Issues & Answers in Digitization

CENDI/FLICC Workshop 5

Moving Image Archiving:

Digitizing Present & Future

Library of Congress

Packard Campus for Audio Visual Conservation

National Audio Visual Conservation Center

Culpeper, Virginia

James Snyder

Senior Systems Administrator, NAVCC

Presenter

- Senior Systems Administrator – Library of Congress NAVCC
- Responsible for:
  - Library of Congress – National Audio-Visual Conservation Center in Culpeper, VA
    - Preserving all forms of recording from the dawn of recording to today's audio, moving image and electronic gaming & learning
    - Converting entire physical archive to digital for access and preservation
- Previous projects:
  - Sarnoff Research Center
  - Public Broadcasting Service (PBS)
  - ABC radio & television
  - Fox News Channel – first all-digital facility launch & shakedown
  - Intelsat
  - MCI/*Verizon*Business Digital Media Technologies Group
  - Design Engineer, LOC NAVCC Culpeper, VA (2006-9)
- Member: AES, IEEE, NATAS, SMPTE & WEBE
- Frequency coordinator – National Capital area

Today's Course

- This workshop was adapted from these seminars:
  - Advanced Television Test Center (1988-1996)
    - Original digital television lab for the U.S. broadcast DTV standardization process
  - PBS Advanced Television Field Test Project (1993-96)
    - Original field tests of the U.S. digital TV standard
  - Harris/PBS DTV Express project (1998-99)
  - PBS DTV educational program (1998-2006)
    - Educated U.S. Public television personnel on media production and broadcast conversion to digital
  - SMPTE, WEBE & USTTI engineering lectures
  - Real world digital imaging implementation experiences
  - Portions of this presentation are used with the permission of the above organizations. The copyrights remain with the original owners.

Agenda

- Introduction
- Basics of analog moving images
  - Video
  - Film
  - And a little audio too!
- Digitization basic concepts
- Basics of born digital moving image content
  - Physical media
  - Files

Agenda

- Evaluation of physical media
  - Human inspection
  - Tape inspectors & cleaners
- Choosing an archival file format
  - Standards vs. vendor proprietary
  - Compressed or uncompressed
- IT Systems
  - Planning factors
  - Quality control
  - Bit-error rate
  - Business IT vs. archival IT

Agenda

- The end of the workflow: storage
- Long term archival storage planning:
- Regular migrations
- Verifying your archive
- Cryptographic checksums
- The future of storage

Agenda

- The end of the workflow: storage
- Long term archival storage planning:
  - Regular migrations
  - Verifying your archive
  - Cryptographic checksums
- The future of storage
  - aka….ummm…what's next….?

Questions?

Please wait for the end of the presentation!

Introduction

TV Studio Technologies

Analog - 1940s & '50s

- Black and white video systems –
  - UK 405 line: on-air from Alexandra Palace, London 1936-9, 1945-1991
  - US 525 line B+W ('NTSC 1'): July 1, 1941 on 6 channels
    - NTSC: National Television System Committee
  - France 819 line B+W: 1947
- Color video:
  - NTSC incompatible 441 line sequential color: 1951 ('NTSC 2')
  - NTSC compatible color standard on air 1954 ('RCA compatible color'; 'NTSC 3')
  - 'Composite' video distribution begins ~1956 (IE B+W & color signals on 1 cable)
- Film recording ONLY until 1956 (aka 'kinescope')
  - Continued widely until the early 1960s
  - Not eliminated completely until 1971 when 'Tonight Show' transferred to Burbank
- First videotape recorder (VTR) introduced April 1956
  - In production at CBS Television City in Hollywood 11/1956
  - Specially modified color quad VTRs at NBC Burbank in 1957 ('RCA mod')
  - Standardized quad VTR color in 1962 (Ampex VR-2000B)

TV Studio Technologies

Analog - 1960s

- PAL/SECAM analog color systems join NTSC on the air
  - PAL: Phase Alternating Line (1964)
  - SECAM: Sequential Colour avec Memoir (1968)
- First 'affordable' VTRs produced:
  - Sony 2" helical VTR (movies on airplanes)
  - Incompatible 1" VTRs introduced by Sony, Ampex, others
  - Incompatible ½" VTRs from Craig, Sony, Panasonic, JVC, others
    - One manufacturers VTRs could not play the others
    - VTRs of the SAME manufacturer could not always play on the same brand's VTRs!
- Color VTRs appear: 2" quad & helical scan VTRs
- 1969: ½" EIAJ B+W format appears in Japan
  - EIAJ: Electronic Industries of Japan
  - 1970: introduced to United States
  - Very commonly used industrial, educational, instructional, and documentary format
  - Color version follows in 1974: EIAJ-2
  - Cassette version – very rare

TV Studio Technologies

Analog - 1970s

- September 1971: Umatic VTR introduced
  - VCR starts to appear as a term: videocassette recorder
- 1972: Electronic news gathering (ENG) begins to appear at the political conventions
  - 1973: WCAU-TV Philadelphia first TV station to test ENG
  - Field acquisition made possible by portable 3/4" VTRs
- 1973: First digital TBC introduced
  - Enable use of unstable VCRs as news acquisition formats
- 1" professional VTRs standardized
  - Type B: mostly in Europe for PAL colour (Bosch Fernseh 1974)
    - About 50 units made for NTSC in U.S. for C-SPAN, KCPT-TV Kansas City
  - Type C: introduced worldwide by Ampex & Sony;
    - first sold to ABC News & Sports divisions;
    - first format able to do slow motion & noiseless still frame

TV Studio Technologies

Analog - 1970s

- September 1971: Umatic VTR introduced
  - VCR starts to appear as a term: videocassette recorder
- 1972: Electronic news gathering (ENG) begins to appear at the political conventions
  - 1973: WCAU-TV Philadelphia first TV station to test ENG
  - Field acquisition made possible by portable 3/” VTRs
- Consumer videocassette recorders start to appear
  - V-Cord (Sanyo 1974 in Europe)
  - Cartrivision (U.S.1972)
  - Sony: Betamax (worldwide 1975)
  - JVC: VHS (Video Home System) worldwide 1977

Starting the Move to Digital

Video Developments - 1970s

- Digital video "islands"
  - Digital *inside* individual pieces of equipment
  - ***digital timebase correctors*** (1973)
  - ***digital frame synchronizers*** (1974)
  - ***digital graphics*** (1975)
  - ***digital video effects generators*** (1978)
- Still composite video distribution
- 1974: NHK & Sony demonstrate first HDTV system in Japan (analog component 1125/60 system)

Starting the Move to Digital

Standards – 1970s & 80s

- **Analog component video standard set: 1978**
  - **Set RGB standard**
  - **Also set component *color difference* standard**
    - **Y, R-Y, B-Y (aka YPbPr)**
- **Digital video sampling standard set**
  - CCIR 601: 1982
  - Set the 8 bit standard for component digital video
    - ***8 bits are used for EACH color channel***
- **Serial digital video interface standards**
  - SDI (SMPTE 259M): 1989
    - Set component video bit depth at 10 bits/channel
    - Extra bits to eliminate visible contouring in colors

Starting the Move to Digital

Video Developments - 1980s

- **ENG takes a step forward:**
  - **Panasonic M-format (1981): *component* analog format**
  - **Sony Betacam format (1982): *component* analog format**
  - **First entry of true component video into everyday production**
- ***First digital VTRs***:
  - D1: First digital *component* VTR (1986): 8-bit 4:2:2 YPbPr
  - D2: First digital *composite* VTR (1988): "4Fsc" (4 times subcarrier) sampling
- ***Non-linear editing begins*** ~ 1986
- Still largely composite video distribution
- 1984: first analog HDTV recorder: Sony HDV-1000 1" VTR
- 1986: Advanced Television Test Center formed to test & document U.S. advanced television standard
- 1988: first digital HDTV recorders:
  - Sony HDD-1000 1" VTR
  - Sony HDV-10 Unihi ½" videocassette recorder
  - Adopted by the U.S. Advanced Television Test Center, U.S. CableLabs, NHK, HDVision & others for first widespread HD production & testing

Starting the Move to Digital

Developments - 1990s

- 1994: MPEG-2 standard adopted (ISO 13818)
  - First high-quality lossy compression system available
    - Lossy: parts of the original essence are discarded and cannot be recovered
  - Allowed full-resolution SD & HD
  - Adopted for U.S. ATSC & European DVB-T digital television broadcasting standards
  - No ‘lossless’ compression feature
- New ***digital VTR*** formats: Digital Betacam, D5, DVCam, DVCPro, MiniDV, Digital-S
- ***Digital video servers debut***
- ***Computer-based non-linear editing and post production***
- Increasing automation

Digital Video Comes of Age

Developments - 1990s

- HD-SDI standard set (1996)
  - SMPTE 292M
  - Enabled 720p60, 1080i29.97, 480p & 576p digital video
- New ***digital VTR*** formats:
  - Digital Betacam, D5, DVCam, DVCPro, MiniDV, Digital-S
- New *digital HD VTR* formats:
  - HD-D5, HDCAM
- **First appearance of *file-based* production & transmission systems**
  - ***Digital video servers widely available***
  - ***Computer-based non-linear editing and post production***
- Increasing automation run by computer software
- 7/1/1996: United Stated digital television broadcasting standard set
  - ATSC: Advance Television System Committee
  - First station on the air (with bits only): WRAL-DT Raleigh, NC on 7/1/1996
  - First station on air with programming: WHD-DT Washington, DC on 7/7/1996 from NBC Washington

Digital Video Comes of Age

Developments – 2000s

- **JPEG2000 standard adopted:** 2000 (ISO 15444)
- **HD 3Gbps serial digital interface standard:** 2007
  - SMPTE 424M & 425M
- ***HD digital VTRs***: D5-HD, HDCam, DVCPro-HD
- **XDCAM & AVCHD *file-based* acquisition formats debut**

Migration to Digital:
Would You Like Audio with your Video?

Audio Developments - 1970s – 2000s

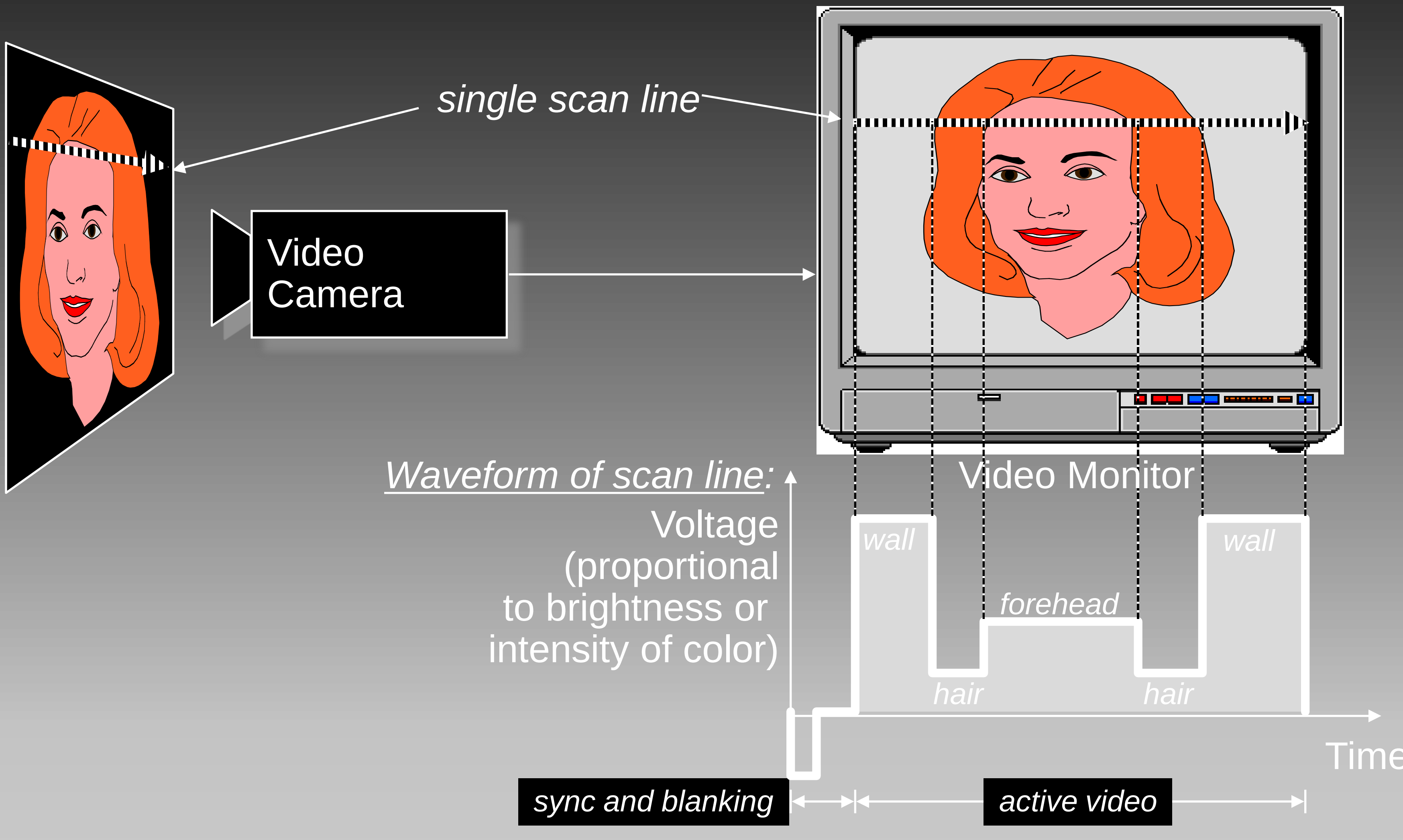
- Digital audio processing “islands”~ 1970’s onwards
  - *digital delays and reverberation units*
  - *digital audio effects*
- Mono → Stereo ~ 1980’s
- ***Digital audio mixers***
- ***Digital audio tape recorders***
- ***Digital hard disk recording and editing***
- ***Digital audio interconnection standard*** 1992 (AES/EBU)
- ***MPEG standard allows high quality audio at very highly compressed bitrates***

