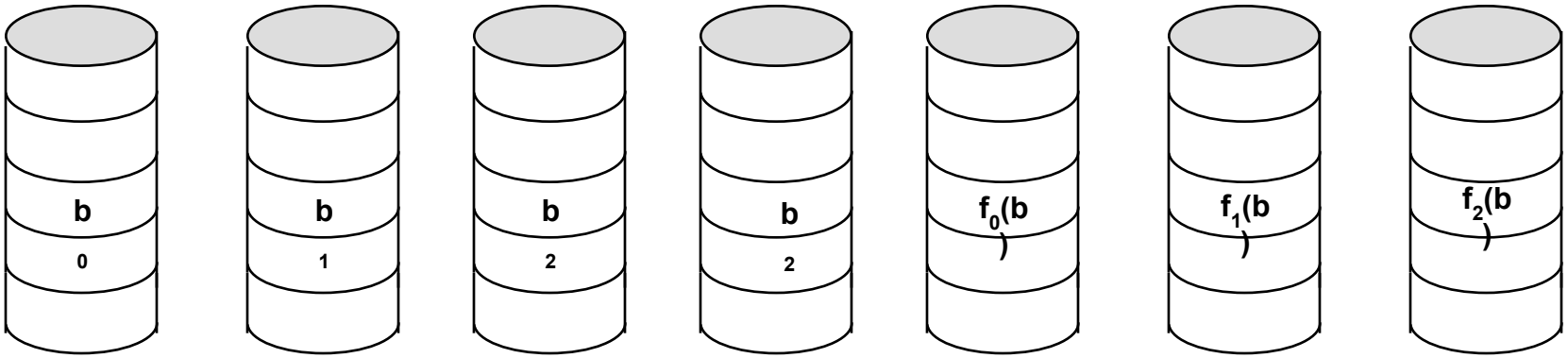
Raid 0+1 (striped+mirrored)



RAID 2

- Small strips, one byte or one word
- Synchronized disks, each I/O operation is performed in a parallel way
- Error correction code (Hamming code) allows for correction of a single bit error
- Controller can correct without additional delay
- Is still expensive, only used in case many frequent errors can be expected

Raid 2 (redundancy through Hamming code)



Hamming code

7	6	5	4	3	2	1	P
1	0	1	*0*	1	*0*	*1*	
*	*	*		*		*	*0*
*	*			*	*		*0*
*		*	*				*0*

Stored sequence
Data: 1011 in 7,6,5,3
Parity in 4,2,1

7	6	5	4	3	2	1	P	
1	1	1	*0*	1	*0*	*1*		
*	*	*		*		*	*1*	
*	*			*	*		*1*	=6
*		*	*				*0*	

Single error can
be repaired

RAID 3

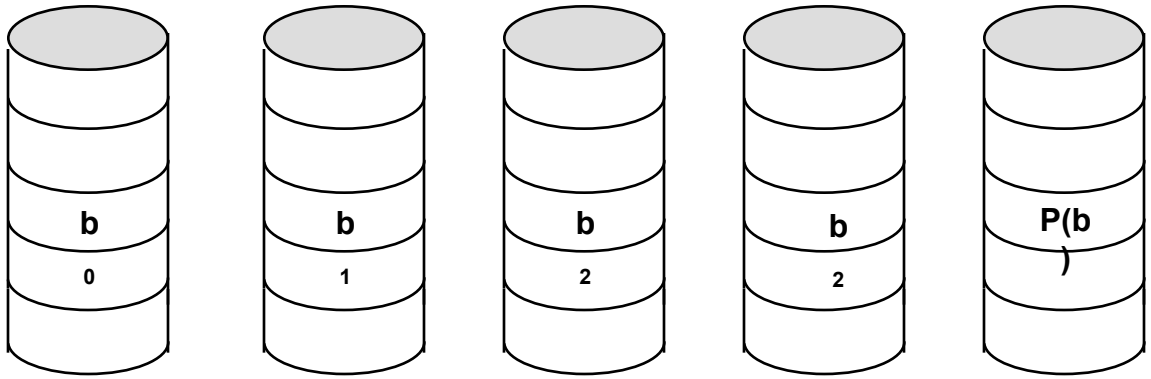
- Level 2 needs $\log_2$(number of disks) parity disks
- Level 3 needs only one, for one parity bit
- In case one disk crashes, the data can still be reconstructed even on line ("reduced mode") and be written (X1-4 data, P parity):

$$P = X1+X2+X3+X4$$

$$X1=P+X2+X3+X4$$

- RAID 2-3 have high data transfer times, but perform only one I/O at the time so that response times in transaction oriented environments are not so good

