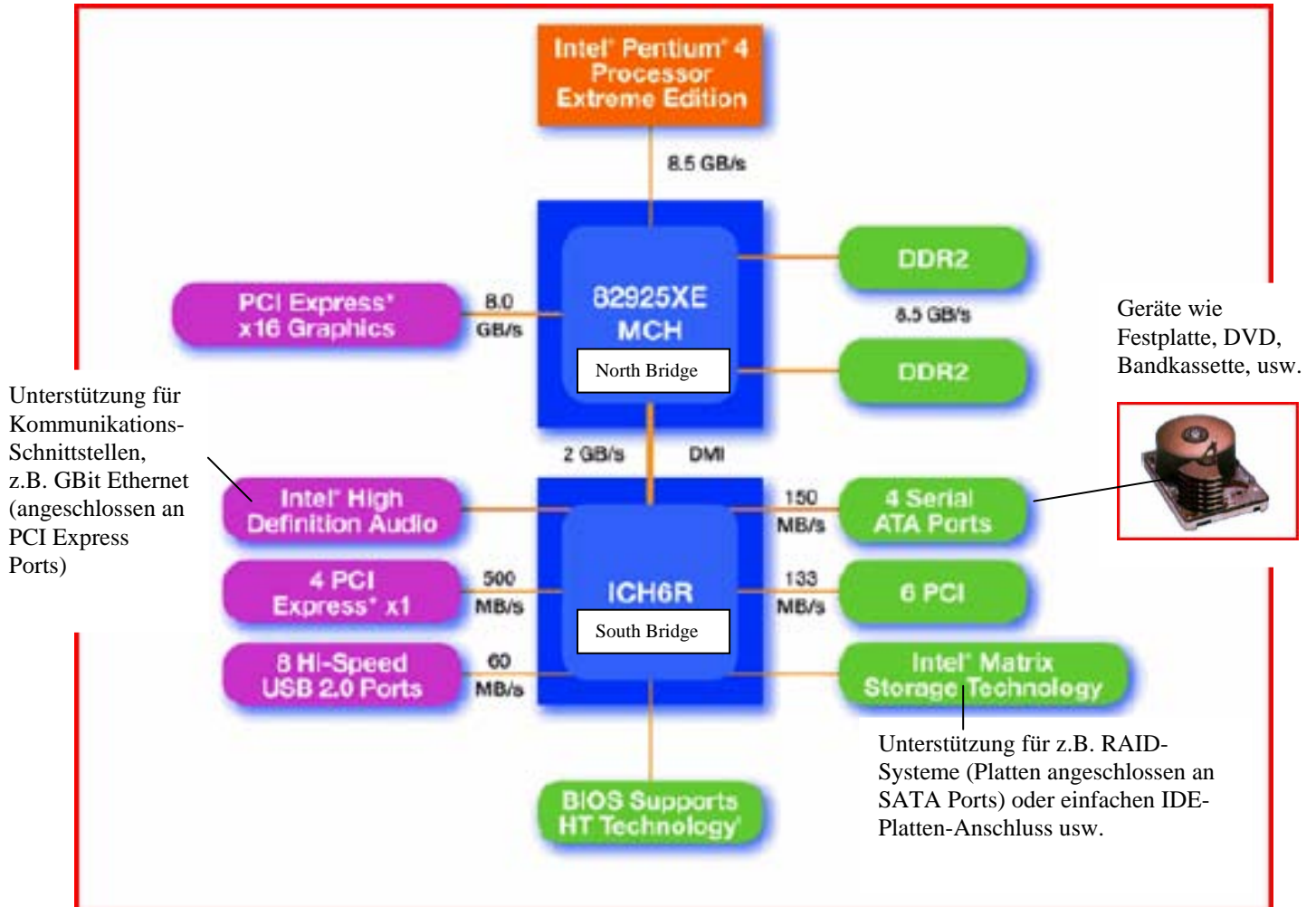


## RECHNER - PERIPHERIE

### Teil 2: Geräte



MCH Memory Controller Hub (Memory Bridge)

DMI Direct Media Interface

ICH6R I/O Controller Hub (I/O Bridge, I=Input, O=Output)

("... 8 multipurpose DMA engines, 4 input and 4 output, support of simultaneous independent data streams")

#### Bild: PC-Architektur (Fa. Intel)

Über die Schnittstellen **PCI**, **SATA** und **USB** sind alle gängigen Peripherie-Geräte (Funktionen) anschließbar.

Weitere Schnittstellen wie **PCI-X**, **Firewire**, **SAS**, **SCSI**, **RS232**, **Enhanced Parallel Port EPP** usw. sind nicht eingezeichnet.

*Frage: wo sind Speichergeräte anschließbar?*

---

## Inhaltsverzeichnis Teil 2

<b>RECHNER - PERIPHERIE</b>	<b>1</b>
<b>1 FESTPLATTE</b>	<b>4</b>
<b>1.1 Aufbau</b>	<b>4</b>
<b>1.2 Datenaufzeichnung</b>	<b>9</b>
1.2.1 Physikalische Ebene	9
1.2.2 Aufzeichnungsverfahren	11
1.2.3 Sektorformat	16
1.2.4 Erzeugen des Sektorformats	18
1.2.5 Selbst-Diagnose	20
1.2.6 Datenrate	21
<b>1.3 Verwaltung des Datenträgers (Dateien)</b>	<b>23</b>
1.3.1 Belegungstabelle der Sektoren	23
1.3.2 Master Boot Record MBR (Boot-Sektor)	24
1.3.3 Sektoranordnung innerhalb einer Datei	26
1.3.4 Fragmentierung und Defragmentierung	27
1.3.5 Abbilden von Sätzen einer Datei auf die Sektoren (Blöcke)	27
<b>1.4 Schnittstelle</b>	<b>29</b>
1.4.1 IDE-Schnittstelle (Integrated Drive Electronic)	29
1.4.2 Kabel	31
<b>1.5 Datenblätter</b>	<b>34</b>
<b>1.6 RAID-Systeme</b>	<b>35</b>
1.6.1 Hardware- und Software-RAID	37
1.6.2 Plattensystem ohne Redundanz (Disk Striping, RAID 0)	40
1.6.3 Spiegelung von Platten (Disk Mirroring, RAID 1)	41
1.6.4 Kombiniertes System RAID0/RAID1	43
1.6.5 System mit Paritätsplatte (RAID 3, RAID 4)	44
1.6.6 System mit rotierender Parität (RAID 5)	47
1.6.7 System mit zwei Paritätsplatten (RAID 6)	50
1.6.8 Technische Realisierung	52
1.6.9 Systembetrachtungen	56
<b>1.7 Speichernetzwerke</b>	<b>57</b>
<b>1.8 Aufgaben</b>	<b>59</b>

<b>2</b>	<b>PORTABLE DATENTRÄGER MIT DIREKT-ADRESSIERUNG</b>	<b>62</b>
2.1	USB Speicherkarten (USB Sticks)	62
2.2	Diskettenspeicher (Floppy Disk, Flexible Disk)	63
2.2.1	1.44MByte-Diskette	64
2.2.2	Disketten mit hoher Kapazität	64
2.3	Optische Speicher	66
2.3.1	Physikalische Ebene bei der Datenaufzeichnung	67
2.3.2	Plattenrotation	74
2.3.3	Datenträgerformat	75
2.3.4	Sektor (Block)-Format CD-ROM	80
2.3.5	Verwaltung des Datenträgers	83
2.3.6	Befehlsschnittstelle CD-R	86
2.3.7	Normen	87
2.4	Aufgaben	88
<b>3</b>	<b>PORTABLE DATENTRÄGER MIT SEQUENTIELLER ADRESSIERUNG</b>	<b>92</b>
3.1	Bandkassetten-Geräte	93
3.1.1	¼ -Zoll-Bandkassetten	93
3.1.2	½ -Zoll-Bandkassetten	93
3.1.3	DAT-Kassette (Digital Audio Tape)	93
3.2	Datenaufzeichnung und Datenträgerformat	94
3.2.1	Datenträger	94
3.2.2	Aufzeichnungsverfahren	94
3.2.3	Bandführung	96
3.2.4	Spur-, Datenblock- und Rahmen-Formate	97
3.2.5	Verwaltung des Datenträgers (Dateien)	100
3.2.6	Technische Daten	102
3.3	Kassetten-Systeme (Auto Changer, Juke Box)	103
3.4	Datensicherung (Backup) und Archivierung	104
3.5	Vergleich Bandgeräte für Backup	106
3.6	Aufgaben	106
<b>4</b>	<b>GEMEINSAME TECHNISCHE DATEN</b>	<b>108</b>
4.1	Zuverlässigkeit (Reliability)	108
4.2	Umgebungstemperatur	111
4.3	Aufgaben	111

# 1 Festplatte

[www.seagate.com](http://www.seagate.com)

(Festplatte, Platte, Magnetplatte, Winchesterplatte, Disk, Disk Drive)

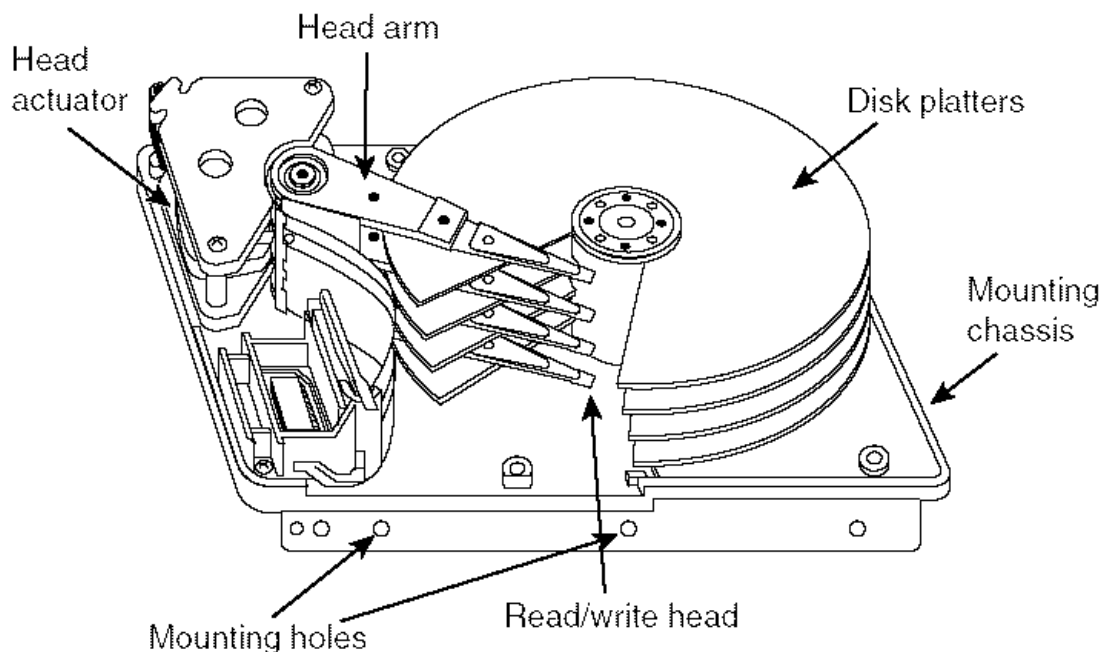
Neben der Geschwindigkeit des Prozessors, der Kapazität und der Zykluszeit des Arbeitsspeichers usw. ... sind die technischen Daten der **Festplatte** (Zugriffszeit, Datenrate, Kapazität) ein wichtiger Faktor für die Leistungsfähigkeit des Computers.

## Vorgehensweise:

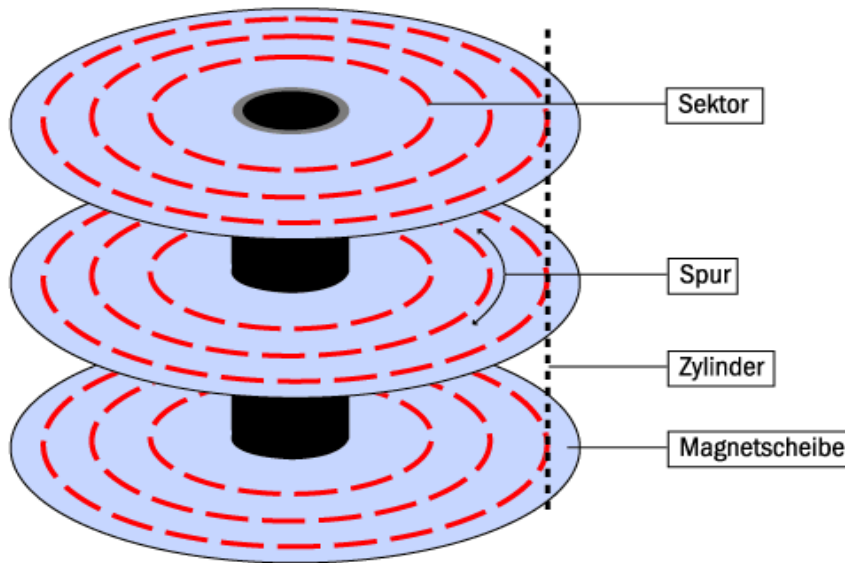
Die Festplatte wird – wie die nachfolgenden Geräte auch – nach folgendem Schichtenmodell diskutiert:

<b>Hardware</b>	Positioniersystem, Servosystem
<b>Aufzeichnungsverfahren</b>	Flußwechselfolge $\leftarrow \rightarrow$ Bitmuster, RLL 2,7
<b>Sektorformat</b>	Zusammenfassung der Bits zu einer <u>adressierbaren Einheit</u> Sektoradresse (Zyl., Kopf, Sektor) Sektorinterleaving
<b>Verwaltung der Platte</b>	<u>Belegungsverzeichnis</u> : Verwaltung (frei, belegt, verkettet) der Sektoren oder entsprechender Zuordnungseinheiten <u>Inhaltsverzeichnis</u> : Zugriff auf Dateien
<b>Abbildung von Sätzen auf Sektoren</b>	Satz- und Blockformate
<b>Geräteschnittstelle (Interface)</b>	

## 1.1 Aufbau



**a) Datenträger:**



© tecChannel

- **Stapel** = bis max. ca. 12 Platten, Umdrehungszahl konstant mit ca. 1.500..15.000 U/min (herstellerabhängig, häufig: 5.400, 7.200, 10.000, 15.000)
- **Platte** (2mm Alu) beidseitig mit Magnetschicht (um), 2 Oberflächen
- **Oberfläche** = Datenoberfläche (Datenspuren + Landespur) oder Servooberfläche (Servospuren + Landespur)
- **Spur** = konzentrischer Kreis, Daten bitseriell beginnend mit MSB eines Bytes (MSB = most significant Bit) bis ca. 100 000 Spuren pro Oberfläche, Spurabstand ca um
- **Datenkapazität** pro Spur:
  - alle Spuren mit gleicher Kapazität (alt):  
Innerste Spur mit höchster Bitdichte (bpi), höchster Flußwechselfichte (FW-Dichte) und damit kleinster FW-Abstand. Innerste Spur stößt an technologische Grenze, äußerste Spur mit kleinerer Bitdichte.
  - alle Spuren mit gleicher Bitdichte (aktuell):  
Damit alle Spuren mit gleicher FW-Dichte. Spurkapazität abhängig von Spurradius. Äußerste Spur hat höchste Datenkapazität. Diese Methode liefert die maximal mögliche Plattenkapazität, da alle Spuren die maximal mögliche Kapazität besitzen.
- **Zylinder** = alle Spuren, die ohne Neupositionieren erreichbar sind.  
Bei z.B. 4 Platten mit je 2 Oberflächen sind maximal  $4 \cdot 2 = 8$  Spuren pro Zylinder möglich. (Ausnahmen bei > 1 Kopf pro Oberfläche)
- **Sektor** = Teil einer Spur = kleinste adressierbare Einheit.

**Sektor-Adresse:**

	Positionieren, ..... ms (Zyl=0 ist außen, Bootsektor)
	Kopfschalten ... us oder ..ms (je nach Verfahren des Servosystems)
	Nr. auf Spur (Drehwartezeit, ..... ms), übliche Sektorgrößen: 128Byte .. 512Byte .. pro Datenfeld

- **Index** = Signal von Geräteelektronik, liefert festen Bezugspunkt auf Spur (Spuranfang).
- **Positioniersystem:**
  - Alle Schreib/Lese-Köpfe und Servo-Kopf starr verbunden ( alle Köpfe auf dem gleichen Zylinder).
  - Von Motor oder Magnetspule auf den Zylinder bewegt (positioniert)
  - Köpfe schweben auf Luftpolster (< 1um) über der Platte. Bei Abschalten Absenken der Köpfe auf Platte (Federkraft). Landespur. Bremse um Schleif-Zeit zu verringern.
- **Servo-Kopf:**  
Liest von Servospuren der Servooberfläche Steuersignale für Positioniersystem. Ermöglicht genaues Positionieren. z.B.: 4 Platten besitzen 7 Datenoberflächen + 1 Servooberfläche. Siehe Kapitel „Positioniersystem“.
- Derzeit: es überträgt immer nur 1 Schreib/Lese-Kopf Daten ( und zusätzlich der Servo-Kopf).

*Frage: wie würde ein paralleler Transfer über mehrere Köpfe funktionieren?  
kennen Sie Platten, bei denen gleichzeitig von mehreren Spuren gelesen wird?*

• **Geräte-Einteilung**

- Festplatte (fixed disk), Wechselplatte, Plattenkassette (disk cartridge), RAM-Disk
- bewegliche Köpfe, (feste Köpfe)
- Formfaktor (.. 1,8'' .. 3½'' ..)

• **Positioniersystem**

- bewegt Köpfe zum gesuchten Zylinder (Positionierzeit  $t_p$ )
- führt Köpfe in Spurmitte: maximales Lesesignal

a) **Geregeltes System (Closed loop system)**

**Regelung:** Servooberfläche = „dedicated servo“

Servokopf liest Steuerinformation von Servospuren

→ Regelung Positioniergeschwindigkeit

→ Kopfführung spurmittig: kleine Spurbreite, hohe Spurdichte

b) **Geregeltes System mit eingebetteter (integrierter) Servoinformation (embedded servo) , Anwendung: Platte**

Integrierte Servoinformation: Servoinformation steht auf den Datenspuren zwischen den Sektoren, siehe Bild.

Ermöglicht höhere Spurdichte als bei „dedicated servo“

Nachteil von „embedded servo“:

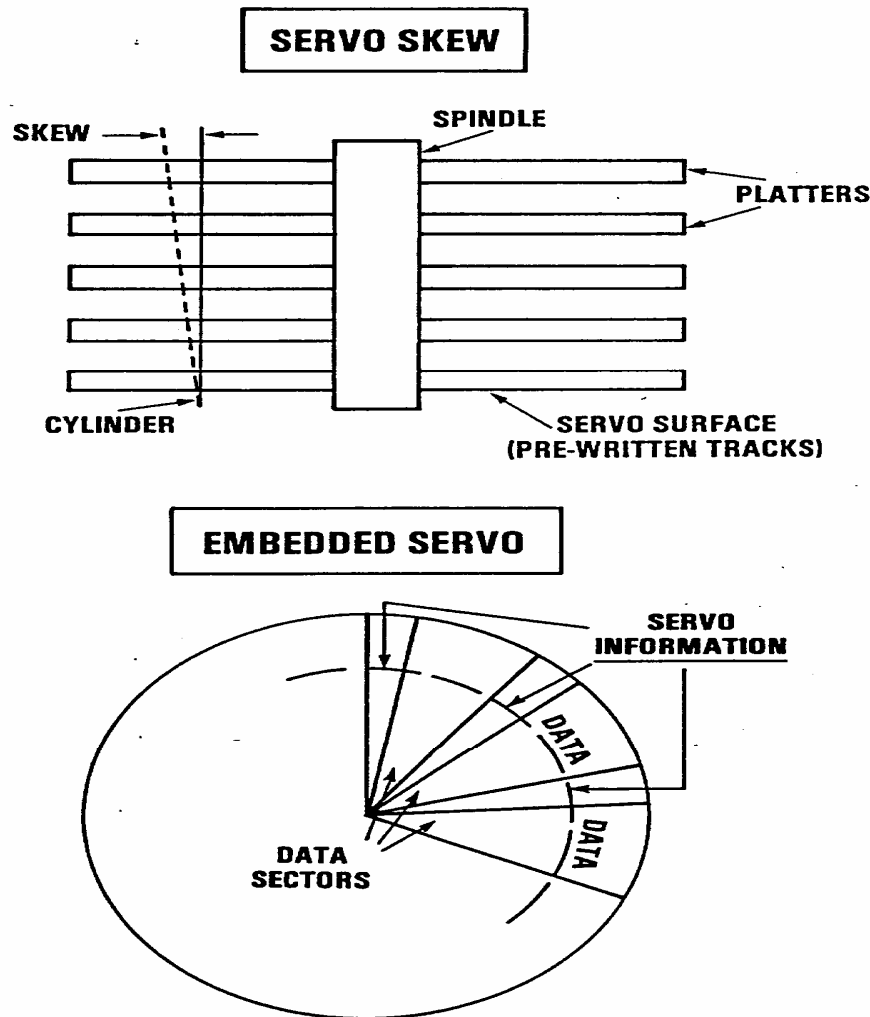
Kopfschaltzeit wesentlich höher (ms anstatt  $\mu$ s), da zuerst Servo-Info zwischen den Sektoren gelesen wird und dann erst der Sektor.

b) **Definitionen:**

Für die Berechnung der Zeiten für „Datei Schreiben“ oder „Datei Lesen“ sind folgende Kennwerte wichtig:

$t_p$ <b>Positionierzeit oder Suchzeit (seek time)</b>	Wartezeit infolge Positionierbewegung des Schreib/Lesekopfes / $t_p$ mittlere Positionierzeit, $t_{pmax} = \max. Pos.$ (z.B. / $t_p = 8ms$ .. Platte)
$t_l$ <b>Drehwartezeit (latency time)</b>	Wartezeit infolge Plattendrehung, $t_{lmax} = t_u$ mit $t_u =$ Umdrehungszeit / $t_l =$ mittlere Drehwartezeit = $\frac{1}{2} t_u$ (z.B. .. 5ms .. Platte)
$t_{acc}$ <b>Zugriffszeit (access time)</b> = $t_p + t_l$	von Beginn Positionieren bis Beginn Schreib/Lesevorgang
$t_{transf}$ <b>Transferzeit</b>	von Beginn erstes Bytes Lesen/Schreiben bis Ende letztes Bytes Lesen/Schreiben

**Platte mit „eingebetteter“ Servoinformation (zu Kap. „Aufbau“):**



Einfluß Positioniersystem auf Plattenkapazität:

**Kapazität** ~ **Bitdichte (bpi) auf Spur** \* **Spurdichte (tpi)**  
 Flusswechseldichte auf Spur + Aufzeichnungsverfahren → **Bitdichte**  
 Positioniersystem → **Spurdichte**

**c) Datenspur: Kapazität und Datenrate → grosser Einfluss auf Datei-Transferzeit**  
 Konstante Bitdichte auf allen Spuren (engl. ZBR = Zoned Bit Recording)

	<b>ZBR (aktuell)</b>
<b>Bitdichte</b>	auf allen Zylindern (Spuren) gleich (= maximal)
<b>Spurkapazität, Sektorzahl pro Spur</b>	steigt nach außen hin, da äußere Spuren länger
<b>Kapazität</b>	maximal bei vorgegebener Bitdichte
<b>Physikalische Datenrate</b>	steigt von innen nach außen, z.B. 20 → 50 MByte/s

**ZBR** = Maximal mögliche Flußwechseldichte und damit maximale Bitdichte auf allen Spuren (Zylindern). Die Sektorzahl auf der Spur hängt hier von der Zylinder-Nr. ( $\cong$ Radius) ab. Die weiter außen liegenden (=längeren) Spuren liefern bei gleicher Umdrehungszahl eine höhere Lineargeschwindigkeit und damit eine höhere Datenrate.

**Auswirkung auf die Datenrate:**

Zyl-Nr. 0 liegt außen, d.h., beim erstmaligen Beschreiben der Platte mit Daten ist zuerst die höhere Datenrate wirksam. Mit zunehmendem „Füllstand“ sinkt die Datenrate, da dann auf den inneren Zylindern abgespeichert wird.

**Aufteilung des Zylinderbereichs einer Festplatte in Zonen (erste 20 GByte):**

Zone	Tracks in Zone	Sectors Per Track	Data Transfer Rate (Mbits/s)
0	624	792	372.0
1	1,424	780	366.4
2	1,680	760	357.0
3	1,616	740	347.6
4	2,752	720	338.2
5	2,880	680	319.4
6	1,904	660	310.0
7	2,384	630	295.9
8	3,328	600	281.8
9	4,432	540	253.6
10	4,528	480	225.5
11	2,192	440	206.7
12	1,600	420	197.3
13	1,168	400	187.9
14	1,815	370	173.8

**Diskussion!**

max. und min. Datenrate in Byte/s ?

maximale Kapazität erreicht?

(Quelle: IBM, Product Manual)

Aufgabe: Plattenspeicher-Kapazität

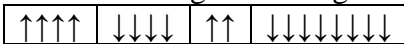
## 1.2 Datenaufzeichnung

### 1.2.1 Physikalische Ebene

Die Ausführungen betreffend magnetische Datenaufzeichnung gelten für alle magnetischen Speichergeräte wie **Festplatte (disk)**, **Diskette**, **Band (tape)**, **Bandkassette (tape cartridge)**. Die Daten sind in der Oberfläche des Datenträgers, bestehend aus einer dünnen ( $\approx \mu\text{m}$ ) magnetischen Schicht, gespeichert. Der Datenträger wird gegenüber einem Schreib/Lese-Kopf bewegt.

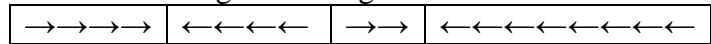
Magnetisierung der Datenoberfläche (Spur):

**vertikale Magnetisierung:**



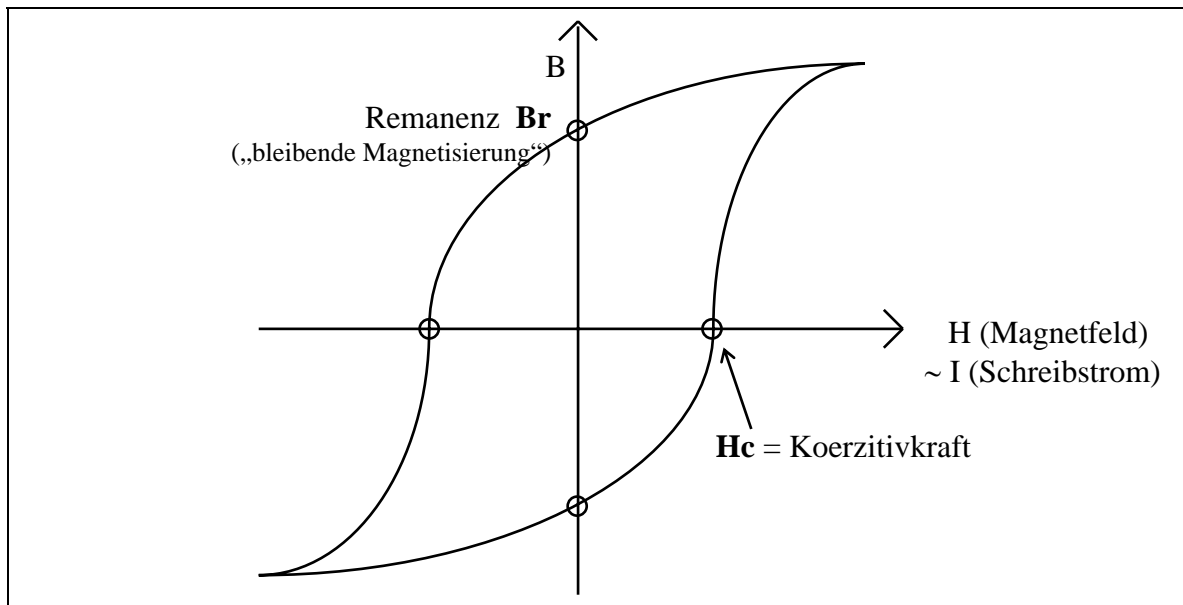
← Spur →

**horizontale Magnetisierung:**



Frage: Vergleich vertikale ↔ horizontale Magnetisierung

### a) Magnetisierungskurve



**Bild: Magnetische Induktion B als Funktion des Magnetfelds H (Schreibstrom I)**

$H$  = von außen angelegtes Magnetfeld (Schreibkopf)

$B$  = magnetische Induktion der Datenträger-Oberfläche

Remanenz  $B_r$ :

Verbleibende Magnetisierung der Datenträger-Oberfläche (Spur) nach dem Schreibvorgang. Sie beeinflusst die Höhe des Lesesignals.

**Lesesignal  $\sim d\Phi/dt$  mit  $d\Phi/dt \sim B_r$ .**

$d\Phi/dt$  ist die zeitliche Änderung des magnetischen Flusses  $\Phi$  in der Spule des Lesekopfes.

Koerzitivkraft  $H_c$ :

Ist ein Maß dafür, wie groß das durch den Schreibstrom erzeugte Magnetfeld sein muß, damit die Magnetisierung der Spur geändert werden kann.

$H_c$  groß: → →



**Für alle magnetischen Datenoberflächen gilt:**

- Information in Form von magnetischen Flußwechseln (FW) gespeichert
- Aufzeichnungsverfahren bestimmt Zuordnung FW-Folge ↔ Bitmuster
- Verfahren: FM, MFM, GCR, etc.
- Flußwechsel liefert Daten-Signal und/oder Synchronisier-Signal für Lesetak im Gerät (Takt-Signal)
- Gespeicherte Information ist nicht flüchtig bei Versorgungsspannung aus
- FW-Dichte , z.B. 50 000fci (flux changes per inch)
- Spur = Aufzeichnungsbereich auf Datenoberfläche unter Schreib/Lese-Kopf
- Bit-Dichte auf Spur = FW-Dichte \* Bit/FW

FW-Dichte = f(magn. Material, Schreib/Lese-Kopf)

Bit/FW = f(Aufzeichnungsverfahren)

**MR-Köpfe (MR = magnetoresistiv)**

Magnetoresistive Leseköpfe erlauben höhere Bitdichten als induktive Leseköpfe, deshalb ist derzeitiger Stand der Technik:

Schreibkopf induktiv, Lesekopf magneto-resistiv.

Konstanter Strom durch magnetoresistive Schicht des Lesekopfes. Der elektrische Widerstand ändert sich durch das Magnetfeld der vorbeidrehenden Spur. Bei gleichem Flußwechselabstand sind die Lesesignale wesentlich höher als beim induktiven Verfahren. Bei gleicher Signalamplitude des Lesesignals können die Bitzellen verkürzt werden und die Spurbreiten verringert werden was höhere Flußwechseldichten (Bitdichten) und höhere Spurdichten liefert und somit zu höheren Kapazitäten führt.

Zusätzlich wurde die PRML-Lesekanal-Technologie eingeführt (PRML=Partial Response Maximum Likelihood). Damit wird ein statistisches Verfahren zur Auswertung der Lesesignale bezeichnet. Die Lesesignalfolge wird mit Sollmustern verglichen und ausgewertet. Dies erhöht die Sicherheit beim Lesen (Lesefehlerwahrscheinlichkeit derzeit ca.  $10^{-14}$ ).

### **1.2.2 Aufzeichnungsverfahren**

#### **a) Einführung**

Das Aufzeichnungsverfahren legt die Zuordnung „magnetische Flußwechselfolge“ ↔ „Bitfolge“ fest.

Magnetische FW-Folge, erzeugt durch aufeinanderfolgende kleine magnetische Bezirke auf der Spur.

Magnetisierung der Bezirke:

"nach rechts" oder "nach links" bei horizontaler Magnetisierung: → oder ←

"nach oben" oder "nach unten" bei vertikaler Magnetisierung: ↑ oder ↓

#### **Forderungen:**

1) Verfahren soll .

2) Verfahren muß

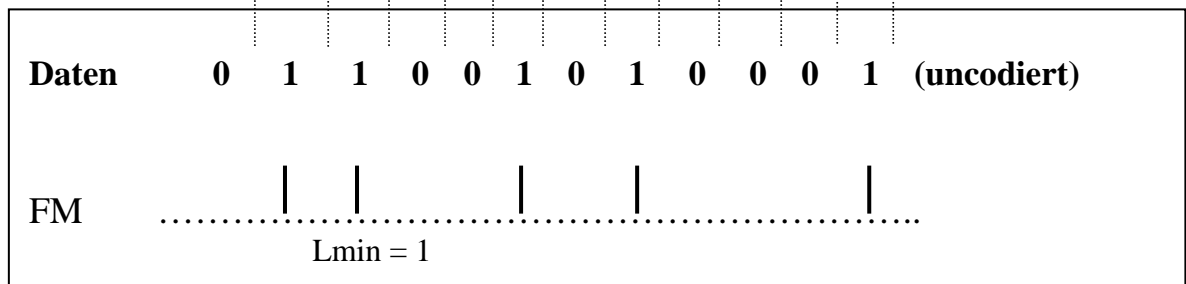
**b) Aufzeichnungsverfahren:**

**(1) „ideales“ Aufzeichnungsverfahren FM (frequency modulation) mit Trennung von Daten-Flußwechsel und Takt-Flußwechsel**

*Zeichnen Sie den Schreibstrom bei FM mit der Vorgabe:*

*Flusswechsel immer zu Beginn einer Bitzelle(Taktflußwechsel) und in der Mitte einer Bitzelle nur, falls Bit=1 (Daten-Flußwechsel)*

*Tragen Sie den minimalen Flußwechselabstand  $L_{min}$  ein*



**(2) „derzeit aktuelles“ Aufzeichnungsverfahren RLL (run length limited)**

--Daten werden codiert:  $n \text{ Bit} \rightarrow 2n \text{ Bit}$

--„1“ ergibt Flußwechsel, „0“ ergibt keinen Flußwechsel **keine** Takt-Flußwechsel  
(Flußwechsel = Umschalten Schreibstrom  $I+ \leftrightarrow I-$ )

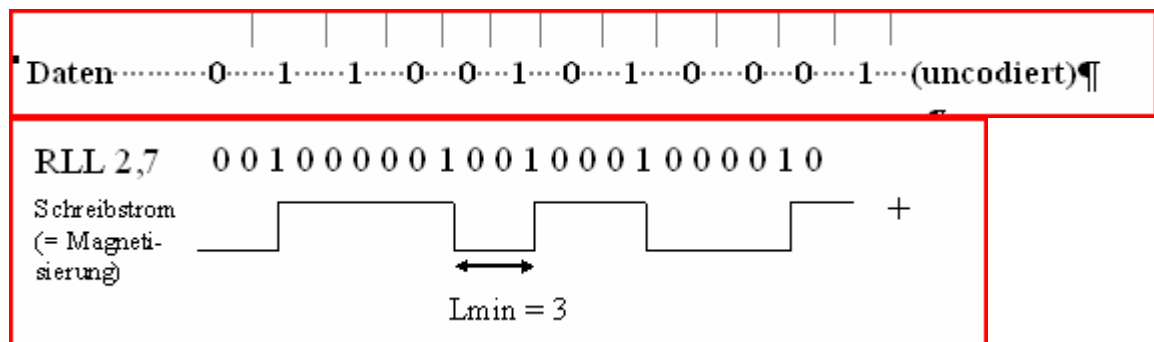
**RLL2,7-Verfahren**

Bitmuster so codiert, daß

(1) immer  $\geq 2$  „0“ zwischen benachbarten „1“  $\rightarrow$  großer FW-Abstand

(2) immer  $\leq 7$  „0“ zwischen benachbarten „1“  $\rightarrow$  spätestens in der achten Bitzelle wieder ein Datensignal (auch als Synchronisiersignal verwendbar). Damit können Synchronisier-FW ganz entfallen.

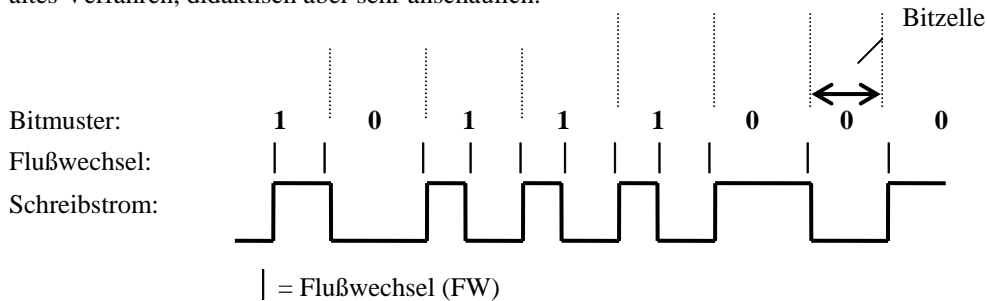
siehe auch Kapitel c) mit Codier-Tabelle und Kapitel d).



Kapazitätsgewinn von RLL2,7 gegenüber FM: Faktor 3 (bei gleicher Technologie)

**c) Detailinformation:**

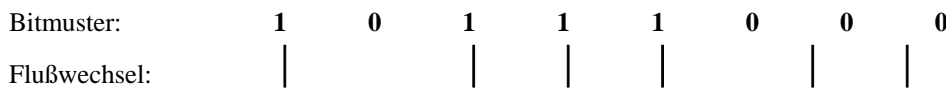
- **FM** (frequency modulation, Wechseltaktschrift):  
altes Verfahren, didaktisch aber sehr anschaulich.