Basics of Analog Moving Images

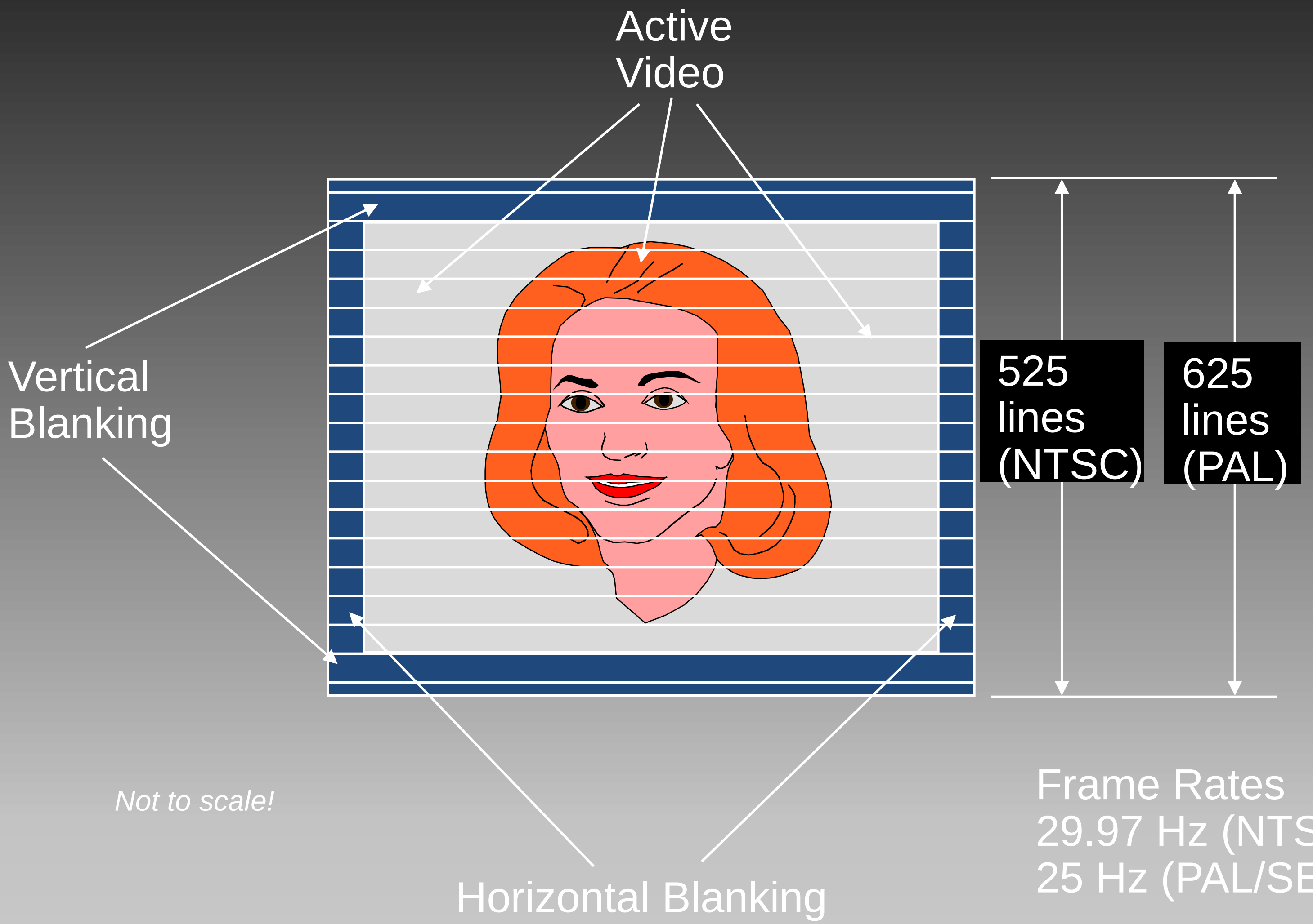
Video

Analog Video

...dissection of image into scanning lines..

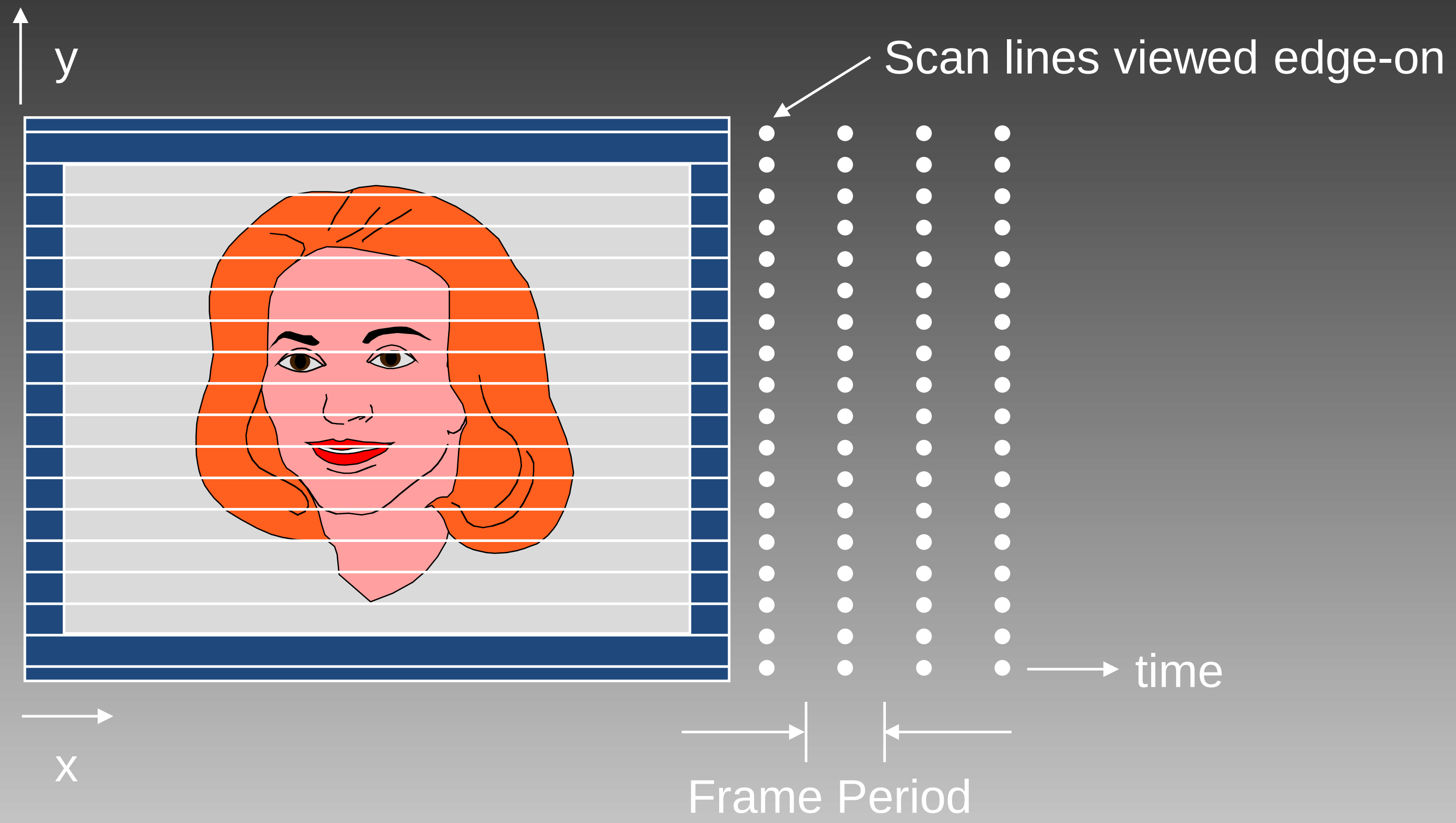


Scanning Raster

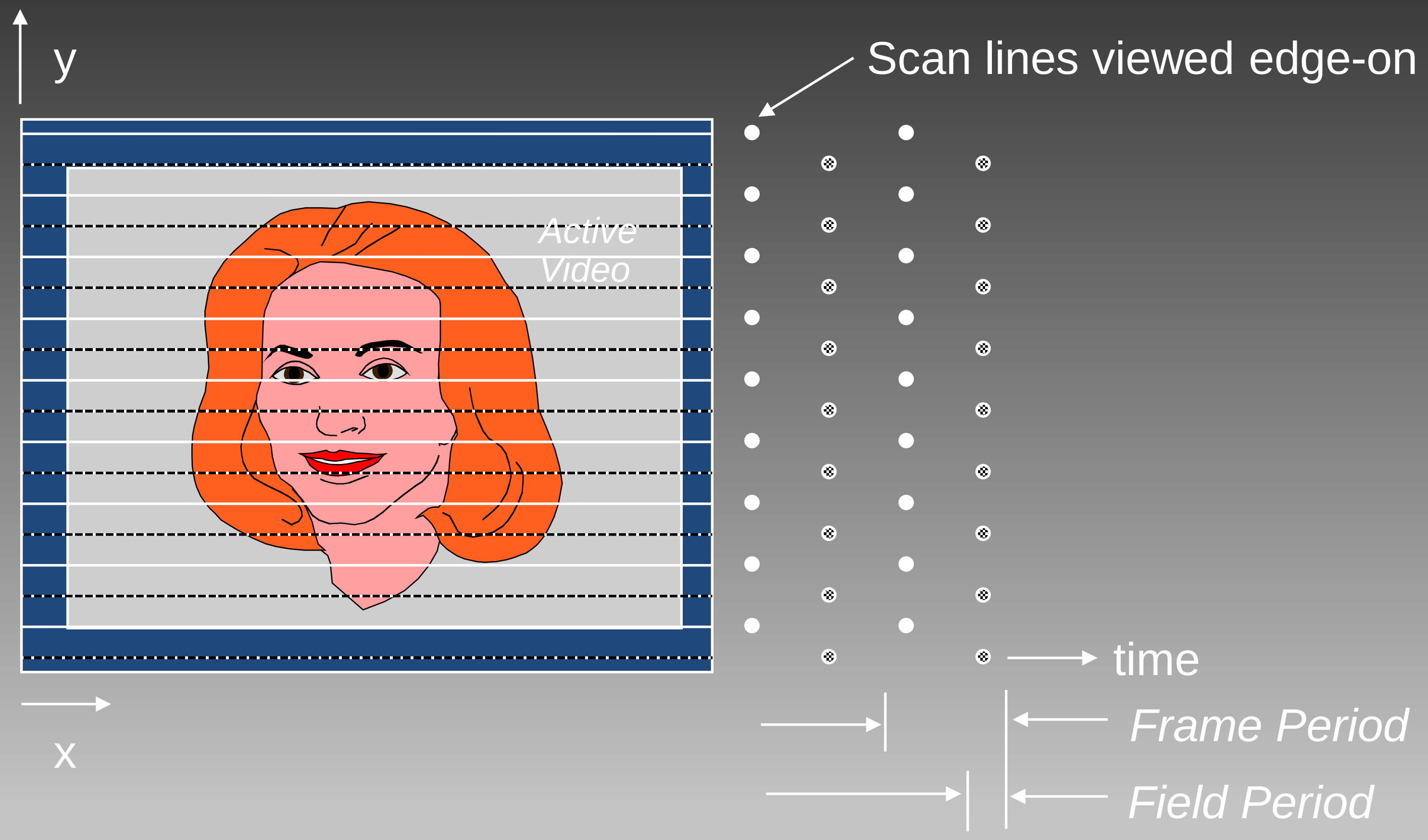


Frame Rates
29.97 Hz (NTSC)
25 Hz (PAL/SECAM)

