Multimedia

Vorlesungsfolien

WS 2014 / 2015

Christian Breiteneder

Institut für Softwaretechnik und Interaktive Systeme
Interaktive und Multimediale Systeme

Technische Universität Wien

Zur Organisation

- Zeit: DO 11:00 - 12:30, 12:50-13.50
- Unterlagen (in *Englisch*)
  Kopien der Vorlesungsunterlagen: Mitte November
  Folien: TISS
- schriftliche Prüfung: letzte Woche im Januar

Contents

- What is a multimedia system? (prologue)
- Media Types (audio, video, ...)
- Media Platforms
- Media Compression (JPEG, MPEG-2, MPEG-4, MP3)
- MM Programming Abstractions
- MM Content-Based Information Retrieval und Indexing
- Content Description (MPEG-7 + MPEG-21)

References

1. *Multimedia Programming. Objects, Environments and Frameworks*, S.Gibbs, D.C. Tsichritzis, Addison-Wesley, ACM Press, 1995.
2. *Multimedia: Computing, Communications and Applications,* R.Steinmetz, K.Nahrstedt, 2nd edition, Prentice Hall, 1999.
3. *Video and Image Processing in Multimedia Systems,* B. Furht, S.W. Smoliar, H. Zhang, Kluwer Academic Publishers, 1995.
4. *Image and Video Compression Standards. Algorithms and Architectures,* V. Bhaskaran, K. Konstantinides, 2nd edition, Kluwer Academic Publishers, 1997.

References

5. *Schall&Klang: Wie und was wir hören.* Georg Eska, Basel; Boston; Berlin; Birkhäuserverlag 1997.
6. *Physikalische und psychoakustische Grundlagen der Musik.* Juan G. Roederer, 3. Auflage, Springer 2000
7. *Fundamentals of Multimedia*. Ze-Nian Li and Mark S. Drew, Prentice Hall, 2004.
8. *The Science of Digital Media*. Jennifer Burg, Pearson Education, 2009.

Prologue:
Multimedia—Definitions

- Multimedia—General Definition
- Media Classification
- *Multimedia Systems Definition*

Multimedia—General Definition

- *Multi-* [lat.: much] many; much; multiple
- *Medium* [lat.: middle]
  - an intervening substance through which something is transmitted or carried on; a means for distribution and presentation of information; e.g., atmosphere, water, text, ..
  - a means of mass communication; e.g., newspaper, magazine, television, ...

Media Classification

- *Perception media* help humans to perceive their environment. *How do humans perceive information in a computer environment?*
- *Representation media* are characterized by internal computer representations of information. *How is the computer information coded?*
- *Presentation media* refer to the tools and devices for i/o of information. *Through which medium is information delivered by the computer, or introduced into the computer?*

Media Classification

- *Storage media* refer to a data carrier which enables storage of information. *Where will the information be stored?*
- The *transmission medium* characterizes different information carriers, that enable *continuous* data transmission (storage media excluded!). *Over what will the information be transmitted?*
- The *information exchange medium* includes all information carriers for transmission, i.e., all storage and transmission media. *Which information carrier will be used for information exchange between different places?*

Representation—Space and Dimension

- *representation spaces* are part of the presentation media for information output; e.g., movie screen, stereo space
- each representation space consist of one or more *representation dimensions*; e.g. computer screen (2), stereophony (3)
- time may occur inside each representation space as additional dimension; *central* to multimedia systems;
  media are divided into 2 types with respect to time in their information space: *time-independent* and *time-dependent*

Representation

- time-independent (discrete) media—information consists exclusively of a sequence of elements or a continuum without any time reference; e.g., text, graphics, ...
- time-dependent (continuous) media—information consists of timed sequences, i.e., element values plus time of occurrence; e.g., audio, video, sensor signals, ...

  processing of these media is *time-critical*

Properties of a MM System

- not every *combination of media* justifies the term *multimedia*; e.g., a text processing program with incorporated images
- the *level of independence* of different media is important; a computer-controlled VCR vs. a text presentation with synchronized audio
- *computer-supported integration*—Computer controlled-data of independent media can be processed and integrated by a programmer or multimedia author.

Multimedia System—Definition

A multimedia system is characterized by computer-controlled, integrated production, manipulation, presentation, storage and communication of independent information, which is encoded at least through a continuous (time-dependent) and a discrete (time-independent) medium. *(Steinmetz/Nahrstedt)*

- MM is often used as an attribute of systems, products, etc., without satisfying the characteristics above. Thus, 2 notions of mm can be distinguished:
  - mm, strictly speaking (above)
  - mm, in the general sense
    used to describe the processing of individual images and text, although no continuous medium is present.

Part I Media Types

Representation, formats and operations of media types

- Text
- Graphics (Images)
- Audio
- Video

I.1. Media Type Text

Text Representations

- ASCII / ISO character sets
- marked-up text — e.g. SGML, HTML, XML, etc.
- hypertext — e.g. HyperCard, World Wide Web

Text Operations

- character and string operations
- editing
- formatting — interactive, non-interactive, page description languages
- pattern matching and searching — indexes, signatures
- sorting
- compression — Huffman, Lempel-Ziv
- encryption
- language specific operations — spelling checking, parsing, statistical analysis of writing style

I.2. Media Type Graphics (Images)

- Achromatic and Colored Light
- Intensities
- Color Models
- Image Representations
- Image Formats
- Image Operations

Light

- chromatic: of color
- achromatic: without color (i.e. B&W)
- color of an object depends upon:
  - object surface: reflectivity, physical properties, composition
  - light source(s) illuminating it
  - color of surrounding environment
  - visual system (eyes, brain, photoelectric cell, ...) of perceiving entity

Achromatic Light

- black, white, usually greys (rare to see pure black or pure white)
- quantity or energy of light is only measurement used for achromatic light
  - intensity, luminance, brightness
  - scale: 0 = black, 1 = white, greys in between
- eye is sensitive to ratios of intensity changes, not the changes in intensity values themselves
- small changes of low intensity seem to have the same effect as larger changes of high intensity
- dynamic range: ratio between min. and max. intensities

Selecting Intensities

- to find 256 intensities starting with the lowest attainable intensity I_0 and going to a maximum intensity 1.0, with each intensity r-times higher, we use:

  $I_1 = rI_0,\ I_2 = rI_1 = r^2 I_0,\ I_{255} = rI_{254} = r^{255} I_0 = 1$ therefore,

  $r = (1/I_0)^{1/255},\ I_j = r^j I_0 = r(1/I_0)^{j/255} I_0$

- example: n=3, I_0= 1/8 gives r=2 and intensity values of 1/8, 1/4, 1/2 and 1
- minimum attainable intensity I_0 for a CRT is between 1/200 and 1/40

Selecting Intensities

- how many intensities are enough?
- answer: r is 1.01 or less, therefore

  $n = \log_{1.01}(1/ I_0)$
- table on next page shows theoretical values

Dynamic Range

display media	typical dynamic range	number of intensities
CRT	50 - 200	400 - 530
photographic prints	100	465
photographic slides	1000	700
coated paper printed in b/w	100	465
coated paper printed in color	50	400
newsprint printed in b/w	10	234

Different Intensity Levels

intensity levels

2

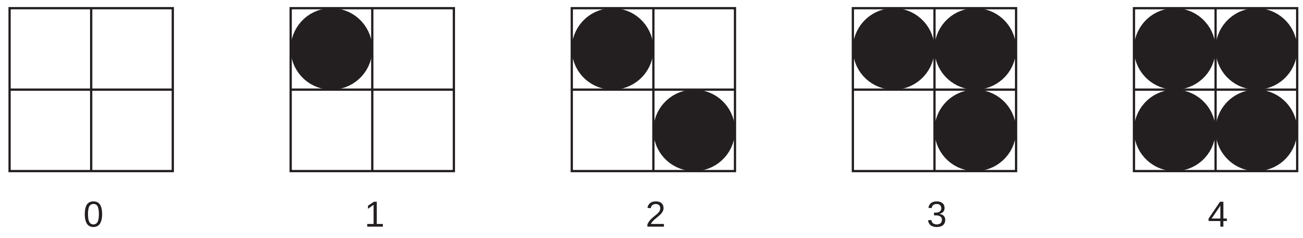
4

16

Halftone Approximation

- dithering (or half-toning): approximating intensities using dot patterns
  - human eye generalizes patterns into shades
  - similar effect with colour: approximate colours using multi-coloured pixel patterns: 8-bit colour graphics does this to "simulate" 24-bit colours (with varying success)
- simple approach: N device pixels per image point/pixel
  - create an N by N dither matrix: defines N*N + 1 intensities
  - to display intensity K, turn on pixels having values < K

Halftone Approximation



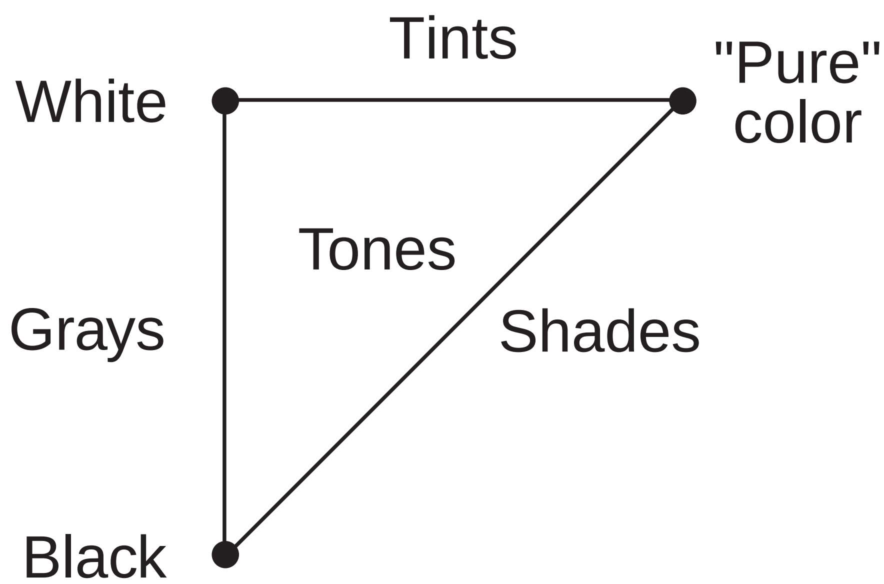
- cluster dithering: dither intensity centered on middle of matrix
  - avoid line patterns or other artifacts
  - patterns should grow - minimizes "contouring" effects
  - patterns should emanate from centre: effect of increasing dot sizes
- cluster dithering with diagonal periodicity

Color

- various models of colour used:
  - printing & graphics industry: colour sample books and codes
  - artists: tint (adding white to pure pigment), shade (adding black), tone (adding both)
  - CRT’s: use hardware models, eg. RGB
  - physicists: use optical models (wavelengths, energy measures)
- human perception of colour depends on how brain reacts to whole visual environment, as well as brain & eyebiological factors

Color

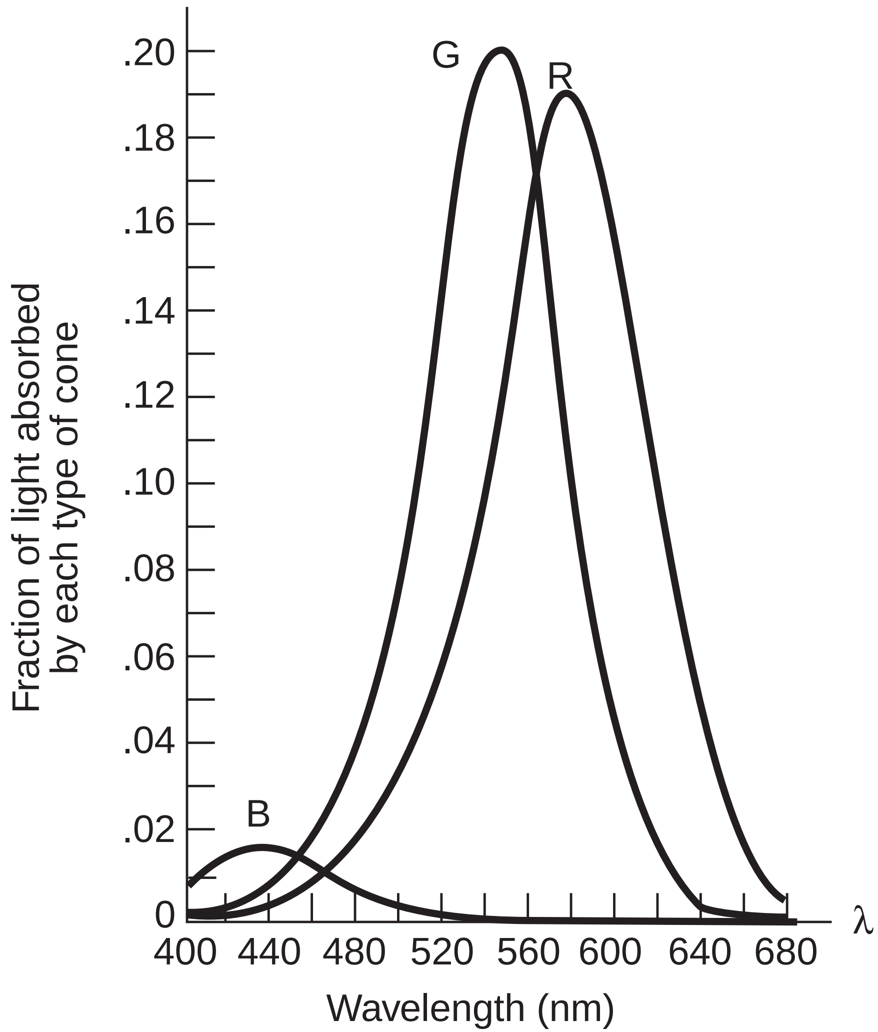
- colour is how we perceive beams of electromagnetic energy that fall on our retina
- we see light between 400 and 700 nanometers in wavelength (10^{-9} meters)

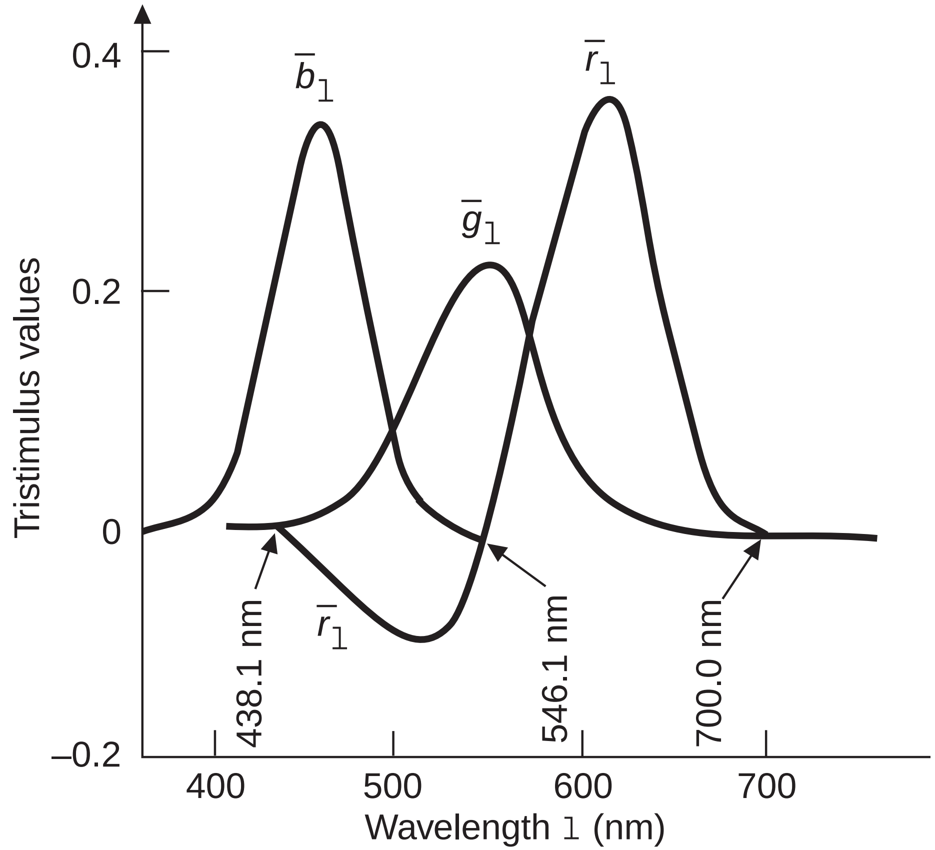


Tristimulus Theory

- the retina contains cones that sense either red, green, or blue light
- 6-7 million cones per eye, concentrated in central area called "fovea"
- rods: sensors that surround fovea, and perceive weak "night" images
- hence, use indirect vision at night to see low-light images
- tristimulus theory isn't totally accurate, because many colors we see are not exact combinations of red, green, blue wavelengths

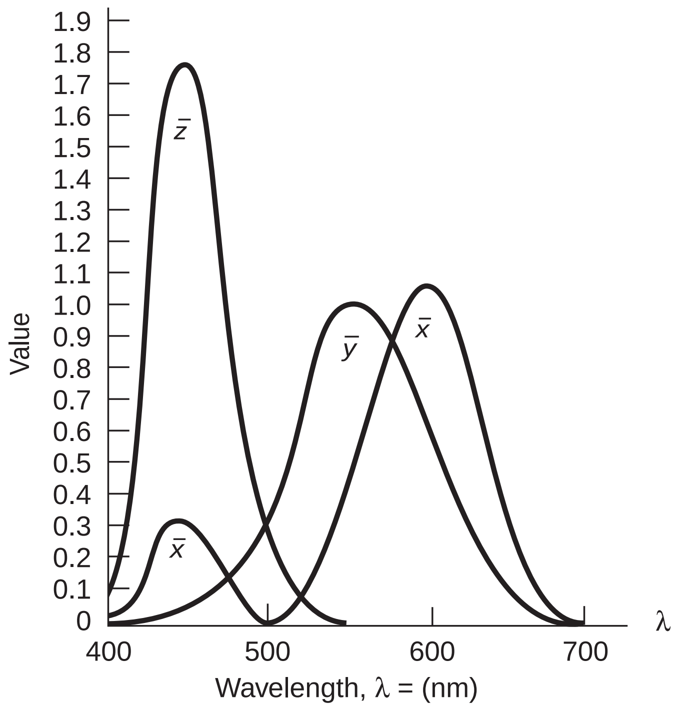
Tristimulus Theory

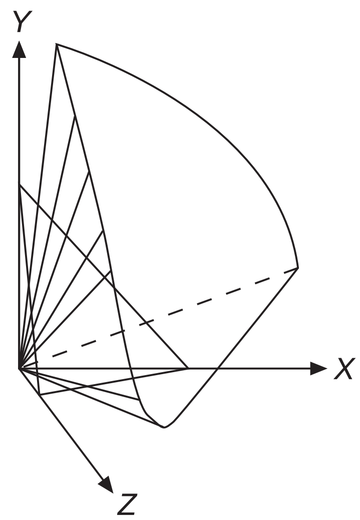


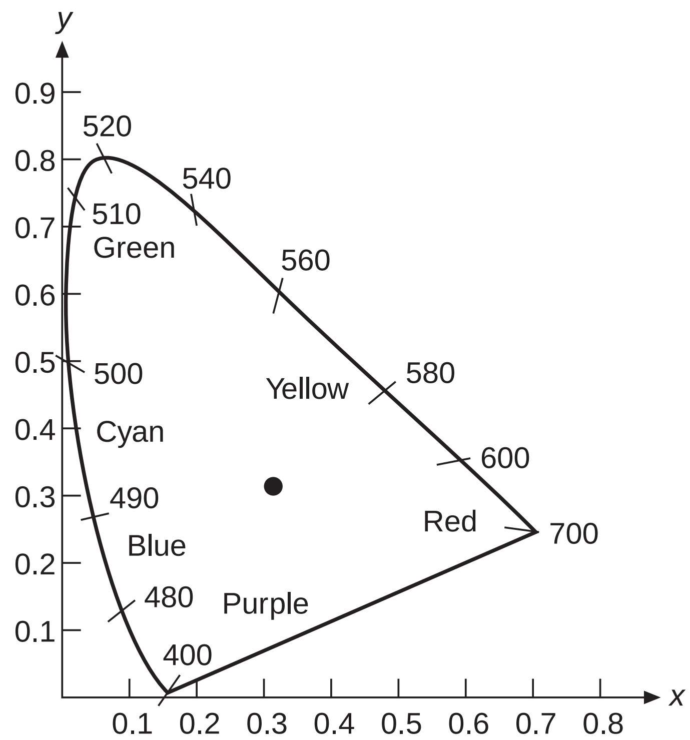


CIE Color Space

- many combinations of dominant wavelength, luminance, saturation appear the same to us
- CIE color space (Commission Internationale de l'Eclairage) used to calibrate other color models; 1931 CIE XYZ



