



AMIA 2022

AV Preservation & Validation Workflows

Presented by
Morgan Oscar Morel



NATIONAL
ENDOWMENT
FOR THE
HUMANITIES



What Is The Point of This Workshop?

- To discuss what AV Preservation is in a digital environment
- Give a brief overview of tools that can be helpful
- Provide some hands-on training with av preservation tools



BAVC Community-Based Preservation Training

- Remote preservation education program. Free for participants
- Provides hands on training with tape playback and monitoring equipment
- Email me for more info about participating!



NATIONAL
ENDOWMENT
FOR THE
HUMANITIES



NATIONAL
ENDOWMENT
FOR THE
HUMANITIES



NATIONAL
ENDOWMENT
FOR THE
HUMANITIES



NATIONAL
ENDOWMENT
FOR THE
HUMANITIES

Morgan Oscar Morel

- Director of Preservation at BAVC Media
- Located in Oakland, CA
- 10 years of experience working in AV Preservation labs and working with open source tools to preserve AV materials



Who Am I?



- Topics
 - AV Preservation Formats
 - FFV1 Validation
 - Station Qualification
- Tools
 - MediaInfo
 - MediaConch
 - FFmpeg
 - QCTools
 - SoX

**What
will be
covered**



- Lecture - Preservation Formats
- Exercise 1 (30 min)
- Exercise 2 (30 min)
- Break
- Short Lecture - FFV1/MKV
- Exercise 3 (30 min)
- Exercise 4 (30 min)
- Break
- Lecture - Audio Station Qualification
- Exercise 5 (30 min)
- Lecture - Video Station Qualification
- Exercise 6 (30 min)
- Final Q&A

Workshop Format

Prepping for the Exercises

Go to the **github repo** and **clone it**

<https://github.com/iamdamosuzuki/AV-Pres-Validation-Workflows/>

or

<https://bit.ly/3ykkn0Y>

● ● ● TERMINAL WINDOW

```
$ git clone  
https://github.com/iamdamosuzuki/AV-  
Pres-Validation-Workflows.git
```

```
$ cd AV-Pres-Validation-Workflows
```



What does Preservation Mean for AV?

Video tape must be reformatted!

Born digital video can be at risk of becoming unsupported if not reformatted

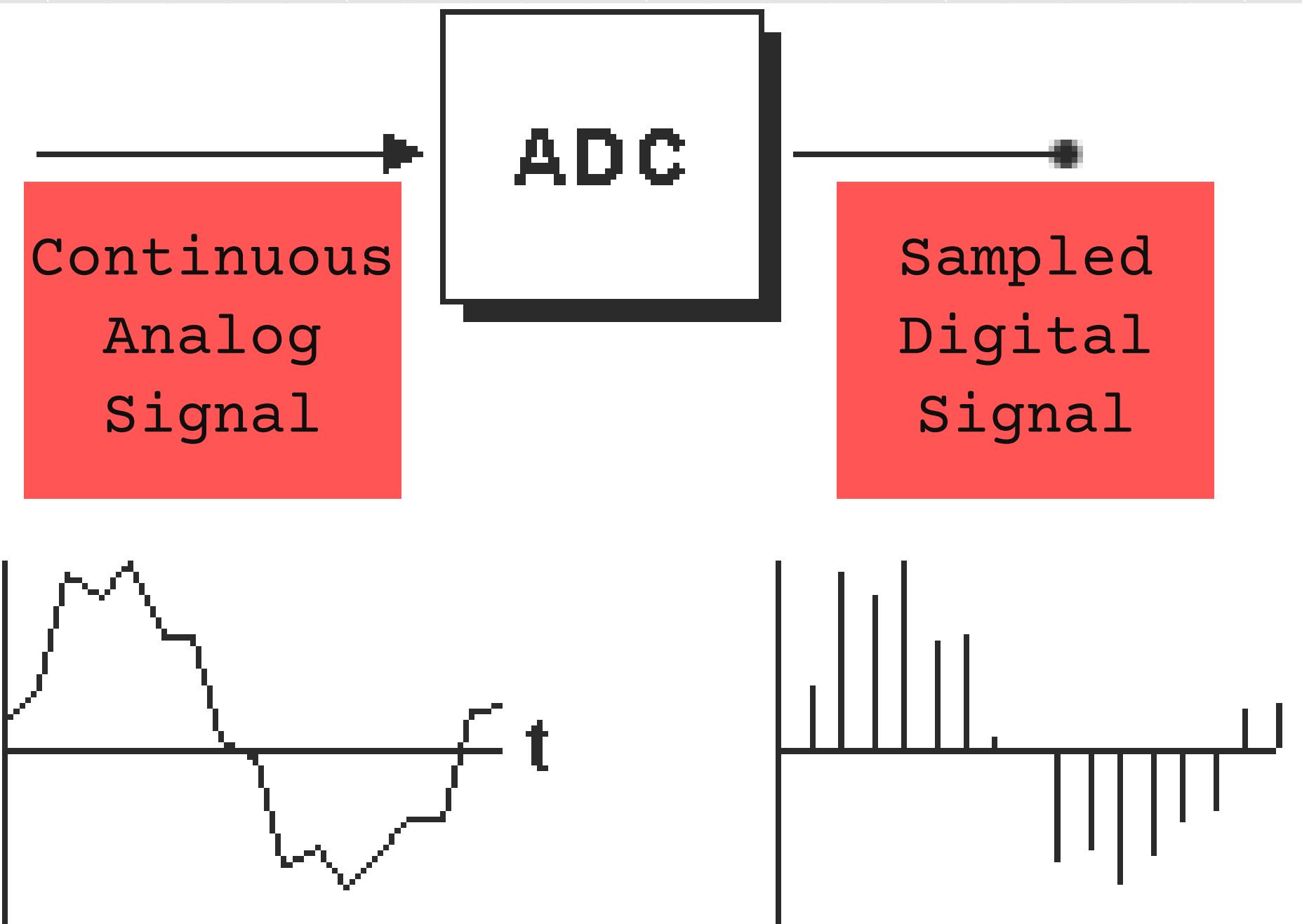
Standard Digital Preservation techniques are needed, plus extra work to deal with AV complexities