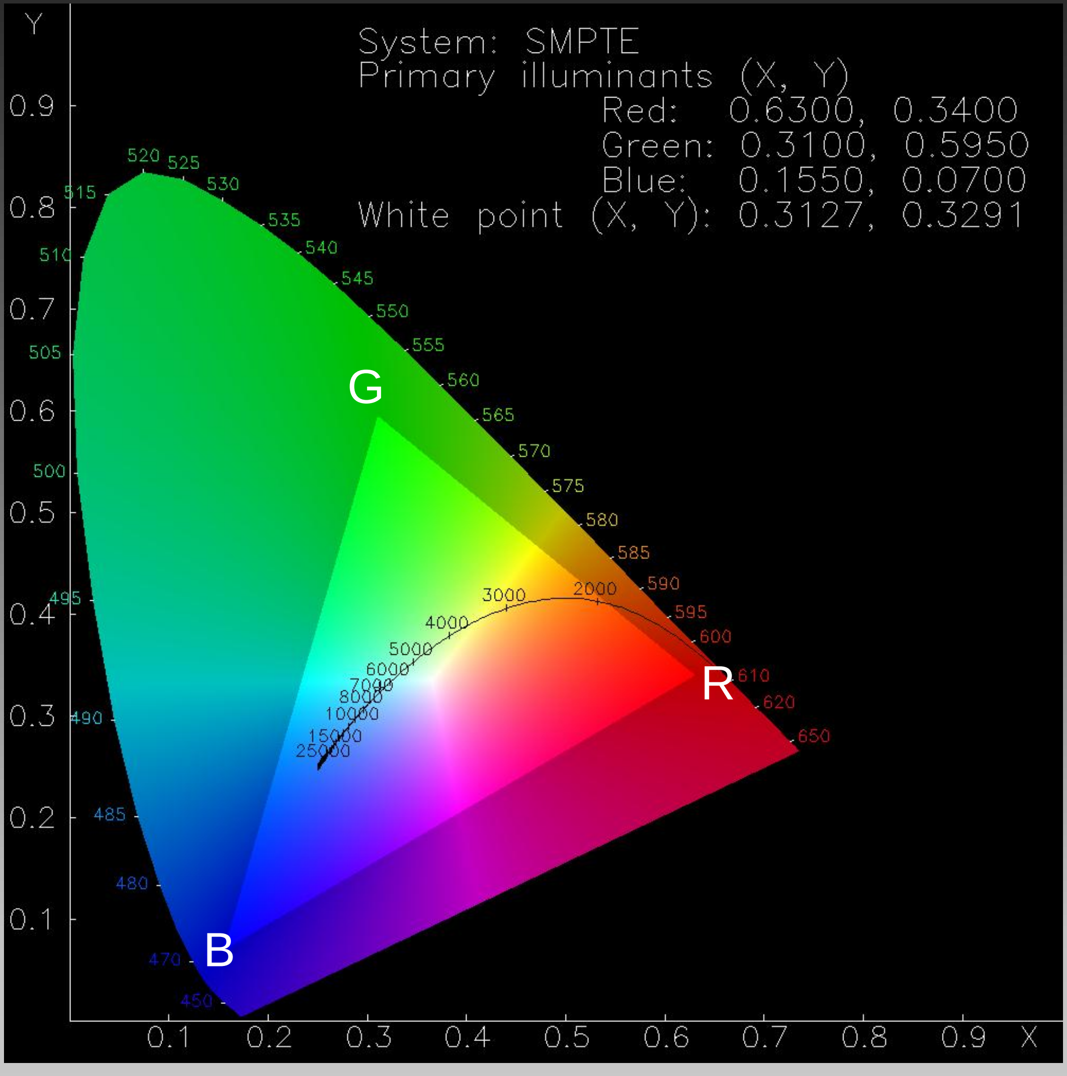
Progressive Raster



Interlaced Raster



Colorimetry



SMPTE CIE x-y color chart

Describes the color values within a photograph / image.

Maps to the voltages created when scanned in an analog form

Maps to the digital values assigned to an analog waveform used to represent the image in digital.

Luminance and Chrominance

- Monochrome, luminance signal known as "Y"

  $Y = 0.3R + 0.59G + 0.11B$

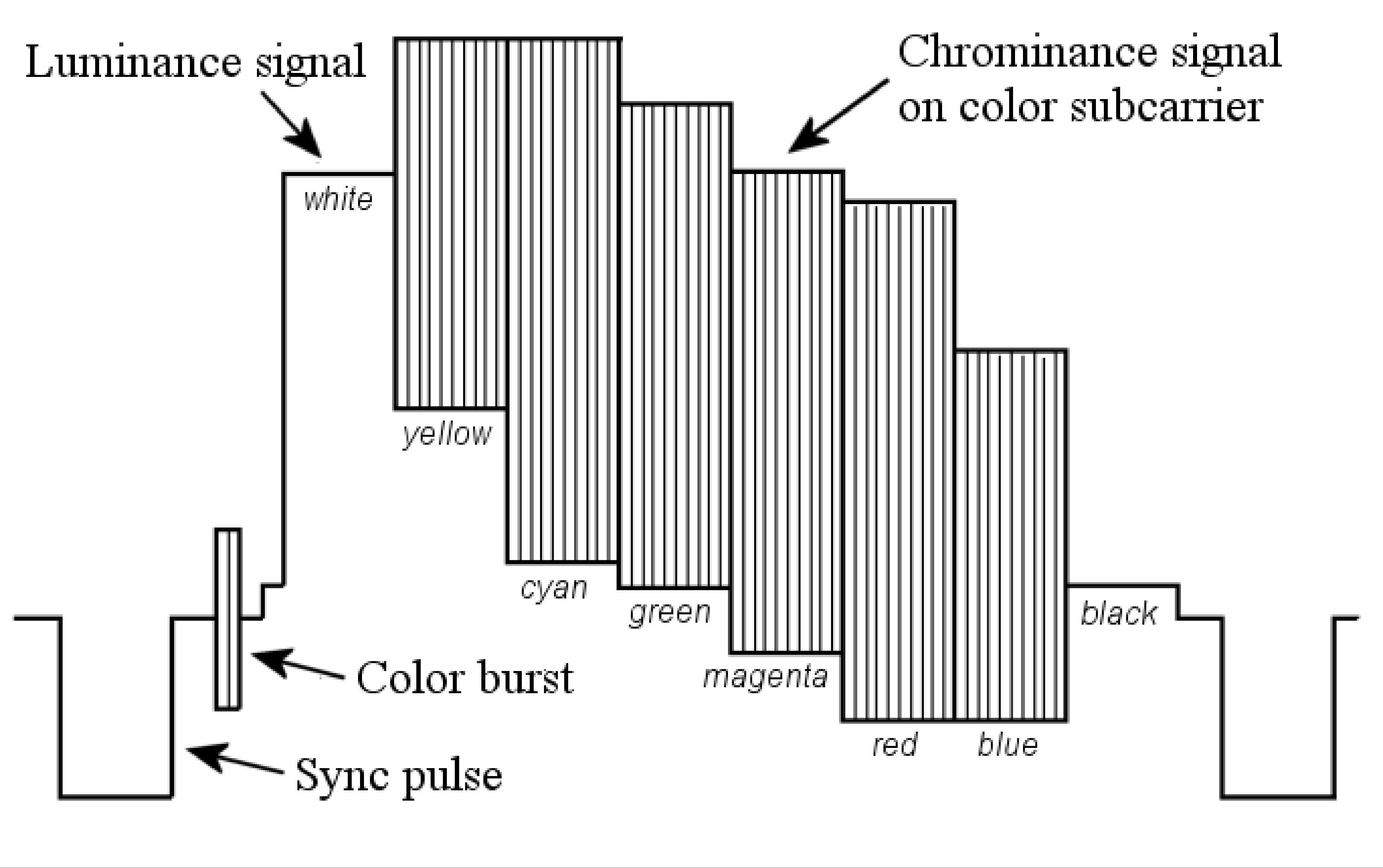
- Color difference signals

  R-Y

  B-Y

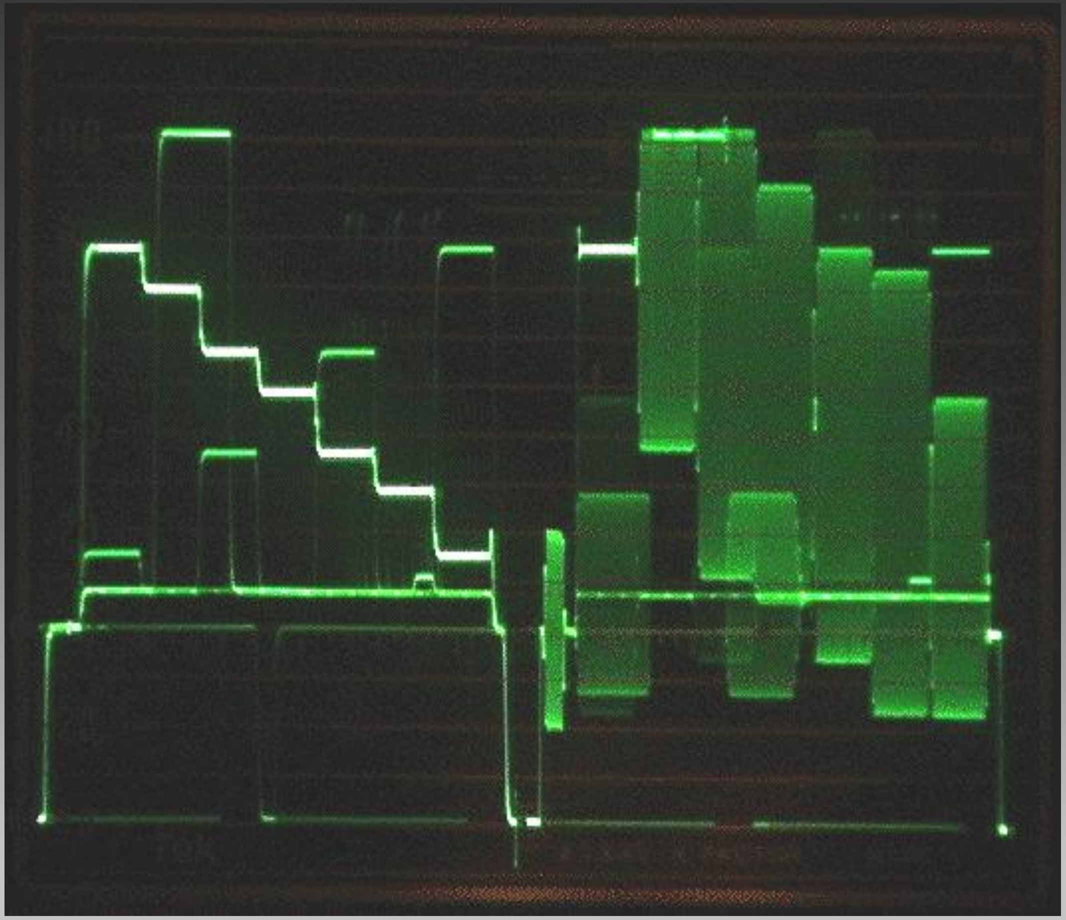
  In NTSC, these are processed to become the I and Q signals that are modulated onto the color subcarrier
  In PAL they become U and V.