RAID 3 (bit-interleaved parity)



RAID 4

- Larger strips and one parity disk
- Blocks are kept on one disk, allowing for parallel access by multiple I/O requests
- Writing penalty: when a block is written, the parity disk must be adjusted (e.g. writing on X1):

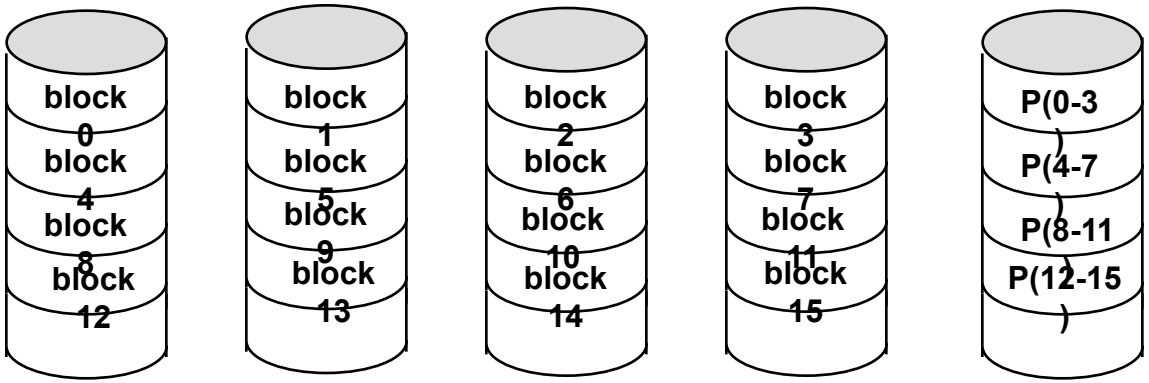
  P =X4+X3+X2+X1

  P'=X4+X3+X2+X1'

  =X4+X3+X2+X1'+X1+X1

  =P+X1+X1'
- Parity disk may be a bottleneck
- Good response times, less good transfer rates

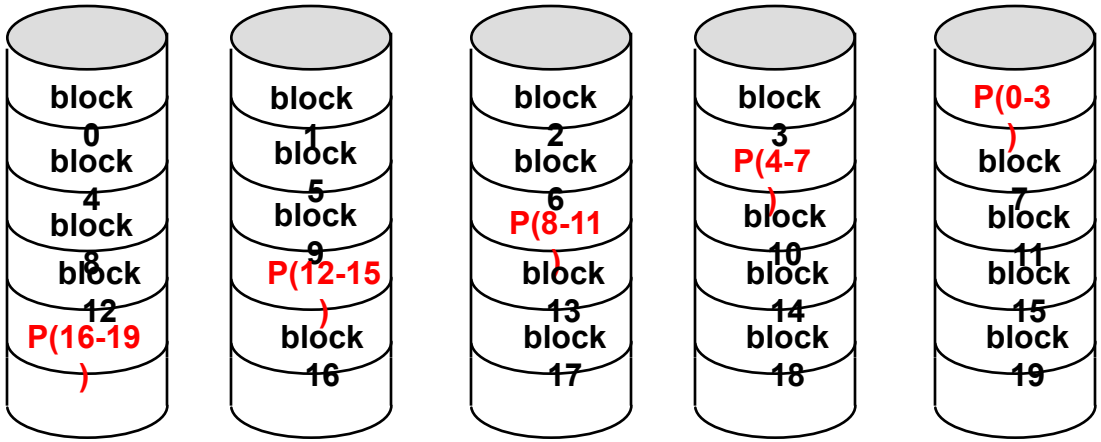
RAID 4 (block-level parity)



RAID 5

- Modo mais utilizado em servidores com um grande número de HDs. O RAID 5 usa um sistema de paridade para manter a integridade dos dados. Os arquivos são divididos em fragmentos e, para cada grupo de fragmentos, é gerado um fragmento adicional, contendo códigos de paridade. Os códigos de correção são espalhados entre os discos. Dessa forma, é possível gravar dados simultaneamente em todos os HDs, melhorando o desempenho.

