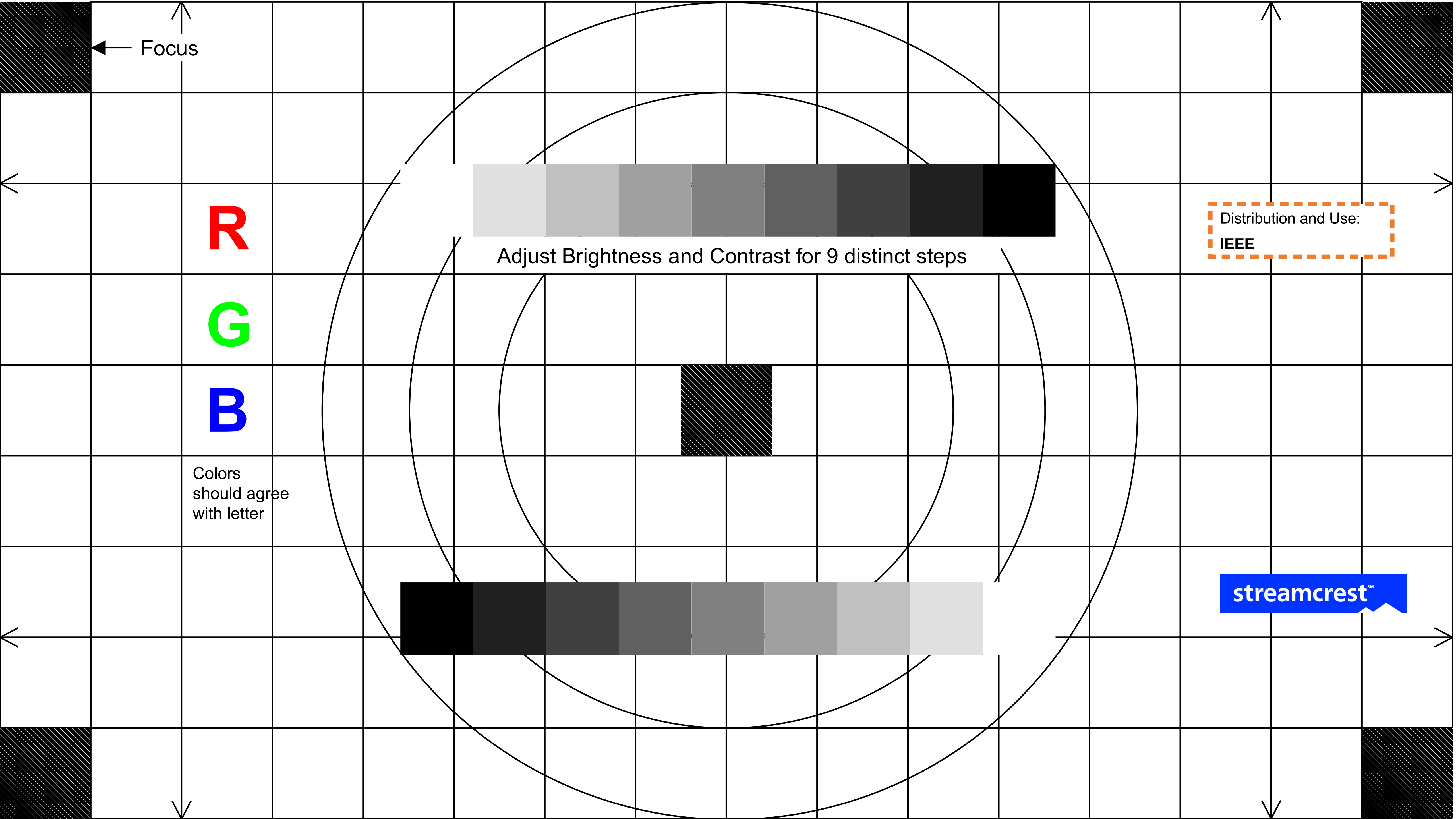

Focus
R
G
B
Colors should agree with letter
Adjust Brightness and Contrast for 9 distinct steps
Distribution and Use:
IEEE
streamcrest℠

From Studio to Sofa:
The Standards Behind Television and Video

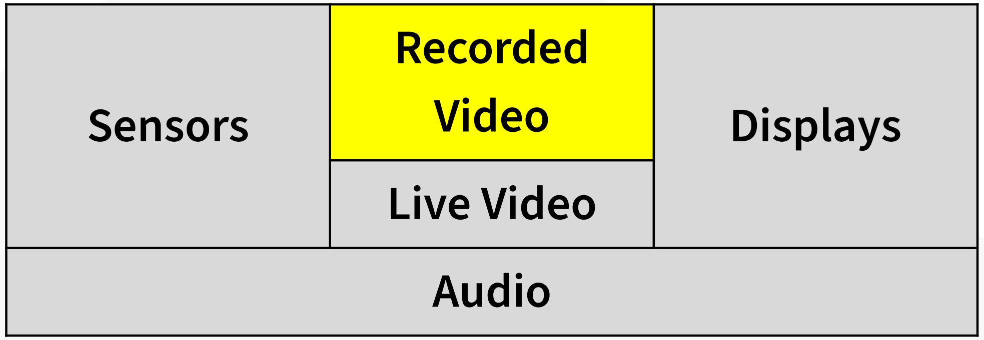
Robert Bleidt – July 7, 2025

Streamcrest Associates

Invited Presentation for Alabama IEEE Section

What we will not cover

- **Sensors, Displays, and Live Production**
  - **Too much for one presentation**
  - **Today we focus on Recorded Video – such as a TV drama episode**
- **Given the scope of the material and that some topics are simplified, some details may be wrong**
- **Slides have a lot of background detail that you can “take home in a doggie bag”**
- **Given enough time, I could present “glass to glass” and mouth to ear**
- **Ask me about dialogue intelligibility…**