Analog Waveform



Analog Waveform

Interlace - as seen on a waveform monitor



1st (odd line) field

2nd (even) field
with color info overlaid

one frame

Most Common Video Formats

- Expressed in horizontal pixels by vertical lines:
  - Don’t forget the frames per second! (fps)
- 720 x 486 (525 lines in analog)
  - 29.97 fps
- 720 x 576 (625 lines in analog)
  - 25 fps
- 1280 x 720 (a digital only format)
  - Can have 23.98, 24, 25, 29.97, 30, 59.94 & 60 fps
- 1920 x 1080 (a digital only format)
  - Can have 23.98, 24, 25, 29.97, 30 fps
  - 59.94 & 60 fps are in development
  - Watch out for the dreaded “1080i60”! It doesn’t exist…except in Sony’s terminology….
    - It’s actually 1080i29.97

Interlace vs. Progressive

- This presentation uses:
- ‘i’ for interlace
- ‘p’ for progressive

- Example:
- 486i29.97
  - 486 lines, interlace scanned, 29.97 frames per second
- 720p60
  - 720 lines, progressive scanned, 60 frames per second
- 1080i29.97
  - 1080 lines, interlace scanned, 29.97 frames per second

Digitizing Decisions

- To maintain THE ORIGINAL QUALITY of analog video:
  - 10 bit sampling must be used
    - 4 times the accuracy of 8 bit video
  - A high-quality analog-to-digital converter must be used that separates the color from the B+W as cleanly as possible
    - Warning: it will NEVER be completely clean!

Inspecting Your Media

What to look for….

Inspecting Your Media

- Tape:
  - Watch out for sticky-shed
  - Pre-"broadcast quality" tapes are showing the worst ravages of time
    - Scotch and Ampex both debuted broadcast quality tapes roughly 1978
    - Anything before 1978 watch for:
      - Sticky-shed
      - Tape leader & splices that fail
      - Mechanical failures
  - If you can afford a tape inspector: BUY IT!!!!

Sticky-shed Treatment

- Only proven treatment is baking
  - Heating the tape in an oven to ~130 degrees for several hours
  - Does NOT permanently stop the condition
  - May make it worse in the long-term
- Use with great caution! There is a good chance you will damage or destroy your tape
- Is there something better?

Sticky-shed Treatment

- Is there something better?
  - We're studying low-humidity treatments
    - storage in low humidity for long periods
    - Treatment in a low humidity chamber with active dry airflow for shorter periods
    - Can we adapt our playback devices to mitigate sticky-shed?
      - Cooling of the head-stacks at the point of contact
  - Stay tuned....

Inspecting Your Media

- Discs:
  - – LaserDisc: watch for disc rot
  - – CED (“RCA Selectavision”): make sure case is OK
    - DO NOT OPEN THE CASE OR TOUCH THE DISC!!!!
  - – VideoCD & DVD:
    - Pressed discs (IE silver or gold color): inspect the surface for any signs of rot
    - ‘burnable’ discs: copy them as soon as possible!!
      - – They are not known to age well

Basics of Analog Moving Images

Motion Picture Film

Film Formats

- Most common formats:
  - 35mm (most common theatre format)
    - Super35: camera negative format with larger negative taking up the space where the optical soundtrack would be for higher quality
  - 16mm
    - very common educational/instructional format
    - Used for TV news & documentaries
    - Sometimes with magnetic strips for sync sound
    - Very common kinescope format
    - Most common distribution format to TV stations before mid-1960s: stations not connected to networks by microwave
    - Super16: camera negative format with larger negative that take up where the optical soundtrack would be for picture area
- 8mm – consumer format, sometime a split 16mm film
- Super8:
  - consumer format starting in the 1960s with smaller sprocket holes that permitted more picture area
- 70mm

Film Formats

- Some uncommon formats:
  - 9.5mm
  - 17.5mm (35mm split in half)
  - 28mm
  - 65mm
    - usually camera original negatives shot for 70mm distribution

Film Resolution

- Resolution is measured in line pairs per millimeter (lp/mm)
  - Roughly equivalent to video lines of resolution
  - Sometimes express in line pairs per picture height to compare electronically scanned equivalent
  - Depends on many factors
    - How the emulsion was made
    - How the film was shot (lenses affect resolution actually captured)
    - How the film was processed
      - Mass produced copies have lower resolution than camera original negatives
      - Forces processing increases grain which decreases resolution
- Film format – rough resolution maximums when scanned:
  - 35mm: up to 16,000 lines on the best negatives
    - 800-1000 lines/frame on theatrical release prints (8 generations from camera negative)
  - 16mm: roughly 4000 lines/frame
  - Super16mm: 6000 lines/frame
  - 8mm/Super8: 2000-3000 lines/frame

Digitizing Decisions

- When converting from analog film to digital:
- Lines of resolution
  - Preservation? Get as close to the resolution of the film as possible
  - Access? HD or SD will do since it can use common production tools
  - Online? Be careful! Choices to go cheap now can come back to haunt you in the future of broadcast-to-the-home!
    - YouTube has a 4k streaming content page!!!
- Bit depth (expressed in bits per channel)
  - Determines the number of color shades that can be represented
  - Anything under 10 bits will show visible contouring
  - Current scanners can manage up to 16 bits/channel
- Color space
- Masking?
  - Do you scan to the edge of the frame?
    - Frames can have very rough looking edges
  - Do you crop while scanning, or in post-processing?
- Many times dictated by how expensive a scanner you can afford

What's This 'Digital Cinema'?

- 'Digital Cinema' standardized by the DCI: Digital Cinema Initiative
  - More info: www.dcimovies.com
- Accepted as the digital motion picture standard by the SMPTE (Society of Motion Picture & Televison Engineers)
  - www.smpte.org
- Two standardized formats, each with two flavors:
  - "2k": 2048 pixels by 1080 lines (D-cinema definition)
    - Hmmm…. widescreen HD! (HD is 1920 x 1080)
    - 2nd variant 2048 x 1556 lines: valid capture format for preservation?

What's This 'Digital Cinema'?

- Two standardized formats, each with two flavors (con't):
  - "4k": 4096 pixels by 2160 lines (D-cinema definition)
    - 4 times the resolution of 2k
    - 24 or 30 frames/second
    - 2nd variant 4096 x 3112: a good preservation format for 16mm?
      - Film scanners today are shipping with this format ability
- Both D-cinema definition formats are DISTRIBUTION formats
- They are COMPRESSED (JPEG2000 Digital Cinema Profile)
- They are usually ENCRYPTED for transmission to theatres
  - Not ideal for long-term archiving

What’s This ‘Digital Cinema’?

- Produced and distributed as files:
  - DSM: Digital Source Master
    - The final edition version of an electronic cinema production
  - DCDM: Digital Cinema Distribution Master
    - The final step before compression and encryption. This is the motion picture in its final form but still at its highest quality.
  - DCP: Digital Cinema Package
    - Compressed, encrypted and keyed to individual projector’s serial numbers

But What Do WE Call It….?

- …if we aren't producing DCI compatible files?
- We use 'Electronic Cinema'
  - Calls attention to the fact that it may NOT be DCI compliant moving image material
  - Also allows us a lot more flexibility in bit depths, color spaces, etc.

Inspecting Your Media

What to look for….

Inspecting Your Media

- Film:
  - Could it be Nitrate?
    - ‘moth balls’ (naptha) or wet dog smell
    - Call LOC’s film lab ASAP unless you know how to properly store and handle Nitrate
    - Most 35mm produced before 1951
      - Most commercial theatre releases were on Nitrate
  - Acetate:
    - Does it smell like vinegar?
  - Polyester:
    - Is it scratched?

Inspecting Your Media

- Film soundtracks:
  - Optical or magnetic?
  - Optical: variable density or variable area?
  - Magnetic: treat with caution! The magnetic strips can delaminate easily
- Film:
  - All film should be cleaned before it is scanned or copied
  - Be careful not to clean off your magnetic soundtrack!
  - Not cheap

Moving Image Digitization

Basic Concepts

What is Digital Video?

- System where video information is carried using digital numbers instead of an analog voltage:
  - Samples of each television line are represented as a series of numbers using binary (digital) coding
  - Sync and blanking pulses are dropped as a tool for raster retrace timing, but the sync timing relationship is maintained using special codes within the digital stream

Binary and Decimal Numbers

Decimal	1	2	3	4	5	6	…	255
Binary	1	10	11	100	101	110	…	11111111

Important Concepts

- Bit depth
- Color space
- Native resolution
- Native frame rate
- Time code
- Keeping your audio in time with your video
- Nyquist limit

Important Concepts

- Bit depth
  - Determines the number of steps from dark to light that are captured: the more steps the smoother the transitions
  - Expressed in bits/channel: 10 bits means 10 each for R, G, B or Y, Pb, Pr, for a total of 30 bits/sample
  - In audio: determines the signal-to-noise ratio of your audio
- Color space
  - Defines how you are representing your picture:
    - RGB, YPbPr?
    - Can affect any processing or post-production of the product
- Native resolution
  - What is the ACTUAL resolution of your material?
  - ANY conversions from the native resolution will be visible in your end product (some more than others)
- Native frame rate
  - What is the ACTUAL frame rate of your material? 24fps? 29.97fps? 60fps?
  - ANY conversions from the native frame rate will be visible in your end product (some more than others); only extremely expensive format converters can hide this well (IE over $500,000)

Important Concepts

- Time code
  - Make sure your time code matches your frame rate (a major problem if you convert to another frame rate)
  - Make sure any ‘NTSC friendly’ frame rates use Drop Frame time code (IE 29.97 or 59.94)
  - 720p59.94 and 720p60: can’t use current time code (SMPTE 12M standard) without issues: watch out!

Important Concepts

- Keeping your audio in time with your video
  - If you convert your video to a different frame rate, you MUST convert your digital audio as well or it will lose ‘lip sync’ (IE the audio goes out of sync with the video)
  - Something as small as 29.97 to 30fps will cause AUDIBLE errors without proper audio conversion

Nyquist Limit

- Mr. Nyquist defined that you needed to sample an analog waveform you wish to represent at AT LEAST twice the highest frequency in the analog signal to prevent unwanted artifacts (called 'aliases')
  - You can sample at HIGHER than the highest frequency without penalty; NEVER lower.
- For example:
  - Video: 6 MHz picture requires at least 12 MHz sampling rate
    - SMPTE standard: 13.5 MHz for SD, 74.25 MHz for HD
  - Audio: 20 kHz audio bandwidth requires at least 40 kHz sampling rate
    - AES standard: 44.1 kHz (commercial CDs) up to 192 kHz (the best for preservation)

Analog to Digital Conversion

Step by Step

Analog to Digital Conversion

- Three basic steps:
  - Sampling
  - Quantizing
  - Coding

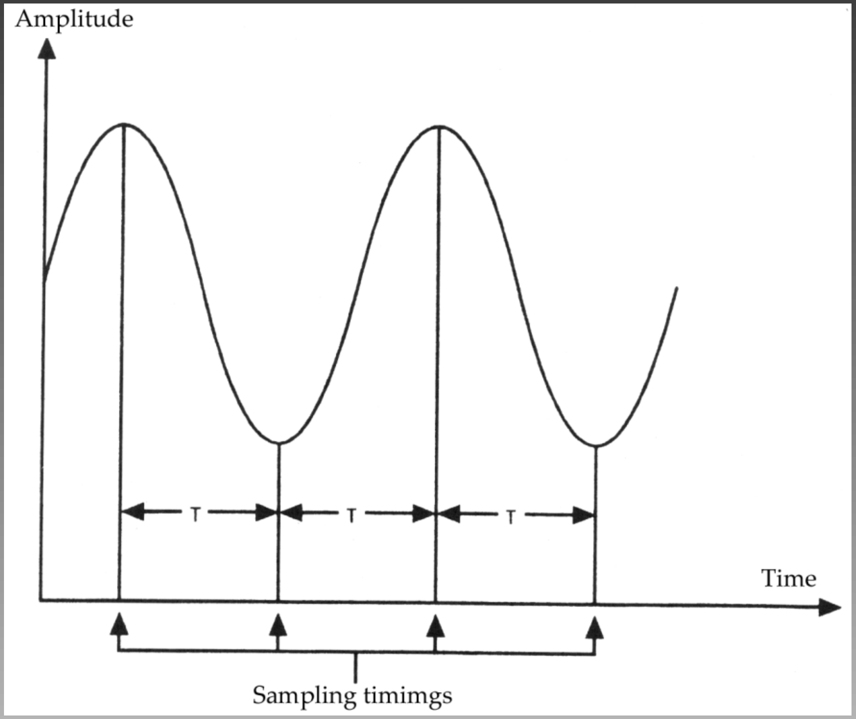
Analog to Digital Conversion

Step One:

- Sampling
  - The analog video waveform amplitude is sampled (measured) at regular intervals and converted to a set of digital values. The number of samples per second is known as the "sampling frequency"
  - The analog signal must be sampled at a minimum of twice the highest frequency in the signal. This is the "Nyquist limit".