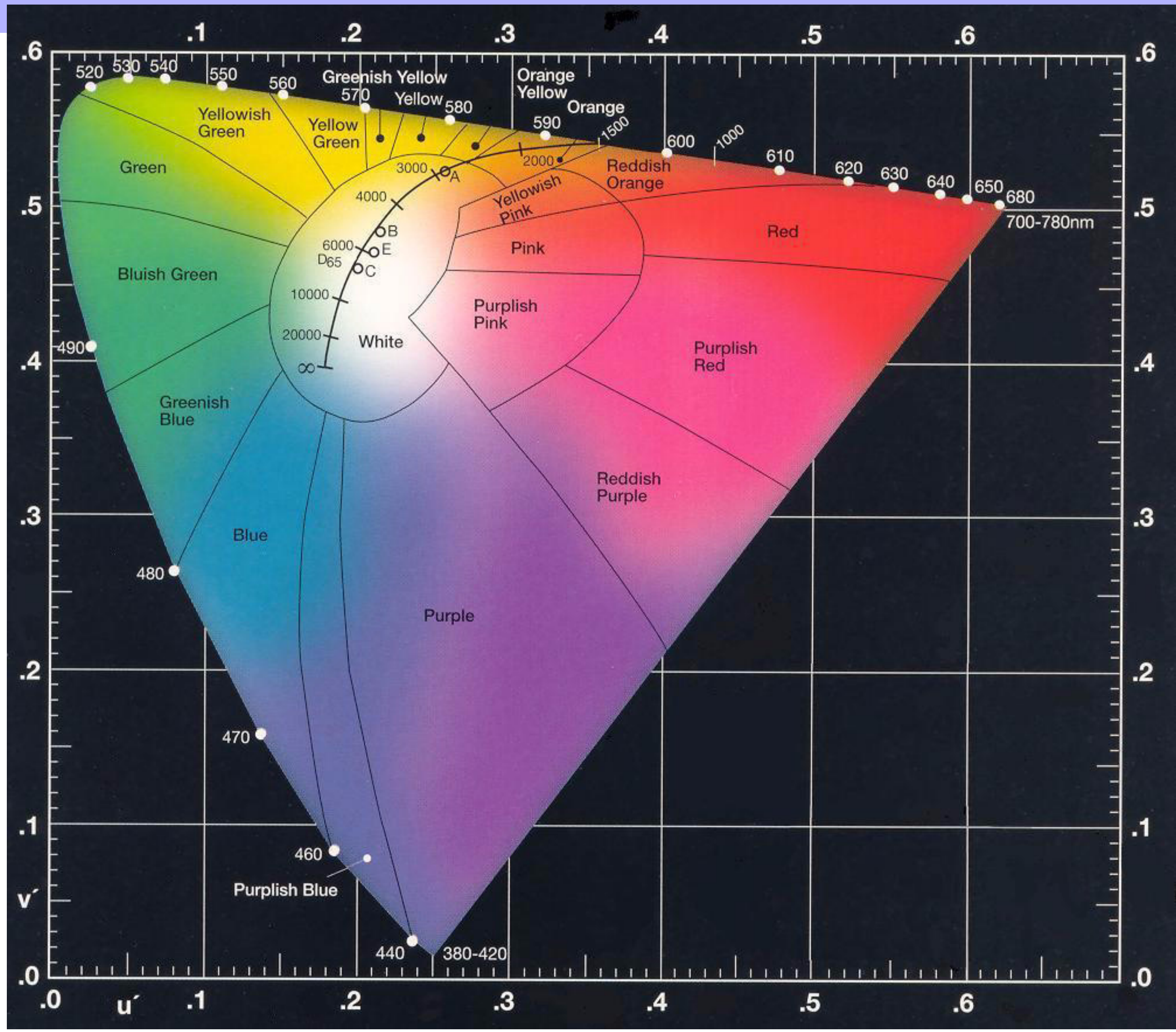




CIE Color Space

- 100% spectrally pure colors on outer curve edge; pure white near centre (at C)
- if color at A, then B is dominant wavelength: you mix B with C to get A
- F’s dominant wavelength defined as the complement of A’s ones
- complementary colors: mix together to get white (e.g.. E and D, or F and A)
- can therefore use CIE to:
  - identify and analyze colors
  - measure CRT performance: how much of CIE color space does a particular CRT cover?

CIE Chromaticity Diagram

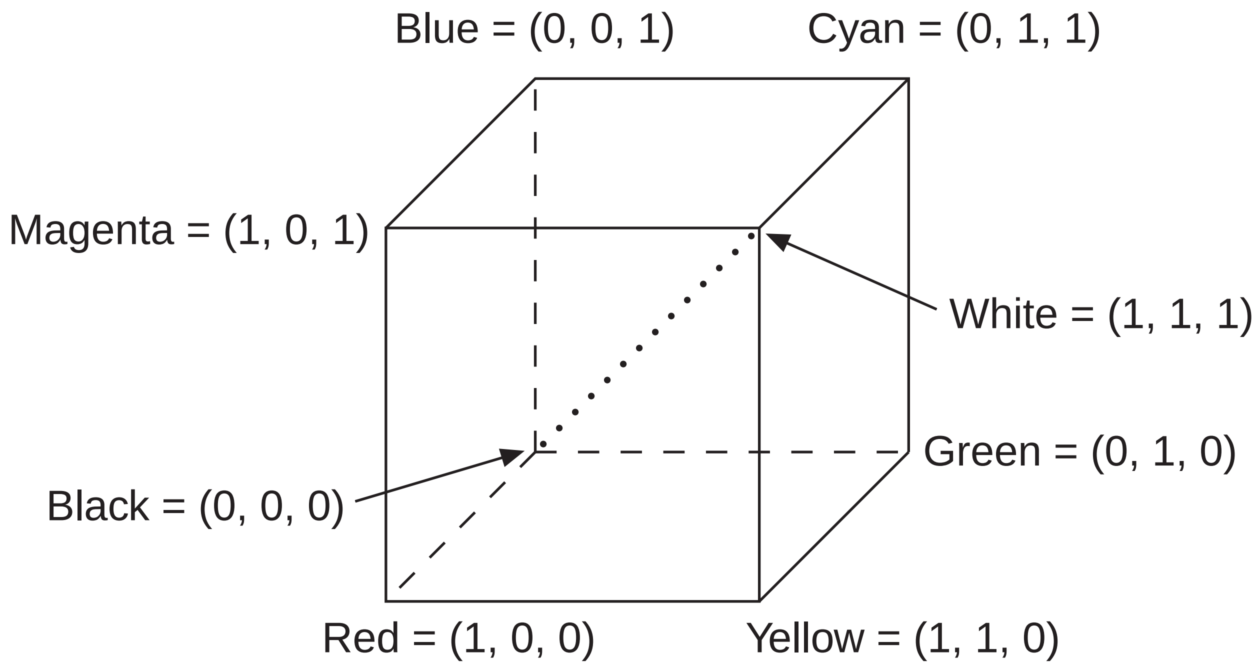


Color Models for Raster Graphics

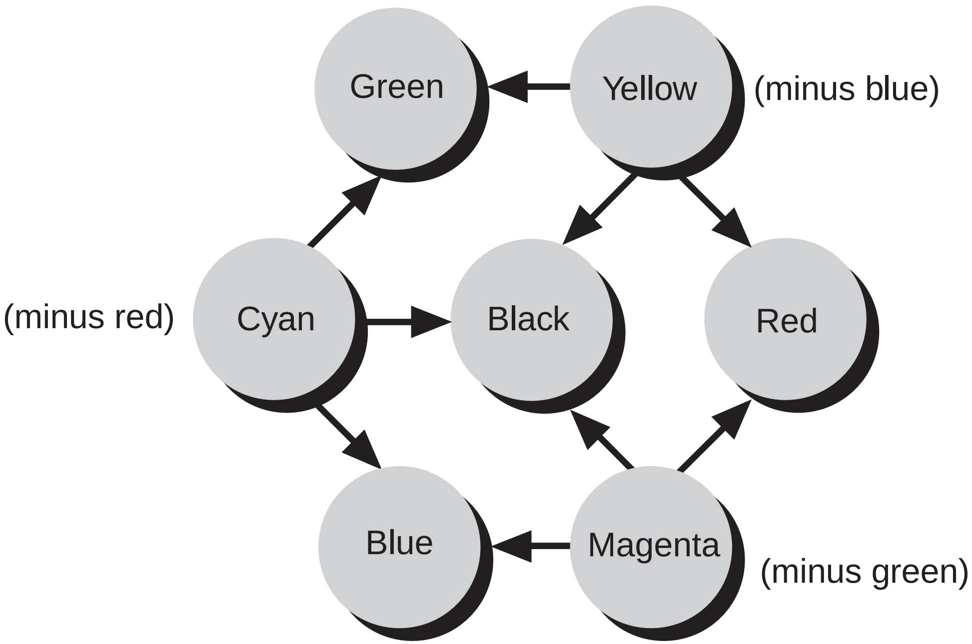
1. The RGB color model

- hardware-oriented model that turns on red, green, and blue pixels at varying intensities
- additive: add red, green, blue together to get overall color
- not all colors are exactly duplicable, and gamut (color range) covered by a CRT depends on performance of its phosphor technology

The RGB Color Model



2. The CMY Color Model

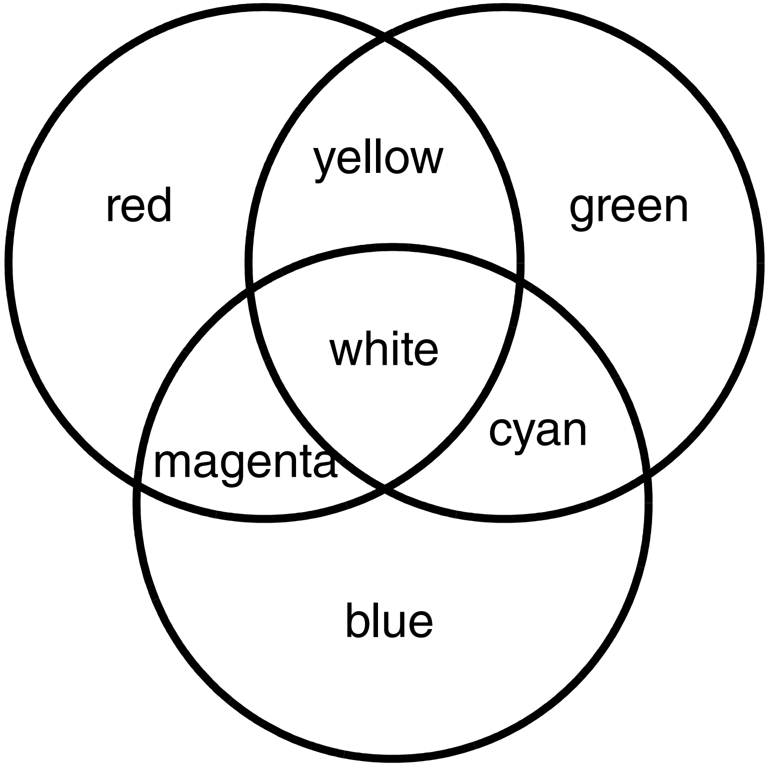


$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} C \\ M \\ Y \end{bmatrix}$$

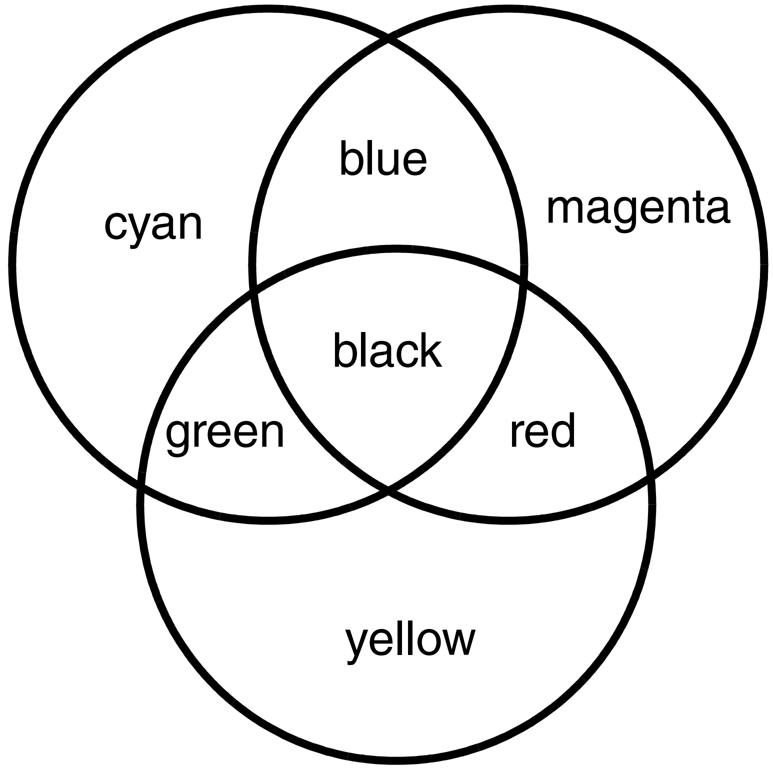
- another related model, CMYK, uses black as 4th color:

  K = min (C,M,Y), C = C - K, M = M - K, Y = Y - K

Direct and Reflected Light



direct light

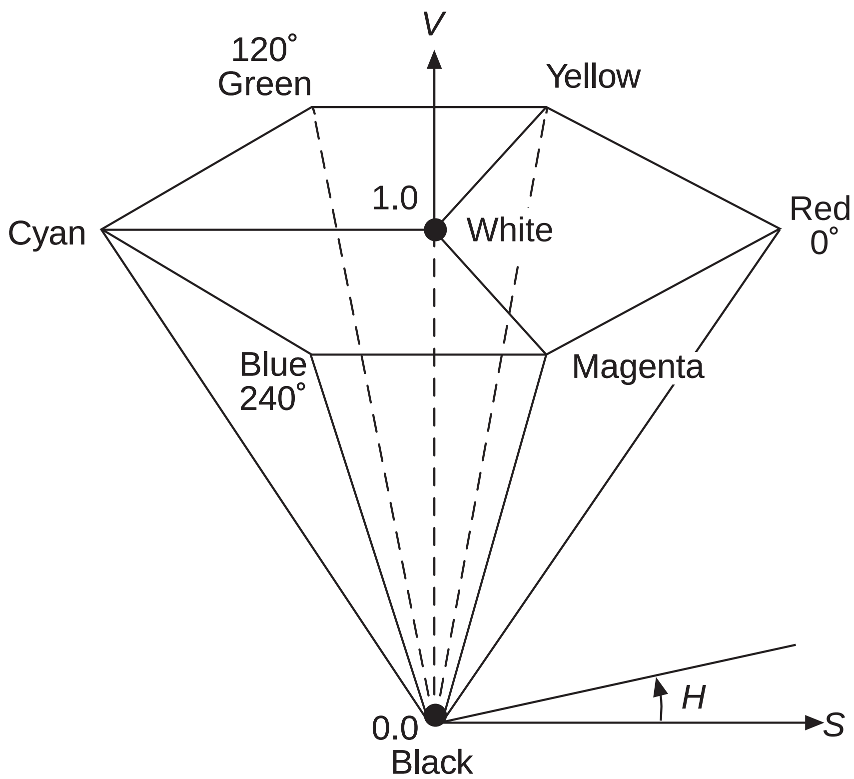


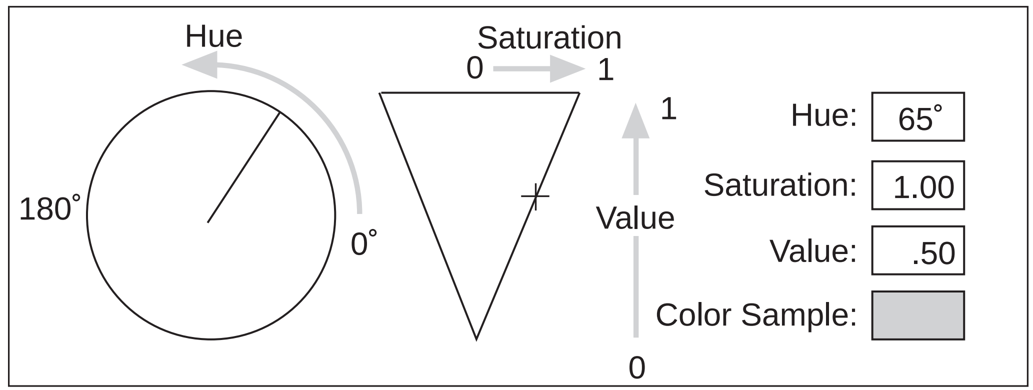
reflected light

The HSV Color Model

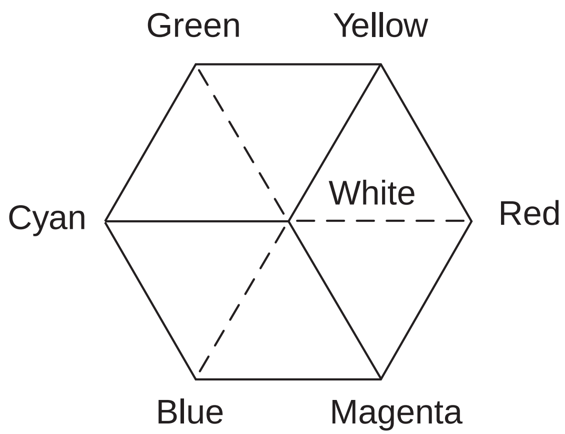
- HSV: “hue saturation value”; based on tint, shade, tone model; defines a hexcone geometry
- V=1: top of cone, bright pure color; V=0: bottom of cone, Black
- H: angle around vertical axis; red = 0 deg, green = 120 deg, ...; complementary colors are 180 degrees from one another
- S = 0: centre line (V axis), S =1 side of cone
- can convert between RGB and HSV using formulae
  - subcubes of RGB fit down the cone
  - tools (xv, Photoshop, Fractal Painter) give both sets of controls to define color

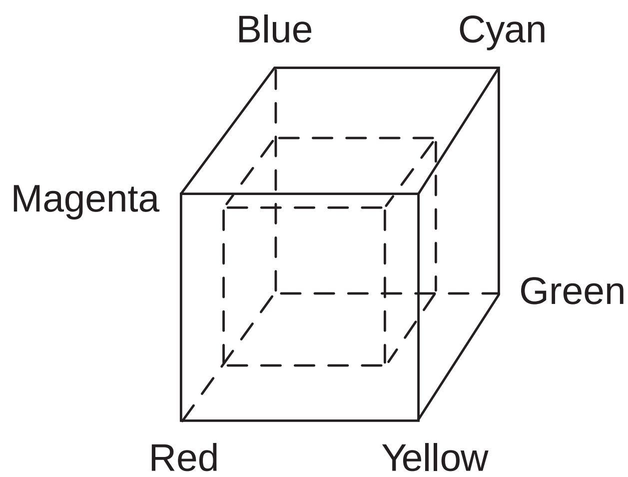
The HSV Color Model





The HSV Color Model





Color Models (Spaces) Overview

- CIE color space (Commission Internationale de l'Eclairage) used to calibrate other color models
  1931 CIE XYZ, *tristimulus* theory
- RGB — for video display drivers
- HSB — hue (dominant color), saturation (amount of gray), brightness (intensity)
- CMYK — cyan, magenta, yellow, black, *(subtractive primaries*)
- YUV — used in television industry, Y (luminance), UV (color difference signals)