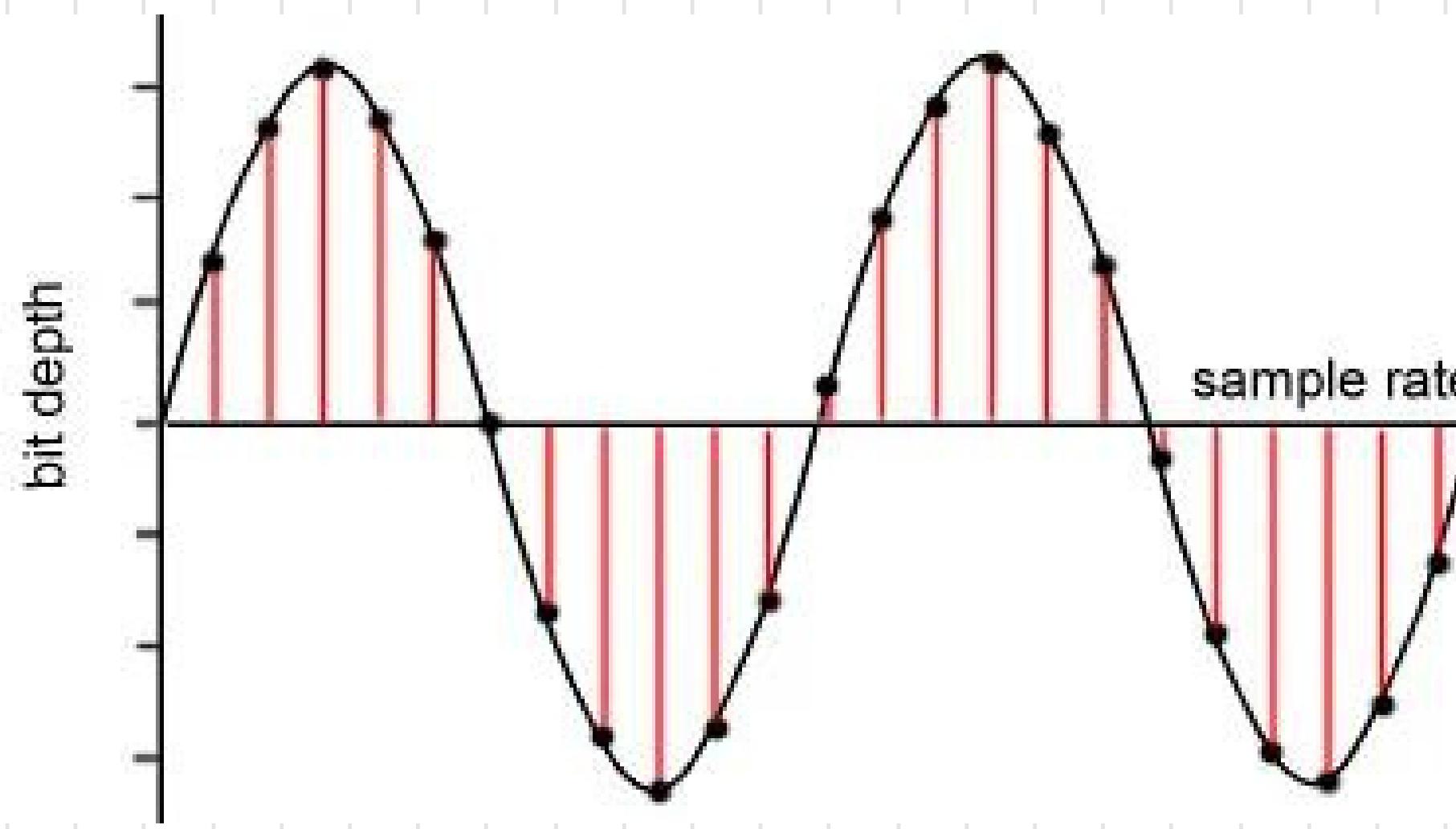
What Preservation Formats need to be able to do

- Store AV stream information and auxiliary data without loss
- Retain original organization of stream information
- Maintain fixity
