

Modul “Digitale Medienproduktion”

Digitale Audio- und Videotechnik – DAVT

MIB-5

Begleitende Folien zur Vorlesung in WiSe-2022/2023

Prof. Dr. Miguel A. García

© Copyright

Diese begleitenden Folien unterliegen einer permanenten Aktualisierung. Einige der verwendeten Quellen sind noch nicht sachgerecht gekennzeichnet, daher sind diese Unterlagen ausschließlich für den internen Gebrauch in der Fakultät Digitale Medien der Hochschule Furtwangen bestimmt. Die Verbreitung außerhalb der Hochschule ist ausdrücklich nicht gestattet.

Digitale AV-Technik - Beschreibung

2 ECTS-Leistungspunkte

Inhalt:

- Digitalisierung Audio- und Videosignale
- Native Audio- und Videosignale
- Digitale Bildformate
- Datenformate im Medienbereich
- Digitale Signalverarbeitung Audio
- Digitale Signalverarbeitung Video
- Media-Encoder und Decoder (Codecs)
- Display- und Projektionstechnologien

Diese Veranstaltung bildet mit der Partnerveranstaltung Datenverarbeitung in der Medienproduktion das integrierte Lernmodul "Digitale Medienproduktion".

Themen:

- Natives Digital Video, Digitale Komponentensignale 
- Farbräume der Postproduction und HDR (*High Dynamic Range Image* – Hochkontrastbild) 
- Datenformate im Medienbereich (Fileformate) 
- Prinzipien der digitalen Signalverarbeitung
- Encodingtechnologien Audio inkl. reale Implementierungen
- Encodingtechnologien Video inkl. reale Implementierungen

Die Veranstaltung schließt mit einer schriftlichen Klausur!

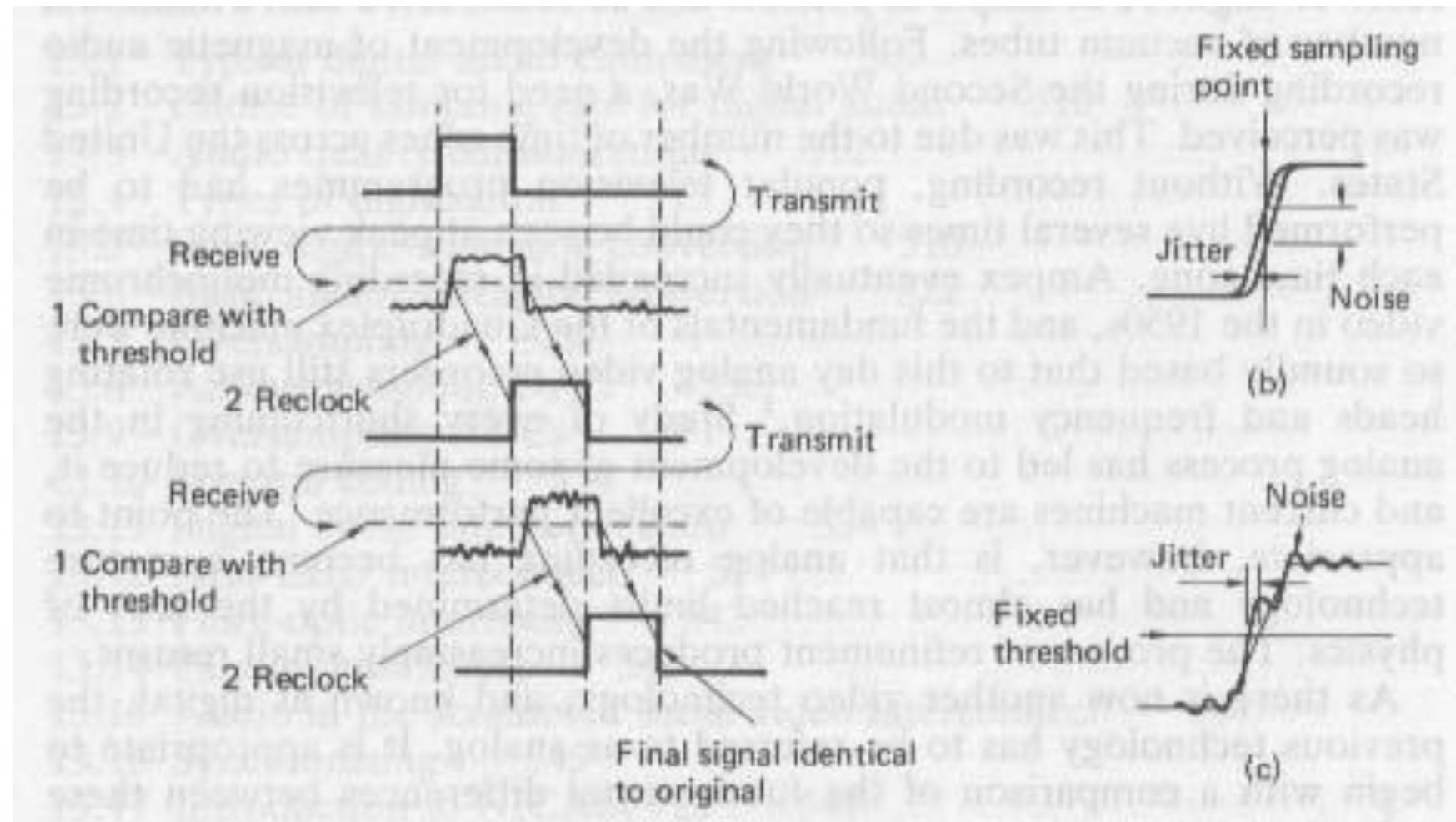
Literatur:

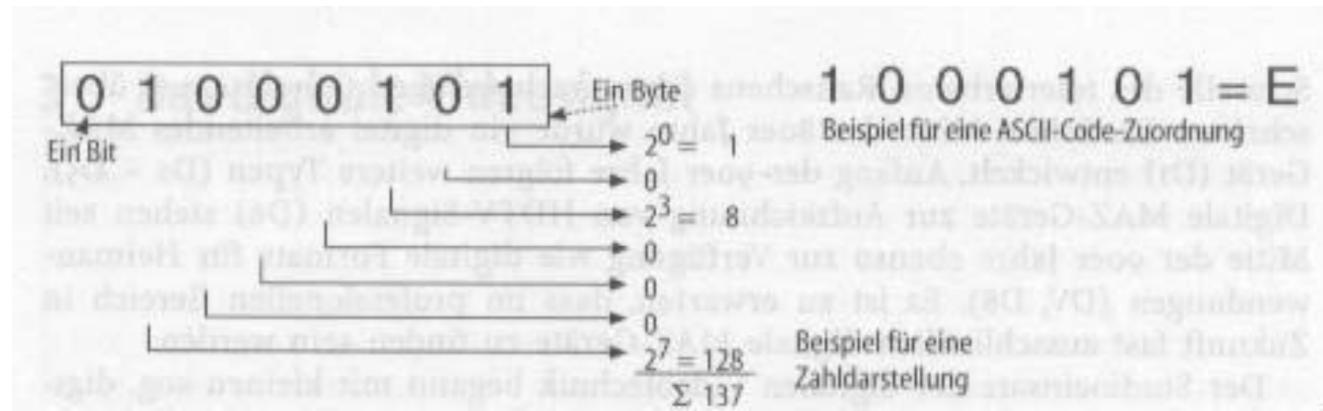
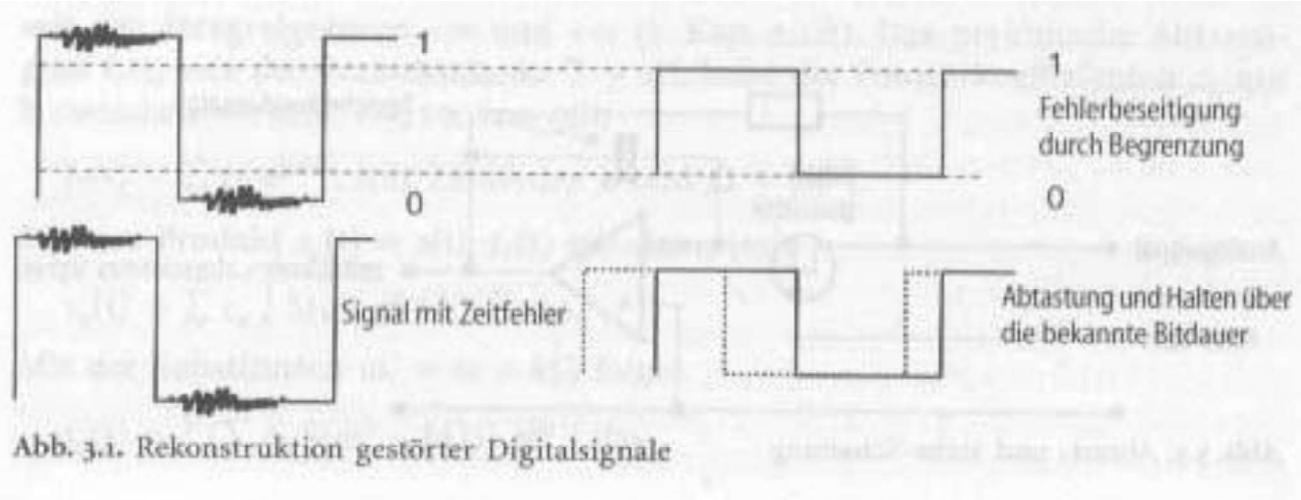
- Ulrich Schmidt "Professionelle Videotechnik", Springer Vieweg; Auflage: 6., 2013, ISBN-13: 978-3642389917
- Charles Poynton, „Digital Video and HDTV“, Morgan Kaufmann Publishers, Second Edition 2012, ASIN: B0092J20YU
zusätzlich: www.poynton.com
- J. Watkinson “MPEG Handbook“, Focal Press, ISBN: 0240516567
- Heyna, Briede, Schmidt, “Datenformate im Medienbereich”, Fachbuchverlag Leipzig, ISBN 3-446-22542-0
- Reimers, Ulrich “Digital Video Broadcasting (DVB)”, Springer, ISBN 3-540-60946-6

Warum digitale Technik?

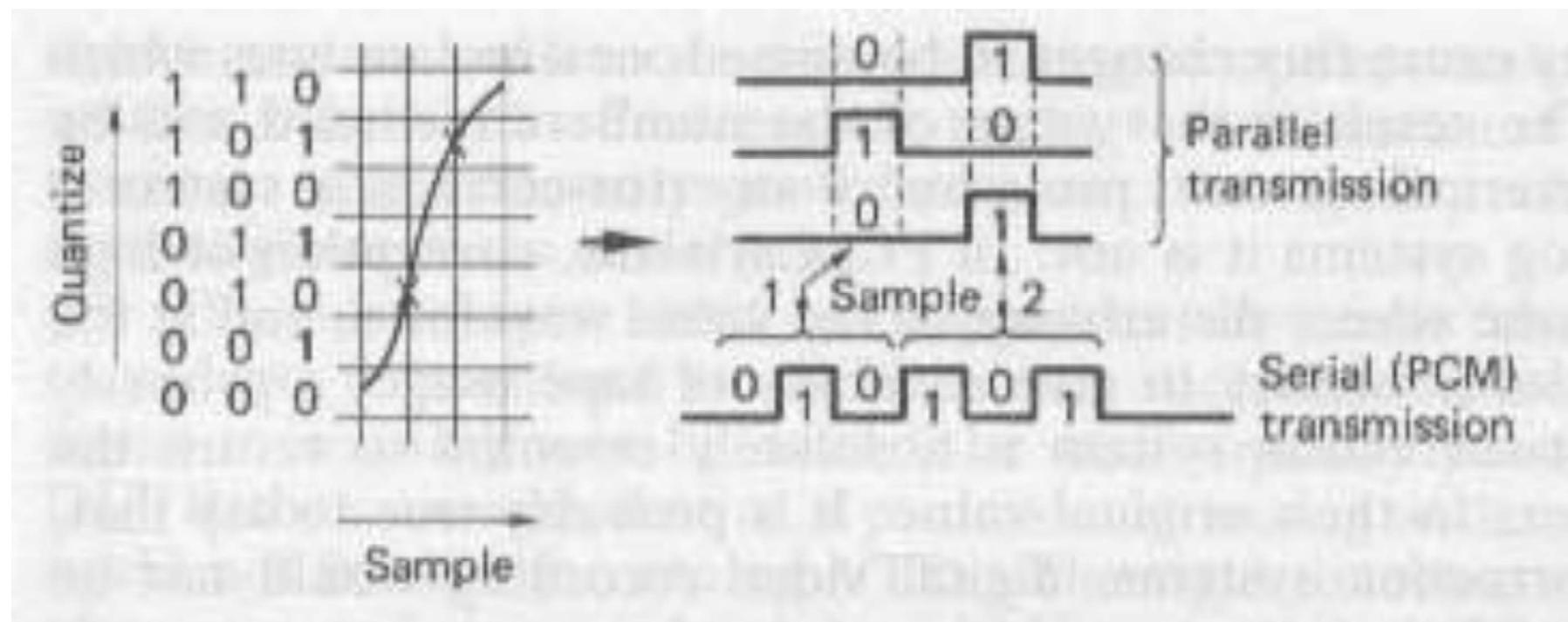
(Vorteile der digitalen Technik!)

- Die Qualität des Signals ist unabhängig vom Kanal
- Digitale Signale sind störunanfälliger
- Keine Verschlechterung des Signals beim Kopieren
- Leichte Speicherung, Verzögerung und Manipulation des Signals
- Fehlererkennung und –korrektur sind möglich
- Datenreduktionsverfahren sind analog nicht realisierbar
- Das Signal-Rausch-Verhältnis bleibt konstant (SNR)
- Gemeinsame Schaltungen für Audio- und Videoverarbeitung





Qualitätsverlust nur beim Umwandlungsprozeß



Hat die digitale Technik Nachteile?

- Qualitätsverlust nur beim Umwandlungsprozeß
- Die Signalqualität wird beim Systemdesign fixiert
- Komplexität der Geräte
- Eine riesige Datenmenge
- Die Signalverarbeitung (DSP) kann Verschlechterung erzeugen
- Bei Audio- und Videosignalen keine Rückwärtswiedergabe möglich

Selbständige Aufgaben:

Lesen Sie bitte aus der Bibliografie die folgenden Kapitel:

- Why digital? Kapitel 1 aus Watkinson, John; *The Art of Digital Video*; Focal Press, 1990. ISBN: 0 240 51287 1
- Why digital? Kapitel 1 aus Watkinson, John; *The Art of Digital Audio*; Focal Press, 1988. ISBN: 0 240 51270 7
- 3.6.2 Die digitale Übertragungsform und 3.6.2.1 Analog- und Digitalübertragung im Vergleich aus Webers, Johannes; *Handbuch der Film- und Videotechnik*, Franzis'-Verlag, 1993. ISBN: 3-7723-7114-0

Von einem analogen zu einem digitalen Signal!

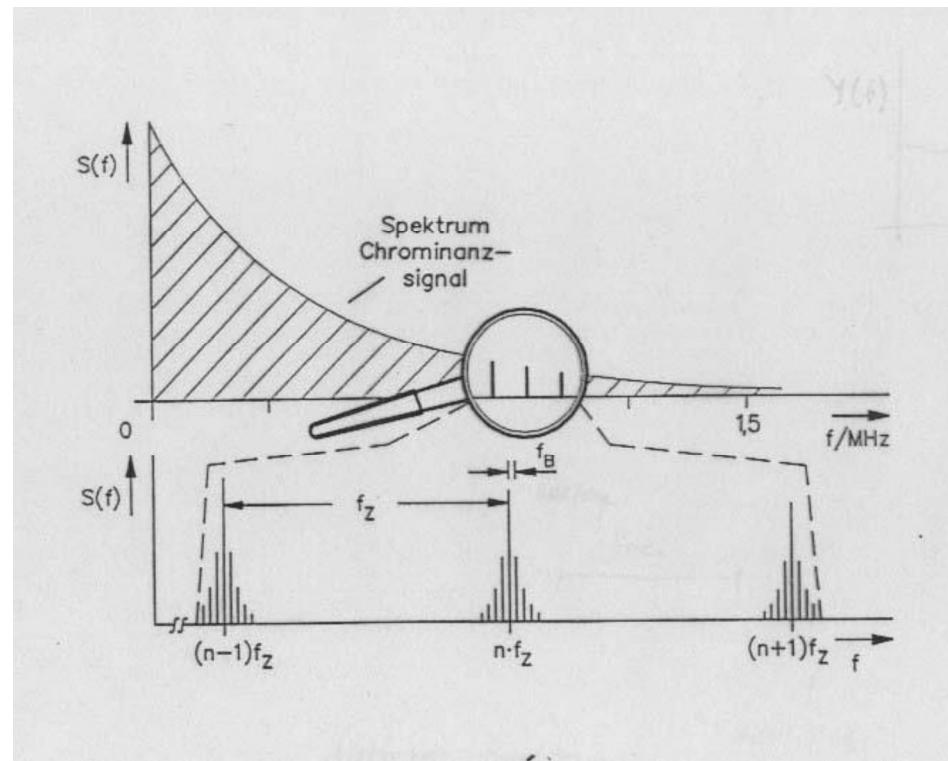
A/D-Umwandlung

0. **Bandbegrenzung**
1. **Abtastung**
2. **Quantisierung**
3. **Codierung**

Frequenzbandbreite

Der Bereich der sich von der tiefsten bis zur höchsten Frequenz erstreckt, die in einem Signal vorkommen können.

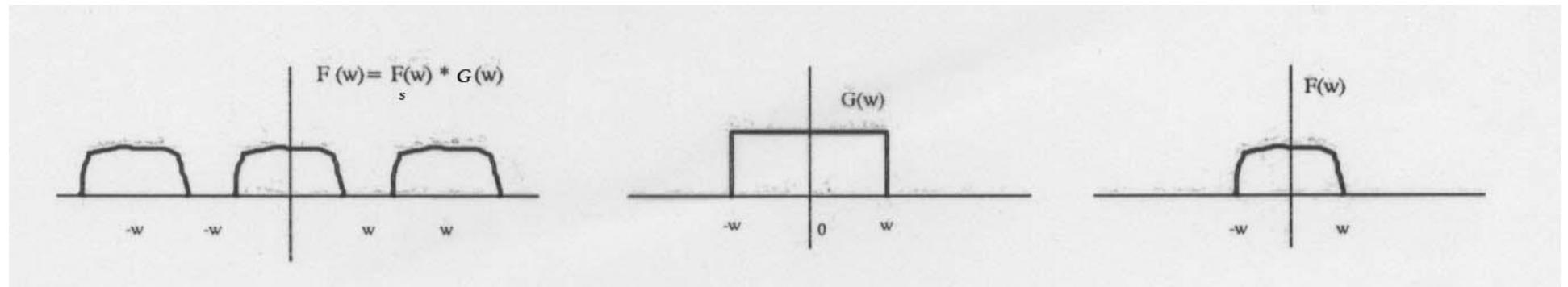
z. B. Audiofrequenzbandbreite, Videofrequenzbandbreite, usw.



(Frequenz-)Bandbegrenzung

Das Eingrenzen der Frequenzbandbreite eines Signals mit Hilfe von sog. Filter.

z.B.: Tiefpaßfilter, Bandpaßfilter, usw.



Abtastung

- Das Abfragen der Amplitudenwerte eines Signals in bestimmten Zeitabständen wird **Abtastung** genannt.
- Diese Amplitudenwerte (sog. Abtastwerte) können in regelmäßigen oder unregelmäßigen Zeitabständen abgefragt werden.
- Aufgrund der Einfachheit in der mathematischen Beschreibung entscheidet man sich für eine **regelmäßige Abtastung**.

Fragen:

- Wie oft soll ein (analoges) Signal abgetastet werden?
- Was möchte man mit dem abgetasteten Signal erreichen?

Antworten:

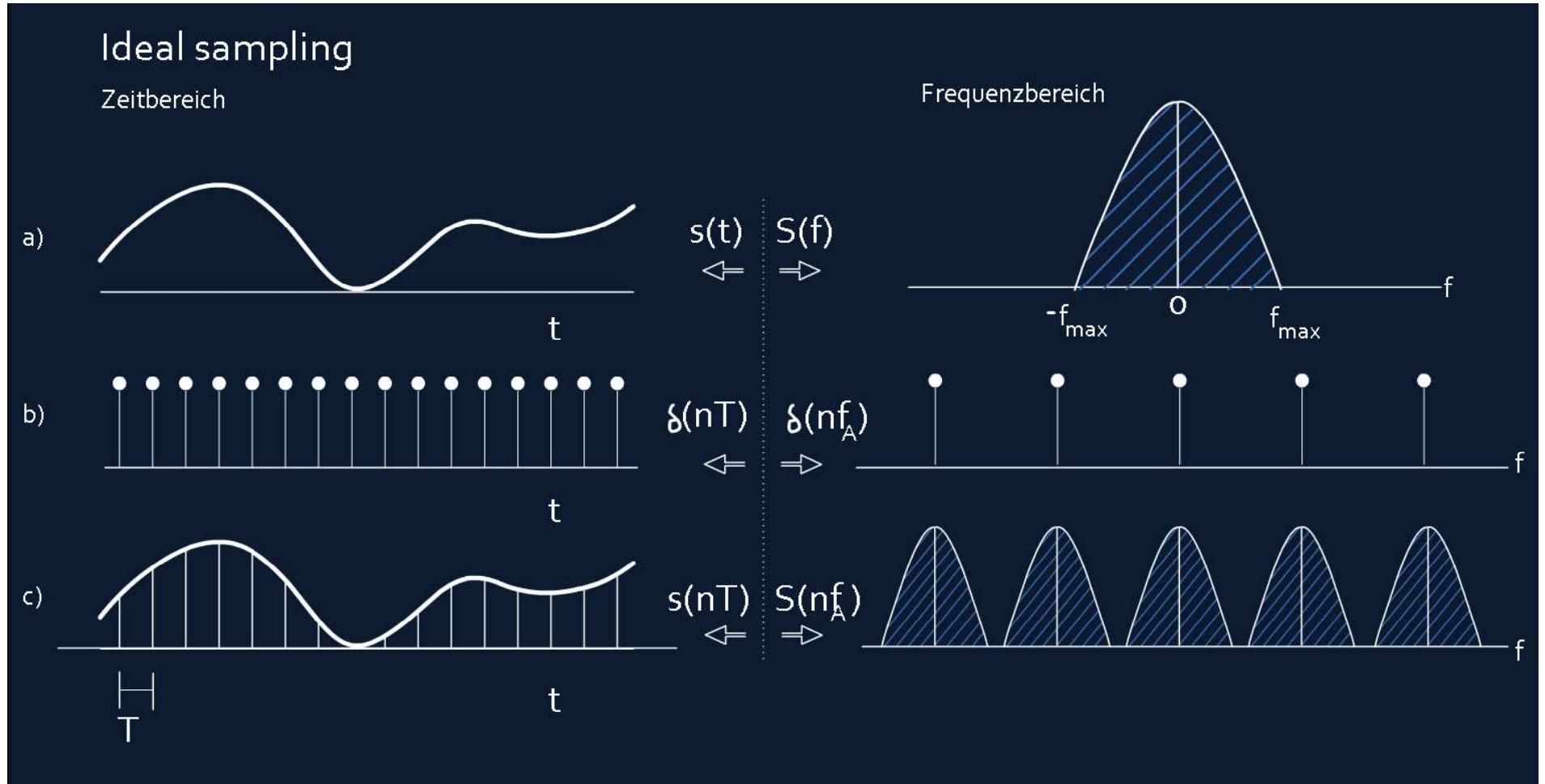
- Das ursprüngliche (analoge) Signal soll rekonstruiert werden.
- Die Anzahl der Abtastwerte soll möglichst gering gehalten werden (geringe Datenmenge).