Media Type Image—Example

image descriptor = {

image width = 640
image height = 480
image depth = 24
color model = RGB
encoding = YUV 8:2:2, JPEG}

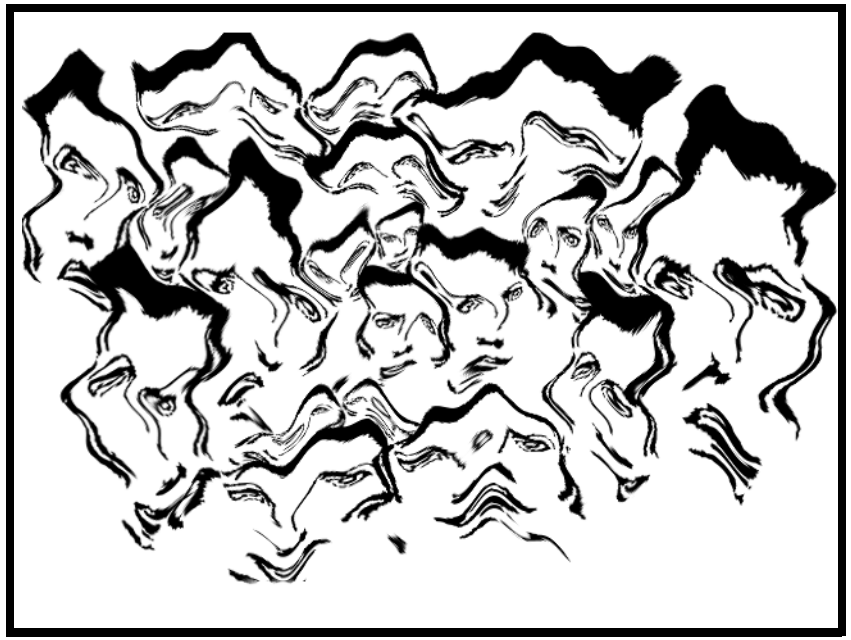
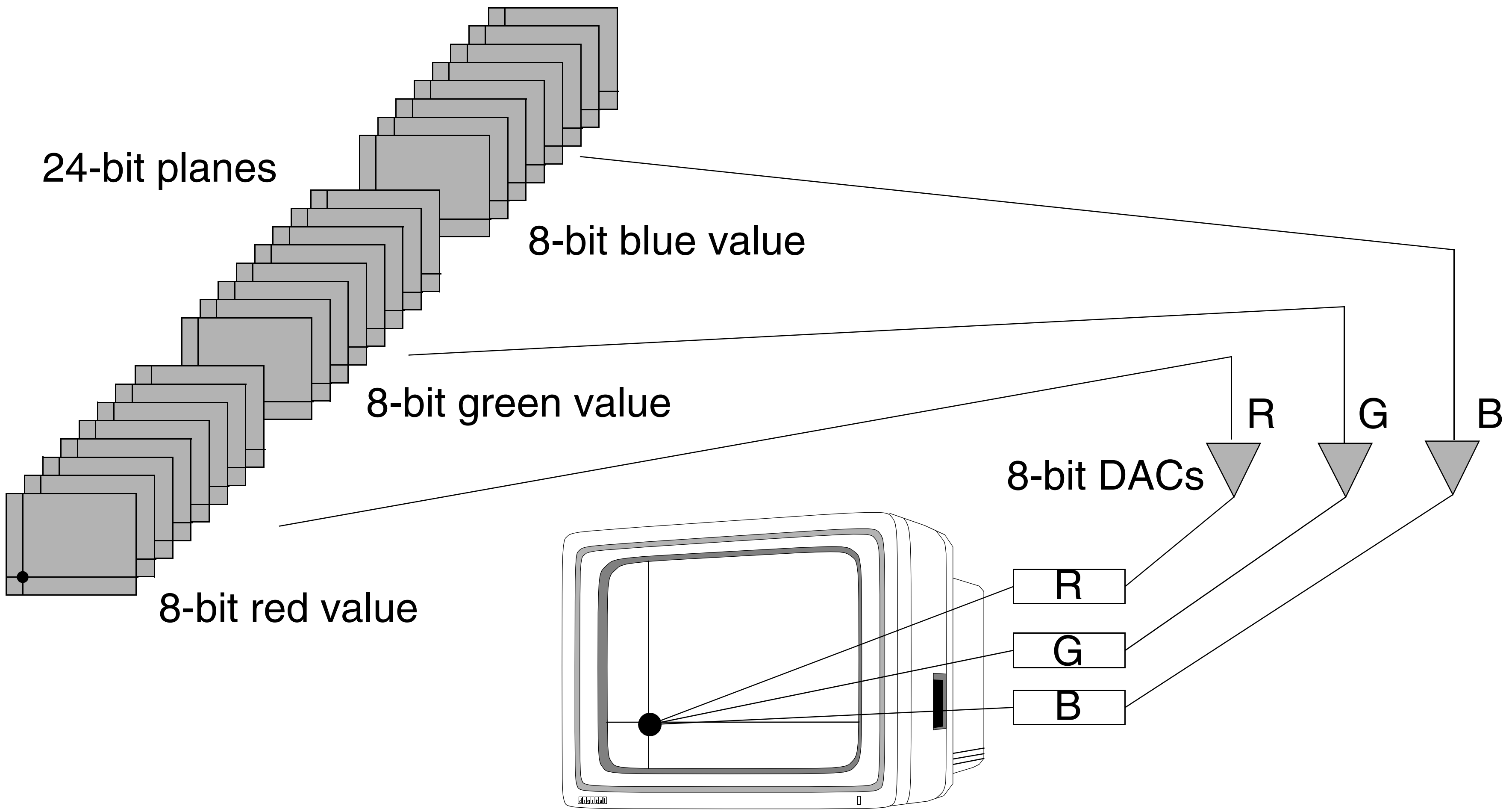


Image Representations

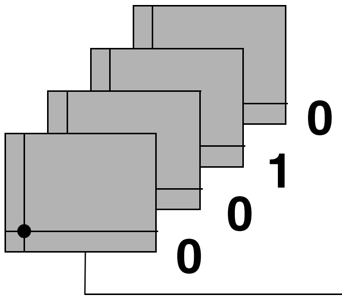
- number of channels — dim. of color model plus alpha channels
- channel depth — number of bits-per-pixel
- alpha — controls transparency; masks and blends
- pixel aspect ratio
- interlacing
- compression — lossless vs. lossy
- indexing — colors are represented by index in *color maps* or *color lookup tables* [CLUT]; map predefined or part of image

Color Display



Color Mapping

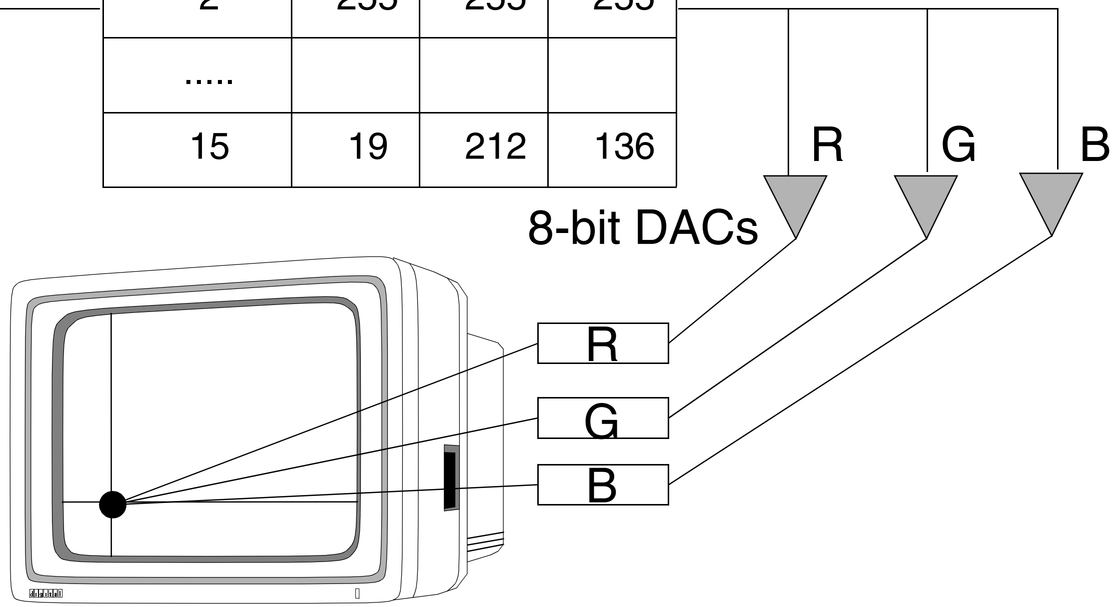
4-bit planes



0010 (2 decimal)

Color Look-up Table

register	R	G	B
0	0	0	0
1	127	123	72
2	255	255	255
.....			
15	19	212	136



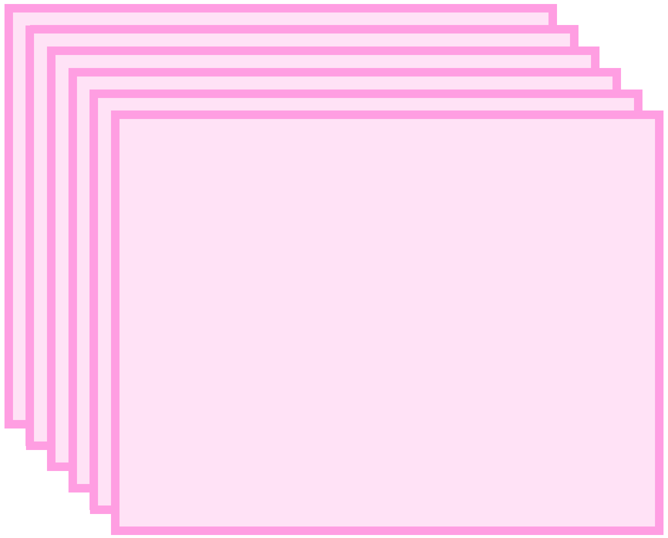
Some Image Formats

- EPS — Encapsulated PostScript
- GIF — Graphics Interchange Format
  CompuServe Information Service, originally for dialup lines
- Group 3 and Group 4 Facsimile
- PICT — may contain geometric data (lines, polygons, etc.)
- JPEG — Joint Photographic Experts Group, lossy

Image Operations

- editing — “paintbrush operations”, selection
- point operations — e.g. thresholding, color correction
- filtering — also values of neighboring pixels used
- compositing — alpha channels frequently used
- geometric transformations — e.g. scaling, warping
- conversions — e.g. format conversion, color separation

I.2 Media Type Video



Video is a

sequence of

FRAMES

Broadcasting History

1893 Telephone Audio Broadcasting (Puskas)

1895 Wireless Communication (Marconi, Popov)

1819 Radio Broadcasting (The Netherlands, Canada)

1935 TV Broadcasting (Germany, Britain)

1941 US b&w TV

1953 US color TV (NTSC)

1963 geostationary satellites

mid ’70s fiber optic transmission

1989 HDTV broadcasting (Japan)

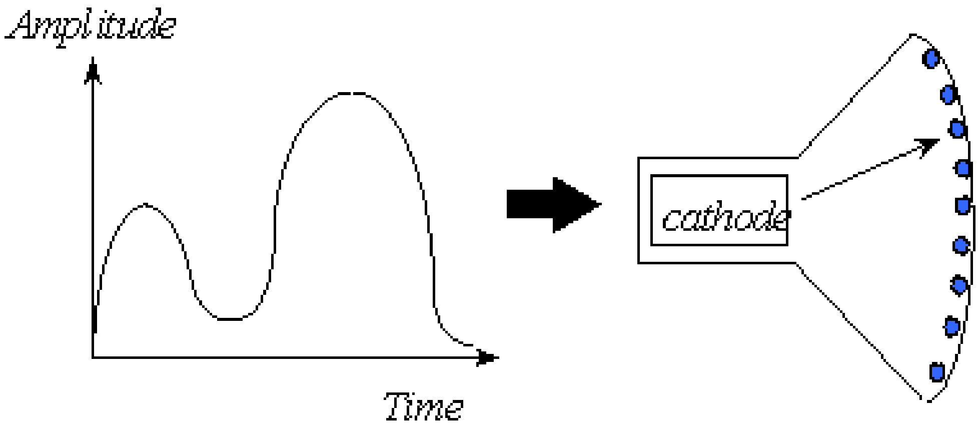
1996 ATSC - digital TV standard

Standards Groups

- ITU - International Telecommunication Union (formerly CCITT)
- SMPTE - Society of Motion Picture and Television Engineers
- MPEG - Motion Picture Experts Group
- ATSC - Advanced Television Systems Committee
- CCIR - Consultative Committee for International Radio
- ......

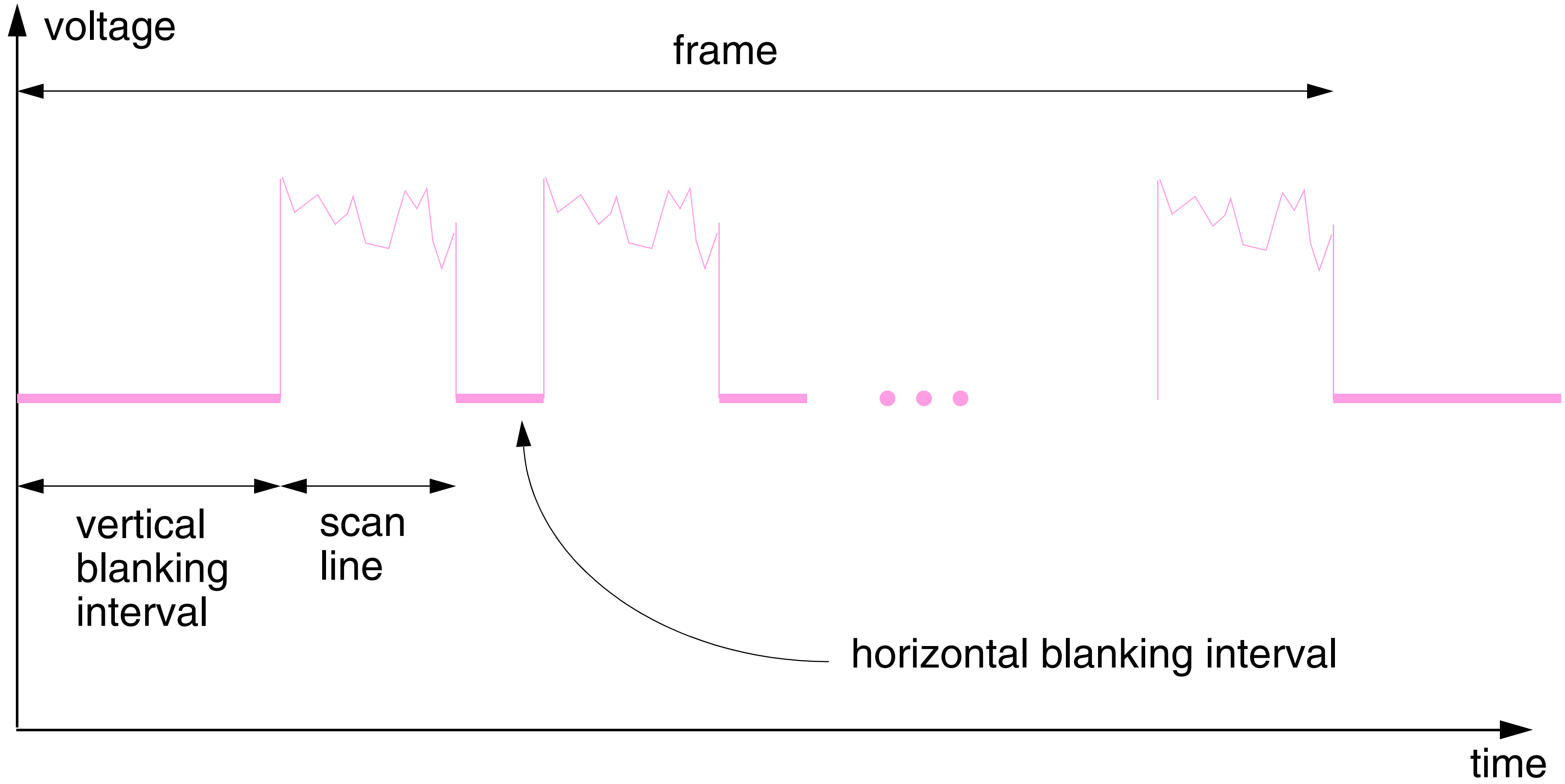
Video Display Scanning

- analog video is a continuous signal that drives a CRT

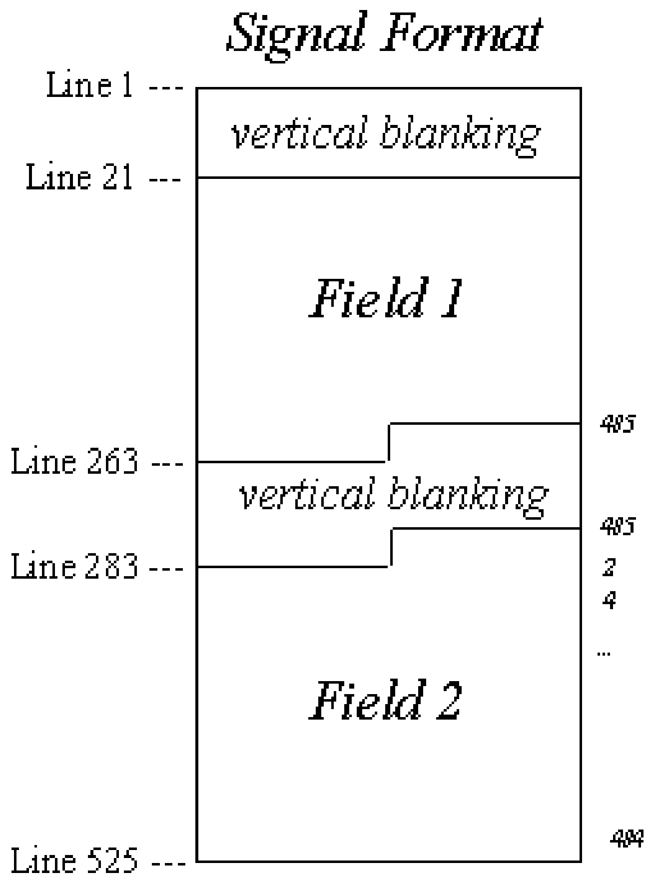


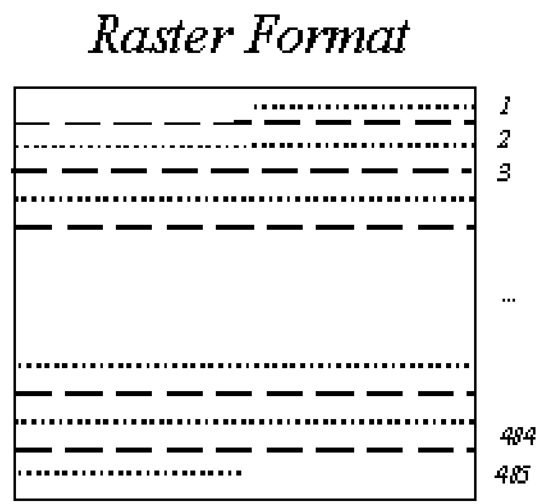
- video composed of luminance and chroma signals