- Retain functionality for playback and transcoding over long periods of time

Storing Data Digitally: Conversion



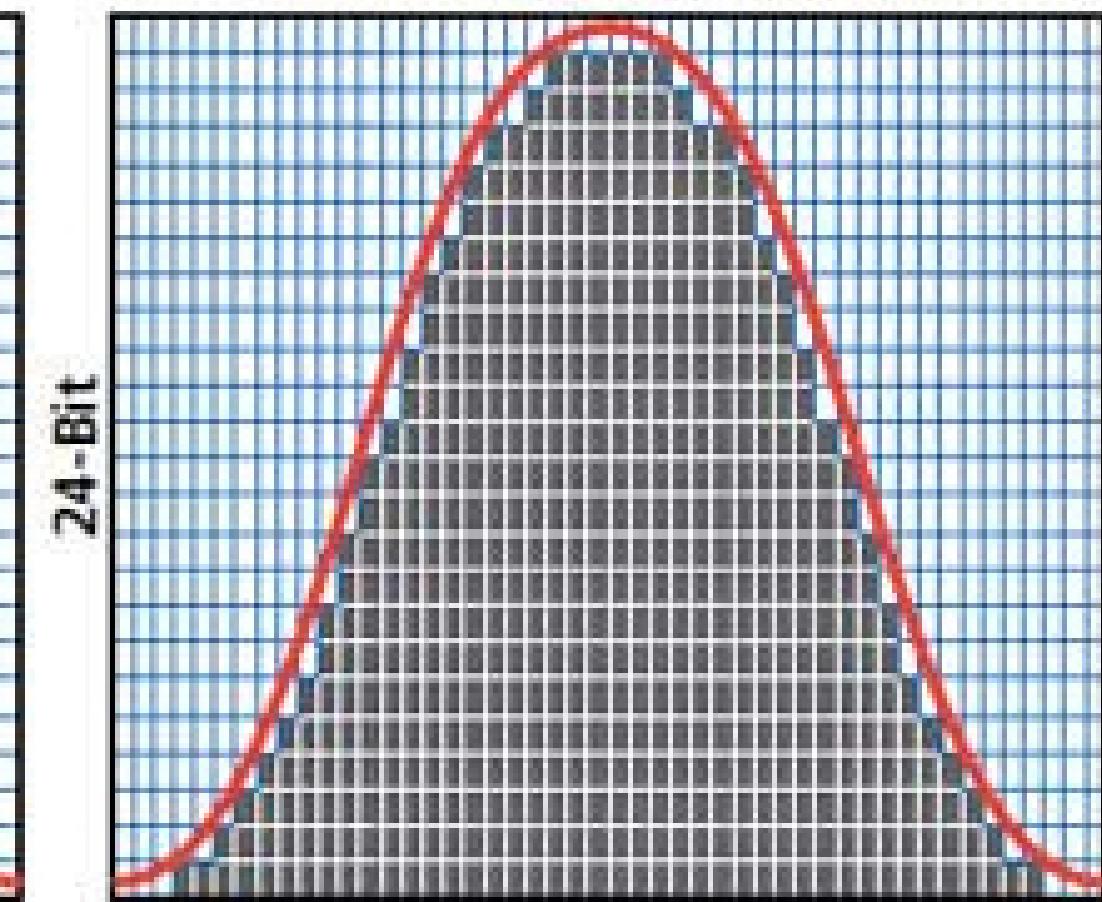
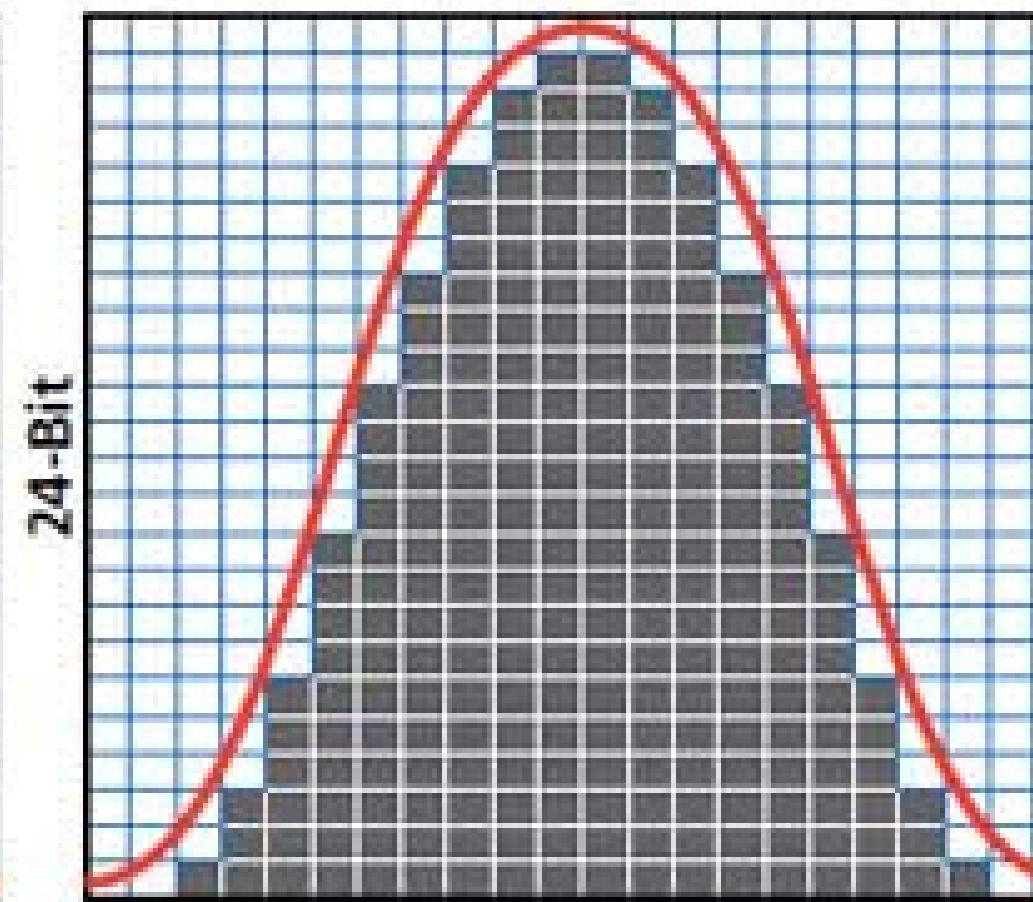
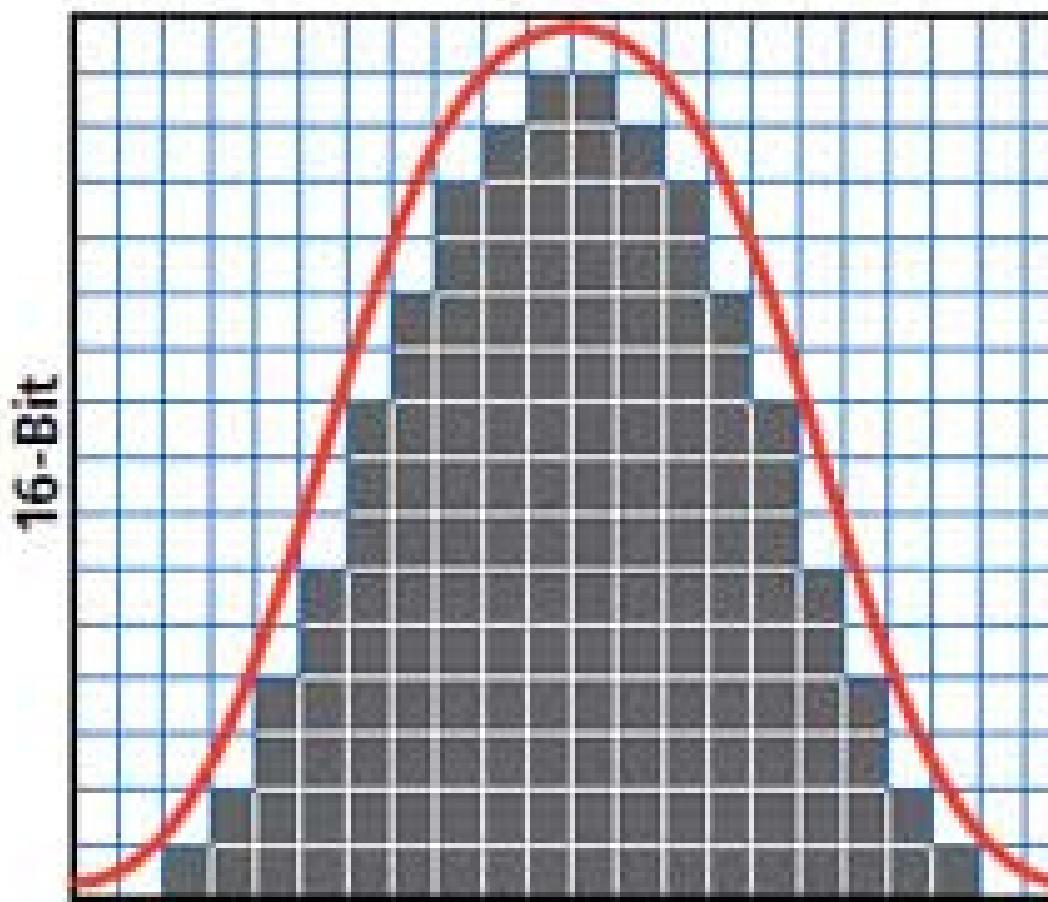
Storing Data Digitally: Conversion