RAID 5: paridade distribuída em nível de bloco



- Distribuição da paridade evita o gargalo
- Usa Round Robin onde

  Parity disk = (-block number/4) mod 5

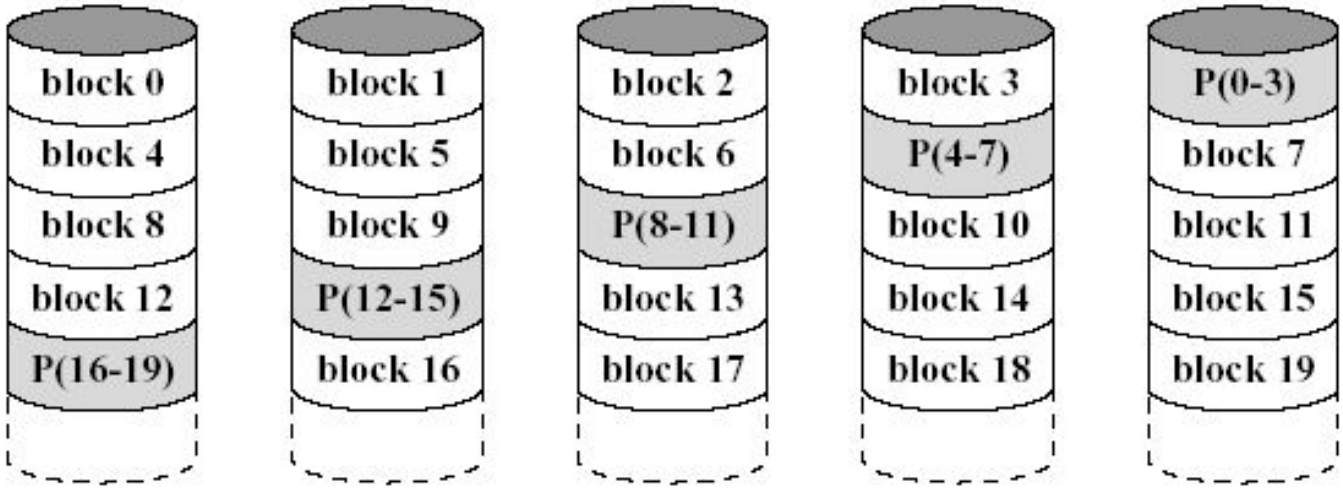
RAID 5

- O RAID 5 pode ser usado com a partir de 3 discos. Independentemente da quantidade de discos usados, sempre temos sacrificado o espaço equivalente a um deles. Em um NAS com 4 HDs de 1 TB, por exemplo, você ficaria com 3 TB de espaço disponível, em um servidor com 10 HDs de 1 TB, você ficaria com 9 TB disponíveis e assim por diante. Os dados continuam seguros caso qualquer um dos HDs usados falhe, mas se um segundo HD falhar antes que o primeiro seja substituído (ou antes que a controladora tenha tempo de regravar os dados), todos os dados são perdidos. Você pode pensar no RAID 5 como um RAID 0 com uma camada de redundância.