Lösung:

Das Abtasttheorem

$$f_A \geq 2 \cdot f_{MAX}$$

Ideale Abtastung



Aliasing

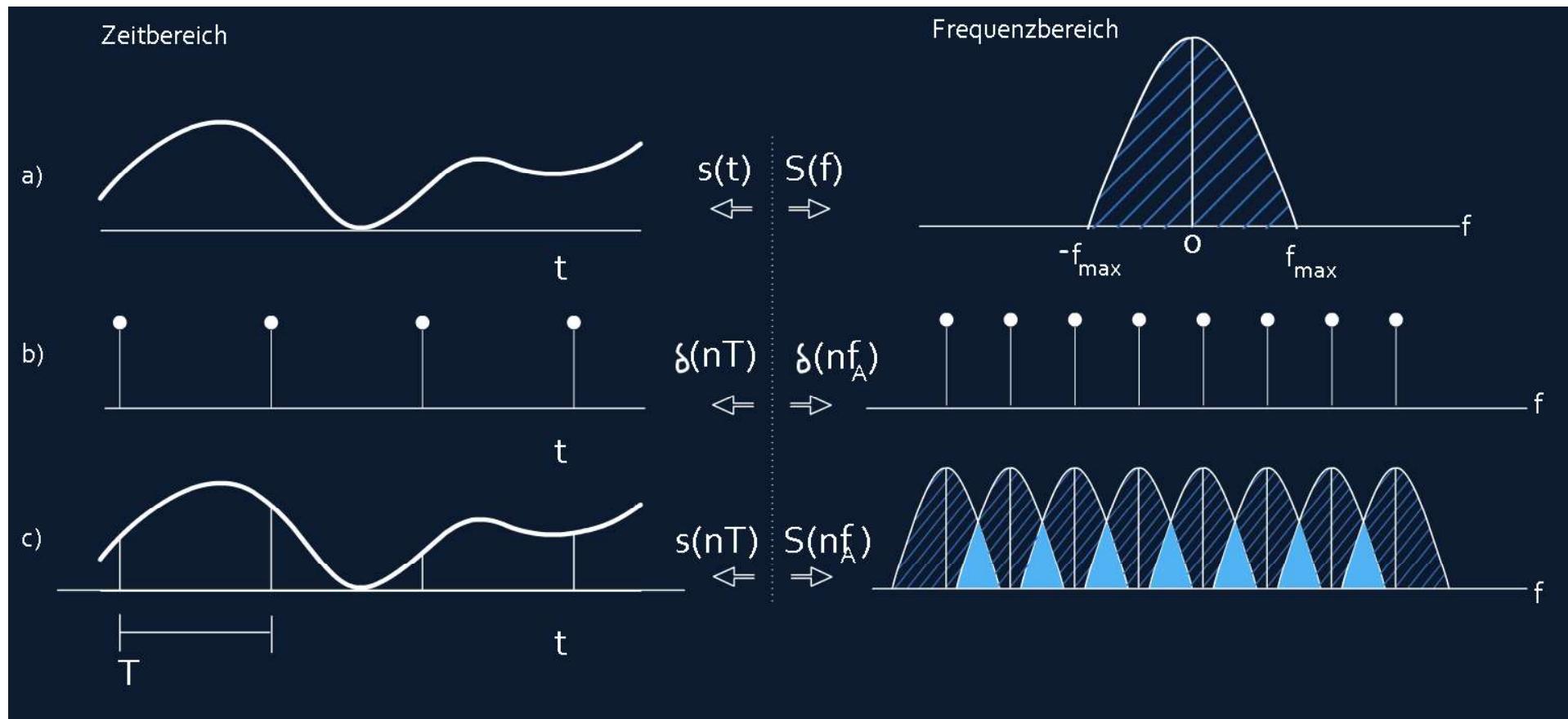
Wird bei der Abtastung eines analogen Signals das Abtasttheorem nicht eingehalten, tritt den sog. Alias-Effekt auf.

Alias-Effekte sind Signalverzerrungen und –Verfälschungen, die bei der Abtastung eines analogen Signals vorkommen können.

Der Alias-Effekt kommt vor, wenn:

- das analoge Signal nicht bandbegrenzt ist
- die Abtastrate zu niedrig ($f_A < 2 \cdot f_{MAX}$) ist

Unterabtastung (*undersampling*) und ihrer Folge: Aliasing



Quantisierung

Beim Digitalisieren eines analogen Signals müssen zwei Informationsdimensionen erhalten bleiben:

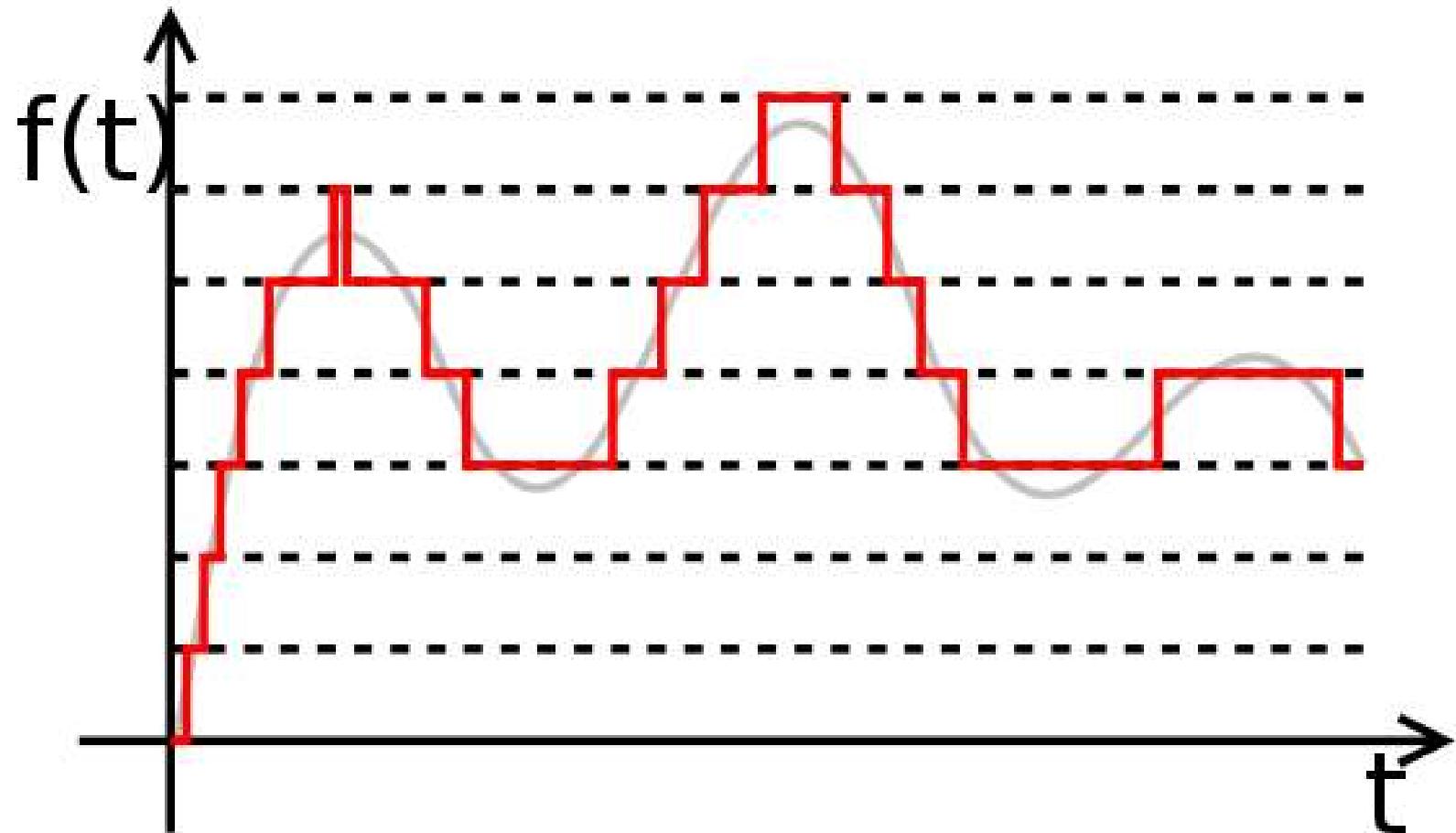
1. die zeitliche Information (sie wird implizit durch die Abtastung erhalten)
2. die Information über die Amplitude des Signals

Die **Quantisierung** wird definiert als die gemessene Amplitude des analogen Signals beim Zeitpunkt der Abtastung.

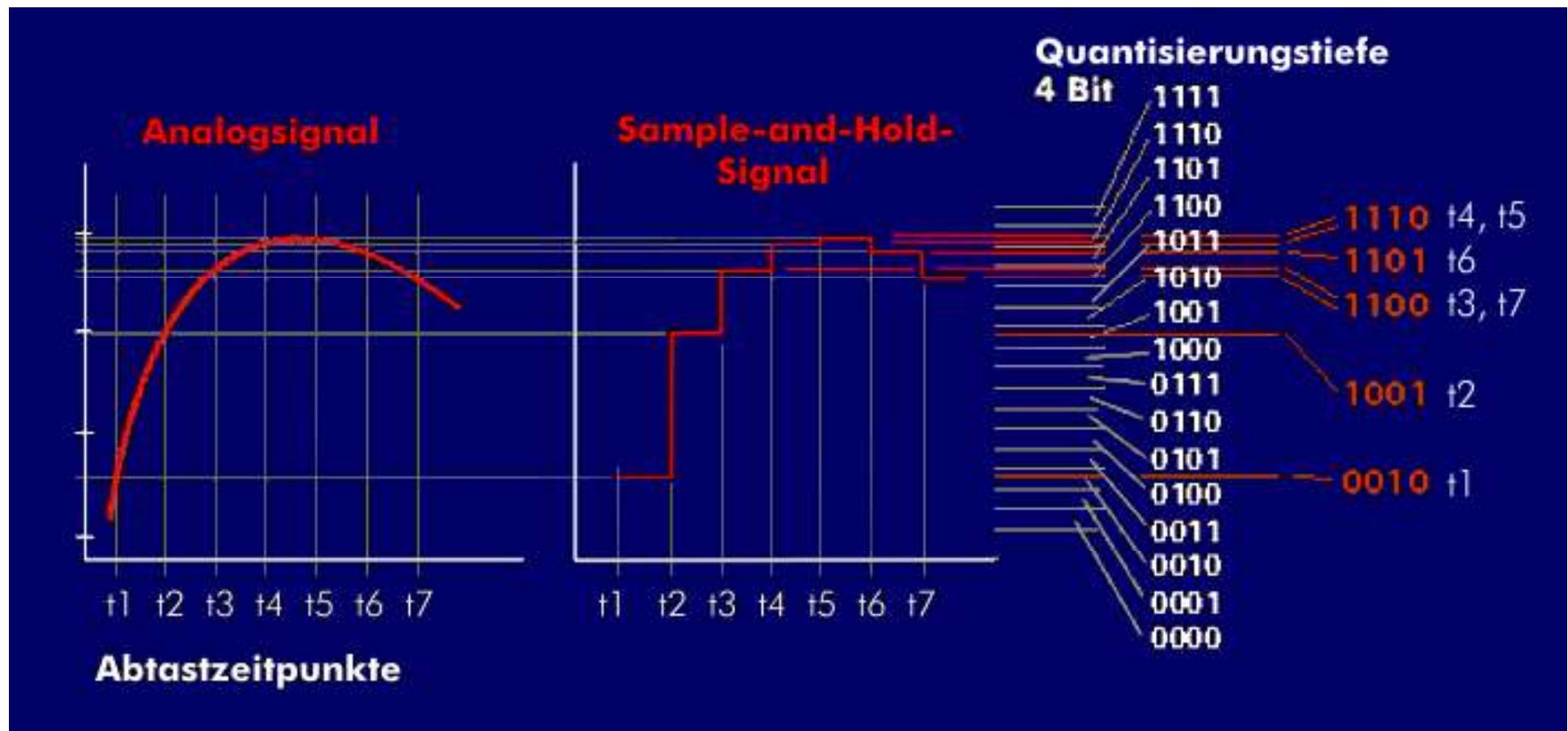
Die Quantisierung wird durch die sog. **Systemauflösung** (Anzahl der zur Verfügung stehenden Quantisierungstufen) begrenzt.

Bei einer n-Bit Auflösung stehen dann: 2^n Stufen zur Verfügung.

Quantisierung des Signals $f(t)$ mit einer 3 Bit Auflösung (8 Quantisierungsstufen)



A/D-Umwandlung mit einer 4-Bit Quantisierungstiefe



Durch die Systemauflösung bedingt wird bei der Quantisierung einen Fehler erzeugt. Dieser Fehler wird **Quantisierungsfehler** genannt.

Da dieser Fehler ähnlich zum Rauschen im analogen Systemen ist, wird er auch **Quantisierungsrauschen** genannt.

Der **Signal-Rausch-Abstand** $[S/N]_{dB}$ bzw. SNR bei einer gleichmäßigen verteilten linearen Quantisierung ist:

$$[S/N]_{dB} = 6,02 \cdot n + 1,76 \text{ dB}$$

$n \equiv$ Bitauflösung bzw. Bitanzahl

oder als Faustregel

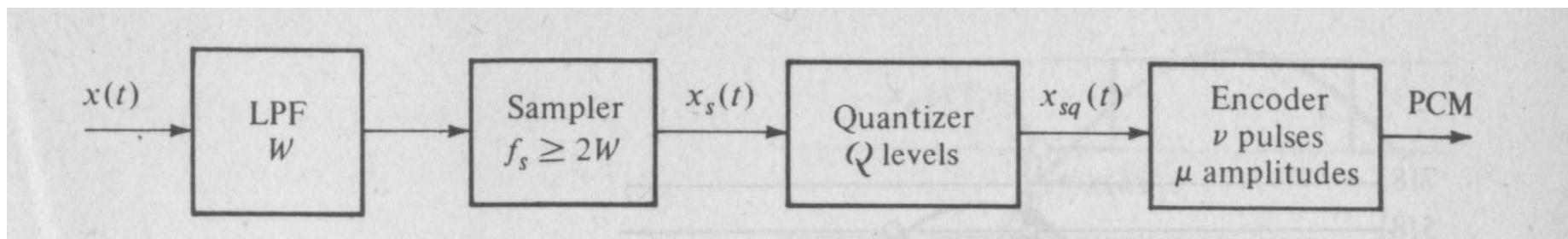
$$[S/N]_{dB} = 6 \cdot n \text{ dB}$$

Codierung

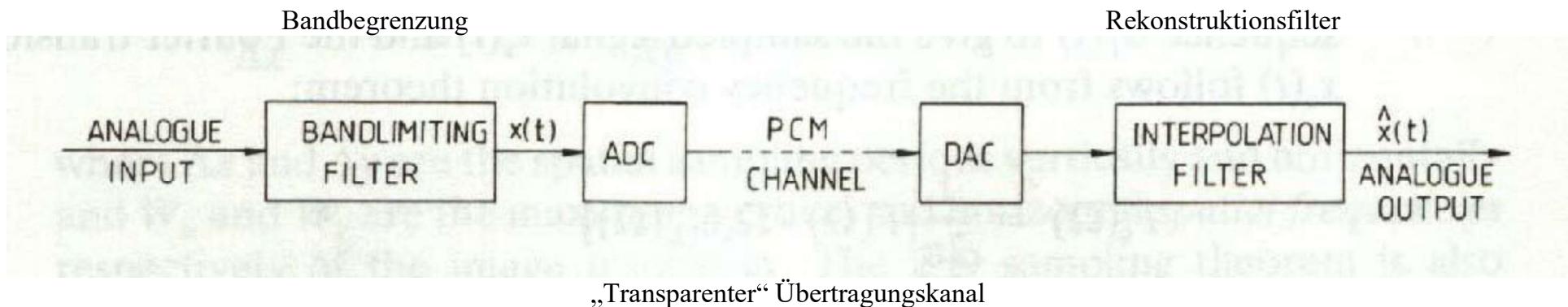
Die gewonnenen abgetasteten und quantisierten Werte des analogen Signals sollen optimal „verpackt“ werden, um möglichst Fehlerfrei übertragen und gespeichert zu werden. Der entsprechenden angesetzten Zuordnungsvorschrift wird Codierung genannt.

Die PCM-Codierung hat sich als die optimale digitale Repräsentation für abgetastete Audio- und Videosignale erwiesen.

Das PCM-Codierungssystem



Das vollständige PCM-Übertragungssystem



A/D-Umwandlung

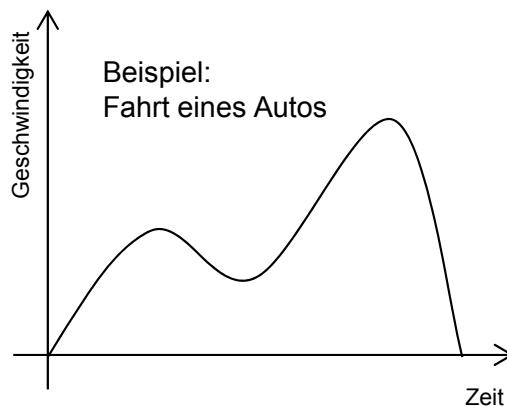
0. **(Frequenz-)Bandbegrenzung**
(Tiefpaß- bzw. BandpaßFilterung; f_{\min} , f_{MAX})
1. **Abtastung** (Abtasttheorem: $f_A \geq 2 \cdot f_{\text{MAX}}$ bzw. $f_A > 2 \cdot f_{\text{MAX}}$)
2. **Quantisierung** (n Bit [Quantisierungstiefe]; $[\text{S}/\text{N}]_{\text{dB}}$)
3. **Codierung** (PCM)

Einleitung

EINSTIEG: Was bedeutete eigentlich nochmal DIGITAL ?

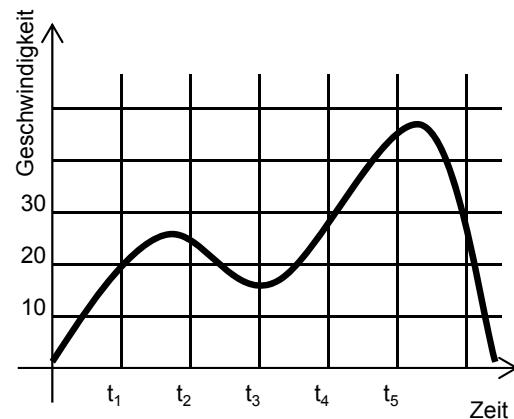
s.a: Audiotechnik im Grundstudium

- Digitalisierung nennt man die Erfassung von zeitkontinuierlichen (analogen) Vorgängen als Zeit- und Werte-diskrete Zahlen¹ und die Umwandlung dieser diskreten Zahlen in einen Datensatz (engl. File)².



Zum Zeitpunkt [sec]:	10 20 30 40 50 ...
Geschwindigkeit [km/h]:	20 25 17 35 50 ...

1



2

Datensatz "Fahrt des Autos":

20, 25, 17, 35, 50,

Daten von links nach rechts lesen.

Zeitbasis: 10 sec.

- Abschließend wird der Datensatz in Binärzahlen umgewandelt.

Binär-Datensatz "Fahrt des Autos":

(20)	(25)	(17)	(35)	(50)
00010100,	00011001,	00010001,	00100011,	00110010,

bit Byte/Oktett word Takt: 10 sec.

- bit = Binary Digit = Binärzeichen
- Byte = kleinstes Datenwort, bestehend aus 8 Binärzeichen (1Byte = 8bit); Oktett
- word = Typdefinition eines logischen Datenwertes (int, float, long, etc.)

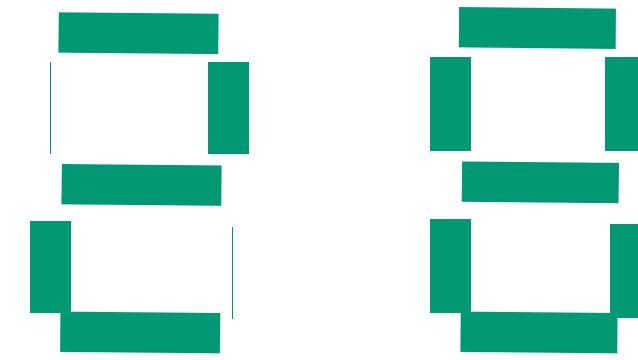
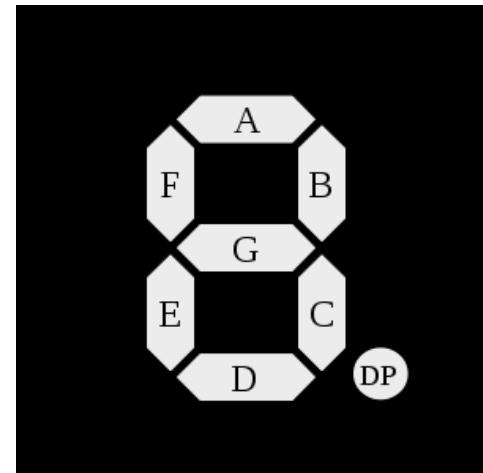
Natürlich lassen sich nach dem gleichen Verfahren auch AV-Signale in "Datensätze" umwandeln (Analog -> Digital - Wandlung).

Exkurs I BCD, EBDIC und die Erfindung des BYTE:

Mit der Definition von IBM im Jahre 1963, acht bit zur kleinsten adressierbaren Dateneinheit 1Byte zusammenzufassen, konnten Ziffern erstmalig optimal prozessiert und gespeichert werden. Diese Lösung wurde Basis für den IBM-EBDIC*-Maschinencode. Danach setzte sich der Sammelbegriff „1 Byte = 8 binary digits“ weltweit durch.

Heute wird in der Digitalen AV-Technik und Netzwerktechnik die präzisere Bezeichnung **OKTETT** verwendet, welcher darauf hinweist, daß mit der REIHENFOLGE („order“) der bits ein WERT verknüpft ist.

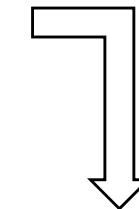
Wir verwenden in dieser Veranstaltung die Leserichtung von rechts nach links, also kleinste Stelle zuerst und höchste Stelle zuletzt („Big Endian“). Die höchste Stelle kann alternativ auch von einem Vorzeichen besetzt sein, dann gilt: „führende 0 links“ = VZ „+“ bzw. „führende 1 links“ = VZ „-“



Datenmenge: 1 Byte
 Leserichtung: ←
 Stelle bedeutet: $2^7 \ 2^6 \ 2^5 \ 2^4 \ 2^3 \ 2^2 \ 2^1 \ 2^0$
 Wert: **218_{dez}** = $128 + 64 + 0 + 16 + 8 + 0 + 2 + 0$

Binäre Datentypen zur Abspeicherung von Abtastwerten:

Zahlenraum	Name	Datenwortbreite [bit]	vorzeichenfähig?
Ganzzahlige Werte	byte	8	unsigned
			signed ±
	integer	16 (short)	unsigned
			signed ±
		32 (int)	unsigned
			signed ±
		64 (long)	unsigned
			signed ±
Gleitkomma-Werte in Mantisse/Exponent-Darstellung	half	16	signed ±
	float	32	signed ±
	double	64	signed ±



Bei allen IEEE-754-Gleitkommatypen gibt das höchstwertige Bit das Vorzeichen der Zahl an; 1 bedeutet minus. Die auf das Vorzeichen folgenden Bits enthalten den Exponenten; bei 16-Bit umfasst er 5 Bit. Gespeichert ist er als vorzeichenlose Binärzahl mit Offset. Dieser auch Bias genannte Wert beträgt in diesem Format 15 und bedeutet, dass man von dem gespeicherten Wert 15 abziehen muss, um auf den tatsächlichen Exponenten zu kommen. Mit fünf Bit lassen sich die Zahlen 0 bis 31 darstellen, wobei 0 und 31 als Sonderfälle betrachtet werden. Vermindert man die Zahlen von 1 bis 30 jeweils um 15, kommt man auf einen Exponentenbereich von -14 bis 15.