Sample Rate: Time Axis Resolution

Bit Depth: Amplitude Axis Resolution

Storing Data Digitally: Conversion

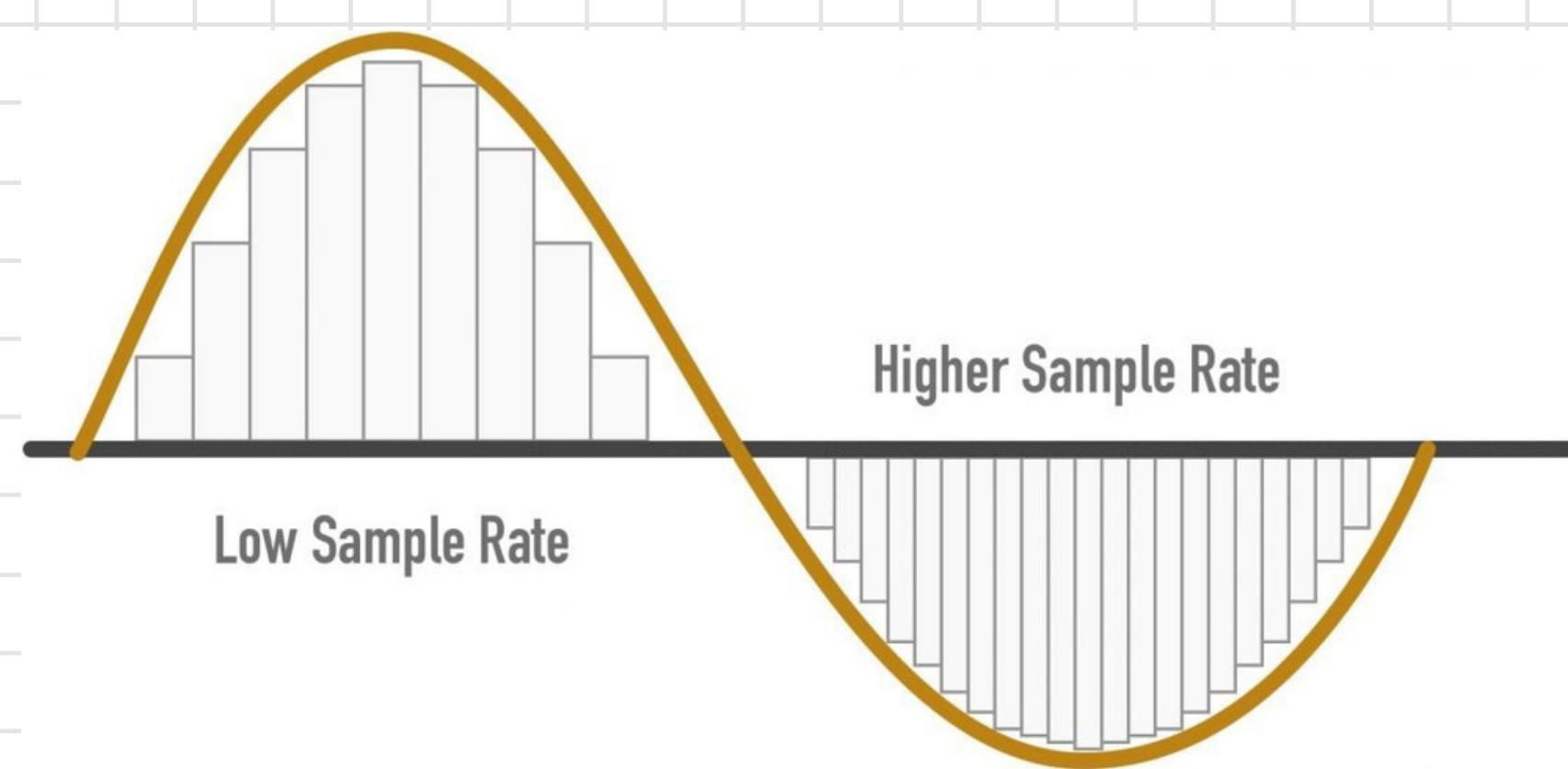


44kHz

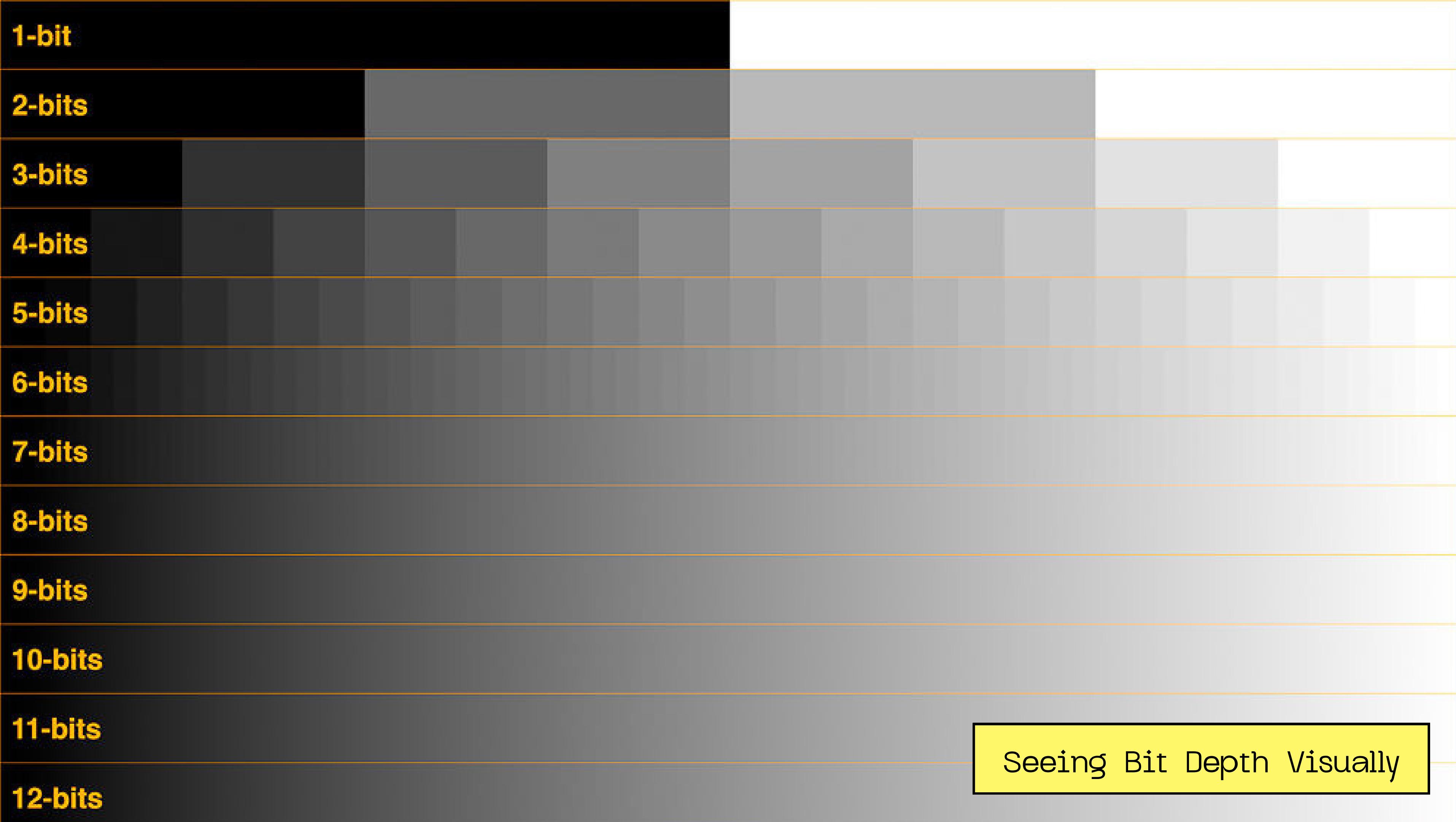
44kHz

96kHz

Sampling Rates



- The higher the sampling rate the higher frequency that can be properly represented by the digital file.
- The highest frequency that can be represented is HALF the sample rate. This is known as the nyquist frequency
- Human hearing caps out at around 20kHz. Need to sample at least 40kKz to capture the upper limits of human hearing