Cinema is a separate subject, though today it’s very similar to 4K TV

Some Background to Get Started

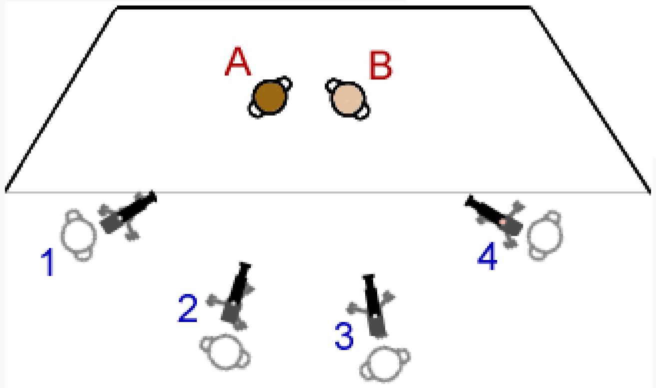
Trends in the Video Entertainment Industry

- **Several years ago we reached the viewing time crossover point between old media (Cable, Satellite, and Broadcast) and new streaming media consumption (Subscription, Free Ad-Supported, and User-Generated Content)**
  - **Example Services:**
    - **Subscription: Netflix, Max, Disney+, YouTubeTV, Amazon Prime Video**
    - **FAST: Roku Channel, Pluto, Samsung TV**
    - **UGC: YouTube, Instagram, TikTok**
- **Old media continues to be disrupted, and new media is still trying to match its technical performance.**
  - **For example, recent issues in carrying live events on streaming services.**
- **The number of services and amount of content produced has grown, but the industry - not really**
  - **Content may be produced outside Hollywood and more quickly**

Production Styles

- “Film Style” productions (features, episodic dramas) are usually shot single camera – actors re-do the scene for different viewpoints. Sound is recorded separately (“double system”) and matched to picture in editing.
- “Video Style” productions (Soap Operas, Game Shows) use multiple cameras to capture a single take including sound. Production is faster but there are often lighting and camera angle compromises.
- Except for special projects (i.e. IMAX 70mm), there is no use of actual photographic film today.
- Film and video worlds have their own vernacular, jobs, and (less so today) hardware.

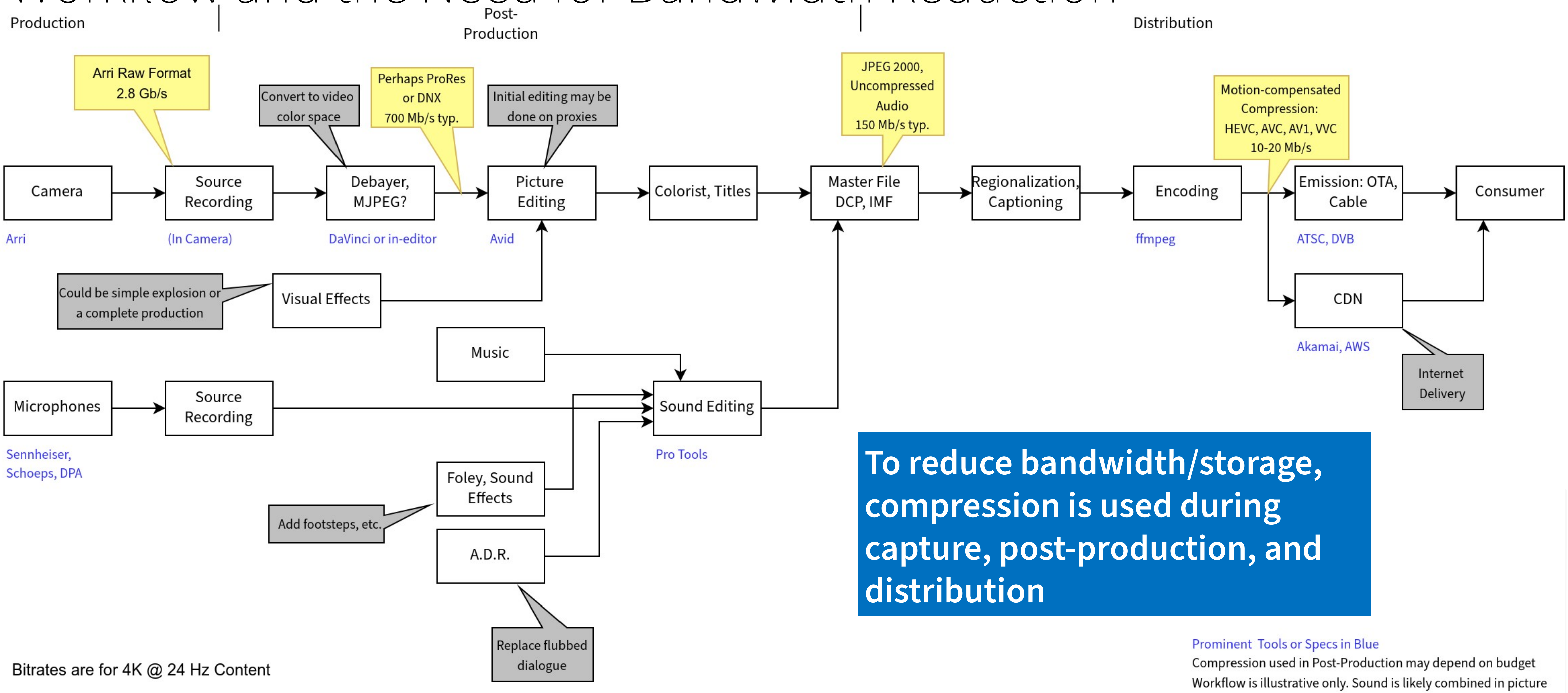
Most of this presentation is about “film style” while live production is always “video style”



Four Camera Production
(3 is more common)

wikipedia

Workflow and the Need for Bandwidth Reduction



To reduce bandwidth/storage, compression is used during capture, post-production, and distribution

Prominent Tools or Specs in Blue

Compression used in Post-Production may depend on budget

Workflow is illustrative only. Sound is likely combined in picture editor, for example.

Who sets standards for TV?