Gleitkommazahlen nach IEEE 754-2008

Größe	Größe Mantisse	Größe Exponent	minimaler Exponent	maximaler Exponent	signifikante Dezimalstellen	kleinste (normalisierte) Zahl	größte Zahl
Binärformate							
16 Bit	10 Bit	5 Bit	-14	15	3 bis 4	$\pm 6,104 \cdot 10^{-5}$	± 65504
32 Bit	23 Bit	8 Bit	-126	127	7 bis 8	$\pm 1,175 \cdot 10^{-38}$	$\pm 3,403 \cdot 10^{38}$
64 Bit	52 Bit	11 Bit	-1022	1023	15 bis 16	$\pm 2,225 \cdot 10^{-308}$	$\pm 1,798 \cdot 10^{308}$

„Und was ist nun so toll an DIGITAL?“

Binäre Daten (= Folgen von 0/1-Zuständen) können im Gegensatz zu analogen Signalen

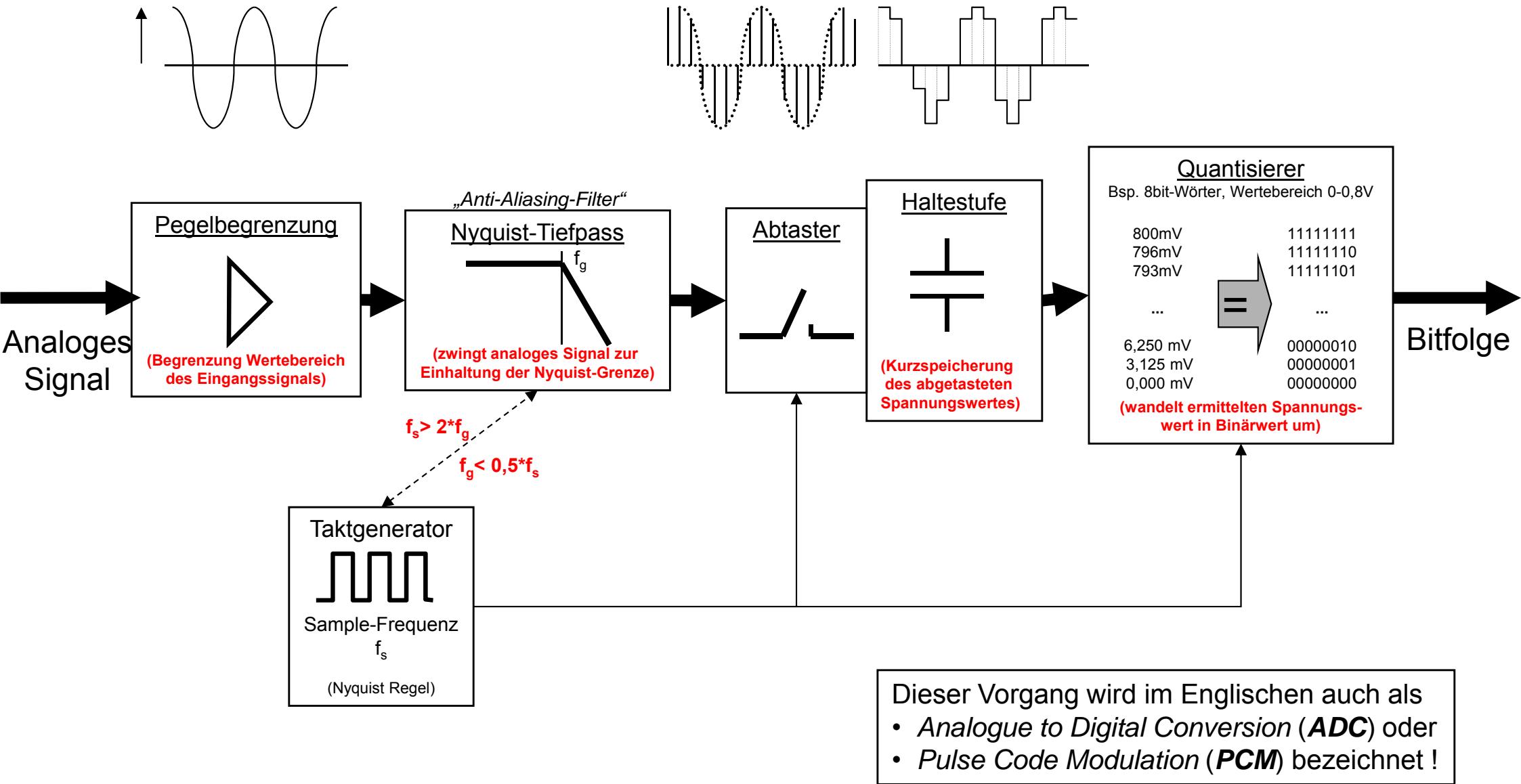
- relativ einfach auf verschiedenen Medien (“Datenträgern”) gespeichert werden
(-> Aufnahme),
- mit Rechnern beliebig manipuliert werden
(-> Bearbeitung),
- beliebig oft kopiert werden,
- auch unter “schlechten” Empfangsbedingungen übertragen werden
(-> Sendung),

... ohne dass es zu Qualitätsverlusten kommt !*

* ACHTUNG:
sofern der Prozess der A/D-Wandlung korrekt durchgeführt wurde!

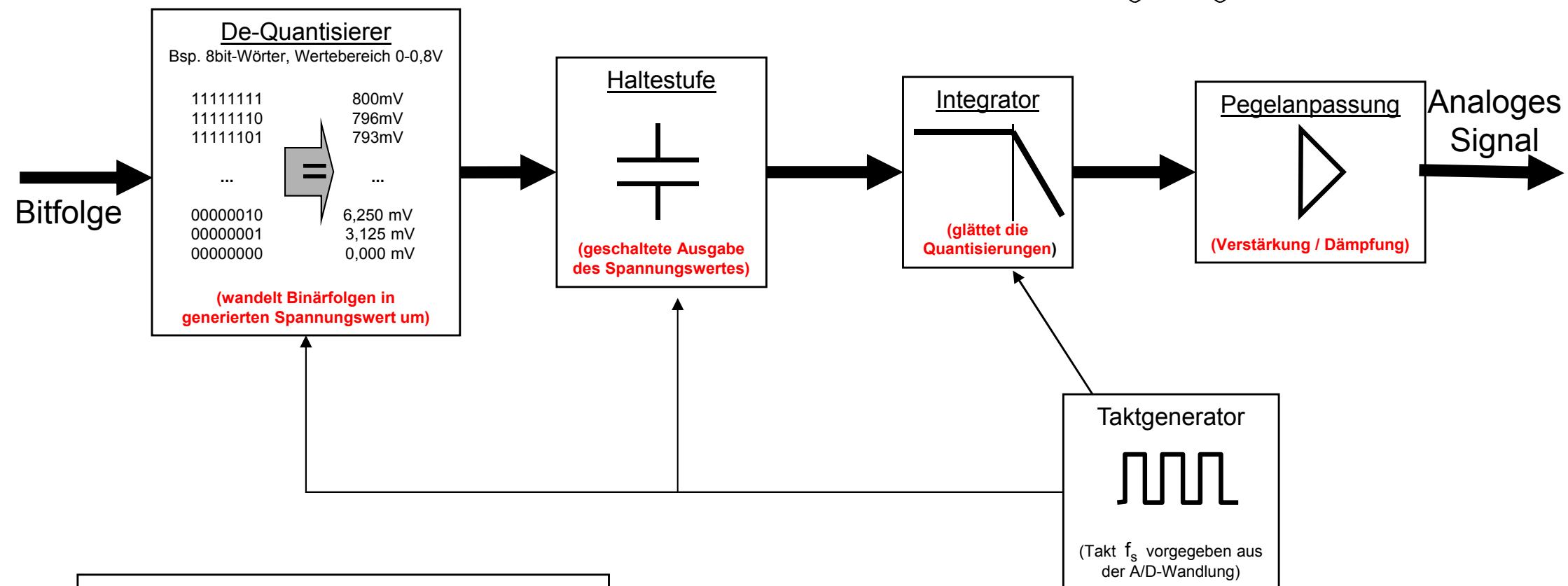
Schaltungstechnisches Prinzip eines Analog/Digital - Wandlers:

s.a: Audiotechnik im Grundstudium



Schaltungstechnisches Prinzip eines Digital/Analog - Wandlers:

s.a: Audiotechnik im Grundstudium



Dieser Vorgang wird im Englischen als
 • *Digital to Analogue Conversion (DAC)*
 bezeichnet!

Allgemeines mathematisches Prinzip eines A/D - Wandlers (I):

s.a: Audiotechnik im Grundstudium



I. **Nyquist-Regel**:

- um ein Signal unverfälscht zu digitalisieren, muss es mindestens mit dem doppelten Takt der höchsten, auftretenden Frequenz abgetastet werden!

Beispiel: **Audio**

Höchste hörbare Frequenz eines Audiosignales (für Menschen): $\approx 20 \text{ kHz}$

International standardisierte Abtastraten:

$f_s = 44,1 \text{ kHz}$ (Audio-CD)

$f_s = 48 \text{ kHz}$ (Audio in Bild/Ton-synchronen AV-Produktionen)

$f_s = 96 \text{ kHz}$ (Tonstudio)

II. Quantisierungs-Regeln:

- Je größer die möglichen Amplitudensprünge, die ein Signal zwischen zwei Abtastzeiten durchführen kann, desto größer muß der Wertebereich des Quantisierers gewählt werden, sonst kommt es zu Verzerrungen! („clipping“)
- Je exakter ein abgetastete Wert binär erfasst / gespeichert werden soll, desto größere Datenwörter müssen gebildet werden (Quantisierung), sonst kommt es bei der D/A-Wandlung zu Rausch- oder Treppen-Effekten durch ungenaue Rundung („Quantisierungsrauschen“)!

Zusammenhang zwischen Datenwortgröße und linearer Signaldynamik:

grobe Abschätzungsformel:

$$\text{Maximal abbildbare Dynamik in Digitalsystemen [dB]} \approx \text{Größe Datenwort} * 6 + 2$$

Man bezeichnet diesen Wert auch als den theoretisch maximal realisierbaren Signal/Rausch-Abstand, engl. Maximum Signal to Noise-Ratio (SNR_{MAX} oder S/N_{MAX})

Standard Datenwortbreiten Audio / Video:

Beispiel: **Audio**

Dynamik-Fähigkeit unseres auditiven Systems (von Hörschwelle bis Hörschmerzen): → **120dB**

International standardisierte Audio-Datenwortgrößen:

$Q = 16 \text{ bit}$ (Audio-CD, DVD, BluRay) $\approx 98 \text{ dB}$

$Q = 24 \text{ bit}$ (Tonstudio) $\approx 146 \text{ dB}$

Beispiel: **Video**

Wieviele unterschiedliche “Stufen” können *gleichzeitig (auf einen Blick)* wahrgenommen werden:

Helligkeit: ca. 1024 Stufen $\equiv 62 \text{ dB}$ *

Chroma + Helligkeit: ca. 256 Stufen pro gammakorrigiertem FARB-Kanal (R,G,B) $\equiv 50 \text{ dB}$ **

Standardlänge Datenwörter Video/digital: Q = 8 o. 10 bit

* **Achtung:** 1. Kompromiss-Wert: Spontane Detailwahrnehmung de-facto 44dB = 128 Stufen. Maximum (nach Adaption durch Iris und Rezeptoren) de-facto 75-80 dB.
2. Diese Stufen gehen davon aus, dass vorab (z.B. in der Kamera) schon eine nichtlineare Gammakorrektur stattgefunden hat. **Siehe auch Folien 29-35.**

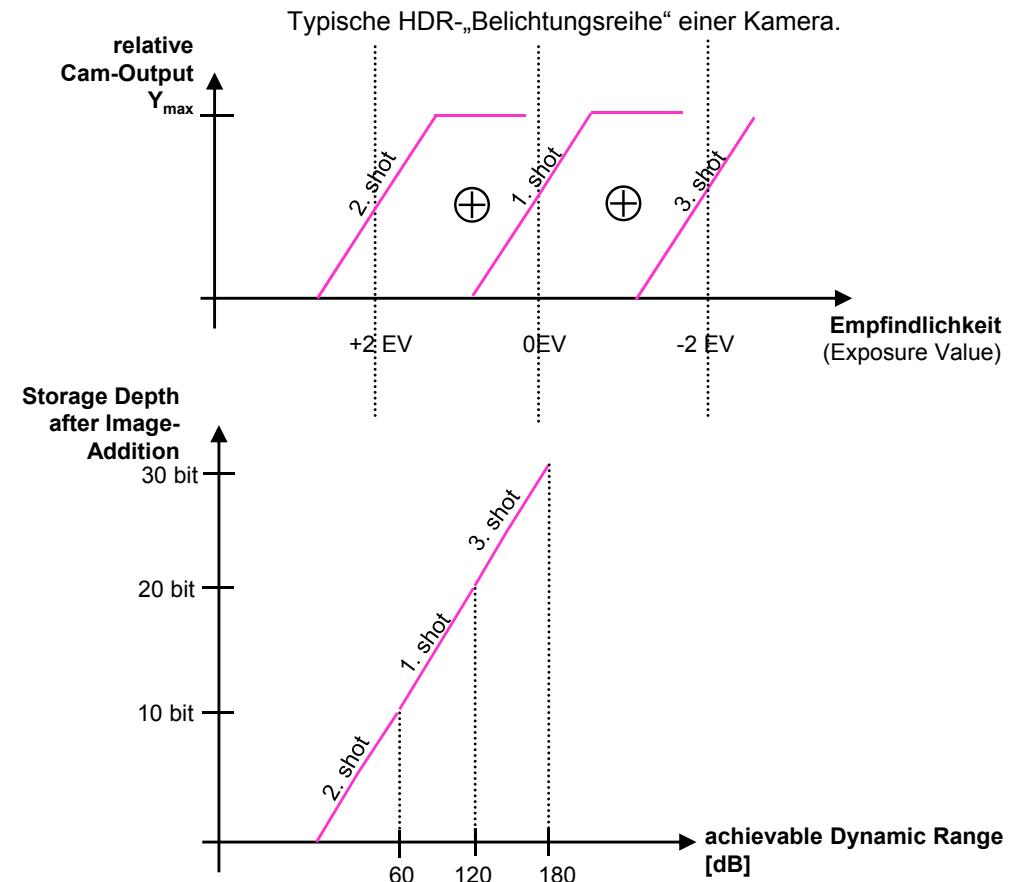
** **Achtung:** eine Farbe = Mischergebnis R+G+B. Jede Primärfarbkomponente muß also mit dieser Quantisierung erfasst werden.
Dies ergibt (256 x 256 x 256 =) 16,7 Mio. unterscheidbare Farbwerte = 24bit Gesamt-Speichertiefe pro erzeugtem Farbpunkt!

Extrem-Quantisierung bei HDR/raw bzw. VFX/Postproduction („scene linear“):

Eine noch höhere Video-Speichertiefe von $Q > 30$ bit benötigen unkomprimierte HighDynamicRange-Bilder (HDRI), bei denen ein Bild z.B. aus drei Belichtungen aufsummiert wird, um den maximal-möglichen Dynamikumfang des Auges (ca. 160dB) unkomprimiert aufzuzeichnen.

Abgespeichert werden diese unkomprimierten HDR-Abtastwerte derzeit meist im **DPX**-Format (Digital Picture Exchange) als 32bit float-Werte oder im **OpenEXR** 16bit half-float oder 32 bit Format s.a. Folie9.

Um in der Video-Postproduction präzise und einschränkungsfrei arbeiten zu können, wird dort ebenfalls gerne im 32bit-float Format gearbeitet. Dadurch erhält man sich in der Postpro die Möglichkeit, abschließend alle tatsächlichen Fähigkeiten späterer Darstellungsgeräte aus einem Masterformat heraus bedienen zu können (s. a. ACES/Folie 65).



*Anmerkung:

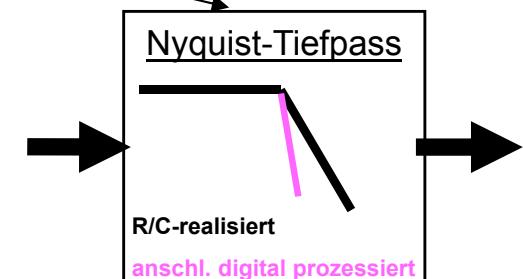
gute LCD-Displays realisieren ca. 60dB,
HDR-Displays ca. 120 dB Dynamic Range.

f_s - Sonderfälle: *Oversampling* und *Upsampling*

Sind spezielle Abtastverfahren, um die **A/D-Wandlung** oder die **D/A-Wandlung** zu optimieren!

Ansatz bei der A/D-Wandlung (*Oversampling*):

- Das analoge, R/C*-basierte Filter am Eingang des A/D-Wandlers (Nyquist-Tiefpass) muß eigentlich von sehr hoher Güte/Flankensteilheit sein, um Klierreffekte (Aliasing) konsequent zu vermeiden.
- Hochqualitative, analoge R/C-Tiefpässe sind aber schaltungstechnisch sehr aufwendig.
- Idee: R/C-Filter bleibt simpel - Signal wird anschließend kurzfristig **überabgetastet** (Abtastrate \gg Nyquistregel).
- Danach findet nochmals eine *digitale Filterung* als einfache mathematische Operation statt.
(s.a. Folie 130)



Vorteil 1: digitale Filter können eine extrem hohe Filtergüte realisieren

Vorteil 2: gleichzeitig wird Rauschen massiv reduziert

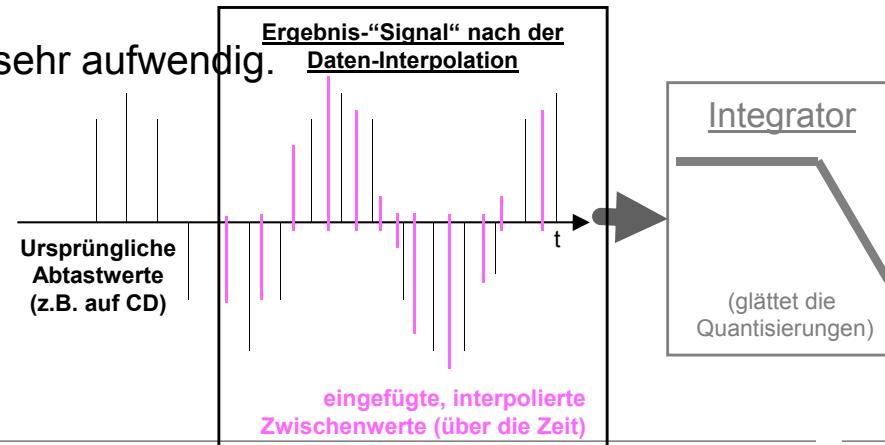
*R/C = Widerstand/Kondensator

Ansatz bei der D/A-Wandlung (*Upsampling*):

- Die analoge, R/C-basierte Integratorschaltung am Ausgang des D/A-Wandlers (Tiefpass) sollte ebenfalls von sehr hoher Güte/Flankensteilheit sein, um die Kanten sauber zu glätten.
- Hochqualitativ, analoge R/C-Integratoren sind aber schaltungstechnisch sehr aufwendig.
- Idee: es werden vom D/A-Wandler interpolierte Zwischenwerte eingefügt (= Reduzierung der Kantensprünge), quasi „Upsampling“.

Vorteil 1: die Integratorschaltung kann äußerst simpel bleiben, ohne dass sich die Klangqualität verschlechtert.

Vorteil 2: Reduzierung des Quantisierungsrauschen



Wahrnehmungsgerechte Quantisierung in *Digitalen Medien*

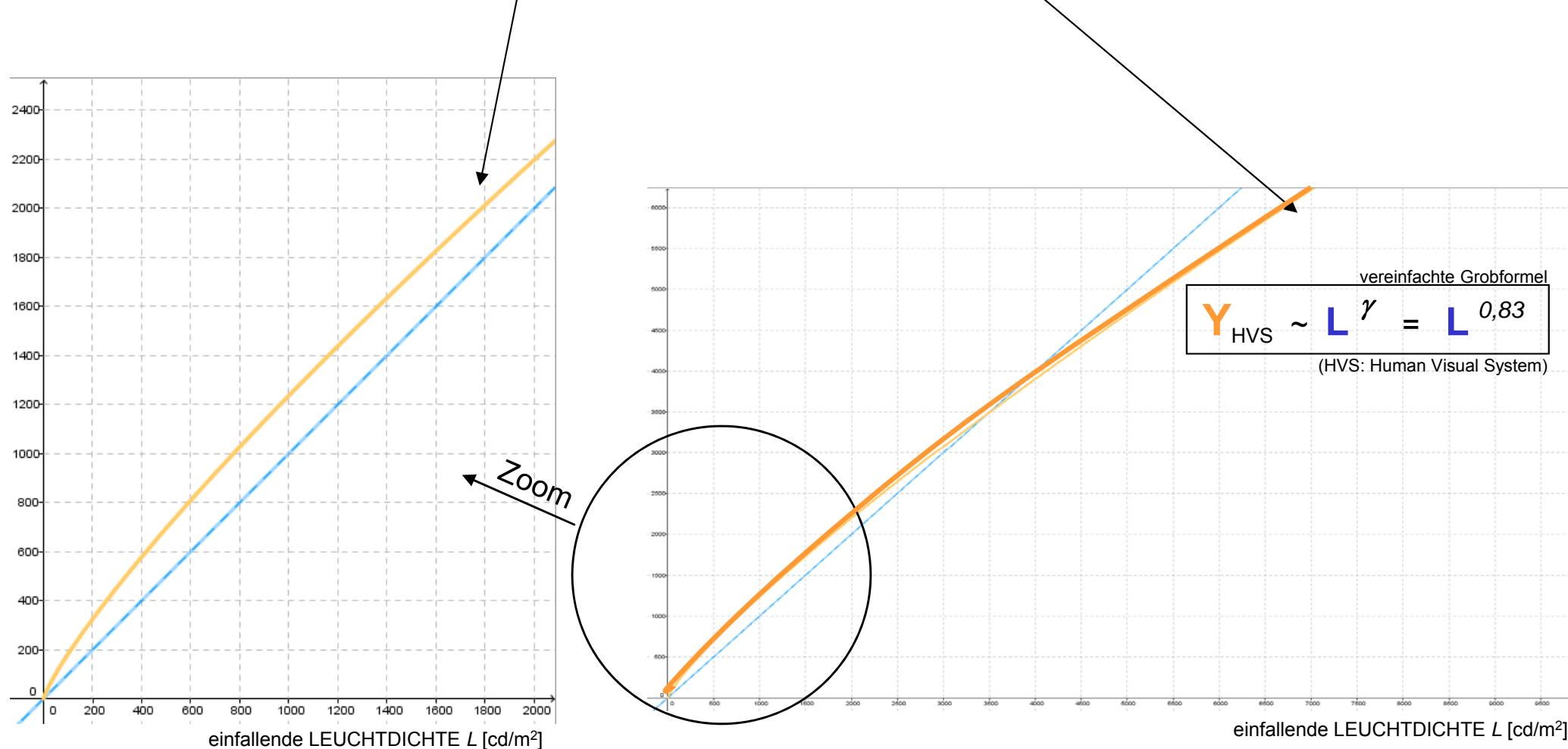
Audio, Video und Computergrafik sind nur für EINE Zielgruppe relevant: den Menschen.

Folglich wird bei der Verbreitung von digitalen Medien
oft gleich „wahrnehmungsgerecht“ quantisiert!

Was war bei Video nochmal

„wahrnehmungsbezogene Gammakorrektur“?:

Die Kontrastwahrnehmung des menschlichen Sehens ist nicht linear sondern eher exponentiell! (Potenzfunktion)
Die empfundene Helligkeit Y steigt in dunklen Bereichen steiler und in stark leuchtenden Bereichen weniger steil an.

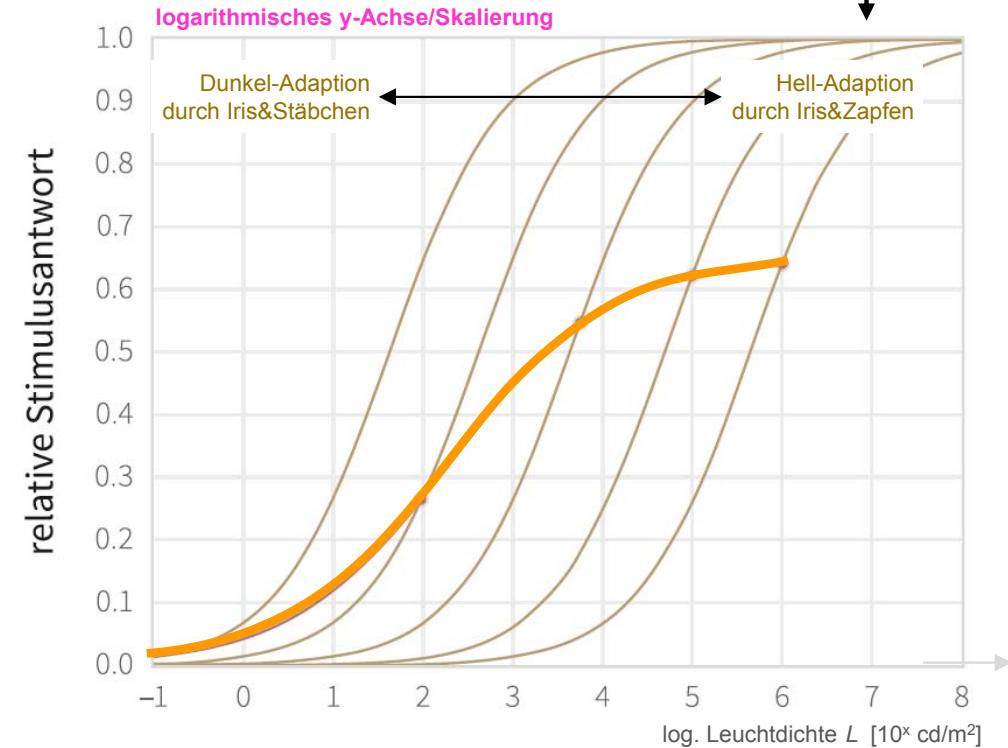
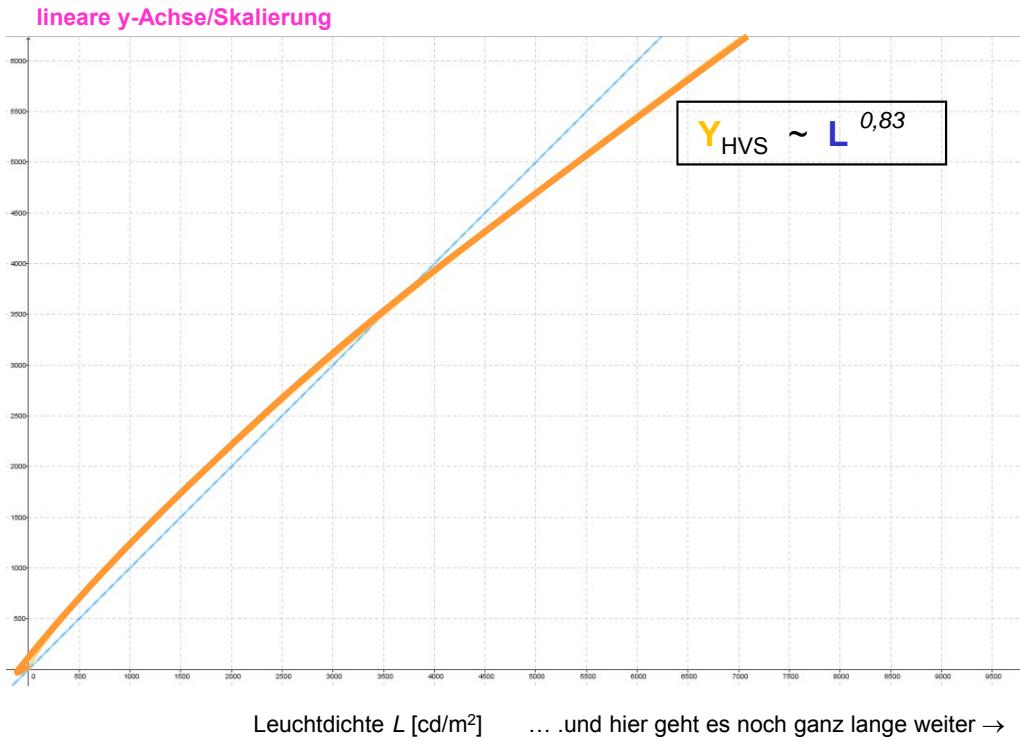


Bildsensoren arbeiten hingegen oft *annähernd* linear (siehe blaue Linie) bezüglich Input zu Output. In der Medienproduktion wird in den Kameras daher intern meist sofort eine „Empfindungs-Korrektur“ durchgeführt, welche als → *wahrnehmungsbezogene Gammakorrektur* bezeichnet wird.