Analog Video Signals



Interlaced Fields





Characteristics of Analog Video

❑ frame rate	(25, ~30)
❑ scan lines	(525, 625, >1000)
❑ aspect ratio	(4:3, 16:9)
❑ interlacing	(2:1, non-interlaced = progressive)
❑ signal quality	(consumer, professional, broadcast)

Characteristics of Analog Video

- composite versus component
- stability (editing degradation)
- audio tracks
- tape size

Video Representations

- component video
  each primary is sent as a separate signal - RGB or a luma/chroma transformation (YUV, YIQ, YC_RC_B, ...)
  requires more bandwidth and good synchronization
- composite video
  signals are mixed into a single carrier wave; interference
- S-video (separated video)
  a compromise between component and composite analog video, 2 lines, one for chroma and one for luma

Analog Video Signal Formats

- NTSC (National Television Systems Committee): North America, Central America and Japan, and some parts of the South Pacific and South Africa.
- PAL (Phase Alternation Line): western Europe , India, China, Australia, and parts of Asia and South America.
- SECAM (Séquentiel Couleur avec Mémoire): France, Eastern Europe, Russia, and parts of Africa and the Middle East.
- RGB: A component video signal format, used for computer displays. No single RGB standard.
- HDTV (High Definition Television): The earliest HDTV system, Japan´s Hi-Vision, is used for daily broadcasts in Japan (also known as MUSE)

Analog Video Signal Format Comparison

Signal format	Comp.	Frame rate (Hz)	Scan lines	Aspect ratio	Interlacing
NTSC	1	29.97	525	4:3	2:1
YUV 525/60	3	29.97	525	4:3	2:1
PAL	1	25	625	4:3	2:1
SECAM	1	25	625	4:3	2:1
YUV 625/50	3	25	625	4:3	2:1
RGB	3	ca. 25-75	ca. 200-1000	Varies	Usually 2:1
1125/60 (Hi-Vision, MUSE)	3	30	1125	16:9	2:1
1250/50 (HD-MAC)	3	25	1250	16:9	2:1

PAL

Phase Alternating Line

- 625 scan lines repeated 25 times per second (40msec / frame)
- interlaced scan lines divide frame into 2 fields each 312.5 lines
- aproximately 20% more lines than NTSC
- YUV color model
- lines reserved for control information at the beginning of each field, only 575 lines of visible data
- each line lasts 64 *u*sec (12 *u*sec blanked)

NTSC

National Television Systems Committee

- 525 scan lines per frame, 30 frames per second (actually 29.97)
- interlaced scan lines divide frame into 2 fields each 262.5 lines
- YIQ color model
- lines reserved for control information at the beginning of each field, only 485 lines of visible data
- each line lasts 63.5 *u*sec (10.5 *u*sec blanked)

Video Representations

- PAL (Phase Alternation Line)

  $Y = 0.299R + 0.587G + 0.114B$

  $U = 0.492(B-Y)$

  $V = 0.877(R-Y)$

- NTSC (National Television Systems Committee)

  $Y = 0.299R + 0.587G + 0.114B$

  $I = 0.596R - 0.275G - 0.321B$

  $Q = 0.212R - 0.523G + 0.311B$

- new ATSC (Advanced Television Systems Committee)

  1920x1080 progressive 16:9

  1280x720 progressive 16:9

Analog Video Tape Formats

signal quality

signal quality	Format	Use
↓	VHS	consumer
	Betamax	
	Video 8	
	S-VHS	
	Hi 8	professional/ industrial
	Umatic	
	Betacam	broadcast
	1 inch	

Composite vs. Component Video

composite video

- single electrical signal
- + easy to broadcast
- + simple cabling
- - poor quality

component video

- multiple electrical signal
- - difficult to broadcast
- - cabling more difficult
- + good quality

examples:

RGB, YUV, Y/C

Video Equipment

- routing switcher (matrix switch) - nxn cross bar circuit switch; analog or digital; can switch several signals; one input can be routed to multiple outputs
- distribution amplifier - split 1 input signal to 2 or more outputs
- timebase correctors (TBCs) - reconstructs signal to remove timing errors introduced by VTR/VCR
- sync generator - master clock to provide sync pulse for all equipment
- frame (delay) buffer - to synchronize external sources
- video production switcher

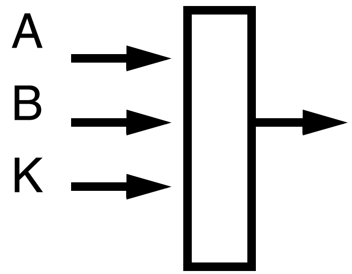
Video Production Switcher

- to produce high quality visual images - n input streams and 1 output stream (live events and post-production)
- special effects - titles, transitions, picture-in-picture, chroma-key, compositing
- implementation
  - conventional solutions use custom-designed hardware
  - modern approaches use more and more software

Video Production Switcher

mixing

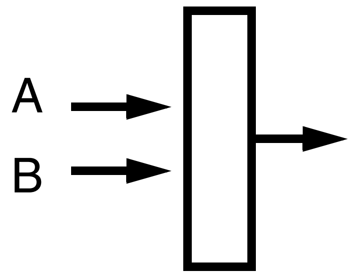
keying



if K then A else B

blending

wipes

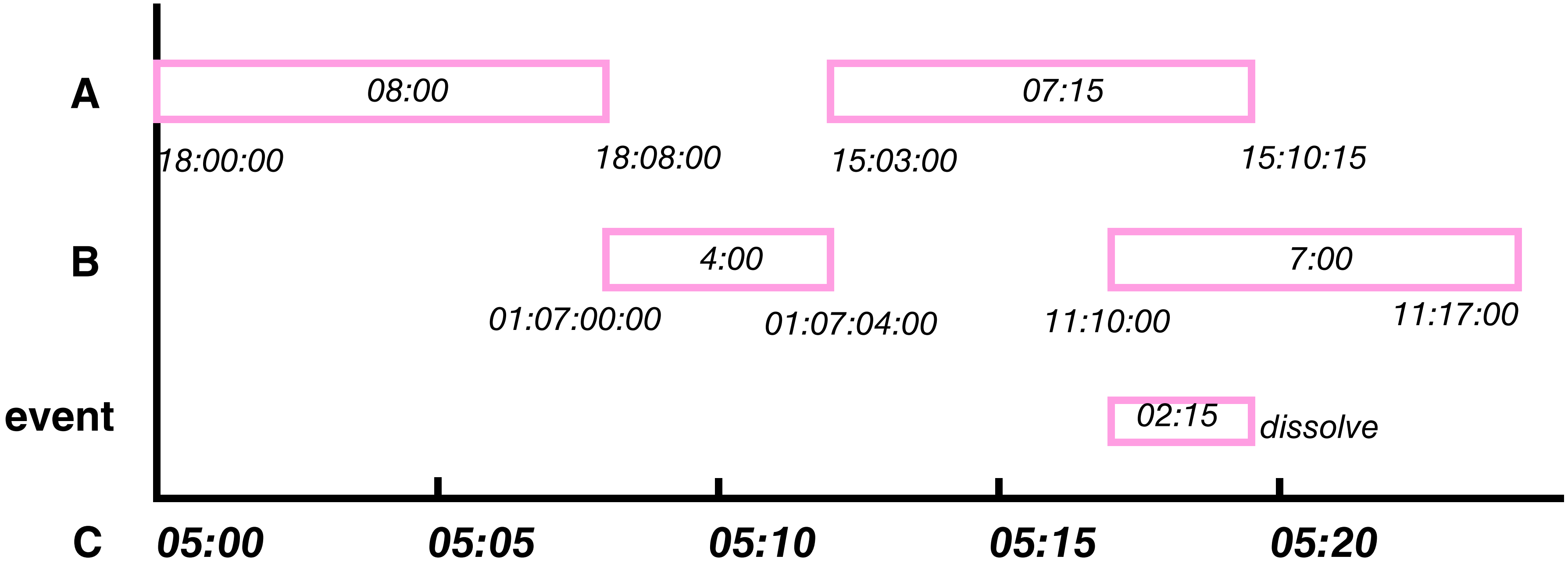


f * A + (1-f) * B

DVEs

key framing

Video Editing / EDL

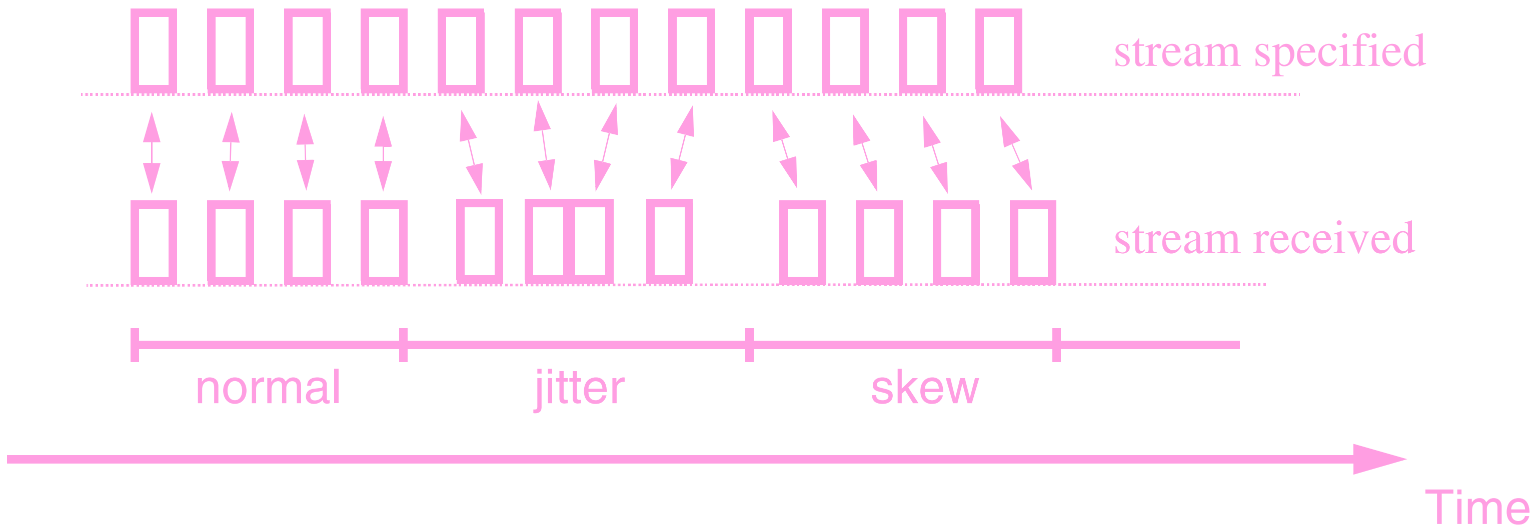


Video Time Codes

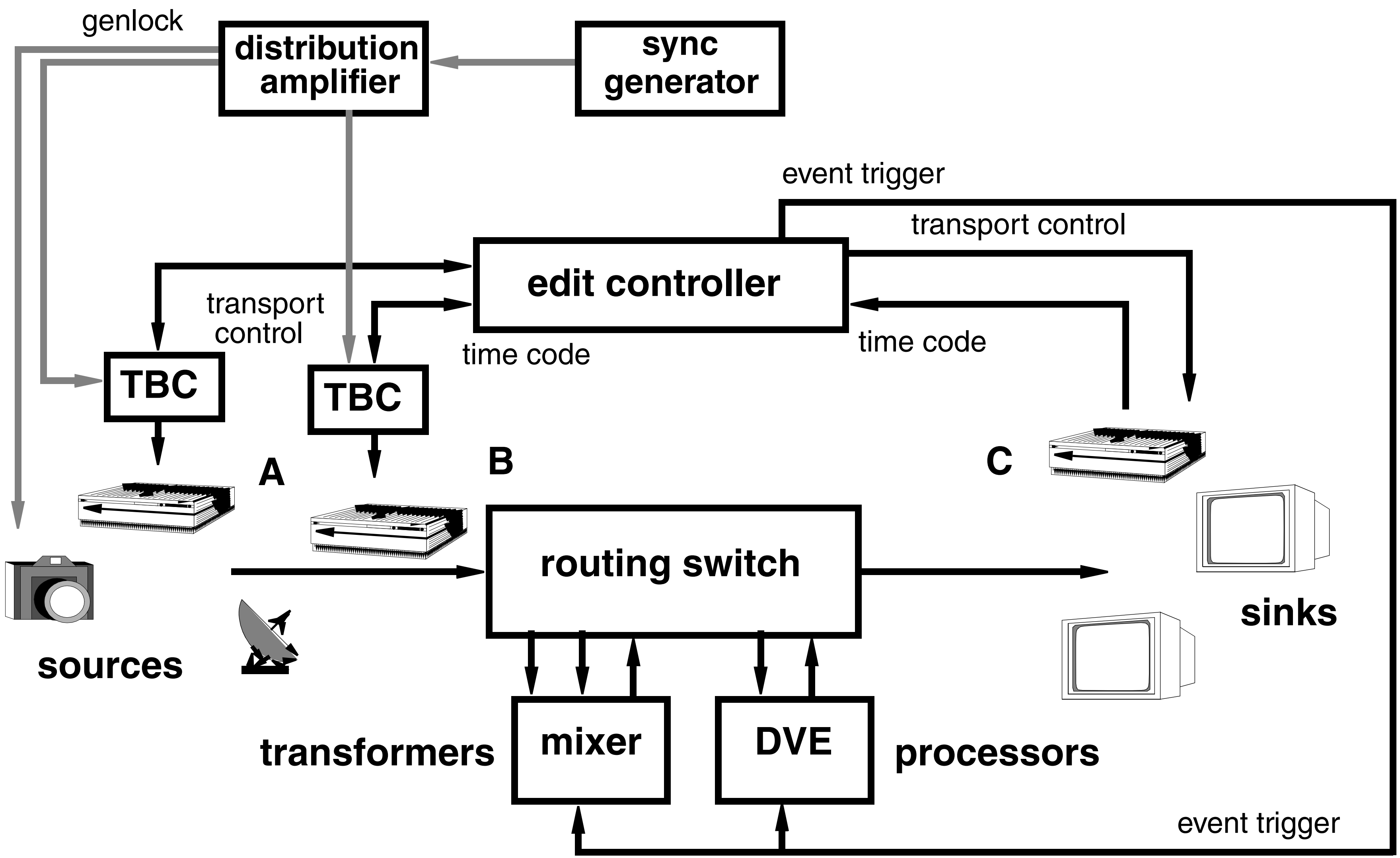
- SMPTE time code (HH:MM:SS:FF)
  - non-drop frame time code - FF in [0, 29]
  - drop frame time code - FF in [2, 29], except every tenth minute [0, 29]
- time code recorded on tape
  - LTC - longitudinal time code, recorded in audio track
  - VITC - vertical interval TC, recorded in vertical blanking
  - RCTC - rewritable consumer time code, Sony Video-8 ad Hi-8

Video Synchronization

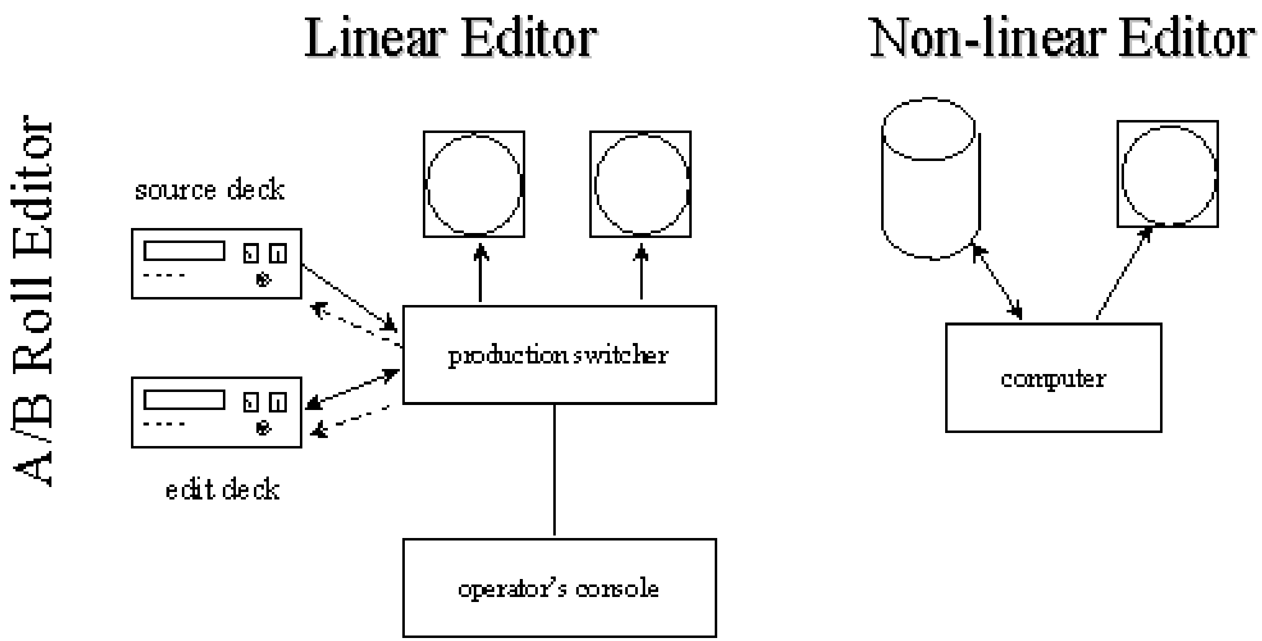
- intra-flow synchronization (skew, jitter)



Video Studio



Video Editing Comparison



Digital Video

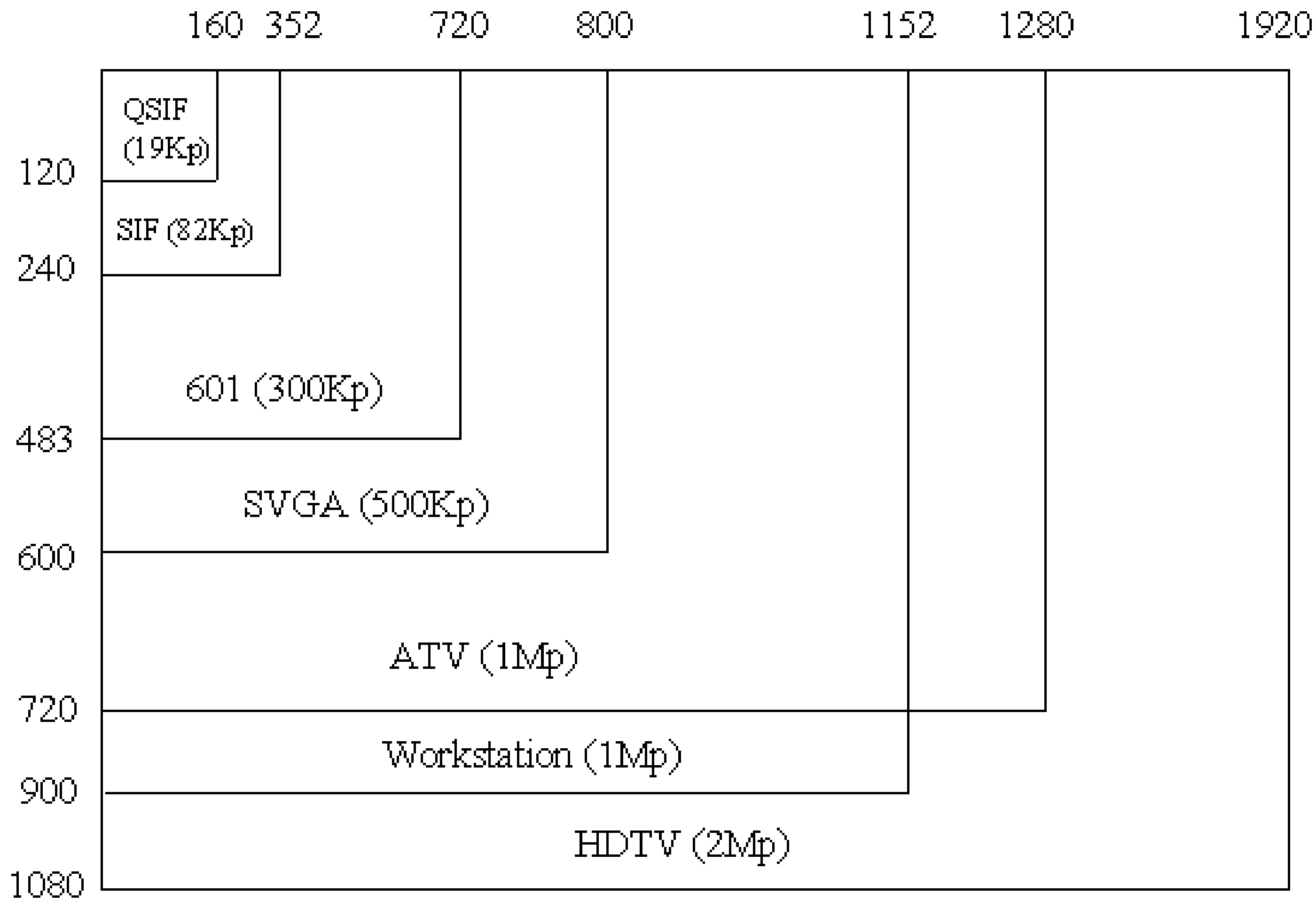
Digital video “sizes”:

- “raw” digital video:
  - 20 Mbyte/sec
  - 70 Gbyte/hr
  - 1 Gbyte = 50 seconds
- “low data rate” digital video
  - 200 kbit/sec – 5 Mbit/sec
  - 1 Gbyte = 30 min – 10 hr

Digital Video

- analog video is a continuous signal
- digital video uses discrete numeric values
- signal is sampled
- small, discrete regions are digitized
- frame is represented by pixel array

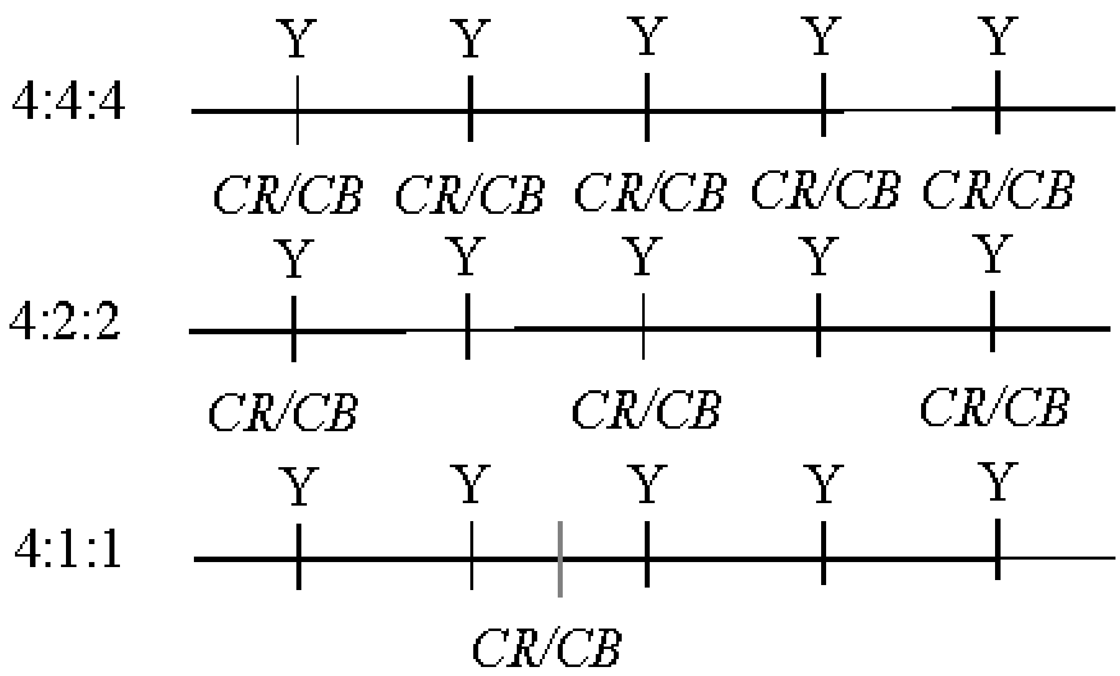
Pixel Arrays



CCIR 601

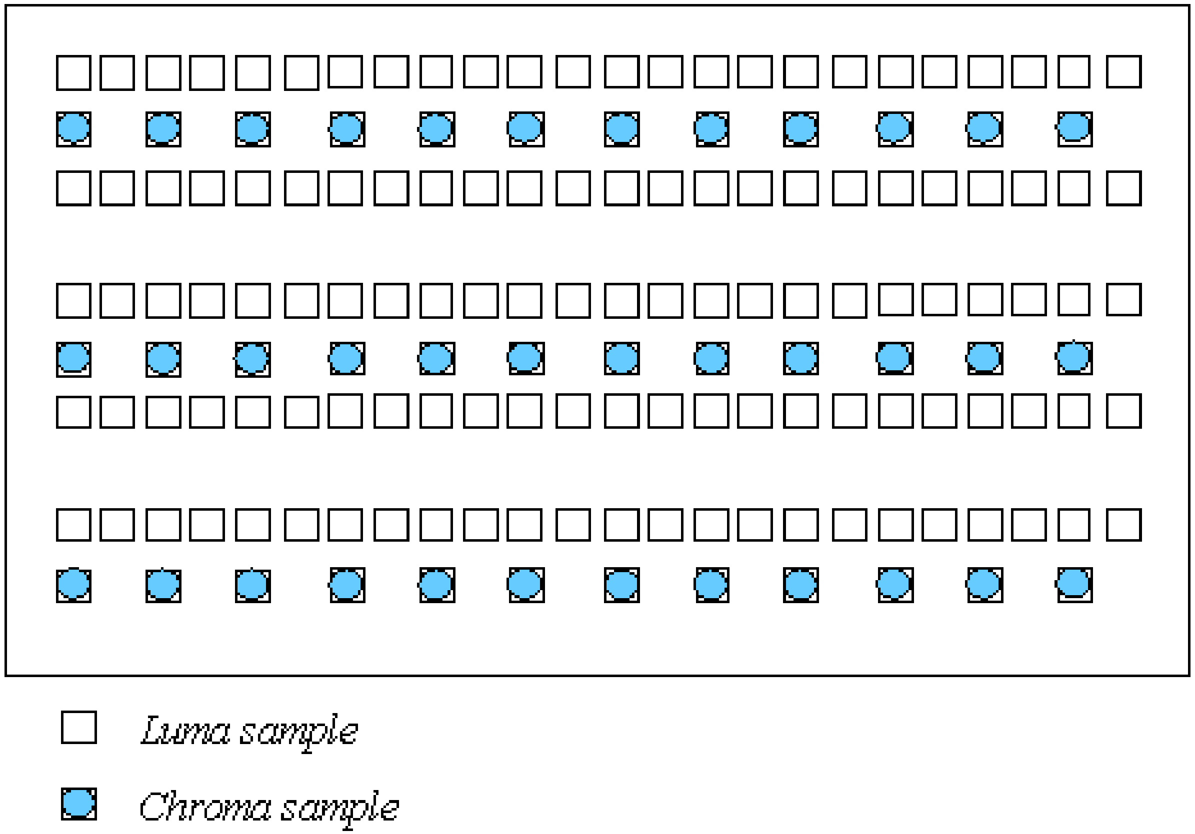
- digital component video - CCIR Recommendation 601
- format is obtained by sampling a component video signal
- base sampling rate 3.375 MHz
- particular member of 601 m:n:l (multipliers of sampling base rate)
- multiplier values: 1, 2, 3 or 4

Line Sampling



- 4:2:2 - broadcast quality; 4:1:1 - VHS quality
- 4:2:0 - 2:1 subsampling in horizontal and vertical direction
- 4:1:1 and 4:2:2 mostly used in MPEG and JPEG

4:2:0 Sampling



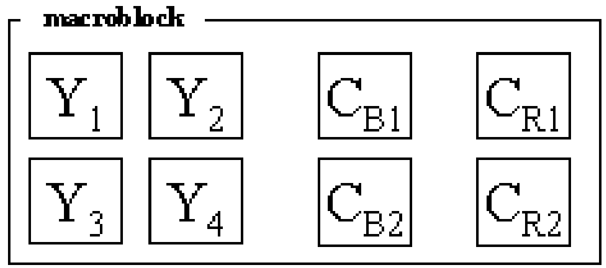
Video Block Structure

- 4:2:2 YC_RC_B

  16x16 macroblock

  8x8 pixel blocks

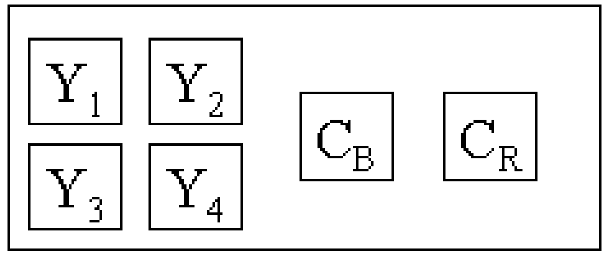
  8 bits/sample = 16 bits/pixel = 4Kbits/macroblock



- 4:1:1 YC_RC_B

  3Kbits/macroblock

  12 bits/pixel



Digital Video Representations

- Digital Composite Video (D2/D3, SMPTE 244M)
  - 14.3 MB/s data rate
  - subsampled color signals 4:2:2
- Digital Component Video (D1/D5, SMPTE RP125)
  - separate signals for luminance and color
  - 27 MB/s data rate
  - subsampled color signals 4:2:2

Video Data Rate

- D1 Digital
  - 720x483 = 347,760 pixels/frame
  - 4:2:2 sampling gives 695,520 bytes/frame
  - 21 MB/sec (168 Mbs)
- ATV
  - ATV
  - 4:2:0 sampling gives 1,382,400 bytes/frame
  - 41 MB/sec (328 Mbs)

Video Formats

- Digital Betacam - 720x480; 4:2:2; CCIR 601
- Sony DV Format - 500 lines; 4:1:1; DCT-based compression
- Panasonic/JVC DV Format (DVCPro) - similar to Sony DV but 4:2:2
- Sony Betacam SX - MPEG coding (IBBI patterns)

- DVI (Intel) PLV + RTV
- QuickTime (Apple)
- px64 / H.261 (CCITT)

Digital Video Formats

Video Format	Analog formats sampled	Sampling rate (Mhz)	Sample size	Appr. video data rate (Mbyte/sec)	Frame resolution
Digital Component (CCIR 601)	525/60 YUV 625/50 YUV	13.5	8/10	30.9, 20.6, 15.4	720 X 500 720 X 600
Digital composite	Compos. NTSC Compos. PAL	14.3 17.7	8	11.2 13.7	768 X 510 948 X 608
CIF QCIF	Various	Various	8	4.5 1.1	360 X 288 180 X 144
Digital HDTV	NA	NA	NA	ca. 125	ca. 1600 X 900

Compression

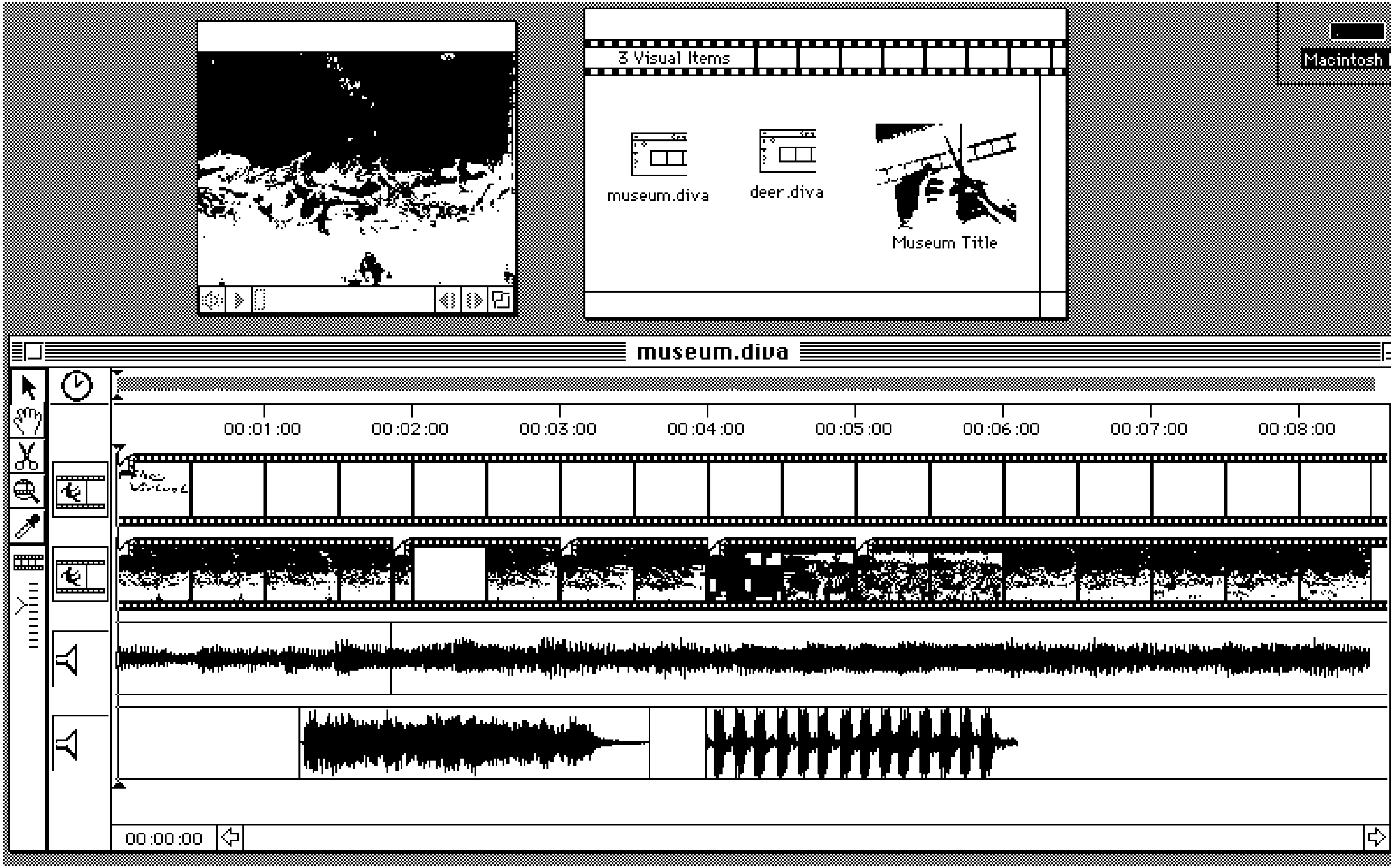
- lossy versus lossless
- real-time (symmetric)
- spatial versus temporal
- scalable
- type of source material

Factors Influencing Video Quality

- frame size and depth
- frame rate, key frame rate
- source material $\Rightarrow$ algorithm
- algorithm parameters, decode time
- compressed data rate
- compressed file size

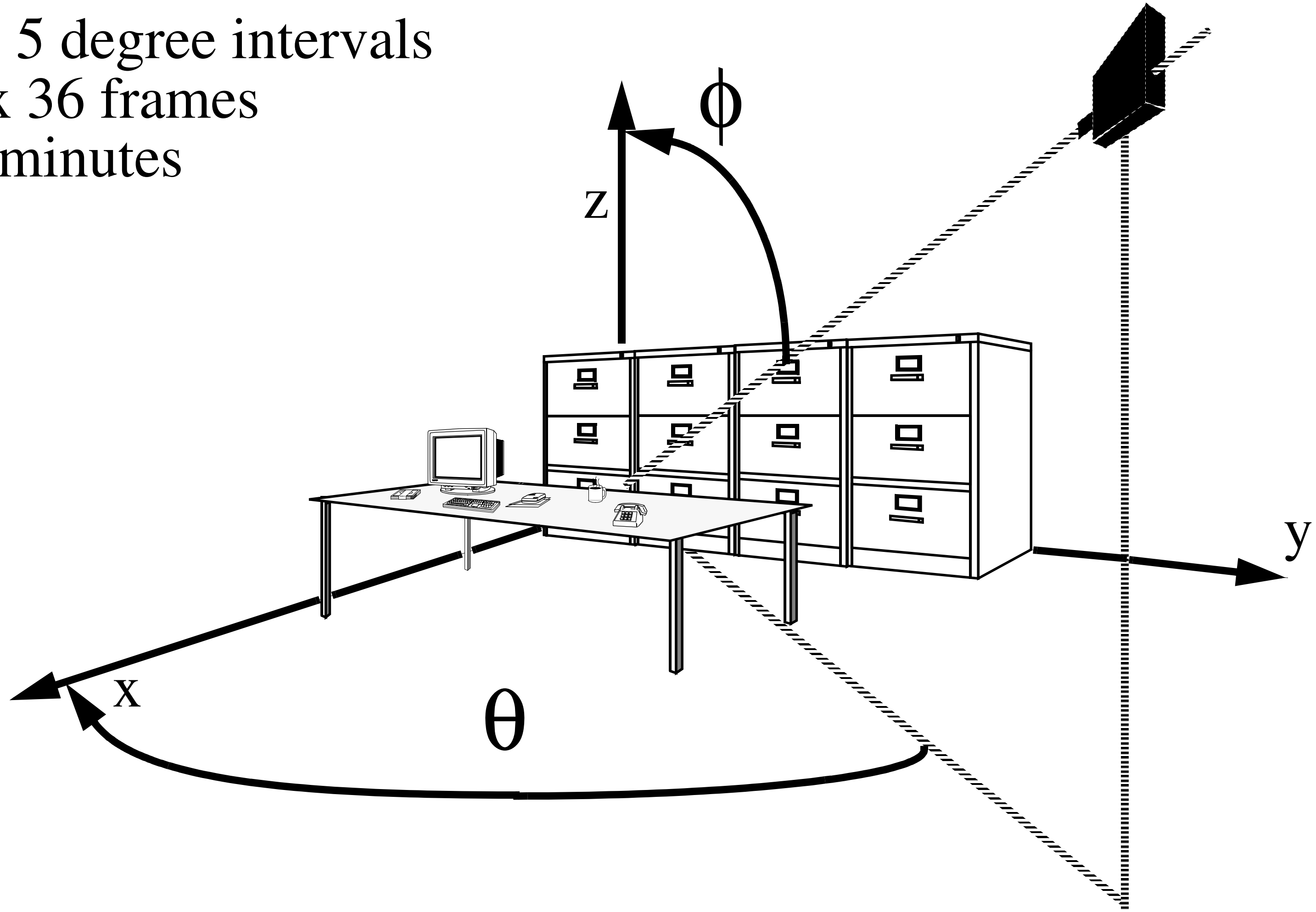
Operations on Digital Video

- storage and retrieval (HD, CD-ROM, RAID, DVTR)
- editing ("non-linear")
- digital video effects (transitions, keying, compositing*..)
- conversion (e.g. NTSC $\Rightarrow$ digital $\Rightarrow$ PAL)
- compression, decompression