- **Over-the-air broadcasting:**
  - ATSC (Advanced Television Systems Committee) U.S.
  - DVB (Digital Video Broadcasting) EU
  - ARIB (Assoc. of Radio Industries and Businesses) Asia except China
  - (These are regional standards adopted by government regulators such as the FCC)
- **Internet Delivery**
  - IETF (Internet Engineering Task Force)
- **In the home:**
  - HDMI (High Definition Multimedia Interface)
  - CTA (Consumer Technology Association)
- **Professional studio and distribution**
  - SMPTE (Society of Motion Picture and Television Engineers)
  - AES (Audio Engineering Society)
  - ITU (International Telecommunications Union)
- **Compression**
  - ISO/IEC groups:
    - MPEG (Moving Pictures Expert Group)
    - JPEG (Joint Photographic Experts Group)
  - Open Source / IETF
- **De-facto standards**
  - Apple, Netflix specs

Evolution of the Industry and Terms to Learn About

	1980’s	2000’s	Today
TV Raster Size (pixels), TV Frame Rate (Hz)	720x483, 60i	1920x1080, 60i 1280 x 720, 60p	3840x2160, 24p, 30p or 60p
Dynamic Range, TV Input	~40 dB	8 bits	10 bits + HDR
Aspect Ratio	4:3	16:9	16:9
Display Color Primaries	NTSC	Rec. 709	DCI-P3, Rec. 2020
Studio Distribution	Analog	HD-SDI	ST 2110, 12G SDI
Emission Standard	Analog	MPEG Transport Stream	DASH, HLS, HTTP
Consumer Channels	OTA, Cable	Cable, Satellite	Internet, Cable
Technology Driver	Professional	Consumer Electronics	Consumer Electronics

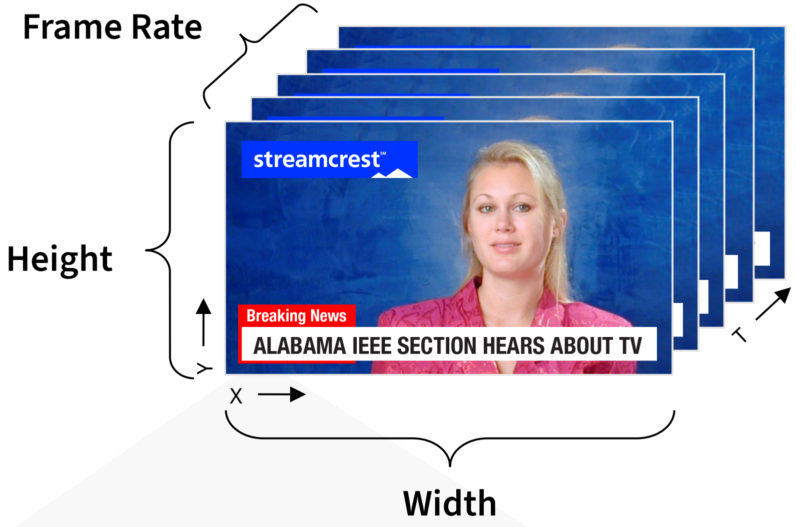
This is just a U.S. region comparative illustration, many details have been over-simplified or omitted

Evolution towards increased realism limited by distribution bandwidth
Distribution technologies influenced by business relationships and content rights

Visual and Video Science

Raster Sizes and Frame Rates[1]

- **TV is presented as a set of Raster Size (width x height) images that change at a Frame Rate (frames per second).**
  - The width /height = aspect ratio
  - Each "dot" in the image is a pixel
- **In the analog TV era, rasters of 768 x 483 or 640 x 480 were often used. Frame rate was 59.94 Hz with two interlaced fields or 29.97Hz. The aspect ratio was 4:3**
- **Film has always been 24 Hz. Gaming may use frame rates above 60Hz.**
- **For historical reasons, frame rates in NTSC countries are usually offset from nominal values by 1000/1001.**
- **ATSC 1.0 (U.S. Digital TV) introduced 1080i and 720p and 16:9 aspect ratio. ATSC 3.0 introduced 4K**
- **Streaming uses a variety of sizes from 360p to 4K**
- Professionals and cinema may use slightly different sizes. Cinema can also have 1.85:1 or 2.39:1 aspect ratios.
- Film content may be converted to "video" rates by "3:2 pulldown" or other ways
- 8K TVs are available, but content is only in Japan (some demos on YouTube[2])
- [1] NTSC region (U.S.) data, exceptions exist [2] Based on media reports – I don't have 8K equipment



Resolution	"NTSC" Region – U.S.	"PAL" Region - EU
QCIF	176 x 144	176 x 144
CIF	352 x 288	352 x 288
VGA	640 x 480	640 x 480
D1	720 x 486	720 x 576 (4CIF)
720p	1280x720	
1080i	1920 x 1080	
"4K" UHD-1	3840 x 2160	
4K DCI	4096 x 2160	

Why my mother did not buy a 4K TV

- **Human eyes can typically resolve details of about 1/60 of a degree. This is about 0.035 inch at 10 feet – about the thickness of a credit card.**
- **So, to take advantage of more picture resolution, either the screen has to be bigger or you have to sit closer.**
- **You would need to sit three feet from an 85 inch 8K TV unless you have 20/10 vision**
- **Some recommendations suggest a horizontal viewing angle of 30 degrees to strike a balance between immersion and differences in perspective**
- **Some people argue we can see more (as do audiophiles who promote higher sample rates)**

	Viewing distance (feet) for 1 arc minute acuity				Viewing distance (feet) for 30° horizontal viewing angle
Screen Diagonal (Inches)	720p	1080p	"4K"	"8K"	
45	8.8	5.9	2.9	1.5	6.3
55	10.7	7.2	3.6	1.8	7.7
65	12.7	8.5	4.2	2.1	9.1
75	14.6	9.8	4.9	2.4	10.5
85	16.6	11.1	5.5	2.8	11.9

F-stops and the Range of Light 'Intensity'

- **Most non-phone cameras have an aperture – an adjustable ring that reduces the light through the lens to control exposure. This is calibrated in "f-stops" and each stop reduces the light by half. The iris / pupil of your eye does the same thing.**
- **These are typically needed since light from a scene can vary over a wide range:**
  - Outdoor noon sun: 100,000 lux of incident light
  - Candle at 1 foot: 10 lux (1 foot-candle in U.S. units)
  - Reflectivity might vary from 2% to 95%
  - Range of perhaps 19 stops
- **The iris of the eye is constantly changing to "adjust the exposure." At any given instant, it can perceive detail over about a 5 ½ stop dynamic range.**

Fun fact: the "brightness" of light emitted from a monitor is measured in "nits" (candela/m^2). Typical TV is about 200 nits. HDR TV about 1000 nits.