Trennung von Daten- und Synchronisier-FW.

- Synchronisier-FW definiert Beginn Bitzelle.
  - Daten-FW in Mitte Bitzelle: „1“ → Flußwechsel, „0“ → kein Flußwechsel
- Ideal für Synchronisierung aber 2 FW pro Bit notwendig → niedrige Bitdichte

- **MF**M (modified frequency modulation oder Miller-Code, modifizierte Wechseltaktschrift):  
angewendet bei 1.44MByte Diskette



Daten-FW auch für Synchronisierung verwendet, dadurch Wegfall der Synchronisier-FW. Nur bei „0“-Folgen noch Synchronisier-FW (zwischen benachbarten „0“ liegt ein Synchronisier-FW).

„1“ → Flußwechsel, „0“ → kein Flußwechsel

--Aufwendiger bei Synchronisierung

--1 FW pro Bit notwendig:

--Bei gleicher Bitdichte wie FM doppelter FW-Abstand.

--Bei gleicher FW-Dichte wie FM: doppelte Bitdichte und damit doppelte Kapazität, doppelte Datenrate (bei gleicher Umdrehungszahl).

- **Aktuell: RLL2,7**-Verfahren (run length limited)

Bitmuster so codiert, daß

(1) immer  $\geq 2$  „0“ zwischen benachbarten „1“ → großer FW-Abstand

(2) immer  $\leq 7$  „0“ zwischen benachbarten „1“ → spätestens in der achten Bitzelle wieder ein Datensignal (auch als Synchronisiersignal verwendbar). Damit können Synchronisier-FW ganz entfallen.

MF	
Daten uncodiert	Daten codiert
0	K0
1	01

K ist Komplement des vorhergehenden Code Bits

RLL2,7	
Daten <u>uncodiert</u> , vom Anwender	Daten codiert, auf Platte abgespeich.
000	000100
10	0100
010	100100
0010	00100100
11	1000
011	001000
0011	00001000

#### d) Vergleich der Aufzeichnungsverfahren: MFM ↔ RLL2,7

Das Aufzeichnungsverfahren wirkt sich bei einem Speichergerät auf Kapazität, Datenrate und die Anforderungen an die elektrischen Eigenschaften (Frequenzverhalten des Schreib/Lese-Kopfes) aus. Es werden aus didaktischen Gründen die Vorteile und die Probleme aufgezeigt, die bei Plattenspeichern durch den Wechsel vom MFM- zum RLL2,7-Aufzeichnungsverfahren (Austausch der Plattencontroller) in der Vergangenheit aufgetreten sind.

##### • Vergleich des Flußwechselabstands:

###### MFM:

Codieren 1 Bit → 2 codierte Bit, siehe Tabelle.

Damit Toggle-Flipflop für Schreibstrom ansteuerbar.

„1“ = Schreibstrom umschalten → Flußwechsel.

{Toggle-FF. T=1 Q=/Qo (=Umschalten), T=0 Q=Qo (= kein Umschalten)}

Bspl. ST506-Schnittstelle (=physikalische Schnittstelle):

vom Anwender (uncodiert): 5MBit/s

codierte Daten: 10MBit/s = 0.1us/Bit

FW-Abstand:  $L_{min}=2*0.1us=0.2us$ ,  $L_{max}=4*0.1us=0.4us$

###### RLL2,7:

Codieren 1 Bit → 2 codierte Bit, siehe Tabelle.

Damit Toggle-Flipflop für Schreibstrom ansteuerbar.

„1“ = Schreibstrom umschalten → Flußwechsel.

vom Anwender (uncodiert): 5MBit/s

codierte Daten: 10MBit/s = 0.1us/Bit

FW-Abstand:  $L_{min}=3*0.1us=0.3us$  !,  $L_{max}=8*0.1us=0.8us$  !

##### • Vergleich von Kapazität und Datenrate:

Gleiche Bitdichte vorausgesetzt:

$L_{RLL} = 0.3us$ ,  $L_{MFM} = 0.2us$  →  $L_{RLL} = 1.5 * L_{MFM}$  !

Gleichen minimalen FW-Abstand vorausgesetzt,  $L_{RLL} = L_{MFM}$  (=gleiche technologische Grenze):

$L_{min}=0.2us$

• Bitdichte auf Spur: Bitdichte-RLL = 1.5 \* Bitdichte-MFM

• Speicherkapazität einer Spur:

Kapazität-RLL = 1.5 \* Kapazität-MFM → 50% Erhöhung!

• Datenrate am Schreib/Lese-Kopf (= codierte Daten):

MFM: 10MBit/s

RLL: 15MBit/s !

MFM (bei ST506-Schnittstelle):

5MBit/s mit 17Sektoren, 512Byte/Sektor

$/R=17*512*60U/s=522\ 240$  Byte/s bei 3000U/s

RLL2,7 (bei ST506-Schnittstelle):

7.5MBit/s mit 26 Sektoren, 512Byte/Sektor

$/R=26*512*60U/s=798\ 720$  Byte/s bei 3000U/s

→ 50% Erhöhung!

• **Notwendiger Frequenzbereich der Schreib/Lese-Elektronik für das Schalten des Schreibstroms:**

**MFM:**

$$L_{\min}=0.2\mu\text{s} \quad \cong T = 2*0.2\mu\text{s} \quad \cong 2.5\text{MHz} = f_{\max}$$

$$L_{\max}=0.4\mu\text{s} \quad \cong T = 2*0.4\mu\text{s} \quad \cong \underline{1.25\text{MHz} = f_{\min}}$$

T=Periodendauer der Grundwelle des Schreibstroms am Schreibkopf

**RLL:**

$$L_{\min}=0.3\mu\text{s}/1.5 \quad \cong T = 2*0.2\mu\text{s} \quad \cong 2.5\text{MHz} = f_{\max}$$

$$L_{\max}=0.8\mu\text{s}/1.5 \quad \cong T = 2*0.53\mu\text{s} \quad \cong \underline{0.9\text{MHz} = f_{\min} !}$$

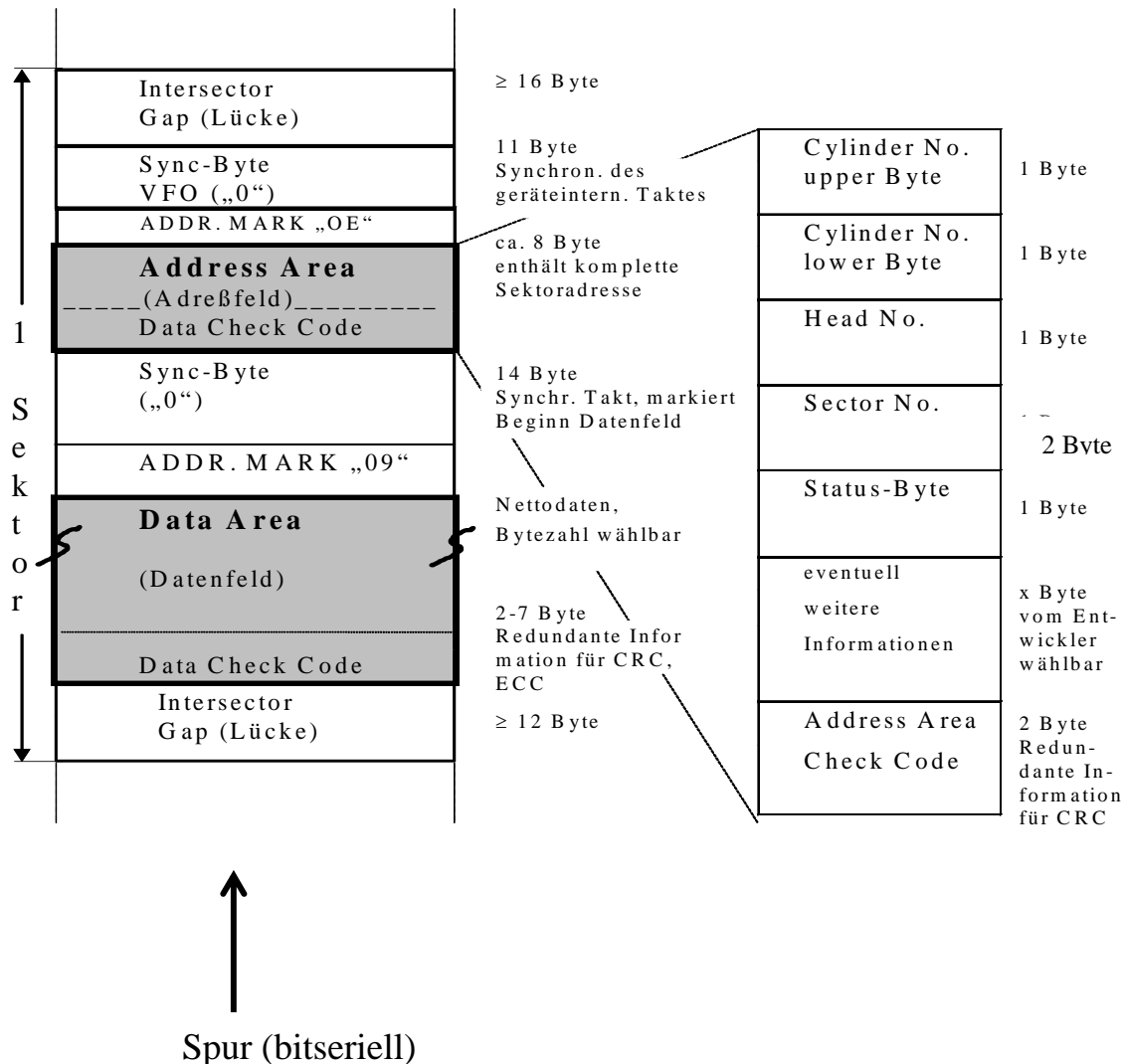
RLL benötigt breiteres Bandpaß-Verhalten (0.9 .. 2.5MHz) der Schreib-Lese-Elektronik und des Schreib/Lese-Kopfes.

Die Datenraten entsprechen den damaligen Werten!

### 1.2.3 Sektorformat

#### a) Standard-Format

(SMD-Schnittstelle, Quelle Fujitsu)



- Das **Adressfeld** enthält die komplette Sektoradresse. Vor jedem Zugriff auf das Datenfeld (Lesen oder Schreiben) wird von der Steuerung zuerst die Sektoradresse gelesen und mit der gesuchten Sektoradresse verglichen. Damit ist sichergestellt, daß nur auf den gesuchten Sektor zugegriffen wird (große Datensicherheit). Positionierfehler würden sofort erkannt werden.

Das Adreßfeld ist CRC-Code gesichert (CRC = cyclic redundant code). Das Adreßfeld enthält am Ende 2Byte redundanten CRC-Code. Nach dem Lesen des Adreßfelds wird in der Steuerung der CRC-Code neu gebildet und mit dem gelesenen CRC-Code verglichen. Bei Abweichung liegt ein fehlerhaftes Adreßfeld vor. Der Sektor wird nicht weiter bearbeitet.

- Das **Datenfeld** ist in der Regel mit einem redundanten Code gesichert, der Fehlererkennung (EDC = error detection code) und Fehlerkorrektur (ECC = error correction code, ca. 7 Byte) erlaubt. Korrigierbar sind meist Büschelfehler bis zu einer Breite von einigen Bits. Bei nichtkorrigierbaren Fehlern im Datenfeld kann der Sektor im Statusbyte des Adreßfelds als defekt gekennzeichnet werden. Ein Verweis auf einen Erstzsektor ist möglich.

Die Größe des Datenfelds liegt üblicherweise bei 128Byte .. 8kByte, bei PC-Platten 512 Byte.

- **Kapazität einer Spur**

- unformatierte Kapazität** =

- (wird auch als Bruttokapazität bezeichnet),  
bpi = Bit per inch, 1 inch = 2.54cm

- formatierte Kapazität** =

- (wird auch als Nettokapazität bezeichnet).

- Das **Sektorformat** und damit auch der Inhalt des Adreßfelds werden beim Formatiervorgang (Low-Level-Formatieren) erzeugt. Der Entwickler der Formatierroutine kann daher festlegen, welche Inhalte das Adreßfeld neben der Sektoradresse noch besitzt, z.B. eine fortlaufende logische Sektornummer etc..

### b) Erweitertes Sektorformat

Das bisherige Sektorformat ist um ein Schlüsselfeld erweiterbar:

Sektor = Adreßfeld+**Schlüsselfeld**+Datenfeld

Das Schlüsselfeld ist ein kleines Datenfeld ( $\leq 255$  Byte). Es enthält Daten, die der Kennzeichnung des Datenfelds dienen, z.B. bei einer Kundendatei das Sortierkriterium (=Kundennummer). Durch Zugriff auf das wesentlich kleinere Schlüsselfeld kann in vielen Fällen entschieden werden, ob das längere Datenfeld überhaupt gelesen werden muß. Dadurch wird unnützes Lesen von Datenfeldern und unnützes Belegen des Lesekanals vermieden.

Die Länge von Schlüsselfeld und Datenfeld sind im Adreßfeld des Sektors angebbbar. Dies ermöglicht z.B. unterschiedliche Längen (Kapazität) der Datenfelder auf der Platte.

### c) Spurformat

Den Datensektoren einer Spur sind eine „Home Address“ und ein „Spurkennblock“ mit Spuradresse, Spurzustandsbyte und Verweisen auf eine eventuelle Ersatzspur vorangestellt. Damit sind nur ganze Spuren als "defekt" kennzeichenbar und nicht mehr einzelne Sektoren.

### 1.2.4 Erzeugen des Sektorformats

Beim Vorgang "Low-Level-Formatieren" (physikalisches Formatieren) wird auf den Spuren der Plattenoberflächen (Plattenstapel) das Sektorformat mit Adreßfeld und Datenfeld erzeugt. Die Spuren sind die vom Schreibkopf magnetisierten Bereiche der Plattenoberfläche; der Spurbstand (Zylinderabstand) auf einer Oberfläche wird durch das Positioniersystem festgelegt.

#### a) Ablauf beim Low-Level-Formatieren

- Beginnend bei Zyl. 0/ Kopf 0/ Sektor 1 bei Position = "Spurbeginn" (Index-Puls):  
Schreiben Adressfeld unter Berücksichtigung des Interleavingfaktors.  
Schreiben Datenfeld eines jeden Sektors mit Testmuster.
- Kontroll-Lesen aller Sektoren (Adreßfeld, Datenfeld). Fehler werden über CRC oder über Datenvergleich in der Steuerung erkannt.  
Fehler: Registrierung der Adresse des defekten Sektors und eventuell Ersatzsektorzuweisung.

Nach dem Low-Level-Formatieren sind alle Datenfelder gelöscht (überschrieben mit Testmuster) und nicht mehr restaurierbar.

#### b) Befehlsparameter beim Low-Level-Formatieren (aus SCSI-2 Norm)

**Befehl „Mode Select“:** Übergabe der Parameter für den Formatier-Befehl

**Befehl „Format Unit“:** Formatiert die Platte, gesteuert von den Parametern

**Befehlsparameter** (Hersteller-spezifisch, Beispiel):

- „**Interleaving-Faktor**“ 1..32
- „**Spur-Versatz** (track skew)“: ..... (Anzahl Sektoren)  
Versatz des Sektors Nr. 1 auf der nächsten Spur. Dauer der Kopfschaltzeit wichtig (embedded servo!).
- „**Zylinder-Versatz** (zyl. skew)“: ..... (Anzahl Sektoren)  
Versatz des Sektors Nr. 1 auf dem nächsten Zylinder. Dauer der Positionierzeit wichtig bei Zylinderwechsel (letzte Spur Zyl. n → erste Spur Zyl. n+1).
- „**Freie Sektoren** (spare sectors)“: 0 .. 3 Freie Sektoren pro Spur für die Einrichtung von Ersatzsektoren.
- „**Reservepuren** (alternative tracks)“: Ersatzspuren
- „**Blockgröße** (block size)“: 256 .. 4096 Byte Kapazität des Datenfelds eines Sektors.

*Diskutieren Sie mögliche Auswirkungen von Spur-Versatz und Zylinder-Versatz auf Ausführungszeiten von Vorgängen wie z.B. "Datei lesen". Welcher Dateityp ist betroffen?*

**Format device page (aus SCSI-2 Norm, Kap. 9.3.3.3) :**

Eine „device page“ entspricht einer **Zone** der Plattenoberfläche, Aufteilung des Datenträgers in Zonen siehe Kap. 1.1 c)

2	(MSB)								
Tracks per zone									
3									(LSB)
4	(MSB)								
Alternate sectors per zone									
5									(LSB)
6	(MSB)								
Alternate tracks per zone									
7									(LSB)
8	(MSB)								
Alternate tracks per logical unit									
9									(LSB)
LUN=0 bis 7									
10	(MSB)								
Sectors per track									
11									(LSB)
12	(MSB)								
Data bytes per physical sector									
13									(LSB)
14	(MSB)								
Interleave									
15									(LSB)
16	(MSB)								
Track skew factor									
17									(LSB)
18	(MSB)								
Cylinder skew factor									
19									(LSB)
20	SSEC	HSEC	RMB	SURF					Reserved
21									
Reserved									
23									

alternate sector = Ersatz-Sektor, skew factor = Versatz

Üblicherweise werden Ersatzsektoren auf der Spur zugewiesen, auf der der defekte Sektor liegt. Eine Ersatzspur-Zuweisung erfolgt, wenn keine Ersatzsektoren mehr verfügbar sind.

**c) Einrichten von Ersatzsektoren für defekte Sektoren**

Eine Ersatzsektor-Zuweisung liefert dem Betriebssystem eine konstante Spurkapazität und eine konstante Plattenkapazität (notwendig ?).

Vorgehensweise bei SCSI-Platte (Datenfeld defekt und Ersatzsektor auf derselben Spur):

**Defekter Sektor:** Adressfeld/ Status-Byte: Bit „defekter Sektor“, Logische Sektor-Nr. = FF  
**Ersatzsektor:** Adressfeld/ Status-Byte: Logische Sektor-Nr.= Nr. des defekten Sektors

Der Ersatzsektor liegt am Ende der Spur. Bei der Ersatzsektor-Zuweisung ist auch ein „Rangieren“ der Sektoren auf der Spur möglich, sodaß die Sektor-Nummern in aufsteigender Reihenfolge sind. (Drehwartezeit bei sequentiellem Zugriff!)

Falls der Ersatzsektor auf einer anderen Spur liegt, muß zusätzlich im Status-Byte "Ersatzspur zugewiesen" stehen und die Adresse der Ersatzspur im Adreßfeld des defekten Sektors enthalten sein.

Mit dem Befehl „Reassign Blocks“ (SCSI-2) veranlasst der Initiator das Target (SCSI-Gerät), den defekten Sektoren, aufgelistet in einer „Defect List“, Ersatzsektoren zuzuweisen.

Von SCSI-2 werden verschiedene „Defect Lists“ unterstützt:

-Primary defect list, Plist: vom Hersteller abgespeichert  
-Target certification list, Clist: während des Kontroll-Lesens bei Format Unit erkannt.  
-Data defect list, Dlist: vom Initiator an das Target (Gerät) gesendet  
-Grown defect list, Glist: Liste aller defekten Sektoren, erkannt vom Target oder gemeldet vom Initiator (nicht Plist).

Aufgabe: Plattenspeicher mit Ersatzsektoren

**The Reassigns Counter** (Quelle: Intel RAID-Controller)

The Reassigns counter records the number of media defects which occur on a hard drive. Defective blocks of a hard drive are assigned substitute blocks (spare blocks) that are either on the same track, or on alternate ones if all spare blocks on the same track are already in use. The administration of the reassignments is carried out by the hard drive through reassignment tables. If you observe that the number of reassigns is constantly increasing, you may assume that something is wrong with the drive.

If a hard disk works with alternate tracks, it is generally no longer suitable for applications with high performance expectations. Whenever a defective block is being accessed, the read/write actuator has to move to an alternate position and this requires extra time.

### 1.2.5 Selbst-Diagnose

Durch Plattenfehler tritt Datenverlust mit den bekannt hohen Kosten auf. Die Hersteller sind daher bemüht, ein „Frühwarnsystem“ in die Platten einzubauen. Ein System heißt Self-Monitoring Analysis and Reporting Technology (S.M.A.R.T). Es wird bei IDE- und SCSI-Platten eingesetzt.

Prinzip: bestimmte Plattenausfälle kommen nicht plötzlich sondern sie kündigen sich an. Der Disk-Controller protokolliert daher während des Betriebs verschiedene Plattenparameter (u. a. auch Temperatur-Sensoren), um den Plattenzustand zu überwachen:

S.M.A.R.T – Commands siehe z.B. [www.seagate.com](http://www.seagate.com) .....Festplatten-Manuals.

Common attributes that degrade over time are identified as:

- power-on hours (minutes)
- raw read error rate
- seek error rate
- re-allocated sector count
- spin-up time
- seek time
- spin retry count
- calibration retry count

Es ist immer eine Software oder ein SMART-fähiges BIOS notwendig, um den Zustand abzufragen. (Protokollierung während des Betriebs oder Test anstossen per Befehl)

### 1.2.6 Datenrate

#### a) Definition Datenrate

- **physikalische Datenrate  $R_{phys}$**  (MBit/s)

Die physikalische Datenrate ist die momentane Datenrate beim Lesen (oder Schreiben) am Schreib/Lesekopf. Sie ist auf verschiedene Arten berechenbar, z.B.

$$R_{phys} = \text{Bitdichte (bpi)} * \text{Schreib/Lesegeschwindigkeit (ips)} \quad (\text{bpi}=\text{bit}/\text{inch}, \text{ips}=\text{inch}/\text{Sekunde})$$

$$R_{phys} = \frac{\text{Spurkapazität unformatiert}}{\text{Umdrehungszeit}}$$

- **mittlere Datenrate  $\bar{R}$**  =  $\frac{\text{übertragene Datenmenge}}{\text{Zeitintervall}}$  (MByte/s)

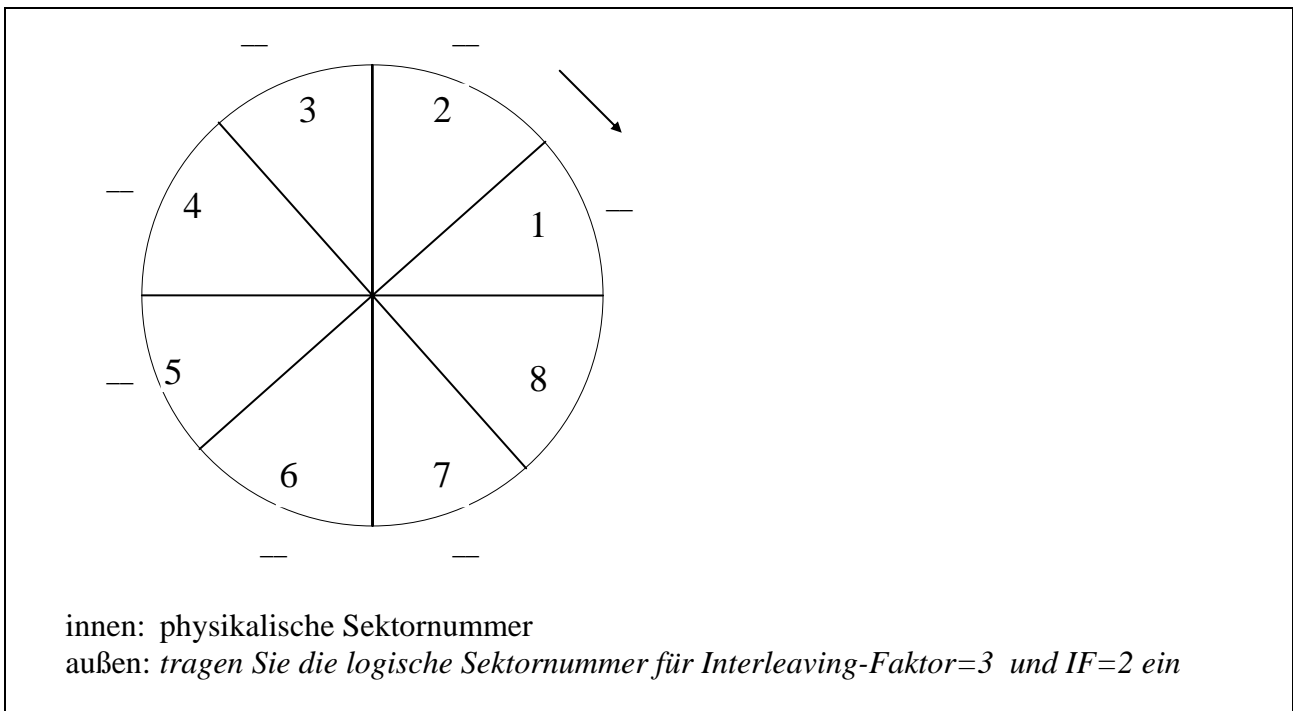
Frage: mittlere Datenrate bei  $IF=3$  (geg.: formatierte Spurkapazität  $K_s$  und Umdrehungszeit  $t_u$ )?

$$\bar{R} \text{ (MByte/s)} =$$

#### b) Sektor-Interleaving

Die Nummerierung der Sektoren auf der Spur - abweichend von ihrer physikalischen Reihenfolge - wird mit Sektor-Interleaving bezeichnet ( $\rightarrow$  Interleaving-Faktor).

SCSI-2: Formatierparameter, übergebbar mit SCSI-Befehl "Mode Select".



#### Bild: Sektor-Interleaving

Interleaving-Faktor = 1: logische Reihenfolge = physikalische Reihenfolge

Frage: wie wirkt sich Sektor-Interleaving aus?

**c) Ablauf "Schreiben/Lesen Datenfeld"**

• Positionieren auf Zylinder	
• Kopf (Datenoberfläche) auswählen	
• Adreßfelder aller "vorbeidrehender" Sektoren lesen bis gesuchter Sektor am Schreib/Lesekopf ist	
• Lesen oder Schreiben (=Umschalten Kopf) Datenfeld und ECC/CRC-Bytes des Datenfelds	
• Bei Lesefehler: Korrektur über ECC oder Lesewiederholung	

### 1.3 Verwaltung des Datenträgers (Dateien)

Hier erfolgt nur eine kurze Erläuterung der Prinzipien. Für eine ausführliche Diskussion wird auf die Betriebssystem-Vorlesungen verwiesen.

#### 1.3.1 Belegungstabelle der Sektoren

Die Sektoren einer Festplatte werden mit Hilfe einer Belegungstabelle (z.B. FAT, File Allocation Table) verwaltet, in der die Verfügbarkeit der Sektoren (frei, belegt, defekt usw.) abgespeichert ist. Manche Betriebssysteme verwalten die Platte in größeren Einheiten als Sektoren, z.B. teilt Windows die Platte in sogenannte Zuordnungseinheiten (Cluster) ein, wobei 1 Cluster mehrere Sektoren umfasst (z.B. Festplatten > 32 GByte: 1 Cluster = 64 Sektoren = 32 kByte). Beim Abspeichern z.B. einer 100Byte-Datei und einer 28 kByte-Datei wird jeweils 1 Cluster belegt, d.h., 32 kByte Speicherplatz. Der nicht benützte Speicherplatz innerhalb eines Clusters ist Verschnitt.

Die Belegungstabelle (FAT file allocation table) enthält für jeden Cluster einer Platte einen Vermerk:

"frei"/ "belegt"/ "defekt"

"belegt" = Cluster-Nr des nächsten Clusters (der Cluster-Kette einer Datei)  
oder Bitmuster "letzter Cluster (der Kette)"

Über das Belegungsverzeichnis sind die weiteren Cluster der Datei verkettet. Die Cluster einer Datei können beliebig über die Platte verteilt sein ("gestreute" Abspeicherung).

Innerhalb eines Clusters sind die Sektoren aufeinanderfolgend (sequentiell) angeordnet.

Ist der Speicherplatz einer Platte in mehrere Partitionen (logische Laufwerke C:, D:, usw.) unterteilt, dann enthält jede Partition eine eigene Sektor-Belegungstabelle und ein Datei-Verzeichnis.

Somit sind Doppelbelegungen eines Clusters ausgeschlossen.

Ausschnitt aus Inhaltsverzeichnis:

<b>Dateiname</b>	.....	.....	.....	<b>Erster Cluster = 0100</b>	.....
------------------	-------	-------	-------	------------------------------	-------

Ausschnitt aus einer Belegungstabelle: übertrieben stark fragmentiert!!!

Cluster-Nr.		Kommentar
0000 0000	<b>0000 0000</b>	
← 0100	<b>0103</b>	Datei-Beginn
0101	<b>0000</b>	frei
← 0102	<b>0000</b>	frei
0103	<b>0104</b>	Datei (Folge-Cluster)
← 0104	<b>0108</b>	Datei (Folge-Cluster)
0105	<b>F7FF</b>	defekter Cluster
0106	<b>0000</b>	frei
← 0107	<b>0000</b>	frei
0108	<b>FFFF</b>	Dateiende (letzter Cluster)
0109	<b>0000</b>	frei

*Erstellen Sie die Belegungstabelle (verkürzt) für eine Datei mit den Clustern 0100, 0103, 0104, 0108.*

*Vorgaben:*

*"Cluster frei"=0000 / "defekt"=FFF7 /*

*"belegt"=Folge-Cluster /*

*"letzter Cluster einer Datei"=FFFF*

*Frage:*

*Wieviele Sektoren umfaßt obige Datei bei Abspeicherung auf Platte unter den Annahmen 64 Sektoren/Cluster, 512 Byte Datenfeld pro Sektor?*

Im Datei-Verzeichnis (Directory) ist der Dateiname und die Adresse (Nummer) des jeweiligen ersten Sektors (Clusters) vermerkt.

Der Zugriff auf obige Informationen erfolgt über den Master Boot Record MBR (Boot Sektor) der Platte (siehe Betriebssystem-Vorlesungen).

**Partitionieren:**

- Aufteilen des Speicherplatzes in Partitionen (C:, D:, usw.)
- Beschreiben des Boot-Sektors mit den Sektor-Adressen der Partitionen (boot partition).  
Damit ist ein Starten von Platte möglich.

**High-Level-Formatieren der logischen Laufwerke:**

- Erzeugen einer Sektor-Belegungstabelle für jede Partition. kein Löschen der Datenfelder!
- Erzeugen eines Datei-Verzeichnisses (Directory).

**1.3.2 Master Boot Record MBR (Boot-Sektor)**

Der Master Boot Record MBR ist der erste Sektor auf der Festplatte (Zyl. 0 / Kopf 0 / Sektor 1). Er enthält die Informationen, die das Betriebssystem über den Datenträger Festplatte benötigt (z.B. BIOS Parameter Block) und ein Ladeprogramm (Boot Loader) für das Laden des Betriebssystems in den Arbeitsspeicher.

Der File System Boot Sector ist der erste Sektor einer Partition.

Die Informationen eines FAT32 Master Boot Records erstrecken sich über 3 Sektoren.

- Sektor 1: BIOS Parameter Block (siehe unten, ab Offset 0B),
- Sektor 2: Informationen über das Datei-System (File System)
- Sektor 3: zusätzlicher Code für den Boot Loader.

(BIOS Parameter Block in Tabelle ab Offset 0B. Ein gültiger Boot-Sektor hat beim Offset 1FE die Kennung AAFH)

Offset in Boot Sector	Length in Bytes	Mnemonic
00	3	Machine Code for jmp over the data
03	8	<a href="#">OEM_Identifier</a> <sup>1)</sup>
0B	2	<a href="#">BytesPerSector</a>
0D	1	<a href="#">SectorsPerCluster</a>
0E	2	<a href="#">ReservedSectors</a> = Anzahl Sektoren vor der FAT → FAT-Adresse
10	1	<a href="#">NumberOfFATs</a> <sup>1)</sup>
11	2	<a href="#">RootEntries</a> <sup>1)</sup>
13	2	<a href="#">NumberOfSectors</a> <sup>1)</sup>
15	1	<a href="#">MediaDescriptor</a> <sup>1)</sup>
16	2	<a href="#">SectorsPerFAT</a> <sup>1)</sup>
18	2	<a href="#">SectorsPerHead</a>
1A	2	<a href="#">HeadsPerCylinder</a>
1C	4	<a href="#">HiddenSectors</a> <sup>1)</sup>
20	4	<a href="#">BigNumberOfSectors</a>
24	4	<a href="#">BigSectorsPerFAT</a> bei FAT32
28	2	<a href="#">ExtFlags</a> <sup>1)</sup>
2A	2	<a href="#">FSVersion</a> <sup>1)</sup>
2C	4	<a href="#">RootDirectoryStart</a> = Stamm-Verzeichnis
30	2	<a href="#">FSInfoSector</a> <sup>1)</sup>
32	2	<a href="#">BackupBootSector</a> <sup>1)</sup>
34	12d	Reserved

### Aufbau MBR Sektor 1 (Boot Sektor) bei FAT32

Ein weiterer Eintrag ist z.B. das [Volume Label](#) (11 Byte Feld), das das Laufwerk identifiziert.

<sup>1)</sup> Erklärungen (unvollständig):

[OEM\\_Identifier](#) is the eight-byte ASCII string that identifies the system that formatted the disk..

[Reserved sectors](#) is the number of sectors before the first FAT including the boot sector.

[Number of FATs](#) is the number of File Allocation Tables. This value is usually two. FATs are consecutive on the disk: the second copy of FAT goes right after the first copy. At least one copy of FAT should be present.

[RootEntries](#) contains the number of the entries in the root directory if root directory is fixed. It is zero if the root directory is not fixed. FAT32 disks should contain zero in this field, indicating that the root directory can be arbitrarily long. Otherwise, this field usually contains 512. Each directory entry takes up 32 bytes. To avoid wasting space, [RootEntries](#)\*32 should be divisible by [BytesPerSector](#).

[NumberOfSectors](#) is the total number of sectors on the disk. If the number of sectors is greater than 65535, then this field is set to zero and the dword at [BigNumberOfSectors](#) contains the actual number of sectors..

[MediaDescriptor](#) describes the device: Floppies

[SectorsPerFAT](#) contains the number of sectors in one FAT. This field is zero for FAT32 drives, and [BigSectorsPerFAT](#) contains the actual value. [HiddenSectors](#) is the number of sectors between the beginning of this partition and the partition table.

[ExtFlags](#), and all fields described below, are defined only for FAT32 disks.

[FSVersion](#) is the version of the file system. The high byte is the major version, the low byte is the minor version.

[RootDirectoryStart](#) contains the number of the first cluster for the root directory.

[FSInfoSector](#) is the sector number for the file system information sector. This sector is new to FAT32.

[BackupBootSector](#) is the sector number for the backup copy of the boot sector. This copy can be used if the main copy was corrupted. It is also nice to compare the two copies on startup. If they do not match, a warning should be issued. They may not be in tact because of corruption or a boot virus.

### ***1.3.3 Sektoranordnung innerhalb einer Datei***

betrifft: Abspeicherung, Zugriffsart

#### **a) Sequentielle Datei**

Die Datei ist in aufeinanderfolgenden Clustern/Sektoren (Datenblöcken bei Magnetband) abgespeichert.

**SAM**-Datei (sequential access method).

Beispiele für sequentielle Anordnung:

#### **b) Gestreute Datei**

Die Clustern/Sektoren einer Datei sind beliebig über den Datenträger verstreut.

**DAM**-Datei (direct access method).

Voraussetzung ist, daß jeder Sektor direkt adressierbar ist, z.B. die Sektoren einer Festplatte. Dateien auf der Festplatte können wegen der Fragmentierung des Speicherbereichs gestreute Dateien sein.

#### **c) Sequentielle Datei mit Index**

**ISAM**-Datei

- Die Datei ist sequentiell abgespeichert, nach Schlüssel sortiert.
- Zusätzlich wird die Datei mit Hilfe einer Index-Liste unterteilt. Dadurch sind Teile der Datei direkt adressierbar (Voraussetzung ist, daß die sequentielle Datei auf einem direkt adressierbaren Speicher wie der Platte abgespeichert ist). Dies ermöglicht einen schnelleren Zugriff zu einzelnen Sätzen. (siehe auch Betriebssystem-Vorlesung)

Die **Index-Liste** orientiert sich z.B. am physikalischen Aufbau der Platte:

Ebene 1: Zylinder-Index (Tabelle)

Ebene 2: Spur-Index pro Zylinder (Tabelle)

### ***1.3.4 Fragmentierung und Defragmentierung***

Bei einer neu formatierten Platte bietet die Platte dem Anwender einen einzigen grossen freien Speicherbereich an. Mit zunehmender Benutzung (speichern, löschen, erweitern von Dateien) wechseln sich belegte und freie Bereiche im Speicherbereich ab.

Eine Datei ist fragmentiert, wenn sie "nicht zusammenhängend" (= nicht an einem Stück) abgespeichert ist, sondern in zwei oder mehrere Teile zerfällt.

Ein Datenträger ist fragmentiert, wenn der freie Speicherplatz durch belegte Bereiche in zwei oder mehrere Teile aufgeteilt ist.

Die Ursache der Datei-Fragmentierung ist die Fragmentierung des Datenträgers. Wird z.B. eine 4MByte große Datei mitten aus einem belegten Bereich gelöscht, so entsteht eine 4MByte große Lücke und der Datenträger ist fragmentiert. Wird jetzt z.B. eine 12MByte große Datei abgespeichert, so wird zuerst die 4MByte-Lücke gefüllt. Der Rest der Datei wird an das Ende des belegten Datenbereichs angehängt, sodaß die Datei in zwei Teile aufgespalten ist und somit fragmentiert ist.