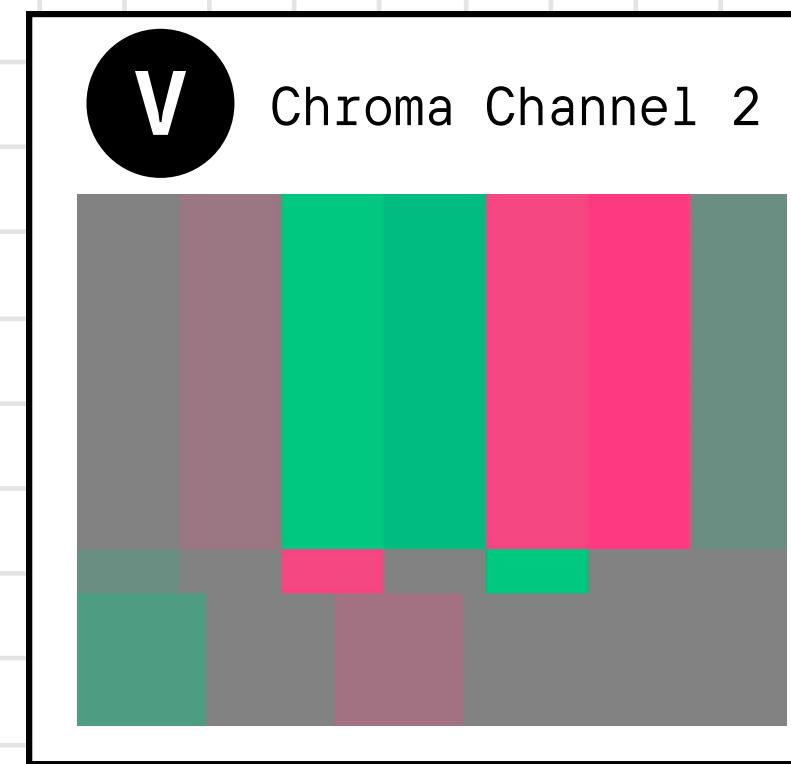
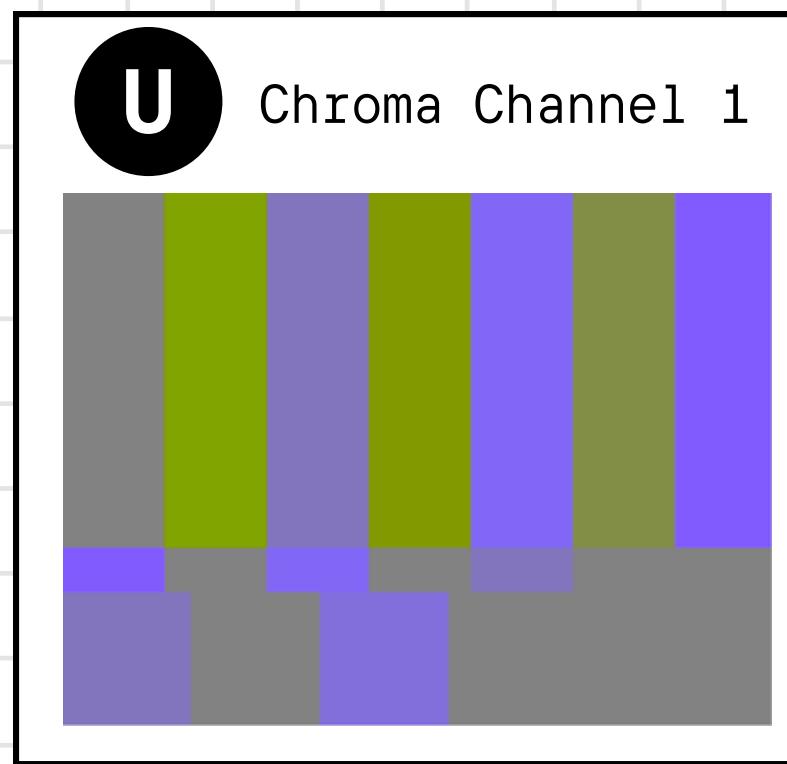
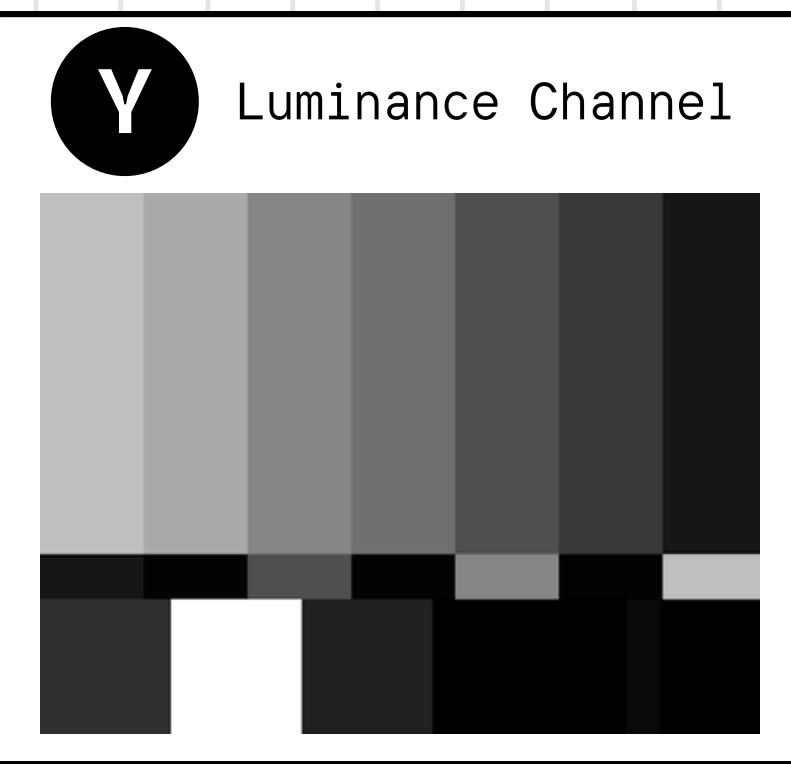
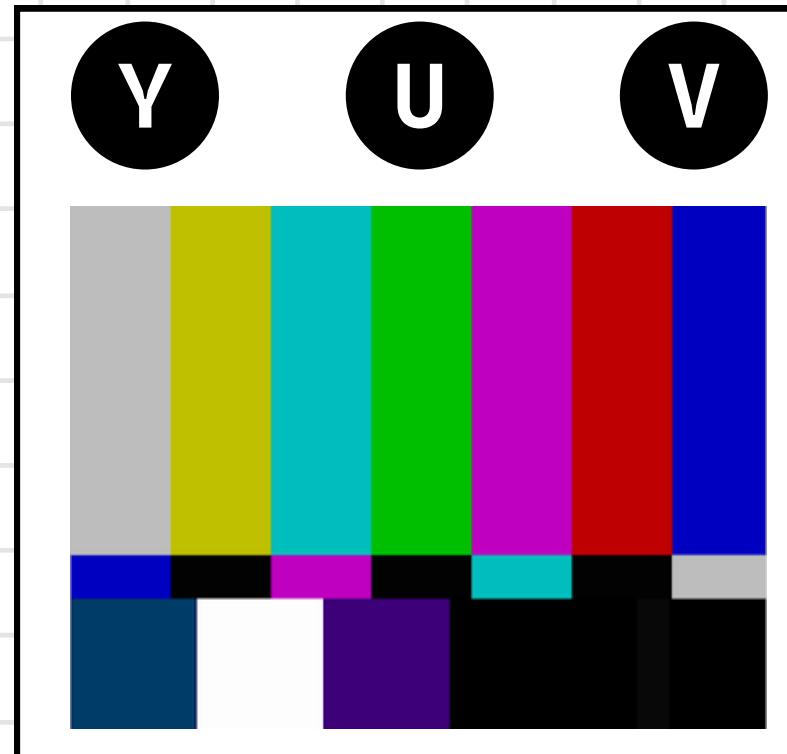
Seeing Bit Depth Visually

1-bit	0	1														
2-bits	00	01	10	11												
3-bits	000	001	010	011	100	101	110	111								
4-bits	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
5-bits																
6-bits																
7-bits																
8-bits																
9-bits																
10-bits																
11-bits																
12-bits																

