

# 01 GRUNDLAGEN

Auf den folgenden Seiten wird versucht, einen kurzen Überblick über für die HD-Produktion wichtigen Grundlagen zu vermitteln. Das notwendige technische Verständnis für den sicheren Umgang mit Film- und TV-Programmen ist mit HD umfangreicher geworden. Immer wieder schwirren Begriffe wie 4:2:2, Quantisierung, Kompressionsverfahren, Codecs und ähnliches bei HD-Produktionen durch den Raum. Wer zuverlässig, sicher und vor allem innerhalb des vorgegebenen Budgetrahmens in HD produzieren will, muss sich heute intensiver mit technischen Fragen beschäftigen.

Überblickhaft, einfach verständlich, schnell – das sind die Merkmale unter denen die folgenden Informationen stehen. Sie sollen dem Nutzer für die Praxis einen ersten Input geben, die wichtigsten Fragen beantworten.

## Abtastformate

Die EBU-Festlegung einer gemeinsamen Schreibweise für die Abtastformate lautet: Es soll zunächst die vertikale Auflösung (aktive Zeilenzahl) genannt werden, gefolgt vom Abtastformat (i oder p). Nach einem Schrägstrich wird dann die Vollbildfrequenz (frame rate) genannt. In vielen Fällen wird insbesondere im Format 1080i/25 immer noch 1080i/50 genutzt. Dies ist jedoch eigentlich falsch. Wir verwenden in diesem Handbuch durchgehend die EBU-Schreibweise 1080i/25.

- 576i/25** „Standard Fernsehen“ (SDTV): Dieses Signal wird heute ausgestrahlt. Es arbeitet mit 576 Zeilen im interlaced Abtastformat und mit 25 Vollbildern pro Sekunde, die sich aber aus 50 Halbbildern zusammensetzen.  
Das eigentliche aktive Bildformat besteht aus 720 x 576 Zeilen.
- 1080i/25** „Hochauflösendes Fernsehen“ (HDTV) mit 1080 Zeilen im interlaced Abtastformat und mit 25 Vollbildern pro Sekunde, die sich aber aus 50 Halbbildern zusammensetzen.  
Das eigentliche aktive Bildformat besteht aus 1920 x 1080 Zeilen
- 720p/50** „Hochauflösendes Fernsehen“ (HDTV) mit 720 Zeilen im progressiven Abtastformat und mit 50 Vollbildern pro Sekunde.  
Das eigentliche aktive Bildformat besteht aus 1280 x 720 Zeilen.
- 1080p/24** „Hochauflösendes Fernsehen“ (HDTV) mit 1080 Zeilen im progressiven Abtastformat und mit 24 Vollbildern pro Sekunde. Das ist dieselbe Anzahl Vollbilder wie beim Kinofilm. Dieses Format wird auch bei Blu-ray Discs benutzt.  
Das eigentliche aktive Bildformat besteht aus 1920 x 1080 Zeilen.
- 1080p/25** „Hochauflösendes Fernsehen“ (HDTV) mit 1080 Zeilen im progressiven Abtastformat und mit 25 Vollbildern pro Sekunde.  
Das eigentliche aktive Bildformat besteht aus 1920 x 1080 Zeilen
- 1080p/50** „Hochauflösendes Fernsehen“ (HDTV): Dieses Signal wird zurzeit noch nicht ausgestrahlt. Es arbeitet mit 1080 Zeilen im progressiven Abtastformat mit 50 Vollbildern pro Sekunde.  
Das eigentliche aktive Bildformat besteht aus 1920 x 1080 Zeilen

## Der Unterschied progressiv/interlaced

*Das Interlaced-Verfahren – Bildformate mit Halbbildern: historisch bedingt und suboptimal*

Bei der Wiedergabe von bewegten Bildern wird die physiologische Eigenschaft des menschlichen Sehvermögens ausgenutzt, da sich durch die Trägheit des visuellen Systems die Änderungen des Bildinhaltes zu einer Bewegung vermischen. Ab 16 bis 18 Einzelbildern pro Sekunde nimmt das menschliche Auge diese als bewegte Sequenz wahr. Eine flimmerfreie Wahrnehmung ist dagegen erst ab etwa 50 Bildern oder Halbbildern pro Sekunde gewährleistet. Beim Kinofilm werden 24 Bilder pro Sekunde aufgenommen, beim Abspielen mit einem Kinoprojektor wird jedes Bild jedoch zweimal dargestellt, um das Flimmern zu vermindern.

Eine Übertragung von 50 Vollbildern pro Sekunde war zur Zeit der Entwicklung des „Standard Fernsehens“ (SDTV) technisch nicht möglich, da die dafür notwendige Übertragungsbandbreite für die erwünschte Bewegungsauflösung nicht zur Verfügung stand. Um dennoch eine flimmerfreie Darstellung mit 50 Hz und ausreichender Bewegungsauflösung zu ermöglichen, wurde ein Kompromiss gefunden: man überträgt Halbbilder. Jedes Bild wird zeilenweise abgetastet (Zeilensprungverfahren); allerdings zu jedem Abtastzeitpunkt nur die Hälfte der Zeilen. Dies erfolgt nach folgendem Schema: In einem Halbbild werden die ungeraden Zeilen sequentiell abgetastet und übertragen; zum nächsten Abtastzeitpunkt, also zeitlich später (unterschiedliche Bewegungsphase), werden im zweiten Halbbild dann die geraden Zeilen sequentiell abgetastet und übertragen. Damit wird die Wiederholrate der Bilder pro Sekunde künstlich verdoppelt. Im Empfänger werden diese 50 Halbbilder wieder zu Vollbildern (25 pro Sekunde) zusammengefügt. Hierdurch können bei der Bewegung und bei der Darstellung auf Flachdisplays (diese arbeiten immer progressiv, es ist daher de-interlacing erforderlich) Artefakte entstehen, die man als „Zähnen“ an sich bewegenden Elementen erkennen kann.