Analog to Digital Conversion

Sampling



Minimum sampling rate

SD video: 13.5 MHz = 13.5 million luminance samples per second is used

HD video: 74.25 MHz = 74.25 million luminance samples per second is used

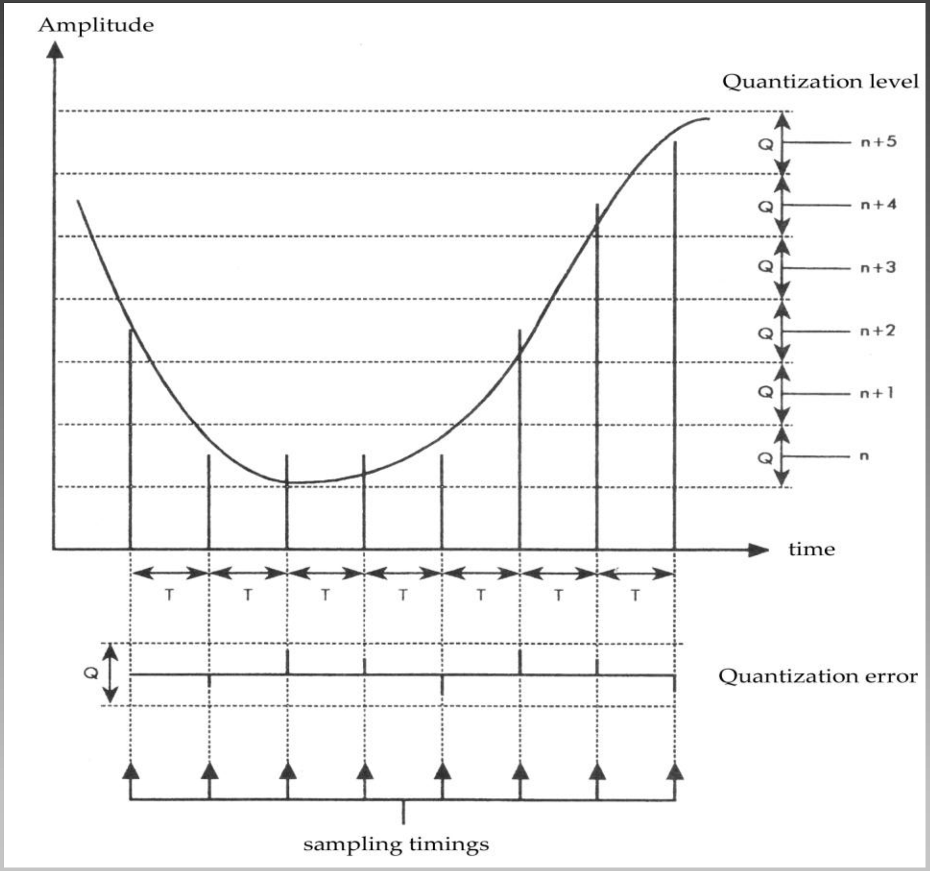
Analog to Digital Conversion

Step Two:

- Quantizing
  - For each sample from the continuous waveform, decisions are made as to what digital value will be assigned for a given voltage:
    - If a voltage value falls between the two numbers available to represent the two closest values, a decision must be made as to whether the higher or lower value will represent that particular value. Functionally the same as "rounding" up or down.

Analog to Digital Conversion

Quantizing



Analog to Digital Conversion

Step Three:

- Coding
  - Assigning numerical values to the sampled waveform. For 8-bit video, black (or darkest) is assigned 16 and white (or lightest) is 235. All other shades between white and black must fit between digital value 16 and digital value 235
  - 10 bit video adds four times more values: 16 to 1016. Smoother graduations with less obvious “stepping” (visible contour lines in the picture where there shouldn’t be lines)
  - Electronic cinema cameras & film scanners are now capable of up to 16 bits/channel
  - Each of three color channels are coded:
    - RGB, XYZ, YPbPr
    - 10 bits per color channel x 3 color channels = 30 bits/sample
    - 16 x 3 = 48 bits/sample

Analog to Digital Conversion

Related steps:

- Decoding composite video to luminance and color difference components before sampling
  - Not needed if signal is already in component form
- Bandwidth filtering of luminance and color difference signals before converting to digital
- Arrangements for sampling the color signals

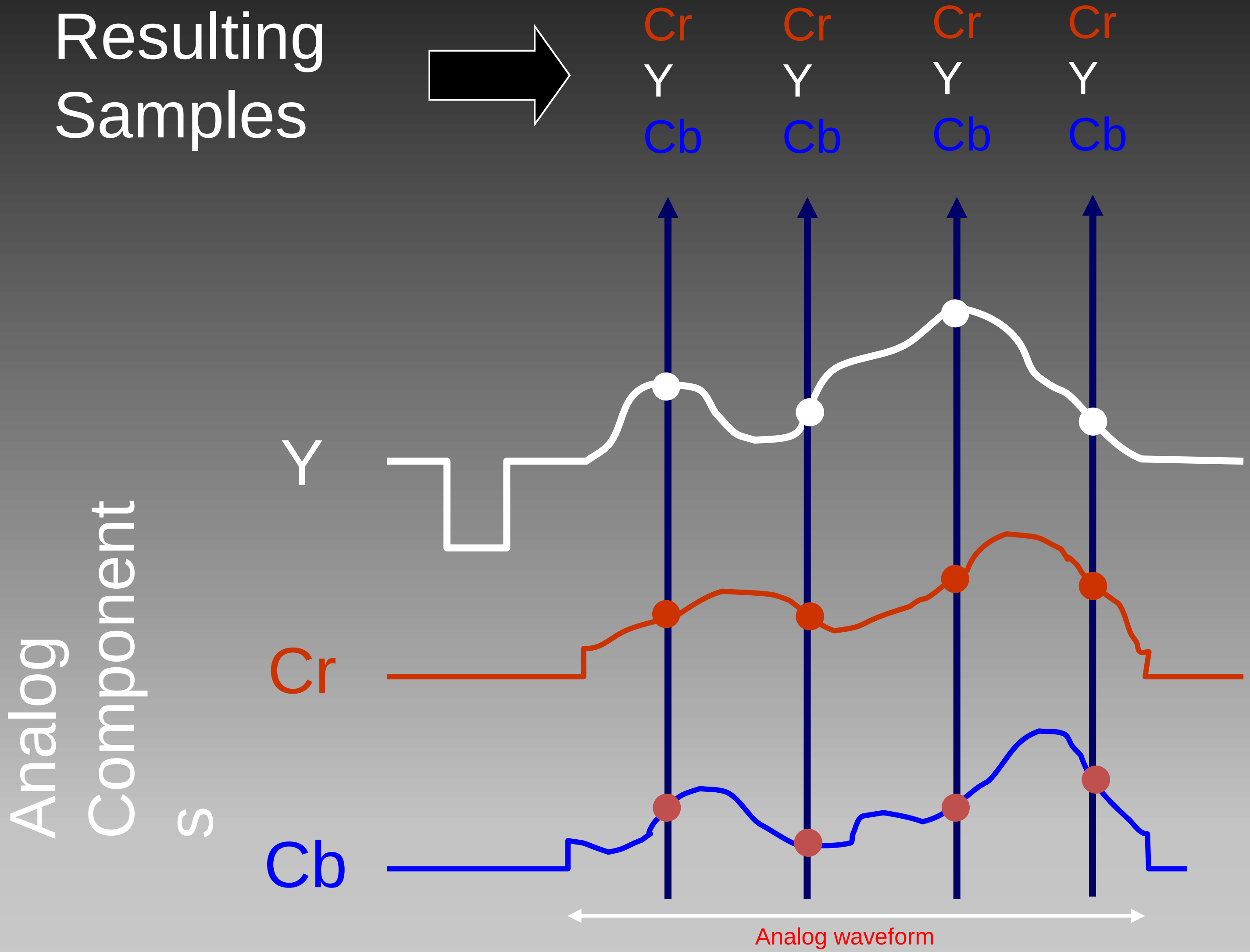
Proper Terminology: Component

- RGB: Component
- Component color difference
  - YPbPr (aka Y, R-Y, B-Y): analog, where all 3 are related to each other mathematically
  - YCbCr: digital version of YPbPr transmitted serially in a single digital signal
- Composite
  - YUV: Used in PAL analog composite; misused to signify component color difference in many contexts
  - YIQ: Used in NTSC analog composite

Color Subsampling

- Color Difference channels may be subsampled
- Easy way to reduce bit rate before applying actual compression in compression systems
- Exploits the human visual system's reduced ability to see color detail
- Digital equivalent of chrominance (color) bandwidth reduction in NTSC and PAL