Binary Values

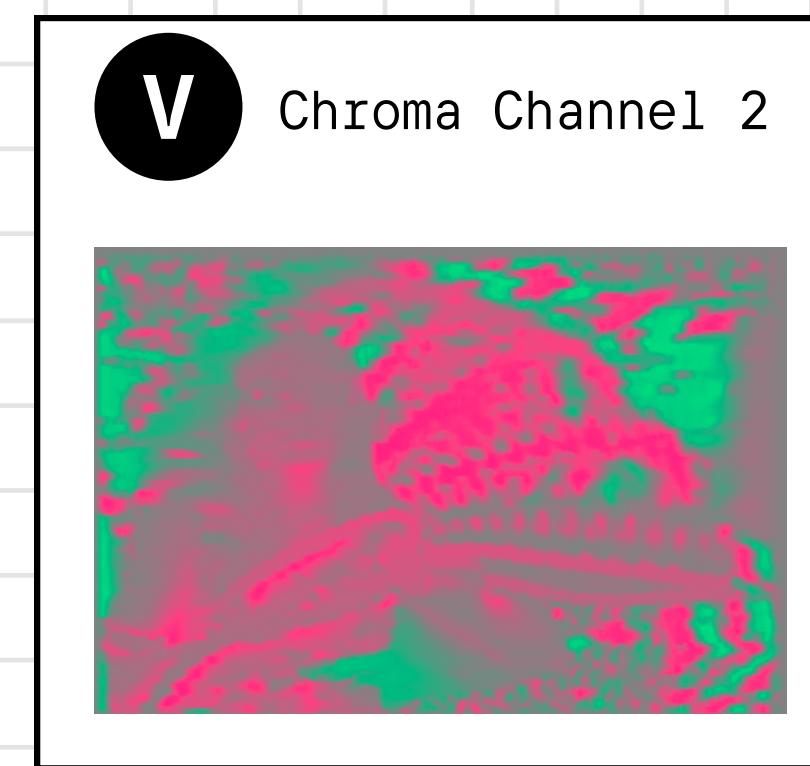
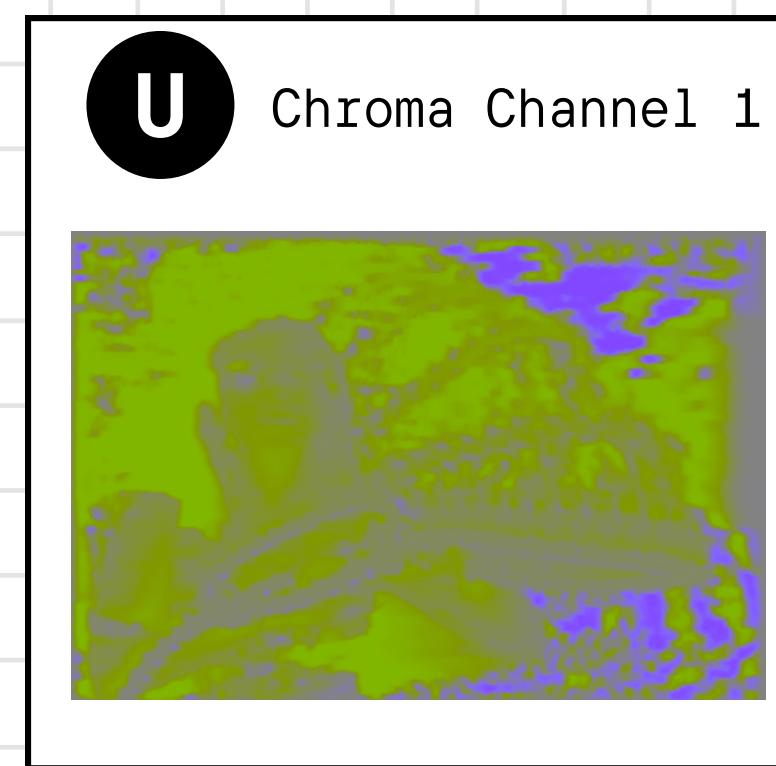
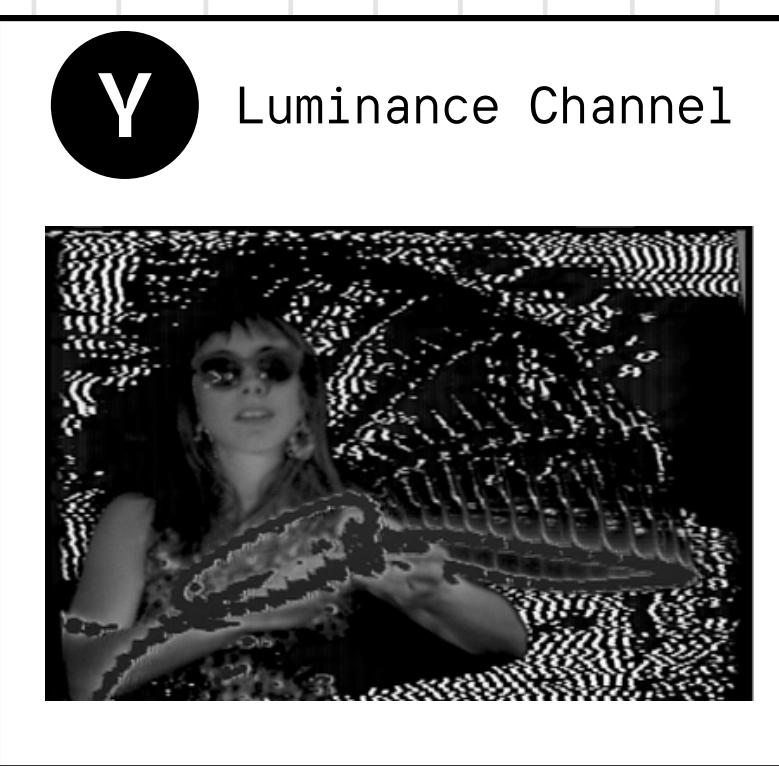
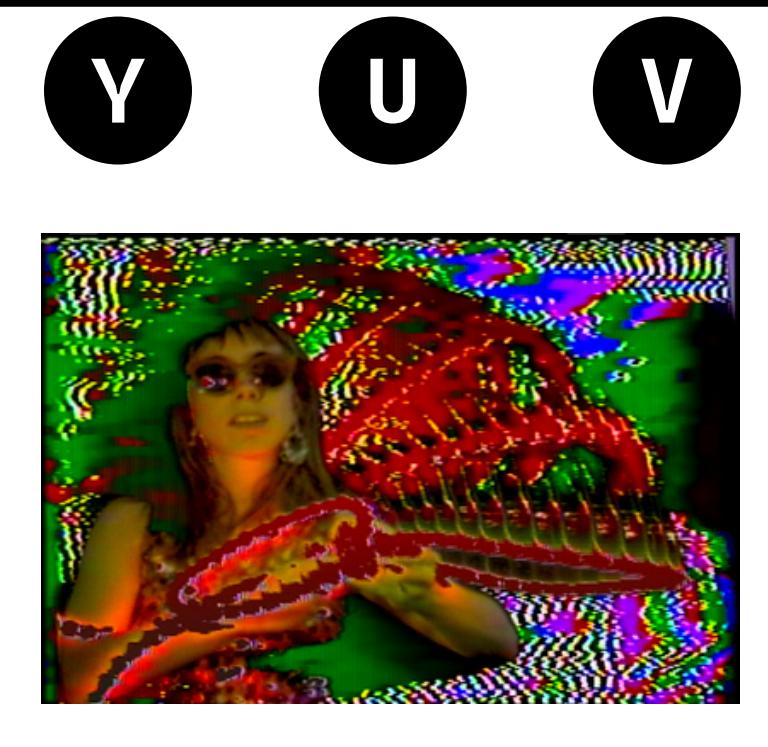