Fragmentieren entsteht also durch Löschen, Verkleinern und Vergrößern von Dateien. Große Dateien können so in viele Teile aufgespalten werden, was u.U. die Transferzeit der Dateien um ein Vielfaches verlängert.

Es empfiehlt sich daher, mit Defragmentierungs-Programmen die Zerstückelung von Dateien rückgängig zu machen. Fragmentierte Dateien vermindern die Systemgeschwindigkeit erheblich.

#### **Beispiel:**

Literaturwert: Bei durchschnittlich 6 Fragmenten pro Datei erhöhte sich die Transferzeit beim Lesen einer Datei auf 280%. (100% = Datei defragmentiert).

Aufgabe: *Fragmentierte Datei*

### ***1.3.5 Abbilden von Sätzen einer Datei auf die Sektoren (Blöcke)***

Eine Datei lässt sich als Summe von Sätzen darstellen (Informatik-Vorlesung).

1 Datei =  $\sum$  Sätze

1 Satz =  $\sum$  Felder (Records)

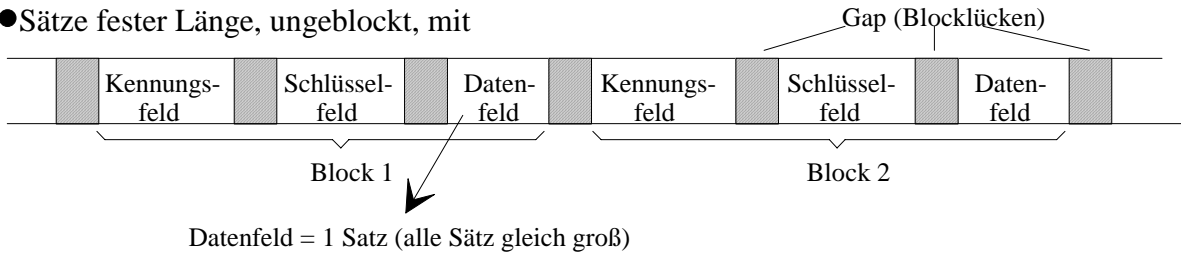
1 Feld =  $\sum$  Bytes

Ein Satz kann kleiner, gleich oder größer als das Datenfeld eines Sektors sein. Sind mehrere Sätze in einem Datenfeld abgespeichert, spricht man von geblockten Sätzen. Der **Blockungsfaktor** gibt an, wieviele Sätze in einem Datenfeld enthalten sind. Ist ein Satz auf mehrere Datenfelder verteilt, spricht man von **Spann-Satz**. Es sind Sätze „fester Länge“ (alle Sätze gleich groß, z.B. Kundendatei) und Sätze „variabler Länge“ (z.B. Eingabedatei vom Terminal, jede Zeile ist 1 Satz) definiert.

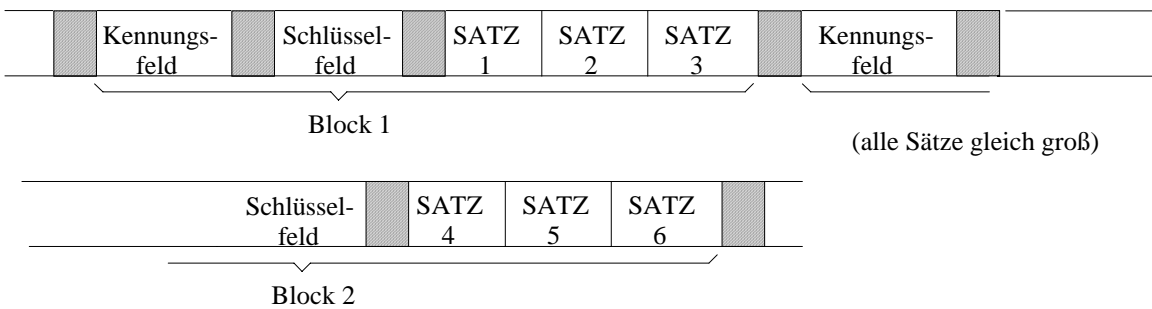
**Bild: Magnetplattenspeicher - Satz- und Blockformate**

siehe auch Betriebssystem-Vorlesung

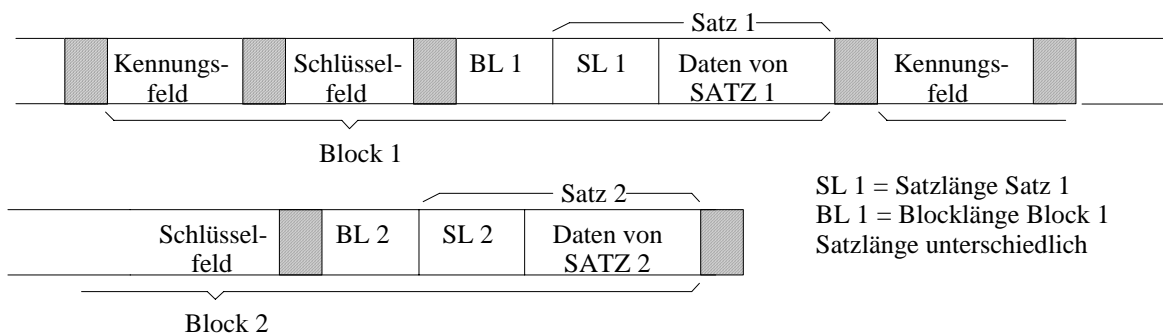
- Sätze fester Länge, ungeblockt, mit



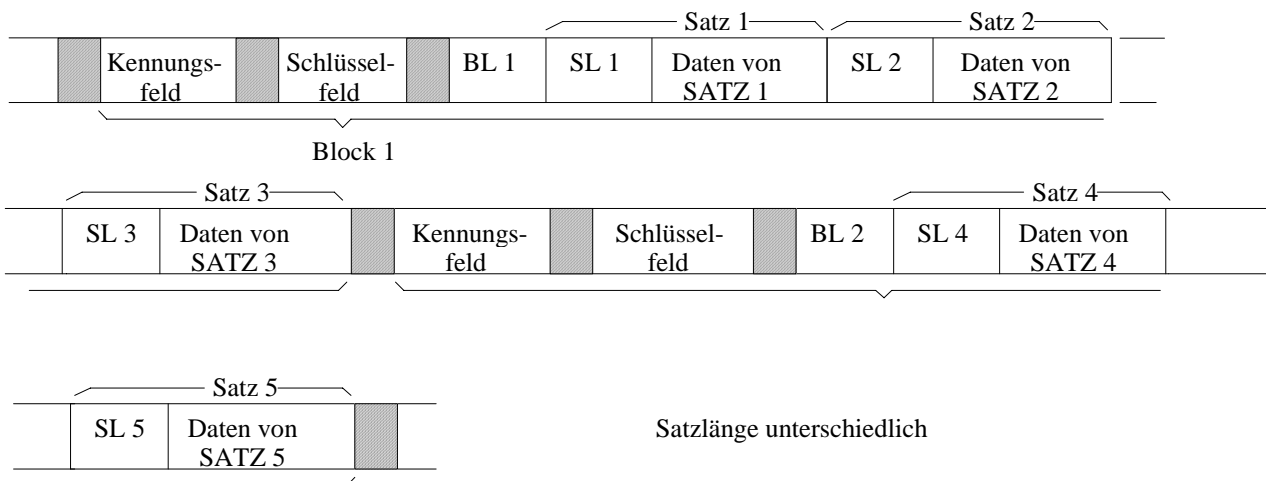
- Sätze fester Länge, geblockt, mit



- Sätze variabler Länge, ungeblockt, mit



- Sätze variabler Länge, geblockt, mit

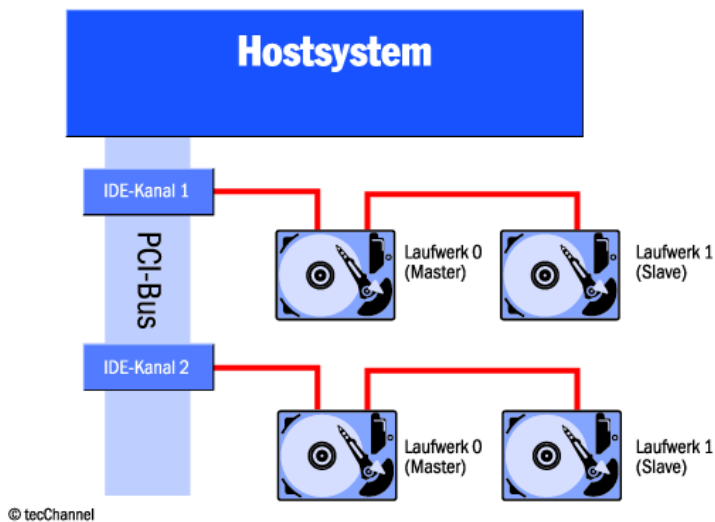


## 1.4 Schnittstelle

### 1.4.1 IDE-Schnittstelle (Integrated Drive Electronic)

<http://www.tecchannel.de/hardware/574/6.html>, <http://www.computer-tutorial.de/workshop/harddisk1.html>

Auch ATA-Schnittstelle, AT-Bus-Schnittstelle oder Conner-Interface genannt, beschreibt Geräte mit integrierter Controller-Elektronik gemäß der ATA-Spezifikation. Dient zum Anschluß von Geräten (Platten) intern im PC. Heutige Chipsätze bieten meist zwei ATA-Kanäle (primärer und sekundärer Port) für die Verwaltung von vier Geräten. Der zweite Port erhält einen eigenen Interrupt zur Adressierung. Das Bild zeigt ein Hostsystem (PC) mit zwei Kanälen und vier Geräten.



Obige Laufwerke könnten sein: Festplatte, ZIP250, CD-RW, DVD-ROM usw.

IDE-Drives sind erste Wahl, wenn niedrige Anschaffungskosten im Vordergrund stehen.

Die IDE-Schnittstelle ist eine Untermenge der Signale des ISA-Busses im PC mit 16 Datenleitungen, 3 Adreßleitungen, Steuerleitungen (/DIOW, /DIOR, /DDMARDY, INTRQ, usw.) sowie einigen Auswahlsignalen (dekodierte Adressen) /CBLID, /CSEL, /CS0, /CS1, ... (40-poliges Kabel, ab 66MHz 80-polig wegen Masseleitungen zur Abschirmung).

Um den gestiegenen Leistungsanforderungen (Datenrate, Kapazitätsgrenze) an die Platte nachzukommen wurde die IDE-Schnittstelle weiterentwickelt zu **Enhanced IDE (EIDE)** oder **Fast ATA** mit Datenraten von 33MByte/s, 66MByte/s, 100MByte/s (**UltraDMA/100, ATA-6-Standard**), 150MByte/s .... und **Serial-ATA** (SATA, 150MByte/s) usw. mit einer Anschlußmöglichkeit an einen schnellen lokalen Bus (PCI-Bus).

Für ATA-Festplatten sind nach dem ATA-5-Standard über 50 verschiedene Befehle spezifiziert. Die Kommandos reichen vom einfachen Schreiben/Lesen über Formatieren und Löschen bis hin zu Powermanagement-Befehlen. Bei den Kommandos gibt es eine Unterscheidung nach Pflichtbefehlen, die jedes Gerät beherrschen muss, und optionalen Befehlen.

{BIOS}

Beim Betrieb von zwei Geräten pro Kanal wird zwischen einem Master-Laufwerk und Slave-Laufwerk unterschieden. Die Bezeichnung Master/Slave ist dabei irreführend, weil sich die beiden Laufwerke weder gegenseitig beeinflussen noch eines von dem anderen kontrolliert wird. In den ATA-Spezifikationen finden sich deshalb die Bezeichnungen Device 0 (Master) und Device 1 (Slave). An den Geräten muss die Einstellung als Device 0 und 1 getroffen sein. Hierfür befinden sich an ATA-Laufwerken immer zwei Steckbrücken für die Konfiguration als Master oder Slave. Die Platte von der der Rechner booten soll (Betriebssystemplatte), ist das Master-Laufwerk am primären Controller.

Die Erweiterung von Enhanced IDE für den Anschluß anderer Peripheriegeräte wie CD-ROM, Bandkassetten, usw. wird **ATAPI** (AT Attachment Packed Interface) genannt. Die Ansteuerung der ATAPI-Laufwerke basiert auf SCSI-Kommandos. Diese werden in ATA-Kommandos eingepackt, daher der Name Packet Interface.



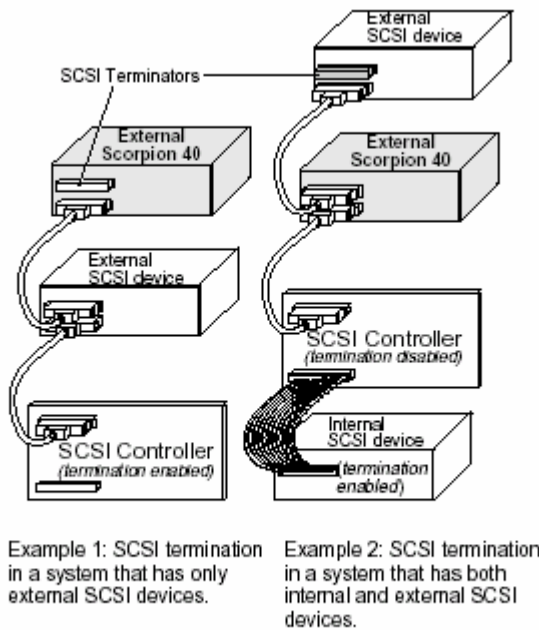
### 1.4.2 Kabel

Es werden allgemein bei allen Geräten (Platte, Diskette, Bandkassette, optische Speicher usw..) zwei Kabelführungen unterschieden:

#### a) geschleifter Anschluß (daisy chain)

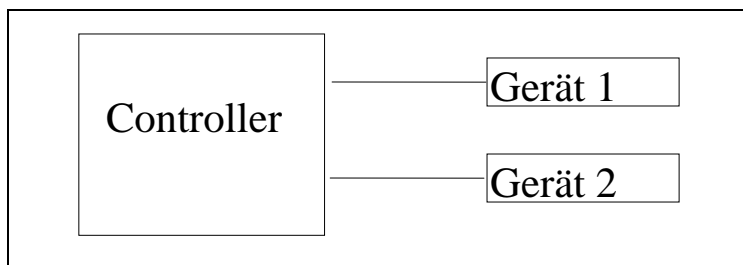
Um den Aufwand klein zu halten wird meistens das Kabel über alle angeschlossenen Geräte eines Controllers geschleift. Dies mindert die Leistungsmerkmale, da überlappende Vorgänge auf dem Kabel nicht möglich sind und überlappende Vorgänge in den Geräten selbst nur begrenzt stattfinden können (z.B. überlappendes Positionieren).

#### Beispiel SCSI-Schnittstelle:



#### b) sternförmiger Anschluß (radial)

Jedes Gerät ist separat angeschlossen, d.h., unabhängig in Hardware und Software vom anderen Gerät. Ist sinnvoll bei Kabeln mit wenig Signalen.



Fragen:

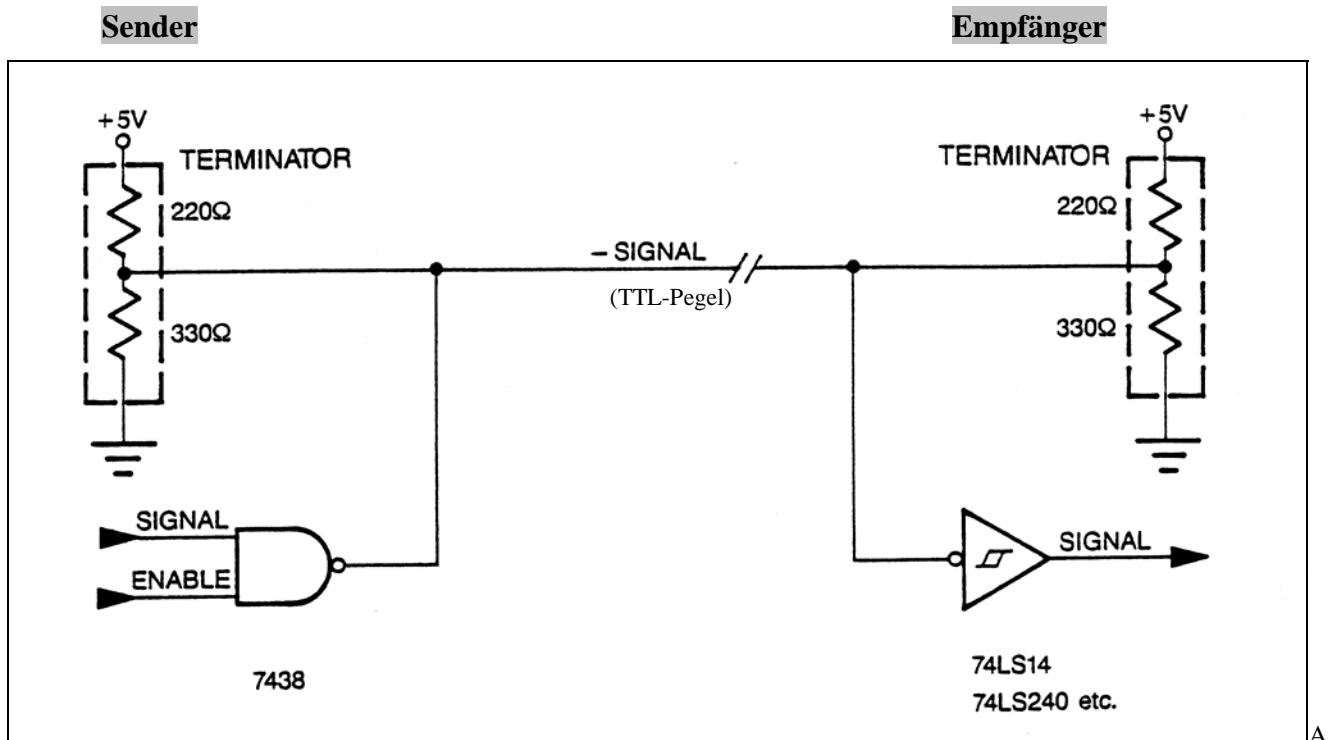
Nennen Sie Vor- und Nachteile von sternförmig/ geschleift  
Nennen Sie Schnittstellen mit geschleifter Kabelführung

**c) Elektrische Übertragung auf dem Kabel**

Es gibt bei allen Gerätetypen zwei Übertragungsarten: .....

**Asymmetrische (single ended):** Übertragung mit TTL-Pegel

Die Signale werden mit TTL-Pegel gegen Masse übertragen. Diese einfache Übertragungsart ist stör anfällig und daher nur für kurze Kabellängen geeignet (max. 6m bei SCSI-Schnittstelle).



typical driver/receiver circuit and termination for one signal of a **single-ended SCSI** bus.

Sender: Typ Open Collector (7438 wegen des hohen Ausgangsstroms)

Empfänger(großer Störabstand): Typ .....

Abschluß Signalader: mit Widerstandsteiler aus 220 und 330 Ω

- **Abschlußwiderstand** = ...  $220 // 330 \Omega = 132 \Omega$  , // = Parallelschaltung  
dies entspricht dem Wellenwiderstand Z der Leitung und etwa dem Ausgangswiderstand des Senders. Damit werden Reflexionen des Signals am Kabelende vermindert.  
(Z einer Leitung:  $Z^2 = L / C$  mit L = Induktivität pro cm, C = Kapazität pro cm)

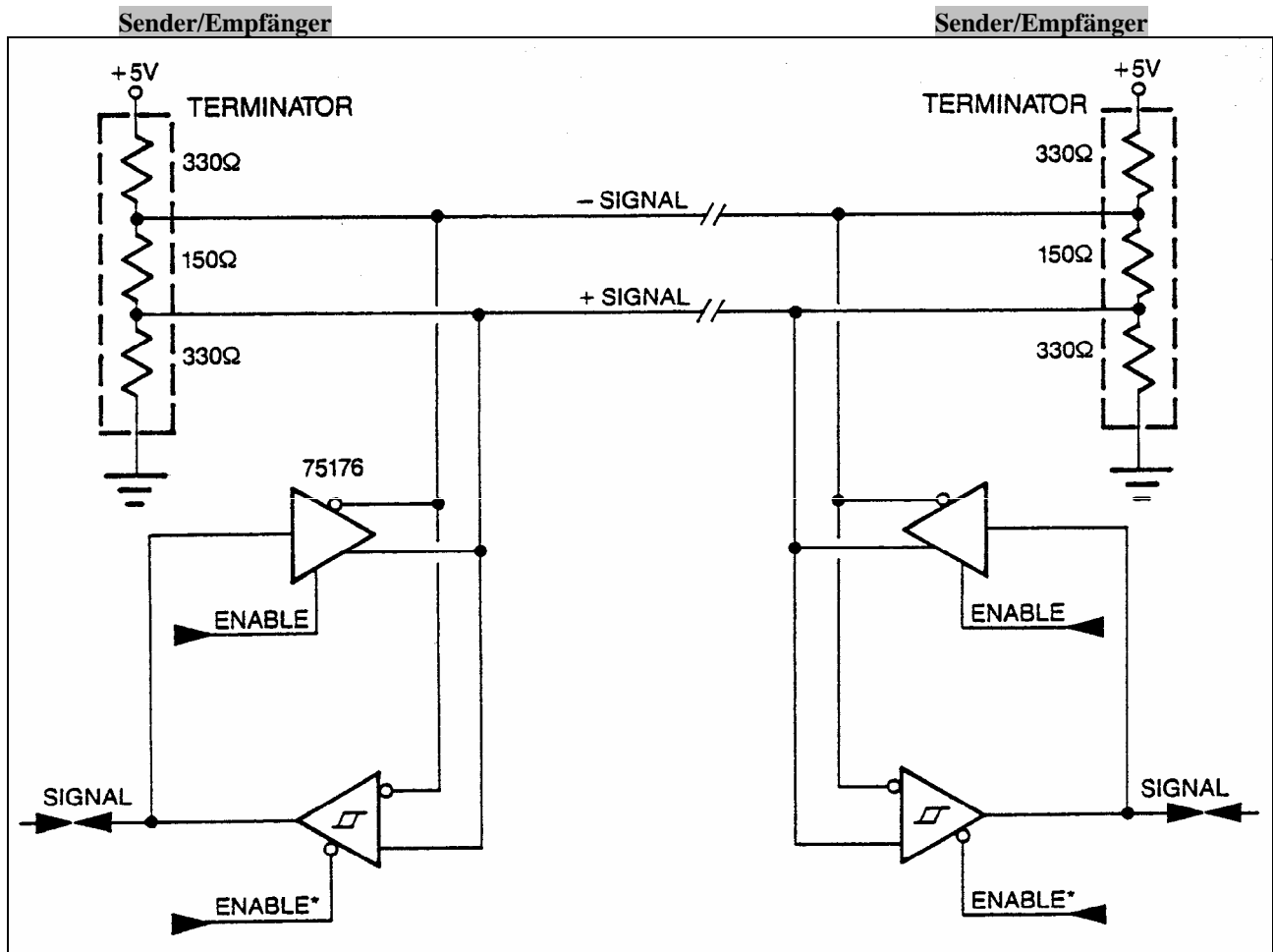
- **Gleichspannungspegel** = .....  $5V * 330 / (220 + 330) = 3V$ , High-Pegel  
(bei hochohmigem Senderausgang = Signal nicht aktiv oder bei „Kabel gezogen“)  
Eine Low-aktive Logik vorausgesetzt entspricht dies dem Zustand „Signal = nicht aktiv“.

**Symmetrische (differential) Übertragung:**

Jedes TTL-Signal wird vom Sender in ein Plus- und ein Minus-Signal zerlegt und übertragen. Ein Differenzverstärker am Eingang wertet die Spannungsdifferenz (oder Stromdifferenz) aus und liefert wieder TTL-Pegel.

$$S_{\text{plus}} - S_{\text{minus}} \geq +200\text{mV} \rightarrow \text{Signal}=1 \text{ (True), Signal=aktiv}$$

$$S_{\text{plus}} - S_{\text{minus}} \leq -200\text{mV} \rightarrow \text{Signal}=0 \text{ (False)}$$



A typical driver/receiver circuit and termination for one signal of a **differential SCSI** bus.

Störsignale, die sich auf beide Signaladern einkoppeln, werden vom Differenzempfänger unterdrückt. Die Übertragung ist daher sehr störicher und wird für lange Kabellängen und hohe Datenraten verwendet (max. 25m bei SCSI-Schnittstelle).

Abschluß der beiden Signaladern: mit Widerstandsteiler aus 330, 150 und 330 Ω

$$\text{Abschlusswiderstand} = 150 // (330 + 330) = 122 \Omega \quad // = \text{Parallelschaltung}$$

$$\text{Gleichspannungspegel} = 5V * 150 / (330 + 150 + 330) = 0.9V$$

Bei nicht vorhandenem Sender liefert der Spannungsteiler am Eingang ....  
für  $S_{\text{plus}} - S_{\text{minus}}$  mit  $5V * (-150/810) = -0,9V$  ("False").

### **1.5 Datenblätter**

Die Leistungsfähigkeit der Platte läßt sich steigern durch den Einsatz schneller Pufferspeicher (Cache-Speicher). Die von der Platte gelesenen Blöcke (Cluster, Sektoren) werden im **Read Cache** zwischengespeichert. Bei erneutem Lesezugriff wird zuerst der Cache auf die gewünschten Daten hin durchsucht. Sind diese im Cache, entfällt der Zugriff zur Platte und die Daten sind ohne Plattenzugriff schnell verfügbar. Beim **Read Ahead Cache** erfolgt ein Vorauslesen von noch nicht angeforderten Blöcken (Es werden die nachfolgenden Blöcke eines Lesezugriffs von Platte gelesen). Beim **Write Back Cache** werden die zu schreibenden Blöcke im Cache zwischengespeichert und erst dann auf Platte geschrieben, wenn keine weiteren Aufträge für die Platte vorliegen. Das Schreiben auf Platte wird dadurch scheinbar schneller. Es besteht aber die Gefahr, daß bei Spannungsausfall die Daten verlorengehen.

"Segmentierter" Cache für die Platte:

z.B. Cache mit 2 Lese-Segmenten, einem Schreib-Segment und einem Firmware-Bereich.

Cache-Speicher (DRAM oder SRAM) sind im Controller der Platte enthalten oder sind ein Teil des Arbeitsspeichers. Betriebssysteme wie z. B. UNIX reservieren einen Teil des Arbeitsspeichers als Cache für die Platte.

#### **Datenblätter:**

z.B.: [www.seagate.com](http://www.seagate.com) Demo „Festplattendaten“

## 1.6 RAID-Systeme

**Die Plattenspeicher-Peripherie eines Computers soll .....und .....usw. sein?**


### **Abschätzen des Ausfalls einer Platte:**

MTBF einer Platte  $\approx 500.000$ h (Betriebsstunden), 100 Geräte

Bei 100 Platten in einer mittelständischen Firma ergibt sich ein mittlerer zeitlicher Abstand zwischen zwei Plattenfehlern von  $MTBF = 500.000 / 100 = 5.000$  h, Ausfallrate  $\lambda \rightarrow 100 \lambda$

d.h., alle 5.000h (1 Jahr = 8760 h) tritt ein Plattenfehler - eventuell verbunden mit Datenverlust – auf.

(Die Garantiezeit für 1 Platte beträgt: 3..5 Jahre).

Eine Verbesserung der Leistungsmerkmale gegenüber einem Einzelgerät - insbesondere auch der Ausfallsicherheit - läßt sich nur durch einen Verbund mehrerer Plattengeräte (=Plattenspeicher-System) erzielen. Das Betriebssystem sieht das Plattensystem als nur ein logisches Laufwerk oder als mehrere logische Laufwerke, je nachdem ob die Platten des Systems in eine oder in mehrere Gruppen (RAID-Gruppen) konfiguriert sind.

Die derzeit auf dem Markt befindlichen Plattenspeichersysteme (Disk Arrays) basieren auf Prinzipien, die im Rahmen einer Studie (1987) an der amerikanischen Universität of Berkeley, California entwickelt wurden:

### **"RAID      Redundant Array of Independent (Inexpensive) Disks"**

In dieser Studie wurden verschiedene RAID-Konzepte RAID1 .. RAID5 entwickelt , bei denen der Ausfall einer Platte nicht zum Datenverlust führt und die auf verschiedene Leistungsmerkmale (Zugriffe/s, Datenrate) hin optimiert wurden. Der Anwender soll nur 1 Platte sehen (virtuelle Platte), obwohl das System aus mehreren Platten besteht.

weiterer Begriff:

JBOD = Just a Bunch Of Disks (Disk Array ohne RAID-Controller, d.h., der Kunde kann sich einen RAID-Controller von einem beliebigen anderen Hersteller dazu kaufen)

Im folgenden werden die technisch wichtigen Systeme (RAID 0), RAID 1, RAID0+1, RAID 3 (RAID4), RAID5 und RAID6 diskutiert. Es werden die erreichbaren **Datenraten** und die **Zahl der Zugriffe/s** berechnet. Dabei handelt es sich aber um theoretische Werte. Die im Betrieb erreichbaren tatsächlichen Werte dürften wegen des "Overheads" (Programmlaufzeiten im RAID-Controller und Hostcomputer usw. ) deutlich darunter liegen.

Bei einer einzelnen Festplatte sind die Werte (Datenrate, Zugriffe/s) für Lesen oder Schreiben einer Datei gleich groß. Bei Plattensystemen können sie sehr stark differieren. Schreibvorgänge dauern je nach RAID-Variante wesentlich länger. Wichtig ist daher der Mix zwischen Lese- und Schreibvorgängen.

**Bild: RAID-System im 19"-Gehäuse**



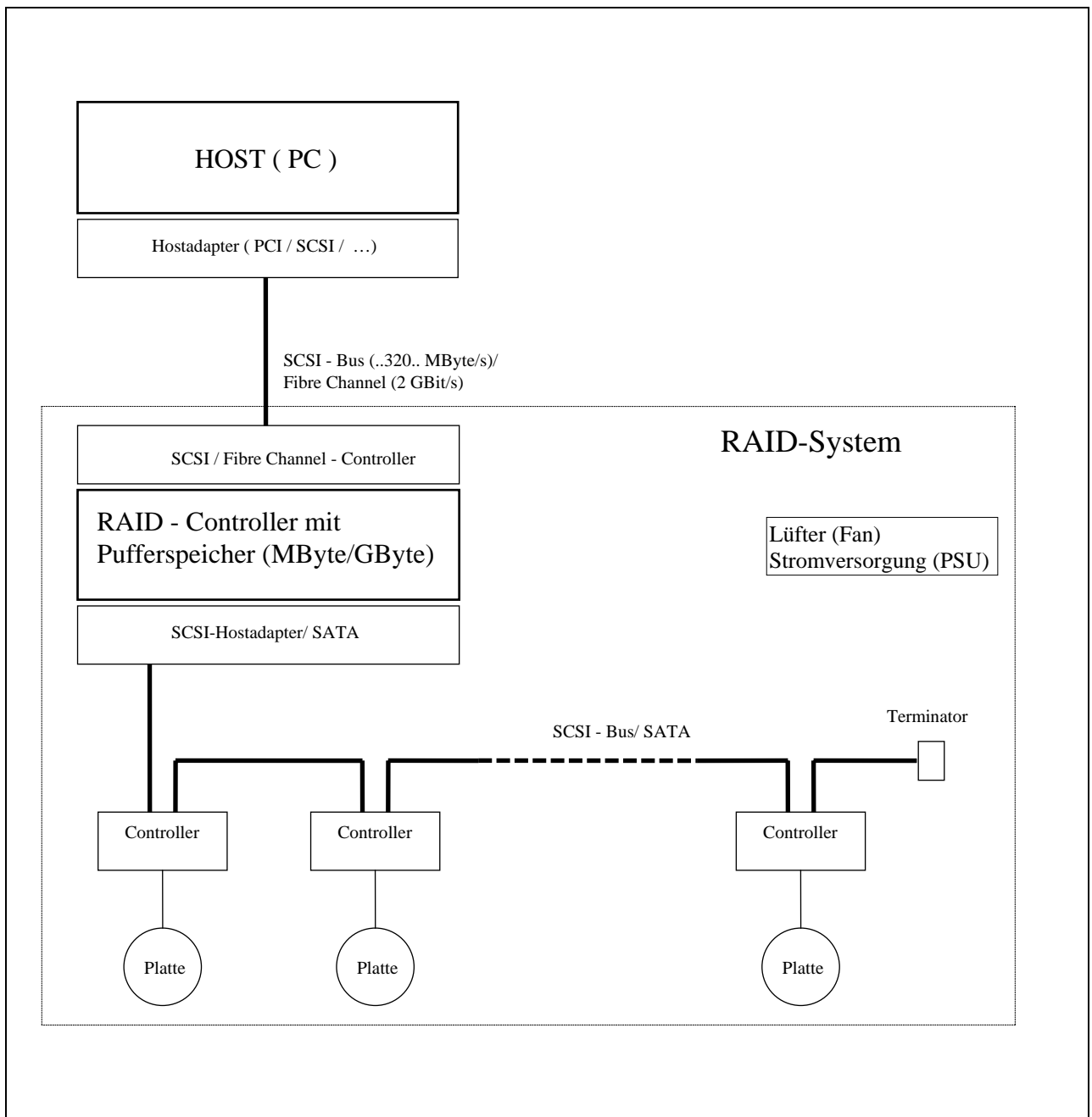
### ***1.6.1 Hardware- und Software-RAID***

Beim Software-RAID sind die einzelnen Festplatten über einfache Festplattencontroller am Computer angeschlossen. Es wird im Gegensatz zum Hardware-RAID kein spezieller RAID-Controller benötigt. Eine auf der CPU des Hosts laufende Software übernimmt die Steuerung des Plattenverbunds. Z.B. unterstützen Windows XP und Linux die gängigen RAID-Konzepte.

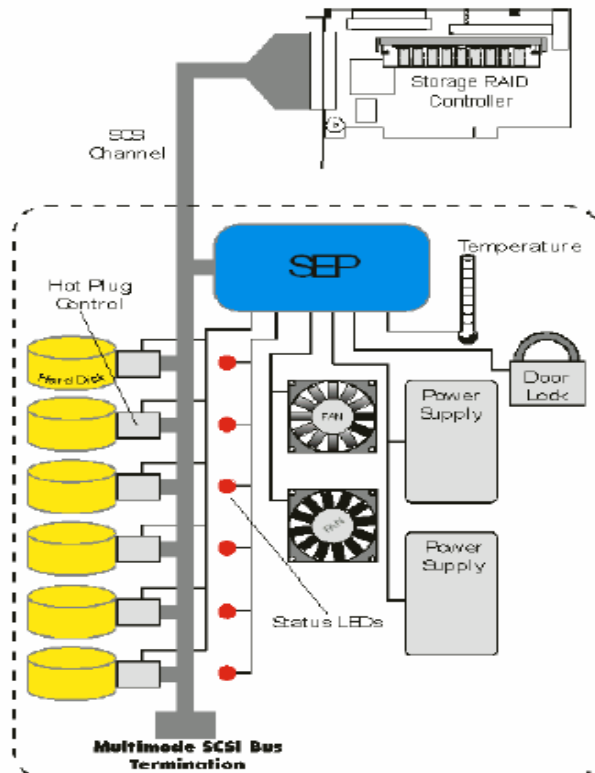
Software-RAID stellt meist die preisgünstigere und einfachere Lösung dar. Andererseits verursacht es eine hohe CPU-Belastung und arbeitet naturgemäß plattform- und betriebssystemgebunden. Auch stehen zur Ansteuerung der Laufwerke meist nur ein oder zwei Anschlüsse zur Verfügung. Dies beschränkt die mögliche Parallelisierung der Plattenzugriffe und damit auch die Leistungsfähigkeit.

Software-RAID unter Linux bietet z.B. die Möglichkeit, einzelne Partitionen einer Festplatte in einen RAID-Verbund einzubauen, um jedoch einen nennenswerten Geschwindigkeitsvorteil oder entsprechende Redundanz zu erzielen, sind generell die RAID-Partitionen auf unterschiedliche Festplatten zu verteilen.

Dagegen übernimmt bei Hardware-RAID ein eigener Controller die Ansteuerung des RAID-Arrays (Plattenverbund). Das bringt eine Entlastung der Host-CPU und eine höhere Leistungsfähigkeit mit sich. Zudem binden RAID-Controller die Laufwerke über mehrere Kanäle an, was gleichzeitige Laufwerkszugriffe und damit höhere Transferraten ermöglicht. Dafür gilt es aber, einen deutlich höheren Preis zu zahlen. Hardware-RAIDs arbeiten plattformunabhängig. Sie benötigen aber zur Verwaltung Software, die natürlich auf ein bestimmtes Betriebssystem zugeschnitten ist.



**Blockschaltbild eines RAID-Systems (Hardware-RAID)**



**RAID-System, angeschlossen mit PCI-Controller**

Intel-RAID, SAF-TE Subsystem (SAF-TE = SCSI Accessed Fault Tolerant Enclosures),  
SEP = SCSI Enclosure Processor (enclosure = Gehäuse)

Durch die Konfiguration des RAID-Controllers wird das RAID-Konzept (RAID0 .. RAID5) festgelegt. Die Information auf den Platten ist natürlich immer - egal welches RAID-Konzept vorliegt - in Sektoren abgespeichert! Die Sektorgröße (Größe des Datenfelds) ist beim Formatieren einstellbar.