Bei klassischen Röhrenmonitoren machte das Nachleuchten zusammen mit der Trägheit des menschlichen Auges diesen Effekt kaum erkennbar. Bei Flachbildschirmen „verschmelzen“ die Halbbilder aus unterschiedlichen Bewegungsphasen dagegen nicht; vor allem bei der Darstellung von sich schräg bewegenden Objekten und Kameraschwenks führt dies zu einer deutlichen Unschärfe und Bildartefakten. Bei der Nutzung einer progressiven Abtastung (Vollbilder) treten diese Artefakte nicht auf.

Aufgrund der derzeitigen Parallelität von SDTV und HDTV sind auch in HD-Formaten Interlaced und Progressive in allen HD-Aufzeichnungsformaten vorhanden (siehe Kapitel 2). Mit konsequenten Workflows in HD wird das interlaced-Verfahren zunehmend verschwinden. Doch: archivierte SD-Inhalte werden in den kommenden Jahren immer wieder Eingang vor allem in nichtfiktionalen HD-Programmen finden (News, Reportage, Dokumentationen). Hier bleibt das Deinterlacing wichtig.

### *Das Progressive-Verfahren*

Bei diesem Verfahren wird das Bild vollständig und in einem einzigen Durchlauf von der ersten bis zur letzten Zeile abgetastet. Die Bilder werden nacheinander dargestellt. Bei HDTV sind verschiedene progressive Bildwiederholungen möglich. 24 Bilder pro Sekunde werden bei Filmmaterial verwendet. Im Fernsehbereich wird diese Bildwiederholrate auf 25 Bilder pro Sekunde umgerechnet (Kompatibilität zu den 50 Hz). 50 Bilder pro Sekunde sind ebenfalls möglich. Aufgrund von Bandbreitenbeschränkungen ist hier jedoch die Bildgröße begrenzt: 1920 x 1080 Pixel mit einer Bildwiederholung von 50 Vollbildern sind derzeit noch nicht übertragbar, weil die benötigte Bandbreite zu groß ist. Deshalb wird entweder die Auflösung beibehalten und mit 25 Bildern pro Sekunde gearbeitet (1080p/25) oder es wird die Auflösung von 1280 x 720 Pixel bei 50 Vollbildern pro Sekunde verwendet (720p/50).

In der elektronischen Filmproduktion für das Fernsehen wird das Verfahren 1080p/25 in der szenischen Feature-Produktion eingesetzt. Jedoch ist die Bildwiederholrate für das menschliche Auge eigentlich zu gering. Deshalb kommt es insbesondere bei schnellen Bewegungen (z. B. Fußball oder schnelle Kameraschwenks) zu einem „Stottern“ im Bild.

### *Deinterlacing – aus Halbbildern Vollbilder machen*

Da aus den vergangenen Jahrzehnten Unmengen von Material im interlaced-Format in den Archiven liegt, aber die neuen Flachbildschirme intern „progressiv“ arbeiten, viele Rundfunkanstalten im progressiven Format ausstrahlen und progressive Formate auch in der Produktion eingesetzt werden, wird es in den kommenden Jahren immer wieder die Notwendigkeit geben, Programme mit Halbbildern in Vollbilder umzurechnen. Dies wird Deinterlacing genannt. Dazu gibt es mehrere Verfahren, leider haben alle Vor- und Nachteile. Neben den semi-professionellen Verfahren (Weave, Bobbing, SkipField) werden im professionellen Produktionsumfeld folgende Verfahren verwendet: lineares, bewegungsadaptives und objektbasiertes Deinterlacing.

Für die **lineare Technik** (das resultierende Bild wird aus einer Anzahl von benachbarten Eingangsbildern berechnet) können unterschiedlich große Aperturen für die Konvertierung zum Einsatz kommen. In der Praxis findet man üblicherweise einen passenden Kompromiss für eine feste Apertur mit unterschiedlicher Gewichtung für innere und äußere Bildbereiche.

Bei der **bewegungsadaptiven Technik** versucht der Konverter die auftretenden Bewegungen zu detektieren und bewegungsabhängig zwischen einer kleineren Apertur (bei Bewegung) und einer größeren Apertur (keine Bewegung) umzuschalten. Dabei wird allerdings ein mehr oder weniger *sichtbares Pumpen* zwischen den unterschiedlichen *Auflösungsbereichen* sichtbar. Bei entsprechenden Bildinhalten kann dabei der Bewegungsdetektor durchaus in einen Dauerkonflikt kommen.

Bei der **objektbasierten Technik** werden sowohl die grundsätzliche Bewegung der Szene als auch die individuellen Bewegungen der Objekte gemessen. Auf Basis der gemessenen Bewegungsvektoren kann für jedes Objekt die volle Auflösung aus dem aktuellen und dem benachbarten Halbbild ermittelt werden. Die im aktuellen Halbbild fehlende vertikale Information wird aus benachbarten Halbbildern entnommen und in der richtigen Position dem aktuellen zugesetzt.

### *Sonderfall PsF – Progressive Segmented Frames*

Jedes Vollbild wird in zwei Halbbilder segmentiert (geteilt). Jedes Teilbild besteht für sich aus den ungeraden bzw. geraden Zeilen. Im Gegensatz zum Zeilensprungverfahren sind beide Halbbilder aus der gleichen Bewegungsphase. So können aus den im Halbbildverfahren gesendeten Teilbildern die ursprünglichen Vollbilder vollständig und verlustfrei rekonstruiert werden. Eine Interpolation ist nicht nötig, der für das Zeilensprungverfahren typische Kammefekt tritt nicht auf.

Das HDTV-Format 1080p/25 wird meist im PsF-Verfahren übertragen. Das HDTV-Display muss in dem Falle jedoch „verstehen“, dass es sich bei dem Bildmaterial um PsF-Material handelt und zwischen den Halbbildern kein zeitlicher Versatz (bei Schwenks und dergleichen) vorhanden ist und einfach aus zwei Halbbildern ein Vollbild zusammengesetzt werden darf. Dabei müssen eventuell im Display vorhandene Deinterlacing-Algorithmen deaktiviert werden. Diese Unterscheidung von *Interlaced-* gegenüber *Progressive segmented Frame-*Material muss von HD-Receivern oder HD-fähigen Fernsehgeräten als solches erkannt werden, sodass 1080p/25-Material ohne Deinterlacingbehandlungen durch die Algorithmen zu sauberem 1080p/25 gewandelt und angezeigt werden kann.