Reducing Dynamic Range and Saving Bits

- **Most scenes naturally have a smaller dynamic range due to "ambient light"**
  - **Reflections from walls, ceiling, and other objects cast light into shadows to make them brighter**
  - **Professionals will control dynamic range with lighting, diffusors, or reflectors**
- **In "Standard Dynamic Range" video, we limit the video to about a 6-7 stop range, which is what the eye can see detail in, and what the monitor can reproduce**
- **In the new "High Dynamic Range" formats, we limit less: reflections off glass or chrome will appear more natural**
- **How many bits per pixel do we need to reproduce this?**

seerveld.com

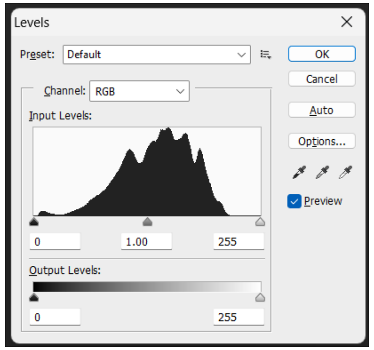
Film crews will rig diffusion fabric overhead to soften shadows

Chet Simmons

Neutral Density gel being used as window film to reduce outside light

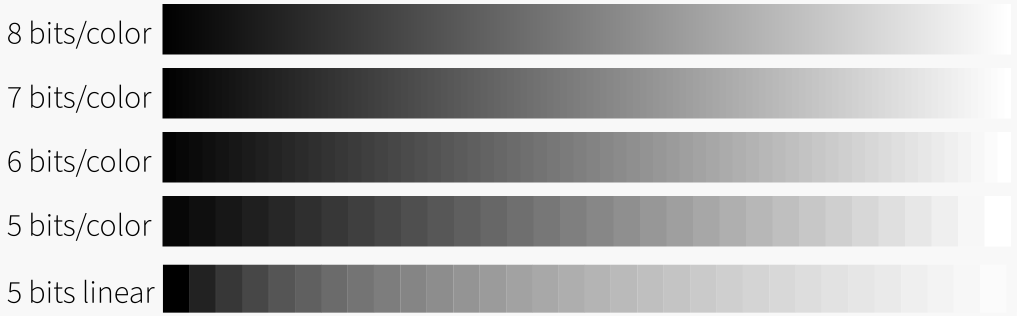
To save bits, TV does not transmit "linear light"

- **The eye has a logarithmic response: It can detect 1% increments over about a 100:1 brightness range. (About 5 ½ f-stops)**
- **To save bits, TV displays have a power-law relationship between input values and light output: gamma.**
  - **Light output = $(\text{input value})^{2.4}$**
  - **This is roughly "perceptually uniform steps" with each increase in the binary video value**
  - **This is the Electro-Optical Transfer Function[2]**
- **Cameras apply an inverse OETF so the system is linear[1] from glass to glass**
- **Computer graphics uses linear light since it needs to model reflections, etc.**
- **Other techniques for High Dynamic Range**



Fun fact: The mid-tone slider in Photoshop's Ctrl-L dialog adjusts gamma

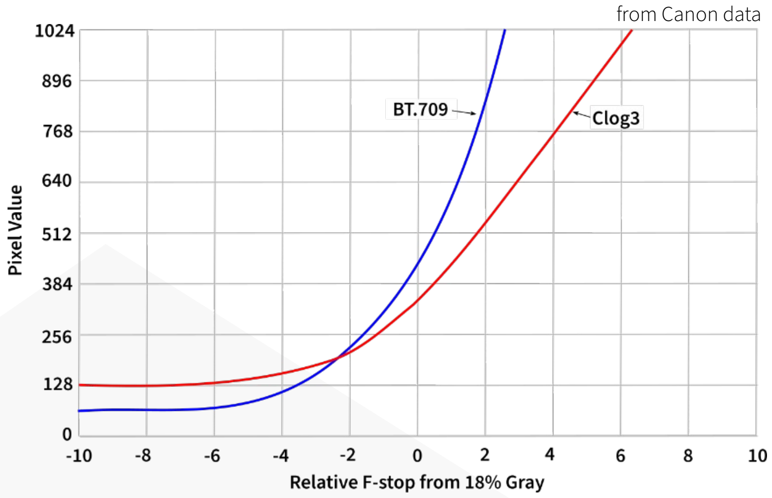
Gamma is used primarily to reduce the bits needed to represent video without contouring or banding. 14 bits of linear light is reduced to 8 or 9 bits. If every pixel is floating point, you don't need it.



[1]This may be altered for creative effect, and we limit shadows and highlights. [2] ITU BT.709 or BT.2020 specify gamma function.

Professional recorded capture today is often “log” format

- **Allows variations in exposure and color to be corrected during post-production**
- **Digital cameras record “linear light” with a single sensor that has a Bayer color filter in front of each sensing element. These are processed to form RGB pixels.**
- **Pixels are then either:**
  - Recorded in RAW format from the sensor for later processing
  - Converted to logarithmic values to reduce storage (commonly professional Apple ProRes or Avid DNX format) for later color correction and editing
  - Converted to the final dynamic range as gamma-corrected values (consumer MPEG formats) if no color or exposure correction is anticipated

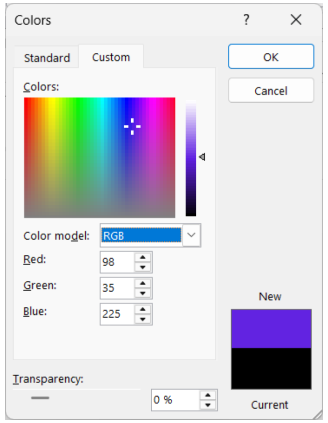


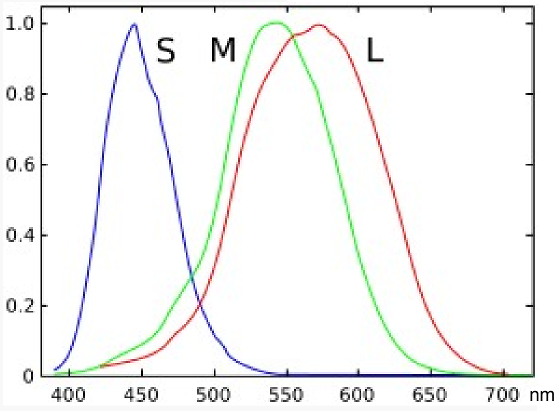
Fun fact: “Brightness” on your TV or monitor is an offset (setup or black level in TV-speak) while “Contrast” is a gain. Both are acting on the overall luminance.