*Frage:*

*Wie kann in einem Plattensystem eine Redundanz realisiert werden, die bei Ausfall einer Einzelplatte einen Datenverlust verhindert?*

### 1.6.2 Plattensystem ohne Redundanz (Disk Striping, RAID 0)

Dieses System aus mehreren Platten ohne Redundanz wurde nachträglich mit RAID 0 bezeichnet und entstammt nicht der RAID-Studie. Die Dateien werden "zerstückelt" (engl. Stripes = Streifen) und parallel (synchron) auf mehrere Platten verteilt. Der Striping-Faktor sagt aus, wie groß die Stücke sind, in die eine Datei zerteilt wird (z.B. 1 Streifen = 1 Byte | 1 Sektor | 1 Block). Üblicherweise umfasst ein Streifen 4kByte bis 128 kByte. Die Plattenspindeln laufen zueinander synchron um unnötige Drehwartezeiten zu vermeiden.

Vorteile/Nachteile?


#### Datenrate R der Anwenderdaten beim Transfer einer Datei:

$R = \frac{R_p}{z}$        $R_p$  = Datenrate einer Platte,  $z$  = Anzahl der Plattengeräte

#### Zugriffe/s:

$t_p = p$ ,

$t_l = t_w/2 = u/2$ ,

$t_{\text{transf}} = t$       Transferzeit einer Datei bei nur 1 Platte  
(Annahme: alle Sektoren innerhalb eines Zylinders)

Zeit pro Zugriff =

Lesezugriffe/s:       $L_0 =$

Schreibzugriffe/s:       $S_0 =$

#### Nutzbare Plattenkapazität:

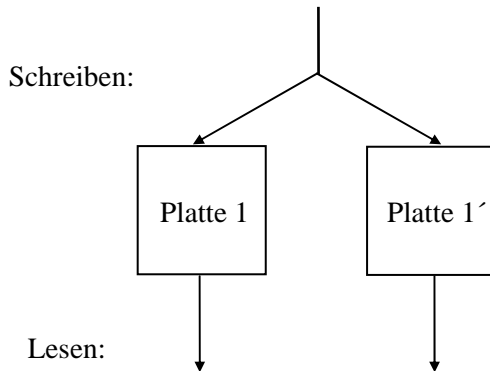
$K = \frac{K_p}{z}$        $K_p$  = Kapazität einer Platte

#### Frage:

Eine Datei besteht aus 6 Blöcken (Datei = ABCDEF). Wie sind die Blöcke bei RAID0 auf zwei Laufwerke verteilt?

### 1.6.3 Spiegelung von Platten (Disk Mirroring, RAID 1)

Bei der Spiegelung wird die Information auf "Platten-Paaren" abgespeichert. Die zweite Platte (Sekundärplatte) ist eine exakte Kopie der ersten Platte (Primärplatte). Im Fall eines Plattendefekts werden beim Lesen automatisch die Daten der anderen Platte verwendet.



Platten 1: Primärplatte  
Platten 1': gespiegelte Platte (1=1')

**System:** 2 Platten

**Schreiben:** gleiche Daten gleichzeitig auf Primärplatte und gespiegelte Platte.

**Lesen:** Primärplatte und gespiegelter Platte arbeiten unabhängig,

*Vorteil: ..*

#### **Fehlerbehandlung:**

- Fehler-Erkennung:

- defekter Sektor:
- ganze Platte defekt:

#### Frage:

Eine Datei besteht aus 6 Blöcken (Datei = ABCDEF). Wie sind die Blöcke bei RAID1 auf die beiden Laufwerke verteilt?

**Zugriffe/s:**

Zeit pro Zugriff =

Zahl der Zugriffe des RAID-Controllers auf eine Platte:

$$N_1 =$$

Spezialfall: nur Lesezugriffe/s:  $L_1 =$

Die Formel gilt unter der Annahme:

Lastteilung beim Lesen, d.h., die Leseaufträge werden auf beide Platten aufgeteilt  
(= optimale Verteilung der Aufträge, liefert maximal möglichen Wert für  $L_1$ )

Unter der Annahme „keine Lastteilung“ gehen alle Aufträge nur an eine Platte.

Es gilt dann:  $L_1 = N_1$

Spezialfall: nur Schreibzugriffe/s:  $S_1 = N_1$

Jeder Schreibzugriff vom Anwender verursacht intern wegen der Spiegelung einen Schreibzugriff auf beide Platten.

**Nutzbare Plattenkapazität:**

$K = =$

$K_p =$  Kapazität einer Platte,  $z =$  Anzahl Platten

**Der Hauptnachteil von RAID 1 ist die schlechte Ausnützung der Platten-Kapazität mit nur 50% (50% Nutzkapazität, 50% Redundanz).**

→ Plattengeräte, Einbauplätze, Lüftungsanteil, SVGs-Anteil

Die Spiegelung von Platten kann auch vom Betriebssystem durchgeführt werden (Software-Spiegelung). In der Regel ist der Datendurchsatz bei Spiegelung durch ein Betriebssystem wesentlich kleiner als bei Spiegelung innerhalb eines RAID-Systems.

Beim hardwaremäßigen Spiegeln (Mirroring) werden zwei gleich große Platten eingesetzt. Das softwaremäßigen Spiegeln setzt keine gleich großen Platten voraus. Die kleinere Platte bestimmt die Kapazität des Systems.

Häufig wird noch zwischen Spiegelung (Mirroring) und Duplexing unterschieden:

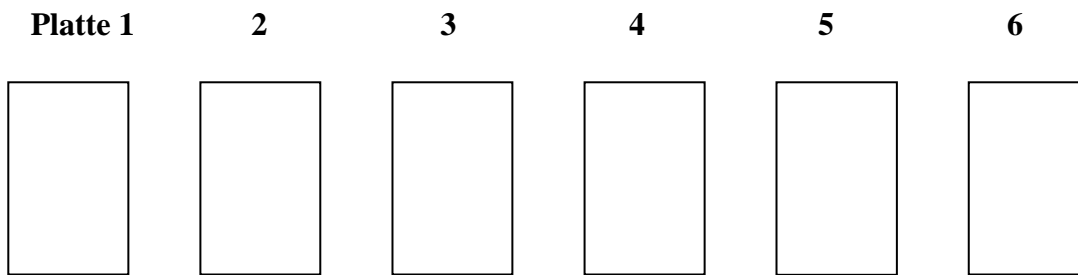
Bei der Spiegelung (Mirroring) hängen beide Platten am gleichen Platten-Controller. Beim Duplexing besitzt zur Erhöhung der Betriebssicherheit jede Platte einen eigenen Controller. Die Verteilung der Aufträge an beide Platten wird vom Betriebssystem (z.B. Novell NetWare) vorgenommen.

### 1.6.4 Kombiniertes System RAID0/RAID1

#### RAID 0+1, auch RAID10 genannt.

Hier werden "Striping" und "Mirroring" kombiniert. Die Dateien werden zerstückelt und parallel auf mehreren Platten abgespeichert. Gleichzeitig werden die Platten gespiegelt. Die hohe Datenrate von RAID0 wird hier mit der Datensicherheit von RAID1 kombiniert. Nachteil ist natürlich wegen der Spiegelung die schlechte Ausnutzung der Plattenkapazität ( Nutzkapazität, Redundanz).

*Aufgabe: Kennzeichnen Sie im Bild, welche der 6 Platten als RAID0-System und welche als RAID1-System arbeiten.*



Frage:

Eine Datei besteht aus 6 Blöcken (Datei = ABCDEF). Wie sind die Blöcke auf die Laufwerke verteilt? ( oben eintragen).

#### Datenrate der Anwenderdaten beim Transfer einer Datei:

$R_p$  = Datenrate einer Platte,  $z$  = Anzahl der Plattengeräte

$R =$

#### Zugriffe/s:

Zeit pro Zugriff =

Schreibzugriffe/s:  $S_{10} =$

Lesezugriffe/s:  $L_{10} =$

Annahme: Lastteilung beim Lesen

(= optimale Verteilung der Aufträge, liefert max. möglichen Wert)

#### Nutzbare Plattenkapazität:

$K =$

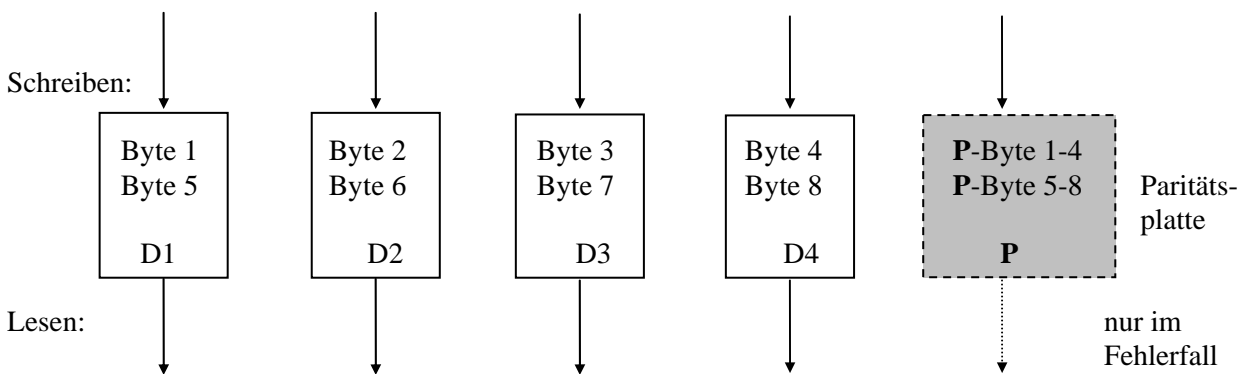
### 1.6.5 System mit Paritätsplatte (RAID 3, RAID 4)

Das RAID3-System umgeht den Nachteil der schlechten Kapazitätsausnutzung bei RAID1 durch Verwendung einer Paritätsplatte anstelle der gespiegelten Platte. Es ist nur 1 Paritätsplatte notwendig, egal wieviele Datenplatten vorhanden sind. Bei z.B. 4 Datenplatten und einer Paritätsplatte beträgt die Kapazitätsausnutzung  $4/5 \cong 80\%$ .

Die Dateien werden in einzelne Bytes aufgesplittet (RAID3) und parallel auf alle Platten verteilt. Bei z.B. 5 Platten wird für je 4 Datenbytes ein Prüfbyte erzeugt und auf der fünften Platte (Paritätsplatte) abgespeichert. Da schon kleinste Datenmengen parallel auf alle Platten verteilt werden, sollten – um die Drehwartezeit möglichst klein zu halten – die Plattenspindeln synchron laufen. Es ist immer nur ein einziger Dateitransfer möglich. Schreib- und Lesezugriffe dauern gleich lang. Bei kleinen Dateien sinkt die mittlere Datenrate des Plattensystems (Die Daten müssen ja in Sektoren/Cluster abgespeichert werden. Dies sind die kleinsten adressierbaren Einheiten auf der Platte.).

Wenn eine Platte ausfällt werden die Daten der defekten Platte aus den restlichen Datenplatten und der Paritätsplatte rekonstruiert. Ein Datenverlust findet nicht statt.

*Frage: Vergleichen Sie die Funktion der Paritätssicherung beim Arbeitsspeicher und beim RAID-System. Wieso ist beim RAID-System mit Paritätssicherung eine Korrektur möglich?*



Beispiel: 4 Datenplatten D1 ... D4  
1 Paritätsplatte P

**System:** 3, 4, 5, 6, ... Platten

Frage:

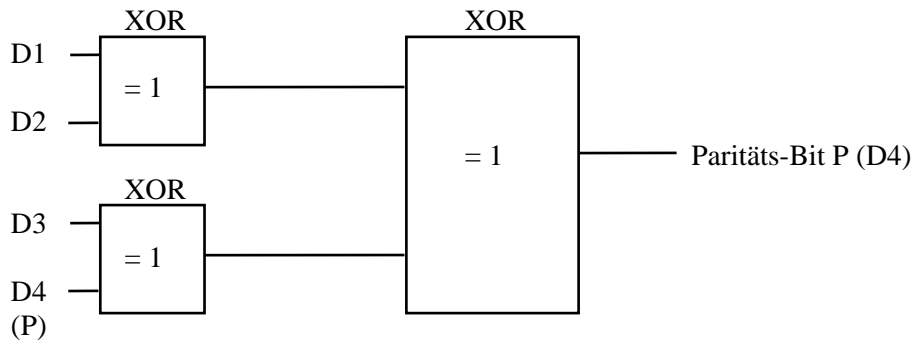
*Eine Datei besteht aus 8 Blöcken (Datei = ABCDEFGH). Wie sind die Blöcke auf die Laufwerke verteilt? (oben eintragen).*

**Fehlerbehandlung:**

Fehler-Erkennung:

Defekter Sektor, defekte Platte:

**Paritätsbit P erzeugen  
oder  
Datenbit D4 aus D1...D3 und P rekonstruieren:**



Wird anstelle von z.B. D4 das entsprechende Paritätsbit P eingespeist (D4 ⇒ P), dann wird anstelle von P das Datenbit D4 erhalten (= Informationsrückgewinnung bei defekter Daten-Platte).

*Aufgabe:*

*Paritätsgewinnung für P-Bit, dann Datenrekonstruktion für D4-Platte aus D1 bis D3 und P. Verwenden Sie einige Bitmuster:*

Paritätsbit erzeugen			Platte D4 defekt: Rückgewinnung D4		
	1. Stufe XOR	Parität P		1. Stufe XOR	D4
D1=0000		<b>P=</b>	D1=0000		<b>D4=</b>
D2=1010			D2=1010		
D3=0011			D3=0011		
D4=1011					

**Datenrate Anwenderdaten beim Transfer einer Datei:**

$R =$   $R_p =$  Datenrate einer Platte

**Zugriffe/s:**

$\text{Zeit pro Zugriff} =$   $t =$  Transferzeit bei nur einer Platte

Zugriffe/s:  $L_3 = S_3 =$  Jede Datei ist über alle Platten verteilt.

**Nutzbare Plattenkapazität:**

$K =$

**Vorteil von RAID 3 gegenüber den gespiegelten Platten von RAID 1:**

----

**Nachteile RAID 3:**

--

### **ähnliche RAID-Varianten, die technisch derzeit aber keine Rolle spielen:**

#### **RAID 4:**

Wie bei RAID3 eine Paritätsplatte. Im Gegensatz zu RAID3 werden die Daten sektorweise oder in Blöcken von Vielfachen eines Sektors (z.B. 128kByte) auf die Datenplatten verteilt. Verschiedene Zugriffe können dadurch gleichzeitig stattfinden.

#### **RAID 30:**

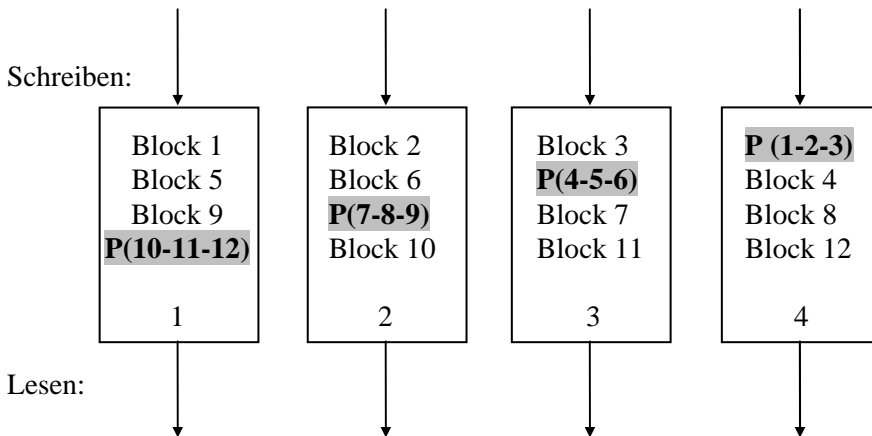
Kombination von RAID 0 und RAID 3, d.h., ein gespiegeltes RAID 3 – System.

#### **RAID2:**

Daten werden bitweise auf die Datenplatten verteilt. Über die Daten wird eine ECC-Information gebildet und auf mehrere ECC-Platten(!) abgespeichert, d.h., anstelle der Paritätsplatte von RAID3 sind mehrere ECC-Platten vorhanden. Z.B. 4 Datenbits und 3 ECC-Bits => 4 Datenplatten +3 ECC-Platten. Wegen des hohen HW-Aufwands wird diese Variante nicht realisiert.

### 1.6.6 System mit rotierender Parität (RAID 5)

Im Unterschied zu RAID3 werden hier die Dateien in Blöcke (und nicht wie bei RAID3 in einzelne Bytes) aufgeteilt und auf die Platten verteilt. Ein Block entspricht meist mehreren Sektoren und sollte mit der Blockgröße (Clustergröße) des Betriebssystems übereinstimmen. Verschiedene Zugriffe (verschiedene Dateien, Lesen und Schreiben gemischt) sind parallel möglich. Die Paritätsinformation ist auf alle Platten verteilt. Damit umgeht man beim Schreiben den Engpaß von nur einer Paritätsplatte und kann beim Lesen von allen Platten gleichzeitig lesen. Zusätzlich werden bei RAID5 auf jeder Platte die Adressen der Paritätsinformationen abgespeichert.



Auf allen Platten: Daten, Paritätsinformation.

{Directory, Belegungstabelle Block frei/ belegt/ defekt/, Tabelle Blöcke ← → zugehöriger Paritätsblock, Tabelle Blöcke <---> Platten-Nr.}

#### Frage:

Beim Überschreiben eines Datenblocks muss der Paritätsblock, z.B. P(1-2-3), neu berechnet werden. Dies erfordert das Lesen aller Blöcke der „Zeile“. Damit sind alle Laufwerke belegt und nicht verfügbar für andere Transfers.

Gibt es eine Lösung, bei der nicht alle Laufwerke belegt sind?

**Schreiben:** 1 Datenplatte belegt + 1 Platte mit Paritätsinformation belegt

**Lesen:** 1 Datenplatte belegt

**System:** 3, 4, 5, 6, ... Platten

Mehrere Dateitransfers parallel möglich. Blockgröße: mehrere Sektoren (.. kByte)

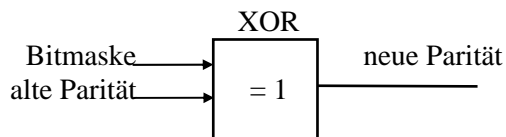
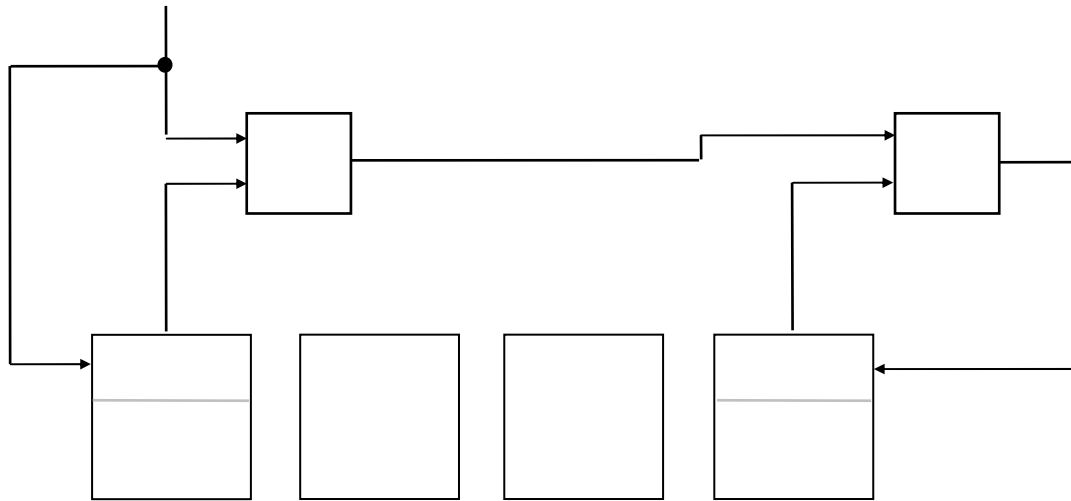
#### Fehlerbehandlung:

Fehler-Erkennung: wie bei RAID 1 über redundante Information im Sektor

Defekter Sektor, defekte Platte: Informationsrückgewinnung über Parität (Bitweise XOR).

*Aufgabe:*

*Welche Aktivitäten sind notwendig beim Überschreiben eines Datenblocks auf Platte 1 ? Ergänzen Sie das Blockschaltbild!*



Bitmaske = 1: → neue Parität = ....

Bitmaske = 0: → neue Parität = .....

Aktivität bei Schreiben eines Datenblocks:

Zeitpunkt t0: Lesen alte Daten, Lesen alte Parität

Zeitpunkt t1: Schreiben neue Daten, Schreiben neue Parität

Die restlichen Platten sind frei.

{Paritätsdisk ist kein Engpaß mehr; aber es ist aufwendig, die Paritätsadressen zu lokalisieren}



### 1.6.7 System mit zwei Paritätsplatten (RAID 6)

Die Fa. Intel propagierte ein „intelligent RAID6“. Bei dieser Variante ist analog zu RAID5 die Paritätsinformation über die Datenplatten verteilt mit dem Unterschied, dass zwei Paritätsblöcke pro „Block-Zeile“ erzeugt werden, d.h., die redundante Paritätsinformation wird im Vergleich zu RAID5 verdoppelt. Der Algorithmus, der die Paritätsblöcke P und Q pro Block-Zeile erzeugt stellt sicher, dass 2 Platten total ausfallen können ohne dass Datenverlust entsteht.

Der Grund für die verdoppelte Redundanz sei, dass das Risiko für einen Datenverlust bei den herkömmlichen RAID-Konzepten zu hoch sei. Fällt eine Platte aus, so dauert das Aktualisieren (= Restaurieren der verlorenen Daten aus den restlichen Platten) der Reserveplatte einige Stunden. Während dieser Zeit würde jeder weiterer Fehler zum Datenverlust führen.

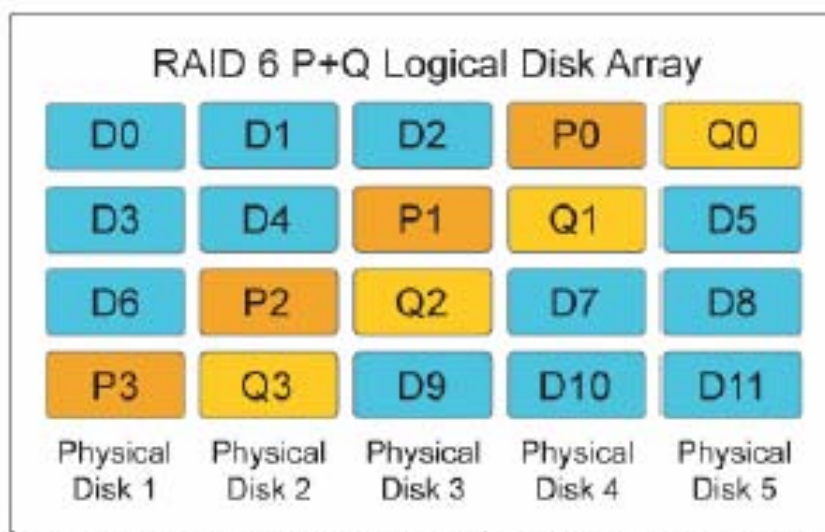


Figure 1. A stripe consists of an equal amount of disk data from each disk, which is addressed contiguously by the RAID controller.

D0, D1, D2, ..... = Datenblöcke  
 P0, Q0, P1, Q1, ..... = Paritätsblöcke

**Nutzbare Plattenkapazität:**  $K = 3 \cdot K_p$        $K_p =$  Kapazität einer Platte

#### How does RAID 6 Work?

RAID 6, as mentioned above, uses two disks of the array for check data, where RAID 5 uses just one. While this leads to a somewhat more expensive array for RAID 6, the result is a significant increase in the reliability of the data stored in the array. The RAID 6 algorithms use two independent equations to compute the stored check values, which enable reconstruction of data when disks and/or blocks fail. Two methods of Intelligent RAID 6 processing are 2D-XOR and P+Q.

The two RAID 6 methods utilize different algorithms, resulting in significantly different behavior and performance for disk I/O processing. This can be shown for writes that span an array stripe. A *stripe*, as shown in **Figure 1**, consists of an equal amount of disk data from each disk, which is addressed contiguously by the RAID controller. Each disk segment is referred to as a *strip*. An array's *stripe size* is therefore equal to the *strip size* multiplied by the number of disks in the array. RAID controllers will also manage the RAID cache in stripe quantities, and so a RAID cache is divided into cachelines equal to the stripe size.

For large I/O writes, or cache write-back operations, all data strips of a stripe are written with new data and the corresponding check values are also overwritten. The P+Q RAID 6 algorithm handles this case efficiently. The 2D-XOR RAID 6 algorithms must process stripe writes as a collection of strip writes, and performs poorly. A higher level of RAID processing is required by 2D-XOR RAID 6 solutions to process cached writes efficiently, and this must be done on a number of stripes of data, creating fewer large cachelines in the RAID cache.

The P+Q RAID 6 algorithm processes a stripe write in the same manner and with nearly the same performance as RAID 5 solutions. The following sequence illustrates the behavior for a large I/O write, but applies equally to a RAID cache write back.

1. All data to be written to disk is first transferred to the intelligent I/O subsystem memory. This is accomplished by a DMA (direct memory access) operation controlled and executed by the I/O processor of the intelligent I/O RAID system.
2. The check values (P and Q) are then computed based on the write data and parameters within the RAID algorithm. This computation is performed on every byte of data, and is executed by the hardware in the I/O processor, thereby eliminating any overhead on the host CPU for this data-processing-intensive operation.
3. The write data and both P and Q check values are then written to the disk stripe, under the control of the RAID algorithm running on the I/O processor.

This sequence illustrates the straightforward processing required to implement the highly reliable P+Q RAID 6 algorithm.

**RAID TP (Triple Parity):**

Wie RAID 6 aber mit 3 Paritätsdisks.

### 1.6.8 Technische Realisierung

Grundsätzlich müßten alle Komponenten eines RAID-Systems redundant sein.

Bei den marktüblichen Systemen ist folgender minimaler Ausbau an Redundanz gegeben:

- 1 Platte: dadurch kein Datenverlust bei Ausfall einer Platte
- 1 Lüfter
- 1 Stromversorgung

Optional (Kostenfrage):

- 1 Reserveplatte (spare disk)

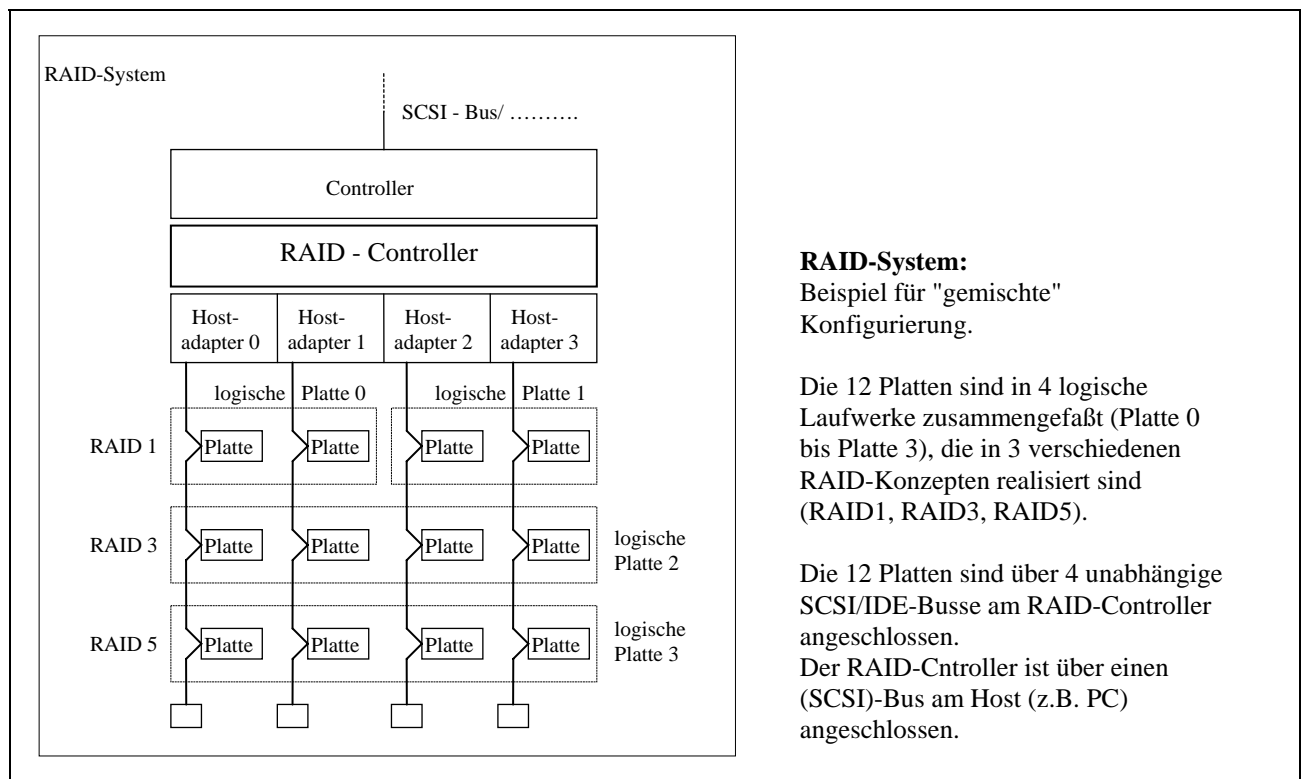
Der Ausfall einer Platte führt zu einer Fehlermeldung. Bei vorhandener Reserveplatte wird diese automatisch mit den Daten aktualisiert, sodaß nach Stunden die Redundanz des Systems wieder gegeben ist. Die defekte Platte wird außer Betrieb genommen. Die Aktualisierung der Daten erfolgt über die restlichen Datenplatten und die redundante Information. Die Aktualisierung erfolgt während des Betriebs. Ersatzplatten können im Betrieb gewechselt werden. Ebenso erfolgt eine notwendige Ersatzsektor-Zuweisung durch den RAID-Controller automatisch im Hintergrund.

- Ein zweiter Bus-Anschluß am Rechner.

Pufferspeicher des RAID-Controllers (Cache):

Kapazität = MByte .. GByte (DRAM, eventuell mit ECC).

Üblicherweise kann ein RAID-System "gemischt" konfiguriert werden. Je nach Anforderung kann z.B. ein Teil der Platten nach RAID1, ein weiterer Teil nach RAID3 und/oder RAID5 konfiguriert sein. Die Konfiguration kann während des Betriebs geändert werden, siehe Bild.



**RAID-System:**  
Beispiel für "gemischte"  
Konfigurierung.

Die 12 Platten sind in 4 logische Laufwerke zusammengefaßt (Platte 0 bis Platte 3), die in 3 verschiedenen RAID-Konzepten realisiert sind (RAID1, RAID3, RAID5).

Die 12 Platten sind über 4 unabhängige SCSI/IDE-Busse am RAID-Controller angeschlossen. Der RAID-Controller ist über einen (SCSI)-Bus am Host (z.B. PC) angeschlossen.

**Netzwerk-fähige RAID-Systeme:**

RAID-Systeme mit entsprechender Schnittstelle (z.B. ..Gbit/s.. Ethernet-Anschluß) und einem Netz-Protokoll (z.B. TCP/IP oder NFS (new file system)) sind direkt an Datennetzen anschließbar. Dies gilt generell für alle netzwerk-fähigen Geräte (Drucker etc.).

zum nächsten Bild:

JBOD = Just a Bunch Of Disks (Disk Array ohne RAID-Controller)

bunch = Haufen, Büschel

Demo „Großes RAID-System“

<http://www.plasmon.com/downloads/pdf/fibrearraydatasheet.pdf>

## FibreArray

### Enterprise Class Fibre-to-Fibre SAN Storage Arrays

Plasmon's FibreArray eXtreme SAN (Storage Area Network) arrays are high-end storage solutions suited to enterprise deployment and mission-critical applications, providing industry leading performance, resilience and scalability with an easy-to-use centralized management interface.

### FibreArray eXtreme 3000

IT professionals are under increasing pressure to carefully manage their resources and provide the demanding service levels required of business critical applications. The FibreArray eXtreme 3000 product line has been specifically designed to accommodate these rigorous, I/O intensive applications, while maintaining a very cost effective price point.

FibreArray systems are highly resilient and provide powerful centralised storage management with the ability to self-diagnose, self-optimize, self-configure and self-heal. For maximum system performance and reliability, FibreArray eXtreme uses a 2Gb, fibre-to-fibre architecture in a high-density enclosure that scales to 32TB.

### SANfinity™ Web Management Software

SANfinity easily manages multiple FibreArray Systems from a single administration console. SANfinity's RAID wizard supports multiple virtual volumes to allow simplified configuration of up to 64 RAID volumes (LUNs). Each volume can have a unique application profile, RAID level and Local or Global Spare assignment. (Refer to the SANfinity Datasheet for full details).

The cost effective FibreArray eXtreme 3000 is the storage solution of choice when business continuity is non-negotiable and maximum performance, availability and functionality is essential.



## Key Features

- Up to 800MB/s, 2Gbps Fibre Channel
- No Single Point of Failure, with cache coherency
- Scalable up to 32TB per solution
- Configuration On Disk (COD)
- Rugged, high-density, 3U 16-bay FC enclosure
- SANfinity web-based management
- FC-Access volume management with LUN masking/sharing

## FibreArray at-a-glance

Features	JBOD		RAID			
	X3001	X3002	X3102	X3204	X3104	X3208
2Gb FC Host Connection	Yes	Yes	Yes	Yes	Yes	Yes
# RAID Controllers	0	0	1	2	1	2
# Host Ports	1	2	1	2	2	4
Max. Performance (MB/s)	200	400	200	400	400	800
Max. RAID Cache	N/A	N/A	1GB	2GB	2GB	4GB
2Gb Fibre Channel Disks	Yes	Yes	Yes	Yes	Yes	Yes
# Expansion Enclosures*	7	7	7	7	7	14
Max. Capacity** (TB)	16	16	16	16	16	32
3U 19" Rackmount	Yes	Yes	Yes	Yes	Yes	Yes

\* Total number of enclosures in a single system.

\*\* 146GB Fibre Channel disks. Disk drive technology is constantly evolving. Contact Plasmon for current list of approved drives.

**Datenblatt:**

Features		JBOD		RAID			
		0	0	1	2	1	2
Storage Controller	Number of RAID Controllers	0	0	1	2	1	2
	Number of Host Ports	1	2	1	2	2	4
Host Interface	2Gb Fibre Channel (FC-AL & Fabric)	.	.	.	.	.	.
	No. of Initiators per Host Port	N/A	N/A	128	128	128	128
Cache Memory	Max. Cache Memory Per Controller	N/A	N/A	1GB	1GB	2GB	2GB
Battery Backup	72 hour Backup of Cache Memory	N/A	N/A	Option	Option	Option	Option
Transfer Speed	Host Data Transfer Speed (MB/sec)	200	400	200	400	400	800
<b>HOST OS SUPPORT</b>							
Operating System	Windows 98/NT/2000/2003; UNIX, Linux, Mac OS, SGI IRIX, Solaris	.	.	.	.	.	.
	AIX, HP UX & Others	.	.	.	.	.	.
<b>RAID</b>							
RAID Levels	0, 1, 10, 3, 5	N/A*	N/A*	.	.	.	.
RAID Sets	Up to 128 RAID Sets Supported	N/A	N/A	.	.	.	.
Disk Drives	Maximum Number of Disk Drives ***	112	112	112	112	112	224
Capacity	Maximum Storage Capacity**	16TB	16TB	16TB	16TB	16TB	32TB
Global Hot Spare	Yes	N/A	N/A	.	.	.	.
Local Hot Spare	Yes	N/A	N/A	.	.	.	.
Automatic Rebuild	Yes	N/A	N/A	.	.	.	.
Disk Interface	2Gb Fibre Channel	.	.	.	.	.	.
Disk Capacity GB	18 / 36 / 73 / 146**	.	.	.	.	.	.
Disk Drive RPM	Up to 15,000	.	.	.	.	.	.
<b>HOT-SWAP &amp; REDUNDANCY</b>							
Disk Drive	16 x Disk Drive Trays	.	.	.	.	.	.
Cooling	3 x Cooling Blowers (n+1)	.	.	.	.	.	.
Power	3 x Power Supply Units (n+1)	.	.	.	.	.	.
Power Entry	3 x Power Inlets	.	.	.	.	.	.
<b>ADMINISTRATION</b>							
GUI Administration	SANfinity Web	Lite	Lite	.	.	.	.
LED Notification	PSU, Temperature, Blowers	.	.	.	.	.	.
Interface Options	Ethernet Management	N/A	N/A	.	.	.	.
	In-band SES Over Fibre Channel	.	.	.	.	.	.
<b>SAN MANAGEMENT SOFTWARE</b>							
SANfinity Web	Open Platform Enterprise Storage Management for a Single Array	.	.	.	.	.	.
	Open Platform Enterprise Storage Management for Multiple Arrays	N/A	N/A	Option	Option	Option	Option
FC Access	Assigns Volume Access to Individual Host Servers, Assigns LUN Numbers	N/A	N/A	Option	Option	Option	Option
<b>ELECTRICAL &amp; ENVIRONMENTAL</b>							
Universal A/C Output	100-240VAC	.	.	.	.	.	.
Nominal Power	650W max. (fully populated)	.	.	.	.	.	.
Operating Environment	-4°F to 140°F (-20°C to 60°C)	.	.	.	.	.	.
	10%-80% Relative Humidity, Non-Condensing	.	.	.	.	.	.
Non-Operating Environment	41°F to 104°F (5°C to 40°C)	.	.	.	.	.	.
	10%-80% Relative Humidity, Non-Condensing	.	.	.	.	.	.
Height	13.2 cm	.	.	.	.	.	.
Width	44.3 cm	.	.	.	.	.	.
Depth	64 cm	.	.	.	.	.	.
Weight Excl. Disks	21kgs	.	.	.	.	.	.