YUV Color Encoding





YUV Color Encoding

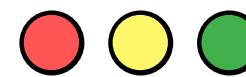


10-Bit Video Data Scope

449	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	
9197	916D	9162	917A	91BB	9206	9237	921D	9297	91F4	91DE	91E9	922F	926B	928A	92B6	92D9	92A5	923C	91C5	9141	91C8	9177	9181	9193	
249	91BB	91BB	91A9	91A9	9192	918E	918E	91A3	91A3	9190	9190	9184	9184	9196	9196	919B	919B	919D	919D	91A0	91A0	91A1	91A1	91A1	
	9245	9245	9245	925A	925A	926B	926B	9270	9270	9264	9264	9285	9285	9292	9292	9285	9285	9286	9286	9273	9273	9273	9273	9273	
	919B	9197	917C	9169	9189	91BD	91EA	9206	920D	9209	9212	9222	9207	91E8	9284	924F	9269	9262	9242	91DB	9138	9137	9137	9137	
250	91BB	91BB	91C2	91C2	91B2	91AE	91AE	919B	919B	9183	9183	9196	9196	91C6	91C6	91C2	91C2	91BC	91BC	917E	917E	918C	918C	917C	
	924A	924A	924B	924B	9249	9256	9256	9254	9254	926A	926A	928F	928F	9294	9294	9270	9270	926D	926D	9293	9293	9283	9283	9283	
	9191	9175	9172	9197	91EA	9226	9238	9210	91DB	91CA	91CE	91DF	91F1	91EE	9201	922D	9240	923D	9234	9215	91C2	9168	9139	914D	918
251	91BC	91BC	91C0	91C0	91CA	91CA	91BE	91BE	91A3	91A3	9197	9197	9198	9198	91BC	91BC	91B9	91B9	91A5	91A5	919C	919C	91BC	91BC	
	9234	923A	9248	9248	924C	924C	924C	9258	9258	9261	9261	9272	9272	9287	9287	926F	926F	9271	9271	9285	9285	9273	9273	9273	
	919B	9160	9169	91C4	9235	9273	9282	923C	91E1	919F	919C	91CD	9200	923C	926E	926B	9256	9252	926F	9297	927A	929D	919A	9158	916
252	91F3	91F3	91E6	91E6	91B6	91C2	91C2	91DF	91DF	91B4	91B4	91BA	91BA	91CB	91CB	91C9	91C9	91B2	91B2	9198	9198	91C	91C	91C	
	9244	9244	9215	9215	922A	922A	924A	924A	9245	9245	9245	9245	926D	926D	9278	9278	926D	926D	9261	9261	9274	9274	9283	9283	926
	91AD	919F	91A7	91E3	9237	925D	9258	921C	91BC	9187	91A6	9207	926C	92BC	92E1	92B1	9268	925A	9282	92BE	92BB	923C	9182	91WD	901
253	91E3	91E3	91D8	91D8	91C2	91C2	91D2	91D2	9200	9200	91D3	91D3	91B9	91B9	91D5	91D5	91E5	91E5	91B4	91B4	9198	9198	91B6	91B6	
	9236	9236	921B	921B	922A	922A	9245	9245	9233	9233	921B	921B	9256	9256	926D	926D	924F	924F	923F	923F	925E	925E	928C	928C	
	9152	915E	9183	91C7	9222	9258	924A	91FF	91B9	91A9	91CB	9218	9298	92D8	92EF	92CA	9298	9273	928E	92D4	92D6	924E	9188	915F	915
254	91EB	91EB	91F3	91F3	9297	9207	91F4	91F4	91BE	91BE	91D9	91D9	91E5	91E5	91D1	91D1	91C8	91C8	91DF	91DF	929D	929D	918A	918A	
	9237	9237	9248	9248	9235	9235	9211	9211	921D	921D	924n	924n	923B	923B	923F	923F	925B	925B	926F	926F	9242	9242	920E	920E	
	9003	9003	9197	9155	91D2	921D	923B	9232	928n	91E7	91FC	9236	928F	92D1	92E6	92CA	9297	9287	929F	92CB	9285	924E	9196	910A	910
255	91E3	91E3	91E1	91E1	91FB	91FB	9200	9200	91EF	91EF	91ED	91ED	91D9	91D9	91CA	91CA	91C8	91C8	91CE	91CE	9200	9200	919F	919F	
	9227	9227	922F	922F	9231	9231	921B	921B	9220	9220	9225	9225	9223	9223	9231	9231	9240	9240	924C	924C	923A	923A	921C	921C	
	9004	9004	9020	9133	91A7	91F9	920C	91EE	91C5	9199	91B3	9200	9272	92BD	92CB	9294	923F	9214	9223	9258	9274	9228	919A	918D	918
256	91E3	91E3	91F3	91F3	91D7	91D7	91D5	91D5	91E6	91E6	91CC	91CC	91B4	91B4	91CE	91CE	91DB	91DB	91C2	91C2	91AD	91AD	91F1	91F1	
	9220	9220	9209	9209	920C	920C	920F	920F	9212	9212	920B	920B	9224	9224	9241	9241	9231	9231	922E	922E	9263	9263	927E	927E	
	9005	9005	9005	9193	91AD	9245	929A	927D	91F3	9153	9130	918C	9211	927B	9287	9235	91D4	9198	91B4	91DB	91F7	91DD	917D	9116	9005
257	9006	9006	9006	91A7	91A7	91FB	91FB	9219	9219	910E	910E	91AE	9186	9186	91CA	91CA	91E7	91E7	91BC	91BC	91A8	91A8	91D7	91D7	
	9210	9210	9207	9207	920F	920F	924A	924A	9202	9202	91E7	91E7	922C	922C	924F	924F	922E	922E	922B	922B	924D	924D	924C	924C	
	92C9	92CD	92C8	92C9	92E8	930E	933F	9328	9299	91DF	9170	9142	916E	91CD	9200	91D1	917B	915E	9184	91C8	91EC	91CB	916E	9113	9005
258	91CE	91CE	91F9	91F9	91EE	91EE	91DF	91DF	91E5	91E5	9200	9200	9200	91FE	91FE	920E	920E	9242	9242	924A	924A	922E	922E	9222	9222
	9216	9216	920E	920E	9200	9200	9200	9200	9215	9215	920D	920D	91FE	91FE	920E	920E	9242	9242	924A	924A	922E	922E	9222	9222	
	9317	92FB	9314	9311	9314	932B	9331	92FB	9266	919F	9192	9191	9119	9173	9173	9140	9133	9167	91C9	921F	9225	91E7	9191	917	
259	91F9	91F9	91EC	91EC	91D0	91D0	91B8	91B8	91E1	91E1	91E1	91E1	9100	9100	91A7	91A7	91BC	91BC	91ED	91ED	91F3	91F3	91C8	91C8	
	9282	9282	91F2	91F2	91FE	91FF	91FF	921C	921C	9205	9205	9205	9205	9200	9200	9234	9234	9250	9250	9240	9240	9217	9217	9200	9200
	92D5	92DC	92DE	92E4	92F1	9300	932A	92EE	9214	9182	9182	9117	9157	9157	913C	912C	9183	91F0	9243	9250	922F	922A	925	925	
260	91D5	91D5	91EF	91EF	91FD	91FD	9203	9203	9200	9200	9200	9200	9200	9200	9200	9200	9200	91D7	91D7	91C9	91C9	91D7	91D7	922	922
	9207	9207	9206	9206	9205	9205	920																		

10-Bit Video Data Scope

02D6	024E	0188	00DB	0007	00C7
02BD	02BD	01BA	01BA	0151	0161
0242	0242	02BE	02BE	0266	0266
02B5	024E	0196	00DA	00B4	009E
026B	026B	019F	019F	0153	0153
023A	023A	021C	021C	027D	027D
0274	0228	019A	01BD	00C9	00B6
01AD	01AD	01F1	01F1	021B	021B
0263	0263	027E	027E	023B	023B