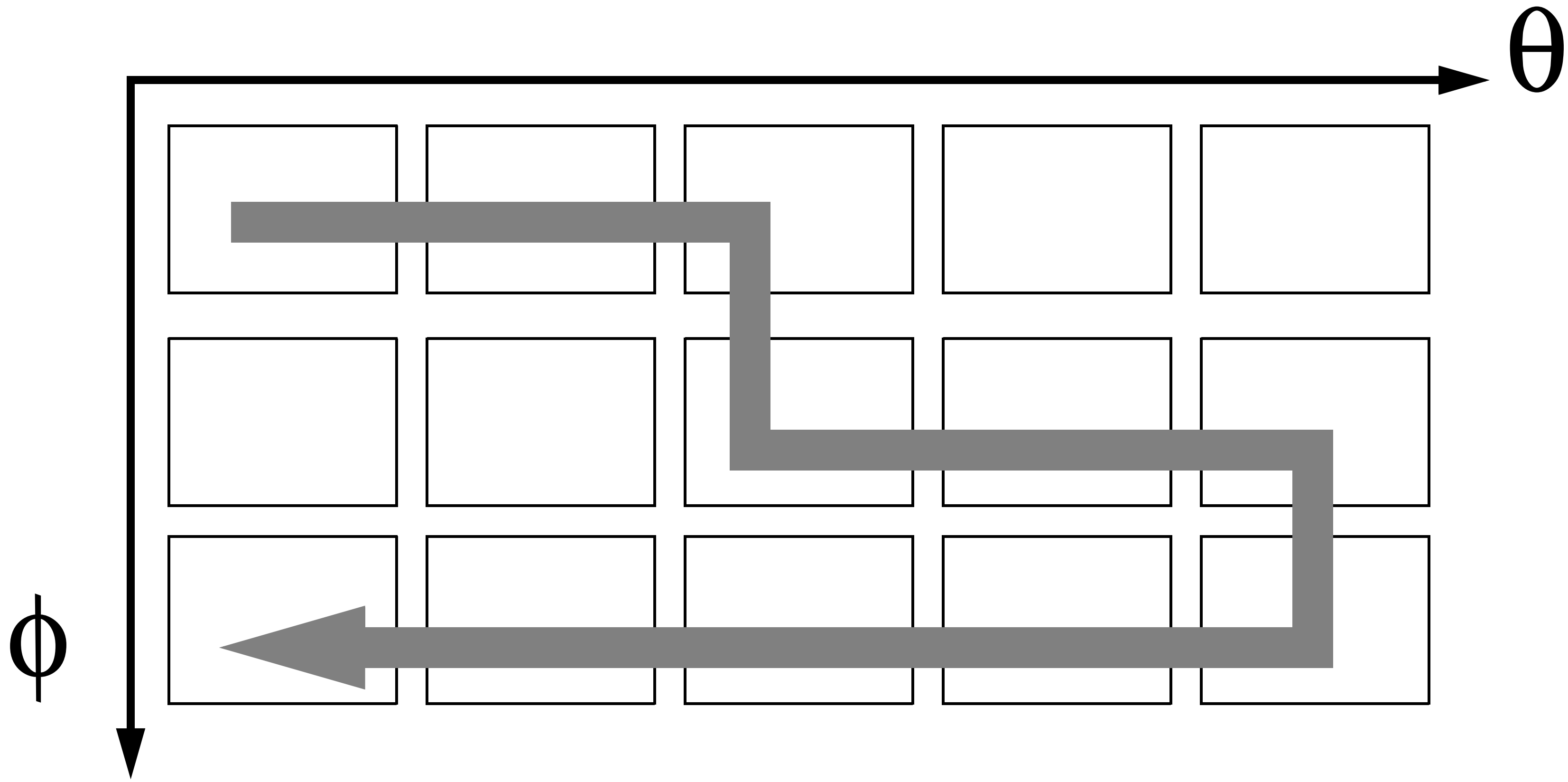

3 Visual Items
museum.diva
deer.diva
Museum Title
Macintosh
museum.diva
00:01:00
00:02:00
00:03:00
00:04:00
00:05:00
00:06:00
00:07:00
00:08:00
00:00:00

Example: “Navigable Video”

θ ϕ: 5 degree intervals
72 x 36 frames
1.5 minutes



Access Function



$$\theta,\phi \Rightarrow \mathrm{frame}(\theta,\phi)$$

I.3 Media Type Audio

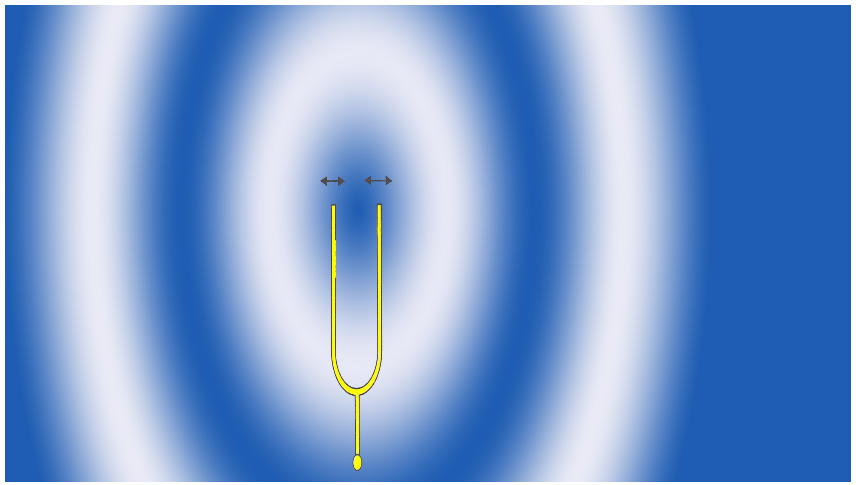
- Grundlagen der Akustik
- Digitalisierung
- Formate
- Operationen
- MIDI (Musical Instruments Digital Interface)

Grundlagen der Akustik

- Was ist Schall?
- Wie wird er mathematisch beschrieben?
- Eigenschaften (Amp., Frequenz) Auswirkung aufs Ohr?
- Ton, Klang, Geräusch
- Frequenzspektrum (diskrete und kontinuierliche Fourierzerlegung)
- Schalldruckpegel
- Schallpegeladdition
- Kurven gleicher Lautstärke
- Lautstärkepegel

Was ist Schall?

- Physikalische Definition: Schall ist die wellenförmige Ausbreitung von Druckschwankungen in elastischen Medien (d.h. Bindungskräfte zwischen Molekülen oder Atomen sind elastisch, wie z.B. in der Luft, Wasser,Knochen, Holz, Metall...). Dabei findet periodisches Komprimieren und Dekomprimieren des Mediums statt.



Stimmgabel überträgt mechanische Schwingung auf angrenzende Luftmoleküle:

Kompression↔**Dekompression**

Druckschwankungen werden an benachbarte Moleküle weiterge-geben ⇒ Schallwelle

Wie wird er mathematisch beschrieben?1

- Da Schall auf Schwingungen beruht, lassen sich Schallereignisse durch ihren zeitlichen Schwingungsverlauf beschreiben. Einfachste Schwingung: Sinusschwingung (alle anderen Schwingungen lassen sich ableiten):
- **Frequenz f:** Anzahl der Schwingungen pro Sekunde
  Einheit : Hertz [Hz]
- **Periodendauer τ**: Dauer einer Schwingung $\tau = 1/f$
  Einheit: Sekunde [s]
- **Amplitude p_0:** Druckwert bei maximaler Kompression
  Einheit: Pascal... [Pa]

Wie wird er mathematisch beschrieben? 2

- **Wellenlänge λ:** der während einer Periodendauer zurückgelegte Weg. Einheit:..[m]

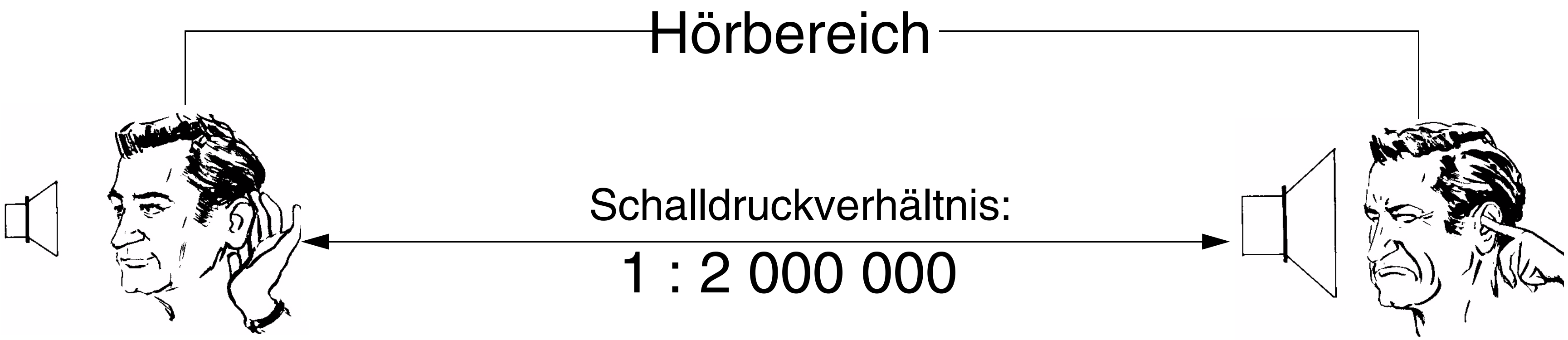
$$\lambda = v x \tau = v / f$$

- **Schallgeschwindigkeit** v: ist abhängig vom Ausbreitungs-medium: Einheit: [m/s]

Medium	v[m/s]	λ [m] 16Hz	λ [m] 100Hz	λ [m] 4kHz	λ [m] 20kHz
Luft (0^o)	331,0	20,68	3,31	0.08	0,0166
Luft (20^o)	343,6	21,48	3,44	0,09	0,0172
Wasser (20^o)	1484	92,75	14,84	0,37	0,0742
Ziegel	3650	228,13	36,50	0,91	0,1825

Eigenschaften und deren Wirkung 1

- **Frequenz:** ein Maß für die Tonhöhe. Je höher Frequenz, desto höher der Ton. Musikalisch entspricht eine Verdopplung der Frequenz einer Tonerhöhung von einer Oktav.
  hörbare Frequenzen: ca. 16Hz - 20000Hz
- **Amplitude der Druckwelle:** ein Maß für die Lautstärke.



Hörschwelle: $p_0=10^{-5}$ [Pa]
Ton gerade noch hörbar

Schmerzgrenze:p=20[Pa]
physiologischer Schmerz

Eigenschaften und deren Wirkung 2

Wellenlänge: Ist nicht direkt hörbar, aber spielt bei folgenden Hörphänomenen eine wichtige Rolle:

- Beugung: Wellen weichen von ihrem geradlinigen Weg ab und "lugen" um die Ecke. Verstärkt bemerkbar, wenn Wellenlänge größer als Hindernis.("um die Ecke hören")
- Interferenz: Überlagerung der direkten Schallwelle und indirekten Wellen, die durch Beugung und Reflexion entstanden sind. Interferenzeffekte, die im Bereich unseres Kopfes stattfinden, lassen uns unterscheiden, ob Schall von vorne oder hinten, oben oder unten kommt.

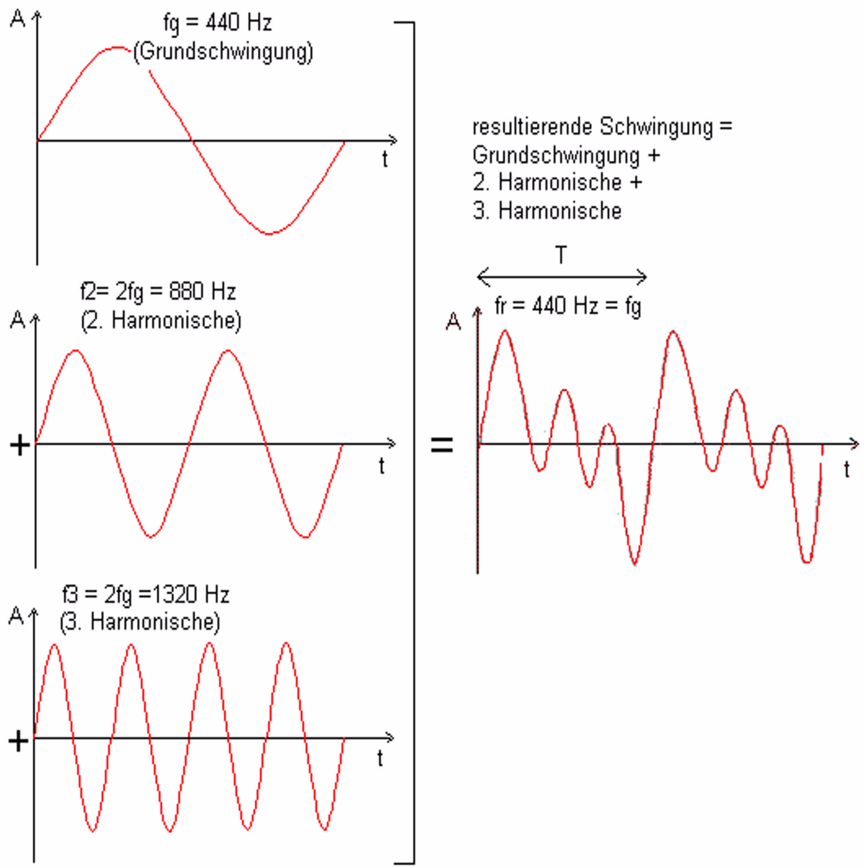
Eigenschaften und deren Wirkung 3

- Brechung: Schallwellen ändern ihre Richtung, wenn sie durch Gebiete mit unterschiedlicher Schallgeschwindigkeit laufen.
- Dispersion: Für verschiedene Tonhöhen ist die Stärke der Brechung verschieden. Dieser Effekt heißt Dispersion.

Ton, Klang, Geräusch

- **Ton:**

Eine einzige Sinusschwingung wird als Ton bezeichnet. p(t) ist sinusfunktion

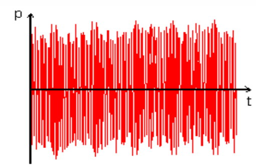


- **Klang:**

Klang entsteht aus einer Überlagerung von Grundton(hören wir alsTonhöhe) und Obertönen (f=n*fg) (hören wir als Klangfarbe). p(t)-Funktion allgemeine periodische Funktion mit Frequenz fg.

- **Geräusch:**

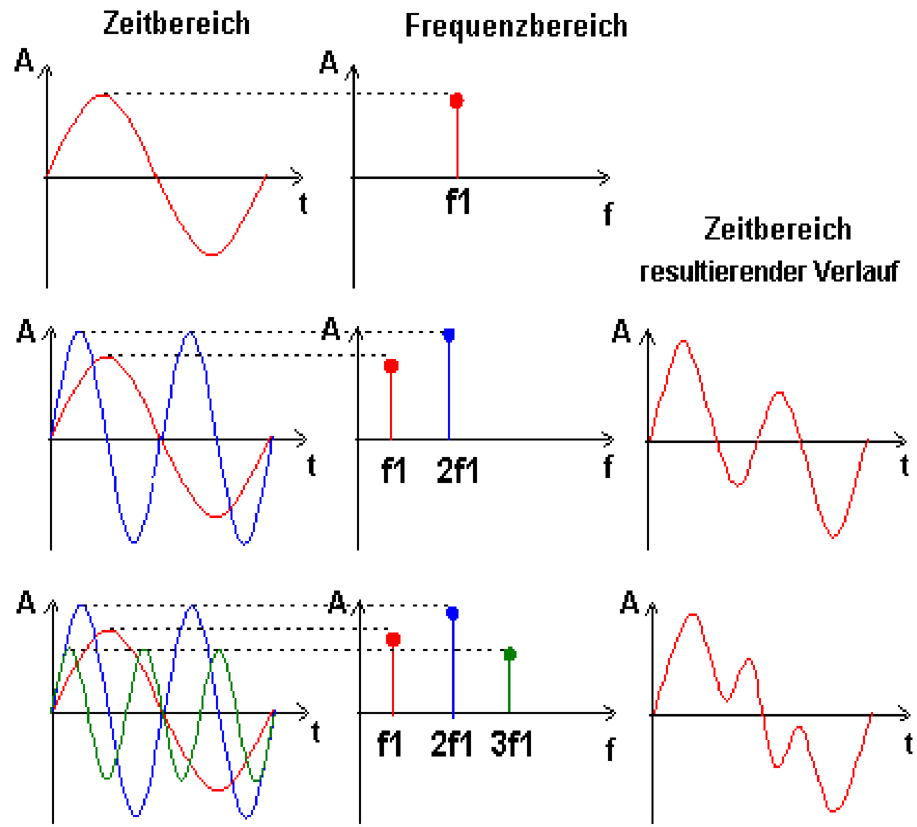
Kein Grundton mehr erkennbar, aperiodische p(t) Funktion



Frequenzspektrum 1/2 [1,6]

- Frequenzspektrum:

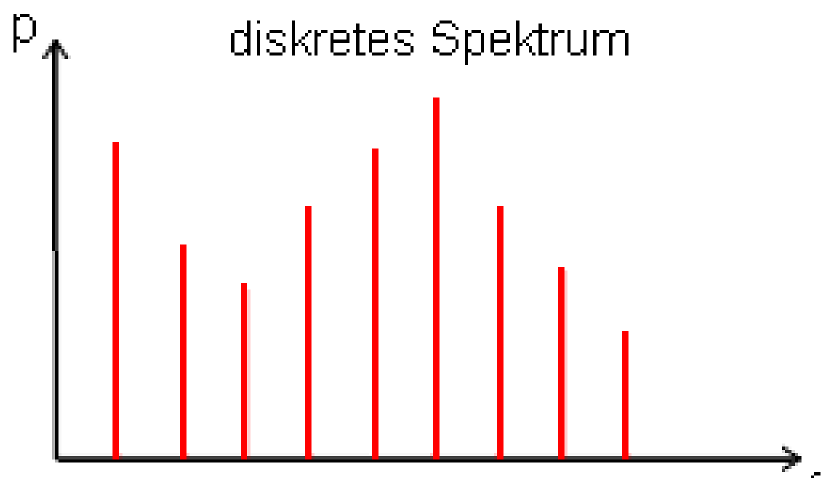
Jede periodische Schwingung beliebiger Wellenform (= Klang) kann als eine Überlagerung von Grundschwingung und Oberschwingungen dargestellt werden. Trägt man in einem Diagramm die Amplituden aller beteiligten Schwingungen in Abhängigkeit der Frequenz auf, so erhält man das sogenannte Frequenzspektrum



Frequenzspektrum 2/2

- Diskretes Spektrum:

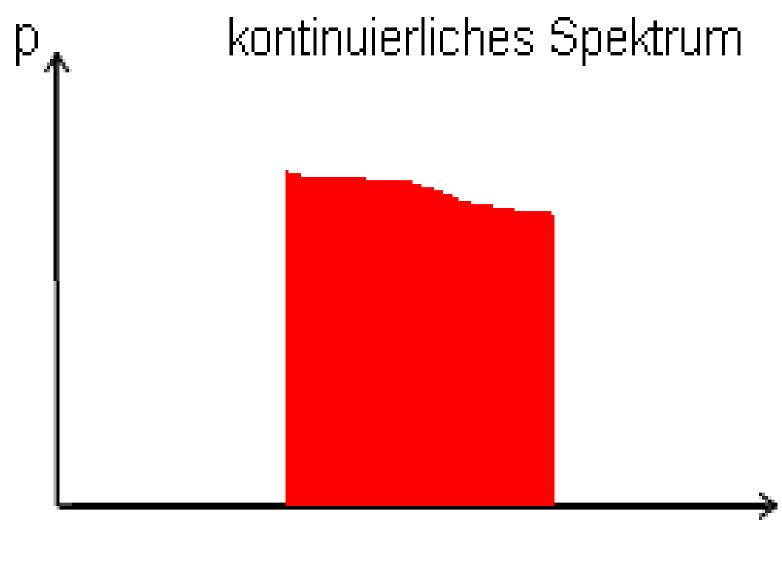
Klänge enthalten nur Schwingungen, deren Frequenzen ein ganzzahliges Verhältnis zur Grundfrequenz aufweisen.