4:4:4 Sampling of One Line



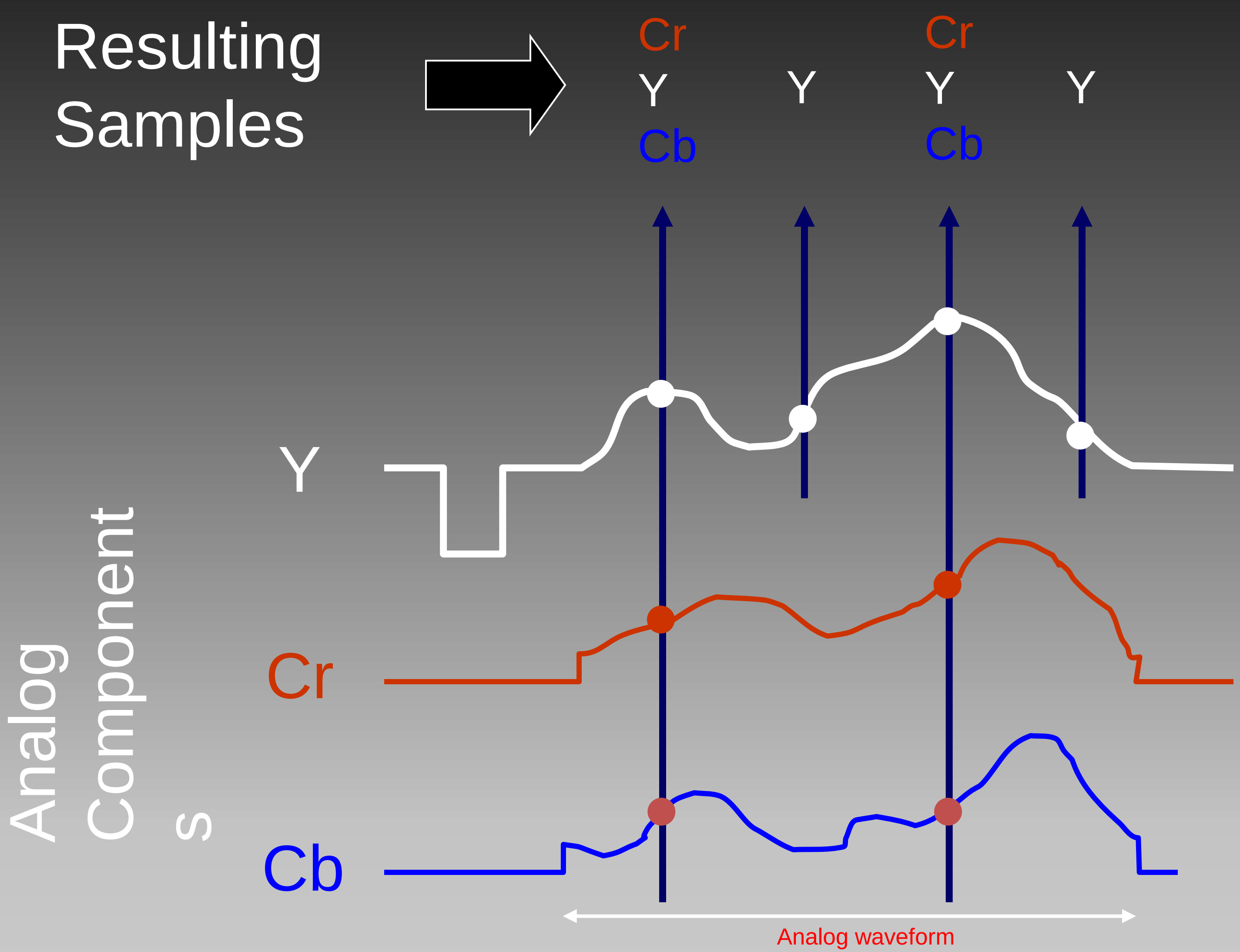
Video is sampled @ 4:4:4

Y: 4 samples

Cr: 4 samples

Cb: 4 samples

4:2:2 Sampling of One Line



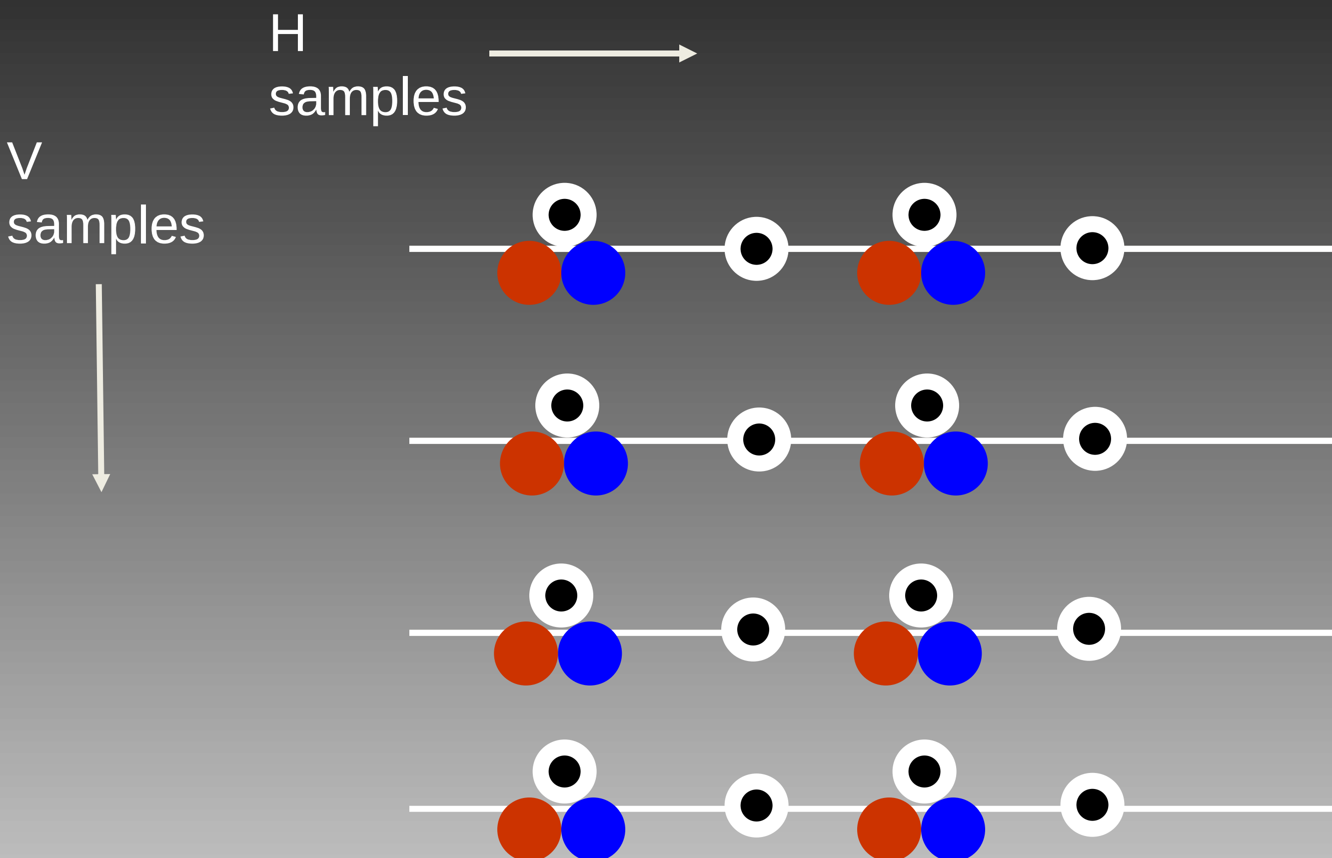
Video is sampled @ 4:2:2

Y: 4 samples

Cr: 2 samples

Cb: 2 samples

4:2:2 Sampling of One Frame



Video is sampled
@ 4:2:2:
on each line

4:1:1 Sampling of One Line

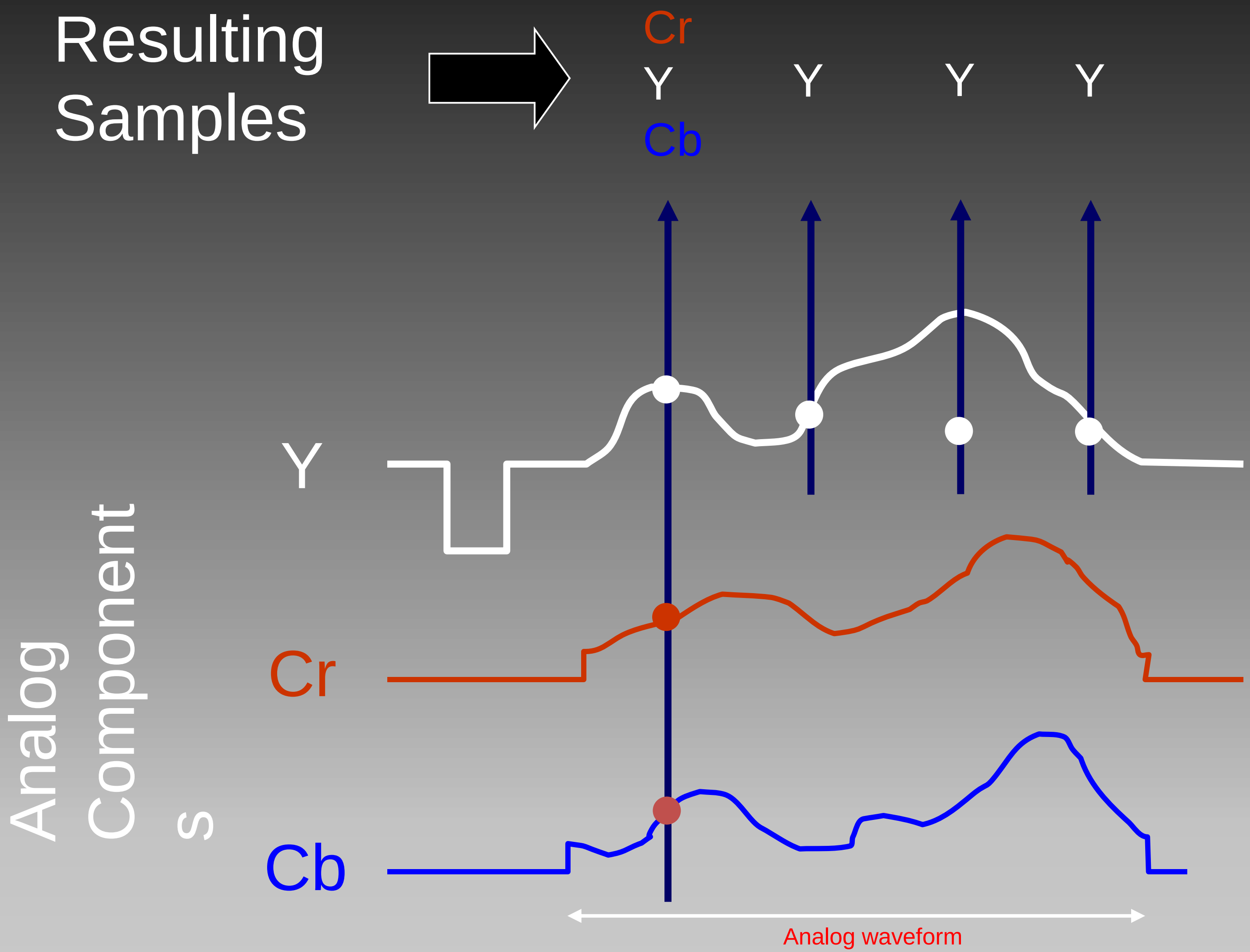
Resulting Samples

Video is sampled @ 4:1:1

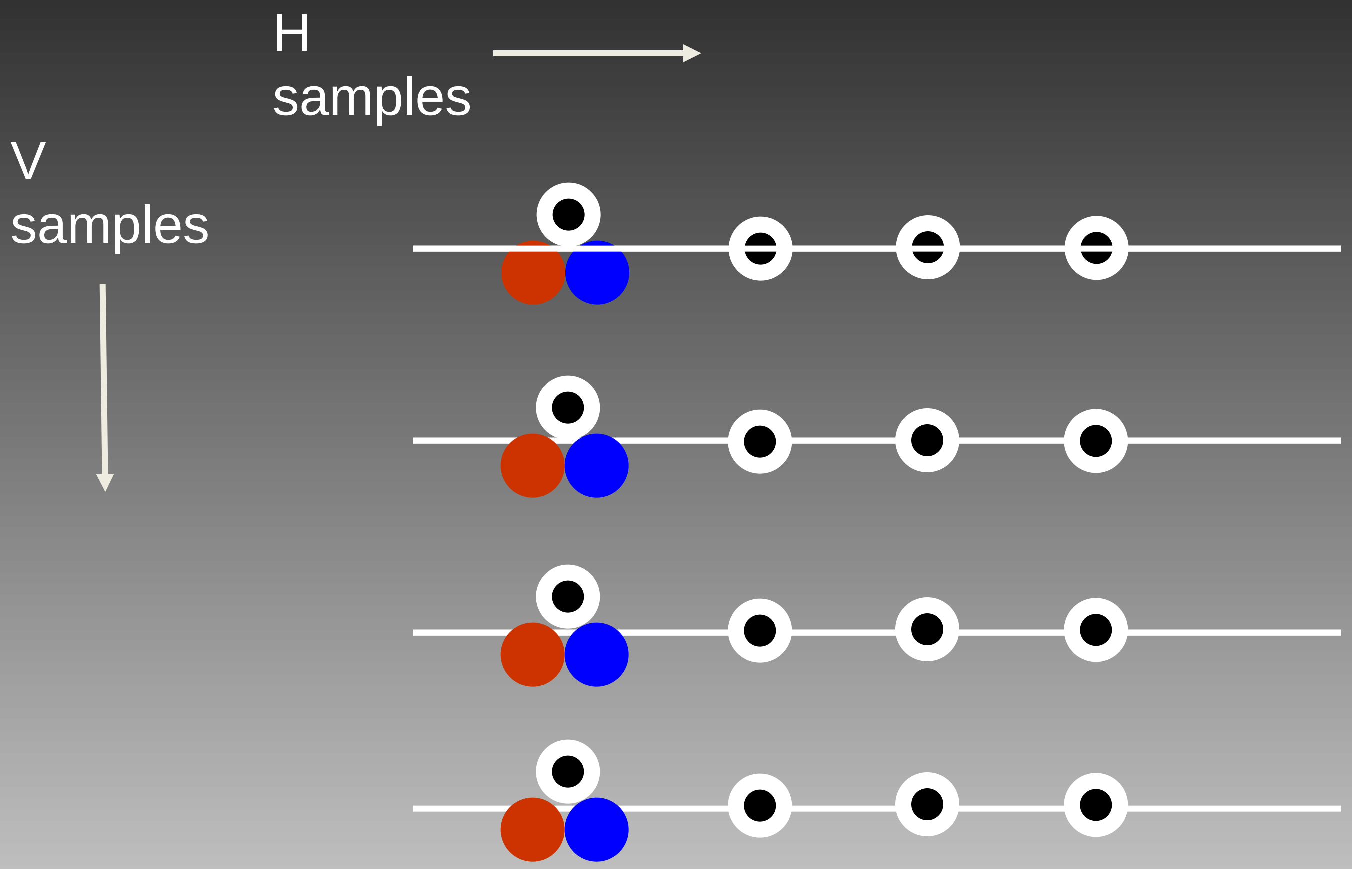
Y: 4 samples

Cr: 1 samples

Cb: 1 samples

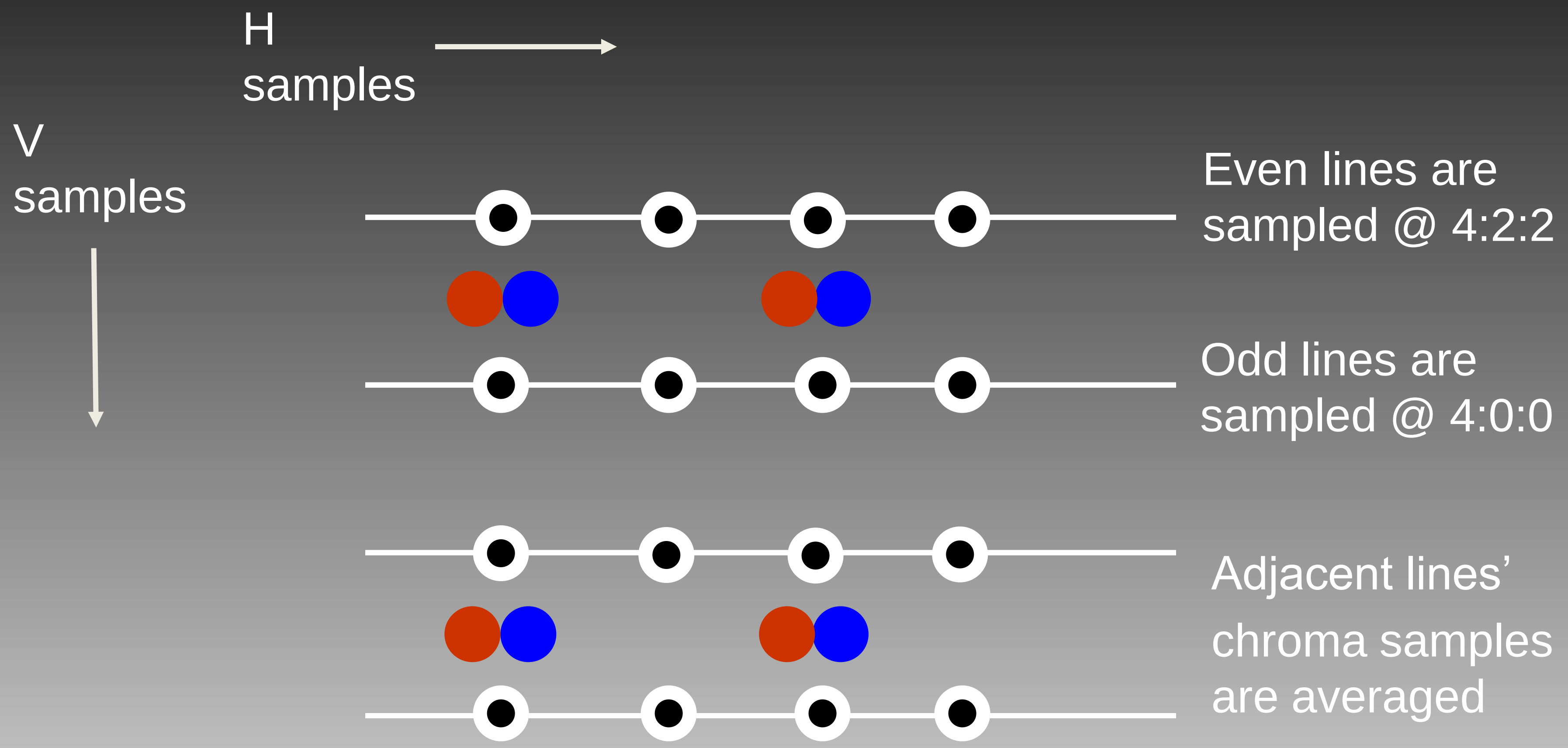


4:1:1 Sampling of One Frame



Video is sampled
@ 4:1:1
on each line

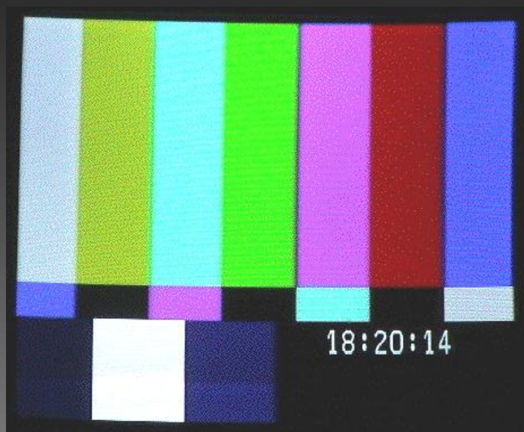
4:2:0 Sampling of One Frame