ISO 9001:2000 Certified Due to continuing research and product improvements, features or product specifications described in this brochure may change without notice. © 2004 All rights reserved. Raidtec and the Raidtec logo are trademarks of Plasmon plc. FibreRAID, FlexArray, FibreArray, UltraRAID, RAIDman, RAIDserver, SNAZ, \_ This symbol indicates that the feature is supported as standard. \* JBOD is compatible with third-party RAID or Host-based RAID controllers. \*\* Disk drive technology is constantly evolving. Contact Plasmon for current list of approved drives. \*\*\* Capacities currently based on 146GB drives.

### **1.6.9 Systembetrachtungen**

*Frage/Aufgabe:*

*Ersetzt ein redundantes Plattenspeichersystem die Notwendigkeit für ein Datensicherungssystem (RAID anstatt Back-up mit Bandkassette)?*

*Diskutieren Sie Fehlerfälle wie "Einfachen Fehlerfall", " Fehler infolge Viren" und "Naturkatastrophen-Fall" usw.*

Plattenspeichersysteme bieten Schutz gegen:

Plattenspeichersysteme bieten keinen Schutz gegen:

Deshalb ist notwendig:

Im "normalen" Fehlerfall (nicht Katastrophenfall) arbeitet ein RAID-System unterbrechungsfrei weiter, während im Hintergrund eine Datenaktualisierung für die neu dazugeschaltete Reserve-Platte über ein Backup/Restore-System (Bandkassette) läuft.

## 1.7 Speichernetzwerke

In Computernetzen (Server+Clients) ist die Speicherperipherie in der Regel zentralisiert. Festplatten und Backup-Speicher sind am Server angeschlossen oder über einen eigenen Server (File Server) direkt am Computernetz angeschlossen.

Man spricht hier von Storage Area Network SAN, wenn eine Speicherperipherie an mindestens 2 Servern angeschlossen ist, und von Network Attached Storage NAS bei Anschluss der Speicherperipherie über einen File Server direkt am Computernetz. Der herkömmliche Anschluss der Speicher an nur einem Server - z.B. über SCSI-Bus - wird als Direct Attached Storage DAS bezeichnet.

Die Zentralisierung der Speicher stellt hohe Anforderungen an ihre Verfügbarkeit. Z.B. bedeutet eine Verfügbarkeit von 99,9% im Jahresdurchschnitt noch eine Ausfallzeit von ca. 9 Stunden. Redundante Festplattensysteme wie z.B. RAID-Systeme liefern eine hohe Verfügbarkeit. Zusätzlich sollten auch die Datenwege zwischen Festplattensystem und Server redundant ausgelegt sein.

Vorteile eines zentralen Speicheranschlusses sind:

- eine bessere Ausnutzung der Speicherkapazitäten (dynamische Zuordnung von Kapazitäten an die jeweiligen Server/Benutzer, gemeinsame Nutzung eines Speichersystems durch mehrere Server) und damit
- geringere Hardware-Kosten,
- ein kleinerer Verwaltungsaufwand (Systemadministration) sowie
- eine hohe Verfügbarkeit und
- besserer Leistungsmerkmale (Performance).

Die Zentralisierung der Speicherperipherie wird auch mit Speicherkonsolidierung (Storage Consolidation) bezeichnet. Speicherkonsolidierung bedeutet die Verwendung von weniger Speicher-Hardware durch Zentralisierung.

Ein einfaches Beispiel ist die gemeinsame Nutzung eines RAID-Systems durch zwei Server (Die Kosten eines RAID-Systems doppelter Kapazität sind wesentlich geringer als die Kosten für zwei RAID-Systeme einfacher Kapazität).

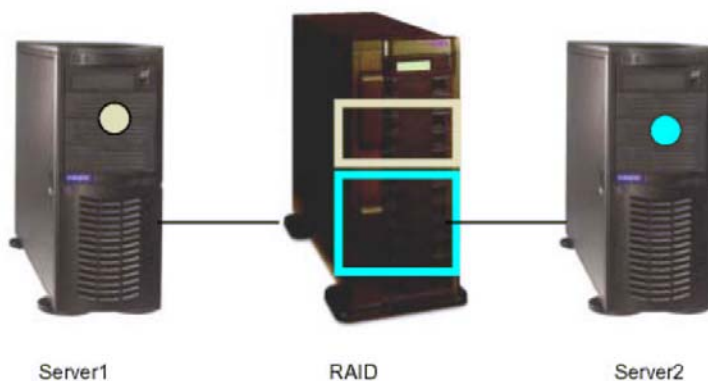


Bild: gemeinsam genutztes RAID-System

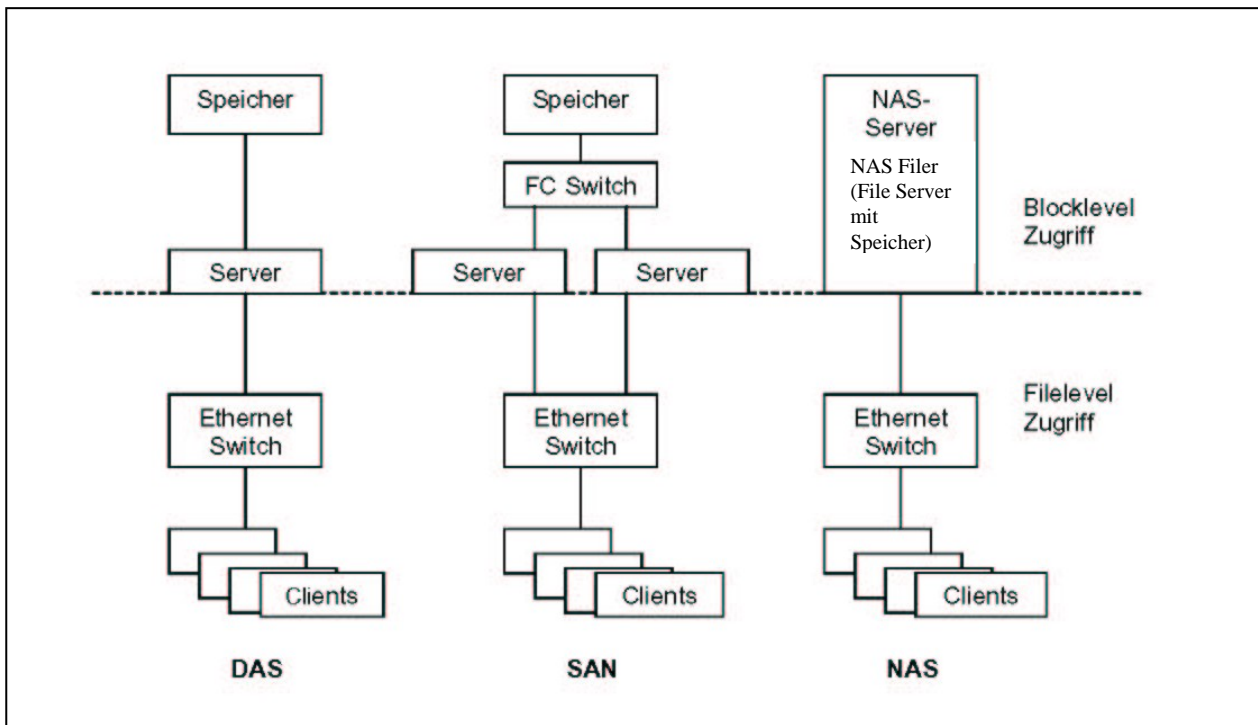


Bild: Speicheranschluss bei DAS, SAN und NAS

**Server (DAS, SAN):**

"allgemeine" Server-Funktionen (z.B. Server für Anwendersoftware wie Modellsim, Leonardo-SW für Schaltungsdesign, Bildschirm-Server für SUNRAY-Terminals usw.) und Dateitransfer-Funktionen

**NAS-Server:**

optimiert für Dateitransfer-Funktionen: zum Ethernet auf Dateiebene (file level), zum Speicher auf Blockebene (block level)

Weitere Informationen dazu in anderen Vorlesungen, z.B. „Computernetze“ usw.

## 1.8 Aufgaben

### **Aufgabe Disk-1: Plattenspeicher, schreiben Datei**

50GByte-Festplatte:

*Datenrate: 28-44 MByte/s (abh. von Zyl.-Position)*

*Sektorgröße: 512 Byte*

*1 Platte/Stapel*

*1 Kopf/Oberfläche*

*7200 U/min*

*Positionierzeit:  $tp=1ms$  Zyl/Zyl*

*Kopfschaltzeit: vernachlässigbar*

1. Wie lange dauert das Abspeichern einer sequentiellen 440 MByte-Datei, wenn Sie vereinfachend nur die Datenrate zur Berechnung verwenden? Die Datei ist auf den äußeren Zylindern abgespeichert. Beurteilen Sie das Ergebnis!
2. Mit welchem Parameter „Zylinder-Versatz“ (Anzahl Sektoren) muss die Platte formatiert sein, damit der berechnete Wert von Frage 1 möglichst erreicht wird? *Annahme: 741 Sektoren/Spur* (siehe SCSI-Befehl "Mode Select" bei Low-Level-Formatieren, Parameter „cylinder skew factor“)
3. Der Hersteller hat den Formatier-Parameter "Zyl.-Versatz" von Frage 2 auf 0 gesetzt. Wie lange dauert das Abspeichern?

### **Aufgabe Disk-2: Low-Level-Formatieren Festplatte**

Festplatte: *Stapel mit 2 Platten, 4 Köpfe, 56 332 Zyl., 7200 U pro min*

Sie wollen Ihre Festplatte abstrafen, indem Sie den Vorgang „Low-Level-Formatieren“ (= Platte plattmachen) starten. Die Platte revanchiert sich, indem der Vorgang „Low-Level-Formatieren“ unendlich andauert. Sie reagieren mit einem Disko-Besuch. Am nächsten Morgen ist der Vorgang „Low-Level-Formatieren“ beendet aber Sie wissen jetzt nicht, wie lange er gedauert hat?

Berechnen sie die Dauer des Vorgangs „Low-Level-Formatieren“ unter den Annahmen

- a) best case
- b) selbst gewählte Annahme in Richtung worst case.

### **Aufgabe Disk-3: Plattenspeicher, abspeichern fragmentierte Datei**

Festplatte:

*Datenrate: 28-44 MByte/s*

*7200 U/min*

*$tp=8ms$*

*Clustergröße=16kByte*

Eine 440 MByte-Datei wird auf einer stark fragmentierten Platte abgespeichert. Die Lücken im belegten Speicherbereich sind im Durchschnitt 10 Cluster groß. Wie lange dauert das Abspeichern?

*weitere Annahmen:*

*Die Datei ist gleichmäßig über den ganzen Zylinderbereich verteilt. Die Speicherbereichs-Lücken befinden sich immer innerhalb eines Zylinders (kein Positionieren beim Abspeichern innerhalb einer Lücke).*

**Aufgabe Disk-4: Plattenspeicher, Ersatzsektorzuweisung**

Eine 400 MByte-Datei ist sequentiell abgespeichert in einer Zone mit 1000 Sektoren pro Spur.

Datenfeld pro Sektor = 0,5 kByte.

$t_u = 8,3 \text{ ms}$

- Berechnen Sie die Zeit  $t_1$  für das Lesen der Datei unter den Annahmen:  
 -- Positionierzeit und Drehwartezeit zum Dateianfang vernachlässigbar.  
 -- Positionierzeit Spur/Spur und Zylinder/Zylinder vernachlässigbar.
- Sie haben die Festplatte bei Ebay ersteigert: Auf jeder Spur ist ein Sektor defekt (entspricht 1% defekte Sektoren, ein absolut faires Angebot). Dafür wird ein Ersatzsektor zugewiesen, der sich am Ende der jeweiligen Spur befindet. Ansonsten gelten die gleichen Annahmen wie bei Frage 1.  
 Berechnen Sie die Zeit für das Lesen der Datei.

Mit welcher Maßnahme könnte man die Zeit für das Lesen der Datei verringern?

- Ihr Kommilitone hat ebenfalls eine Festplatte der gleichen Qualität bei Ebay ersteigert. Auch bei ihm ist pro Spur ein Sektor defekt. Der Ersatzsektor ist bei ihm der erste Sektor auf einer Spur in einem eigenen Ersatzsektor-Zylinderbereich. Die Positionierzeit dorthin beträgt  $t_p = t_u$ .  
 (Ansonsten gelten die gleichen Annahmen wie bei Frage 1.)  
 Berechnen Sie die Zeit für das Lesen der Datei.

**Aufgabe Disk-5: Plattenspeicher, Aufteilung in Sektoren**

Gegeben ist die Tabelle mit der Aufteilung in Spuren (Tracks) und Sektoren.

Zone	Tracks in Zone	Sectors Per Track	Data Transfer Rate (Mbits/s)
0	624	792	372.0
1	1,424	780	366.4
2	1,680	760	357.0
3	1,616	740	347.6
4	2,752	720	338.2
5	2,880	680	319.4
6	1,904	660	310.0
7	2,384	630	295.9
8	3,328	600	281.8

Sie brauchen eine Partition auf der Festplatte, bei der die Datenrate mindestens 44 MByte/s beträgt.

1. Welcher Spurbereich eignet sich dafür?  
*Der Spurbereich beginnt in Zone 0 mit Spur 0  
„Tracks in Zone“ = Anzahl Spuren auf einer Oberfläche in der jeweiligen Zone*
2. Welche maximale Kapazität in GByte kann die Partition besitzen (Faktor 1000 verwenden)?  
*Die Festplatte hat einen Stapel mit 2 Platten. Alle Oberflächen sind Datenoberflächen.  
Sektorgröße=0,5 kByte (Datenfeld)*
3. Wie lange dauert das Abspeichern einer sequentiellen 10 MByte-Datei in dieser Partition?  
*Positionierzeiten vernachlässigbar. Zonen-Grenzen werden nicht überschritten.*

### **Aufgabe Disk-6: Plattensystem mit RAID-Konzept**

*Das RAID-System ist über einen SCSI-Bus angeschlossen. "Datei Lesen" oder "Datei Schreiben" erfordert von der CPU einen Zugriff auf dem SCSI-Bus.*

*$t_p = 10$  ms mittlere Positionierzeit, 7200 U/min, Transferzeit  $t$  pro Zugriff =  $2,5 \cdot$  Umdrehungszeit  
2 Plattengeräte*

1. Wieviel Lese-Zugriffe/s  $L=?$  kann bei RAID1 die CPU mit dem Plattensystem maximal durchführen?
2. Wieviel Zugriffe/s  $N=?$  (Lesenzugriffe+Schreibzugriffe) kann die CPU bei RAID1 mit dem Plattensystem durchführen unter der Annahme: 1/3 Schreibzugriffe, 2/3 Lesezugriffe ?
3. Ab wieviel Platten liefert das RAID3-Konzept mindestens 60% mehr nutzbare Plattenkapazität als das RAID10-Konzept bei gleicher Plattenzahl? Berechnung!

## 2 Portable Datenträger mit Direkt-Adressierung

Fragen: portabel = ? Direkt-Adressierung = ? Gegensatz dazu ? Beispiele ?

### 2.1 USB Speicherkarten (USB Sticks)



#### **Technische Daten:**

Kapazität	.. GByte .....	
Datenrate	Lesen: .... 25 MByte/s (USB2.0)...	Anzahl Schreib(Lösch)-Zyklen begrenzt, z.B. 1 Mio (vergrößert durch chip-interne Redundanz), siehe auch Kap. „Flash-Chip“ Siehe USB-Kapitel .....
	Schreiben: .... 15 MByte/s....	
Schnittstelle	USB 2.0	
Temp.-Bereiche	Betriebstemperatur.: -10°C .. +40°C	
	Datenstabilität: -20°C .. +70°C	Datenbestandszeit: ..10..20.. Jahre

#### **Dateisystem:**

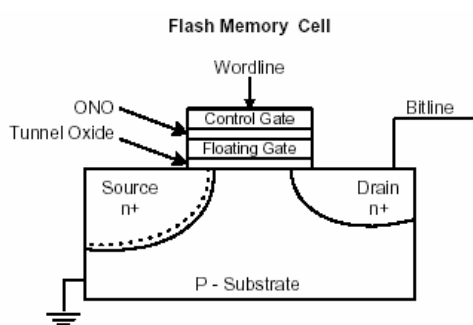
Auf dem USB-Stick werden vom Betriebssystem ein Belegungs-Verzeichnis und ein Datei-Verzeichnis angelegt. Damit kann auf Dateien direkt zugegriffen werden. Unter Windows kommt FAT16 oder FAT32 (ab 2 GByte) zum Einsatz. Partitionen sind möglich. Bootfähig.

#### **Kompatibilität:**

Die Formate FAT16 und FAT32 werden auch von anderen Betriebssystemen wie Linux und Mac OS erkannt.

#### **Flash-Chip:**

##### **Speicherzelle und -matrix:**



##### **Betriebsart und Blockstruktur:**

Die Betriebsarten (modii) beim Flash-Chip sind Löschen, Programmieren (Schreiben) und Lesen.

Die Speichermatrix ist in Blöcke unterteilt, die nur als ganzes gelöscht (erase) und neu programmiert werden können. Jeder Block ist nahezu beliebig oft löscht- und beschreibbar.

Das Lesen erfolgt wesentlich schneller.

Bspl.: 128 MBit-Chip = 16 MByte, organisiert als 16MByte x 1 oder 8MByte x 2  
interne Organisation: 128 Blöcke zu 128 kByte, Wortbreite = 1 oder 2 Byte  
Löschzeit pro Block  $\approx$  ..1s., Programmierzeit pro Byte  $\approx$  ..1us..  
Ändern Block  $\approx$  ..100 000.. mal  
Lesen: Zykluszeit  $\approx$  ..20ns.. (z.B. 66MHz-Betrieb ohne Wait States möglich).

Durch Redundanz im USB-Stick erhöht sich die Anzahl der Schreibzyklen.  
Höhere Datenraten des USB-Sticks sind durch Parallelschalten von Flash-Chips möglich.  
Blockstruktur → blockweises Löschen/Programmieren, Cluster (FAT)

**Software-Schnittstelle:**

Die Chipdaten (Blockgröße, Wortbreite, dynamische Kennwerte) sind in internen Festwert-Registern gespeichert und vom Treiberprogramm abfragbar. Für den Anstoß von Lösch-, Programmier- und Lesezyklen sind im Scalable Command Set (SCS) Befehle definiert, die für unterschiedliche Flash-Chips gültig sind und somit den Treiberaufwand in der Software reduzieren (Common Flash Interface CFI).

**2.2 Diskettenspeicher (Floppy Disk, Flexible Disk)**

[www.imation.com](http://www.imation.com), [www.iomega.com](http://www.iomega.com)

### 2.2.1 1.44MByte-Diskette

Die Diskette ist eine flexible Kunststoffscheibe, beidseitig magnetisiert. Das Laufwerk ist ein stark vereinfachtes billiges Plattenlaufwerk. Im Unterschied zur Platte besitzt es einen Löschmagnet zu beiden Seiten des Schreib/Lesekopfes, um vor dem Schreiben die linke und rechte Flanke einer Spur zu löschen (Schreiben auf verschiedenen Laufwerken mit unterschiedlicher Zentrierung). Mechanischer Schreibe Schutz (Schalter, Kerbe offen/geschlossen) vorhanden. Für den Austausch kleiner Datenmengen gut geeignet, da fast an jedem Computer vorhanden.

	<b>Diskette</b>	<b>Platte</b> (zum Vergleich)
Positionierzeit <b>tp</b>	3ms Spur/Spur	
mittlere Positionierzeit	120ms	10ms
Kopfladezeit <b>tload</b>	10ms	---
Kopfberuhigungszeit <b>ts</b>	10ms	---
Umdrehungszahl (U/min)	300..360	2500..15 000
Mittlere Drehwartezeit <b>t<sub>d</sub></b>	100ms (bei 300 U/min))	5ms (bei 6000)
Datenrate	62.5kByte/s (3½" Diskette)	20..50MByte/s
<b>Ergebnis:</b>	<b>Geschwindigkeit der Platte um den Faktor ..100.. höher</b>	

Diskettenkapazität (3½", DS, HD):

80 Spuren/Seite \* 2 Seiten \* 18 Sektoren/Spur \* 0.5 kByte/Sektor = 1 440 kByte formatiert

#### Sektorformat der DOS-Diskette (Soft-sektoriert):

Das Sektorformat **soft-sektoriert** Disketten besteht wie bei der Platte aus Adreß- und Datenfeld. Der Sektorbeginn wird durch eine Adreßmarke am Beginn des Adreßfelds markiert. Wie bei der Platte ist der Sektor die kleinste adressierbare Einheit der Diskette:

Adresse = Zylinder-Nr. (Spur)/ Kopf-Nr. (Seite)/ Sektor-Nr.

Zusätzlich steht im Adreßfeld die Länge N des Datenfelds:

Länge (Kapazität) =  $128 * 2^N$  Byte (N = 2 liefert  $128 * 4 = 512$  Byte)

Es existiert noch ein altes Sektorformat, das sogenannte **hard-sektorierte** Format. Hier sparte man sich aus Kapazitätsgründen das Adreßfeld. Ein Sektor besteht nur aus dem Datenfeld sowie einigen Synchronisierbytes. Der Sektorbeginn wird über Sektorlöcher am Diskettenumfang mit einer LED abgetastet. Die Sektorzahl und Größe sind fest.

#### Geräteanschluß: Intern: IDE, extern: USB

### 2.2.2 Disketten mit hoher Kapazität

Die Weiterentwicklung führte zu verschiedenen Gerätetypen wie der optischen Diskette ("Floptical"), der LS-diskette und des ZIP-Laufwerks mit Kapazitäten bis ca. 1 GByte.

#### a) Optische Diskette

**Kapazität:** bis 120 MByte

**Kompatibilität:** In den Laufwerken können auch herkömmliche 1,44MByte-Disketten (3 1/2") gelesen, formatiert und beschrieben werden.

Magnetische Datenaufzeichnung:

Vertikale Magnetisierung. Dadurch sehr hohe Bitdichte (.. 36000 .. bpi) auf der Spur mit entsprechend vielen Sektoren pro Spur. MFM-Aufzeichnung.

Optische Spurführung:

Zwischen den Datenspuren sind optische Servospuren (Spurbreite ca. 4µm) eingeprägt, die mit einer LED abgetastet werden (anderer Reflexionsfaktor) und somit eine genauere Spurführung ermöglichen. Am Schreib/Lesekopf ist zusätzlich ein optischer Kopf mit LED und Photodetektor integriert. Die Spurdichte erhöht sich dadurch um den Faktor .. 10 ..

#### b) ZIP-Laufwerk (Quelle: IOMEGA)

3½" - Diskette mit Kapazitäten von derzeit: 100 MByte, 250 MByte, 750 MByte

Technische Daten:

- 100/ 250/ 750 MByte formatierte Kapazität
- ca. 3000 U/min, 29ms („seek time“)
- Datenrate: 0.79 (innen) .. 1.40 (außen) MByte/s
- Schnittstelle:
  - extern: USB (Stromversorgung über USB), FireWire, Parallel Port, SCSI
  - intern: IDE/ ATAPI

*Frage: Ist die Zahl der Sektoren pro Spur konstant?*

**c) LS-Diskette**

LS = Laser Servo mechanisme

Technische Daten: 3½" - Diskette

- 120/ 240 MByte formatierte Kapazität (UHD Mode)
- ca. 1440 U/min, 70ms („seek time“)
- Datenrate: 0.75 (innen) .. 1.360 (außen) MByte/s
- Schnittstelle: Intern: IDE, extern: USB, usw.
- kompatibel (lesen und schreiben) mit 1.44 MByte Diskette (HD Mode)

## 2.3 Optische Speicher

<http://www.t10.org/ftp/t10/drafts/mmc5>, <http://www.tecchannel.de/hardware/821/index.html>, HP.com, c't 22, 2005

Der Technologie-Fortschritt (Verkleinerung der Strukturen und damit höhere Speicherkapazitäten) fand bisher in 3 Schritten statt:

alt: <b>CD</b>	→ <b>DVD</b>	→ <b>BD, HD DVD</b>
compact disk 0,5 GByte	digital versatile disk 4,7 GByte SL, 8,5 GByte DL .....	HD DVD = high density DVD 15 GByte SL, 30 GByte DL BD = Blu-ray Disk 25 GByte SL, 50 GByte DL
Bei allen wird beim Datenträger unterteilt in Read Only, Recordable und Rewritable:		
CD-ROM CD-R CD-RW	DVD-ROM DVD-R, DVD+R DVD-RW, DVD+RW	-ROM -R HD DVD: -RW BD: -RE (rerecordable)

SL= Single Layer, DL= Double Layer

*Frage: Die obigen Typen sind wie geordnet?*

Die Kapazitätswerte verdoppeln sich jeweils bei 2 Schichten pro Seite (double layer DL) und bei Aufzeichnung auf beiden Seiten (double sided DS).

### HD DVD:

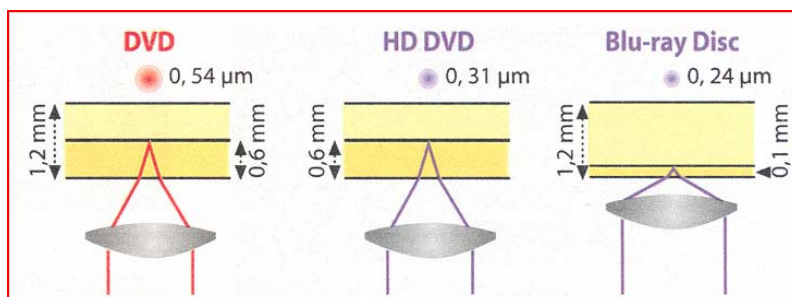
Wichtiges Anliegen war es, den physikalischen Schichtaufbau der DVD möglichst beizubehalten, um Investitionskosten bei der Umstellung von DVD auf HD DVD bei der Disk-Herstellung gering zu halten.

### BD Blu-ray Disk:

Die Strukturen wurden technisch ausgereizt, daher die höhere Kapazität gegenüber HD DVD. Hat einen besseren Kopierschutz als die HD DVD.

### Hybrid-Disk:

Eine DVD-Schicht und eine HD DVD-Schicht (bzw. BD-Schicht).

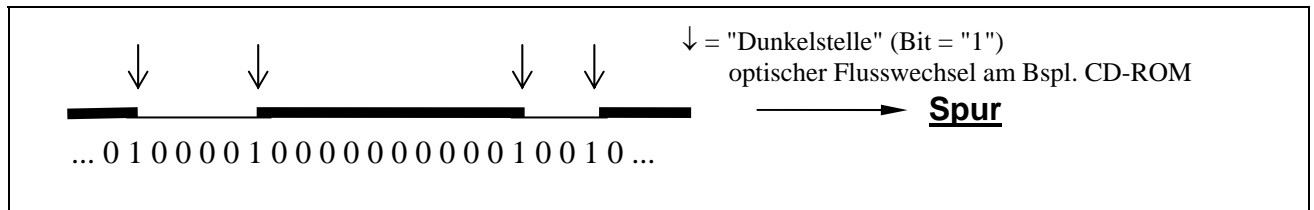


Die optischen Speicher sind als Audio-CDs, als Video-CDs und als Daten-CDs (Computertechnik) verwendbar. Die folgenden Ausführungen konzentrieren sich auf Daten-CDs.

Optische Techniken sind generell einsetzbar im Positioniersystem (z.B. optische Diskette) und bei der Datenaufzeichnung auf der Spur.

### 2.3.1 Physikalische Ebene bei der Datenaufzeichnung

Für alle CD-Typen gilt beim Aufzeichnungsverfahren:



Die Daten sind bitseriell auf der Spur abgespeichert. Beim Lesen (=Abtasten mit Laserstrahl) der Spur verursachen Änderungen im Reflexionsverhalten der Spuroberfläche einen Intensitätswechsel (hell/dunkel) des reflektierten Lichts, der von der Leseelektronik als "1" interpretiert wird. Das Erzeugen von Zonen unterschiedlichen Reflexionsverhaltens beim Schreiben (oder Pressen) einer CD ist bei den jeweiligen Gerätetypen (CD-ROM, CD-R usw.) verschieden.

Durch Codierung der Daten (z.B. RLL 10,2-Code) werden lange "0"-Folgen und direkt aufeinanderfolgende "1" verhindert. Der Bit-Abstand wird durch den Geräte-Takt festgelegt. Beim Lesen wird der Gerätetakt über die "1"-Signale synchronisiert.

Die Mindestlänge der Zonen gleichen Reflexionsverhaltens (z.B. 0,4µm) begrenzt die Bitdichte auf der Spur, der Mindestabstand der Spuren (z.B. 0,7µm) die Spurdichte.

#### Fragen:

Wieso sind lange "0"-Folgen und direkt aufeinanderfolgende "1"-Bits nicht erwünscht?

Wie groß sind minimaler und maximaler Abstand zweier optischer Flusswechsel (in Bitpositionen)?

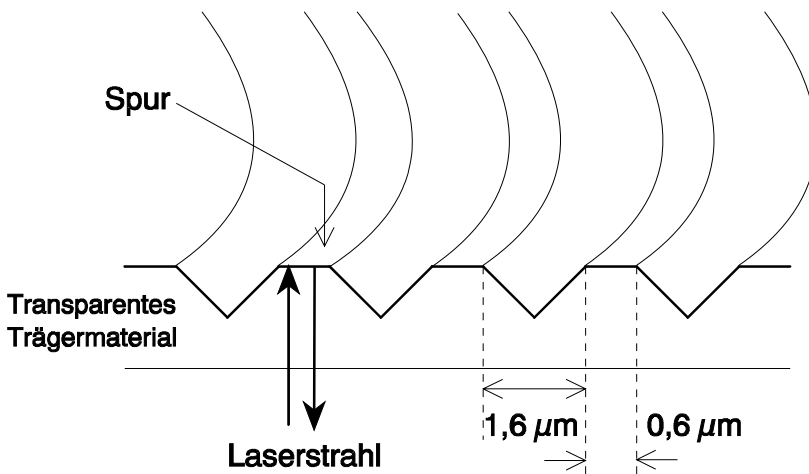
#### 2.3.1.1 "ROM"-Platten (Read Only Memory) (Quellen: Maxell, Philips)

Die Platten sind nur lesbar, nicht beschreibbar. Die Information wird über einen mechanischen Fertigungsprozeß beim Hersteller aufgebracht (Spritzguß).

- **CD-ROM:** 0,5 GByte bei 5¼ Zoll, 74 Minuten für 650 MB oder 80 Minuten für 700 MB
- **DVD-ROM:** 4,7 GByte bei 5¼ Zoll, einseitig+Einschicht-Verfahren (als Musikkassette ca. 480 min).
- **BD-ROM:** 25 GByte bei 5¼ Zoll, einseitig+Einschicht-Verfahren

**a) CD-ROM:**

**Querschnitt CD-ROM:**



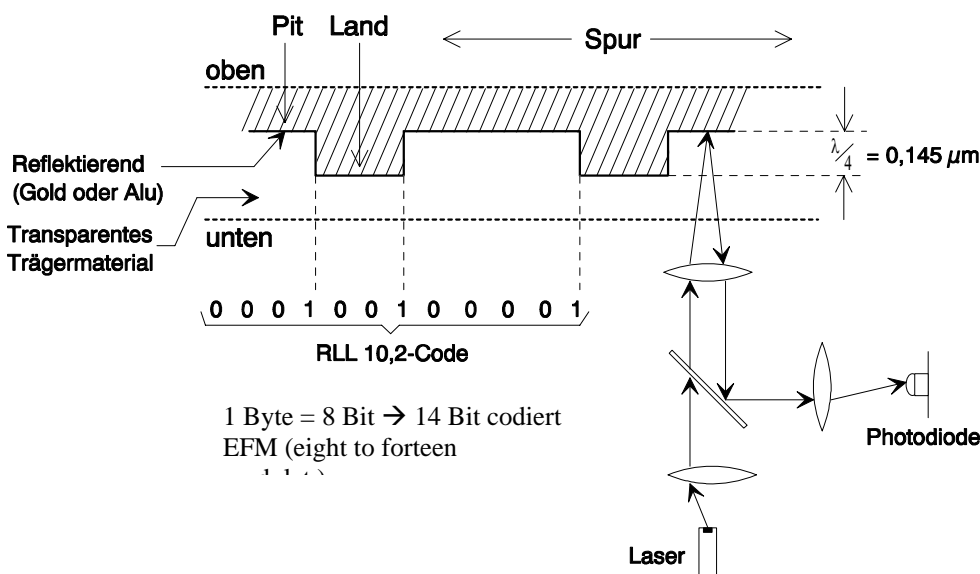
Spur schon bei Herstellung festgelegt (nicht wie bei Magnetplatte durch Formatieren)

Optische Spurführung verhindert Abgleiten von Spur (Schräge reflektiert weniger Licht)

Spur-Maß: 1.0um+0.6um=1.6um (Fehler in Bild)

Optische Spurführung (Spurdichte 15875 tpi, HP 6100)

**Schnitt längs der Spur:**



$\lambda$  = Wellenlänge  
= 0,58 μm  
(589 nm)

An der Grenze Pit/Land: Auslöschung des reflektierten Lichts, da  $\Delta x = 2 * \lambda/4 = \lambda/2 \cong '1'$  bei Übergang Pit ↔ Land

'1' = dunkel, '0' = hell

(analog zur NRZI-Aufzeichnung: '1' = FW, '0' = kein FW)

Pit = Vertiefung (aus Sicht des abtastenden Lichtstrahls), Land = Erhöhung

**Datenaufzeichnung (channel modulation):**

DVD:	EFM+, RLL(2,10)	eight-to-fourteen modulation
HD DVD:	ETM, RLL(1,10)	eight-to-ten
BD:	1.7PP	

**Fragen:**

1) Welche Lichtquellen liefern Licht einer Wellenlänge (monochromatisches Licht) ?

2) Wie erkennt das Gerät beim Lesen 4 aufeinanderfolgende "0"-Bits ? 3) Mit Hilfe welcher Technik werden die kleinen Strukturen (um) erzeugt ?

Berührungsloses Abtasten durch Laser → unempfindlicher (als magn. Platte) gegen Schmutzpartikel.

In der Regel tastet ein einziger Laserstrahl die Spur ab. Bei manchen Laufwerken teilt das Linsensystem den Laserstrahl in mehrere Strahlen auf, sodaß parallel mehrere Windungen gleichzeitig gelesen werden. Dieses Prinzip liefert bei kleineren Drehzahlen hohe Transferraten.

**Herstellung:**

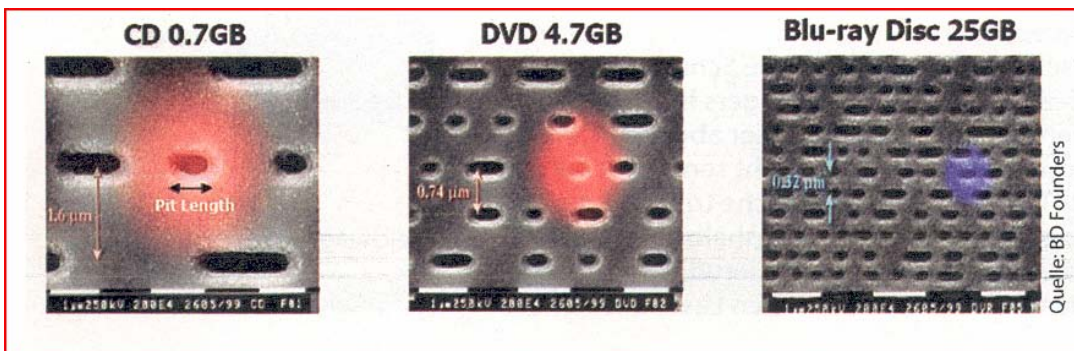
Aus einer Masterplatte (Glas mit lichtempfindlicher Schicht, Belichtung mit 100 mW-Laser) wird ein Nickelabdruck zur Herstellung von Matrizen für die Fertigung der Trägerplatten erstellt. Nach der Pressung der Einzel-CD wird auf das spiralförmige Pit-Muster ein metallischer Niederschlag aufgebracht und die Oberseite mit einem Schutzlack versiegelt.

**b) DVD-ROM, HD DVD-ROM und BD-ROM:**

Die Kapazitätserhöhungen werden u.a. durch eine höhere Spurdichte und eine kleinere Pitlängen auf der Spur erreicht.