ACHTUNG: nicht reinlegen lassen!

Der gesamte Helligkeits-Wahrnehmungsbereich des Menschen verläuft von ca. 10^{-3} cd/m² bis ca. 10^{+5} cd/m²
 $\rightarrow 160$ dB (100.000.000 : 1).

Aus diesen Gründen wird die Gammafunktion oft auch „relativ über eine logarithmische Achse“ dargestellt.



Beachten Sie die Auswirkungen der Gammakorrektur auf die Werte-Rasterung z.B. bei einer A/D-Wandlung:
 s.a. nachfolgende Folie

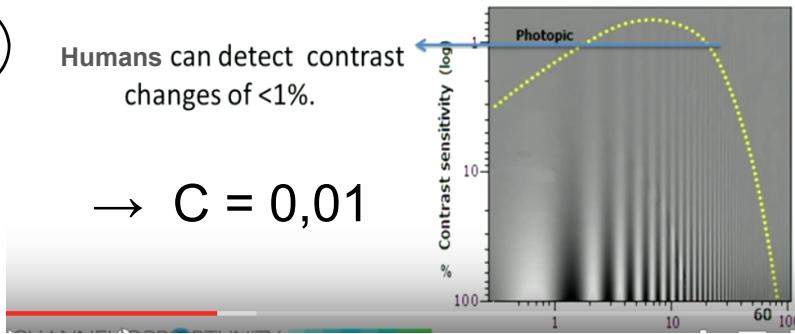
- um die **Sensorwerte** über den gesamten, menschlich-wahrnehmbaren Leuchtdichtebereich korrekt *linear* abzubilden („scene linear“) benötigt man mindestens 16-20 bit pro Datenwort (z.B. Kamerasensor).
- um die **Y**-Werte über den gesamten, menschlich-wahrnehmbaren Leuchtdichtebereich *wahrnehmungsbezogen, relativ* abzubilden benötigt man nur 8-12bit.

Ablauf Rechenschritte optimale Video-Datenwortbreiten für ideale (wahrnehmungsbezogene) Bildqualität:

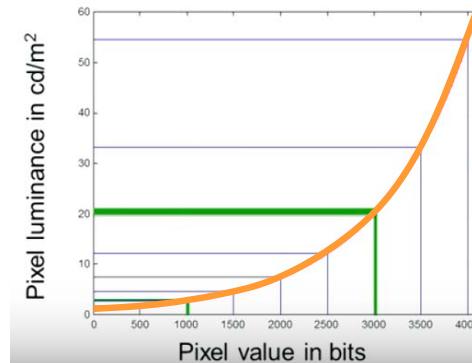
1.

Humans can detect contrast changes of <1%.

$$\rightarrow C = 0,01$$



2.



Weber's law suggests luminance should be an exponential function of pixel value.

A 1-bit change is then equally (in)visible at all levels.

3.

Standard-LCD-Display:

$$L_{\max} = 200 \text{ cd/m}^2, L_{\min} = 0,1 \text{ cd/m}^2$$

$$P = \frac{1}{0,01} \ln \left(\frac{200}{0,1} \right)$$

$$= 760 \approx 2^{9,57} \rightarrow \text{empfohlen: 10bit Datenwortbreite}$$

- How many pixel-values do we need for a 1-bit change to be just invisible?

HDR-Display:

$$L_{\max} = 20000 \text{ cd/m}^2, L_{\min} = 0,005 \text{ cd/m}^2$$

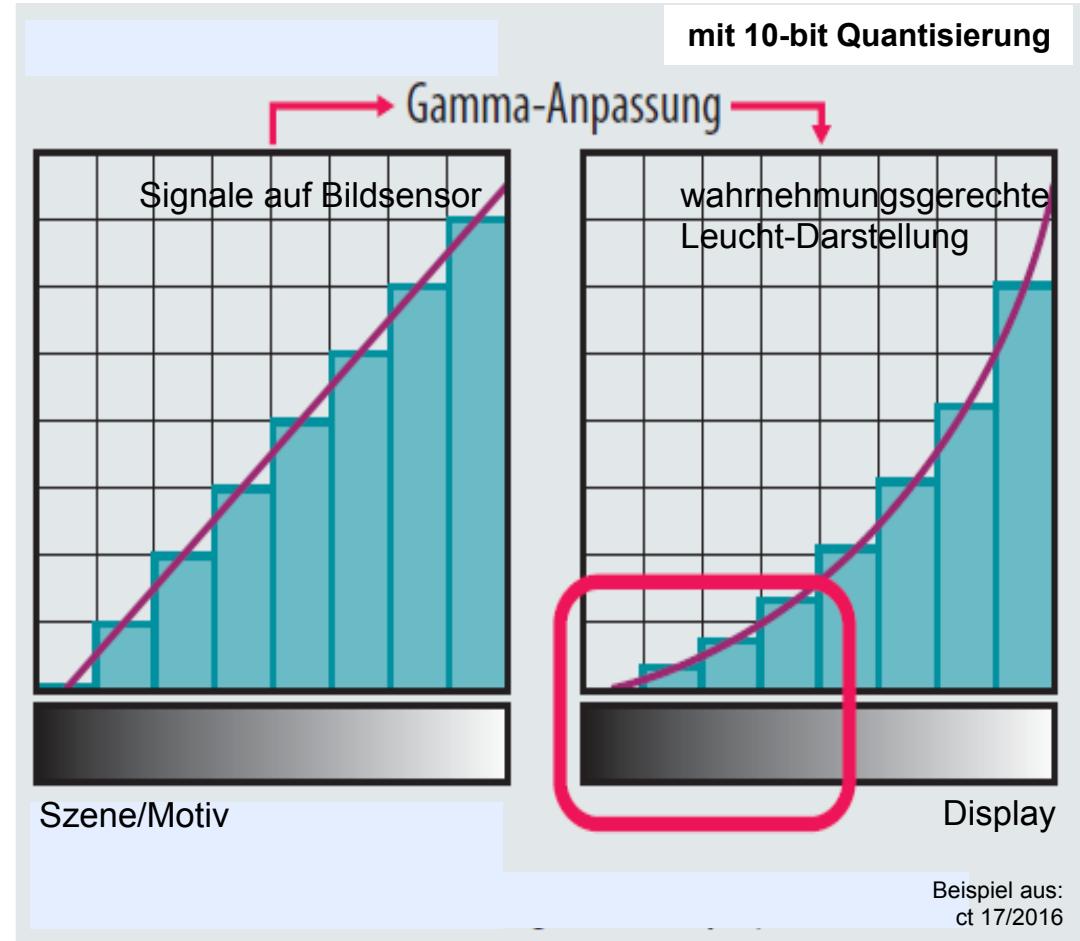
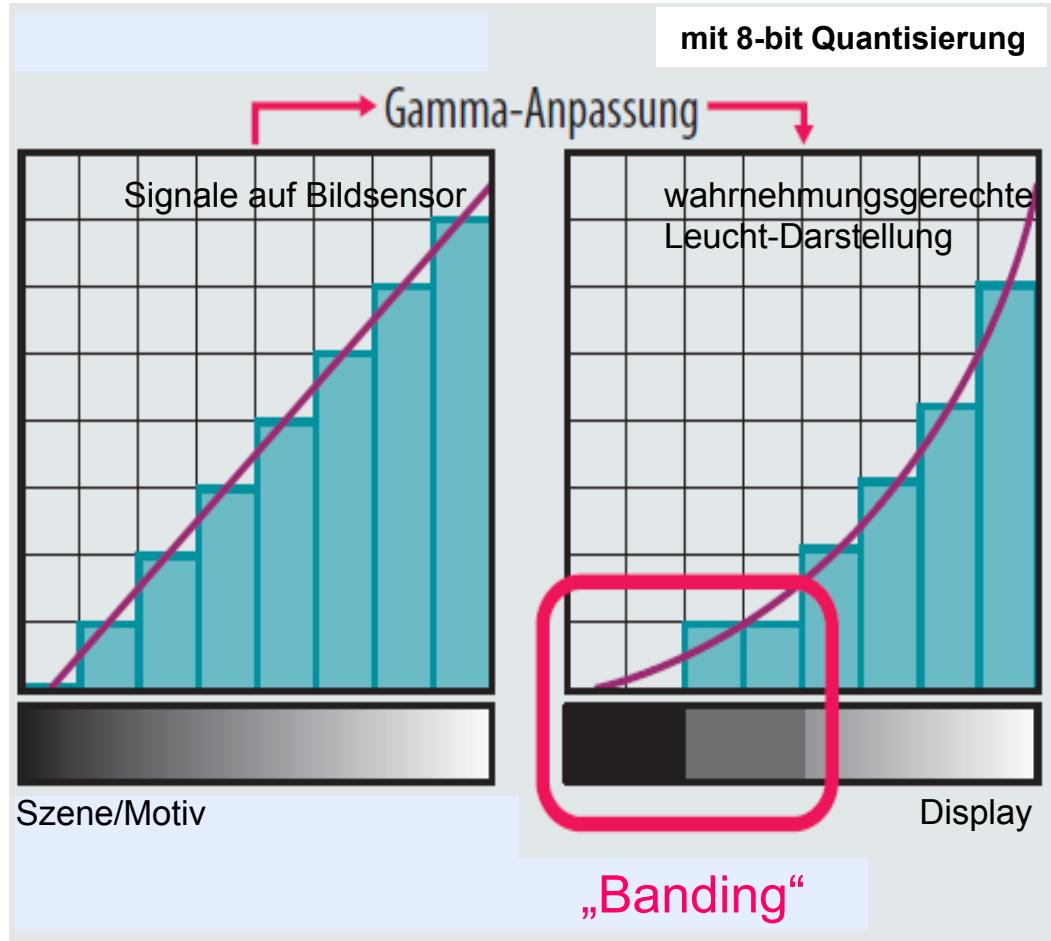
$$P = \frac{1}{0,01} \ln \left(\frac{20000}{0,005} \right)$$

$$= 1520 \approx 2^{10,57} \rightarrow 11 \text{bit Datenwortbreite} \\ (\text{empfohlen: 12bit, bei HDR-Streaming: oft nur 10bit})$$

$$P = \frac{1}{C} \ln \left(\frac{L_{\max}}{L_{\min}} \right)$$

Number of levels per color (color depth)
Contrast threshold ~ 0.01
"Contrast ratio"
Natural logarithm, log to base e

Visuelle Verdeutlichung niedriger Datenwortbreiten:



Exkurs: „10bit-Displays“ in der Praxis

Viele „10-Bit-Monitore“ nutzen kein echtes 10-Bit-Panel, nehmen aber am Digitaleingang Signale mit 10 Bit pro Farbkanal entgegen. Intern verarbeiten sie die Bilder mit 14 oder 16 Bit und übergeben sie anschließend mit 10 Bit Farbtiefe ans 8-Bit-Panel, das daraus per Frame Rate Control (FRC) 1,07 Milliarden Tonwertnuancen erzeugt.

The screenshot shows the appleinsider website with the URL <https://appleinsider.com/article/?id=3910>. The page title is "Apple hit with another "millions of colors" lawsuit". The article is by Slash Lane and published on Monday, March 31, 2008, at 12:00 pm PT (03:00 pm ET). The text discusses Apple being sued for misrepresenting the quality of its notebook screens and iMac displays, using a software technique called 'dithering' to create the illusion of millions of colors.

Apple hit with another "millions of colors" lawsuit

By [Slash Lane](#)

Monday, March 31, 2008, 12:00 pm PT (03:00 pm ET)

Just days after settling a lawsuit in which it was charged with misrepresenting the quality of its notebook screens, Apple Inc. has been hit with a second class-action suit ripe with similar allegations over one of its iMac displays.

According to the new suit, filed in a San Jose court Monday by Kabateck Brown Kellner, LLP, Apple is deceiving consumers by concealing that its new 20-inch iMac monitors are inferior to the previous generation's and those of the new 24-inch iMac.

...

Apple last week agreed to a settlement in a similar class-action lawsuit brought on by two professional photographers, which charged that the company's Intel-based notebooks were only suited to display the "illusion of millions of colors through the use of a software technique referred to as 'dithering,' which causes nearby pixels on the display to use slightly varying shades of colors that trick the human eye into perceiving the desired color even though it is not truly that color."

The terms of that settlement were not made public.

FRC

Frame Rate Control (FRC), is a method of emulation of the missing colors when the color of a pixel is changed slightly with every frame. For example, the monitor has to output the color RGB:{154;154;154}, and the matrix doesn't physically support it, but it supports the two neighboring colors, i.e. RGB{152;152;152} and RGB{156;156;156}. If we were outputting these two colors alternately with the frequency of the refresh rate, the similarity of these colors and the inertia of the human eye (which doesn't perceive flickering at a frequency of 60 hertz) as well as of the matrix itself (which is "smoothing" the moment when the colors are being switched) would give us what our eyes would perceive as some in-between color, i.e. the required RGB:{154;154;154}.

Of course, this is still an emulation, which cannot match the true "true color" reproduction, so such monitors are often specified to output 16.2 millions of colors – if you see this number in the specs, know that you are dealing with an 18-bit matrix. Unfortunately, the text "16.7 millions of colors" doesn't mean anything – many manufacturers write this in the specs of their 18-bit matrices, too. More complex FRC mechanisms are also used in practice, accompanying the more traditional dithering (when the necessary color is formed by several neighboring pixels of slightly different colors), i.e. changing not just the color of a single pixel each frame, but of a group of four pixels, for example. This allows for a more precise reproduction of colors the matrix doesn't support, but the hitch remains the same – such matrices can hardly be considered "full-color" ones.

Typische Gamma-Settings in der Filmproduktion:

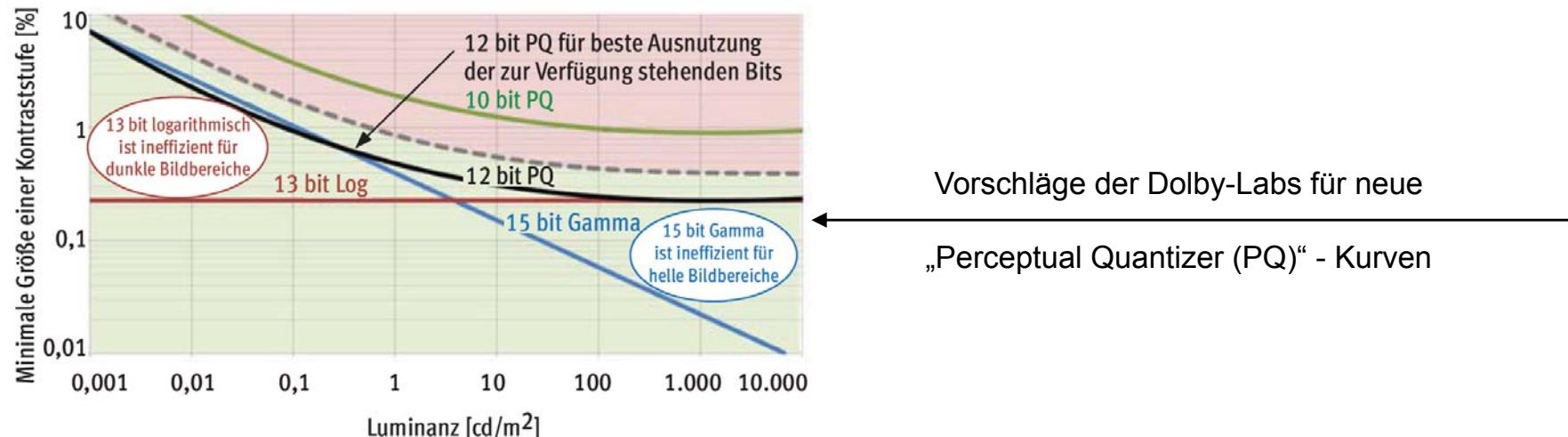
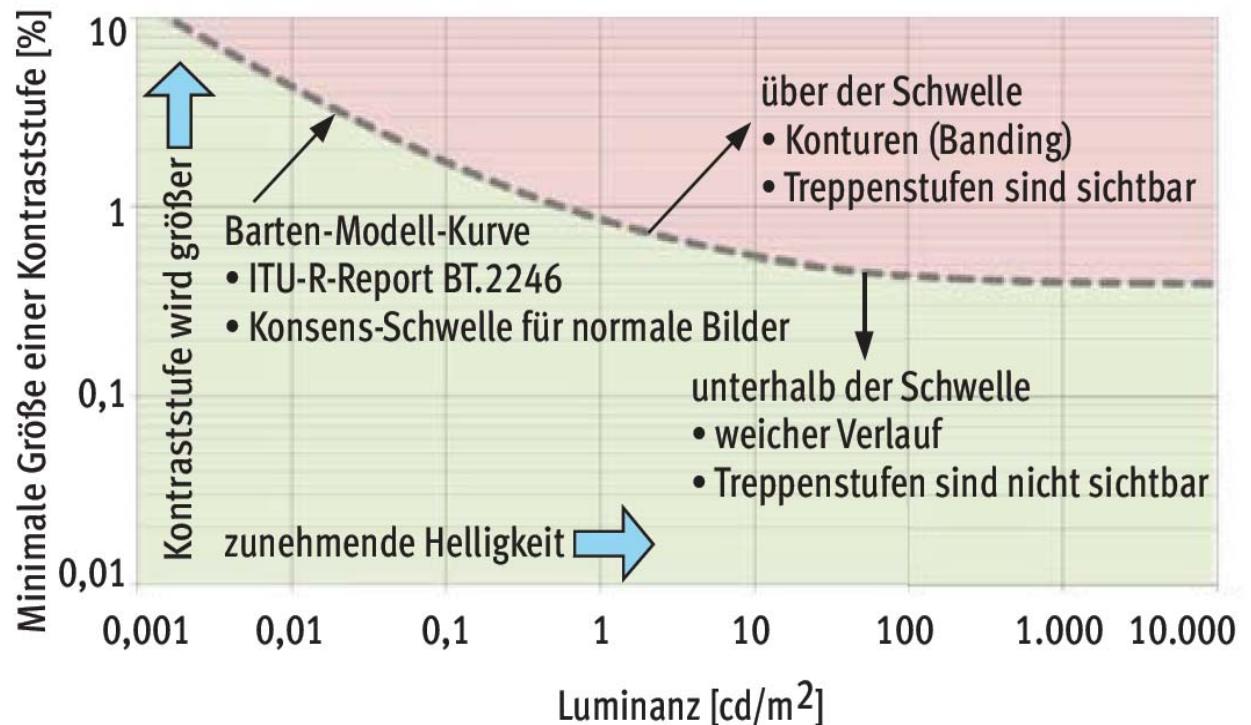
HINWEIS: In der Norm **ITU-R BT.1886** für HD-Displays ist eine Potenzfunktion mit dem Exponenten $\gamma = 2,4$ festgelegt. Im DCI-P3-Standard für Kinoprojektoren ist eine Potenzfunktion mit einem Exponenten von 2,6 vorgeschrieben und im Webstandard sRGB ein Exponent von 2,2.



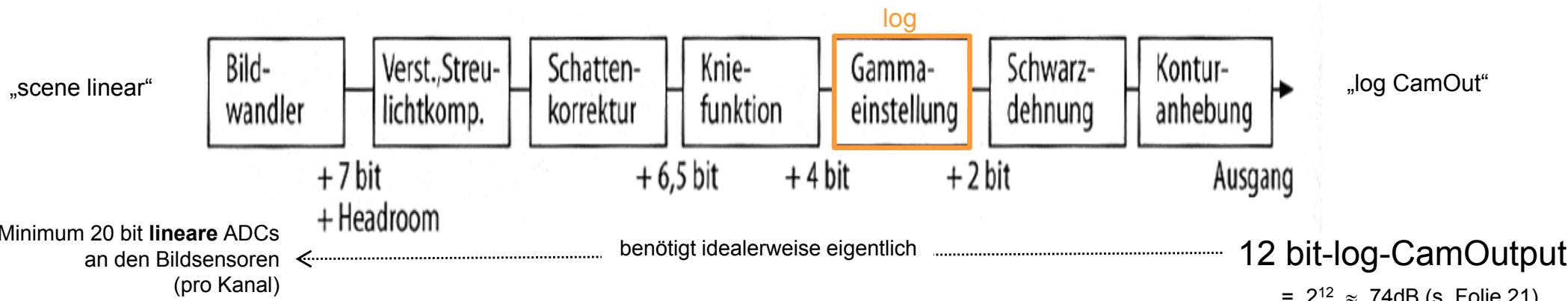
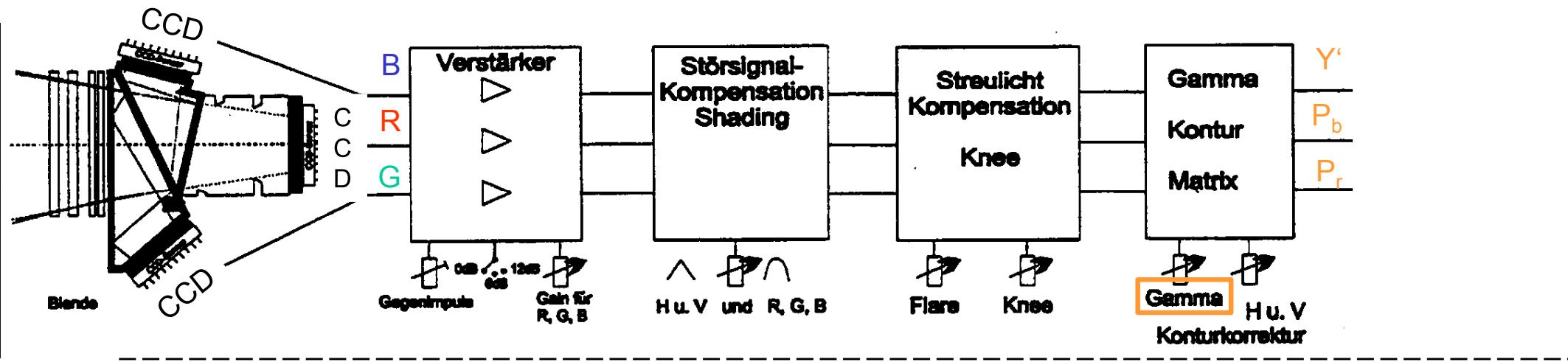
Aktuelle Ansätze für „wahrnehmungs-präzise“ γ –Kurven in HDR-Videos:

Messung der menschl.
Kontrastwahrnehmung
über Grundhelligkeit

(nach Peter G. J. Barten:
"Contrast Sensitivity of the Human Eye and its
Effects on Image Quality", SPIE PRESS / 1999)