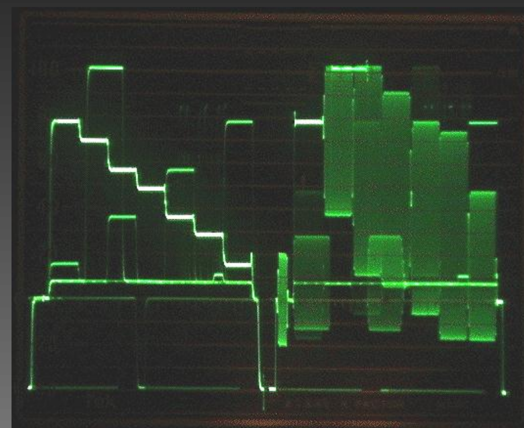
Component Video Signals

Familiar, yet different….

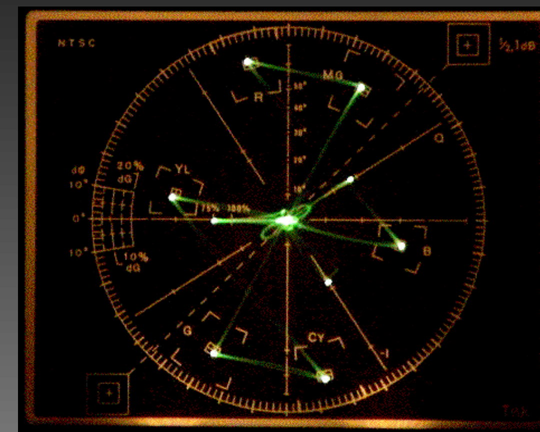
Waveforms: NTSC and Digital Component



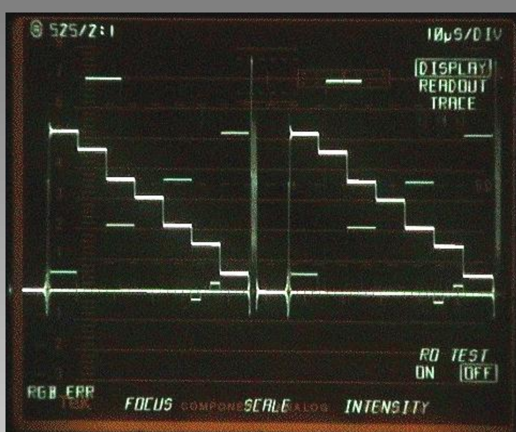
Original picture



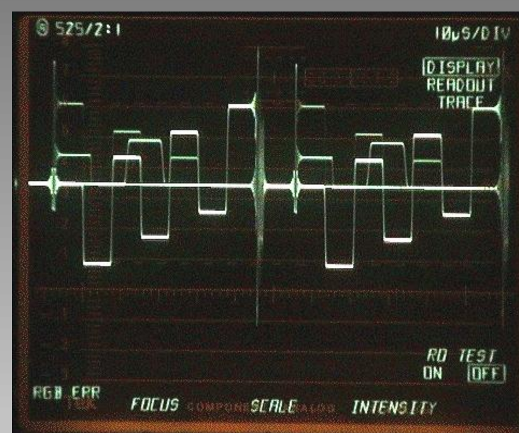
Composite waveform



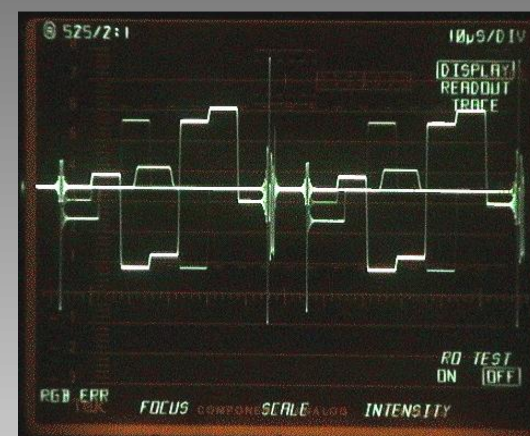
Composite vectorscope



Y



Cr



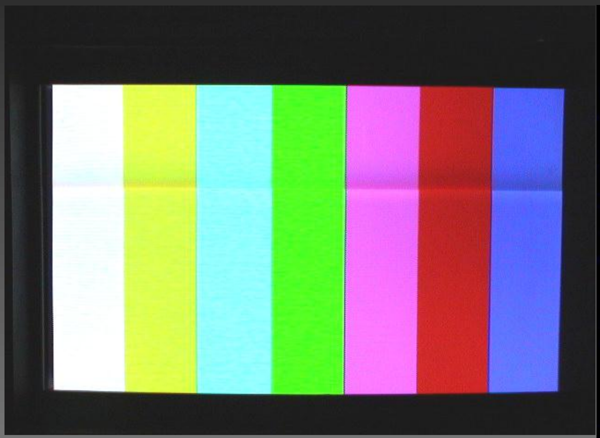
Cb

Waveforms: Old & New

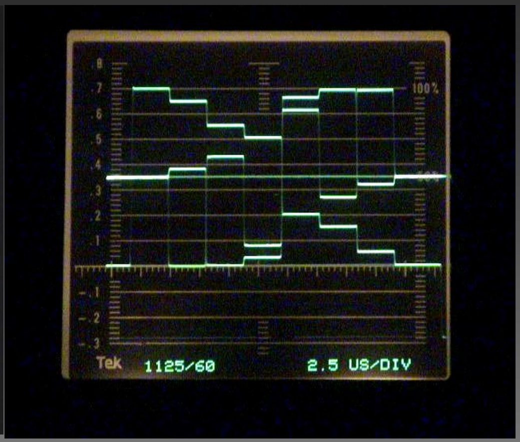
- NTSC:
  - Waveform monitor measures luminance
  - Vectorscope measures chrominance (I and Q in the 3.58 subcarrier)
- Component video:
  - Three signals:
    - Y is the luma component
    - Cr is the R-Y (Red minus luma) component
    - Cb is the B-Y (Blue minus luma) component