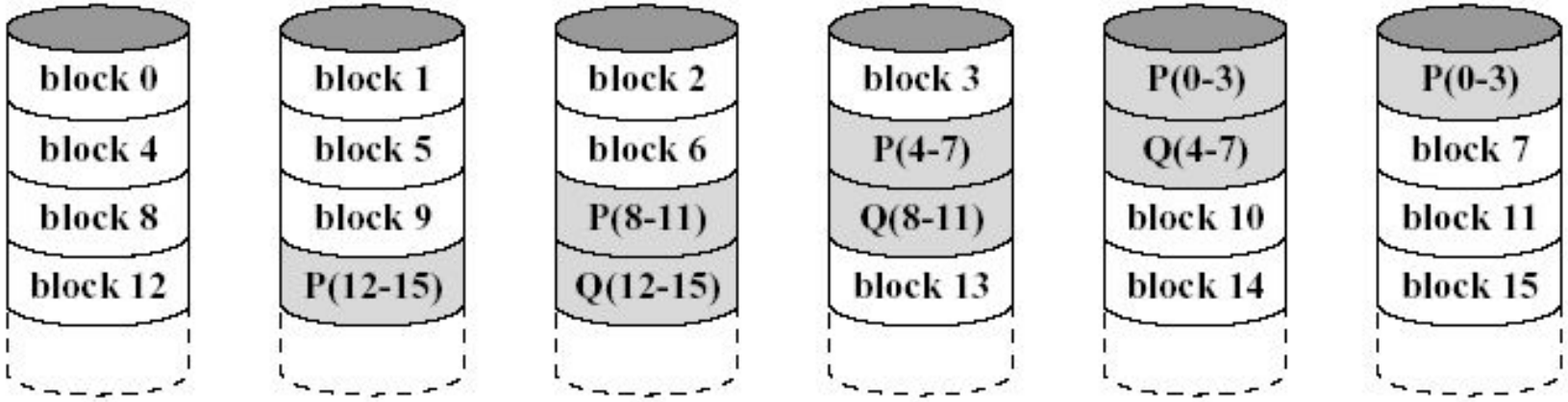
RAID 6

- **RAID 6**: O RAID 6 dobra o número de bits de paridade, eliminando o ponto fraco do RAID 5, que é a perda de todos os dados caso um segundo HD falhe. No RAID 6, a integridade dos dados é mantida caso dois HDs falhem simultaneamente, o que reduz brutalmente as possibilidades matemáticas de perda de dados.
- A percentagem de espaço sacrificado decai com mais discos, tornado- progressivamente mais atrativo. No caso de um grande servidor, com 20 HDs, por exemplo, seria sacrificado o espaço equivalente a apenas dois discos, ou seja, apenas 10% do espaço total. O maior problema é que o RAID 6 exige o uso de algoritmos muito mais complexos por parte da controladora, e não é suportado por todos os dispositivos.

Comparação Raid 5 e 6

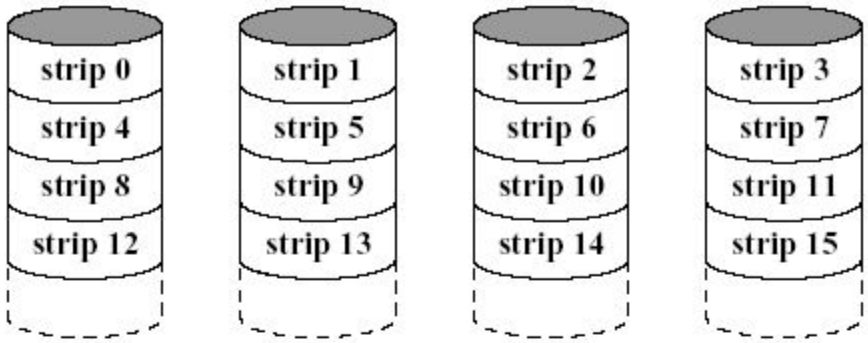


(f) RAID 5 (block-level distributed parity)

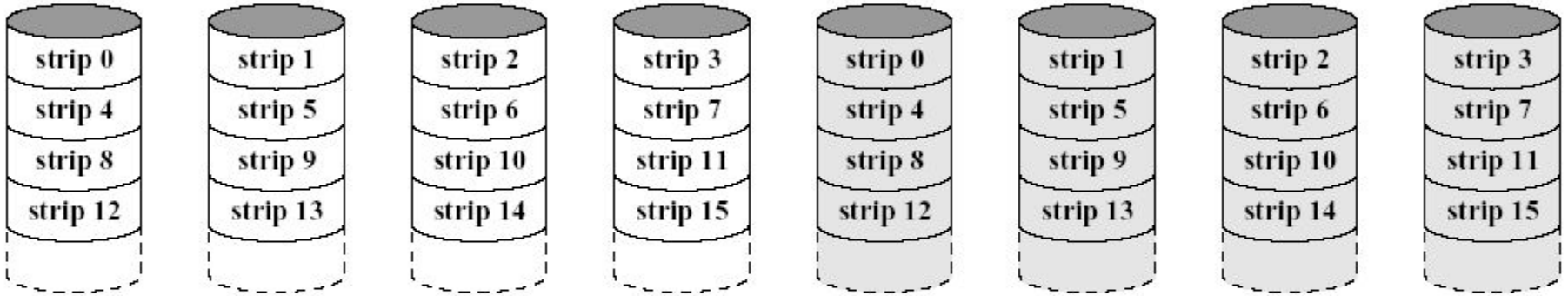


(g) RAID 6 (dual redundancy)