# Quantisierung, Abtastrate, Datenstrom, Signal

## Abtastrate und Datenstrom

Ein digitales Signal besteht im Gegensatz zu einem analogen Signal aus lediglich zwei Zuständen:

Signal aus = 0

Signal an = 1

Diese beiden Zustände werden im Informationsträger 1 Bit zusammengefasst.

Die Quantisierung (Amplitudenauflösung) wiederum ist die Anzahl der darstellbaren Zustände eines bestimmten Bildpunktes:

$2^8 = 256$  mögliche Werte    8 Bit

$2^{10} = 1024$  Werte            10 Bit

$2^{12} = 4096$  Werte            12 Bit

Im Audiobereich ist die angestrebte Quantisierung sehr viel größer, weil das Gehör im Gegensatz zu den Augen sehr viel empfindlicher ist:

$2^{16} = 65.536$  Werte        16 Bit

$2^{24} = 16.777.216$  Werte    24 Bit

Im Videobereich bezieht sich die Quantisierung auf die Granularität der abbildbaren Graustufen, im Audiobereich bezieht sie sich auf die Granularität der reproduzierbaren Lautstärke. Je höher die Quantisierung bei Video also ist, desto feiner können die Abstufungen von Grauwerten dargestellt werden (mit 1024 möglichen Grauwerten ist die Abstufung „feiner“ als mit 256 Grauwerten).

Mit der **Abtastrate** wird ein Signal kontinuierlich in gleichen zeitlichen Abständen abgetastet, ausgedrückt in Abtastungen pro Sekunde. Dies wird als Abtastfrequenz (auch Samplingfrequenz oder Samplingrate) bezeichnet und in der Einheit Hertz oder Samples per Seconds dargestellt. Quantisierung und Abtastrate zusammen ergeben eine Datenrate.

Die Datenrate eines HD-Signals ist sehr viel höher als bei den bislang bekannten SD-Signalen, weil die Auflösung in HD ein Vielfaches der Auflösung in SD beträgt (SD = 720 Bildpunkte auf 576 Zeilen, bei HD sind es z. B. 1920 Bildpunkte auf 1080 Zeilen oder 1280 Bildpunkte auf 720 Zeilen).

**Allgemein:**

Datenrate [Bit/s] =

Quantisierung [Bit] x Abtastrate [1/s]

**Beispiel aus der Praxis – Video-Datenrate eines 720p/50 Signals:**

Datenrate =

$$\{( \text{Quantisierung} \times \text{Pixel/Y-Zeile} \times \text{Y-Zeilenanzahl} ) + 2( \text{Quantisierung} \times \text{Pixel/Chroma-Zeile} \times \text{Chroma-Zeilenanzahl} )\} \times \text{Vollbildrate}$$

Datenrate [Bit/s] =

$$\{(10 \text{ Bit} \times 1280 \times 720) + 2(10 \text{ Bit} \times 640 \times 720)\} 50 [1/s] = 0,9216 \text{ [GBit/s]}$$

Das Bauteil, das in Gerätschaften der Fernsehproduktion (z. B. Kameras) die Digitalisierung vornimmt, ist der sogenannte A/D-Wandler, der analoge Signale in digitale Signale wandelt. Üblicherweise wird das analoge Signal mit einer Quantisierung von 8 oder 10 Bit abgetastet. Aber auch eine feinere Quantisierung kann angewandt werden. Die meisten A/D-Wandler in heutigen professionellen HD-Kameras arbeiten mit einer Quantisierung von mehr als 10 Bit, z. B. 12 Bit oder 14 Bit.

**Exkurs in die History der Videosignale****RGB-Signal**

Prinzipiell wird in einer Farbkamera das aufzunehmende Bild in die drei Grundfarben Rot, Grün und Blau zerlegt und für jede Farbe gesondert die entsprechende Lichtstärke, die wir Helligkeit nennen, in die Signale R, G und B umgesetzt.

Im Prinzip wäre es möglich, für jede der drei Grundfarben einen Übertragungskanal zur Verfügung zu stellen. Dies wäre aber mit mehreren Nachteilen, z. B. hohe erforderliche Bandbreite, (15 MHz in SDTV), hohen Kosten, Speicherproblemen verbunden. Insbesondere bei der Einführung des Farbfernsehens musste auch auf eine Kompatibilität zum bestehenden Schwarz-Weiß-Signal geachtet werden (zumindest in der Ausstrahlung). Daher musste die Farbinformation innerhalb der verfügbaren Bandbreite des Luminanzsignals (5 MHz) möglichst ohne wechselseitige Beeinträchtigung untergebracht werden.

**YUV-Signal**

Mit dem gewählten Lösungsweg für das analoge PAL-Fernsehen erzeugt man aus den Signalen R, G und B das sogenannte Leuchtdichte- oder Luminanzsignal Y und die zwei Farbdifferenzsignale U und V:

$$Y = 0,299 R + 0,587 G + 0,114 B$$

$$U = 0,493 (B-Y)$$

$$V = 0,877 (R-Y)$$

Vorteilhafterweise kann die Bandbreite der Farbdifferenzsignale begrenzt werden, da beim menschlichen Auge das Schärfeempfinden für Helligkeitssignale in etwa doppelt so hoch ist wie für Farbsignale. Für das analoge PAL-Fernsehen wurde für U und V eine Begrenzung auf 1,3 MHz gewählt.

In einem weiteren Schritt wurde ein Hilfsträger mit den U- und V-Farbdifferenzsignalen Quadratur-moduliert. Der modulierten Hilfsträger wird als Farbartsignal F bezeichnet. Sinnvollerweise wurden die Frequenz dieses Farbträgers und seine Verkopplung mit dem Luminanzsignal so optimiert, dass möglichst wenig wechselseitige Beeinflussung mit dem Luminanzsignal auftritt. Im Ergebnis wurde das BAS-Signal (Bild-Austast-Synchron-Signal) des Schwarz-Weiß-Fernsehens zu einem FBAS-Signal (Farb-Bild-Austast-Synchron-Signal) modifiziert.