CD	DVD	HD DVD	BD
0,5 GByte	4,7 GByte (Musik ca. 480 min).	15 GByte 30 GByte (double layer)	25 GByte 50 GByte (double layer)
Physikalische Ebene:			
$\lambda=0,780 \text{ um} = \text{Infrarot}$	0,650 um = Rot	0,405 um = Blau	0,405 um = Blau
fokus. Lichtstrahl: $\varnothing= 2,1 \text{ um}$	$\varnothing=0,54\text{u}$	$\varnothing =0,31\text{u}$	$\varnothing = 0,24\text{u}$
(Strahlfläche: $4,41 \text{ um}^2$ )	Fläche=1 → <b>Faktor1</b>	Fläche=0,32 →Faktor=3,1	Fläche=0,19 →Faktor=5,3

Die Struktur-Verkleinerung von DVD zu BD beträgt – aus Sicht der Laser-Lichtstrahlfläche, mit der abgetastet wird – etwa einen Faktor 5,3. Dementsprechend erhöht sich bei gleicher Fläche des Datenträgers die Speicherkapazität.



**Bild: Vergleich der Strukturen Pit und Land**

**Aufbau bei 2 Schichten pro Seite:**

Die Pit/Land-Struktur der ersten Schicht wird auf eine "Halbdisk" gepreßt. Darüber wird eine halbdurchlässige Schicht (dünne Goldschicht) aufgetragen; danach eine Kunststoffschicht, in die die Pit/Land-Struktur der zweiten Informationsschicht gepreßt wird. Nach dem Aushärten wird eine reflektierende Alu-Schicht aufgetragen. Zwei Halbdisks werden Rückseite an Rückseite zu einer Disk miteinander verbunden. Das Hauptproblem ist die Dimensionierung der halbdurchlässigen Schicht so, daß von beiden Schichten ein ausreichendes Lichtsignal zurückkommt (Fokussierung auf innere Schicht und auf äußere Schicht).

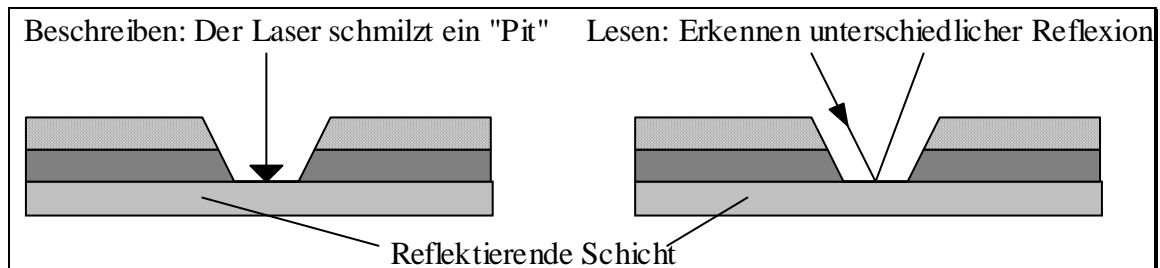
**Variante: CD-R (recordable) 0,5GByte**

Laserstrahl schreibt Information auf die Spur, Datenformat auf der Spur dasselbe wie bei CD-ROM. Die CD-R ist 1x beschreibbar (Schreibverfahren wie bei CD-WORM, Format wie CD-ROM).

### 2.3.1.2 "Einmal" beschreibbare Platten

- **CD-R (Recordable)**
- **DVD-R**
- **HD DVD-R**
- **BD-R**
- **Exot: CD-WORM (Write Once Read Many):** 0.5 GByte bei 5¼ Zoll (älterer Typ)

#### CD-R:



#### Schreiben (Brennen):

Ein Laser erzeugt Pits.

- Er brennt Löcher in ein Metallsubstrat (s. Bild)  
oder

- er heizt die Metallschicht auf und erzeugt in der darunterliegenden Schicht (zersetzbare Polymer) eine Blase.

Bei beiden Varianten wird der Reflexionsfaktor gegenüber den nicht bestrahlten Teilen der Spur ("Lands") verändert.

#### Lesen:

Ein Laser geringerer Leistung tastet die Spur ab. "Pits" und "Lands" reflektieren das Laserlicht unterschiedlich stark.

#### **a) CD-R, DVD-R, BD-R:**

Spurformat wie CD-ROM (spiralförmig). Die Spur ist auf dem Rohling vorgegeben, die Information auf der spiralförmigen Spur in Form von Pits und Lands wird beim Brennen aufgebracht.

#### **b) CD-WORM:**

Die Blöcke sind in konzentrischen Spuren (wie bei Magnetplatte) angeordnet. Die Umdrehungszahl ist konstant (CAV).

#### **Anwendung:**

Archivierung von Daten. Die Archivierungsgarantie (Zeit für Datenerhalt) der Hersteller liegt bei 10 .. 40 Jahren, d.h., in dieser Zeit ist kein Auffrischen (wie beim Magnetband) notwendig.

### 2.3.1.3 "Beliebig oft" beschreibbare Platten

Die Platten sind beschreib- und lesbar.

- **CD-RW (ReWritable), HD DVD-RW, BD-RE:** "Phase-Change"-Aufzeichnungsverfahren (optisch).
- **DVD-RAM:** "Phase-Change"-Aufzeichnungsverfahren (optisch).
- **Exot: MO-Disk (Magneto-Optical):** "Magneto-optisches" Aufzeichnungsverfahren.

Da diese Platten beschreib- und lesbar sind, stehen sie theoretisch in Konkurrenz zu den Magnetplatten. Beim Datenzugriff sind diese optischen Platten (ca. 3600 U/min, mittlere Zugriffszeit ca. 35 ms) derzeit jedoch deutlich langsamer als die magnetischen Festplatten (ca. 10 000 U/min, 5 ms mittlere Zugriffszeit).

#### a) CD-RW, DVD-RW (DVD-RAM), HD DVD-RW, BD-RE:

##### "Phase-Change"-Aufzeichnungsverfahren

Die unter der Reflexionsschicht liegende Aufzeichnungsschicht wird durch den Schreiblaser punktuell mit zwei verschiedenen Temperaturen erhitzt. Der Abkühlungsvorgang ist von der dieser Erhitzung abhängig und liefert entweder eine kristalline ("land") oder eine amorphe ("pit") Oberfläche. Die Spur ist auf dem Rohling vorgegeben.

- kristalline Zone mit guter Reflexion (entsteht durch geringe Erhitzung und langsame Abkühlung)
- amorphe Zone mit schlechter Reflexion (entsteht durch starke Erhitzung und schnelle Abkühlung)

Bei Platten mit Phase-Change-Verfahren heißt "beliebig oft" beschreibbar derzeit ca. 1000x Neubeschreiben. In Plattenbereichen, die häufig verändert werden (Inhaltsverzeichnis der Platte) ist dies noch sehr hinderlich. Mit der Methode, defekte (=nicht mehr überschreibbare) Sektoren zu markieren und den Inhalt an eine andere Stelle auf der Platte zu schreiben, kann diese Beschränkung teilweise umgangen werden (Ersatzsektoren).

Das Verfahren erlaubt im Gegensatz zum magneto-optischen Verfahren ein Schreiben ohne vorheriges Löschen (nur 1 Umdrehung pro Spur).

#### b) MO-Disk: "Magneto-optisches" Aufzeichnungsverfahren

##### Prinzip:

Die Magnetisierungsrichtung der Datenoberfläche ( $\uparrow$  oder  $\downarrow$ ) verändert die Polarisationssebene des reflektierten Lichts (Kerr-Effekt). Ein Strahlteiler separiert das Licht nach Polarisationssebene ( $\perp$  oder  $\parallel$ ). Die Intensität des reflektierten Lichts einer Polarisationssebene hängt damit von der Magnetisierungsrichtung der Datenoberfläche ab und liefert optische Flusswechsel die von der Leseelektronik als Bit="1" interpretiert werden.

##### Schreiben:

Die Magnetschicht hat eine sehr hohe Koerzitivfeldstärke von ca. 8000 Oersted. Damit ein Ummagnetisieren (=Schreiben) der Spur möglich ist, wird durch lokale Erhitzung der Spur mit einem Laserstrahl auf ca. 200 °C (=Curie-Punkt) die Koerzitivfeldstärke an dieser Stelle auf ca. 150 Oersted abgesenkt. Somit läßt sich mit einem schwachen Magnetfeld die Spur an den erhitzten Punkten ummagnetisieren. Das Erhitzen mittels Laserstrahl erlaubt kleinste Strukturen.

1. Umdrehung: die ganze Spur in Richtung  $\downarrow$  magnetisieren (=löschen).

2. Umdrehung: Spurbereiche entsprechend Bitmuster in Richtung  $\uparrow$  magnetisieren.

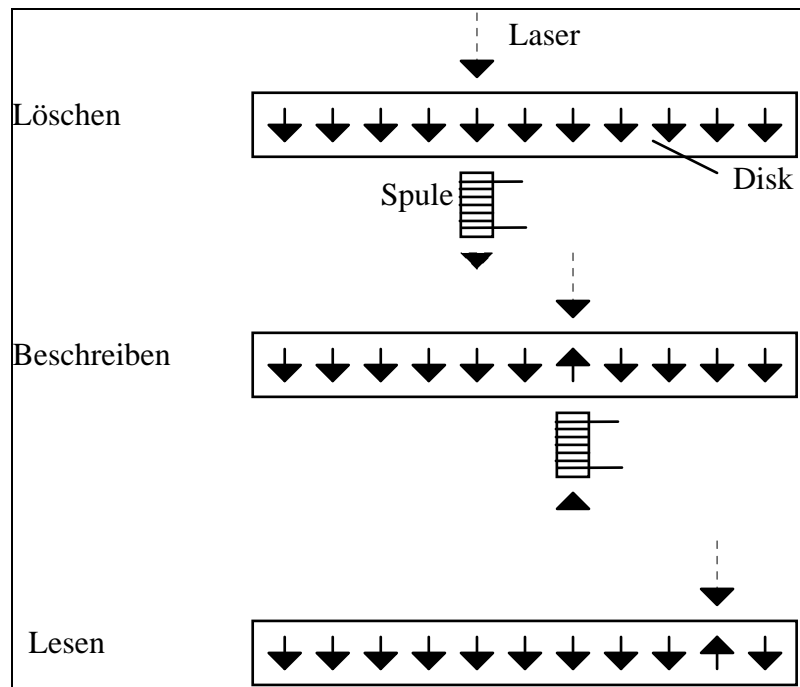
(Grund für die zwei Umdrehungen: Trägheit des Magnetfelds von ca. 50us Umschaltzeit. Damit sind Änderungen entsprechend 200ns/Bit nicht möglich)

##### Lesen:

Die Magnetisierungsrichtung der Datenoberfläche ( $\uparrow$  oder  $\downarrow$ ) verändert die Polarisationssebene des reflektierten Lichts (Kerr-Effekt). Ein Strahlteiler separiert beim Lesen das Licht nach

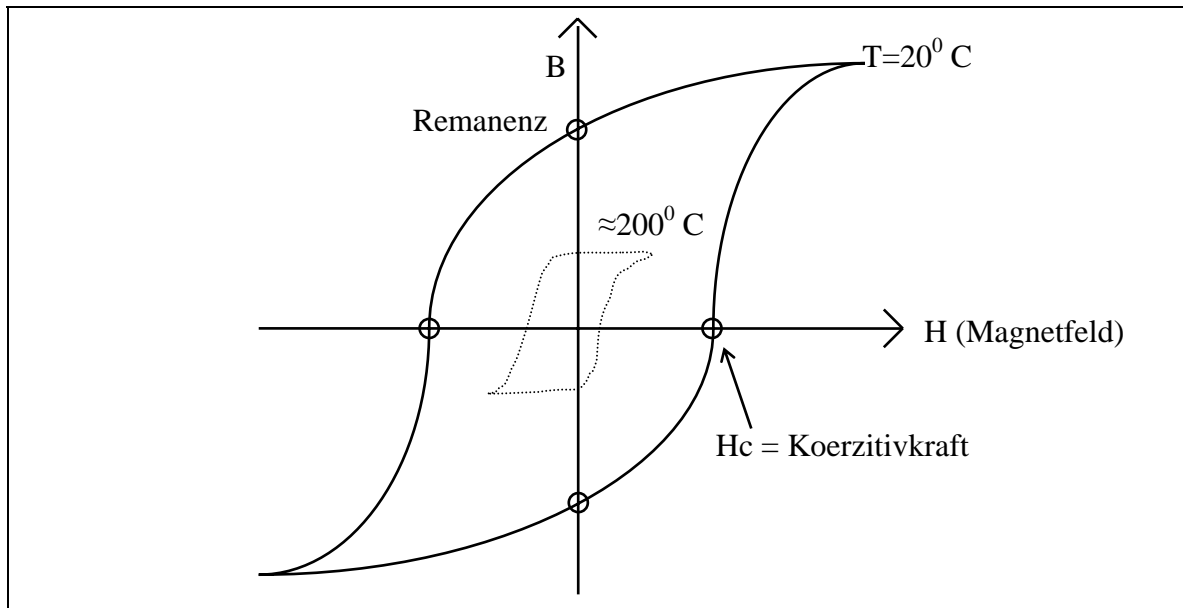
Polarisationsebene ( $\perp$  oder  $\text{—}$ ). Die Intensität des reflektierten Lichts einer Polarisationsebene hängt damit von der Magnetisierungsrichtung der Datenoberfläche ab und liefert optische Flusswechsel die von der Leseelektronik als Bit="1" interpretiert werden.

Datenstabilität: > 10 Jahre (40 Jahre)



Frage: Lesen 40-fach, erstmaliges schreiben 8-fach, ändern 4-fach ?

**Vergleich der magnetischen Eigenschaften Magnetplatte (Diskette) ↔ optische MO-Platte:**



Die Koerzitivkraft  $H_c$  ist ein Maß für die Beeinflußbarkeit der Magnetisierung durch externe Magnetfelder:

HD-Diskette:  $H_c \approx 635$  Oersted

Magneto-optische Platte:  $H_c \approx 8000$  Oersted

Die magneto-optische Platte ist wesentlich unempfindlicher gegen magnetische Störfelder als eine herkömmliche magnetische Diskette.

Einfluß der Verschmutzung: Magnetplatte (Diskette) ↔ optische Platte:

Wegen der transparenten Schicht von ca. 1mm Dicke wirkt sich eine Störung der Oberfläche bei der optischen Platte nicht so stark aus, da an der Oberfläche der Strahldurchmesser noch ca. 1mm groß ist.

### 2.3.2 Plattenrotation

Die Umdrehungszahl (U/min) ist entweder konstant (wie bei der Magnetplatte) oder sie wird in Abhängigkeit von der Kopf-Position nachgeregelt.

a) **CLV** (Constant Linear Velocity):

Lesen mit konstanter Lineargeschwindigkeit (= Geschwindigkeit des Datenträgers gegenüber dem Lesekopf. Die Datenrate beim Lesen ist damit konstant (konstante Bitdichte auf allen Spuren vorausgesetzt). Die Umdrehungszahl ist bei CLV abhängig von der momentanen Position des Lesekopfes. Diese Betriebsart kommt hauptsächlich bei Audio-CDs vor.

Beispiel: 540 - 210 U/min (innen/außen)

**ZCLV** (Zonen-CLV):

Der Datenträger ist in Zonen gleicher Lineargeschwindigkeit eingeteilt.

b) **CAV** (Constant Angular Velocity):

Lesen mit konstanter Winkelgeschwindigkeit (= konstante Umdrehungszahl). Nachdem die äußeren Spuren länger sind, ist dort die Datenrate am höchsten (konstante Bitdichte auf allen Spuren vorausgesetzt). Auf den inneren Spuren ist sie am niedrigsten. Diese Betriebsart kommt hauptsächlich bei Daten-CDs vor.

Laufwerke mit CAV verringern u.U. ihre Winkelgeschwindigkeit wenn starke Vibrationen auftreten (Vibrationssensoren) und/oder im Falle von auftretenden Lesefehlern.

Beispiel: 2,7 - 6,4 MByte/s (innen/außen), Mittelwert = 4,5 MByte/s (40-fach Laufwerk, Messwerte)

Partial CAV: Im inneren Spurbereich wird mit CAV (Datenrate variabel) gelesen, außen mit CLV (Datenrate=konstant).

DVD	HD DVD	BD
CLV/ CAV	ROM/R: CLV RW: ZCLV (Zonen-CLV)	CLV

In der Regel (z.B. bei CD-ROM) werden die Daten beginnend mit der innersten Spur aufgezeichnet. Ist eine CD nur zum Teil mit Daten gefüllt (was häufig zutrifft) und wird mit CAV gelesen, dann wird hauptsächlich nur mit der niedrigeren Datenrate gelesen!

Maßzahlen für die Datenrate:

<b>CD:</b> $R=n \cdot x_{CD}$ ,	$x_{CD} = 150 \text{ KByte/s}$ ,	CD-ROM, 40-fach = 6 MByte/s
<b>DVD:</b> $R=n \cdot x_{DVD}$ ,	$x_{DVD} = 1,35 \text{ MByte/s}$ ,	DVD-ROM, 10-fach = 13,5 MByte /s.
<b>HD DVD, BD:</b>	$x_{HD \text{ DVD}} = 4,5 \text{ MByte/s}$ ,	HD DVD-ROM, 4-fach = 18 MByte /s.

Aufgabe: Vergleich der Datenraten von CAV- und CLV-Gerät

### 2.3.3 Datenträgerformat

Die Ausführungen gelten für DVD, HD DVD und Blu-ray Disk (Multi-Session)

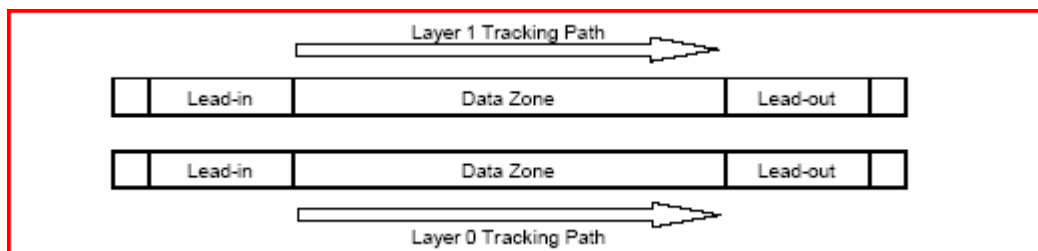
Es sind 4 Strukturen möglich:

- Single sided, single layer
- Single sided, dual layer
- Two sided, single layer,
- Two sided, dual layer.

Die Daten sind auf spiralförmigen Spuren aufgezeichnet. Die Aufzeichnungsrichtung auf der ersten Schicht (Layer 0) ist von innen nach aussen. Bei 2 Aufzeichnungsschichten unterscheidet man zwischen Parallel-Spur- Verfahren (parallel-track-path, **PTP**) und Gegen-Spur-Verfahren (opposite-track-path, **OTP**).

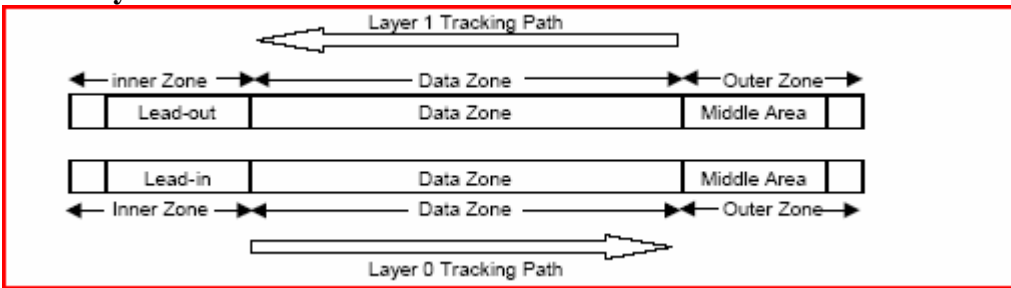
#### Dual Layer PTP Structure:

Die 2 Schichten sind voneinander unabhängige Informationszonen, jede mit einer Single Layer Struktur.



Es sind 2 getrennte Adressräume für Layer 1 und Layer 2..

**Dual Layer OTP Structure:**

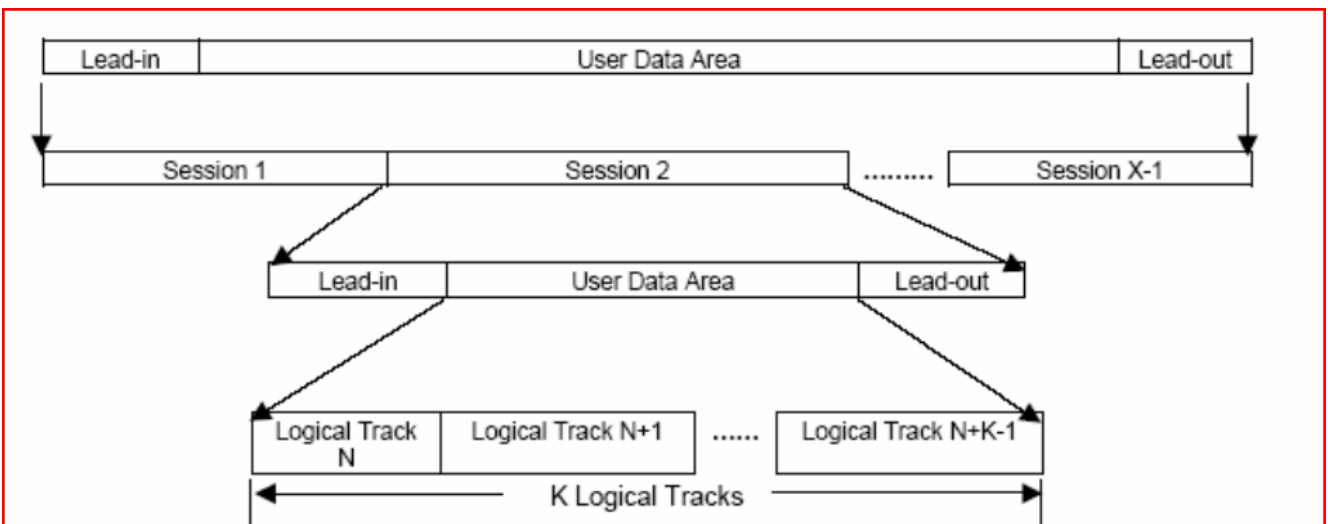


Die Adressierung der Sektoren innerhalb einer Zone (mit n Sektoren) beginnt in Layer 0 mit Sektor# = 0 und endet in Layer 1 mit Sektor# = n-1.

Mit dem SCSI-Befehl SEND DISC STRUCTURE kann die Lage des Umschaltbereichs (Middle Area) festgelegt werden.

In der Regel kommt die OTP-Struktur zur Anwendung.

**Unterteilung des Speicherraums einer Disk:**



Information über die Disk Struktur: abrufbar über SCSI-Befehl READ DISC INFORMATION.

Information über die Session Struktur: abrufbar über SCSI-Befehl READ TRACK INFORMATION.

**Sessions:**

Der gesamte Speicherraum ist in Sessions unterteilt, beginnend mit Session 1. Eine Session besteht aus mehreren logischen Spuren.

Session = Daten + Inhaltsverzeichnis (Im Lead In)

Eine Disk ist „geschlossen“ (finalized), wenn die letzte Session geschlossen ist (und keine neue Session angehängt werden kann).

**Logische Spuren (Logical Tracks):**

Eine logische Spur besteht aus mehreren Untereinheiten (Blöcken, Sektoren). Die Bezeichnungen für eine logische Spur sind je nach Media-Typ (Disk-Typ) unterschiedlich:

MM Disc Type	Media Specific Name
CD-ROM/-R/-RW	Track
DVD-R/-RW, DL DVD-R, HD DVD-R	RZone
DVD+R, DL DVD+R	Fragment
BD-R	SRR

MM Disc = MultiMedia Disk, DL = Dual Layer

SRR = sequential recording range, RZone = recording zone

BD-R = Blu-ray Disk Recordable

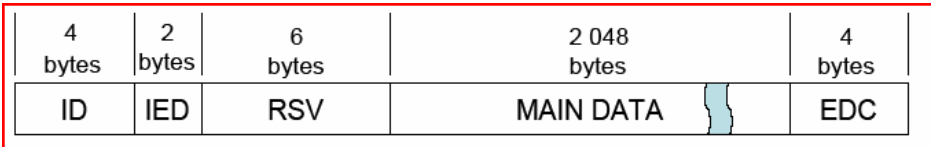
Die Beginnadresse einer logischen Spur ist die logische Blockadresse des ersten Blocks (Sektors) in dieser logischen Spur.

**Unterteilung einer logischen Spur (Track) in Blöcke am Beispiel der HD DVD Disk:**

Jede logische Spur besteht aus einer Folge von ECC Blocks. Jeder ECC Block enthält 16 (32 ?) Sektoren mit je 2048 Byte Datenfeld, d.h., 64 kByte Anwenderdaten und zusätzlich 5 kByte ECC-Code. Die Sektoren sind durchnummeriert mit einer 24-Bit-Adresse (Physical Sector Number, PSN). Die ECC Blöcke sind sequentiell angeordnet. Um einen einzelnen Sektor zu lesen muss der ganze ECC Block zuerst gelesen werden, wobei ggf. über den ganzen ECC-Block ein Fehlerkorrekturverfahren angewandt wird.

**Datensektor:**

Ein HD DVD Datensektor enthält 2064 Bytes, davon 2048 Bytes Datenfeld und 16 Bytes Formatinformation.



EDC 4 Bytes Error Detection Code (EDC) für den ganzen Sektor.

IED ID Error Detection (Redundanz über ID-Feld)

RSV Reserve

ID **Bits (31 – 24): Sektorinformation → Tabelle,**

**Bits (23 – 0): physikalische Sektornummer (Adresse),**

**Tabelle Sektorinformation:**

Sector Information ID bits 31 through 24							
31	30	29	28	27	26	25	24
Reserved				Area Type		Data Type	Layer Number
Area Type	00b	When the sector is in the Data Area					
	01b	When the sector is in the System Lead-in Area or Data Lead-in Area					
	10b	When the sector is in the Data Lead-out Area or System Lead-out Area					
	11b	When the sector is in the Middle Area					
Data Type	0b	indicating read-only data					
	1b	Reserved					
Layer Number	0b	When the sector is in the layer 0					
	1b	When the sector is in the layer 1					

**ECC Block Struktur bei HD DVD:**

Ein 2064 Byte Sektor ist in 6 Reihen zu 344 Bytes aufgeteilt. 16 (32 ?) sequentiell angeordnete Sektoren sind in eine Struktur von 208 Reihen zu 182 Bytes angefügt. Error Correction Redundanz sind entsprechend angeordnet, siehe folgende Tabelle.

		User Data					ECC Parity on Rows				
User Data	B0,0	B0,1	B0,2	...	B0,171	C0,0	C0,1	C0,2	...	C0,9	
	B1,0	B1,1	B1,2	...	B1,171	C1,0	C1,1	C1,2	...	C1,9	
	B2,0	B2,1	B2,2	...	B2,171	C2,0	C2,1	C2,2	...	C2,9	
	B3,0	B3,1	B3,2	...	B3,171	C3,0	C3,1	C3,2	...	C3,9	
	...	...	...	...	...	...	...	...	...	...	
	B190,0	B190,1	B190,2	...	B190,171	C190,0	C190,1	C190,2	...	C190,9	
B191,0	B191,1	B191,2	...	B191,171	C191,0	C191,1	C191,2	...	C191,9		
ECC Parity	R0,0	R0,1	R0,2	...	R0,171	C192,0	C192,1	C192,2	...	C192,9	
	R1,0	R1,1	R1,2	...	R1,171	C193,0	C193,1	C193,2	...	C193,9	
	R2,0	R2,1	R2,2	...	R2,171	C194,0	C194,1	C194,2	...	C194,9	
	...	...	...	...	...	...	...	...	...	...	
	R15,0	R15,1	R15,2	...	R15,171	C207,0	C207,1	C207,2	...	C207,9	

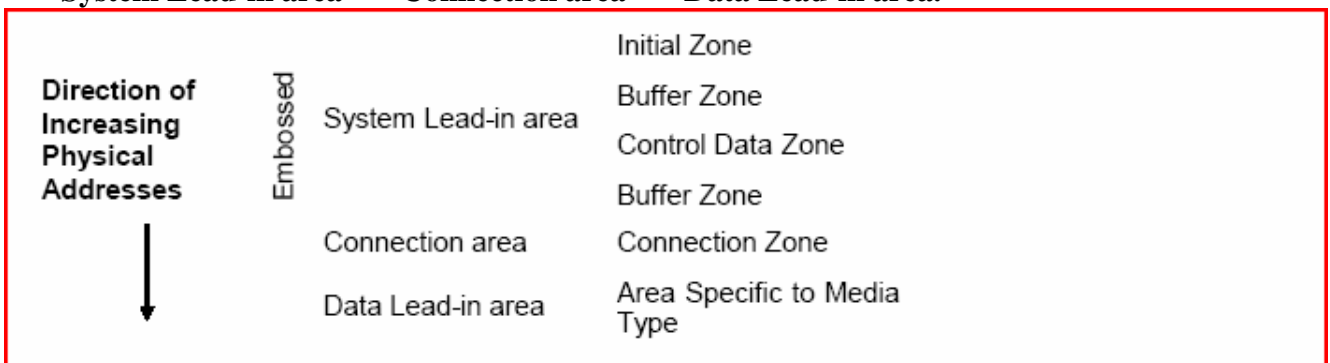
R = Row Parity (Reihe), C = Column Parity (Spalte)

**Lead-In/Out Area**

Lead-In und Lead-Out Area sind Disk-spezifisch, sie enthalten u.a. das Inhaltsverzeichnis bzw. einen Verweis auf das nächstfolgende (aktuelle) Inhaltsverzeichnis.

Der Lead-In Bereich besteht aus 3 Teilen:

- System Lead-in area
- Connection area
- Data Lead-in area.



**Control Data Zone:** The Control data zone is a structure recorded in the lead-in area of a DVD and HD DVD discs that contains information concerning structure of the disc. The DVD Control Data Zone is comprised of 192 ECC blocks in the System Lead-in Area of a DVD disc. The content of the 16 sectors in each ECC block is repeated 192 times. The HD DVD Control Data Zone is comprised of 192 ECC blocks in the System Lead-in Area of a HD DVD disc. The content of the 32 sectors in each Block is repeated 192 times.

**HD DVD:**

Die HD DVD Disk wurde definiert durch das DVD Forum: HD DVD-ROM (Read Only), HD DVD-R (write-once Recordable) and HD DVD-RW (Rewritable). Das HD DVD-Format basiert auf dem gegenwärtigen DVD-Format.

HD DVD RAM: Sektoren fest vorgeprägt, Disk in 18 Zonen aufgeteilt (wie DVD RAM).

### **Blu-ray Disc BD:**

Ein Datensektor enthält 2066 Bytes, davon 2048 Bytes Datenfeld und 18 Bytes Formatinformation. 32 Sektoren bilden einen Cluster. Für die Cluster existiert eine Belegungstabelle (Space Bitmap), die angibt, ob eine Cluster belegt oder frei ist.

Dual layer BD is defined only for OTP discs.

#### **{Definitions**

##### 4.5.5.4.2.2.1 Logical Blocks

A Logical Block is the smallest logically addressable unit of data that is readable by the Host. For BD-R, the Logical Block size is 2 048 bytes. This value is specified in the Logical Block Size field in the Random Readable Feature Descriptor.

##### 4.5.5.4.2.2.2 Recordable Units

A recordable unit is the smallest physically writable collection of contiguous Logical Blocks. For BD-R the recordable unit size is 32 Logical Blocks, one Cluster. This value is specified in the Blocking field of the Random Writable Feature Descriptor.

##### 4.5.5.4.2.2.3 Logical Track: Sequential Recording Range (SRR)

A Logical Track is a set of sequential recordable units. Logical Tracks are numbered consecutively, starting with number one. On BD-R, the Logical Track is defined as a Sequential Recording Range (SRR). The physical extents of each SRR and the status of each SRR is defined in the SRR Information (SRRI) structure. The SRRI is a structure in the TDMS that is maintained in the Temporary Disc Management Areas (TDMA).

BD-R Logical Tracks bounds are defined only in the SRRI. No overhead blocks are used in the definition of a Logical Track on BD-R. }

### 2.3.4 Sektor (Block)-Format CD-ROM

Literatur [1]

Die Information ist in Blöcken (Sektoren) auf einer spiralförmigen Datenspur mit ca. 20 000 Windungen abgespeichert, am inneren Ende der Spirale beginnend.

#### a) Sektorformat bei der Audio-CD (=CD-DA), (Sektor = 1/75 Sekunde):

<b>User Data (Datenfeld)</b> <b>2352 Byte</b>	<b>Q-Parität</b> 392Byte Level 2	<b>P-Parität</b> 392Byte Level 1	<b>Control</b> 98 Byte
--	--	--	---------------------------

Die Informationen für die Unterkanäle (subchannels), wie z.B. Zeit-Code, Musiktrack (=Musikstück) etc. sind in den 98 Control-Bytes des Sektors enthalten.

Die 2352 Byte Information des Datenfelds (Sektors) sind in 98 Rahmen (Frames) aufgeteilt und in codierter Form abgespeichert. Ein Rahmen umfasst 33 Byte und enthält 24 Byte "Nutzdaten".

**detailliert:**

#### a1) Die 33 Datenbytes eines Rahmens enthalten:

33 Datenbytes=24 Byte "Nutz"-Daten+4 Byte Q-Parität+4 Byte P-Parität+1 Kontroll-Byte (entspricht C1- und C2-error correction, Red Book)

**Nutzdaten pro Sektor (uncodiert): 2352 Byte = 24 Byte/Rahmen x 98 Rahmen**

**Daten insgesamt pro Sektor (codiert): 3234 Byte = 33 Byte/Rahmen x 98 Rahmen**

#### a2) Codierung:

Jedes Byte wird in 14 Bit codiert (= sogenannte Kanal-Bits) entsprechend einem RLL 10,2-Code (2..10 x '0' zwischen benachbarten '1', siehe Kap. Magnetplatte), sodaß auf der Spur die maximale bzw. minimale Pit/Land-Länge 11 bzw. 3 Bitpositionen beträgt. Diese Rahmen benötigen codiert 588 Bit; sie werden auch EFM-Frames (Eight-to-Fourteen Modulation) genannt.

1 Sektor (= 2352 Bytes)=98 EFM-Frames

→ **Red Book, level 0**

1 EFM-Frame (Rahmen) = 588 Bit

= 24 Sync-Bits+14x33 Daten-Bits (=33 Daten-Bytes)+3x33 Merging-Bits+3 Ende-Bits

(Erläuterung:33 Datenbytes, codiert mit 14 Bit pro Datenbyte, 3 Merging-Bits pro Datenbyte, zusätzlich erforderlich für die Codierung)

**Fehlerrate:**  $10^{-5}$  ohne Redundanz,  $10^{-11}$  mit Q- und P-Parität (Red Book)

#### b) Sektorformat bei der Daten-CD (=CD-ROM):

Die Sektoren müssen direkt adressierbar sein. Deshalb besitzt jeder Sektor ein Adressfeld (Header) Der Sektor ist die kleinste unabhängig adressierbare Einheit auf einer CD.

Sektor-Adresse im Header = Minute/Sekunde/Block-Nr. (Sektor-Nr)

1 Minute = 60 Sekunden (0..59)

1 Sekunde = 75 Blöcke (0..74)

Durch Einfügen von redundanter EDC (error detection code) – und ECC (error correction code) – Information im Sektor wird die Datensicherheit gegenüber der Audio-CD wesentlich gesteigert. Die Fehlerrate beträgt  $10^{-15}$  (Yellow Book). Die Sektoren werden wie bei der Audio-CD codiert in 98 Rahmen abgespeichert.

**Sektorformat CD-ROM für Computer-Daten, ohne Rahmen-Formatbytes:**

**Mode 1, nur "Nutz"-Daten dargestellt** (Norm: Yellow Book, Datenkapazität der CD-Rom= 552 MByte)

<b>Sync</b> 12 Byte	<b>Header</b> 4 Byte Min/Sek/ Block-Nr <b>Mode = 1</b>	<b>User Data (Datenfeld)</b> 2048 Byte	<b>EDC</b> 4 Byte Error Detection Code	<b>Zero</b> 8 Byte	<b>ECC</b> 276 Byte Error Correctio n Code
← 2352 Byte →					

Mode = 2: kein EDC,kein ECC (→ Datenkapazität = 630 MByte)

Mode = 0: kein EDC,kein ECC, keine Daten (Markierungsblock für Datei-Format)

**Vergrößertes Sektorformat CD-ROM mit Rahmen-Formatbytes (=Mode 1):**

<b>Sync</b> 12 Byte	<b>Header</b> 4 Byte	<b>User Data (Datenfeld)</b> 2048 Byte	<b>EDC</b> 4 Byte  Level 3	<b>Zero</b> 8 Byte	<b>ECC</b> 276 Byte  Level 3	<b>Q-Parität</b> 392Byte  Level 2	<b>P-Parität</b> 392Byte  Level 1	<b>Control</b> 98 Byte
← Rahmenformat →								

**Die 33 Datenbytes eines Rahmens enthalten:**

33 Datenbytes=24 Byte "Nutz"-Daten+4 Byte Q-Parität+4 Byte P-Parität+1 Kontroll-Byte (entspricht C1- und C2-error correction, Red Book)

"Nutz"-Daten sind alle Daten des obigen Sektors:

12 Sync-Bytes + 4 Header-Bytes + User-Daten + 4 Byte EDC + 8 Byte Zero + 276 Byte ECC (entspricht C3-error corr. Yellow Book).