Interstitial: Pantone Book, color checker

How do we transmit color?

- Color is fundamentally a sense of the wavelength of light, about 400-700 nm. You can hear two musical notes at once, but see only one color – vision is different.
- Combinations of red, green, and blue light combine to form a perceived color. You’ve seen this in the dialogs of Word and Powerpoint, etc. Paint and ink don’t work this way and CYMK is used.
- In TV, we transmit signals of light detected through three Color Primary optical filters: red, green, and blue. Three similar wavelengths are generated by the display based on the RGB signals.
- The eye is less sensitive to detail in color compared to luminance (the monochrome part of the scene, denoted Y), so color signals are often sub-sampled (4:2:2 or 4:2:0) to lower sampling frequencies. A linear transformation of the R,G,B signals to Y ,R-Y, and B-Y signals is done to allow this. For this, the combination of R,G,B that makes white must be defined.



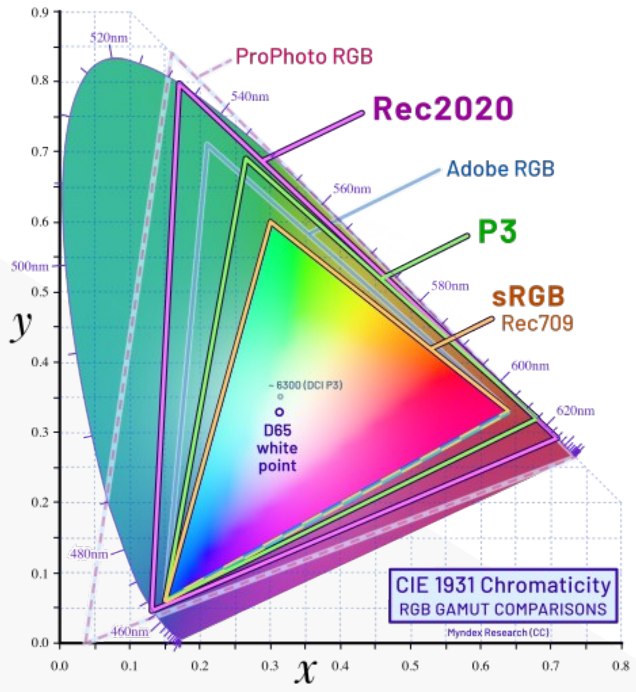


Normalized sensitivity of cones in the eye

How do we transmit color? - 2

- **Psychologists developed a uniform mapping of all perceived colors, the CIE x-y value. Pure wavelengths of light fall on an arc in this space. Signals from the three Color Primaries can produce any perceived color in a triangle defined by the Primary colors.**
- **Primary colors were originally limited by available CRT display phosphors. The ITU-R BT.2020 color space is much larger – the Primary colors are pure wavelengths. Today's quantum dot displays can come close to this space.**

Wikipedia, Andrew Somers



Plot of range of color primaries in popular standards. Ordinary computer monitors typically use sRGB.

Visual and Video Science Summary

- **TV uses several techniques to reduce the data to be stored or transmitted:**
  - **Raster size appropriate to resolution of the eye**
  - **Gamma function to map bit steps to eye's amplitude response**
  - **Three R,G,B color channels that allow the eye to perceive color**
    - **Often a transformation to Y, R-Y, B-Y signals where Y is the monochrome info. and the other two components are half bandwidth**
- **For professional use, our edited output is perhaps 500-700 Mb/s**

Video Compression

Professional bitrates are too big for consumer transmission

- **How do we fix that? Video (and audio) compression.**
- **I have offered the analogy "Compression is the Instant Coffee of Media"**
- **Encoders "freeze dry" the signal by removing information you won't notice and cleverly packing the rest**
- **The resulting compressed file is easy to store and transmit**
- **Decoders "add water" by undoing the packed format and reconstructing the picture**
- **Just like instant coffee, it's close enough to the real thing some of the time**

- Data Compression Techniques
  - PK Zip, Huffman Coding
  - Arithmetic Coding
- Frequency Domain Transforms
  - JPEG
  - Motion JPEG, Apple ProRes
- Motion-Compensated Transforms
  - MPEG-1,-2,-4, AVC,HEVC, VVC
  - VP9, AV1

How does compression work?

- **By removing redundancies in the representation of images**
  - **Data Compression Techniques**
- **By removing information (mostly) beyond human perception**
  - **Frequency Domain Techniques**
    - **(or wavelet domain in the case of DCP or IMF)**

Compression for a given perceived quality is always a trade-off between:

- **Coding Efficiency (bits/second)**
- **Computational Complexity (CPU cycles)**
- **Latency (delay)**

Data Compression

Original Data:
00 00 00 00 11 23 23
Run Length Coding:
04 00 01 11 02 23

Repeat Count (Run Length)

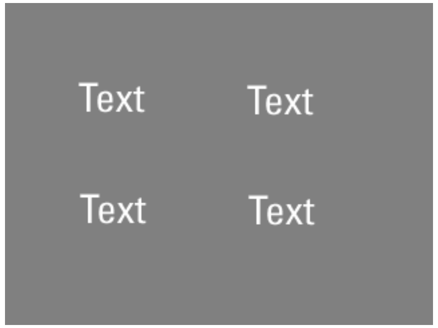
Value

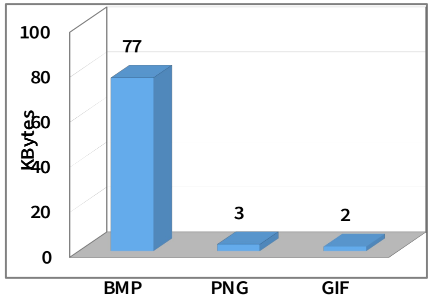
Typical Codes, English Letters:

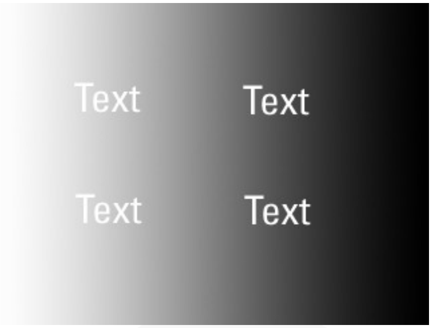
	ASCII	Huffman
E	01000101	100
A	01000001	0100
Z	01011010	1110111111