### Y, B-Y, R-Y-Signale

In der weiteren Evolution der Fernsehsignale wurde versucht, die Einschränkungen durch das PAL-Signal (Luminanz- und Chrominanzsignale können nicht verlustlos separiert werden, begrenzte Bandbreite der Chrominanzsignale) durch eine getrennte Übertragung und Aufzeichnung der Luminanz- und Chrominanzsignale zu überwinden. Der entsprechende Standard für Europa ist in EBU N-10 festgelegt. Nachteilig war, dass es nach wie vor analoge Signale waren und zudem für eine Verbindung drei Videokabel erforderlich waren. Die damals eingeführten Aufzeichnungsformate wie Betacam SP haben in diesem analogen Komponentenformat aufgezeichnet. In der Praxis haben sich für die Bezeichnung dieser drei Signale die Begriffe Y, B-Y und R-Y durchgesetzt. Formal richtig müssten sie allerdings mit  $E'_Y$ ,  $E'_{CR}$  und  $E'_{CB}$  bezeichnet werden (siehe EBU N10):

$$E'_Y = 0.299 E'_R + 0.587 E'_G + 0.114 E'_B$$

Wobei  $E'_R$ ,  $E'_G$  und  $E'_B$  die primären Signale nach der Gamma-Korrektur sind.

Die Farbdifferenzsignale werden aus dem  $E'_Y$  Signal und dem jeweiligen primären Signal erzeugt:

$$E'_{CR} = 0.713 (E'_R - E'_Y)$$

$$E'_{CB} = 0.564 (E'_B - E'_Y)$$

Wichtig dabei ist zu wissen, dass zur Zeit der Festlegung der analogen Komponenten auch bereits der digitale Komponenten-Standard ITU-R BT.601 (hie damals

CCIR-601) fertig war. Die Spezifikationen für die analogen Komponentensignale weisen somit die gleiche Festlegung für die Bandbreiten der primären Luminanz- und Chrominanzsignale auf wie für die digitalen Komponenten.

Luminanzsignal: 5,75 MHz

Chrominanzsignale: 2,75 MHz.

### Digitale Komponentensignale

Die digitalen Komponentenensignale für SDTV sind in ITR-R BT.601 spezifiziert.

Die entsprechende Schnittstelle dazu (SDI-Interface) ist in IT-R BT.656 festgelegt.

Die digitalen Komponentensignale  $Y$ ,  $C_R$ ,  $C_B$ , werden aus den primären (analogen) Signalen  $E'_R$ ,  $E'_G$  und  $E'_B$  erzeugt. Nachfolgend sind die entsprechenden Vorschriften für ihre Erzeugung aufgeführt:

$$E'_Y = 0.299 E'_R + 0.587 E'_G + 0.114 E'_B$$

und für die normalisierten Farbdifferenzsignale

$$E'_{CR} = 0.713 (E'_R - E'_Y) = 0.500 E'_R - 0.419 E'_G - 0.081 E'_B$$

$$E'_{CB} = 0.564 (E'_B - E'_Y) = -0.169 E'_R - 0.331 E'_G + 0.500 E'_B$$

### Videosignale für HDTV

Die Spezifikationen der Videosignale für die unterschiedlichen HDTV-Abtastformate muss den jeweiligen Standards entnommen werden, da eine detaillierte Erläuterung an dieser Stelle zu weit führen würde. Prinzipiell ist der Zusammenhang von den RGB-Signalen und den  $Y$ ,  $C_R$ ,  $C_B$ -Signalen wie bei den oben aufgeführten digitalen Videosignalen für SDTV.

Nachfolgend ein kurzer Überblick der für Europa relevanten vier HD-Formate, wie sie in EBU Tech 3299 spezifiziert aufgeführt:

- **System 1 (S1)** mit 1280 aktiven horizontalen Abtastwerten und 720 aktiven Zeilen in progressiver Abtastung mit einer Vollbildrate von 50Hz, 16 x 9 Bildformat.  
**Abkürzung: 720p/50**
- **System 2 (S2)** mit 1920 aktiven horizontalen Abtastwerten und 1080 aktiven Zeilen in interlaced Abtastung mit einer Vollbildrate von 25Hz, 16 x 9 Bildformat.  
**Abkürzung: 1080i/25**

- System 3 (S3) mit 1920 aktiven horizontalen Abtastwerten und 1080 aktiven Zeilen in progressiver Abtastung und einer Vollbildrate von 25Hz, 16 x 9 Bildformat.  
Abkürzung: 1080p/25
- System 4 (S4) mit 1920 aktiven horizontalen Abtastwerten und 1080 aktiven Zeilen in progressiver Abtastung mit einer Vollbildrate von 50Hz, 16 x 9 Bildformat.  
Abkürzung: 1080p/50

HDTV Systeme EBUTech 3299	Horizontal samples	Active lines	Frame Rate	Coloursampling/ Quantisation (Bit)	Net image Bit Rate [Gbit/s]
System 1 720p/50	1280	720	50	4:2:2	10 0,9216
System 2 1080i/25	1920	1080	25	4:2:2	10 1,0368
System 3 1080p/25	1920	1080	25	4:2:2	10 1,0368
System 4 1080p/50	1920	1080	50	4:2:2	10 2,0736

Empfohlene HDTV Systeme für Europa (EBU Tech 3299)

## Die Abtaststrukturen

### *RGB (4:4:4) und YCRCB (4:2:2)*

Das RGB-Signal bietet die volle Farbauflösung mit einem Verhältnis der Abtastfrequenzen von 4:4:4. Dieses Verhältnis beschreibt den Zusammenhang der jeweils verwendeten Abtastfrequenzen für die drei Farbsignale R, G und B. Die 4:4:4-Abtastung wird überwiegend für die szenische (High-Quality) Produktion (z. B. Kamera, Filmabtastung, Farbkorrektur) eingesetzt.

Im Gegensatz dazu arbeitet das YCRCB-Signal mit einem Verhältnis der Abtastfrequenzen von 4:2:2. In diesem Fall beschreibt das Verhältnis den Zusammenhang der Abtastfrequenzen des Luminanz-Signals Y und der Chrominanz-Signale  $C_R$  und  $C_B$ . Im Falle einer 4:2:2-Abtastung besitzen die Chrominanz-Signale (nur jeder zweite Bildpunkt wird abgetastet) nur die halbe Farbauflösung (Bandbreite) des Luminanz-Signals.