Bei der Audio-CD entfallen die EDC- und ECC-Bytes, da keine so hohe Datensicherheit notwendig ist wie bei einer Daten-CD. Das Datenfeld und die Kapazität werden dementsprechend größer.

**c) Sektorformat für Multimedia-Anwendungen:**

Das CD-ROM/XA Sektorformat (XA = extended Architecture) erlaubt einen Mixed Mode von Computer-Daten und Audio/Video-Daten. Dieser Mixed Mode ist notwendig, um ein Verschachteln (interleaving) von Computerdaten-Sektoren und von Audio/Video-Sektoren bei Multimedia-Anwendungen zu ermöglichen. Es verwendet "Mode2 Form1-Sektoren" für Computerdaten und "Mode2 Form2-Sektoren" für Audio/Video-Daten.

**Sektorformat CD-ROM/XA für Computerdaten:**

**Mode 2 Form 1** (Green Book)

<b>Sync</b> 12 Byte	<b>Header</b> 4 Byte Min/Sek/ Block-Nr Mode	<b>Sub Header</b> 8 Byte	<b>User Data (Datenfeld)</b> 2048 Byte	<b>EDC</b> 4 Byte	<b>ECC</b> 276 Byte
← 2352 Byte →					

Wie bei der CD-ROM wird der Sektor in 98 Rahmen (Frames) aufgeteilt und in codierter Form abgespeichert.

**Sektorformat CD-ROM/XA für Audio- und Video-Daten:**

**Mode 2 Form 2 (Green Book)**

<b>Sync</b> 12 Byte	<b>Header</b> 4 Byte 1 Min/1 Sek/ 1 Block-Nr 1 Mode	<b>Sub Header</b> 8 Byte	<b>User Data (Datenfeld)</b> <b>2324 Byte</b>	<b>EDC</b> 4 Byte
------------------------	---	-----------------------------	--	----------------------

← 2352 Byte →

Wie bei der CD-ROM wird der Sektor in 98 Rahmen (Frames) aufgeteilt und in codierter Form abgespeichert.

*Frage: Unterschied der beiden obigen Sektorformate?*

**Subheader:**

Byte 1 (5): Datei-Nummer (0 bis 255). Sektoren verschiedener Dateien können verschachtelt sein.

Byte 2 (6): Kanal-Nummer. Für verschiedene Audio-, Video- und Datenkanäle,

Beispiel: Video mit 4 Sprachen

Channel 0 = Video-Sektoren, Channel 1 bis 4 sind Sprach-Sektoren (4 verschiedene Sprachen). Video Recorder wählt einen der Sprach-Kanäle aus.

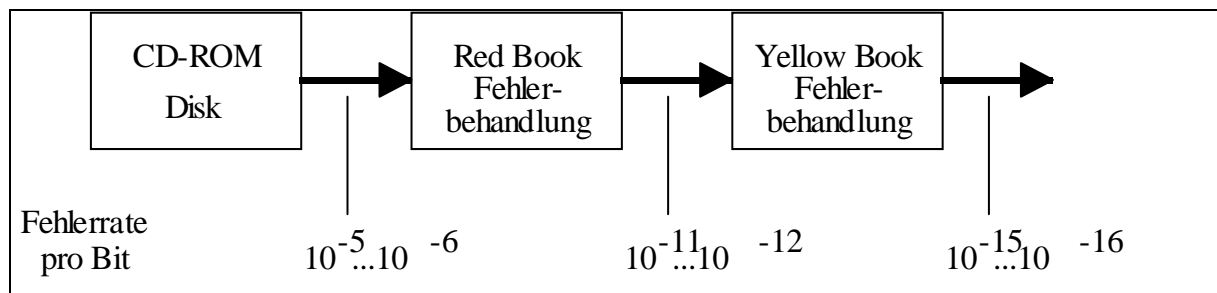
Byte 3 (7): Submode Bits: Letzter Sektor einer Datei (EOF)  
Realzeit-Sektor (wird bei Fehler nicht korrigiert)  
Datensektor oder  
Audio-Sektor oder  
Video-Sektor  
Letzter Sektor eines Datensatzes (EOR),  
etc.

Byte 4 (8): Kodierungsart bei Audio- und Videoinformationen (z.B. Mono- oder Stereo-Mode)

**d) Fehlerkorrektur:**

Durch die hohe Bitdichte haben CD-ROMs eine hohe Fehlerrate ( $10^{-5}$  bis  $10^{-6}$

Fehlerwahrscheinlichkeit pro Bit). Durch viel Redundanz (EDC error detecting code und ECC error correcting code) wird die Fehlerrate reduziert.



### 2.3.5 Verwaltung des Datenträgers

#### a) Datei-Systeme

##### **Dateisystem ISO9660:**

Ziel war ein möglichst systemunabhängiges (DOS, UNIX, WIN usw.) und damit kompatibles Dateisystem. Dabei beschränkte man sich auf den kleinsten gemeinsamen Nenner. CDs mit ISO9660 als Dateisystem sind auf allen Betriebssystemen lesbar. Es erlaubt Dateinamen mit 8+3 Zeichen und Verzeichnisnamen mit 8 Zeichen. Dabei dürfen nur die Zeichen A-Z, 0-9 und der Underscore ( \_ ) verwendet werden. Die maximale Schachtelungstiefe von Verzeichnissen ist auf 8 Ebenen (inkl. Root) beschränkt.

(Ergänzung ECMA 168 "Frankfurt Proposal" ??, ECMA168 Redirector Software für Sector by Sector incremental Write).

##### **Joliet:**

Dies ist eine Erweiterung von ISO9660. Datei- und Verzeichnisnamen dürfen bei "Joliet" bis zu 64 Stellen lang sein, es gibt keine Einschränkungen in der Verzeichnistiefe, der Zeichensatz ist erweitert usw. (Die ISO 9660 Verzeichnisstruktur wird nach den bekannten festgelegten Regeln erstellt. Wo keine Jolietvorschriften greifen, gilt der ISO-9660-Standard auch für Joliet).

##### **UDF (Universal Disk Format/File System):**

Erweitertes Dateisystem, das es erlaubt, Dateien in einzelnen kleinen Paketen zu schreiben ("packet writing"). Damit lassen sich CD-RW- und CD-R-Medien aus jedem Programm heraus wie Festplatten ansprechen. Darin besteht der Unterschied der Multisession-Methode zum UDF-Format. Egal, wie voll der mit UDF beschriebene Rohling ist, er erlaubt fortwährend Einsicht in die gespeicherten Daten. Der Rohling wird vor der Nutzung formatiert.

**HFS (Hierarchical File Structure):** für den MAC-PC.

#### b) Single- und Multi-Session

Stehen genügend Fotos, Scans oder Daten bereit, um die CD komplett zu füllen, empfiehlt sich das Single-Session Verfahren. Dabei wird zunächst ein Lead-In, eine Einleitung geschrieben, dann kommen die Daten und schließlich das Lead-Out. Zusätzlich wird in den am Anfang freigelassenen Bereich das TOC (Table of Content = Inhaltsverzeichnis) geschrieben. Es enthält am Ende einen Abschlussvermerk. Auch wenn auf der CD noch Speicherplatz vorhanden ist, können Daten nicht mehr hinzugefügt werden.

Vorteil des Single-Session-Verfahrens ist es, dass der Platz auf der CD optimal genutzt wird.

Session	Lead-In	Einleitung
	TOC	Inhaltsverzeichnis
	Tracks	Daten, max. 99 Tracks
	Lead-Out	Session und CD geschlossen

##### **Multi-Session (Track at Once):**

Dieses Verfahren erlaubt, daß ein CD-Recorder die Datenspirale in mehreren Arbeitssitzungen (Sessions) stückweise beschreiben kann. Jedes Spiralstück, das in einer separaten Session beschrieben wird, enthält jeweils folgendes Session-Format:

-- "**Lead-In**"-Bitmuster (Start) mit **TOC** (Table of Content = Inhaltsverzeichnis), ca. 10MByte  
-- **Daten** und

--"Lead-Out"-Bitmuster (Endekennung), 4 MByte

Die TOCs der verschiedenen Sessions sind miteinander verkettet, sodaß der Anwender nicht merkt, daß der Datenbestand aus mehreren Bruchstücken mit Verzeichnissen an unterschiedlichen Stellen besteht. Die zuletzt geschriebene TOC ist immer eine Aktualisierung ("update") der vorher geschriebenen. Das aktuelle Inhaltsverzeichnis steht immer in der letzten Session. Man hat immer Einblick in das Inhaltsverzeichnis aber keinen Zugriff auf die Daten, bis die Multisession-CD abgeschlossen wird. Zwischen den einzelnen Brennvorgängen einer Multisession können die Daten nicht gelesen werden.

Der Vorteil der Multi-Session-Methode ist, daß man auf dem nur einmal beschreibbaren und nicht löschbaren Medium CD-R Dateien wie auf einer beliebig wiederbeschreibbaren Wechsellplatte erweitern kann. Für jede Session fallen jedoch ca. 15 MByte Verwaltungsinformationen auf der CD an.

Für das tägliche Backup kleinerer Datenmengen steht das sog. "Incremental Packet Writing"-Verfahren zur Verfügung, das mit wesentlich weniger zusätzlicher Verwaltungsinformation auskommt. Der Brenner schreibt die Daten in kleinen Paketen auf den Rohling. Nach jedem Paket lässt sich der Schreibvorgang unterbrechen, wobei das Schreiben auf den Rohling jederzeit wieder fortgesetzt werden kann. Ein sogenannter "buffer underrun" kann nicht auftreten.

Session 1	Lead-In	Einleitung
	TOC	Inhaltsverzeichnis
	Tracks	Daten, max. 99 Tracks
	Lead-Out	Session geschlossen, Verweis auf das nächste TOC
Session 2	Lead-In	Einleitung
	TOC	Inhaltsverzeichnis
	Tracks	Daten
	Lead-Out	Session geschlossen, Verweis auf das nächste TOC
Session n	Lead-In	Einleitung
	TOC	aktuelles Inhaltsverzeichnis
	Tracks	Daten
	Lead-Out	Session geschlossen, CD geschlossen

**Datenträger-Belegung:**

innen

außen

PCA	PMA	Lead-In 1 TOC 1	Recorded Area 1 (Datensektoren)	Lead-Out 1	Lead-In 2 TOC 2	Recorded Area 2 (Datensektoren)	Lead-Out 2

**PCA** (program calibration area): Laser-Sender (Energie) wird calibriert für Schreiben

**PMA** (program memory area): Track-Nummern der Titel mit Start- und Stop-Adresse. Wird verwendet, solange die CD noch nicht voll ist (TOC noch nicht geschrieben).

**Lead-In:** Inhaltsverzeichnis TOC wird erst geschrieben, wenn die Recorded Area voll ist. Es wird dann anstelle des PMA von der Software verwendet.

**Lead-Out:** Kennzeichnung des Endes des Datenbereichs

Nach dem letzten Track (= Datei bei einer Daten-CD, Musiktitel bei einer Audio-CD) wird die CD fixiert (SCSI Command "Fixation"), d.h., Lead-In mit TOC (Start/Stop-Adressen aller Tracks) und Lead-Out werden geschrieben.

*Frage: Wieso kann man bei der nur 1x beschreibbaren CD-R das Inhaltsverzeichnis aktualisieren?*

**Adreßvolumen einer CD:**

	<b>Adresse</b>	
Erster Block (=innerster Block):	00min/00s/Block 00	PMA Inhaltsverzeichnis ???
Beginn Anwenderbereich:	00min/02s/Block 00	<b>Lead-In Session 1</b>
	bis 00min/02s/Block 15	
	00min/02s/Block 16	PVD primary volume descriptor
		PT Pfadtabelle: enthält Directories
		Root Directory
		Directory A
		Dateien von Directory A
		Directory B
		Dateien von Directory B
		<b>Lead-Out Session 1</b>
		<b>Lead-In Session 2</b>
		usw.
		<b>Lead-Out Session 2</b>
Ende Anwenderbereich:	60min/01s/Block 74	
Letzter Block:	60min/03s/Block 74	

PVD primary volume descriptor: enthält Adresse des Hauptverzeichnisses

### 2.3.6 Befehlsschnittstelle CD-R

SCSI-Laufwerk (Philips CD-R CDD522)

**Messages:**

Code(hex)	Message	in=Initiator←Target out=Initiator→Target
00	command complete	in
02	save data pointer	in
04	disconnect	in
06	abort	out
07	message reject	in/out
08	no operation	out
0C	bus device reset	out
80	identify	in/out

**Group 0 Commands:**

Opcode (hex)	Command
00	test unit ready
01	reset(rezero) unit
03	request sense
08	read
0A	write
0B	seek
12	inquiry
15	mode select
16	reserve
17	release
18	copy
1A	mode sense
1B	start/stop unit
1D	send diagnostics
1 <sup>E</sup>	prevent/allow medium removal

**Group 1 & 2 Commands (10 Bytes):**

Code (hex)	Command
25	read capacity
28	read
2A	write
2B	seek
2F	verify
35	flush cache
42	read sub channel
43	read disc info

**Vendor Unique commands:**

Code (hex)	Command
E2	first writable address
E4	reserve track
E5	read track info
E6	write track
E7	medium load/unload
E9	fixation (write lead-in and lead-out)
EC	recover
EB	send absorption control errors

### 2.3.7 Normen

Die Block(Sektor-)Formate der Compact Disks wurden in sogenannten "bunten Büchern" niedergelegt.

<b>Plattentyp</b>	<b>Norm-Schrift</b>
CD Audio = CD-DA (Compact Disk Digital Audio)	<b>Red Book</b>
CD-ROM CD-ROM/XA ISO9660, Single Session Disk	<b>Yellow Book</b>
CD-I	<b>Green Book</b>
CD-R CD-WO CD-MO CD-RW Multi Session Disk Packet writing	<b>Orange Book</b>
Video-CD	<b>White Book</b>

## 2.4 Aufgaben

### Aufgabe PD-1: USB-Stick, Vergleich mit Festplatte

Festplatte:

Positionierzeiten:  $t_p=10ms$ ,  $t_{Zyl/Zyl} \approx 0ms$ ,  $t_f=5ms$ ,  $R=21\text{ MByte/s}$

USB-Stick (USB 2.0):  $R_{lesen}=7\text{ MByte/s}$ ,  $R_{schreiben}=2\text{ MByte/s}$

Bei welchen Dateigrößen ist der USB-Stick schneller?

- a) Lesen einer sequentiellen Datei (Annahme: Positionier- und Drehwartezeit nur beim Transferbeginn)
- b) Schreiben einer sequentiellen Datei (Annahme wie bei a))
- c) Lesen einer fragmentierten Datei (Clustergröße=16kByte, 10 Cluster pro Fragment)

### Aufgabe PD-2: optische Speicher, Technologiefortschritt

1. Den Technologie-Fortschritt kann man in 3 Schritten beschreiben.

Ergänzen Sie die Tabelle:

alt	derzeit	zukünftig
CD		a) b)
Abkürzung = compact disk		a) b)
Kapazität = 0,5 GByte		a) b)

2. CLV und CAV kennzeichnen zwei Zugriffsverfahren bei den optischen Speichern.

- a) Ergänzen Sie
  - Abkürzung auf Englisch
  - technische Erklärung
- b) Bei welchem Verfahren ist die Datenrate beim Lesen konstant?

### Aufgabe PD-3: CD-ROM, Vergleich der Datenraten bei CLV und CAV

Vergleich zweier CD-ROM Laufwerke:

Auf der Oberfläche sind 20 000 Windungen (=Spuren).

Die Windung(Spur) 0 liegt innen. Eine Datei ist beginnend bei Windung 0 abgespeichert.

Laufwerk 1:

12-fach CLV (Lesen mit konstanter Lineargeschwindigkeit)

Datenrate = 1,8MByte/s

Laufwerk 2:

24-fach CAV (Lesen mit konstanter Winkelgeschwindigkeit)

Datenrate:

1,5 MByte/s - 3,6 MByte/s

(Die Datenrate steigt von der innersten zur äußersten Spur linear an).

1. Zeichnen Sie in ein x-y-Diagramm die Datenraten R als Funktion der Windungszahl.  
Kennzeichnen Sie die Bereiche  $R_{clv} > R_{cav}$  und  $R_{clv} < R_{cav}$ .
2. Berechnung: In welchem Spurbereich ist das 12-fach CLV-Gerät schneller (=höhere Datenrate) als das 24-fach CAV-Gerät ?

3. *Lesen einer Datei, beginnend von Spur-Nr. 0.*  
Bei welchen Dateikapazitäten (MByte) liefert das 12-fach CLV-Gerät kürzere Transferzeiten als das 24-fach CAV-Gerät ?  
*Gesamtkapazität eines Laufwerks = 552 MByte.*  
*Vereinfachende Annahme: Alle Spuren haben gleiche Spurkapazität*
4. *Die Annahme bei der vorherigen Aufgabe 2: "Alle Spuren haben gleiche Spurkapazität" führt zu einem ungenauen Ergebnis.*  
Wie würde die richtige Annahme lauten? Wie würde sich das Ergebnis bei realistischer Annahme verändern?

#### **Aufgabe PD-4: optische Speicher, physikalische Ebene**

*Bei den mehrfach beschreibbaren optischen Platten unterscheidet man zwischen CD-RW (DVD-RW) und MO-Disk. Erklären Sie die Aufzeichnungsverfahren bei beiden Plattentypen:*

##### **1. CD-RW (DVD-RW)**

- Name des physikalischen Aufzeichnungsverfahrens?
- Erklären Sie die wesentlichen Merkmale dieses Aufzeichnungsverfahrens.
- Wie viele Umdrehungen der Platte sind mindestens notwendig, um eine Spur zu ändern?

##### **2. MO-Disk**

- Name des physikalischen Aufzeichnungsverfahrens?
- Erklären Sie die wesentlichen Merkmale dieses Aufzeichnungsverfahrens.
- Wie viele Umdrehungen der Platte sind mindestens notwendig, um eine Spur zu ändern?





### **3 Portable Datenträger mit sequentieller Adressierung**

[www.tandberg.com/de/](http://www.tandberg.com/de/),

Sie umfassen die **Bandkassettengeräte** (QIC-Kassette, DAT-Kassette, DLT-Kassette, ..). Diese sind die Nachfolger der älteren **Magnetbandgeräte** (NRZI-, PE- und GCR-Geräte).

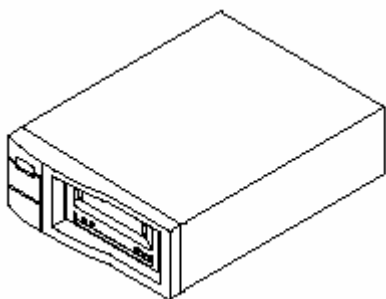
Bandkassetten werden heute hauptsächlich für die Datensicherung von Servern eingesetzt. Während optische Speicher auch mit DVD-Datenträgern noch zu teuer sind, um als günstiges Backup-Medium zu dienen, gewinnt die Datensicherung auf billigen Ultra-ATA- oder SATA-Festplatten immer mehr an Bedeutung (Wogegen bieten „räumlich entfernte“ Festplatten als Backup-Medium Sicherheit?).

Bandkassetten speichern Daten sequentiell auf Blockebene. Dementsprechend liegt ihre Stärke im Schreiben grosser Datenmengen in einem kontinuierlichen Datenstrom. Für den schnellen Zugriff auf bestimmte Dateien oder verstreut gespeicherte Informationen sind Bandkassetten hingegen denkbar ungeeignet.

Die **Magnetbandgeräte** haben - obwohl es Neuentwicklungen gibt - wegen der extremen Normung den technischen Anschluß hinsichtlich z.B. der Kapazität verpaßt. Auf Magnetbändern sind - je nach Aufzeichnungsverfahren - Datenmengen bis maximal 20 MByte (NRZI-), 40 MByte (PE-), 140 MByte (GCR-) und ca. 700 MByte (GCRXC-Aufzeichnungsverfahren) abspeicherbar, d.h., als „Back-up“-Speicher für GByte-Platten sind diese Kapazitätswerte zu klein. Dennoch werden Magnetbandgeräte an DV-Anlagen weiterhin eingesetzt.

(Die Normung war notwendig, weil die Magnetbänder früher die einzige Möglichkeit waren, größere Datenmengen zwischen DV-Anlagen auszutauschen. Deshalb mußte alles herstellerunabhängig bis „ins letzte“ genormt sein, was den technischen Fortschritt sehr hinderte, da die festgeschriebenen Werte für Bitdichte, Spurzahl usw. anscheinend nur zögerlich dem technischen Fortschritt folgten.)

**Bandkassettengeräte wie QIC-, DAT-, DLT- und LTO-Kassette** sind aktuelle Geräte der Magnetbandtechnik. Auf einer Magnetbandkassette sind - je nach Gerätetyp - derzeit bis zu 800 GByte Daten (z.B. LTO-Kassette) maximal abspeicherbar. Bandkassetten eignen sich daher als „Backup“-Speicher für GByte-Platten.



DAT-Kassetten-Laufwerk, 3 1/2 "  
40 GByte  
Helical Scan Datenaufzeichnung  
2,75 MByte/s  
SCSI-Schnittstelle

#### **Anwendung:**

- Archivierung von Daten (1x pro Jahr ist Wiederauffrischung = Lesen+Schreiben notwendig).
- Erstellen von Sicherungskopien („Backup“) von der Platte.
- Datenträger für Massendaten (Terabyte), z.B. Satelliten-Empfangsdaten

Konkurrierend bei obigen Anwendungen sind optische Platten wie DVDs, HD DVDs und BDs.

### **3.1 Bandkassetten-Geräte**

#### **3.1.1 1/4 -Zoll-Bandkassetten**

##### **Streamer-Kassette (QIC-Cartridge, QIC-Kassette)**

QIC=Quarter inch compatibility (1/4-Zoll Band Kompatibilität). Dies soll deutlich machen, daß die Formate abwärtskompatibel sind, d.h., in einem z.B. 1GByte-Laufwerk können auch Kassetten mit z.B. 120MByte beschrieben und gelesen werden.

Bandbreite = 1/4 Zoll  $\approx$  6.3mm

Kapazität = 60 MByte .. 5 GByte .. 25 GByte .., in Stufen genormt

Die Streamer-Kassette zeichnet die Daten serpentinartig linear auf, so dass die Spuren parallel zur Kante des Bandes liegen, siehe Bild. Es werden 1/4-Zoll-Magnetbänder in kompakten Kassetten beschrieben.

#### **3.1.2 1/2 -Zoll-Bandkassetten**

##### **a) DLT-Kassette (Digital Linear Tape)**

Die Daten werden wie beim QIC-Streamer serpentinartig aufgezeichnet, sodaß die Spuren parallel zur Bandkante verlaufen. Eine Dateneinheit (Rahmen) besteht z.B. aus 24 4KByte-Blöcken, die letzten 5 Blöcke sind ECC-Blöcke für die Fehler-Erkennung und -Korrektur.

Kapazitätsbereich (mit Datenkompression):

bis 70 (320) GByte und bis zu 32 MB/s Transferrate.

##### **b) LTO-Kassette (Linear Tape Open)**

Die LTO-Technologie wurde von Seagate, HP und IBM entwickelt. Aufgrund der breiten Herstellerunterstützung hat sich LTO sehr schnell am Markt durchgesetzt. Ähnliche Technologie wie DLT-Kassette mit vergleichbaren Kapazitäten und Datenraten.

#### **3.1.3 DAT-Kassette (Digital Audio Tape)**

Die DAT-Technologie ging aus dem Einsatz von digitalen Aufzeichnungsverfahren im Audio-Bereich hervor. Die Daten werden im Helical-Scan-Verfahren (Schrägspuraufzeichnung) auf das Band (8mm, 4mm) aufgezeichnet. Das Band wird schräg zur Kopftrommel geführt, was höhere Spurdichten erlaubt(?). Die Datenspuren sind schräg zur Laufrichtung des Bandes liegende Spurstücke.

Normierung: DDS-Format (Digital Data Storage) mit Datenkompression.

Die Formate unterscheiden sich in Bandlänge, Kapazität und Datenrate, z.B.

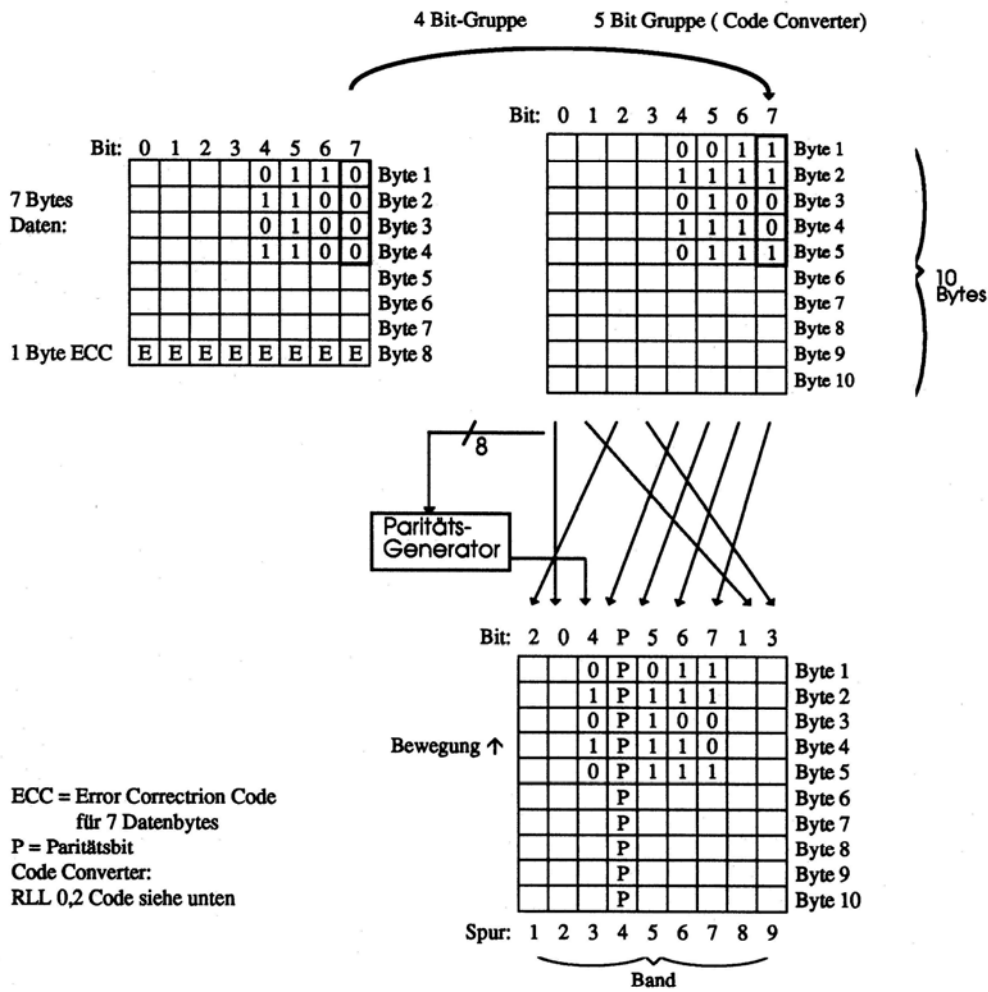
DDS-2 Format: bis 8 GByte, 1MByte/s (mit Datenkompression 2:1)

DDS-3 Format: bis 24 GByte, 2 MByte/s

DDS-4 Format: bis 40 GByte, 2 MByte/s, 72 GByte usw.



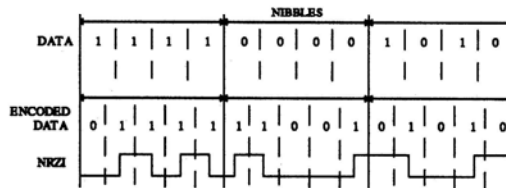
**GCR- Aufzeichnungsverfahren beim Magnetband**



ECC = Error Correction Code für 7 Datenbytes  
P = Paritätsbit  
Code Converter:  
RLL 0,2 Code siehe unten

**RLL 0,2-Code:**

DATA NIBBLE	RUN-LENGTH CODE	DATA NIBBLE	RUN LENGTH CODE
0000	11001	1000	11010
0001	11011	1001	01001
0010	10010	1010	01010
0011	10011	1011	01011
0100	11101	1100	11110
0101	10101	1101	01101
0110	10110	1110	01110
0111	10111	1111	01111



**Alte Verfahren:**

- NRZI (non return to zero change on ones): 1 = FW, 0 = kein FW, Querparität (ungerade) liefert Synchronisiersignal
- PE (phase encoded): 1 = positiver FW, 0 = negativer FW

### 3.2.3 Bandführung

Grundsätzlich unterscheidet man zwischen dem Start-Stop-Betrieb (alt) des Bandes und einem Betrieb mit kontinuierlicher Bandbewegung ("streaming", aktuell):

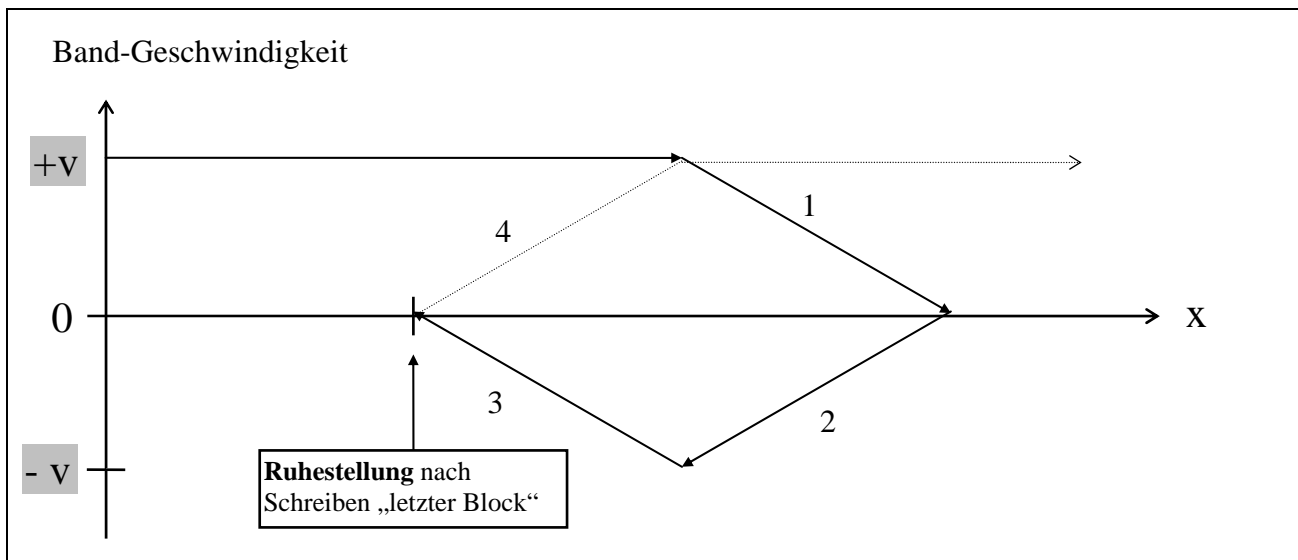
#### Kontinuierliche Bandbewegung ("streaming"):

Ziel der neuen Bandtechnik ist es, große Datenmengen zu lesen oder zu schreiben. Dadurch fällt die Startzeit nicht sehr ins Gewicht. Das Band wird „sanft“ bis auf Lesegeschwindigkeit beschleunigt, was eine einfache Mechanik ermöglicht (billig), und es werden mit gleichmäßiger Bandgeschwindigkeit („streaming“) Blöcke und Blocklücken geschrieben ohne in den Blocklücken zu stoppen.

Das Stoppen (Bremsen) des Bandes erfolgt „sanft“ wie beim Starten und benötigt eine längere Bandstrecke. Damit diese Bandstrecke für die Speicherung von Daten nicht verloren geht, kommt es nach dem Stoppen zu einer Repositionierbewegung (Repositionierschleife), d.h., das Band bewegt sich um die Strecke zurück, die es zum Stoppen brauchte und zusätzlich um die Strecke, die es zum erneuten Starten braucht. Nach Beendigung des letzten Blockes (z.B. Schreiben) wird folgende Bewegung vom Band ausgeführt:

$$t_{\text{stop}} \text{ (vorwärts)} + t_{\text{start}} \text{ (rückwärts)} + t_{\text{stop}} \text{ (rückwärts)}$$

1
2
3



**Bild: Abstoppen und Repositionieren**

Nach einem erneuten Starten wird das Beschreiben des Band infolge der Repositionierschleife unmittelbar hinter dem zuletzt (vor dem Stoppen) geschriebenen Block fortgesetzt. Wird eine Aktivität beendet, z.B. Schreiben letzter Block, und geht das Band in die Repositionierschleife, dann dauert es eine Zeit  $t_{\text{tot}}$ , bis der erste Block der nächsten Aktivität wieder geschrieben werden kann:

$$t_{\text{tot}} = t_{\text{stop}} \text{ (vorwärts)} + t_{\text{start}} \text{ (rückwärts)} + t_{\text{stop}} \text{ (rückwärts)} + t_{\text{start}} \text{ (vorwärts)}$$

1
2
3
4

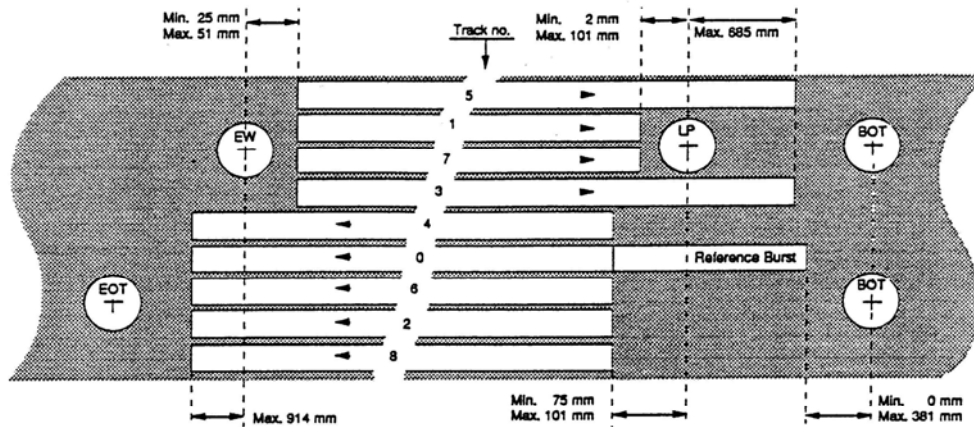
z.B.  $t_{\text{stop}} = t_{\text{start}} = 300\text{ms}$ ,  $t_{\text{tot}} = 4 \times 300\text{ms} = 1.2\text{s}$

### 3.2.4 Spur-, Datenblock- und Rahmen-Formate

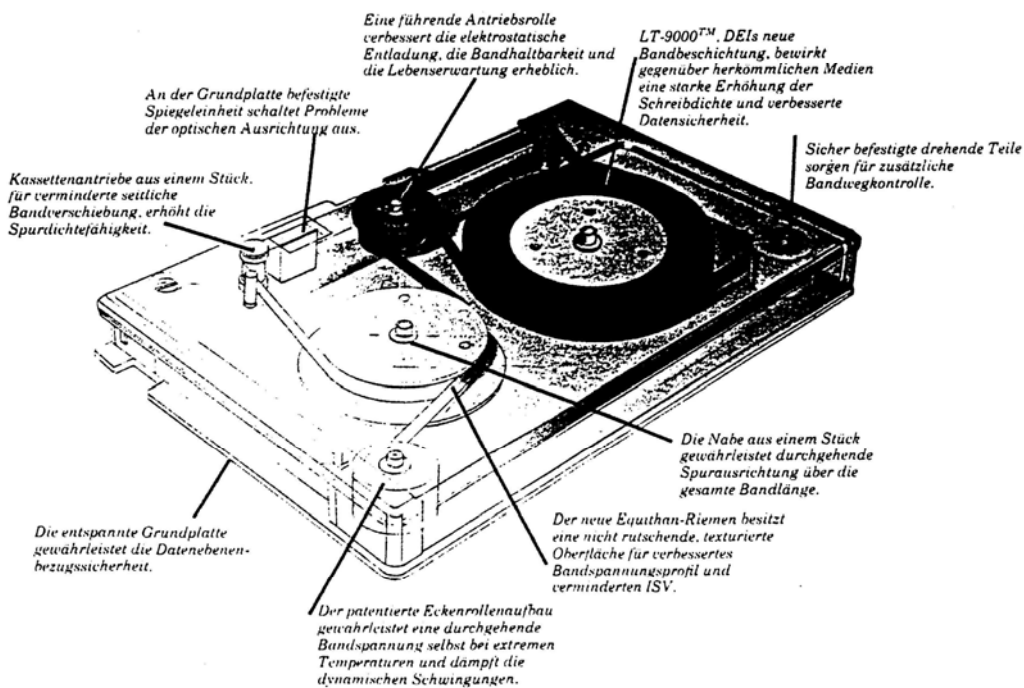
#### Spuranordnung (bei bitserieller Aufzeichnung):

**1/4-Zoll-Bandkassette mit QIC-Format**  
(QIC-Cartridge, Quelle : Fa. Tandberg)

- a) Spuranordnung QIC-24 mit 9 Spuren (QIC - 2 G Byte: 42 Spuren)
- b) Abwärtskompatibilität
- c) Aufbau Kassette

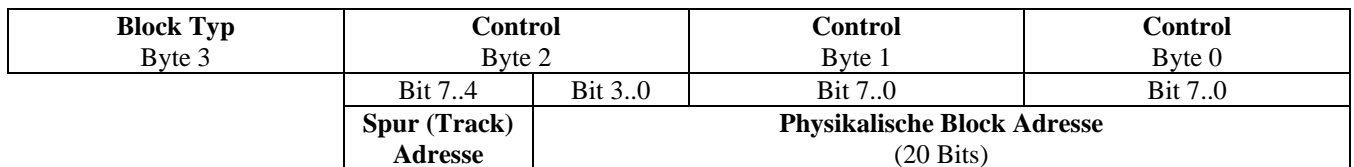
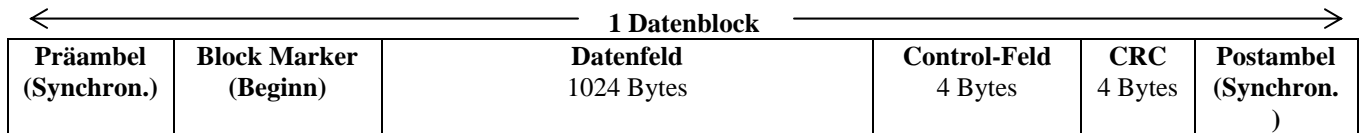


Tape Format	Capacity	Write	Read	
QIC-1000 - 30 tracks	QIC-1000	1.0 GByte	x	x
QIC-525 - 26 tracks	QIC-525	525 MByte	x	x
QIC-150 - 18 tracks	QIC-150	155 MByte	x	x
QIC-120 - 15 tracks	QIC-120	125 MByte	x	x
QIC-24 - 9-tracks	QIC-24	60 MByte		x



**Block-Format:**

Das folgende Blockformat gilt für QIC-, DLT- und LTO-Kassetten. DLT- und LTO-Kassetten enthalten größere Datenfelder (DLT: 4 kByte). (Die Zahlenwerte gelten für QIC-1GByte-Kassetten).