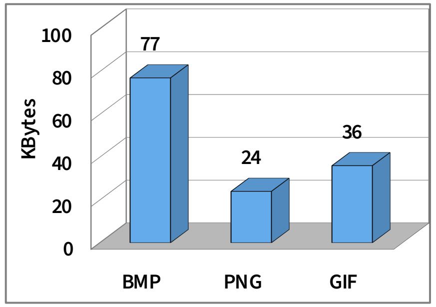
- **Instead of sending our data, we send a coded version that is shorter**
- **With Run Length Coding, we replace a repeated "run" of a value with a count and the value**
- **Huffman coding replaces a value with a variable-length code based on statistics. Popular values get a shorter code**
- **"ZIP" files use Limpel-Ziv and Huffman encoding**
- **All this is lossless – the decoded file matches the original – with about 50% reduction on audio or video files**

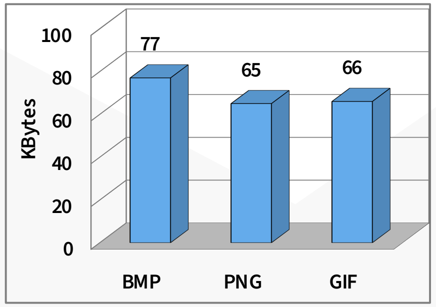
Applying Data Compression: PNG, GIF





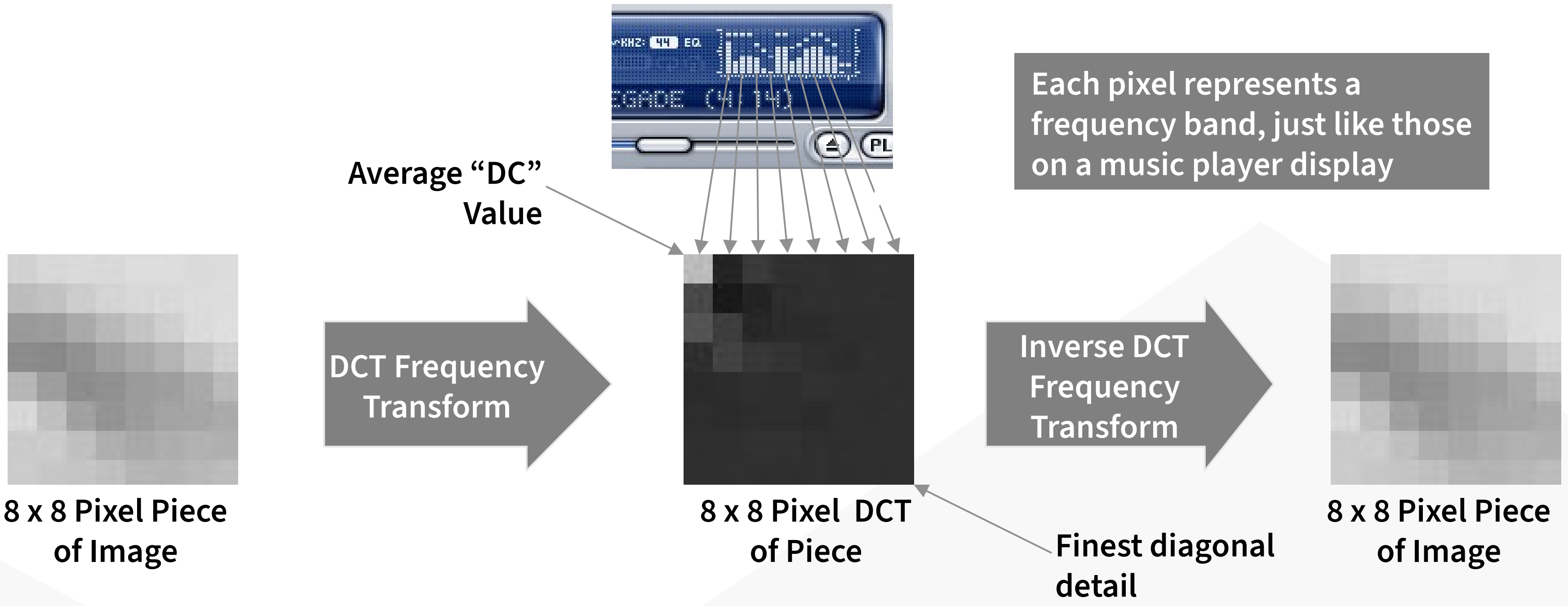






- **Simple images with many repeated pixels can be compressed 20-40x in PNG or GIF formats**
- **Photo or video images can’t be compressed much since they have few repeating patterns – a GIF of random noise is larger than the original**