Das entspricht in etwa dem Verhalten des menschlichen Auges, bei dem das Schärfeempfinden für Helligkeitssignale in etwa doppelt so hoch ist wie für Farbsignale. Das 4:2:2-Signal benötigt gegenüber einem 4:4:4-Signal weniger Bandbreite (Datentransferrate) und erfüllt somit sowohl die qualitativen als auch die wirtschaftlichen Anforderungen für den Regelfall in der Fernsehproduktion.

### *Weitere Unterabtaststrukturen (Subsampling)*

Wird zur 4:2:2-Abtastung (Regelfall) eine zusätzliche Unterabtastung (Subsampling, z. B. 4:2:0 oder 3:1:1) angewandt, dann wird die benötigte Bandbreite (Datentransferrate) weiter reduziert. Dies führt aber gegenüber dem Regelfall 4:2:2 sowohl zu einer Reduzierung der Farbauflösung und/oder Luminanzauflösung, als auch zu einem reduziertem Qualitätspotential für mehrfache Bearbeitungsschritte.

#### **Abtaststruktur 4:1:1**

In dieser Abtaststruktur werden die Farbinformationen im Gegensatz zum 4:2:2 weiter verringert, um Übertragungsbandbreite einzusparen. Der Farbraum 4:1:1 erstellt an jedem vierten Y-Sample-Punkt jeder Zeile entsprechende  $C_R$ - und  $C_B$ -Samples. Diese Abtaststruktur wird vor allem im Format DVCPRO eingesetzt, also im SD-Aufzeichnungsformat. Trotzdem hat diese Abtaststruktur mehr Chrominanzdetails als PAL oder NTSC.

#### **Abtaststruktur 4:2:0**

Hier werden die Chromianzwerte  $C_R$  und  $C_B$  jeweils aus zwei benachbarten Zeilen berechnet. Diese Berechnung wird an jedem zweiten Bildpunkt durchgeführt. MPEG2-Codierungen arbeiten mit dieser Abtaststruktur, die heute nicht mehr den technischen Lieferanforderungen für HD entspricht. Fast alle Sendeanstalten Europas akzeptieren Aufzeichnungsformate mit dieser Abtaststruktur nur in absoluten Ausnahmefällen.

#### **Abtaststruktur 3:1:1**

Wie die beiden vorher genannten Abtaststrukturen ist auch 3:1:1 ein sogenanntes Subsampling. Diese wird insbesondere dazu eingesetzt, um die zu speichernde oder zu überragende Datenrate bereits vor einer Videokompression zu verringern. Die Abtaststruktur 3:1:1 wird vornehmlich bei HDCAM eingesetzt. DVCPRO HD arbeitet mit einer Abtaststruktur von 3:1,5:1,5. Mit der daraus resultierenden reduzierten Farbauflösung sind z. B. Keys nur wenig zufriedenstellend umsetzbar.

### Abtaststruktur 4:4:4

Im Gegensatz zu  $YC_{\text{B}}C_{\text{R}}$ -Signalen, die im Regelfall eine 4:2:2-Abtaststruktur aufweisen, wird für RGB-Signale eine Abtaststruktur von 4:4:4 angewandt. Diese Abtaststruktur benötigt eine sehr hohe Datenrate und wird derzeit ausschließlich im Premiumbereich eingesetzt (z. B. Redcam oder Arri D21). Die Abtaststruktur 4:4:4 wird aktuell außerdem für Abtastungen von Film und in der High-End-Postproduktion genutzt. Film wird in der Regel auf RGB 4:4:4 gescannt, um bei digitalen Intermediates bestes Arbeiten bei der Farbkorrektur zu ermöglichen. HDCAM SR kann mit einer Abtaststruktur von 4:4:4 aufzeichnen und wird deshalb bei Filmabtastungen als Speichermedium genutzt. Es ist anzunehmen, daß mit fortschreitender technischer Entwicklung und verbesserten Möglichkeiten der Datenspeicherung diese Abtaststruktur auch eine größere Rolle spielen wird.

### Zusammenfassung

HD hat eine sehr viel höhere Datenrate als SD. In der folgenden Tabelle sind einige Beispiele für Datenraten im Kontext der jeweiligen Auflösung genannt. Es wird dabei von unkomprimiertem HD mit 10 Bit Quantisierung sowie einer Abtaststruktur von 4:2:2 ausgegangen.

Abtastformat	Auflösung	Video-Datenrate	Datenmenge für 2 Stunden Material
720p/50	1280 x 720	0,9216 Gbit/s	0,892 TB
1080i/25	1920 x 1080	1,0638 Gbit/s	0,957 TB
1080p/25	1920 x 1080	1,0638 Gbit/s	0,957 TB
1080p/50	1920 x 1080	2,0736 Gbit/s	1,914 TB

Die hohe Datenrate ist das Hauptproblem, mit dem HD derzeit kämpft. Alle Fragen der Datenreduktion, in entsprechenden Verfahren und mit entsprechend entwickelten Codecs, hängen damit zusammen, daß die hohen Datenraten heute noch nicht wirtschaftlich effizient speicherbar und verarbeitbar sind. Um es populär zu sagen: wir erzeugen wunderbare Bilder in HD, „zerstören“ diese jedoch sofort wieder, um sie speichern oder übertragen zu können. Diese „Zerstörung“ muss sehr durchdacht sein, denn sie darf möglichst nur soweit gehen, daß die zerstörten Daten auch wiederherstellbar sind. Deshalb kommt heute in Akquise und Verarbeitung von HD-Material der Datenreduktion und damit der Encodierung und Decodierung eine so zentrale Rolle zu. Darauf wird auf den folgenden Seiten überblickhaft eingegangen.

## Datenreduktion

Die heute in der Akquise erzeugten HD-Signale sind in transparenter Form nicht praktikabel speicherbar. Dies ist das eigentliche Grundproblem, mit dem HD zu kämpfen hat. Wäre eine originäre Aufzeichnung möglich und vor allem im weiteren Produktionsprozess verarbeitbar (Schnitt), müssten wir auf die folgenden, zum Teil komplizierten Vorgänge nicht eingehen. Es ist anzunehmen, daß die Technik in der Zukunft die völlig unkomprimierte Speicherung und Verabreichung originärer HD-Signale auch kostengünstig möglich machen wird. Im Moment jedoch werden wir uns mit verschiedenen Formen der Datenreduktion weiter befassen müssen, um HD-Signale zu speichern.