Konsequenz: Kamerainterne-Datenwortbreiten



Minimum 20 bit **lineare** ADCs
an den Bildsensoren (pro Kanal)

benötigt idealerweise eigentlich 12 bit-log-CamOutput
= $2^{12} \approx 74\text{dB}$ (s. Folie 21)
≈ 12 abbildbare Blendenstufen bei 2000lx Grundlicht

Minimum 18 bit **lineare** ADCs
an den Bildsensoren (pro Kanal)

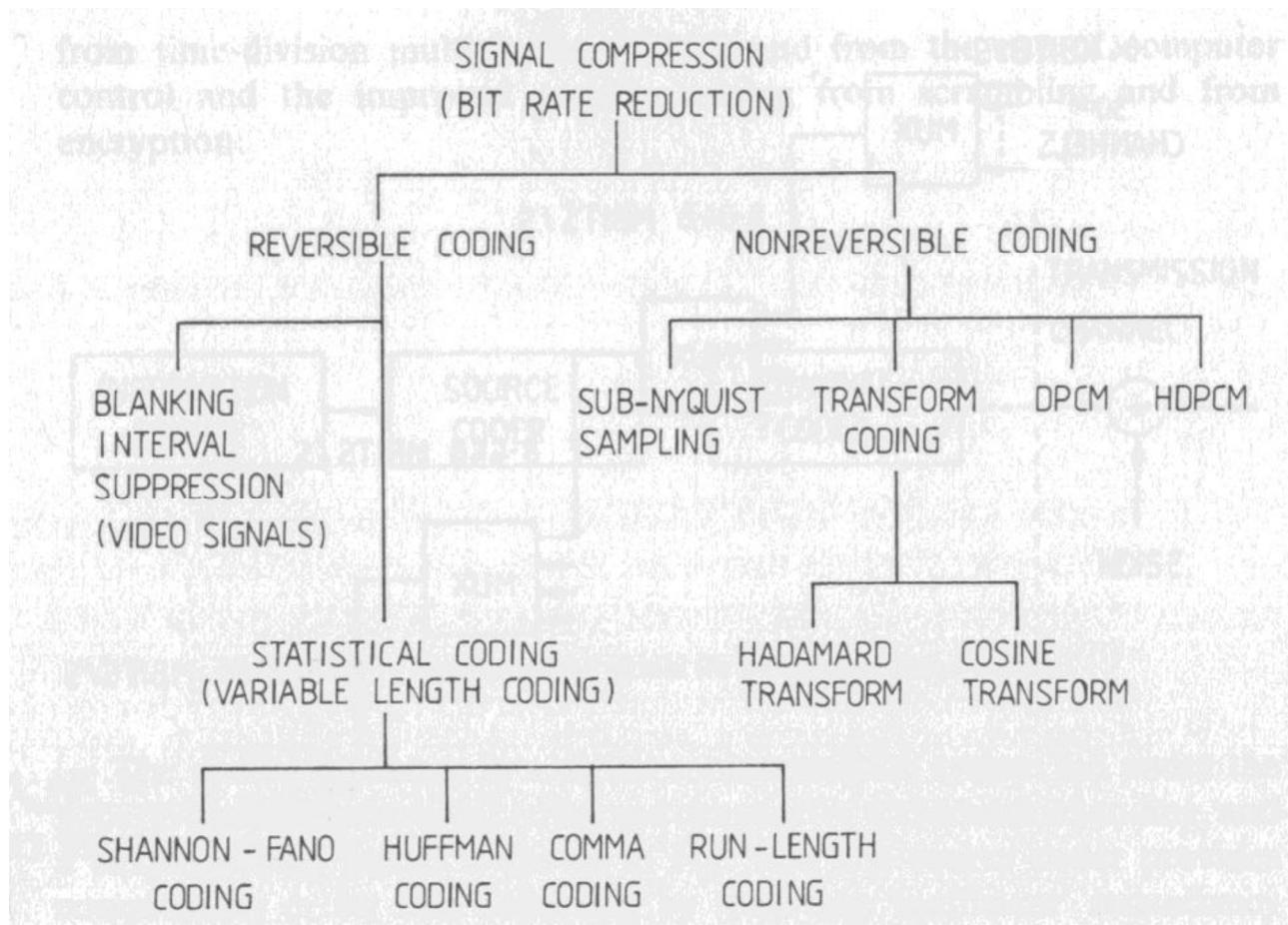
benötigt idealerweise eigentlich 10 bit-log-CamOutput
= $2^{10} \approx 62\text{dB}$ (s. Folie 21)
≈ 10 abbildbare Blendenstufen bei 2000lx Grundlicht

Minimum 16 bit **lineare** ADCs
an den Bildsensoren (pro Kanal)

benötigt idealerweise eigentlich 8 bit-log-CamOutput

beachte Praxisbeispiel Folie 85!

Compressionsmethoden (Datenreduktionsverfahren)



DPCM ≡ *differential* PCM
HDPCM ≡ *hybrid* DPCM

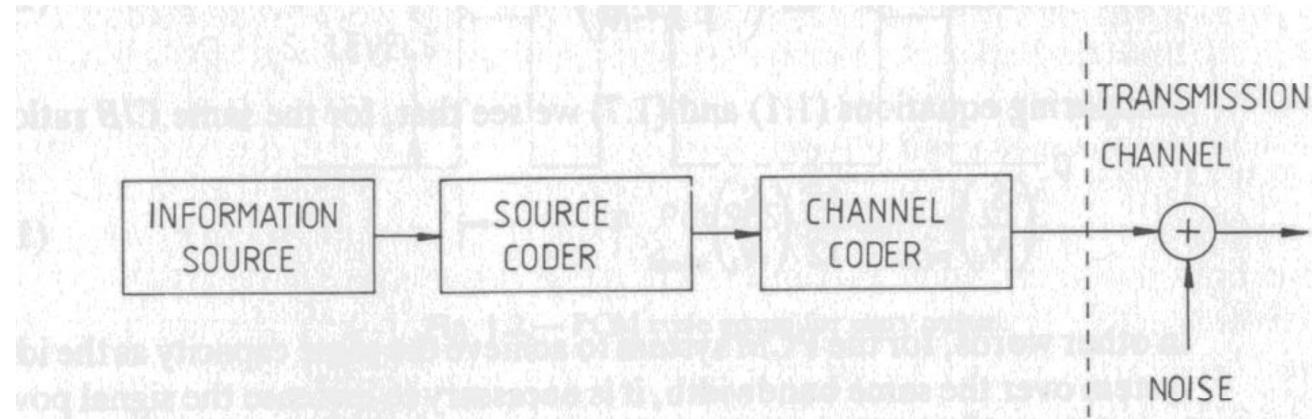
Prinzipien der Datenreduktion

1. **Redundanzreduktion:** Das Herausnehmen von überflüssiger, weil bereits bekannter Information.
(verlustfreie Codierung, statistische Codierung, *reversible coding*)

2. **Irrelevanzreduktion:** Das Weglassen von unwesentlicher und meist vom menschlichen Sinnesorgan (Auge, Ohr) nicht wahrnehmbarer Information.
(verlustbehaftete Codierung, *nonreversible coding*)

Datenreduktion (*compression coding*)

1. **Quellencodierung (*source coding*):** erzeugt eine Datenreduktion mittels einer Redundanzreduktion.
2. **Kanalcodierung (*channel coding*):** passt das zu übertragende Signal zum Kanal an und fügt „effiziente“ Redundanz in Form von Fehlerkorrektur ein.



Redundante Information beim Videosignal:

- Spatiale Redundanz (nebeneinander oder übereinander liegende Bildpunkte)
- Temporale Redundanz (aufeinander folgende Teilbilder, Ausschnitte der Teilbilder)

Irrelevante Information beim Videosignal:

- Chrominazuflösung
- Details der Bilder

Sub-sampling

*Preprocessing
(Chroma-sub-sampling)*

Die Video-Eingangsdaten liegen als sog. "Natives Video" nach dem ITU-R 601 Standard vor.

Das heißt, daß die Videodaten mit der Abtastfamilie 4:2:2 gewonnen worden sind.

Das *chroma-sub-sampling* ist ein einfaches Datenreduktionsverfahren, das eine Dezimierung der Daten aufgrund der physiologischen Eigenschaften und der Farbwahrnehmung des menschlichen Auges, durchführen kann, ohne eine sichtbare Verschlechterung der Videoinformation hervorzurufen.

Eingangsdaten:	Pre-processing:	Ergebnis:
<i>Natives Video</i>	<i>Sub-sampling</i>	
4:2:2	"4:2:0"	4:2:2 und 4:0:0

The table illustrates the sub-sampling process. The first column shows the input data as 'Natives Video' in 4:2:2 format. The second column shows the pre-processing step, which is 'Sub-sampling'. The third column shows the resulting output, which consists of two parts: 4:2:2 (the original YUV data) and 4:0:0 (the chroma data after sub-sampling).

Input Data (4:2:2):

```
<C_b, Y_1, C_r>, Y_2, <C_b, Y_3, C_r>, Y_4, <C_b, Y_5, C_r>, Y_6,  
<C_b, Y_1, C_r>, Y_2, <C_b, Y_3, C_r>, Y_4, <C_b, Y_5, C_r>, Y_6,  
<C_b, Y_1, C_r>, Y_2, <C_b, Y_3, C_r>, Y_4, <C_b, Y_5, C_r>, Y_6,  
<C_b, Y_1, C_r>, Y_2, <C_b, Y_3, C_r>, Y_4, <C_b, Y_5, C_r>, Y_6,  
...
```

Output Data (4:2:0):

```
<C_b, Y_1, C_r>, Y_2, <C_b, Y_3, C_r>, Y_4, <C_b, Y_5, C_r>, Y_6,  
<C_b, Y_1, C_r>, Y_2, <C_b, Y_3, C_r>, Y_4, <C_b, Y_5, C_r>, Y_6,  
<C_b, Y_1, C_r>, Y_2, <C_b, Y_3, C_r>, Y_4, <C_b, Y_5, C_r>, Y_6,  
<C_b, Y_1, C_r>, Y_2, <C_b, Y_3, C_r>, Y_4, <C_b, Y_5, C_r>, Y_6,  
...
```

Result (4:2:2 und 4:0:0):

```
<C_b, Y_1, C_r>, Y_2, <C_b, Y_3, C_r>, Y_4, <C_b, Y_5, C_r>, Y_6,  
<-, Y_1, ->, Y_2, <-, Y_3, ->, Y_4, <-, Y_5, ->, Y_6,  
<C_b, Y_1, C_r>, Y_2, <C_b, Y_3, C_r>, Y_4, <C_b, Y_5, C_r>, Y_6,  
<-, Y_1, ->, Y_2, <-, Y_3, ->, Y_4, <-, Y_5, ->, Y_6,  
...
```

Transformationscodierung

- Die Transformationscodierung ist eine unumkehrbare (*non reversible*) Verlustbehaftete Codierungsalternative, die sowohl auf Video- wie auf Audiosignale angewendet werden kann.
- Die TC erzielt eine Kompression der Daten dadurch, dass nur eine endliche Anzahl der Koeffizienten einer Transformation (z. B. Fourier-Transformation) übertragen werden.
- Der Empfänger führt eine inverse Transformation, um das ursprüngliche Signal zu rekonstruieren.

In der Praxis werden andere Transformationen als die DFT eingesetzt.

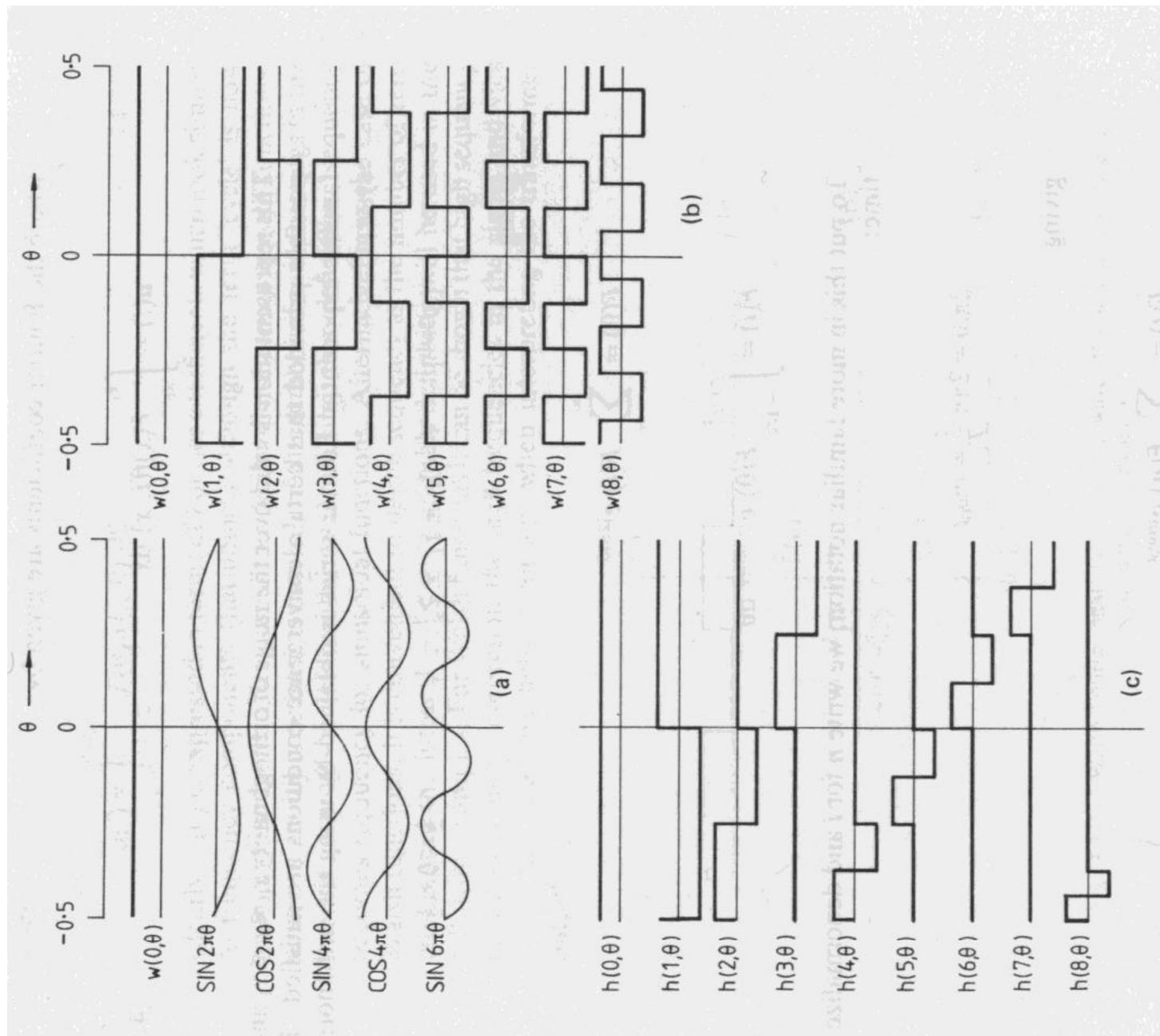
Beispiele dafür sind:

Die **DCT** (*Discrete Cosinus Transformation*)

Die **Walsh-Hadamard** Transformation
(zeichnet sich durch geringen Rechenaufwand aus)

Die **Haar** Transformation
(Sie lässt sich ebenfalls schnell berechnen)

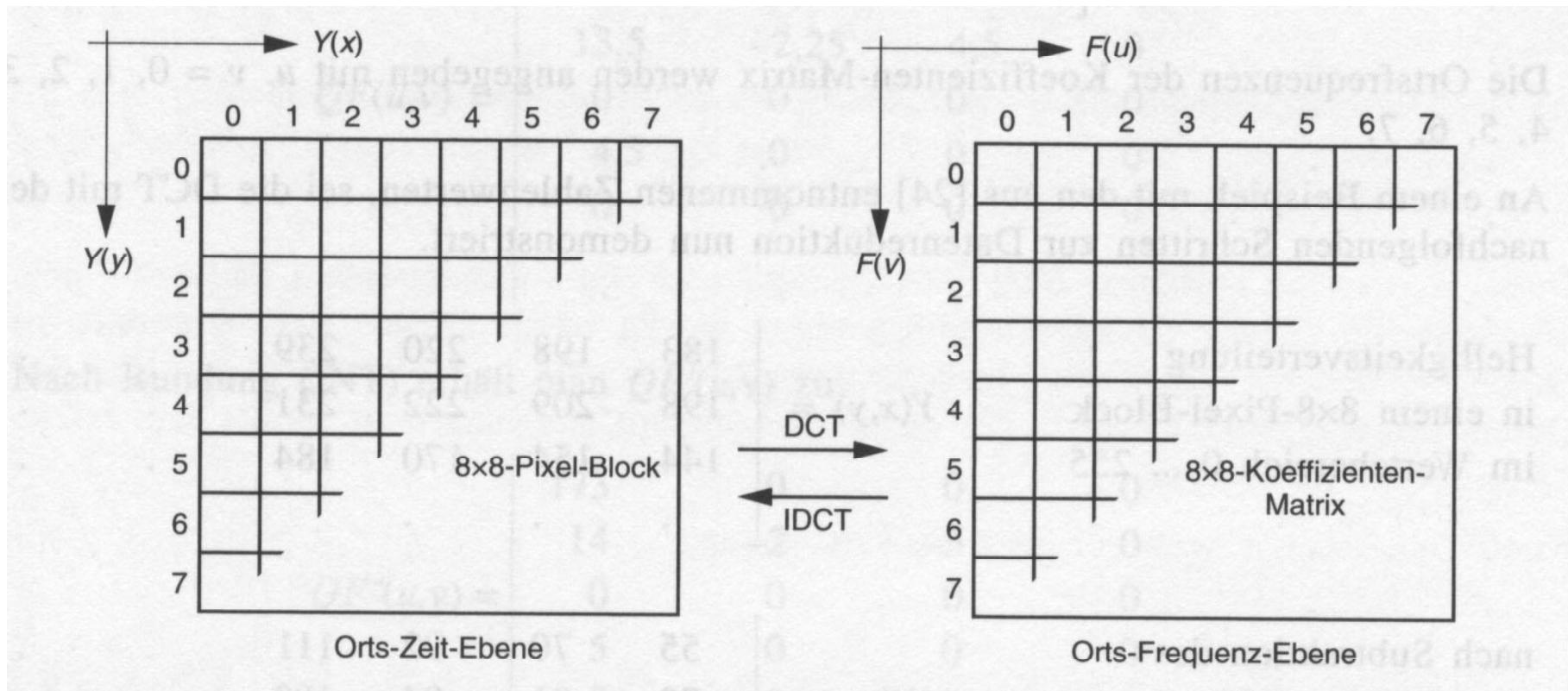
Usw.



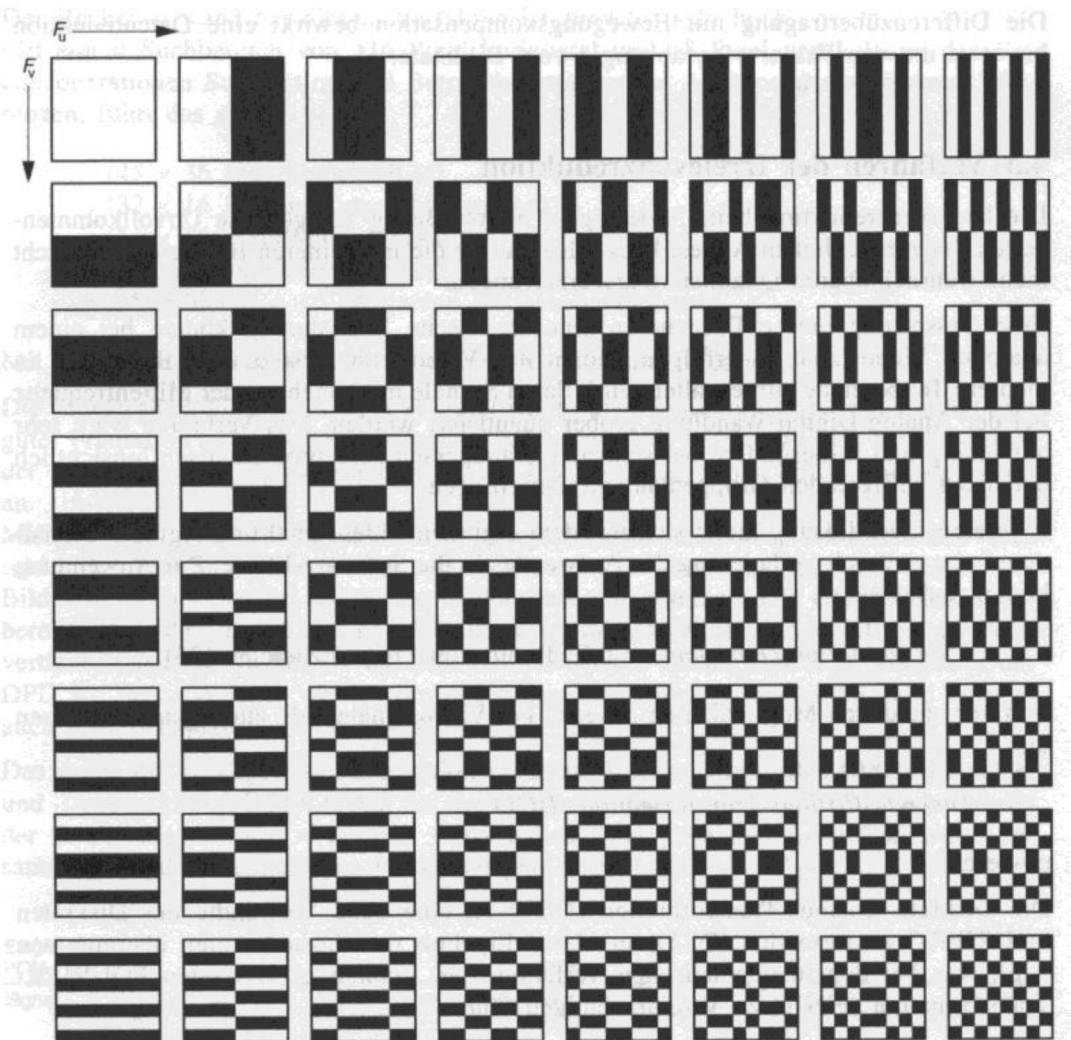
Basisfunktionen (Orthogonale Funktionen) für:

- a) die Cosinus und Sinus Transformation
- b) die Walsh-Hadamard Transformation und
- c) die Haar Transformation

Die *Discrete Cosinus Transformation* (DCT) und die IDCT

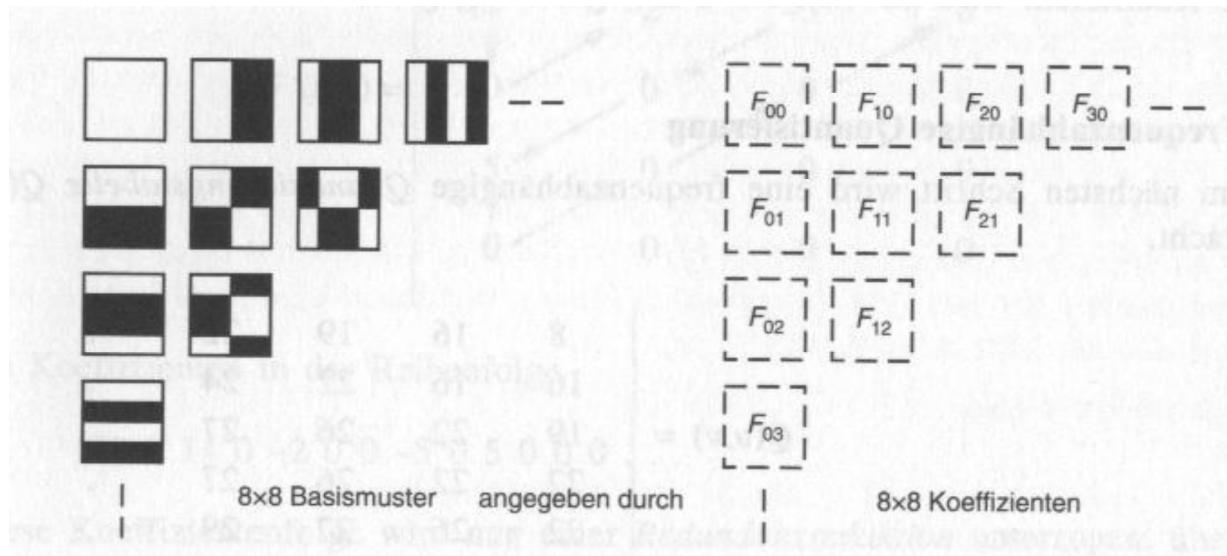


DCT und IDCT



DCT-Basismuster

Zuordnung der Basismuster auf die DCT-Koeffizienten



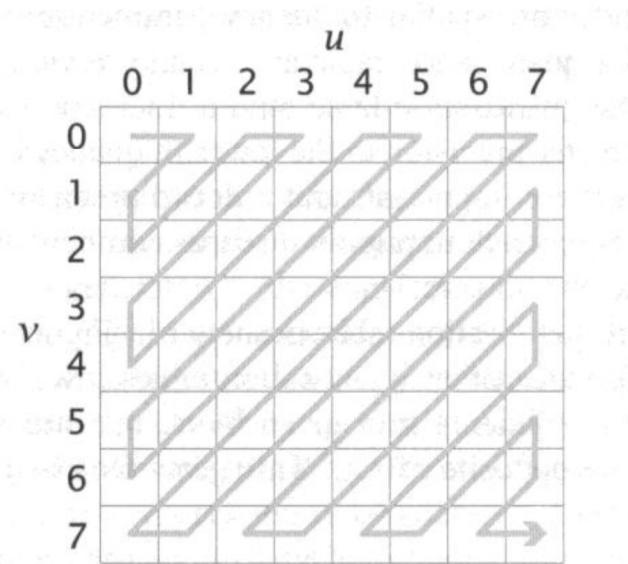
Die DCT-Koeffizienten werden eine frequenzabhängige Quantisierung unterzogen und durch die nachfolgende Rundung erfolgt eine Irrelevanzreduktion.