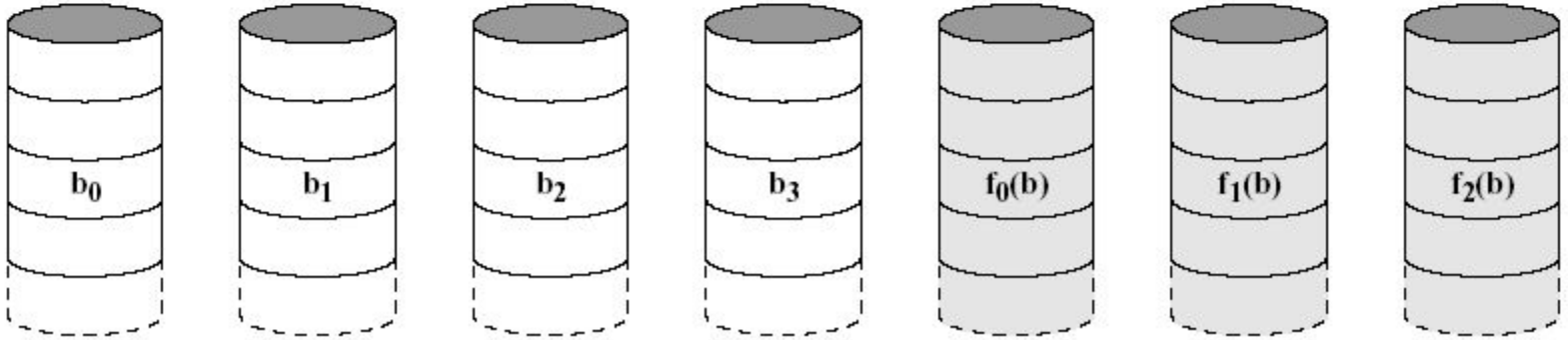
Resumo Raid 0, 1 and 2



(a) RAID 0 (non-redundant)

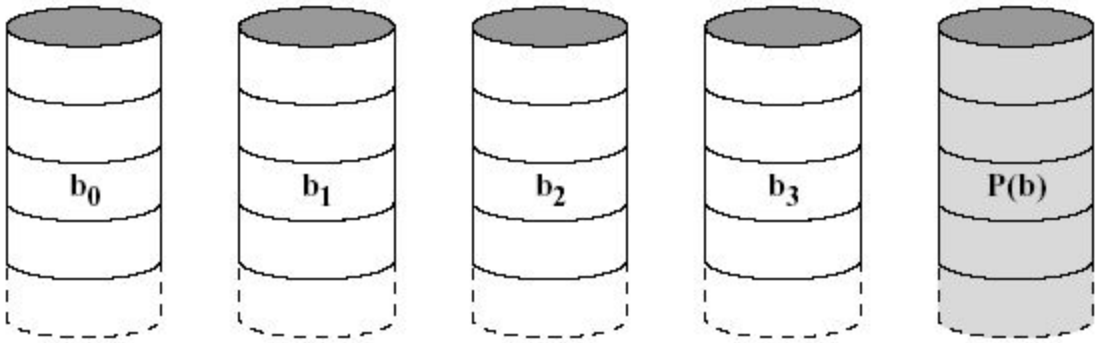


(b) RAID 1 (mirrored)

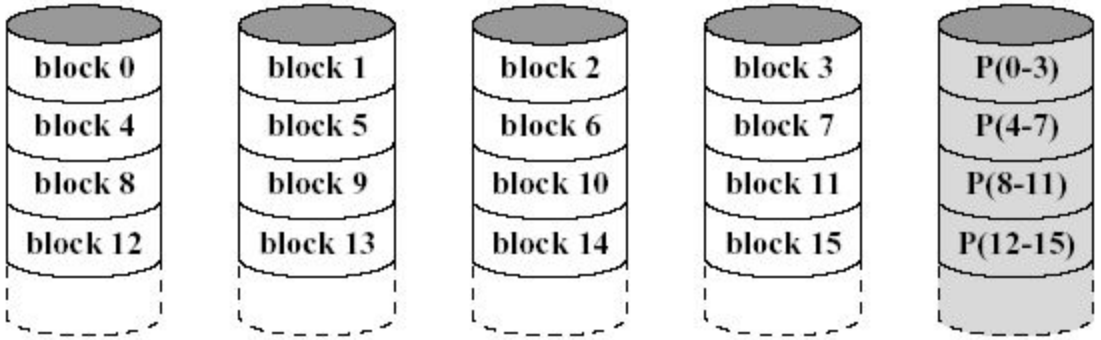


(c) RAID 2 (redundancy through Hamming code)

Resumo Raid 3 e 4



(d) RAID 3 (bit-interleaved parity)