- Kontinuierliches Spektrum:

Allgemeine Schallereignisse (Geräusche) enthalten eine unendliche Anzahl von Einzelschwingungen, deren Frequenzwerte sich kontinuierlich über die x-Achse erstrecken. Es entsteht eine kontinuierliche mathematischeFunktion:

p=p(f)



Schalldruckpegel

- **Definition des Schalldruckpegels:**

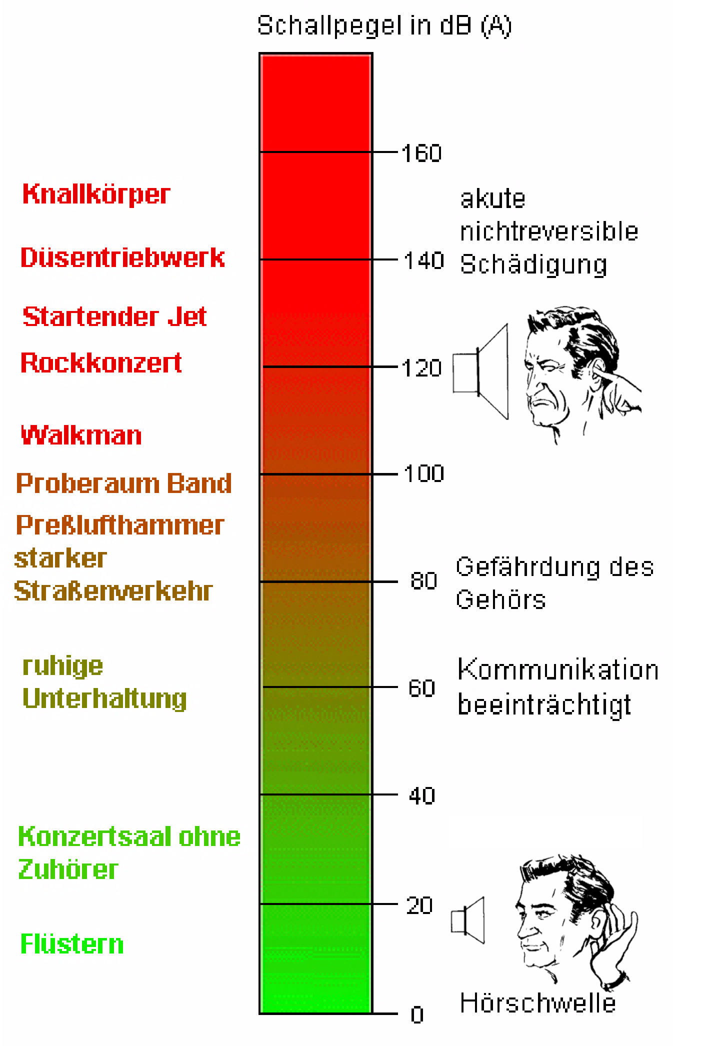
$L=20*\log(p/p_0)$ [dB]

L...Schalldruckpegel

p...Schalldruck der betrachteten Schallwelle

p_0...Bezugsschalldruck: definitionsgemäß der Schalldruck eines 1000Hz Tones, der gerade noch hörbar ist. $p_0=2*10^{-5}$[Pa]

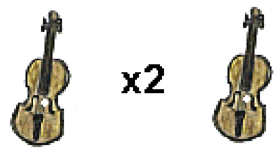
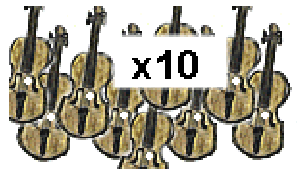
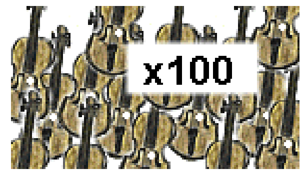
dB...Dezibel, Einheit des Schallpegels



Wozu Schalldruckpegel?

- Damit man den gigantischen Hörbereich (Hörschwelle / Schmerzgrenze 1 : 2 000 000!!) sinnvoll beschreiben und darstellen kann, Abbildung in einen kleineren Wertebereich (Hörschwelle=0dB,Schmerzgrenze= 120dB)
- Experimente zeigen, daß unser Empfinden von Lautstärkedifferenzen einem logarithmischen Gesetz folgen
- der logarithmische Schalldruckpegelwert wird daher unserem Lautstärkeempfinden mehr gerecht als die Angabe eines absoluten Druckamplitudenwertes(*siehe Schallpegeladdition)*.

Schallpegeladdition

Anzahl Schallquellen				
Schalldruck-änderung	x1	x1,4	x3	x10
Schallpegel-differrenz	0dB	+3dB	+10dB	+20dB
Subjektives Hörempfind-en	Grund-lautstärke	etwas lauter	doppelt so laut wie eine einzige Geige	viermal so laut wie eine einzige Geige

Schallpegeladdition

Addition von Schallpegeln gleicher Intensität

- Überlagern sich die Schallintensitäten von n Schallpegeln mit gleicher Schallintensität, so gilt für die am Immissionsort gemessene Schallintensität

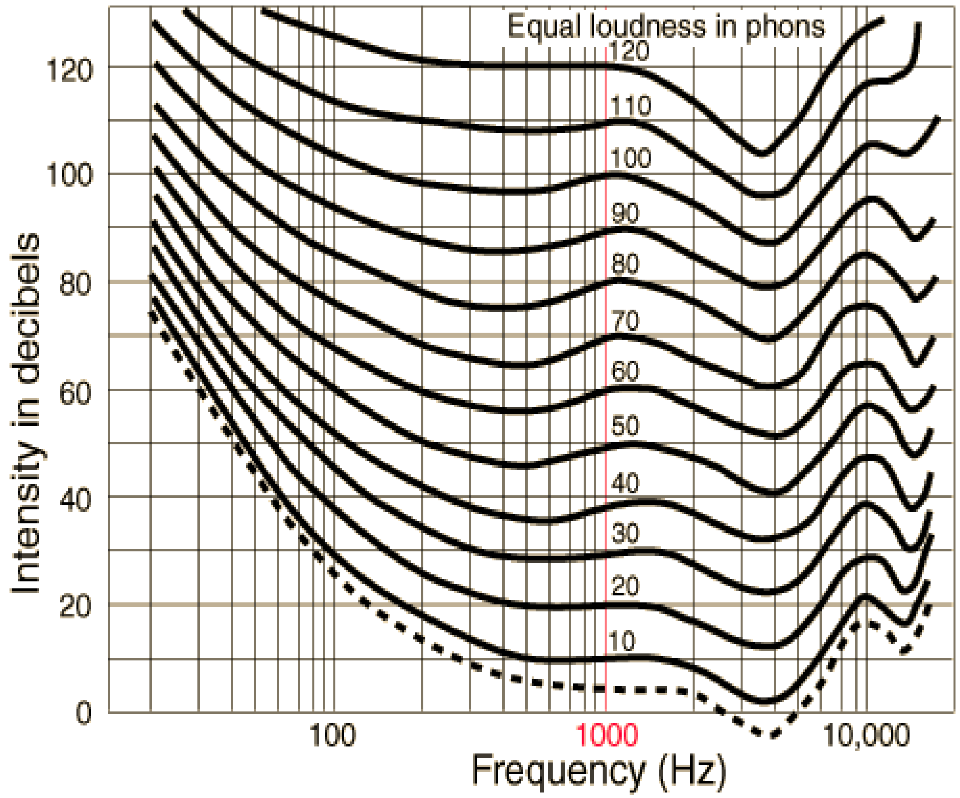
$$I_{ges} = I_1 + I_2 + I_3 + ... + I_n = n*I_1$$

- Für den Summenpegel gilt:

$$L_{ges} = 10*\log(I_{ges}/I_0) = 10*\log(I_1/I_0) + 10*\log n$$

Kurven gleicher Lautstärken

- Töne mit verschiedenen Frequenzen, aber selbem **L** beurteilt unser Ohr verschieden laut. Die Kurven gleicher Lautstärke geben an, wie hoch im Vergleich zum Pegel eines 1000Hz-Vergleichtons der Schallpegel eines Tones variabler Frequenz sein muß, damit dieser Ton als gleichlaut empfunden wird.



Lautstärkepegel

- Damit alle Töne, die für unser Ohr gleich laut klingen, auch mit einem gleichen Pegelwert beschrieben werden, wurde der **Lautstärkepegel L_N** definiert: Einheit: [Phon]

  L_N eines Tones mit der Frequenz f ist gegeben durch den Schalldruckpegel L eines Tones mit 1kHz, der als gleich laut gehört wird.

- Den Lautstärkepegel eines beliebigen Tones kann man leicht aus den Kurven gleicher Lautstärke herauslesen:

- Bsp: Ton mit 100Hz und Schalldruckpegel L=60dB: 50*Phon*

Einführung Digitales Audio

Mikrofon wandelt Schallwelle in elektrisches Signal: Spannungsverlauf analog dem Druckschwankungsverlauf an der Membran des Mikrofons.

- Analoge Audiotechnik:
  analoges Signal gespeichert auf Magnetband, Schallplatte...
  Editieren: mechanisches Schneiden, überspielen
- Digitale Audiotechnik:
  analoges Signal wird mittels A/D Konverter in binäre Darstellung umgewandelt, gespeichert, am Computer editiert; mittels D/A Konverter wieder als Analogsignal am Lautsprecher ausgegeben.

Vorteile / Nachteile DA

Vorteile *(digital)*:

- weniger Rauschen
- höhere Dynamik
- Verlustloses Kopieren
- Höhere Linearität: Frequenzgang, Übertragungskurve (Klirrfaktor)
- Temperatur unempfindlich
- keine Gleichlaufschwankungen

Nachteile *(digital)*:

- anfällig für Datenverlust
- höhere Übertragungsbandbreiten
- aufwendigere Hardware

Digitization in General

- microphones, video cameras produce analog signals (continuous-valued voltages)
- to get audio or video into a computer, it must be converted into a stream of numbers; *discrete sampling* (both time and voltage)
- sampling—divide the horizontal axis (the time dimension) into discrete pieces; uniform sampling is ubiquitous
- quantization—divide the vertical axis (signal amplitude) into pieces; sometimes, a non-linear function is applied

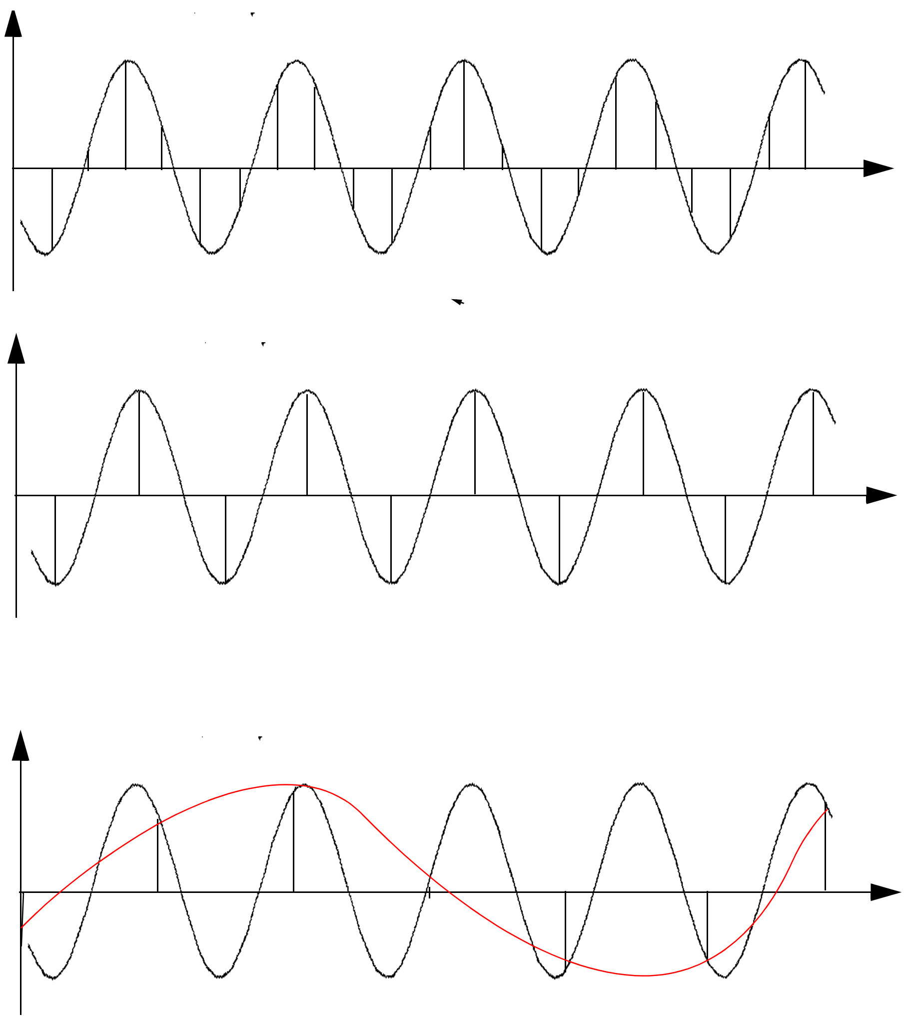
Digitization in General

questions for producing digital audio (Analog-to-Digital Conversion):

- how often do you need to sample the signal?
- how good is the signal?
- how is audio data formatted?

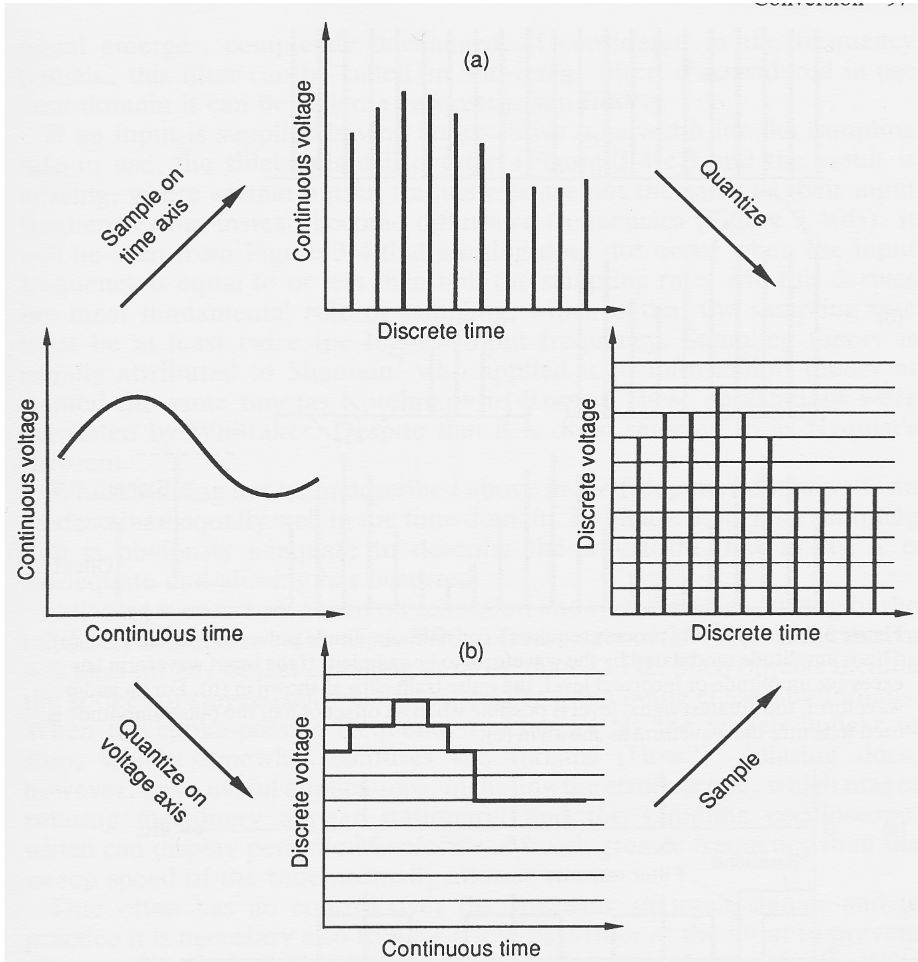
- suppose we are sampling a sine wave. How often do we need to sample it to figure out its frequency?
- if we sample at 1 time per cycle, we can think it's a constant
- if we sample at 1.5 times per cycle, we can think it's a lower frequency sine wave --> alias

Aliaskomponente



Grundlagen Digitalisierung

- Eingang: zeit- und spannungskontinuierliche Wellenforml
- Abtasten- Quantisieren: sind voneinander unabhängig → zeitliche Reihenfolge egal
- Ergebnis: zeit- und spannungsdiskretes Format



(a) Eingangssignal wird abgetastet, Samples werden quantisiert (üblich bei Audio)

(b) Eingangssignal wird quantisiert, dann erst abgetastet (üblich bei Video)

Abtasten

- Prozedur: in periodischen Zeitintervalen wird dem analogen Eingangssignal eine Probe (=sample) entnommen.
- Wie oft muß ein analoges Signal abgetastet werden?

Nyquist-Shannon Theorem:
Ein abgetastetes Signal läßt sich nur dann ohne Informationsverlust rekonstruieren, wenn