Sowohl in der Umgangssprache als auch in der Literatur wird in der Rundfunk-Domäne inzwischen der Begriff „Datenreduktion“ mit dem Begriff „Kompression“ gleichgesetzt und benutzt. Dabei sollte man allerdings nie vergessen (insbesondere im Dialog mit IT-Kollegen), dass „Kompression“ ursprünglich für eine *verlustlose* Komprimierung und Dekomprimierung steht. Wogegen in der Rundfunk-Domäne fast alle eingesetzten Kompressionsformate (eigentlich Datenreduktionsverfahren) eine *verlustbehaftete* Komprimierung und Dekomprimierung beinhalten.

Bei Datenreduktionsverfahren wird vornehmlich versucht, die nötige Datenrate zu verringern. Hierbei sollte beachtet werden, dass in der praktischen Umsetzung häufig bereits vor den eigentlichen Datenreduktionsverfahren (Videokompressionsformate) eine Reduzierung der Datenrate in den angelieferten Signalkomponenten vorgenommen wird. Das kann über die Reduzierung der folgenden Parameter geschehen:

- Auflösung (z. B. zusätzliches Subsampling von 1920 x 1080 Pixel nach 1440 x 1080 Pixel)
- Abtaststruktur (Reduzierung von z. B. 4:2:2 auf 3:1:1 oder 4:2:0)
- Quantisierung (z. B. von 10 Bit auf 8 Bit)

Die Anwendung dieser Möglichkeiten hat bereits Einfluss auf die ursprüngliche Signalqualität (Bildqualität) und die Qualitätsreserve für nachfolgende Bearbeitungsschritte kann dadurch unter Umständen erheblich verringert werden. Die eigentlichen Datenreduktionsverfahren von Bildern nutzen vor allem die folgenden Bildeigenschaften:

**Relevanz:** relevante Bildinformation, die zwingend notwendig ist  
**Redundanz:** zeitlich/örtlich gleiche Bildteile im gleichen Bild oder in aufeinanderfolgenden Bildern.

**Irrelevanz:** Bildteile die ohne Qualitätsverlust weggelassen werden können.

Bei den im Folgenden dargestellten Verfahren (Bausteine) zur Datenreduktion wurde stark simplifiziert, um verständlich zu bleiben. „Eingefleischte“ Techniker werden Informationen vermissen, doch die Ausführungen sollen der Praxis helfen und nicht noch mehr Verwirrung stiften. Nachfolgende Verfahren der Datenreduktion die Bestandteil vieler Kompressionssysteme sind werden thematisiert:

- die DCT (Diskrete Cosinus Transformation) eines digitalen Videosignals, z. B. SDI, HD-SDI,
- die Quantisierung der DCT-Koeffizienten,
- die variable Längencodierung VLC (Variable Length Coding)

## Das Verfahren DCT

Die diskrete Kosinus-Transformation (DCT) ist ein Verfahren zur Übertragung der Werte der Pixel von der Ortsebene in die Frequenzebene. Dieses Verfahren wird als Baustein vieler digitaler Videokomprimierungssysteme wie JPEG, MPEG2 und MPEG4 verwendet. DCT führt zwar nicht unbedingt direkt zu einer Verringerung der Daten, aber die Bildinformationen werden so strukturiert, dass die Daten verringert werden können. Dabei werden die diskreten Werte eines Pixels umgewandelt in die spektrale Verteilung. Da die Darstellung als Frequenz bei einem Pixel keinen Sinn macht, werden Blöcke definiert. In den meisten Fällen werden Pixelblöcke (Makroblöcke) von 8x8 Pixel untersucht. Die Bildinformation in dem Block wird dann mit Hilfe der DCT in den Frequenzbereich transformiert. Somit gibt es 64 Informationen über den Pixelblock, die genauso wieder umgerechnet werden können.

Wichtig ist dabei, dass die Transformation selbst weder Informationen aus dem ursprünglichen Signal entfernt, noch Informationen hinzufügt. Die DCT-Codierung selbst bedeutet keinerlei Datenreduktion. Die DCT-Transformation kann als verlustlos bezeichnet werden, obwohl es in der Praxis einige mathematische Rundungsfehler gibt. Diese sind auf eine real machbare Begrenzung der Rechengenauigkeit zurückzuführen. Mögliche Datenreduktionsverluste hängen also nur von den nachfolgenden Prozessen ab, denen das DCT-Signal durch das jeweilige Kompressionsverfahren unterworfen wird.

## Verfahren Quantisierung der DCT-Koeffizienten

Die nun entstandenen DCT-Koeffizienten werden mit Hilfe der anschließenden Quantisierung gewichtet. Diese Gewichtung berücksichtigt für jeden einzelnen Koeffizienten die spezifischen Eigenschaften unseres Sehannes. Die derart gewichteten Koeffizienten werden dann auf ganze Zahlen gerundet. Die Kombination von Gewichtung und Rundung wird als „Quantisierung“ der DCT-Koeffizienten bezeichnet. Je genauer die entsprechende Quantisierung stattfindet, desto genauer kann das Signal wiederhergestellt werden. Man kann die Quantisierung bewusst ‚schlecht‘ einstellen, erreicht damit eine sehr kleine Datenrate, muß dafür aber Bildqualitätseinbußen in Kauf nehmen. Die Quantisierung ist ein entscheidender Schritt innerhalb eines Kompressionsverfahrens, denn er ist mit unwiederbringlichen Verlusten behaftet!

## Das Verfahren VLC

Die eigentliche Datenreduktion/Kompression wird erst durch die variable Längencodierung (VLC) erzielt. Der VLC-Prozess ist vollständig verlustlos. Dieses Verfahren untersucht den anliegenden Datenstrom auf Werte, die mit hoher Wahrscheinlichkeit öfter vorkommen. Diese Werte werden mit einem kürzeren Codewort versehen; die selteneren Werte mit einem längeren Codewort. Die variable Längencodierung wird als ein Baustein in einem Codec verwendet und immer im Zusammenhang mit der Quantisierung benutzt. Die variable Längencodierung ist eine Redundanzreduktion.