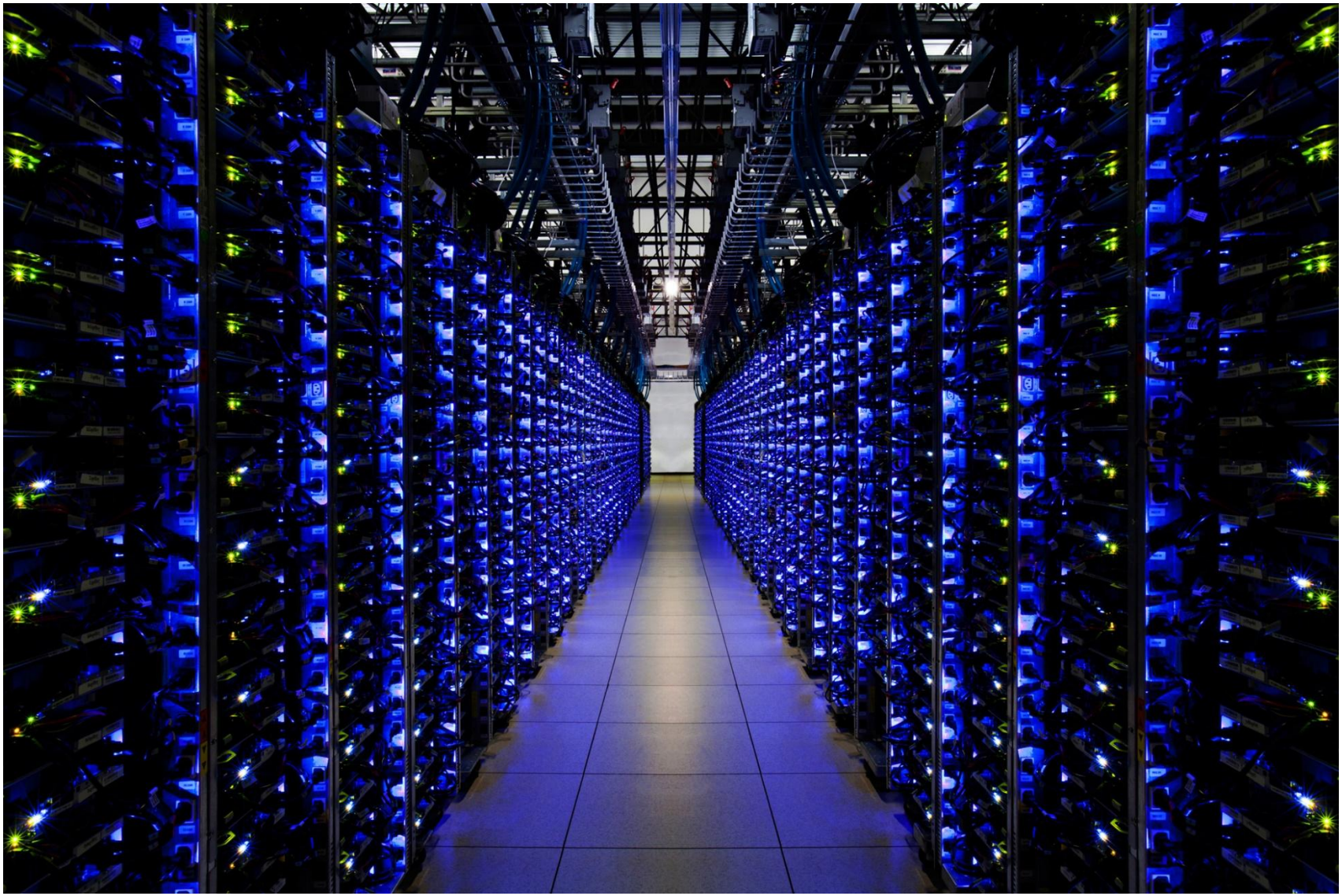


(e) RAID 4 (block-level parity)

JBOD (Just a Bunch Of Disks)

- No JBOD 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 (nesse caso arquivos completos, e não fragmentos como no caso do RAID 0).
- Não existe qualquer ganho de desempenho, nem de confiabilidade,
- Apenas junta vários HDs de forma a criar uma única unidade de armazenamento.
- Não é uma boa opção para armazenamento de dados importantes,

Google Data Center



Google
google.com/datacenters

Google Data Centers

- Street view like
- Virtual Tour

https://youtu.be/ROzBgQURjhs

http://www.google.com/about/datacenters