Die bekannteste Quantisierungsmatrix mit einem guten Kompromiss zwischen Bildqualität und Komprimierung ist:

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Typische Frequenzabhängige Quantisierungstabelle (Vorgang F/Q)

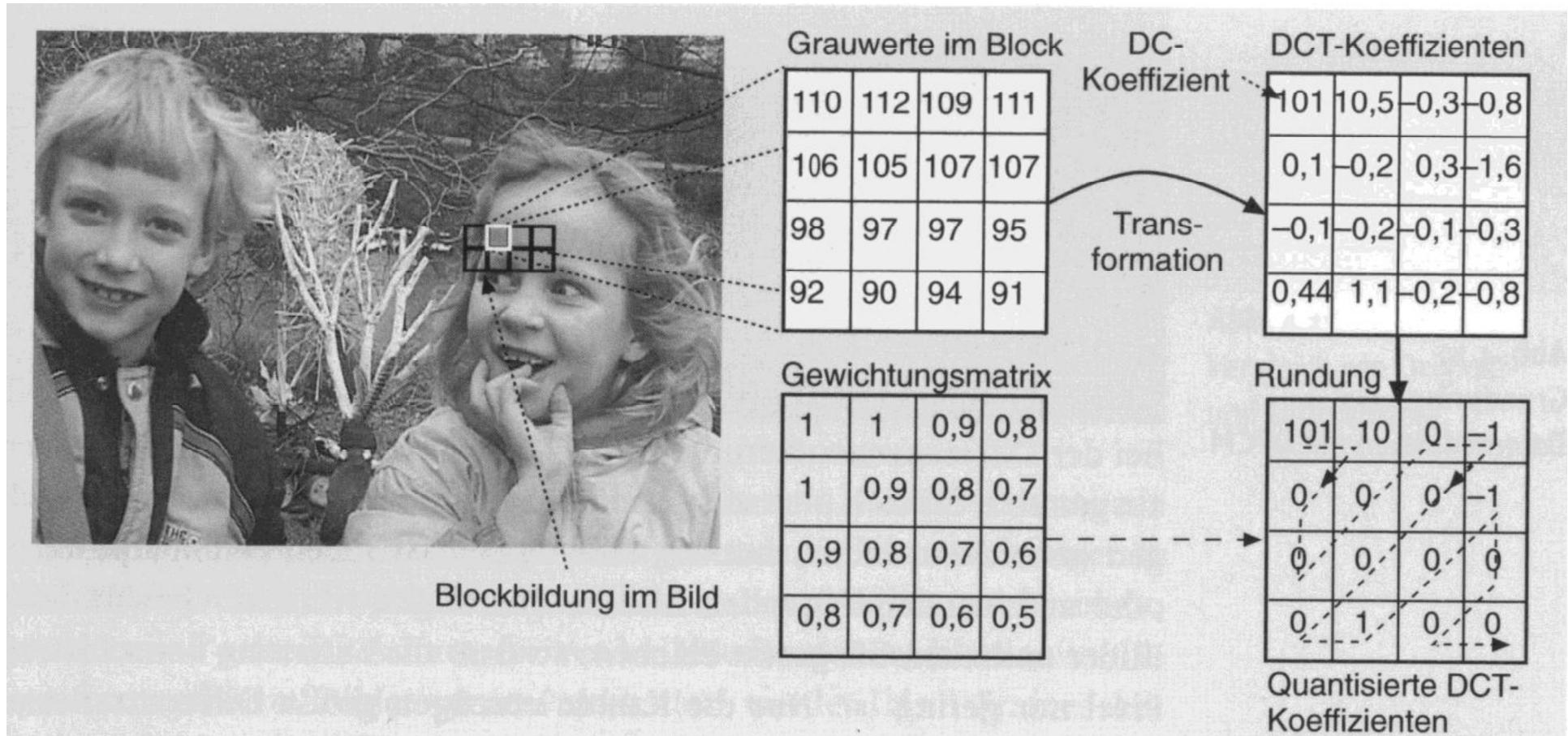
Umordnen der quantisierten DCT- Koeffizienten



Zick-Zack Umordnen der DCT-Koeffizienten

1. Dieses Umordnen ergibt die eindimensionale Reihenfolge:
 $F(0,0), F(1,0), F(0,1), F(0,2), F(1,1), \dots, F(6,7), F(7,7)$
2. Diese Reihenfolge wird mittels *Run Length Coding* einer Redundanzreduktion und anschließen einer *Variable Length Coding* mit Hilfe der Huffman Codierung unterzogen.

Blockbildung und Transformation der Blockwerte bei der DCT



Aus Professionelle Videotechnik, U. Schmidt, 2. Auflage, Springer, 2000.

Kompressionstandards für Videosignale

Standard	Digitalisierungsformat	Bitrate	Anwendungsbeispiel
H.261	CIF/QCIF	x 64 kbps	Videotelefonie /-konferenz über Telefonkanäle und LANs
H.263	S-QCIF/QCIF	< 64 kbps	Videotelefonie /-konferenz und Überwachung über <i>low bit rate</i> Kanäle
MPEG-1	SIF	< 1,5 Mbps	Videospeicherung auf CD-ROM in VHS-Qualität
	Low	< 4 Mbps	Videoaufzeichnung in VHS-Qualität
MPEG-2	4:2:0	< 1,5 Mbps	Digital Video Broadcasting
	Main	< 20 Mbps	
	4:2:2	< 60 Mbps	HDTV (4:3)
High	4:2:0	< 80 Mbps	
1440	4:2:2	< 80 Mbps	
	4:2:0	< 100 Mbps	HDTV (16:9)
	4:2:2	< 100 Mbps	
MPEG-4	Verschiedene	5kbps bis Zehnfache Mbps	Versatile multimedia coding standard
MPEG-7	—	—	Struktur- und Inhaltsbeschreibung komprimierter Multimedainformationen in <i>search engines</i>

Kompressionstandards für Audiosignale

Standard	Bitrate [kbps] (komprimiert)	Qualität	Anwendungsbeispiel
MPEG Audio	Layer 1	32 – 448	Hi-fi (192 kbps) Digital audio cassettes
	Layer 2	32 – 192	Near CD (128 kbps) Digital Audio und Video Broadcasting
	Layer 3	64	CD CD-Qualität über <i>low bit rate</i> Kanäle
Dolby audio coders	AC-1	512	Hi-fi Rundfunk und TV Satellitenrelays
	AC-2	256	Hi-fi PC-Sound Karten
	AC-3	192	Near CD Digital Video Broadcasting

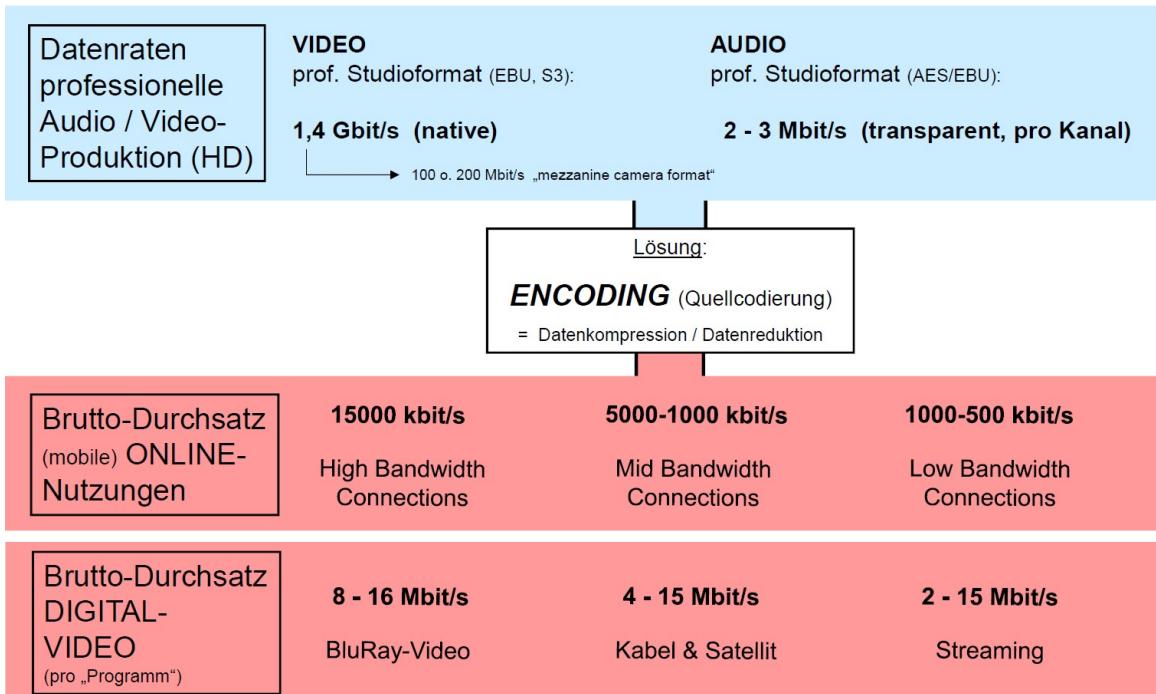
Digitaler Studiostandard ITU-R-601		
Parameter	Spezifikation	
System	525/60	625/50
Codierte Signale	Y, C_R, C_B	
Abtastfrequenzen: - Helligkeitssignal (Y) - jedes Farbdifferenzsignal (C_B, C_R)	(4:2:2) 13,5 MHz 6,75 MHz	
Anzahl der Abtastwerte pro Zeile: - Helligkeitssignal (Y) - jedes Farbdifferenzsignal (C_B, C_R)	858 429	864 432
Anzahl der Abtastwerte für den Bildinhalt pro Zeile: - Helligkeitssignal (Y) - jedes Farbdifferenzsignal (C_B, C_R)	720 360	
Abtaststruktur	Orthogonal über Zeile, Halbbild und Vollbild. Die C_R - und C_B -Werten stehen kohärent zu den ungeradzahligen Y -Werten in jeder Zeile	
Art der Quantisierung	Gleichmäßige Quantisierung mit 8 Bit (optional 10 Bit) pro Abtastwert (256 Quantisierungsstufen), für das Helligkeits- und jedes Farbdifferenzsignal	
Beziehung zwischen Videosignalpegel und Quantisierungsstufenzahl: - Helligkeitssignal (Y) - jedes Farbdifferenzsignal (C_B, C_R)	220 Quantisierungsstufen, wobei 0 Volt (Schwarz) auf Wert 16 und 0,7 Volt (Weiß) auf Wert 235 liegt. 225 Quantisierungsstufen liegen symmetrisch um den Wert 128, der 0 Volt repräsentiert.	
Art der Codierung	Gleichmäßig quantisierte PCM, 8 Bit (optional 10 Bit) pro Abtastwert, für das Helligkeits- und jedes Farbdifferenzsignal	
Benutzte Codeworte für Video	Nur die Werte 1 bis 254 stehen zur Codierung des Videosignals zur Verfügung. Die Werte 0 und 255 dürfen für das Videosignal nicht benutzt werden.	

Digitaler Standard ITU-R-709 (HDTV)

Parameter	Spezifikation	
Codierte Signale:	Y, C_R, C_B	
Frequenzbandbreite:	Y: 30 MHz	C_R, C_B je 15 MHz
Abtastfrequenzen: - Helligkeitssignal (Y) - jedes Farbdifferenzsignal (C_B, C_R)	(4:2:2) 74,25 MHz 37,125 MHz	
Anzahl der aktiven Zeilen pro Bild:	1080	
Bildseitenverhältnis (<i>aspect ratio</i>):	16 : 9	
Pixelseitenverhältnis (<i>pixel aspect ratio</i>)	1 : 1 (quadratische Pixels)	
Anzahl der Abtastwerte für den Bildinhalt pro Zeile: - Helligkeitssignal (Y) - jedes Farbdifferenzsignal (C_B, C_R)	1920 960	
Abtaststruktur:	Orthogonal über Zeile, Halbbild und Vollbild. Die C_R - und C_B -Werten stehen kohärent zu den ungeradzahligen Y-Werten in jeder Zeile	
Art der Quantisierung:	Gleichmäßige Quantisierung mit 8 Bit (optional 10 Bit) pro Abtastwert (256 bzw. 1024 Quantisierungsstufen), für das Helligkeits- und jedes Farbdifferenzsignal	
Beziehung zwischen Videosignalpegel und Quantisierungsstufenzahl: - Helligkeitssignal (Y) - jedes Farbdifferenzsignal (C_B, C_R)	220 Quantisierungsstufen, wobei 0 Volt (Schwarz) auf Wert 16 und 0,7 Volt (Weiß) auf Wert 235 liegt. 225 (16 bis 240) Quantisierungsstufen liegen symmetrisch um den Wert 128, der 0 Volt repräsentiert.	
Art der Codierung:	Gleichmäßig quantisierte PCM, 8 Bit (optional 10 Bit) pro Abtastwert, für das Helligkeits- und jedes Farbdifferenzsignal	
Benutzte Codeworte für Video:	Nur die Werte 1 bis 254 stehen zur Codierung des Videosignals zur Verfügung. Die Werte 0 und 255 dürfen für das Videosignal nicht benutzt werden.	
R, G, B auf Y, C_R, C_B Matrizierung:	$Y = 0,2126 R + 0,7152 G + 0,0722 B$ $C_R = (R-Y) = 0,7874 R - 0,7152 G - 0,0722 B$ $C_B = (B-Y) = -0,2126 R - 0,7152 G + 0,9278 B$	

MEDIA Encoding

Problemstellung Datenraten (I):



Problemstellung Datenraten (II):

Zur Erinnerung:

Wie kommen eigentlich die hohen (nativen) Datenraten zustande?

$$\text{Datenmenge} = \text{Abtastrate} \times \text{Quantisierung}$$

z.B. (Audio) $44,1 \text{ kHz}$ \times 16 bit $= 705 \text{ kbit/s}$

Theoretische
Ansatzpunkte
der Daten-
reduktion:



Abtastrate
(„Auflösung“)
reduzieren

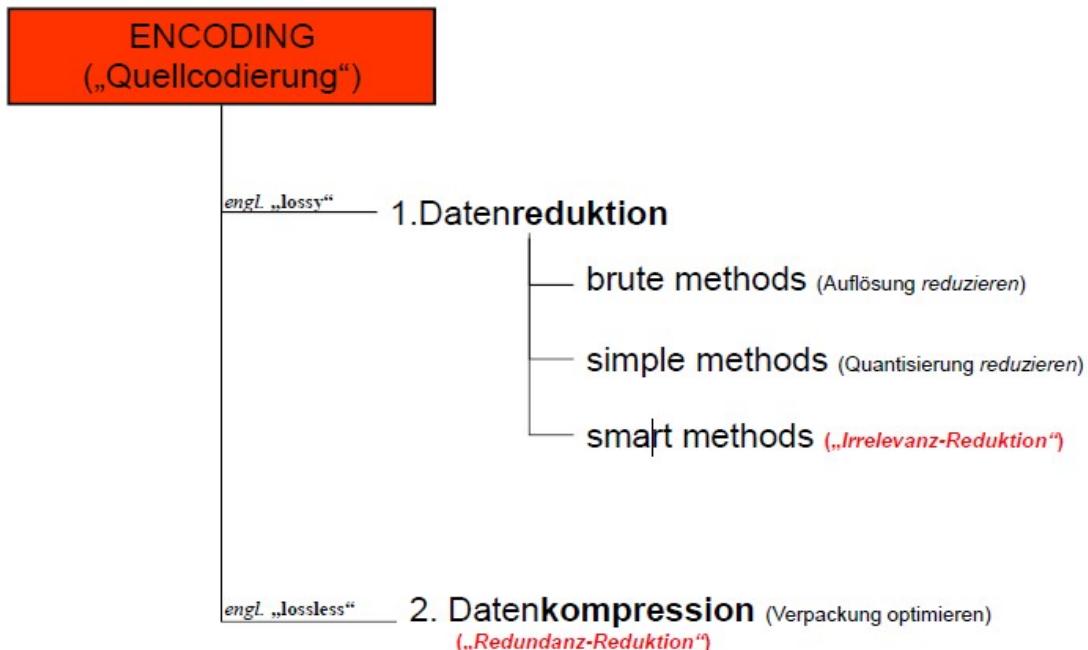


Quantisierung
(„Genauigkeit“)
reduzieren



Verpackung
optimieren
(„komprimieren“)

Grobunterscheidung Encodingphilosophien:



MPEG-Codierung

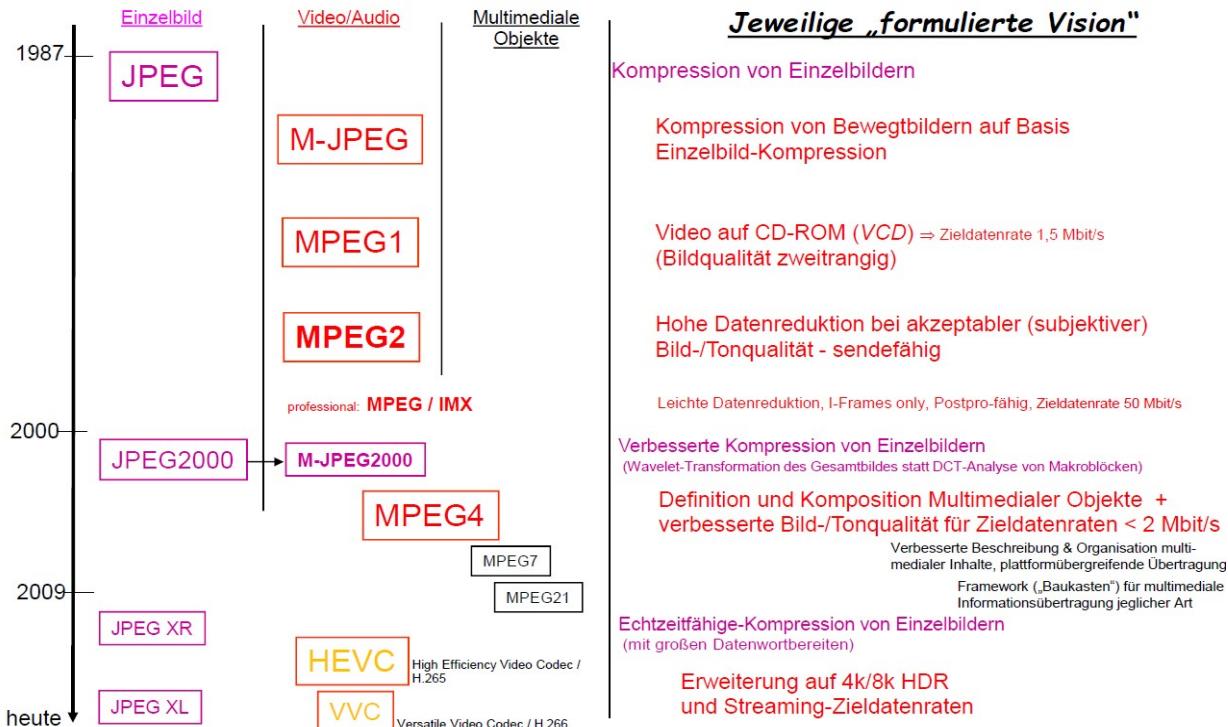
Zusammenfassung:

Vollständige Kennwerte für Digital Video

Jedes digitale Videoformat ist vollständig beschrieben durch folgende Kenngrößen:

- **Bildwechselfrequenz,** (temporale Auflösung)
- **Pixelraster,** (spatiale Auflösung)
- **Zeilenaufbau,**
- **beschreibender Farbraum,**
- **Sampling / Chroma-Subsampling,**
- **Quantisierung / Gamma,**
- **Bildcodierverfahren** („Codec“).

(Historische) Entwicklung der xPEG-Verfahren:



Der MPEG Standard

- *Motion Picture Expert Group (ISO Committee)*
- Definiert ein Standard für die codierte Darstellung bewegter Bilder, Audio und deren Kombination.
- Definiert den Datenstrom (*bit stream*), d. h. die Daten, die zum Decoder übertragen werden
- Definiert ebenfalls die Architektur des verwendeten Decoders

MPEG definiert nicht:

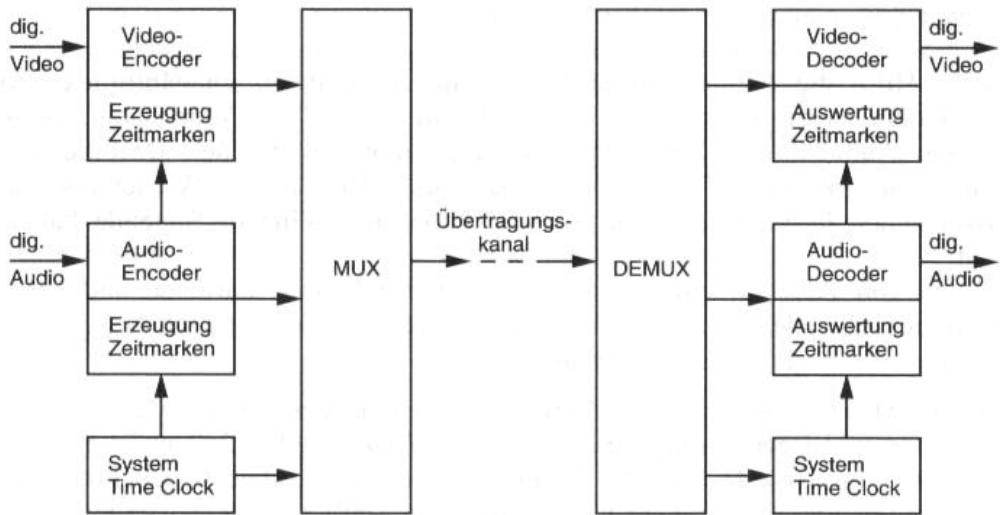
- Die Architektur des Encoders
- Die digitalen Medien
(MPEG ist nicht an eine bestimmte Anwendung oder Implementierung gebunden)

MPEG bildet den Rahmen für eine breite Palette audiovisueller Applikationen vom *Consumer*- bis in den professionellen Bereich auf der Basis digitaler Speicherung und Übertragung von Bild- und Tondaten.