Als weiterer Baustein zur Reduktion der resultierenden Datenrate wird neben der Redundanz innerhalb eines Bildes (Intra-Frame) zusätzlich auch die Redundanz zwischen aufeinanderfolgenden Bildern (Inter-Frame) ausgenutzt. Dies ist insbesondere für solche Varianten von Kompressionsformaten relevant die für die digitale Ausstrahlung genutzt werden (geringere Datenrate = geringere Kosten). Bekannteste Beispiele dafür sind die sogenannten Long-GOP Varianten von MPEG-2 und MPEG-4.

## Verfahren Long-GOP

Mit Hilfe von Bildspeichern versucht das Kompressionssystem, das nächste, am Kompressionseingang erwartete Bild aus den schon bekannten Informationen der bereits übertragenen Bildinformationen vorherzusagen. Nur die Differenz zwischen der Vorhersage und dem tatsächlichen Bild wird dann der DCT-Transformation zugeleitet. Je exakter die Vorhersage, desto kleiner wird die Differenz und

lässt sich deshalb auch auf eine wesentlich geringere Datenrate komprimieren, als das vollständige Bildsignal. Die Vorhersage beinhaltet auch eine Bewegungsanalyse. Daraus resultierende „Bewegungsvektoren“ müssen dem komprimierten Ausgangssignal wieder hinzugefügt werden, damit der Decoder bei der Rekonstruktion eine identische Vorhersage eines Bildes durchführen kann.

Anzumerken ist hier, dass jede Ungenauigkeit in der Bewegungsschätzung nicht zwangsweise zu Störungen im decodierten Bild, aber in jedem Fall zu einer Erhöhung der Datenrate am Ausgang des Encoders führt.

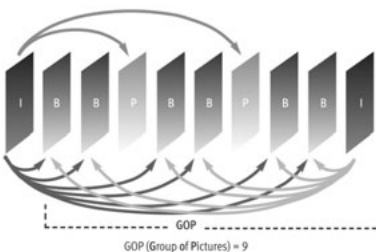
Wesentlich ist auch der Fakt, dass ein Decoder die Bilder nicht nur aus übertragenen Differenzbildern rekonstruieren kann. Von Zeit zu Zeit muss daher ein vollständig komprimiertes Eingangsbild, das keine Prädiktion beinhaltet, zum Decoder übertragen werden. Diese Bilder werden als I-Frame (Intra-Frame) bezeichnet. Bei dieser Technik können in einer Gruppe von Bildern (Group of Pictures GOP) drei Bilderarten (Frames) definiert werden:

- **I-Frames** (Intra-Frame): Referenz- oder Stützbild, das ohne einen Bezug zum vorangegangenen bzw. nachfolgenden Bild codiert wurde,
- **B-Frames** (Bidirektional-Frame): Ein vorhergesagtes Bild, das sowohl aus dem vorangegangenen als auch dem folgenden Bild berechnet wurde,
- **P-Frames** (Predicted-Frames): Ein einseitig vorhergesagtes Bild, das aus dem vorhergehenden Bild berechnet wurde.

Der Abstand zwischen zwei I Frames darf nicht zu groß sein, denn sonst werden B-Frames und P-Frames ungenau. Die Anordnung der Frames nennt man „Group of Pictures“ (GOP). Es gibt unterschiedliche Zusammenstellungen dieser GOP. Zum Beispiel:

I B P B I = 4er GOP

I B B P B B P B B I = 9er GOP, alle 9 Frames ein I Frame.



Beispiel für eine GOP-Struktur aus den drei Frame-Typen (MPEG-2)

## Verfahren Long-GOP in der Fernsehproduktion

Long-GOP Verfahren nutzen die Redundanz zwischen aufeinanderfolgenden Bildern eines Videosignals und stellen damit ein hocheffizientes Kompressionsverfahren für die Programmausstrahlung dar, z. B. MPEG-2, MPEG-4. Zu beachten ist, dass sie in erster Linie nicht für den elektronischen Bildschnitt entwickelt wurden. Ihre Anwendung in der Fernsehproduktion erfordert zusätzliche ergänzende Festlegungen. Zum Beispiel ist die Einführung von 4:2:2-Profilen zur Vermeidung der 4:2:0 Unterabtastung der Chrominanz ein Schritt in diese Richtung. Auch konventionelle Bearbeitungsschritte wie Editing und Umschalten zwischen zwei komprimierten Signalen bedürfen zusätzlichen Aufwands. Ein Bildschnitt innerhalb einer GOP-Sequenz erfordert z. B. aufwendige Splicing-Techniken und entsprechende Leistungsfähigkeit, z. B. Rechenleistung, der beteiligten Gerätschaften. Aufgrund dieser relativ komplexen Handhabung wurden in der Vergangenheit Kompressionsformat-Varianten die mit einer GOP-Struktur arbeiten in der professionellen Fernsehproduktion nicht eingesetzt.

Mit der Einführung von HD in die Fernsehproduktion sind neue bandlose Akquisitionssysteme verfügbar, die eine Long-GOP Kompressionsformat-Variante verwenden. Zum Beispiel verwendet das XDCAM-HD422 Format von Sony eine MPEG-2 Long-GOP Variante mit einer 4:2:2 Abtaststruktur, 8Bit-Quantisierung und einer Videodatenrate von 50 Mbit/s. Damit wird für die professionelle HD-Produktion in einem IT-basierten Umfeld eine attraktiv „niedrige“ Datentransferrate unterstützt.

## Das Verfahren Wavelet

Beim Wavelet-Verfahren werden im Gegensatz zur DCT keine Blöcke gebildet, sondern es wird jeder Bildpunkt für sich betrachtet. Das zu komprimierende Bild wird erst zeilenweise transformiert und anschließend in der Vertikalen (spaltenweise) bearbeitet. Nach einem Filterprozess werden die entstandenen Koeffizienten quantisiert und in einen Datenstrom eingefügt. Das Wavelet-Verfahren wird heute z. B. in der JPEG2000-Komprimierung verwendet und erfreut sich aufgrund seiner sehr hohen Qualität zunehmenden Einsatzes. Im digitalen Kino wie auch in neuen Kameras wie z.B. der REDCAM wird das Verfahren bereits eingesetzt.