High Definition Video

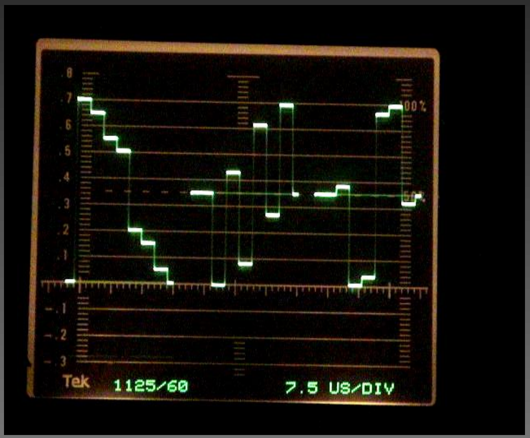
HD Digital Component Waveforms



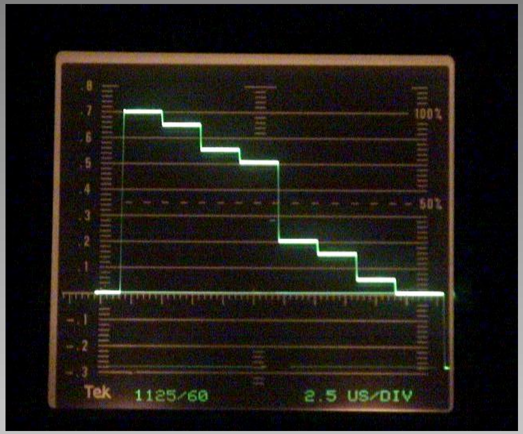
HD color bars



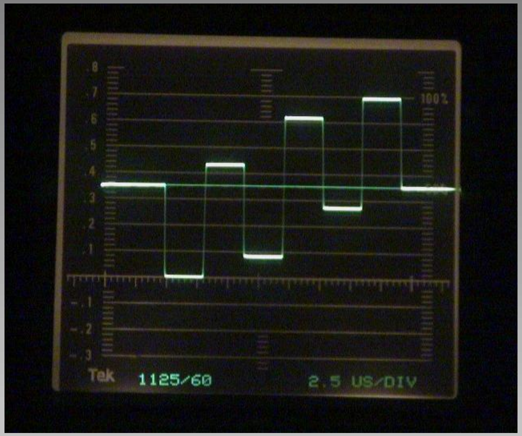
YCrCb overlay



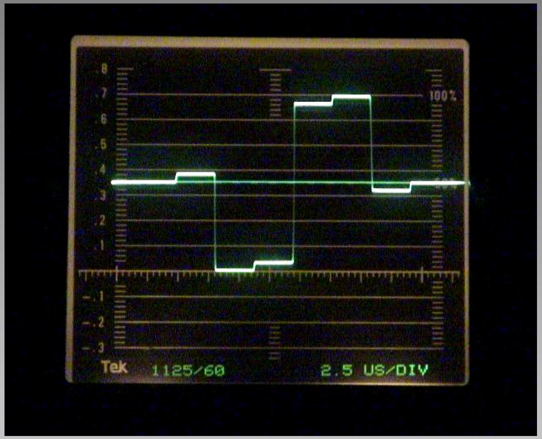
YCrCb side-by-side



Y



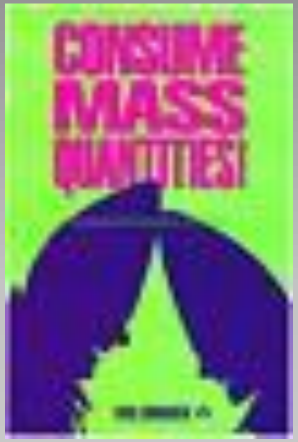
Cr



Cb

Digitizing Everything

Consume Mass Quantities!



Overall Digitizaton Goals

- Archive must be maintained 'for the life of the Republic plus 4000 years'
  - We have a different perspective on the word 'longevity'
- Files should be able to stand alone without references to external databases
  - Lots of metadata wrapped in the files!
- Digitize items in their original quality
  - Audio at highest bandwidth in the recording
  - Video at its native resolution
  - Film scanned at resolution equal to the highest available on film
    - 4k x 4k for 16mm; 8k x 8k for 35mm in development

Overall Digitizaton Goals

- Use as much off-the-shelf products as feasible
- Invent items or write custom software only when a commercial product cannot fit the need
- Use industry standards as much as possible
- Create realistic long-term goals & budgets
- Calculate long-terms needs and timelines for realistic goal-setting

Overall Digitizaton Goals

- Audio (for moving image content):
  - 48 kHz/24 bit BWF (Broadcast Wave Format) for video
  - 96 kHz/24 bit BWF (Broadcast Wave Format) for film
  - Multiple multichannel mixes must be planned for
- Video (SD & HD):
  - JPEG2000 ‘lossless’ (reversible 5x3) (ISO standard 15444)
  - MXF OP1a (SMPTE standard 377M-2004)
- Film: (in development)
  - Short-term expediency: DPX with BWF audio
  - Goal: JPEG2000 lossless with PCM audio with MXF OP1a wrapper

Basics of Born Digital Moving Image Content

Born Digital Moving Image Content

- Two basic types:
  - Physical media:
    - Videotape
    - Optical discs
  - Files:
    - Content created as files
  - Hybrid:
    - Files saved on solid state media

Born Digital Moving Image Content

- The easy part:
  - The program material is already in one of the picture & audio formats we've discussed!
- The hard part:
  - It can be in any of dozens of different flavors of compression
  - Just as with tape formats, over time it will become difficult or impossible to play back certain formats or 'flavors' of common compression formats

What LOC is Doing

- Keep the original file 'for the record'.
- Convert the essence and key metadata to JPEG2000 Lossless, MXF OP1a file with embedded metadata
  - Native format & frame rate
  - Will allow access to material even if the ability to replay the original file is lost

Why JPEG2000 Lossless?

The first practical compression standard that doesn't throw away picture content in any way.

A Few Numbers….

- Uncompressed SD: 270 Mbps
- Uncompressed HD: 1.485 Gbps
- Now do the math:
  - 270 Mbps x 3600 sec/hr = 972,000 Mb/hour (972 Gb/hr)
    - Divide by 8 bits/byte = 121.5 GB/hour
      - This is your storage requirement for ONE HOUR of uncompressed SD video
  - 1.485 Gbps x 3600 sec/hr = 5376 Gb/hour
    - Divide by 8 bits/byte = 668.25 GB/hour
      - This is your storage requirement for ONE HOUR of uncompressed HD video
- Ouch!

A Few Numbers….

- JPEG2000 Lossless (“reversible 5/3”)
  - Averages 2.3:1 compression (depending on program content)
    - All black compresses over 100:1!!
      - Might be a bit boring, though….
- Compare uncompressed SD:
  - 121.5 GB/hour / 2.3 = 52.83 GB/hour
- Compare uncompressed HD:
  - 668.25 GB/hour / 2.3 = 290 GB/hour
- When you have millions of hours, 2.3:1 means real money.

Where Do We Store All This Material?

Archive vs. Repository

Archive

- Contents are not meant to be accessed except when absolutely necessary
- Contents must be preserved for long periods of time
  - 1000s of years?
- Will require multiple migrations over time
- Can be part of a larger digital repository
- The bulk (90%+) of MBRS' repository is the archive

Repository

- Contains all of the information and documents created by business processes
- An archive can be a subset
- Non-archive data can be discarded after certain period in many cases

Our Digital Repository

- 200 TiB SAN
  - Staging area for transmission to backup site and the tape library
  - Backup site has identical SAN & tape library
- Tape library
  - StorageTek robot with 9000 slots currently installed; 37,500 planned by 2013
  - Currently using T10000 tapes with 1TiB/tape current capacity (9PiB available; 37.5PiB without refresh)
  - Upgrade path to ~48TiB/tape by 2019

The Digital Repository

- SAM-FS v4.6 file system currently
- 1016 TiB on tape as of Friday (1/21/2011)
- Increasing at approx. 30 TiB/week now
- First PiB happened on Tuesday 1/18/2011 @ 10:30pm!
- First ExiB anticipated between 2013 & 2017
  - We're in a race to get the tape technology refreshes to keep up before we run out of space!

What the….?

MiB? GiB? TiB? PiB?!?!

Get Your Terms Correct!

- kilo, Mega, Giga, Tera, Peta, Exa
  - All base 10 measurements
- kibbi, mibbi, gibbi, tibbi, pibbi, Exi
  - All 1024 base measurements
- It matters when you get to the PiB level!
  - Like using inches when measuring millimeters: the difference gets to be drastic if you aren't precise!
  - Accuracy matters! Remember the Mars probe that crashed because NASA used inches and contractor used Metric?

Why T10000 tape?

Bit error rate matters!

Digital Repository Requirements

- Data set is essentially permanent
- Archive contents must stand on their own (no external databases required to know all about a file)
- Must be file format agnostic
- Must be scalable to very large size (EiB+)

Bit Error Rate Matters!

- When you get to the PiB level:
- 10^{-17} bit error rates is GiB of errors in your repository!
- T10k has best current error rate: 10^{-19}
  – IBM HPSS system also achieves 10^{-19}
- All other storage: currently the best is 10^{-17}
  – 2 orders of magnitude worse error rate!
- When you are migrating every 5-10 years your entire library, BIT ERROR RATE MATTERS!!!!

Issues

- Most commercial IT equipment has bit error rates of 10^{-14}, including Ethernet backbone equipment: what good is storage BER of 10^{-17} when your system's best BER is 10^{-14}
- How often to check data integrity?
  - Continuous above a certain size
  - Reading the data can also damage it!
- How often to migrate?
  - Individual files: every 5-10 years
  - Update the metadata when you migrate

What About….?

- Optical disc formats: none are considered archival
  - Non-recoverable errors in large % of discs at 5 years!
- Spinning disc?
  - Not scalable: Exabyte size would be physically huge
  - Power requirements:
    - Large if you keep the disks spinning constantly
    - 30%-60% less if you keep them at low power when not using them
  - Not considered archival!
    - EVERY SINGLE DISC will be replaced in under 7 years

What About….?

- Solid state?
    - Not able to scale to large sizes affordably
    - Worth watching
    - Solid state historically has not aged well

Cryptographic Checksums

- Numbers that tell us what the original size of our file was down to the bit
- LOC currently uses SHA-1
  - 160 bit checksum which can show a bit-flip up to file sizes of 2^{61} bits.
  - Other checksums:
    - MD5, SHA-2 (256 bits), SHA-3 (512 bits)

When to Check Your Repository?

- Up to you:
  - Cost in personnel, time, machinery
  - Chance of damaging the tapes just by reading them
- LOC/MBRS:
  - We’ve already checked the entire archive once
    - No bit flips in over a PiB of stored material
  - Probably every 12-24 months

Future Challenges?

Future Challenges

- Finding enough equipment to keep the migrations going
  - Hey buddy, can you spare a piece of equipment?
- Growing the Digital Repository into the exabyte realm…and beyond?
  - Ettabyte….yettabyte….then what….?
- Developing the knowledge and training needed to make sure the employees working on your project are adequately trained with proper documentation
- Getting the most bang for the bucks spent
- Finding the bucks to spend

Future Challenges

- Dealing with how our collection continues to age
- Studying how our digital collection's physical storage ages

Come See Us in Culpeper!

- Movies shown free to the public Thu, Fri & Sat every week
- www.loc.gov, search for “packard campus” or “NAVCC”
- Call before you come!

Suggested Resources

- **SMPTE: Society of Motion Picture & Television Engineers**
  - **www.smpte.org**
- **ATSC: Advanced Television Systems Committee**
  - **www.atsc.org**
- **DVB: Digital Video Broadcasting**
  - **www.dvb.org**
- **NAB: National Association of Broadcasters**
  - **www.nab.org**
- **AES: Audio Engineering Society**
  - **www.aes.org**
- **IEEE: Institute of Electrical & Electronics Engineers**
  - **www.ieee.org**
- AMWA: Advanced Media Workflow Association
  - www.amwa.org
- AMIA: Association of Moving Image Archivists
  - www.amianet.org
- IASA: International Association of Sound & Audiovisual Archives
  - www.iasa-web.org

Suggested Resources

- JPEG2000 standard (ISO 15444)
  - www.jpeg.org/jpeg2000
  - www.iso.org
- Dolby Laboratories website:
  Extensive documentation on Dolby AC-3, Dolby E & multichannel implementation issue; also provides regular DTV audio related email
  - www.dolby.com

Suggested Resources

- Manufacturers White Papers & books
  - Find a copy of Nvision's "The Book" & "The Book II" paperbacks on digital audio
  - Snell & Wilcox books (www.snell.com)
  - Tektronix (go to www.tektronix.com)
    - "A Guide to Standard & High Definition Digital Video Measurements"
    - "PAL Measurements", "NTSC Measurements"
- Suggested reading on digital:
  - "Broadcast Engineering Tutorial for Non-Engineers, 3rd Edition" - NAB
  - "Digital Video and HDTV" – Charles Poynton
  - "How Video Works, Second Edition: From Analog to High Definition" - Weise
  - Publishers: Tab/McGraw-Hill & Focus Books
  - www.nabstore.com for many titles; www.amazon.com has many more

Acknowledgements

- Harris Broadcast Communications
- Harris/PBS DTV Express project
- Graham Jones – DTV Express & the NAB
- Michel Proulx – Miranda
- Cody Claxton – DTV Express
- Gerry Field - WGBH/NCAM
- Mike Isnardi & Sarnoff Corporation
- KHOU, KNPB, KTVU, WCVB, WRAL, Paramount Television, the Harris/PBS DTV Express for video and audio materials

jsny@loc.gov

Thank You!

James Snyder
Senior Systems Administrator
National Audio Visual Conservation Center
Culpeper, Virginia
jsny@loc.gov 202-707-7097