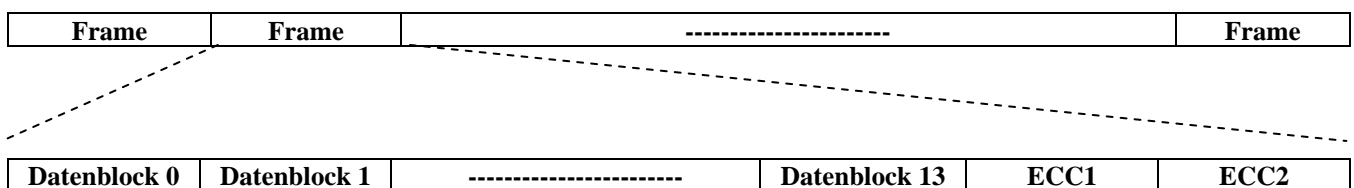


**Blockformat:**

- Präambel bis 700 Flußwechsel
- Block Marker Beginn Datenfeld
- Datenfeld 1024 Byte
- Kontrollfeld:
  - Byte 2..0: Adreßinformation  
Blockadresse:  $2E20 \text{ Blöcke} * 1 \text{ kByte} = 2E20 * 2E10 = 2E30 = 1 \text{ Gbyte}$   
Spuradresse:  $2E4 * 2 \text{ Spuren/Adresse (2 Köpfe)} = 32 \text{ Spuren maximal}$
  - Byte 3: Block-Typ, siehe später

**Rahmenformat:**

Das folgende Rahmenformat gilt auch für DLT- und LTO-Kassetten. Nur besitzen diese andere Anzahlen von Daten-Blöcken und Fehlerkorrektur-Blöcken pro Rahmen.



**Rahmenformat:**

1 **Rahmen** (Frame) = 14 Datenblöcke (je 1 kByte) + 2 ECC-Blöcke (für Datensicherung).  
Die Rahmen werden automatisch vom Gerätecontroller erzeugt.

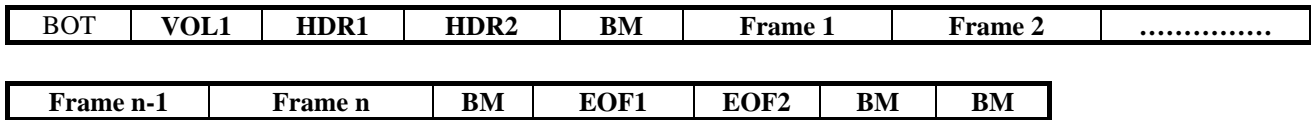
**Blocktypen** (Byte 3 im „control field“):

- Datenblock
- Bandmarke (file mark): für Aufbau eines Dateiformats
- Löschrückblock (cancel block): erklärt vorangehende Bandmarke für ungültig
- Satzmarke (setmark block). für Unterteilen des Bandes in "Sets".  
1 Band =  $\sum \text{Sets}$ , 1 Set =  $\sum \text{Dateien}$
- Füllblöcke (filler): Da nur ganze Rahmen zulässig sind, werden unvollständige Rahmen mit Füllblöcken aufgefüllt..



### 3.2.5 Verwaltung des Datenträgers (Dateien)

Beim Band sind die Blöcke nicht direkt adressierbar. Die Dateien sind daher sequentiell abgespeichert, d.h., in aufsteigender Reihenfolge der Blöcke.



**Bild: Physikalische Anordnung der Blöcke einer Datei auf dem Band. Die Datei enthält Datenblöcke, angeordnet in n „Frames“.**

**BOT** (Begin of Tape)                      Marke am Beginn des Aufzeichnungsbereichs.  
**EOT** (End of Tape)                        Warnmarke, daß bald Bandende erreicht ist.

**VOL1**    Bandanfangs-Kennsatzblock (Volume 1),  
           enthält Benutzerkennung, Zugriffsrechte, .. für den Datenträger  
**HDR1, HDR2,**                      Datei-Kennsatzblöcke (Header) am Dateianfang  
**EOF1, EOF2**                        Datei-Kennsatzblöcke (End of File) am Dateiende  
 Die Kennsatzblöcke sind kleine Datenblöcke mit einem Datenteil von 80Byte.

HDR1    enthält Dateiname, Erstelldatum, Verfallsdatum, Zugriffsrechte, ..  
 HDR2    enthält Satzformat, Blocklänge, Satzlänge (falls fest), ..  
 EOF1=HDR1  
 EOF2=HDR2

**BM**      Bandmarke (Tape mark), trennt  
           • Datenblöcke von den Kennsätzen und  
           • die Kennsätze aufeinanderfolgender Dateien  
           { Spuren 3, 6, 9 gelöscht }

Zwei aufeinanderfolgende Bandmarken bilden eine sog. **doppelte Bandmarke**, die das „**logische Bandende**“ (=Ende der gültigen Datenauf dem Band) darstellt. Die doppelte Bandmarke kann vor oder nach der EOT-Marke liegen.

Vorsicht: Eine sog. Leerdatei (=Datei ohne Datenblöcke) erzeugt ebenfalls eine doppelte Bandmarke, die aber nicht das logische Bandende darstellt. Das logische Bandende wird z.B. benötigt, um für den Test des Geräts einen Bereich zu finden, in dem Testmuster geschrieben werden können.

Auf Bankassetten ist dasselbe Dateiformat möglich wie auf dem Band.

#### Aufgabe: Magnetband, Dateiaufbau

Analog zur Platte existieren die entsprechenden Satz- und Blockformate. Im Unterschied zur Platte wird kein Schlüsselfeld unterstützt.

**Korrektur von Schreibfehlern:**

<b>Block n-1</b>	<b>Block n</b>	<b>Block n+1</b>	<b>Block n+2</b>	<b>Block n</b>	<b>Block n+1</b>	<b>Block n+2</b>	<b>Block n+3</b>
<b>o.k.</b>	<b>Schreibfehler</b>		<b>Schreib-Stop, da defekter Block n erkannt</b>	<b>Schreiben Block n jetzt o.k.</b>	<b>Schreiben Block n+1</b>	<b>Schreiben Block n+2</b>	

**Aufzeichnungsrichtung** →

**Schreibfehler:**

Direkt nach dem Schreiben eines Blocks erfolgt automatisch ein Kontroll-Lesen. Durch den Lesekopf wird ein Schreibfehler sofort erkannt (genau: 2 Blöcke später wegen des räumlichen Versatzes Schreibkopf/Lesekopf). Eine Schreibwiederholung auf derselben Stelle wird wegen der langen Repositionierzeit nicht gemacht, sondern es werden die Blöcke, beginnend mit dem defekten Block (Block n) nochmals geschrieben. Da im Gerät ein Datenpuffer vorhanden ist, sind die Daten noch vorhanden:

**Block n (defekt) --- Block n+1 --- Block n+2 --- Block n (wiederholt) --- Block n+1 --- usw.**

**Fehlerstrategie:**

- Schreibfehler:

Die oben beschriebene Schreibwiederholung wird mehrmals durchgeführt (max. 16x, einstellbar über den SCSI-Befehl "Write Retry Count"). Wenn ohne Erfolg, dann Abbruch und Meldung an den Host.

- Lesefehler:

Fehlererkennung über die CRC-Bytes des Blocks. Es erfolgt eine Fehlerkorrektur über die 2 redundanten ECC-Blöcke. Maximal dürfen 2 Datenblöcke pro Rahmen defekt sein. Bei  $\geq 3$  defekten Datenblöcken wird eine Lesewiederholung der Blöcke durchgeführt mit seitlichem Versatz des Lesekopfes aus der Spurmitte. Insgesamt 24x (8x Kopfposition spurmittig, 8x seitlicher Versatz  $+\Delta x$ , 8x seitlicher Versatz  $-\Delta x$ ). Danach Abbruch und Meldung "harter Fehler".

**Adressierung:**

Ab QIC-525 (525 MByte) und SCSI-2 ist eine Direktadressierung von Blöcken möglich, was wesentlich kürzere Zugriffszeiten liefert (quick file access, QFA). Die letzte Spur wird als Directory-Spur verwendet, die das Inhaltsverzeichnis enthält.

Ablauf bei Direktadressierung (max. 40s bei QIC-1000):

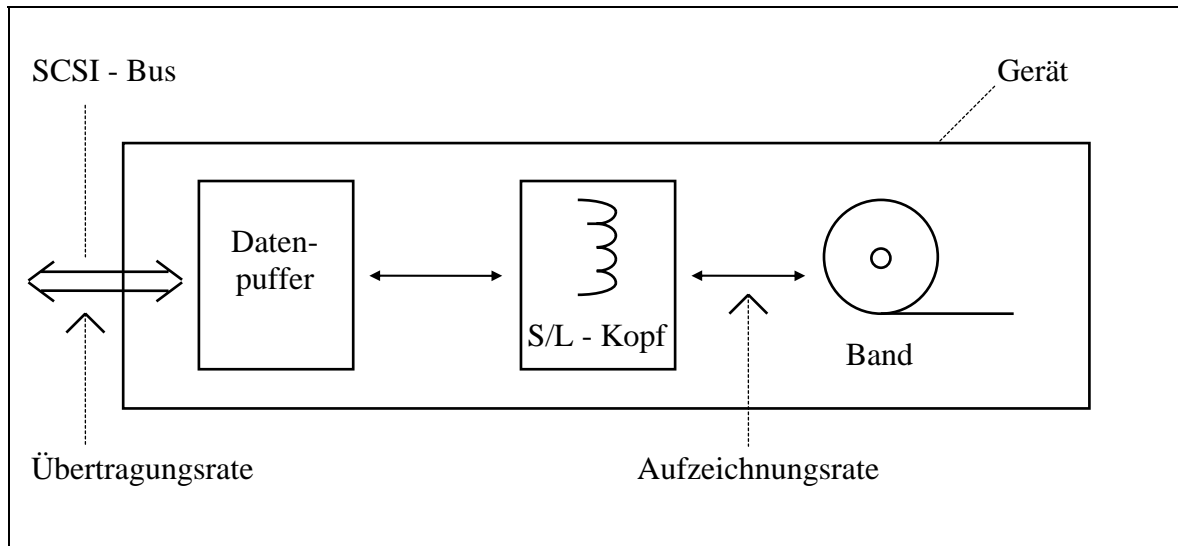
- Inhaltsverzeichnis lesen (Spur 29): → Positionieren auf Datenspur (0..28): → Lesen
- Blockadressen (Blockadresse= Spur-Nr.+ Block-Nr.) → Lesen/Schreiben Datenblock.

### 3.2.6 Technische Daten

Magentbandkassetten sind mit den üblichen Geräteschnittstellen ausgestattet (ATA, SCSI, USB, Fire Wire).

#### Datenraten:

- Aufzeichnungsrate=Datenrate am Schreib/Lesekopf (z.B. bei QIC-Kass. 300-600 kByte/s)
- Übertragungsrate=Datenrate des Pufferspeichers im Gerät (z.B. .... MByte/s)



**Bild: Datenraten bei der Kassette**

*Frage:*

*Sie speichern eine Datenmenge von 1 GByte ab. Welche Datenrate ist wirksam ?*

*(Datenpuffer=0,5 MByte, Übertragungsrate=20 MByte/s, Aufzeichnungsrate=0,4 MByte/s)*

Abspeichern von Daten auf der Kassette:

- Bei Datenmengen  $\leq$  Pufferkapazität des Geräts: Die Übertragungsrate bestimmt die Übertragungszeit Arbeitsspeicher  $\rightarrow$  Gerät (Pufferspeicher zu Beginn leer).
- Bei Datenmengen  $\gg$  Pufferkapazität des Geräts: Die Aufzeichnungsrate bestimmt die Übertragungszeit Arbeitsspeicher  $\rightarrow$  Gerät.

Aufgabe: Streamerkassette, Datenrate

#### **Beispiel: Datenblatt**

### **3.3 Kassetten-Systeme (Auto Changer, Juke Box)**

Beim Einsatz der Kassette als Massenspeicher kann durch Kassettensysteme die Kapazität vervielfacht werden (TeraBytes). Ein Roboter wechselt die Kassetten (Bspl.: 6 s/Wechsel). Es wurden spezielle SCSI-Befehle für Kassettensysteme genormt.

Um die Zugriffszeiten beim Schreiben von Dateien auf ein Kassettensystem zu verbessern (=verkürzen) können in Kassettensystemen Festplatten als Zwischenspeicher (Hard Disk Caching) zwischen dem Rechner und den Kassetten verwendet werden. Schreibdaten werden zunächst auf der Platte abgespeichert. Der Kassetten-Controller sucht dann einen freien Platz auf einer Kassette und speichert die Daten dort ab (Hintergrundoperation). Die Festplatte arbeitet hier als "Schreib-Cache" für die Kassetten.

Festplatten werden generell als schnelle Pufferspeicher für langsamere Geräte eingesetzt wie z.B. für DVD-Server (z.B. System mit 112 DVDs).

Anwendungsbeispiele:

Im Fachhochschul-Rechenzentrum sind DLT-Kassettensysteme im Einsatz mit ca. 100 Kassetten zu ca. 35GByte, das ergibt  $3500\text{GByte} = 3,5\text{Terabyte}$  Speicherkapazität. Im Kassettensystem sind 4 Laufwerke installiert. Der Roboter erkennt eine Kassette im Magazin über einen Bar-Code.

Das Software-Haus SAP AG (Anwendersoftware R/2 und R/3) speichert jede Nacht ca. 2 Terabyte Daten.

### 3.4 Datensicherung (Backup) und Archivierung

#### **a) Backup**

Unter Backup (Sicherung) versteht man das regelmäßige Erstellen von Sicherheitskopien vordefinierter Datenbestände mit begrenzter Lebensdauer (!)

Die Sicherung soll die Restauration der Datenbestände im Falle von HW oder SW-Fehlern, Bedienfehlern (Löschen von Dateien) usw. ermöglichen.

Systemabsturz/Datenverlust verursacht durch?

#### **Begriffe:**

-- Full Backup (Voll-Sicherung, Vollkonserve)

Hier werden alle Daten (oder bestimmte Datenpfade) eines Rechnersystems unabhängig von Alter einzelner Dateien gesichert.

Vorteil: Einfaches und schnelles Restaurieren ganzer Verzeichnisebenen oder eines ganzen Rechnerbestandes;

Nachteil: Lange Backupzeiten.

-- Differential Incremental Backup (Teil-Sicherung):

Es werden alle Daten gesichert, die seit der letzten Datensicherung (Voll-Sicherung oder Teil-sicherung) geändert oder neu angelegt wurden.

Vorteil: kurze Backupzeiten, wenig Backupvolumen, geringe Netz- und Rechnerlast;

Nachteil: bei Wiederherstellen ganzer Verzeichnisebenen bzw. kompletter Rechnerdatenbestände müssen alle inkrementellen *Backups* seit dem letzten vollen bzw. differentiellen *Backup* eingelesen werden. Dies führt zu höherem administrativem und zeitlichem Aufwand beim Restore.

-- Cumulative Incremental Backup (Teil-Sicherung):

Es werden alle Daten gesichert, die seit der letzten vollen Datensicherung (Vollsicherung) geändert oder neu angelegt wurden.

Vorteil: Die Wiederherstellung ganzer Verzeichnisebenen bzw. kompletter Rechnerdatenbestände ist relativ einfach, da nur der jeweils aktuelle „kumulativ inkrementelle Backup“ und der letzte „Full Backup“ zurückgespielt werden. Beim Einsatz kumulativ inkrementeller Backups kann man die Full Backups in längeren Intervallen durchführen.

Nachteil: Je länger der Zeitraum zwischen den Full Backups wird, desto stärker wächst das Volumen der kumulativ inkrementellen Backups (im Vgl. zu den differentiell inkrementellen Backups).

Eine Mischung der drei Strategien (Full backup, inkrementelles Backup, differentiell Backup) gewährleistet in der Regel eine ausgewogene und zuverlässige Datensicherung.

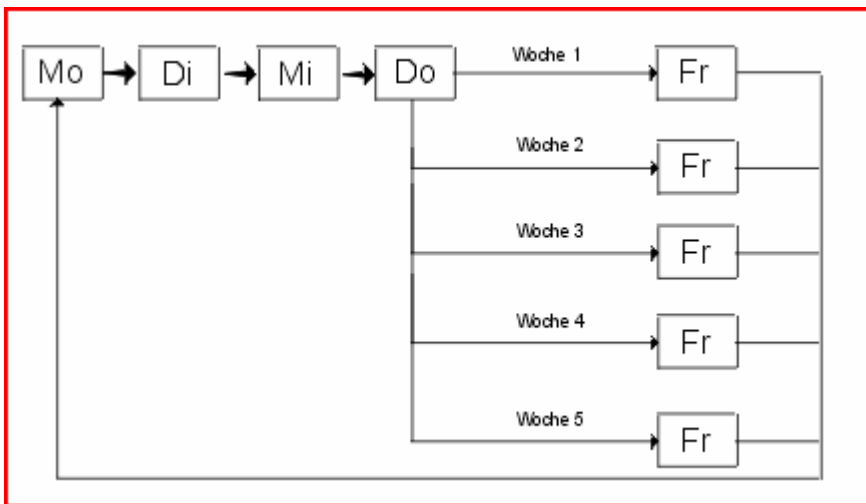
- Online B. zu sichernde Datenbank geöffnet
- Offline B. zu sichernde Datenbank geöffnet heruntergefahren
- Physical B.
- Logical B.

### Beispiel für eine "Backup"-Methode:

Die "Backup"-Methode hängt u.a. davon ab, wie häufig sich die Daten ändern, wie groß diese Änderungen i.allg. sind und wie "wichtig" die gespeicherten Daten sind. Gesichert werden i. allg. nicht die Daten von lokalen Arbeitsplätzen sondern nur die Festplatten der File Server der lokalen Netze. Das "Backup" sollte systematisch durchgeführt werden.

Für die verschiedenen Backup- und Restore-Abläufe stellen sog. "Storage-Manager"-Programme automatisierte Abläufe zur Verfügung. Firmengeheime Daten können vom Backup mit "exclude"-Anweisungen ausgeschlossen werden.

a) Beispiel für regelmäßige Monatssicherung (Fa. Tandberg)



*Anzahl Kassetten = ?  
Welcher Backup-Typ  
wäre sinnvoll?*

b) 10-Bandkassetten-Methode (Firma HP)

3-monatige Datensicherung unter Verwendung von täglichem, wöchentlichem und monatlichem Backup. Nach 1 Jahr sind neue Bandkassetten zu verwenden, siehe Aufgabe.

### Platten-Backup:

Zunehmend werden Festplatten beim Backup eingesetzt. Während die komplette Datensicherung am Wochenende nach wie vor auf Bandkassetten erfolgt, erfolgt die Datensicherungen unter der Woche auf RAID-Systemen (sog. „**tapeless Backup**“). Dies ermöglicht die schnelle Wiederherstellung einzelner Server oder Dateien im Alltag, während der Schutz gegenüber Totalausfällen durch die Bandkassetten nach wie vor gegeben ist.

## b) Archivierung

Unter Archivierung ist die in der Regel vom Benutzer initiierte Funktion zu verstehen, Daten/ Dateien langfristig auf ein externes Speichermedium auszulagern und gegebenenfalls nach Jahren innerhalb weniger Minuten wiederherzustellen.

Bei einer Archivierung werden nach erfolgreicher Auslagerung die Originaldateien automatisch von der Magnetplatte des Nutzers gelöscht. Damit bietet sich dem Nutzer nicht nur die Möglichkeit zur langfristigen Aufbewahrung von Datenbeständen (und ggf. Programmen, mit denen auf die Daten zugegriffen wird), sondern auch die Chance, den eigenen Plattenplatz ökonomisch zu verwalten bzw. Plattenplatz einzusparen.

Zur Erhöhung der Datensicherheit werden Archivierungsdaten in zweifacher Kopie auf verschiedenen Datenträgern abgelegt. Im Gegensatz zum Backup läuft die Archivierung nicht automatisch ab.

## 3.5 Vergleich Bandgeräte für Backup

### 3.6 Aufgaben

#### **Aufgabe SeqD-1: Streamerkassette, Datenrate**

*Kassette (TDC 4100):*

*80 ips = Schreib/Lese-Geschwindigkeit, ips=inch per second*

*45000 bpi = Bitdichte (codiert) auf Spur bpi=Bits per inch*

*RLL 0,2-Code: 4Bit → 5Bit*

*0,3 s = Start- bzw. Stoppzeit*

*1 kByte = Datenfeld eines Blocks*

1. Berechnen Sie die Datenrate  $R_K$  (kByte/s), mit der die codierten Daten auf der Kassette abgespeichert werden unter der Annahme, daß kein Repositionieren stattfindet und die Datenfelder 70% der Spur belegen (Startzeit vernachlässigbar).
2. Welche Datenrate resultiert daraus für den Anwender?  
Berechnen Sie die Datenrate in kByte/s und in MByte/min!
3. Der Datenträger ist beschädigt. Jeder 100-ste Block kann beim Lesen erst mit 5 Repositionier-Schleifen fehlerfrei gelesen werden. Auf welchen Wert sinkt die Datenrate ab (kByte/s und MByte/min)?

#### **Aufgabe SeqD-2: Magnetband, Dateiaufbau**

1. Skizzieren Sie das Format einer Leerdatei auf Magnetband:
2. Bei einem Magnetband bilden die siebte und achte Bandmarke (von BOT an gezählt) eine doppelte Bandmarke. Kennzeichnet diese doppelte Bandmarke das logische Bandende? Begründung!
3. Läßt sich durch einfache Überlegung auf Grund des Ergebnisses von Frage 2. sagen, ob z.B. die 37. und 38. Bandmarke das logische Bandende sein können? Begründung!

**Aufgabe SeqD-3: Backup mit Bandkassetten**

Entwickeln Sie eine 10-Bandkassetten-Methode für eine 3-monatige Datensicherung unter Verwendung von täglichem (Mo bis Do), wöchentlichem und monatlichem Backup. Geben Sie an, wann Sie Voll-Sicherung und wann Teil-Sicherung verwenden.

Vorgehensweise:

Markieren sie auf der Zeitachse für 4 Monate (Einteilung: Monate, Wochen, Tage), wann welches Backup durchgeführt wird und markieren Sie die Bandkassetten 1 bis 10.

**Aufgabe SeqD-4 : Dateiorganisation**

Erstellen Sie für die nachfolgende Datei den Zylinder-Index und den Spur-Index (index-sequentielle Datei)

*Datei: Besteht aus 63 Sätzen fester Länge, geblockt (Blockungsfaktor = 3), mit Schlüsselfeld.*

*Aufsteigende Reihenfolge der Schlüssel (= Satz-Nummern) = 1,2,3,4...*

*Die Datei ist auf Platte abgespeichert, beginnend mit:*

*Zyl.Nr. = 78, Spur-Nr. = 7, Sektor-Nr. = 9*

*Platte: 9 Sektoren/Spur (Sektor 1 bis Sektor 9)*

*8 Spuren/Zyl. (Spur 1 bis Spur 8)*

Tragen Sie die Ergebnisse in die nachfolgenden Felder ein:

Zyl.-Index:

Zylinder	

Spur-Index:

Spur	

Spur	

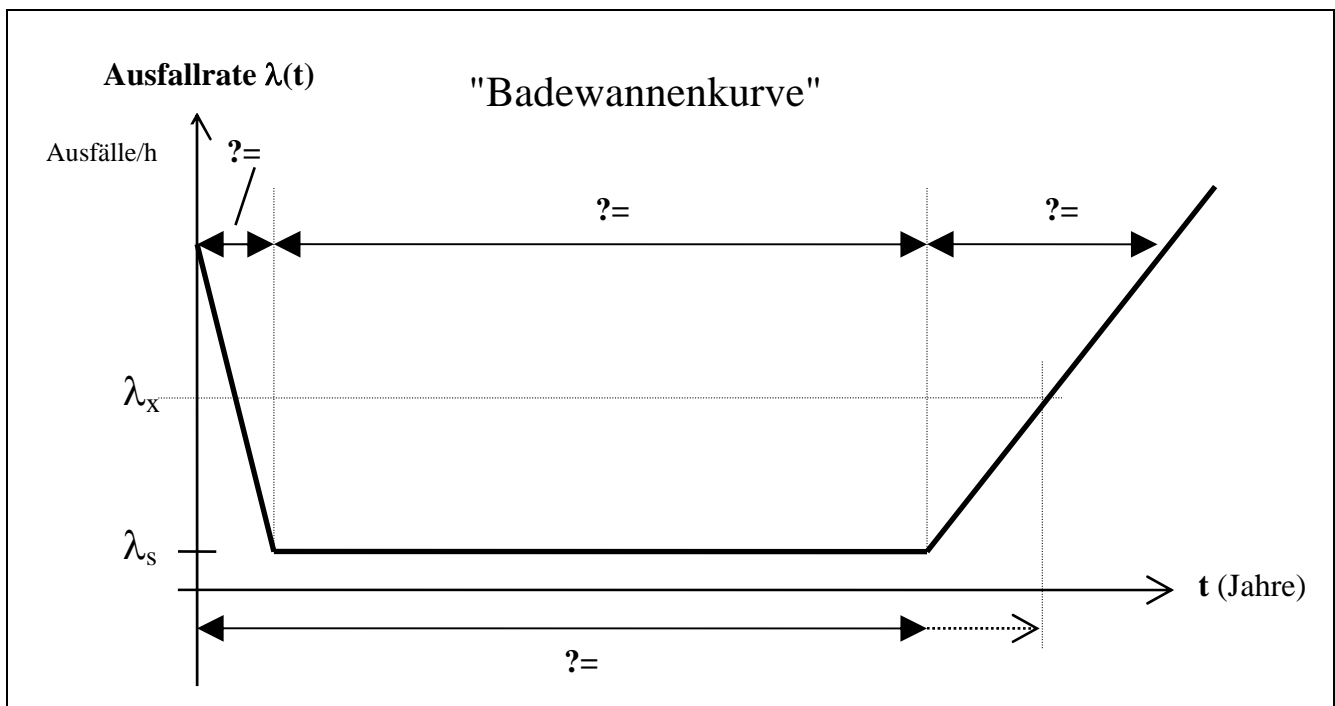
## 4 Gemeinsame technische Daten

### 4.1 Zuverlässigkeit (Reliability)

Die Zuverlässigkeit spielt eine große Rolle bei der Planung von ausfallsicheren Systemen wie z.B. von Steuerungen bei Kernkraftanlagen, Zentralrechnern von Vermittlungsstellen (Ausfallzeit maximal 2h pro 10 Jahre), Steuerrechnern eines Flughafenleitsystems usw. . Aus den Zuverlässigkeitswerten (MTBF, MTTR, Ausfallwahrscheinlichkeit, ..) der Komponenten lassen sich die Ausfallwahrscheinlichkeit und/oder die Verfügbarkeit des Gesamtsystems berechnen. Auch bei einfacheren Systemen wie z.B. bei lokalen Netzen mit Servern für die zentrale Datenhaltung ist eine Abschätzung der Ausfallwahrscheinlichkeit und daraus resultierend eine Datensicherungsstrategie enorm wichtig, da ein Datenverlust in einer Firma zu Mehrkosten, Terminverzug, Mehrarbeit usw. führen kann.

### Lebensdauer und mittlerer Fehlerabstand MTBF (Mean Time Between Failure)

Der Zusammenhang zwischen den verschiedenen Begriffen wie Lebensdauer, MTBF, usw. läßt sich am besten an der "Badewannenkurve" erklären.



Frage: Wie heißen die 3 Bereiche der Badenwannen-Kurve?  
Wie heißt das Zeitintervall auf der x-Achse?

### Frühausfälle:

Unmittelbar nach der Produktion fallen relativ mehr Geräte aus aufgrund von fehlerhaften Bauelementen oder Produktionsfehlern. Solche Fehler können nie gänzlich ausgeschlossen werden. Damit solche Ausfälle nicht beim Kunden auftreten führen die Hersteller mit jedem Gerät einen sogenannten "Burn-in" durch. Dabei werden die Geräte solange betrieben (ggf. bei höherer Betriebstemperatur) bis die Phase der Frühausfälle vorbei ist.

### Statistische Ausfälle:

Nach der Phase der Frühausfälle bleibt die Ausfallrate  $\lambda$  konstant über einen langen Zeitraum auf einem niedrigen Niveau  $\lambda_s$ . Die Geräte fallen nur noch selten aus. Man spricht von der "statistischen" Ausfallrate. Der Kehrwert der Ausfallrate  $\lambda_s$  ist die MTBF (h). Der Wert für die MTBF gibt den mittleren zeitlichen Abstand zwischen zwei auftretenden Fehlern an.

$$MTBF = 1/\lambda_s$$

$\lambda_s$  = **Ausfallwahrscheinlichkeit pro h**

$$\lambda_s = \frac{n_{\text{Ausfall}}}{(n_{\text{gesamt}} * t)}$$

$n_{\text{gesamt}}$  = Gesamtzahl der gleichzeitig betriebenen Geräte,  $t$  = Beobachtungszeitraum (h),  
 $n_{\text{Ausfall}}$  = Zahl der Ausfälle

#### Beispiel 1:

geg.:  $n_{\text{Ausfall}} = 1000$  Geräte,  $t = 100\text{h}$ ,  $n_{\text{Ausfall}} = 2$   
gesucht: MTBF

#### Beispiel 2:

geg.: MTBF = 1000h

Sind 1000 Geräte in Betrieb, so fällt pro Stunde im Durchschnitt 1 Gerät durch Fehler aus (mathematisch genau nur, wenn jedes ausgefallene Gerät sofort durch ein neues ersetzt wird).

Sind 10 Geräte in Betrieb, so fällt pro 100 Stunden 1 Gerät durch Fehler aus.

#### Beispiel 3:

geg.: MTBF Gerät = 50.000h, 10 Geräte, betrachtetes Zeitintervall = 1h

gesucht: Ausfälle  $n$  pro h

### Verschleißausfälle:

Nach einer bestimmten Zeit treten Verschleißerscheinungen auf (z.B. Mechanik), die immer häufiger zu Ausfällen führen (z.B.  $\lambda_x > 5\lambda_s$ ). Das Gerät hat seine "Lebensdauer" überschritten. Die Ausfallrate steigt an.

z.B. Lebensdauer eines Geräts = 5 Jahre

Bei einer Ausfallrate während der Lebensdauer von  $\lambda_s = 10^{-5}$  (1/h) errechnet sich die MTBF zu:  $MTBF = 1/\lambda_s = 10^5 \text{ h} \approx 11,4$  Jahre.

### Bedeutung von Lebensdauer und MTBF in der Praxis

Ist ein möglichst störungsfreier Betrieb nötig, sollte ein Gerät vor Ablauf der Lebensdauer ausgetauscht werden.

Über die MTBF kann abgeschätzt werden, wieviele Geräte vor Ablauf der Lebensdauer pro Zeiteinheit (Monat, Jahr etc.) ausfallen.

**MTTR** (mean time to repair)

Mittlere Reparaturzeit/ Fehlerbehebungszeit/ Ausfallzeit

Üblicherweise wird bei Geräten MTTR=1h angegeben. In dieser Zeit ist ein Gerät natürlich nicht zu reparieren; es handelt sich hier nur um einen Austausch von ganzen Moduln (z.B. Leiterplatte) oder des ganzen Geräts.

**Verfügbarkeit A (availability)  $A = MTBF / (MTBF + MTTR)$**

z.B. MTBF=100 000h, MTTR=1H

$$A = 100\,000 / 100\,001 = 0.9999.. \rightarrow 99.99..%$$

**Beispiel: Zuverlässigkeitsangaben bei einer Festplatte** (Quelle IBM, Blatt 028):

- MTBF=800 000h (Betriebsdauer)
- CDF=3.75% (cumulative distribution failure rate)  
Ausfallwahrscheinlichkeit in 5 Jahren bei 500h Betriebszeit im Monat  
Berechnung:  
5 Jahre \* 12 Monate \* 500h / Monat = 30 000h Betriebsdauer in 5 Jahren  
30 000h / 800 000h = 0.0375  $\rightarrow$  3.75% = Wahrscheinlichkeit, daß 1 Gerät in 5 Jahren ausfällt.

**(Bit)Fehlerraten:**

• **Weicher Fehler (soft error)**

Fehler, die bei Wiederholung des Vorgangs nicht mehr auftreten, z.B. Lesefehler, der durch erneutes Lesen behebbar ist.

1 weicher Fehler bei  $10^{10}$  gelesenen Bits entspricht einer Fehlerwahrscheinlichkeit von  $10^{-10}$ .

• **Harter Fehler (hard error)**

Fehler, der durch Wiederholung des Vorgangs nicht behebbar ist, z.B.

1 harter Fehler bei  $10^{11}$  gelesenen Bits (Fehlerwahrscheinlichkeit  $10^{-11}$ )

Die Grenze zwischen weichem Fehler und hartem Fehler hängt von der Zahl der Wiederholungen ab, die man bei Auftreten eines Fehlers vornimmt. Häufig wird bei Platten bei Auftreten eines Fehlers während des Lesen eines Sektors wie folgt verfahren:

- 9 x Lesen maximal, davon
- 3 x mit Phasenlage Lesetakt  $\Delta\phi = 0$
- 3 x mit Phasenlage voreilend  $+\Delta\phi$
- 3 x mit Phasenlage nacheilend  $-\Delta\phi$

Mit der Änderung der Phasenlage des Lesetaktes verschiebt man die Lage des Taktes zum Lesesignal und erreicht somit bei vorher fehlerhafter Phasenlage u.U. ein höheres Lesesignal.

Bei Magnetbandkassetten verschiebt man bei Wiederholungen die Position des Lesekopfes etwas ( $\pm\Delta x$ ) aus der Spurmitte um eine eventuelle Falschlage zu korrigieren.

Bei Geräten mit Fehlerkorrektur (ECC) werden Einzelbitfehler oder kleine verfälschte Bitgruppen über die Fehlerkorrekturlogik korrigiert. Wiederholungen des Vorgangs kommen erst dann, wenn die Fehlerkorrekturlogik überfordert ist (zu viele verfälschte Bits).

**Beispiel: Zuverlässigkeitsangaben bei einer Festplatte** (Quelle IBM, Blatt 028):

- korrigierbare (Bit)Fehler  $< 10$  in  $10^{11}$  gelesenen Bits  $\rightarrow 10^{-10}$  Fehlerwahrscheinlichkeit.
- nicht korrigierbare (Bit)Fehler  $< 10$  in  $10^{15}$   $\rightarrow 10^{-14}$  Fehlerwahrscheinlichkeit.

## **4.2 Umgebungstemperatur**

Für die Umgebungstemperatur der Geräte im Betrieb sind Bereiche vorgeschrieben, bei deren Überschreitung die Fehlerwahrscheinlichkeit ansteigt.

Platte, Diskette, Bandkassette	ca. 5..45 °C
Band (Speichermedium)	16..32 °C
Optische Speicher	ca. 5..40 (50) °C

Die Temperaturbereiche der Speichergeräte sind also wesentlich enger als die der Logikbausteine.

## **4.3 Aufgaben**

<b>Aufgabe 12: Zuverlässigkeit</b>
------------------------------------

**Bitte nur kurze Antworten (Stichpunkte)**

1. *Die Lebensdauer einer Platte beträgt 5 Jahre, ihre MTBF 50 Jahre.*  
Ist das ein Widerspruch? Begründung!
2. Erklären Sie den Begriff "Soft Error" und wie ist die Grenze zum "Hard Error" festgelegt?