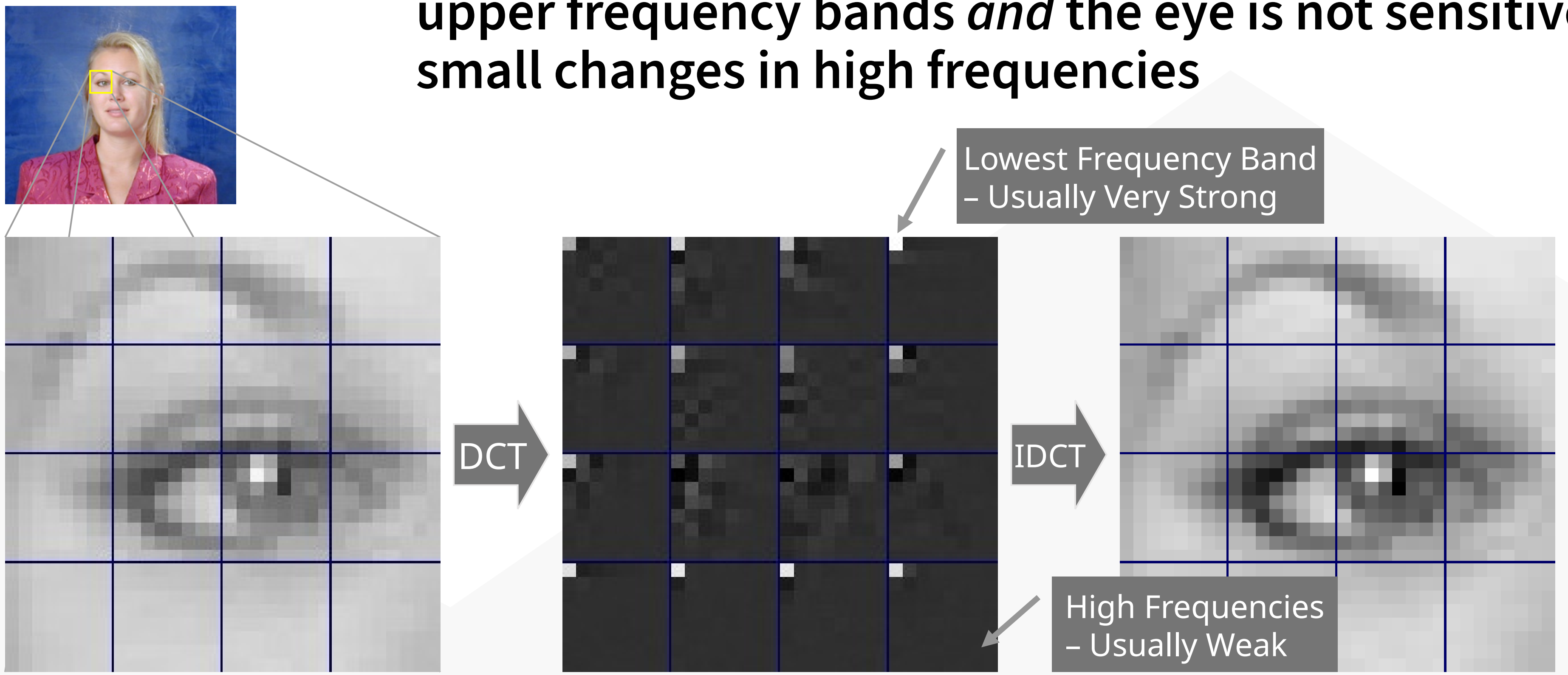
Transforming Images Into The Frequency Domain



- A mathematical transform, or formula, lets us convert a block of pixels into another block where pixels represent the strength of frequency bands in the image
- An inverse transform on this new block exactly re-creates the original image

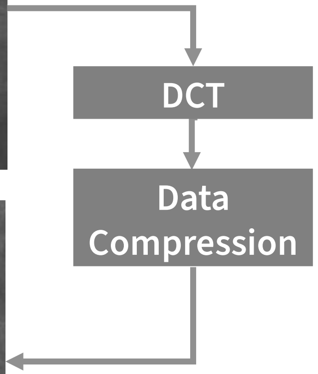
Frequency Domain Shifts Bits Where Some Can Be Quantized

- **There is no magic in using the frequency domain – it's just that *most* images have little amplitude in the upper frequency bands *and* the eye is not sensitive to small changes in high frequencies**

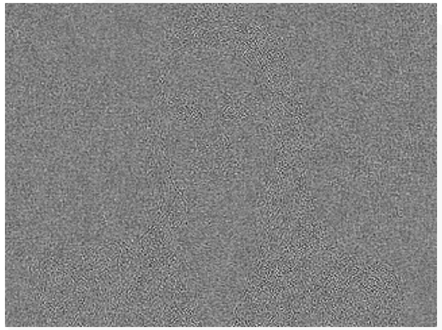


Application: JPEG

Uncompressed
BMP
77 KB (8.1 bpp)



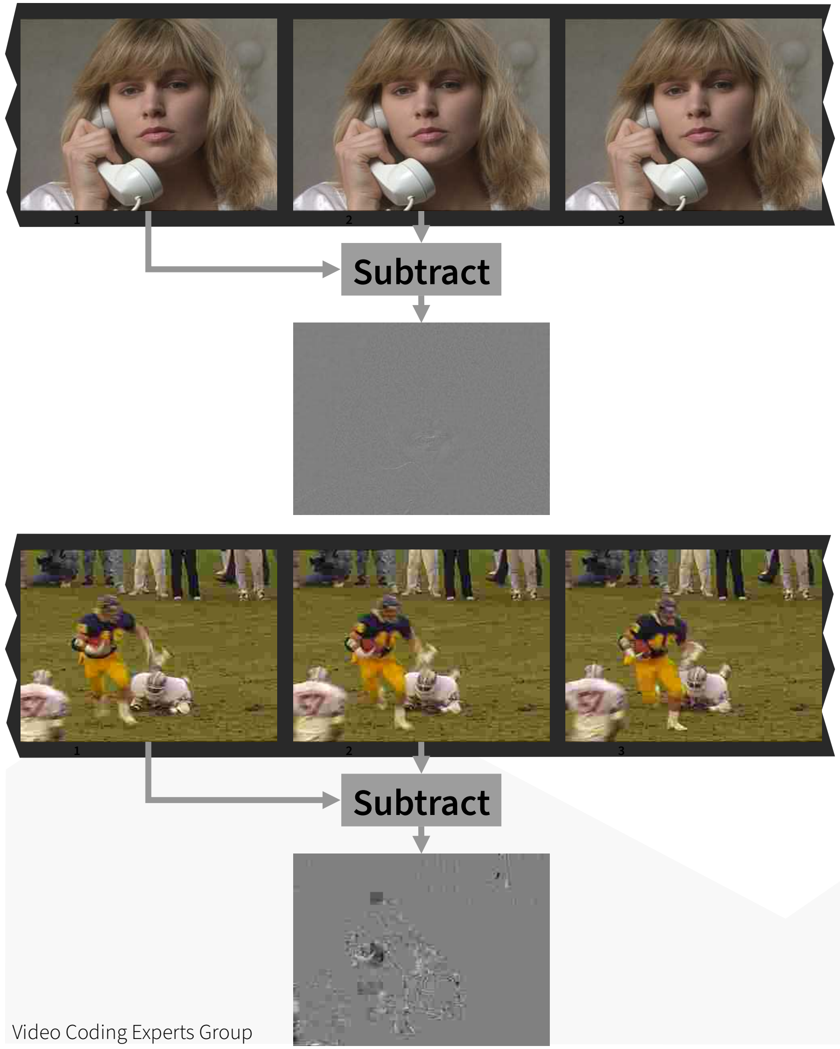
JPEG from
Photoshop Q=9
18 KB (1.9 bpp)



Errors multiplied
by 16

- The frequency transform doesn't reduce the number of pixels, but it does redistribute their values so that high frequencies usually have small values.
- We can use a lower number of bits to represent these high frequency pixels since they are small and the eye is not very sensitive to slight errors in their value. In some cases, we can just set them to zero.
- Data compression techniques can then compress these pixels very efficiently as there are then many repeated zero values and a smaller number of bits per pixel.
- This is the basics of the JPEG standard.