## Komprimierte Datenraten

Die mit oben genannten Verfahren komprimierten Datenraten unterteilen sich in:

### Feste Datenrate:

z. B. eine Voraussetzung für Bandspeicherung. Hier besteht die Gefahr, dass eine zu niedrige Datenrate nicht für alle relevanten Bildteile ausreicht. So können deutlich sichtbare Qualitätsverluste entstehen (z. B. HDV mit 25 Mbit/s).

### Variable Datenrate:

Günstig für eine effiziente Nutzung der Speicherkapazität, z. B. Festplatten oder Blu-Ray. Innerhalb der jeweils vorhandenen Begrenzung kann die Datenrate der Komplexität des zu komprimierenden Bildinhalts effizient angepasst werden. Dadurch kann bei einfachen Bildvorlagen Speicherkapazität eingespart werden, die dann für schwierige Bildvorlagen zusätzlich zur Verfügung steht.

## Interoperabilität von Kompressionsformaten

Zu beachten ist insbesondere, dass einige Kompressionsformate keine feste Regeln für die Kompression (Encodieren) von Fernsehsignalen vorschreiben, sondern einen „Werkzeugkasten“ anbieten, mit dem unterschiedlichste Versionen erstellt werden können (z. B. MPEG-2, MPEG-4). Diese Flexibilität kann bei der Ausstrahlung und Verteilung von Fernsehprogrammen einen großen Vorteil darstellen, z. B. wenn die verfügbare oder bezahlbare Bandbreite beschränkt ist. Diese Flexibilität kann jedoch für Anwendungen in der Fernsehproduktion auch erhebliche Nachteile im erforderlichen Zusammenspiel von beteiligten Gerätschaften verursachen. Dies beruht im Wesentlichen auf der Tatsache, dass in manchen Standards, z. B. MPEG-2, MPEG-4, nur der Decoder spezifiziert ist.

Daher ist es sinnvoll, für die Anwendung in der Fernsehproduktion, in der Datenströme von unterschiedlichen Encodern an „einer“ Stelle zusammentreffen (z. B. Editing) diese Flexibilität zu begrenzen. Das bedeutet, daß Einschränkungen vereinbart werden müssen, die eine Interoperabilität innerhalb eines Kompressionsformats für Encoder unterschiedlicher Hersteller sicherstellen.

Die Zukunft wird zeigen, inwieweit die Anwender dafür Sorge tragen werden, dass ihre Anforderungen von den Herstellern auch umgesetzt werden!

## Codecs

Der Begriff CODEC steht für *CO*mpression and *DE*Compression. Da es hier ganz unterschiedliche Wege gibt, die beschrieben werden können, gibt es auch sehr verschiedene Codecs (Nutzung unterschiedlichster Kompressionsverfahren). Sinn eines Codecs ist es zum einen, ein Bild zu komprimieren, um Daten zu sparen und damit die Übertragung der Bildinformationen bezahlbar zu machen. Zum anderen sorgt ein Codec auch für die Dekomprimierung, also die Wiederherstellung aller zwischenzeitlich weggelassenen Bildinformationen. Im besten Fall ist dieser Vorgang verlustfrei (siehe Datenreduktionsverfahren). Ein häufiges Kodieren/Dekodieren geht zu Lasten der Bild- und Tondaten.

Insbesondere im Bereich der Codecs bieten heute verschiedene Hersteller in einem bemerkenswert hohen Tempo immer wieder neue Lösungen an. Die Erfahrungen der vergangenen Jahre führen dazu, daß sich Codecs immer weiter verbessern. Doch Vorsicht: nicht jeder neue Codec ist auch immer der beste Weg für die Komprimierung/Dekomprimierung.

Es gibt aktuell eine recht unüberschaubare Anzahl von Codecs. Das erschwert professionelles Arbeiten – und die Inkompatibilität von Codecs ist ein zentrales Problem bei HD-Workflows. Kompliziertes Transcodieren, also die Wandlung von Bildinformationen aus einem Codec in einen anderen, ist die Folge. Auch sind bei weitem nicht alle Codecs, die heute auf dem Markt sind, auch für professionelle Arbeit tauglich. Zunehmend verweisen TV-Anstalten darauf, welche Codecs ihren technischen Anforderungen entsprechen und welche nicht (siehe Kapitel Technische Richtlinien).

Mit Codecs wird heute an allen Stellen einer Produktionskette gearbeitet – und meistens mit verschiedenen. Ein Camcorder zeichnet mit einem bestimmten Codec auf, das Material wird dann im Schnittplatz zumeist transcodiert und dann wieder für einen Master-Datenstrom (Tape oder File) dekomprimiert. Je weniger Komprimierungs/Dekomprimierungsvorgänge über den gesamten Produktionsprozess erfolgen, umso besser ist es für das Material. Deshalb ist es manchmal sogar sinnvoller, Material mit einem bestimmten Codec stärker zu komprimieren, den gesamten Produktionsprozess über aber mit diesem Codec konsequent zu arbeiten, als Material fast unkomprimiert aufzuzeichnen und in der Postproduktion dann immer wieder durch verschiedene Komprimierungs/Dekomprimierungsverfahren zu „jagen“.

Auf den folgenden Seiten sollen die Ende 2009 gängigen professionellen Codecs kurz vorgestellt werden. Auch hier gilt: diese Ausführungen sind bewußt simpel und überblickhaft gehalten und keine technische Fachdarstellung. Es wird versucht, die Unterschiede zwischen den einzelnen Codecs zu verdeutlichen und Codecs damit für die Praxis und Entscheidungsfindungen handhabbar zu machen. Die folgenden Codecs sollen kurz angerissen werden:

<b>Codec</b>	<b>genutzt in Aufzeichnungs-Formaten</b>
MPEG-2	HDV, XDCAM, XDCAM HD, XDCAM HD422
MPEG-4	HDCAM SR
AVC-Intra/H.264	P2HD
JPEG2000	RAW
DNxHD	AVID