Die MPEG Standards

- **MPEG-1** mit einer Datenrate von 1,5Mbit/s
(Speicherung von digitalisierten Bewegtbildern auf Medien wie CD-ROM, DAT, MOD und Computer-Discs. Die Datenreduktion beim Audiosignal erfolgt nach dem MUSICAM-Verfahren mit Layer 1)
- **MPEG-2** mit einer Datenrate von bis zu 100 Mbit/s
(günstig für Fernsehsignale, weil er nicht nur Vollbilder sondern auch das *interlace*-Verfahren berücksichtigt. Die Codierung erfolgt nach ITU-R-601 mit 4:2:2- oder 4:2:0-Chrominanzauflösung)
- **MPEG-4** für niedrige Datenraten bis 64 Kbit/s
(findet Anwendung bei Computer-Animation, mobile Bildkommunikation, Videokonferenzen und Internet)

Der MPEG-2 Standard



Prinzip des MPEG-2-Encoders und -Decoders

MPEG-2-Standard ISO/IEC 13818

Generic Coding of Moving Pictures and Associated Audio

Dieser Standard besteht aus den drei Teilen:

- ISO/IEC 13818-1 *Systems*
- ISO/IEC 13818-2 *Video*
- ISO/IEC 13818-3 *Audio*

MPEG-Systems

- Diesem Teil sorgt dafür, daß das Format von allen Empfängern verstanden wird, die MPEG unterstützen.
- Der Systems-Teil definiert die syntaktische Regeln (Zusammenstellung) und die semantische Regeln (Bedeutung) für die Generierung der Bitströme

Syntax des Datenstroms

- Packet Structure
- Pack Layer (Folge von Paketen)

Wie arbeitet MPEG-Video

- Preprocessing (*sub-sampling*)
- Inter pictures motion compensation
- Diskrete Cosinus Transformation
- Quantisierung
- Run Length Coding
- Variable Length Coding

- **Preprocessing**

Sub-sampling

Chroma-sub-sampling

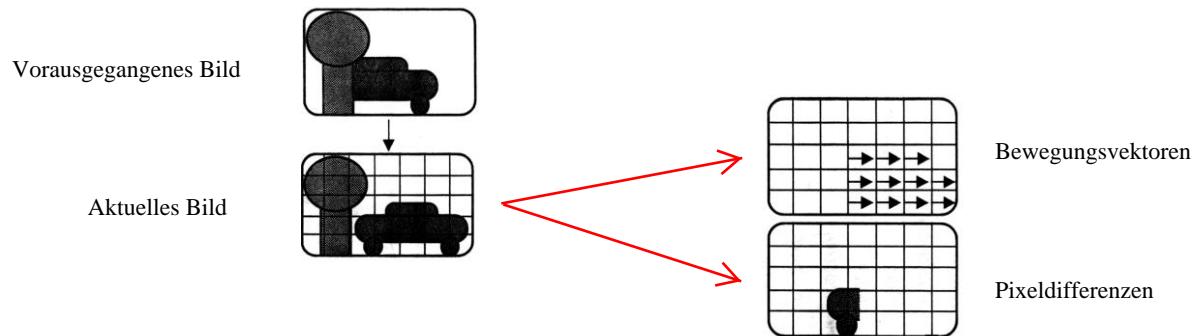
Die Video-Eingangsdaten liegen als sog. "Natives Video" nach dem ITU-R 601 Standard vor. Das heißt: die Videodaten sind mit der Abtastfamilie 4:2:2 gewonnen worden. Das *chroma-sub-sampling* ist ein einfaches Datenreduktionsverfahren, das eine Dezimierung der Daten aufgrund der physiologischen Eigenschaften und der Farbwahrnehmung des menschlichen Auges, ohne eine sichtbare Verschlechterung der Videoinformation hervorzurufen, durchgeführt werden kann.

Eingangsdaten:	Pre-processing:	Ergebnis:
<i>Natives Video</i>	<i>Sub-sampling</i>	<i>Dezimierte Daten</i>
4:2:2	"4:2:0"	4:2:2 und 4:0:0
<C _b , Y ₁ , C _r >, Y ₂ , <C _b , Y ₃ , C _r >, Y ₄ , <C _b , Y ₅ , C _r >, Y ₆ , <C _b , Y ₁ , C _r >, Y ₂ , <C _b , Y ₃ , C _r >, Y ₄ , <C _b , Y ₅ , C _r >, Y ₆ , <C _b , Y ₁ , C _r >, Y ₂ , <C _b , Y ₃ , C _r >, Y ₄ , <C _b , Y ₅ , C _r >, Y ₆ , <C _b , Y ₁ , C _r >, Y ₂ , <C _b , Y ₃ , C _r >, Y ₄ , <C _b , Y ₅ , C _r >, Y ₆ ,	<C _b , Y ₁ , C _r >, Y ₂ , <C _b , Y ₃ , C _r >, Y ₄ , <C _b , Y ₅ , C _r >, Y ₆ , <C _b , Y ₁ , C _r >, Y ₂ , <C _b , Y ₃ , C _r >, Y ₄ , <C _b , Y ₅ , C _r >, Y ₆ , <C _b , Y ₁ , C _r >, Y ₂ , <C _b , Y ₃ , C _r >, Y ₄ , <C _b , Y ₅ , C _r >, Y ₆ , <C _b , Y ₁ , C _r >, Y ₂ , <C _b , Y ₃ , C _r >, Y ₄ , <C _b , Y ₅ , C _r >, Y ₆ , ...	

Inter pictures motion compensation

Das bei MPEG angewandte Kompressionsverfahren nutzt Redundanzen in aufeinanderfolgenden Bildern, um eine Datenreduktion zu erreichen. Objekte, die sich in mehrere aufeinanderfolgenden Bildern wiederholen, können bei der Codierung der entsprechenden Daten eingespart werden. Hierfür ist es aber notwendig eine sog. **Bewegungsschätzung** durchzuführen.

- Bewegungsschätzung zwischen vorausgegangenem und aktuellem Bild
- Codierte Bewegungsvektoren und Pixeldifferenzen nach der Bewegung.



Bewegungsschätzung (cont.)

Ein Grundsätzliches Problem bei der Bewegungsvorhersage ist die Tatsache, dass dort, wo sich etwas wegbewegt, ein **neuer Hintergrund** freigegeben wird. Da dieser Hintergrund verdeckt war, ist sein Aussehen mit der Bildinformation der vorangegangenen Bilder nicht zu ermitteln. Diese Situation wird mit der sog. **Bidirektionalen Prädiktion** gelöst. Der Grundgedanke dabei ist es sowohl zurückliegende als auch zukünftige Bilder bei der Vorhersage zu berücksichtigen.

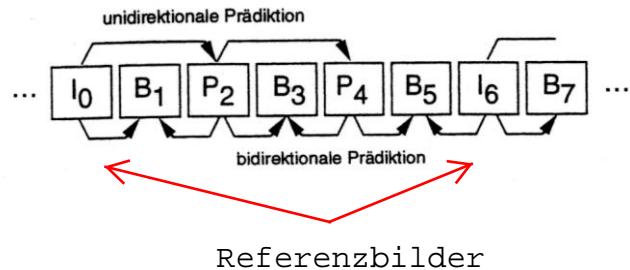
MPEG definiert daher drei verschiedene Bildtypen (MPEG-Syntax):

- **I:** *intraframe coded picture*
- **P:** *predictive coded picture*
- **B:** *bidirectionally predictive coded picture*

I: *intraframe coded picture*

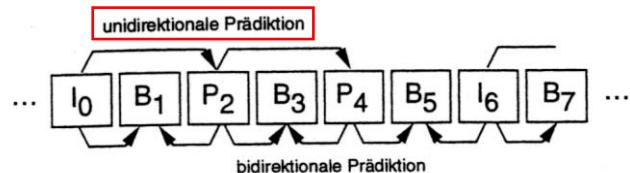
Die Referenzbilder vom Typ I werden mit einem JPEG-ähnlichen Algorithmus codiert.

Die Kompression erfolgt in mehreren Schritten und ist Zeitaufwendig. Mit 10 bis 20 liegt die Kompressionsfaktor dieses Bildtyps am niedrigsten. Die Zahl der I-Bilder pro Sekunde ist frei wählbar, allerdings sollten es nicht zu vielen sein, da sonst die Kompressionsrate zu gering ist.



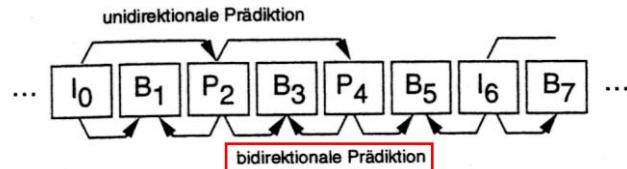
P: *predictive coded picture*

P-Bilder werden aus dem vorangegangenen I- oder P-Bild berechnet. Für die Berechnung wird das unter Bewegungsschätzung beschriebene Verfahren eingesetzt. Die Bilder werden in Blöcke zerlegt, deren Bewegung jeweils mit einem Bewegungsvektor beschrieben wird. Änderungen des Bildinhalts werden in einem zusätzlichen Datenpaket gespeichert.

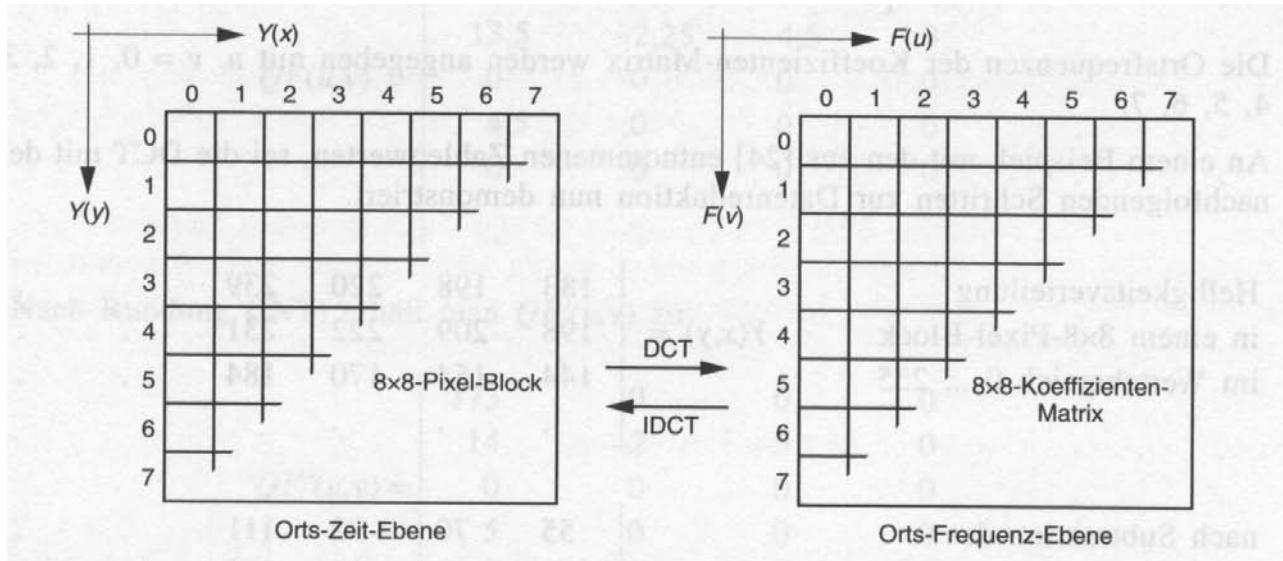


B: *bidirectionally predictive coded picture*

B-Bilder werden aus vorhergehenden und nachfolgenden I- und P-Bildern ermittelt. Damit ist der durch eine Bewegung freiwerdende Hintergrund rekonstruierbar. Der Kompressionsfaktor liegt bei B-Bildern am höchsten. Die Zahl der B-Bilder zwischen den als Referenz dienenden I- und P-Bildern ist variabel. Viele B-Bilder ergeben eine hohe Datenkompression, mehr I-Bilder erhöhen hingegen die Bildqualität und den direkten Zugriff.



Die *Discrete Cosinus Transformation* (DCT) und die IDCT



DCT und IDCT

Die *Discrete Cosinus Transformation* (DCT)

$$F(u,v) = \frac{1}{4} C(u) C(v) \sum_{x=0}^7 \sum_{y=0}^7 f(x,y) \cos \left[(\pi/16)(2x+1) u \right] \cos \left[(\pi/16)(2y+1) v \right]$$

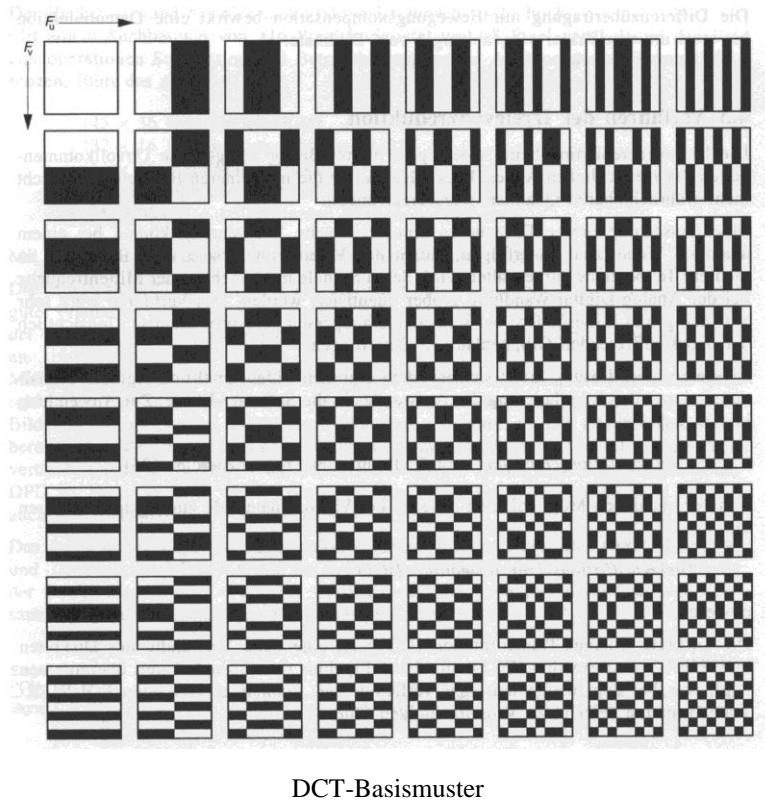
Das Eingangssignal liegt aus einem 8x8 Pixel-Block mit $x, y = 0, 1, 2, 3, 4, 5, 7$ vor

$$C(u) \text{ bzw. } C(v) = \begin{cases} 1/\sqrt{2} & \text{wenn } u=0 \text{ bzw. } v=0 \\ 1 & \text{wenn } u \neq 0 \text{ bzw. } v \neq 0 \end{cases}$$

Die *Inverse Discrete Cosinus Transformation (IDCT)*

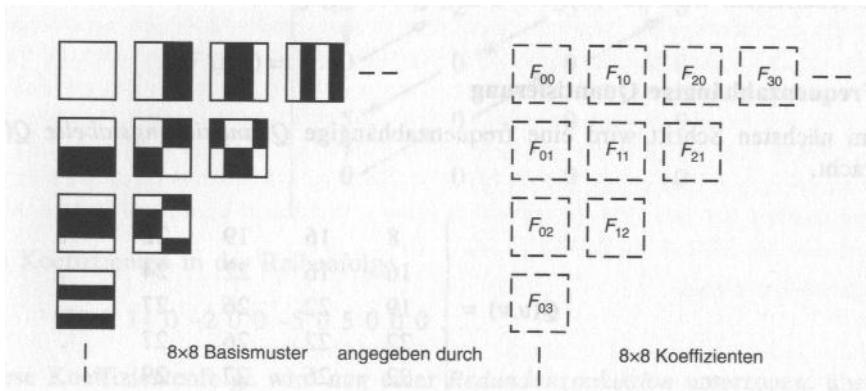
$$f(x,y) = \frac{1}{4} \sum_{u=0}^7 \sum_{v=0}^7 C(u) C(v) F(u,v) \cos [(\pi/16)(2x+1) u] \cos [(\pi/16)(2y+1) v]$$

$$C(u) \text{ bzw. } C(v) = \begin{cases} \frac{1}{\sqrt{2}} & \text{wenn } u=0 \text{ bzw. } v=0 \\ 1 & \text{wenn } u \neq 0 \text{ bzw. } v \neq 0 \end{cases}$$



DCT-Basismuster

Zuordnung der Basismuster auf die DCT-Koeffizienten



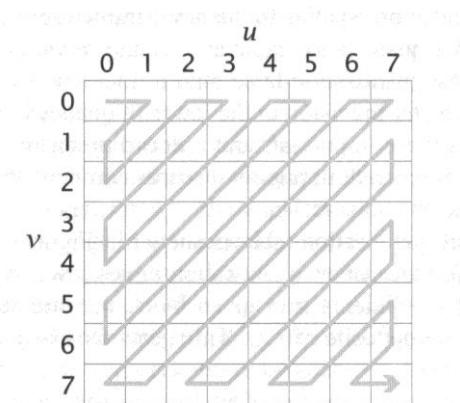
Die DCT-Koeffizienten werden eine frequenzabhängige Quantisierung unterzogen und durch die nachfolgende Rundung erfolgt eine Irrelevanzreduktion.

Die bekannteste Quantisierungsmatrix mit einem guten Kompromiss zwischen Bildqualität und Komprimierung ist:

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Typische Frequenzabhängige Quantisierungstabelle (Vorgang F/Q)

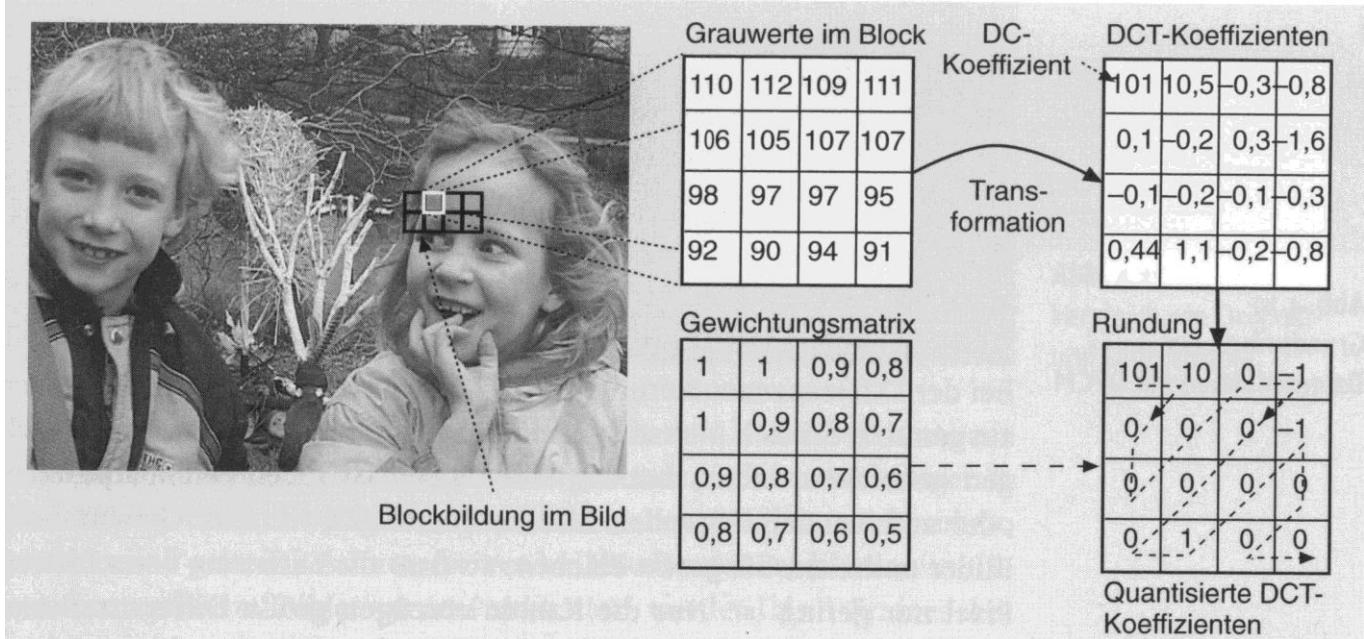
Umordnen der quantisierten DCT- Koeffizienten



Zick-Zack Umordnen der DCT-Koeffizienten

1. Dieses Umordnen ergibt die eindimensionale Reihenfolge:
 $F(0,0), F(1,0), F(0,1), F(0,2), F(1,1), \dots, F(6,7), F(7,7)$
2. Diese Reihenfolge wird mittels Run Length Coding einer Redundanzreduktion und anschließen einer Variable Length Coding mit Hilfe der Huffman Codierung unterzogen.

Blockbildung und Transformation der Blockwerte bei der DCT



MPEG-Video

Der MPEG-2-Standard wurde geschaffen um komprimierte Videobilder für verschiedene Qualitätsebenen zu speichern oder zu übertragen.

Um den unterschiedlichen Aufgaben auf einfache Weise gerecht zu werden, wurde eine Konfiguration mit *Profiles* und *Levels* formuliert.

Level max. Bildwechsel	Low Level 30/s	Main Level 30/s	High Level 1440 60/s	High Level 1920 60/s
Profiles				
Simple Profile 4:2:0 (keine B-Frames)		720 x 576 Pixel < 15 Mbit/s		
Main Profile 4:2:0	352 x 288 < 4 Mbit/s	720 x 576 < 15 Mbit/s	1440 x 1152 < 60 Mbit/s	1920 x 1152 Pixel < 80 Mbit/s
SNR Scalable Profile 4:2:0	352 x 288 < 4 (3) Mbit/s	720 x 576 < 15 (10) Mbit/s		
Spatially Scalable Profile 4:2:0			1440 x 1152 (720 x 576) < 60 (40) Mbit/s	
High Profile 4:2:0 oder 4:2:2		720 x 576 (352 x 288) < 20 (15) Mbit/s	1440 x 1152 (720 x 576) < 80 (60) Mbit/s	1920 x 1152 Pixel (960 x 576) < 100 (80) Mbit/s

Werte in Klammern geben die Alternativen bezgl. der Skalierung an

Profiles und Levels im MPEG-2

Level →	Low Level (LL)	Main Level (ML)	High 1440 Level (H1440L)	High Level (HL)	Profile ↓
Bildformat Bildwechsel/s. Datenrate		720 x 576 30 < 20	1440 x 1152 60 < 80	1920 x 1152 60 < 100	High Profile
Bildformat Bildwechsel/s. Datenrate		720 x 576 30 < 15	1440 x 1152 60 < 60	1920 x 1152 60 < 80	
Bildformat Bildwechsel/s. Datenrate		352 x 288 30 < 4	720 x 576 30 < 20	960 x 576 30 < 25	
Bildformat Bildwechsel/s. Datenrate			1440 x 1152 60 < 60		Spatially Scalable Profile
Bildformat Bildwechsel/s. Datenrate			1440 x 1152 60 < 40		
Bildformat Bildwechsel/s. Datenrate			720 x 576 30 < 15		
Bildformat Bildwechsel/s. Datenrate	352 x 288 30 < 4	720 x 576 30 < 15			SNR Scalable Profile
Bildformat Bildwechsel/s. Datenrate	352 x 288 30 < 3	720 x 576 30 < 10			
Bildformat Bildwechsel/s. Datenrate	352 x 288 30 < 4	720 x 576 30 < 15	1440 x 1152 60 < 60	1920 x 1152 60 < 80	Main Profile
Bildformat Bildwechsel/s. Datenrate		720 x 576 30 < 15			Simple Profile

Bildformat [Pixel], Bildfrequenz [Hz] und Datenrate [Mbit/s]

Studio Profile:

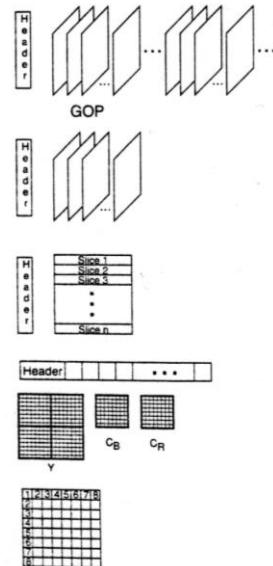
MP@ML-4:2:2 oder 4:2:2-Profile@ML
Bitrate: 50 Mbps

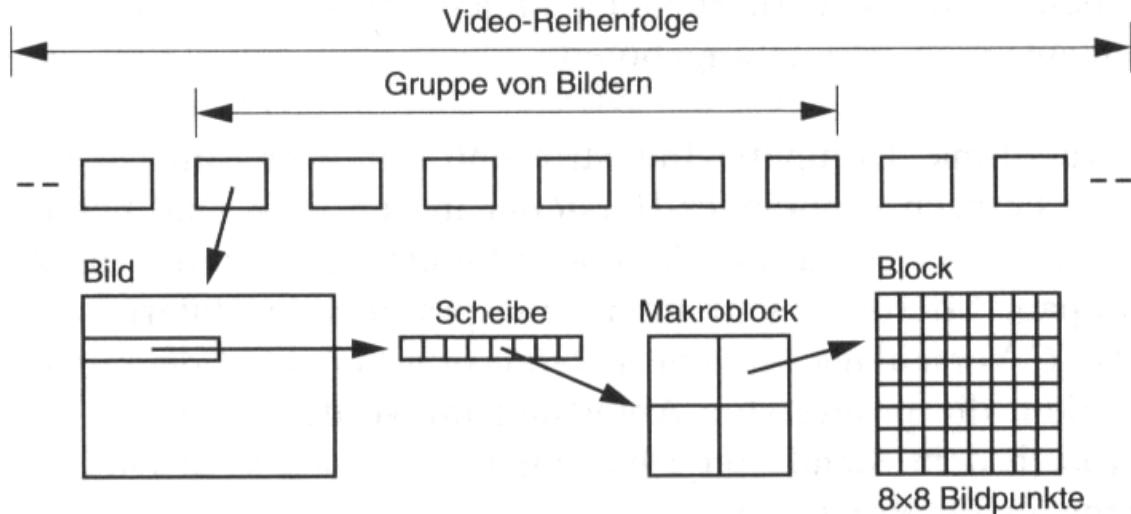
Vergleich der in EUR und USA möglichen Kombinationen:

MPEG-2-Level	Europa	USA
High	1920-1152-25(P), 16:9 50(I)	1920-1080-24(P), 30(P), 16:9, (quadr) 60(I)
High 1440	1440-1152-50(I) 16:9	1280-720- 24(P) , 30(P), 16:9, (quadr) 60(P)
Main	720-576-50(I) 4:3, 16:9	704-480- 24(P) , 30(P) , 4:3, 16:9 60(I) , 60(P) 640-480-24(P), 30(P), 4:3,(quadr) 60(I), 60(P)

Hierarchische Struktur des Datenstroms (Layers beim Videodatenstrom)

- Sequence
- Group of Pictures
- Picture
- Slice
- Macroblock
- Block

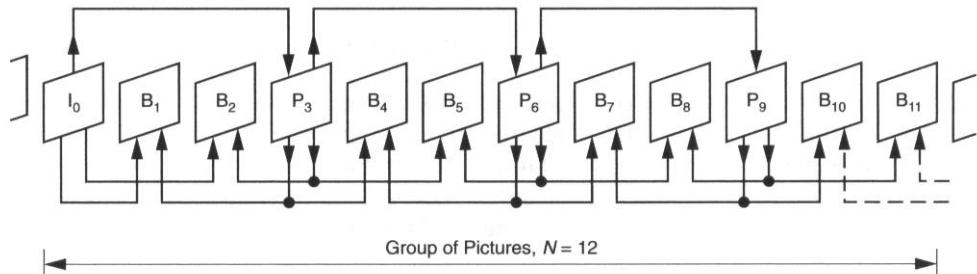




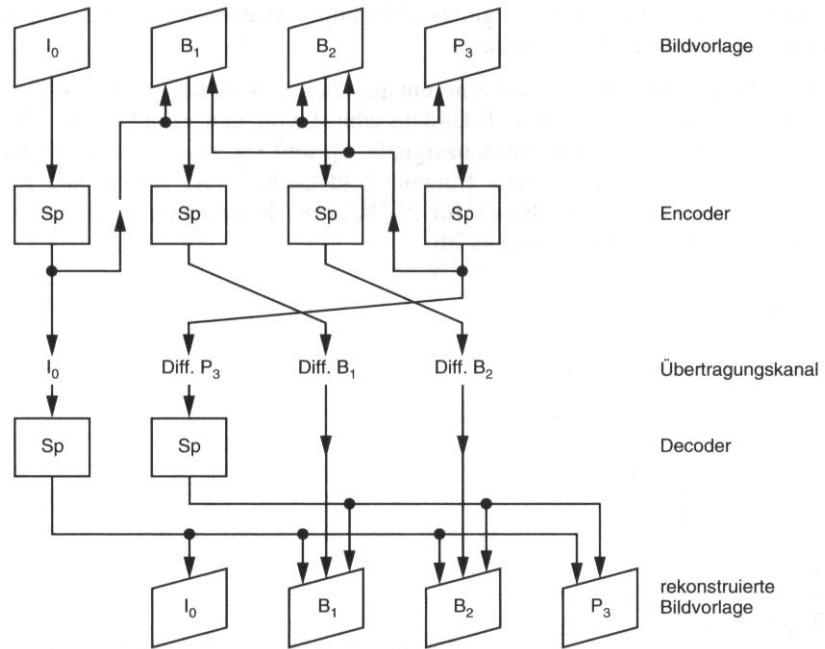
Zuordnung der Layers innerhalb einer Video-Sequenz

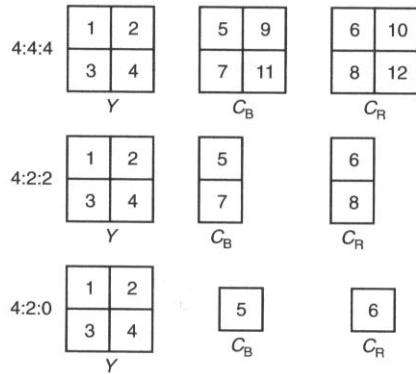
N	M	Bildsequenz																	
12	3	Cod	I_0	B_1	B_2	P_3	B_4	B_5	P_6	B_7	B_8	P_9	B_{10}	B_{11}	I_{12}	.	.	.	
		Übtr				I_0	P_3	B_1	B_2	P_6	B_4	B_5	P_9	B_7	B_8	I_{12}	B_{10}	B_{11}	.
6	2	Cod	I_0	B_1	P_2	B_3	P_4	B_5	I_6	
		Übtr	I_0	P_2	B_1	P_4	B_3	I_6	B_5	

Beispiele verschiedener Kombinationen von I-, P- und B-Bildern

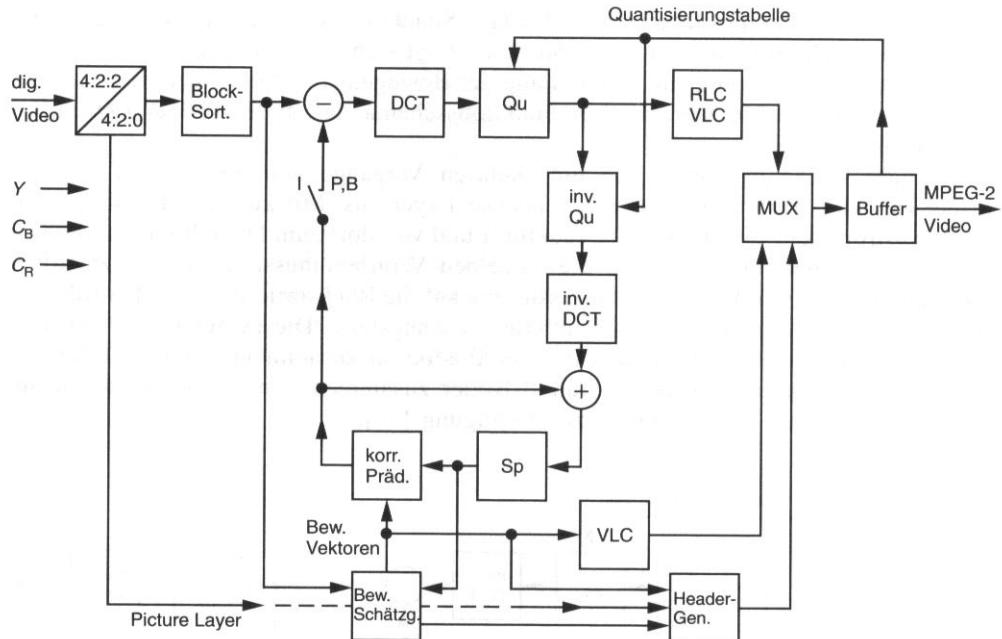


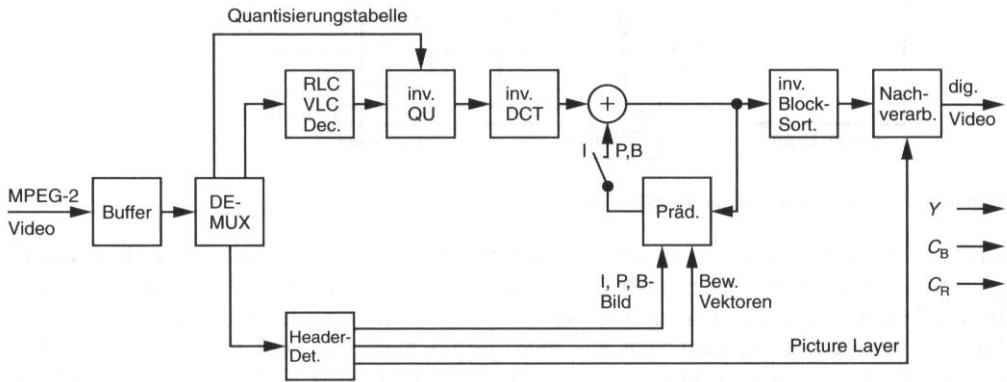
Group of Pictures (GOP): $N=12$ und $M=3$





Makroblock-Block-Zuordnung bei verschiedenen Abtastraster (4:4:4; 4:2:2 und 4:2:0)





Blockschaltbild des MPEG-2-Video-Decoders

MPEG-Audio

MPEG-1-Audio Mono, L-R-Stereo	MPEG-2-Audio Multichannel-Erweiterung
Layer I	Layer II MC
Layer II	Layer II MC
Layer III	Layer III MC

Zusätzliche Abtastfrequenzen für die Codierung mit sehr niedrigen Bitraten:
16 KHz, 22,05 KHz und 24 KHz.

Neben der Teilbandcodierung wird das AAC-Verfahren verwendet.

Audiokompressionsverfahren

Die Frequenzbandbreite des Audiosignals beträgt etwa 20 KHz (15 KHz)

20 Hz (30 Hz) bis 20 KHz (15 KHz)

Die dafür verwendete Abtastfrequenzen (f_A) betragen:

- a) 32KHz (Rundfunk)
- b) 44,1 KHz (CD)
- c) 48 KHz (Studio, DAT)

bei einer Codierung mit 16 Bit pro Codewort bekommt man eine Bitrate von:

$$r_{\text{Bit}} = 48 \text{ KHz} \cdot 16 \text{ Bit} = 768 \text{ KBit/s pro Audiokanal}$$

Für ein Stereosignal ist dann die Bitrate: $768 \text{ KBit/s} \cdot 2 = 1,536 \text{ MBit/s}$

Verfahrensansätze AUDIO – Encoding:

Redundanz - Reduktion ⇒ Gleiche Bit - Informationen zusammenfassen und einfacher codieren !

Ansatz: **ENTROPIE – Codierung**

Irrelevanz - Reduktion ⇒ Welche Signalanteile sind für den Menschen „unhörbar“ ?

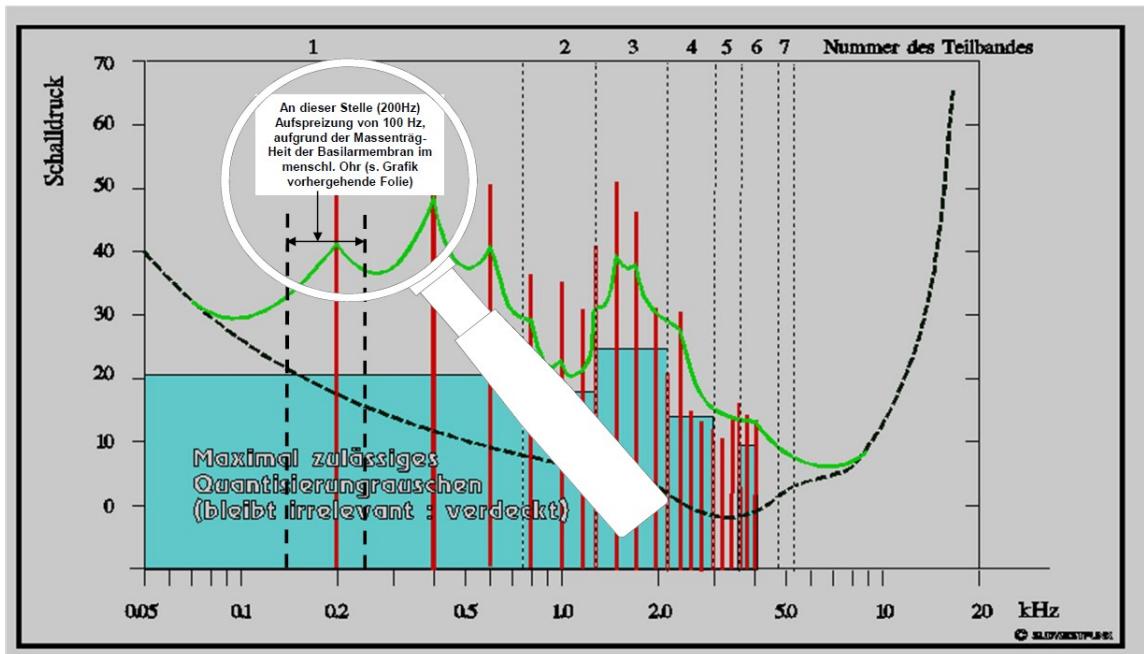
Ansatz: **PERCEPTUAL – Coding**

a) im Frequenzbereich

b) im Zeitbereich

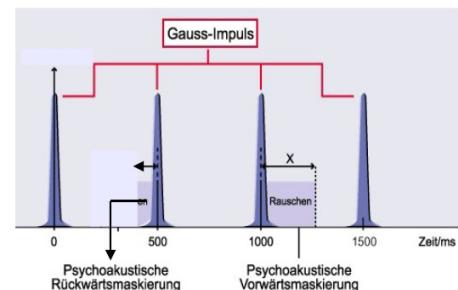
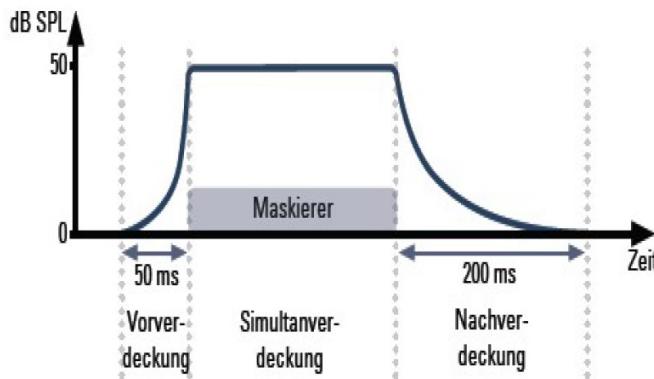
Verdeutlichung des spektralen Verdeckungseffekts:

Beispiel: Veränderung der Hörschwelle (dunkelgrün zu grün) beim Auftreten des Vokales „A“ (rote Spektrallinien)



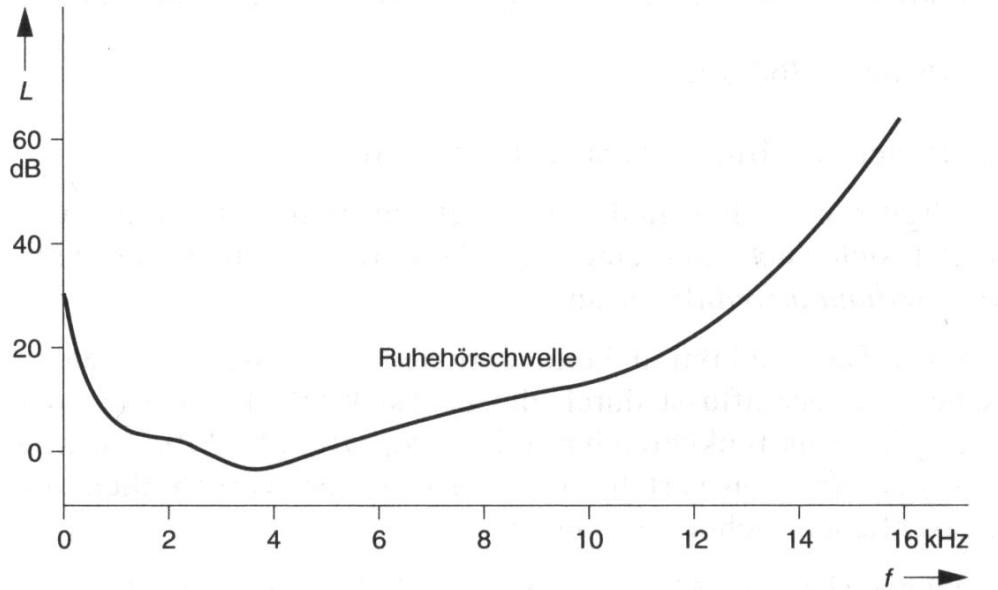
Beispiel für temporalen Verdeckungseffekt:

Das menschliche Gehör verarbeitet Schallereignisse mit hohem Schalldruckpegel schneller als Schallereignisse mit niedrigerem Schalldruckpegel [B08]. Diese Eigenschaft kann bewirken, dass leise Schallereignisse von lauteren Schallereignissen gewissermaßen überholt werden. Dieser als Vorverdeckung bezeichnete Effekt wirkt für einen Zeitraum von etwa 50 ms vor Einsetzen des verdeckenden Schallsignals [B02]. Nachdem ein Maskierer nicht mehr erklingt, schwingt die Wanderwelle auf der Basilarmembran aus, kehrt in die Ruheposition bzw. auf die Empfindlichkeit der Wahrnehmungsschwelle zurück. Treten während der Zeit des Ausschwingens leisere Signale in umgebenden Frequenzbereichen auf, werden diese aufgrund der zu geringen Amplitude nicht wahrgenommen. Dieses Verhalten des menschlichen auditiven Systems erklärt den Effekt der Nachverdeckung, dieser kann bis zu 200 ms andauern [B02]

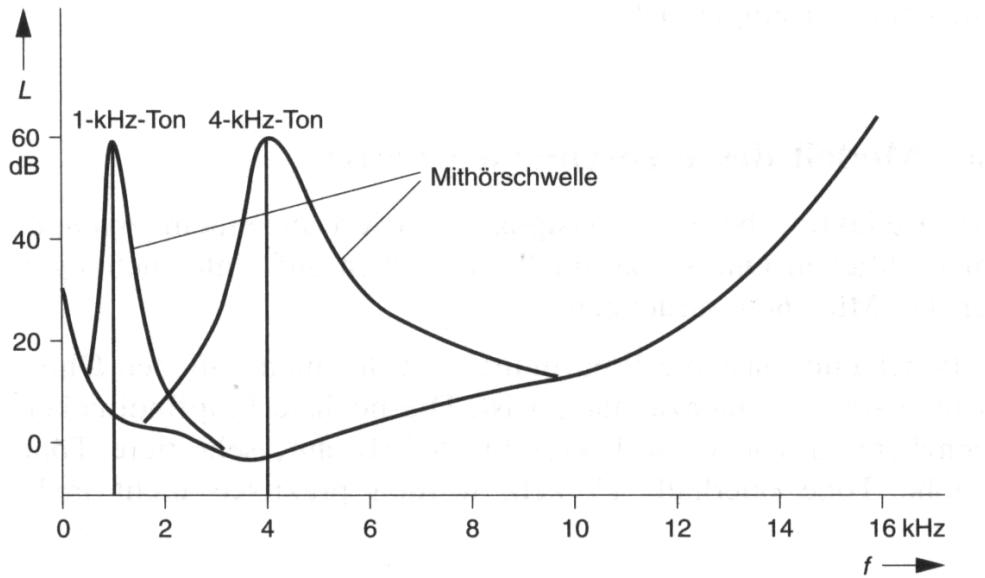


H. Fastel, E. Zwicker - „Psychoacoustics: Facts and Models“

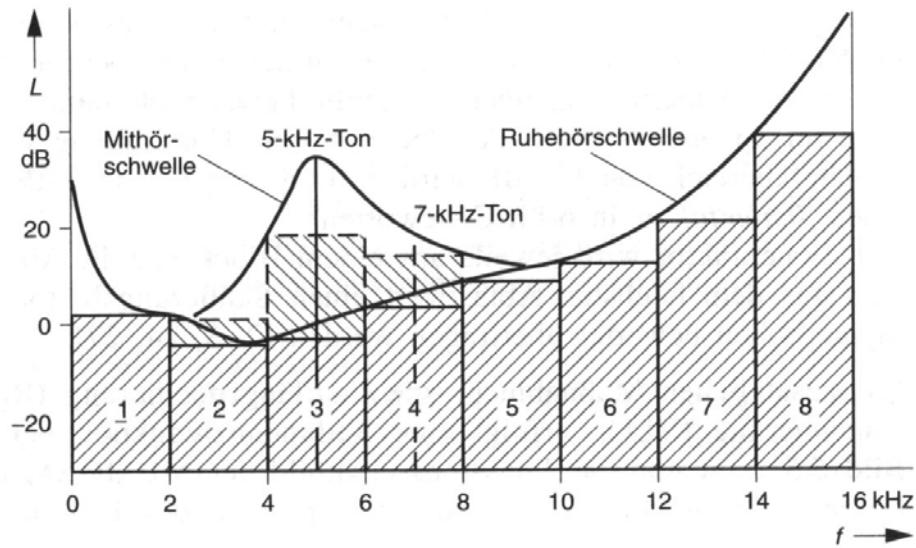
Ruhehörschwelle

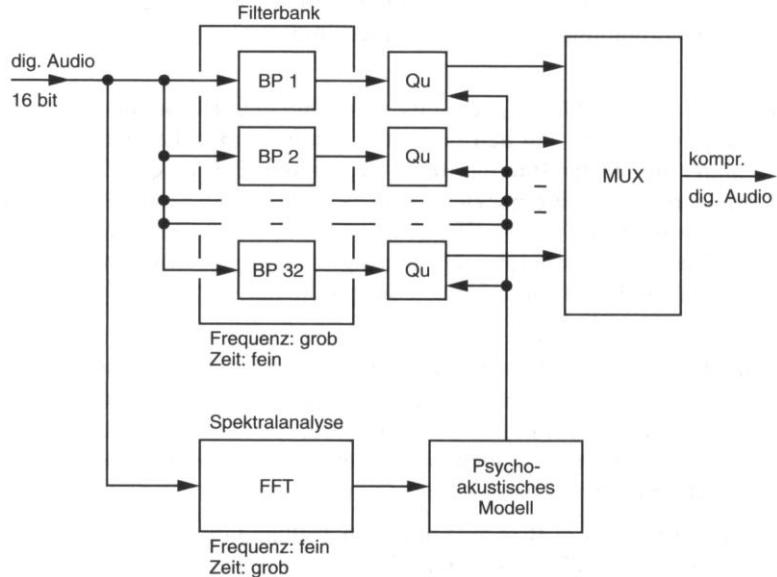


Dynamische Mithörschwelle

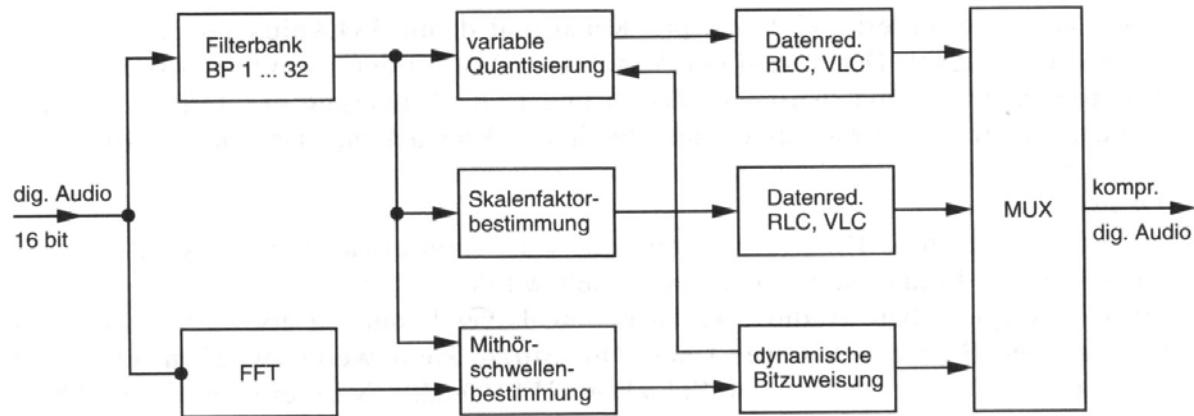


Teilbandcodierung

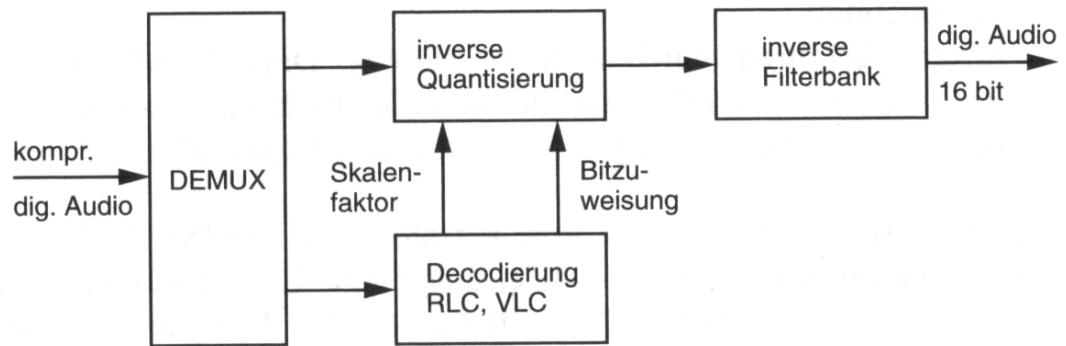




MUSICAM-Encoder



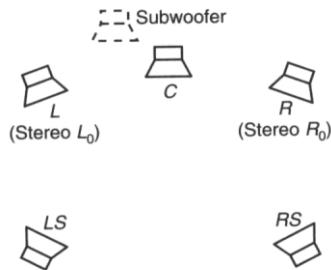
MUSICAM-Decoder



Layers bei den Codierverfahren

- Layer I (Pre-MUSICAM)
- Layer II (MUSICAM)
- Layer III (AC-3, MP-3, AAC)

Mehrkanalcodierung



Stereo-kompatiblen Surround-Sound-Signal