Frame Differences



Video Coding Experts Group

- **Most videos have little motion from one frame to the next**
- **We can subtract the previous frame from a frame and encode only the differences, which are usually quite small**
- **As a refinement, we can move a piece of the previous frame a few pixels before subtraction to reduce the resulting difference image (The amount of movement is sent with the encoded image)**

Frame Differences – Application: MPEG

Original Image (12.7KB JPEG, Q=5)

Previous Frame Subtracted from Image (6.3KB JPEG, Q=5)

- **The difference frame usually has few details and is easy to encode with JPEG-like frequency domain encoding.**
- **If we just JPEG the difference frame using Photoshop, it is about ½ the size of a JPEG of the original.**
- **A combination of other minor tricks, plus a reduced "JPEG" quality setting makes frame difference coding have an efficiency of about 0.5 to 0.05 bits/pixel.**
- **Fundamentally, this is how all common video codecs (such as MPEG) work today.**
- **They differ in refinements (i.e. motion search, block size) with varying computational complexity.**

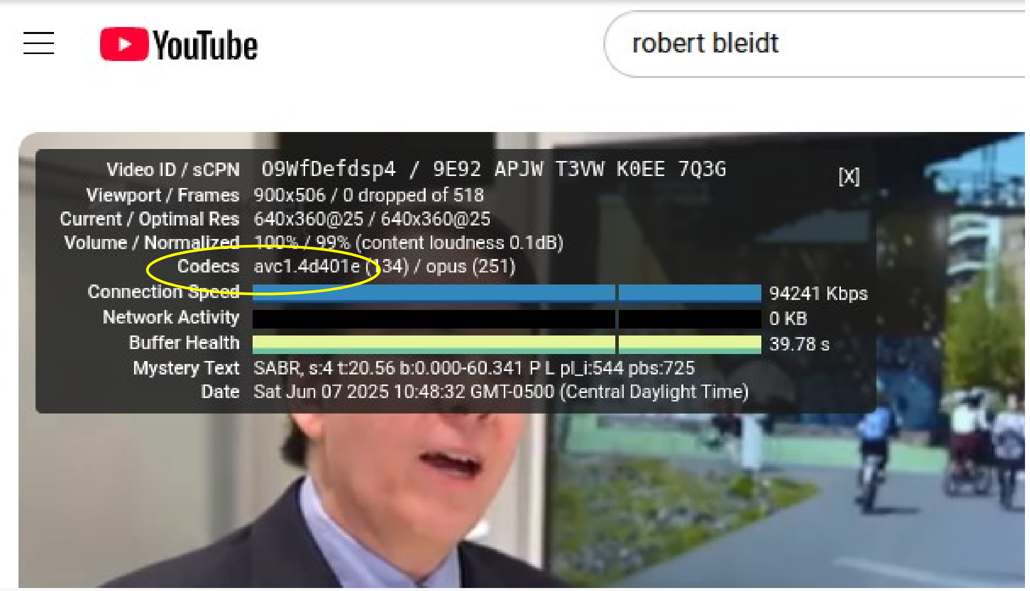
Timeline of Standardized Video Coding

	ITU – International Telecommunications Union Telephone, Radio, TV	**ISO –** International Standardization Organization Photography, Computer, Consumer Electronics	**Open Source –** Web and Internet - Sometimes IETF (Internet Engineering Task Force)	**Relative Bitrate** Streaming Media Magazine and MSU Estimate for 4K (could be wrong)
1990	H.261 – Video Conferencing			
1993		MPEG-1 – Video CD		
1994	(H.262)	MPEG-2 – Digital Cable and Satellite TV		
1995	H.263 – Improved Video Conferencing			
1997		ATSC – (MPEG-2) U.S. HDTV		
1999		MPEG-4		
2002	AVC (H.264)	AVC (MPEG-4 Part 10)		150
2013			VP9	100
2013	HEVC (H.265)	HEVC (H.265) ATSC 3.0 TV		100
2018			AV1	62
2020	VVC (H.266)	VVC (H.266) – Brazil HDTV		50

Future codecs will likely use AI technologies for better compression

Homework Experiment – You Tube Stats for Nerds

- **Right-click on a YouTube Video and open "Stats for Nerds" and you can see which codec YouTube is using for your connection for a video**
- **This can vary depending on operating system, hardware, and content.**



streamcrest℠

Video Compression Summary

- **Lossless "zip file" compression helps reduce the data rate a little**
- **To get stronger compression, we quantize the high frequency (fine detail) information more coarsely than the lower frequencies**
- **Instead of sending each picture, we encode the difference from the previous picture**
- **We may move segments of the previous picture to account for motion before subtracting to get the difference. We then send Motion Vectors along with the picture**

Transmission to the Consumer

- **In order to minimize memory and latency, a program's data has to be sent to the consumer just before the decoder needs it.**
- **Codecs can be Variable Bit Rate ("constant quality") or Constant Bit Rate (quality varies with scene encoding difficulty)**
- **Old broadcast approach is: playback is locked to timing information sent in the bitstream**
  - **Example: MPEG-2 Transport Stream – used for cable, satellite, antenna TV**
- **New unicast approach is: playback device requests small (10 seconds typ.) segments of program as it needs them**
  - **Example: MPEG DASH or Apple HLS streaming**
  - **Each request can be for a different bitrate in case of network congestion**

Future Developments

- **Current trend towards "virtual production" where green/blue screen shots are replaced by LED walls that can track with the camera. *The Mandalorian* was an early adopter.**
- **Today, "virtual reality" is only "three degrees of freedom" video. To walk around in VR requires 6 DoF, limited to computer graphics today. Volumetric capture sort-of enables live action video in 6 DoF.**
- **Lightfield capture senses the direction of light impinging on the camera sensor surface, not just its overall amplitude. This means focus and depth of field can be adjusted in post.**
- **Holographic displays deliver a real "holodeck" experience – at orders of magnitude more bandwidth.**

The End

- **Want the slides? Email rb720@streamcrest.com**