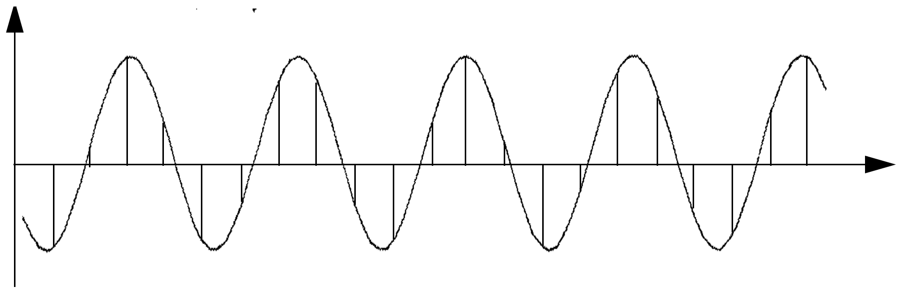
$$f_s \geq 2 \; x \; f_{max}$$

f_s .....Abtastfrequenz

f_{max}...die höchste im Signal vorkommende Frequenz

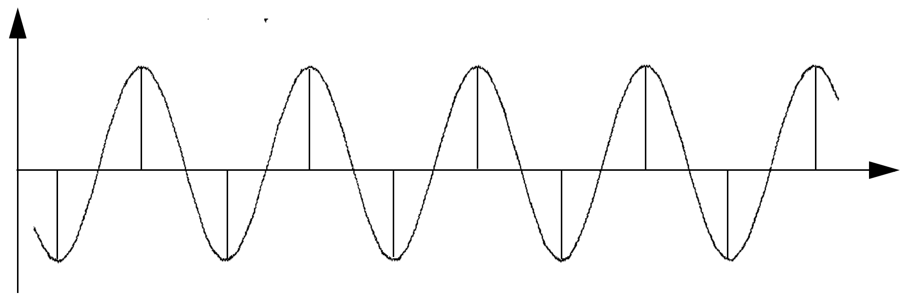
Nyquist Rate, Nyquist Frequenz

Aliaskomponente 1



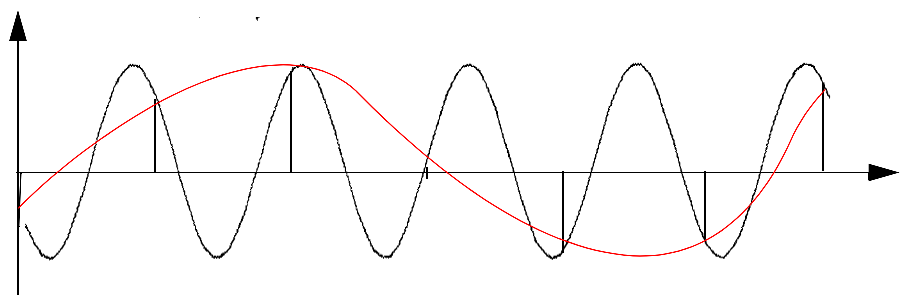
- $f_s > 2\ x\ f_{max}$

  Rekonstruiertes Signal mit ursprünglichem ident



- $f_s = 2\ x\ f_{max}$

  Rekonstruiertes Signal gleiches f wie ursprünglich



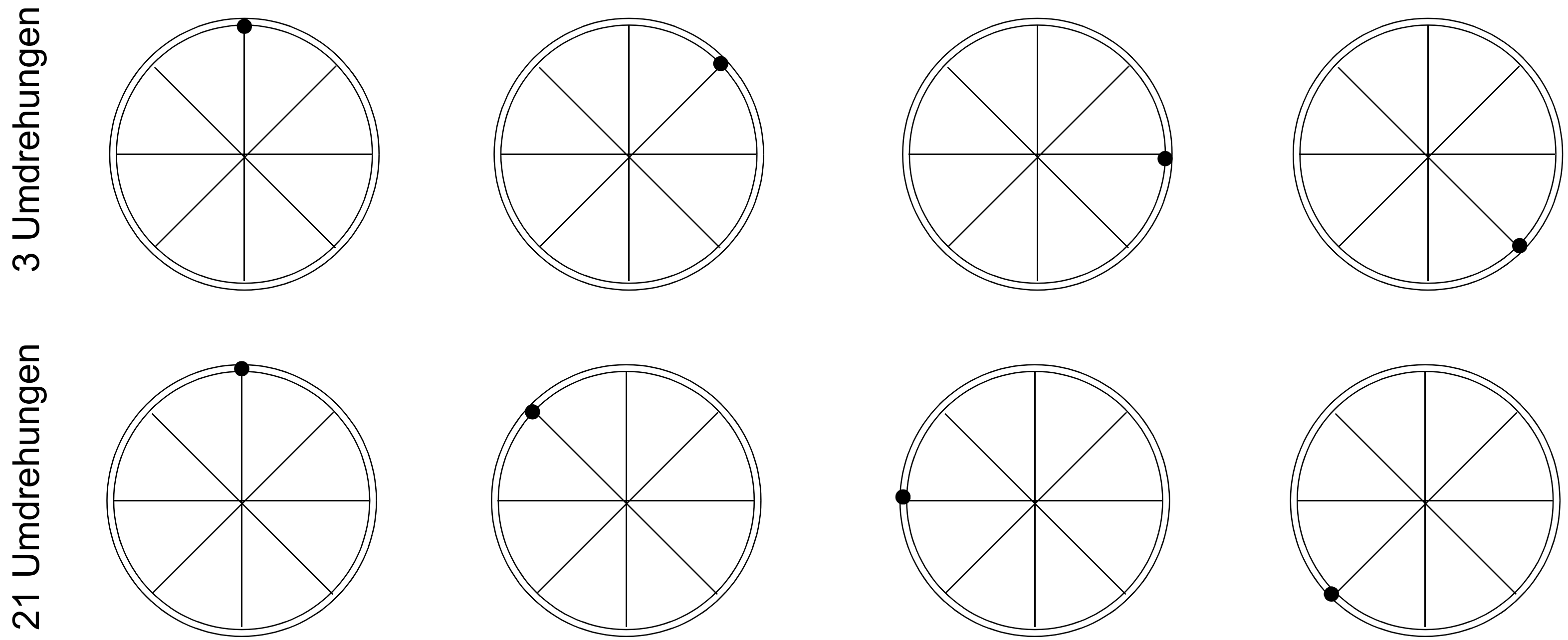
- $f_s < 2\ x\ f_{max}$

  Störfrequenz "Aliaskomponente"

Bsp.: $f_s = 48kHz$, $f_{ein} = 30kHz \rightarrow f_{alias} = 18kHz$

Aliaskomponente 2

- Bsp: drehendes Rad im Film
  - Sampling Frequenz ist dabei fix (FS = 24)
  - Anzahl der Umdrehungen (u) des Rades variabel



Beispiele

Abtastfrequenz und max. Frequenz

Format	Abtastfrequenz	Frequenzbereich
Telefon	8 kHz	200-3400 Hz
Audio- CD	44,1 kHz	20-20000 Hz
DAT, prof. Audio	48 kHz	20-20000 Hz
Sat- Radio	32 kHz	20-15000 Hz

Quantisierung

- Abgetasteten Spannungswerten werden diskrete Zahlenwerte zugeordnet: Gesamtspannungsbereich in Quantisierungsintervale Q unterteilt $\rightarrow$ kontinuierlicher Wert nächstgelegenem Zahlenwert zugeordnet.
- Qualität: Q soll möglichst klein sein, ergibt sich aus der Länge des Datenworts:

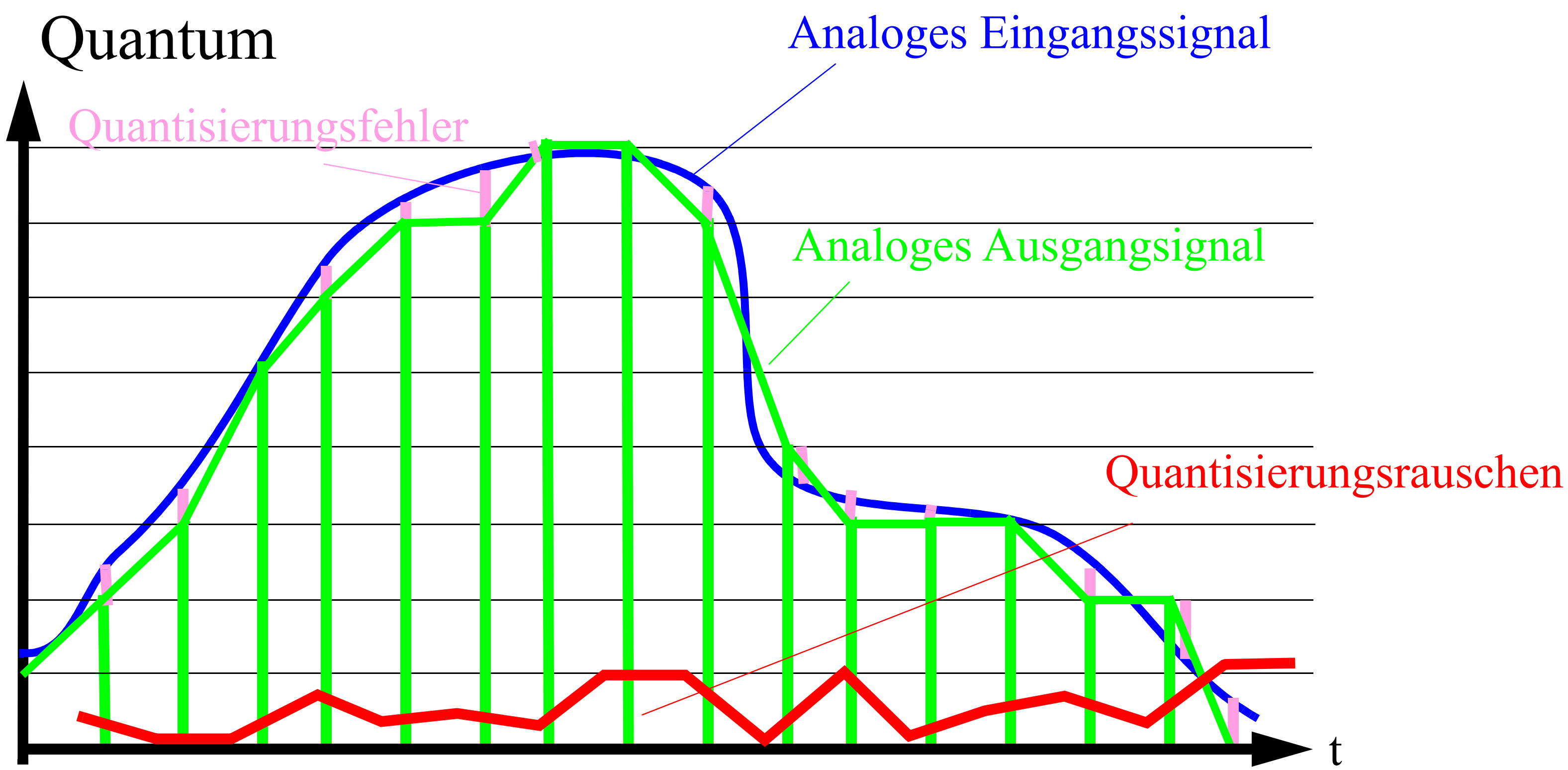
  8 bit *(256 Intervale)*: $Q = V_{pp} / 2^8 = (4x10^{-3}) \; x \; V_{pp}$

  16 bit *(65536 Intervale)*: $Q = V_{pp} / 2^{16} = (15x10^{-6}) \; x \; V_{pp}$

mit $V_{pp} = 2 \; x \; abs(V_{max})$ V_{pp} *... peak to peak Bereich*

Quantisierungsfehler

- Zahlenwert weicht vom abgetasteten Wert um $\leq 0{,}5\ Q$ ab. Bei Rekonstruktion Fehler als "Quantisierungsrauschen" hörbar.



Dynamikbereich

- Schallwiedergabe:

  Quantisierungsrauschen erzeugt Schalldruckpegel $\mathbf{L_{noise}}$.

  Maximales Nutzsignal Schalldruckpegel $\mathbf{L_{max}}$.

- Dynamikbereich:

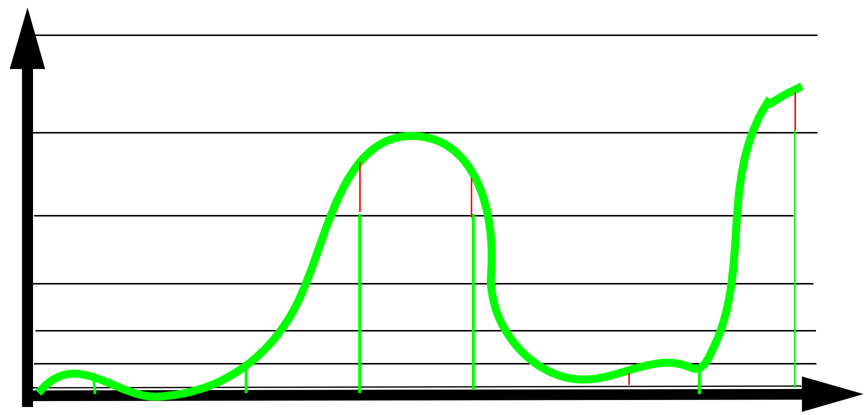
  $\mathbf{SNR = L_{max} - L_{noise}}$

- **SNR**..Signal to noise ratio [dB]

Auflösung	8 bit	12 bit	14 bit	16 bit
Dynamik	49,8 dB	73,7 dB	85,7 dB	97,6 dB

Lineare, nicht lineare Quantisierung

- Lineare Quantisierung: Q konstant; übliche Methode in Audiotechnik
- Nichtlineare Quantisierung: Q *von unterschiedlicher Größe.*
  kleine Werte $\rightarrow$ kleiner Quantisierungsfehler
  große Werte $\rightarrow$ großer Quantisierungsfehler
- Rauschen von großem Nutzsignal maskiert; $\Rightarrow$ auch mit kleiner Auflösung passable Qualität
  Verwendung: für Systeme mit geringer Übertragungsbandbreite

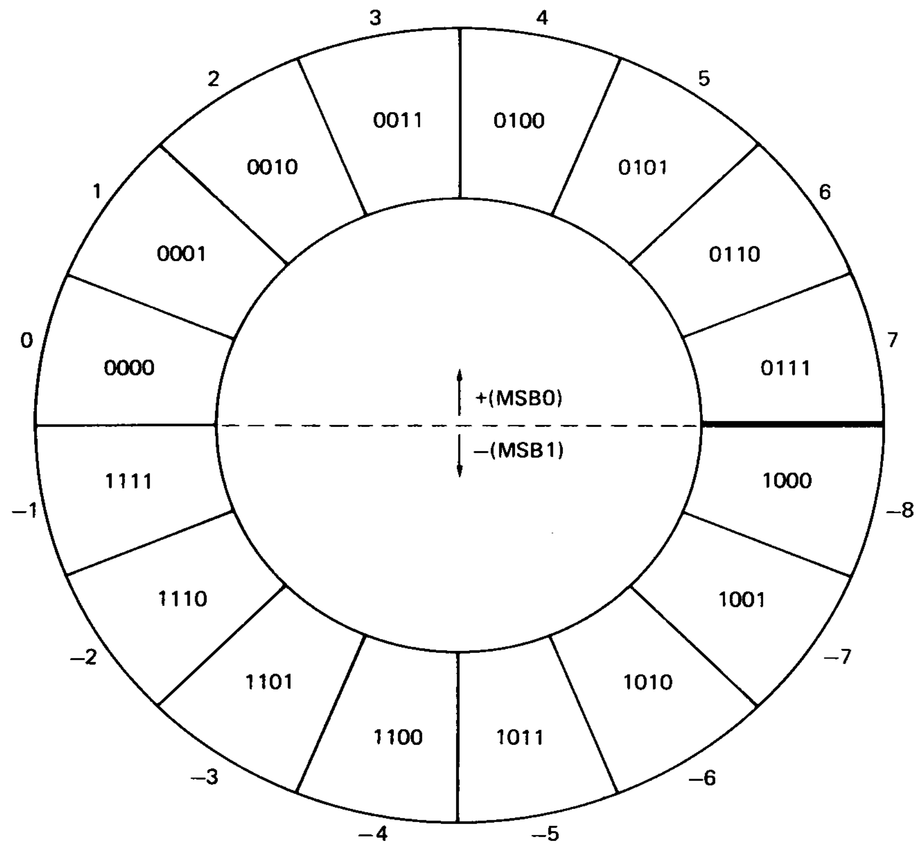


Codierung

Signale von Audio bipolar (pos. und neg.) ⇒

Zweierkomplementcode:

- MSB=1...neg. Wert
  MSB=0...pos. Wert
- bei Addition zweier Signale kein Offset (=Nullpunktverschiebung)



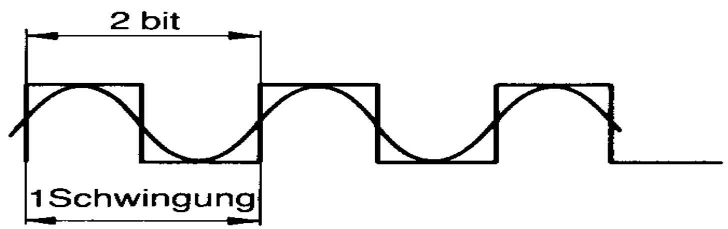
Bsp.: Zweierkomplementcode für 4-bit Daten

Pulscodemodulation

- Prozedur: Binärwerte werden seriell in Form von modulierten Spannungspulsen über eine einzige Leitung übertragen.
- Datenrate:

  Abtastfrequenz[Hz] x Bitauflösung[bit] = Datenrate[bit/s]

  Bsp.: CD-Audio(mono): 44,1kHz;16 bit ⇒ 705600 bit/s
- Bandbreite: nötiger Frequenzbereich, um PCM verlustlos zu übertragen:



  Bandbreite[Hz] = Datenrate[bit/s] : 2

  Bsp.: CD-Audio(mono) ⇒ 352,8 kHz (vgl.: analog 20 kHz !!)

Characteristics of Digital Audio

- sampling frequency (rate)
- sample size / quantization
- number of channels (tracks)
- interleaving
- sample representation (negative values)
- coding/compression

Example Audio Formats

	DAT	CD audio	CD-I (Level B)	A-law[a] μ-law
sampling rate (kHz)	48[b]	44.1	37.8	8
sample size (bits)	16	16	4	8
quantization	uniform	uniform	uniform	log
no. of channels	2	2	1-8	1
data rate per channel (10^3 bit/sec)	768	705	174	64
encoding	PCM	PCM	ADPCM	PCM
quality	very high	very high	“Mid Fi”, “FM”	“telephone”

a. A-law is used within European telephone systems, while μ-law is North American.
b. DAT has a number of audio formats – three sampling frequencies are possible (32 kHz, 44.1 kHz and 48 kHz) and the 32 kHz rate may use either 16 bit linear quantization or 12 bit nonlinear, in the second case two or four tracks may be recorded. The numbers listed are for the highest quality DAT format.

Operations on Digital Audio

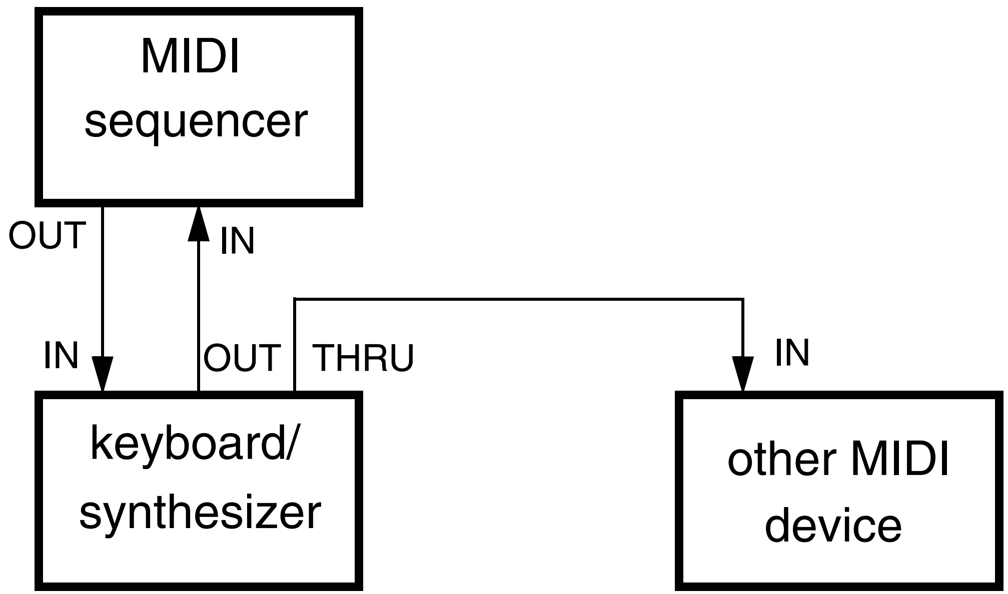
- storage and retrieval (CD, DAT)
- editing (non-linear, non-destructive, play-lists)
- digital audio effects (delay, equalization, noise reduction, time compression and expansion, pitch shifting ...)
- conversion

Audio Processing—Examples

- reshape impulse response to simulate a different room
- move perceived location from which sound comes
- locate speaker in 3D space using microphone arrays
- cover missing samples
- mix multiple signals (i.e. conference)
- echo cancellation

Musical Instruments Digital Interface

example MIDI system configuration



MIDI Terminology

- MIDI—a protocol that enables computer, synthesizers, keyboards, and other musical device to communicate with each other
- synthesizer—a sound generator (various pitch, loudness, tone color); often has a microprocessor, keyboard, control panels, memory, etc.
- sequencer—a stand-alone unit or a software program; used to be a storage for MIDI data; nowadays more a software music editor; has one or more MIDI INs and MIDI OUTs.

MIDI Terminology

- track—used to organize the recordings; can be turned on or off on recording or playing back.
- channel—used to separate information in a MIDI system; there are 16 MIDI channels in one cable; channel numbers are coded into each MIDI message.
- voice—the portion of the synthesizer that produces sound; synthesizers can have many (16, 20, 24, 32, 64, etc.) voices; each voice works independently and simultaneously to produce sounds of different timbre and pitch.

MIDI Terminology

- key number — notes are identified by 128 key numbers
- controller — controller values specify the operational characteristics of a MIDI device
- patch / program — the control settings that define a particular timbre

Important MIDI Concepts

- timing clocks — MIDI sequencer time stamps messages, timebase measured in *parts per quarter note,* tempo in *beats per minute*
- MIDI synchronization — external or internal sync
- MIDI Time Code (MTC) — used to synchronize with film or video