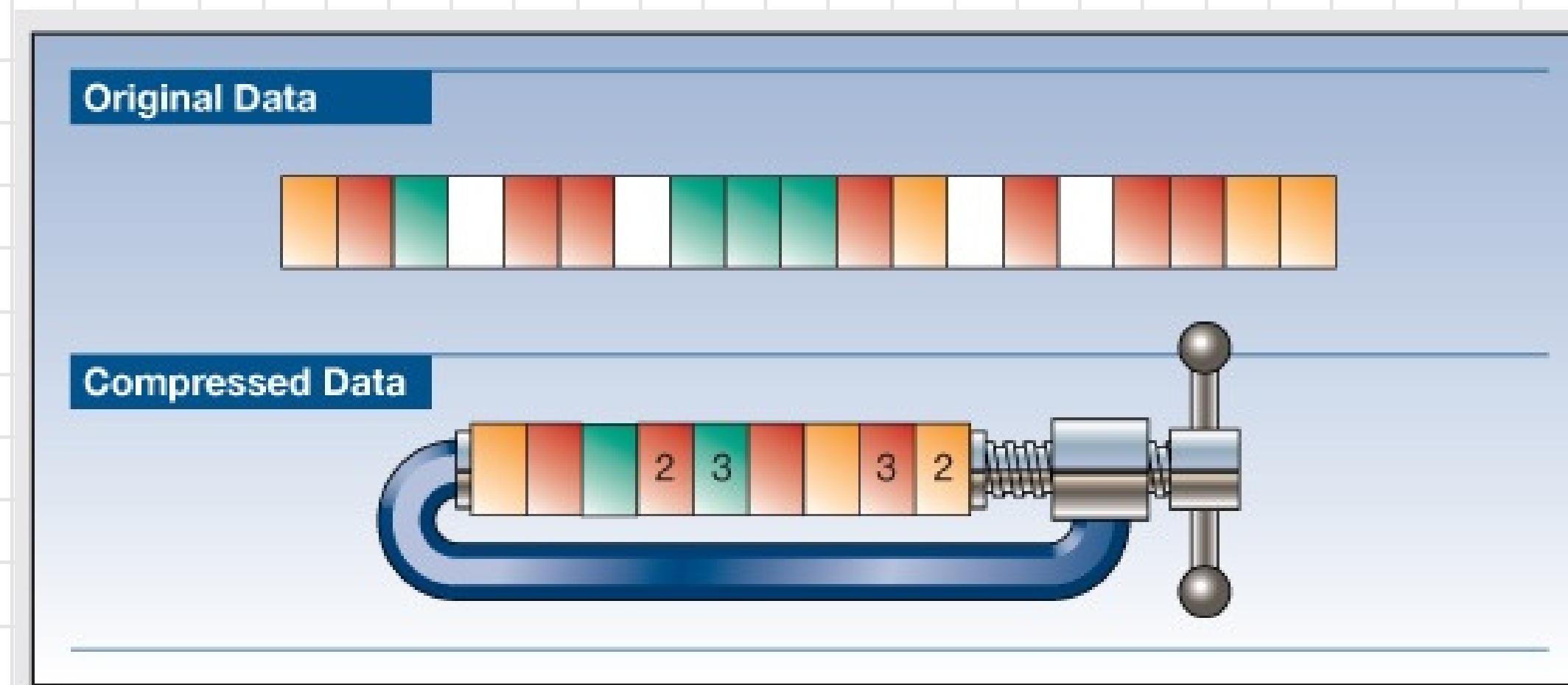


Codec vs Container

- Codec: Defines what the 1's and 0's mean. The actual av "essence"
 - PCM, FLAC, MP3, AAC
 - V210, FFV1, H.264, ProRes, J2K, DV25
- Container: Holds one or more streams, contains metadata about the organization and presentation of the av info
 - WAV, OGG, MOV, AVI, MKV, MXF, MP4

What is Compression

The processes of modifying, encoding or converting a set of binary bits in order to make it smaller





Compression and Preservation

- Preservation often uses Uncompressed Files to ensure no data is lost, but these files are huge
- Compression can save space, and increase accessibility

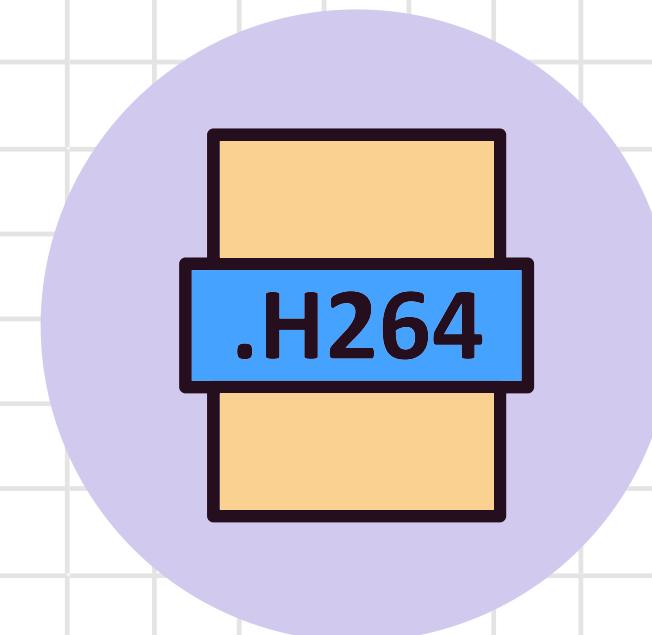


- Lossy
 - Some data is lost during compression and cannot be recovered
 - Common for "access files"
- Lossless
 - No data is lost during compression
 - All original data can be recovered

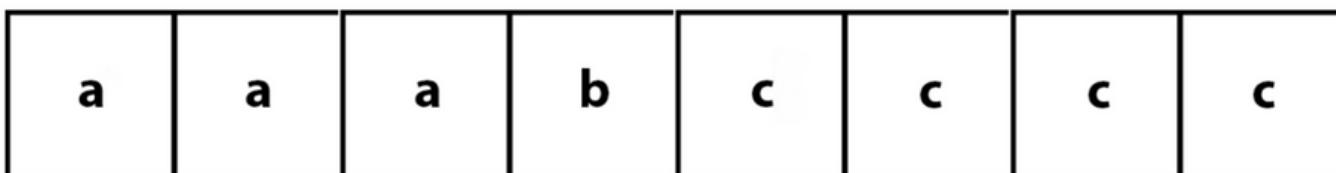
Codecs:
Lossy
vs
Lossless

Lossy Compression

- Lossy does NOT mean low quality
- Data is removed to make the file smaller.
- Perceptual encoding: Removes information that the human ear and eye cannot actually hear or see
- Throwing Away information to lead to better accessibility!



Lossless Compression

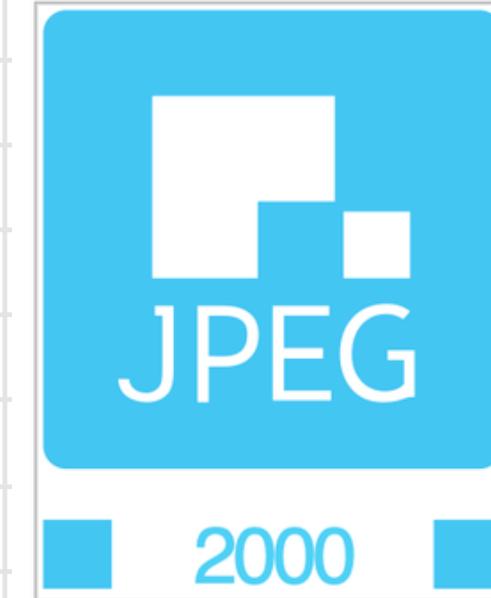


run-length encoding



a	3	b	1	c	4
---	---	---	---	---	---

- All compressed data can be retrieved
- It's not magic, it's math!
- Simple Example: Run Length Encoding (RLE)





Stream Organization In Codec and Wrapper

- Scan type: Interlacing
- Proper number of audio channels
- Timecode, continuous or not
- Extra information in VBI and VANC

- At the most basic: file level checksums
- FFV1 and MXF allow for frame-level or stream-level checksums

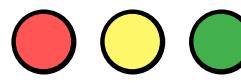
Maintaining Fixity



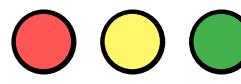
Format Longevity and Adoption

- Open Source presents a possible advantage here
- MKV is flexible enough as a wrapper to include a player inside the file!
- LoC primarily uses J2K/MXF but is not accepting FFV1/MKV files
- IU digitized over 300k hours of video content to FFV1/MKV

Video Preservation Format Comparison							
Encoding / Wrapper	Bit Depth	Compression	Flexible Streams	Frame-Level Checksum	Open Source	Broad Technical Support	Conclusion
H.264 / MP4	Depends, but often 8 Bit	Lossy	Medium	No	No	Yes	1.5 / 6
DV25 / DV	8 Bit	Lossy	No	No	No	No	0 / 5
ProRes / MOV	10 Bit	Lossy	Medium	No	No	Yes	2.5 / 6
UYVY / MOV	8 Bit	None	Medium	No	No	Medium	2 / 6
V210 / MOV	10 Bit	None	Medium	No	No	Yes	3.5 / 6
J2K / MXF	10 Bit	Lossless	Yes	Yes	No	No	4 / 6
FFV1 / MKV	10 Bit	Lossless	Yes	Yes	Yes	No	5 / 6



Exercise 01: Compression Comparison



Exercise 01: Compression Comparison

Q&A

Exercise 02: Media Mystery Meat



Exercise 02: Media Mystery Meat

Q&A



TAKE A
BREATHER



Break Time!

What is FFV1 MKV?

- **FFV1 -> Codec**

- Open Source
- Lossless
- Intra-Frame
- Developed by FFmpeg Community

- **MXF -> Container**

- Flexible and open source format
- Broad support of av codecs
- Highly configurable streams with stream-level checksums

FFV1 contains the **Video information**
ONLY

MKV is the **container** which holds
the video, audio, subtitles,
timecode, and any other ancillary
data

How do
FFV1 and
MKV Work
Together?

FFV1 MKV For Video Preservation

The standardization efforts for both FFV1 and MKV were developed concurrently as part of a wider effort to move the archival field to an accessible lossless compressed/open alternative

Exercise 03: Let's Break an FFV1 File

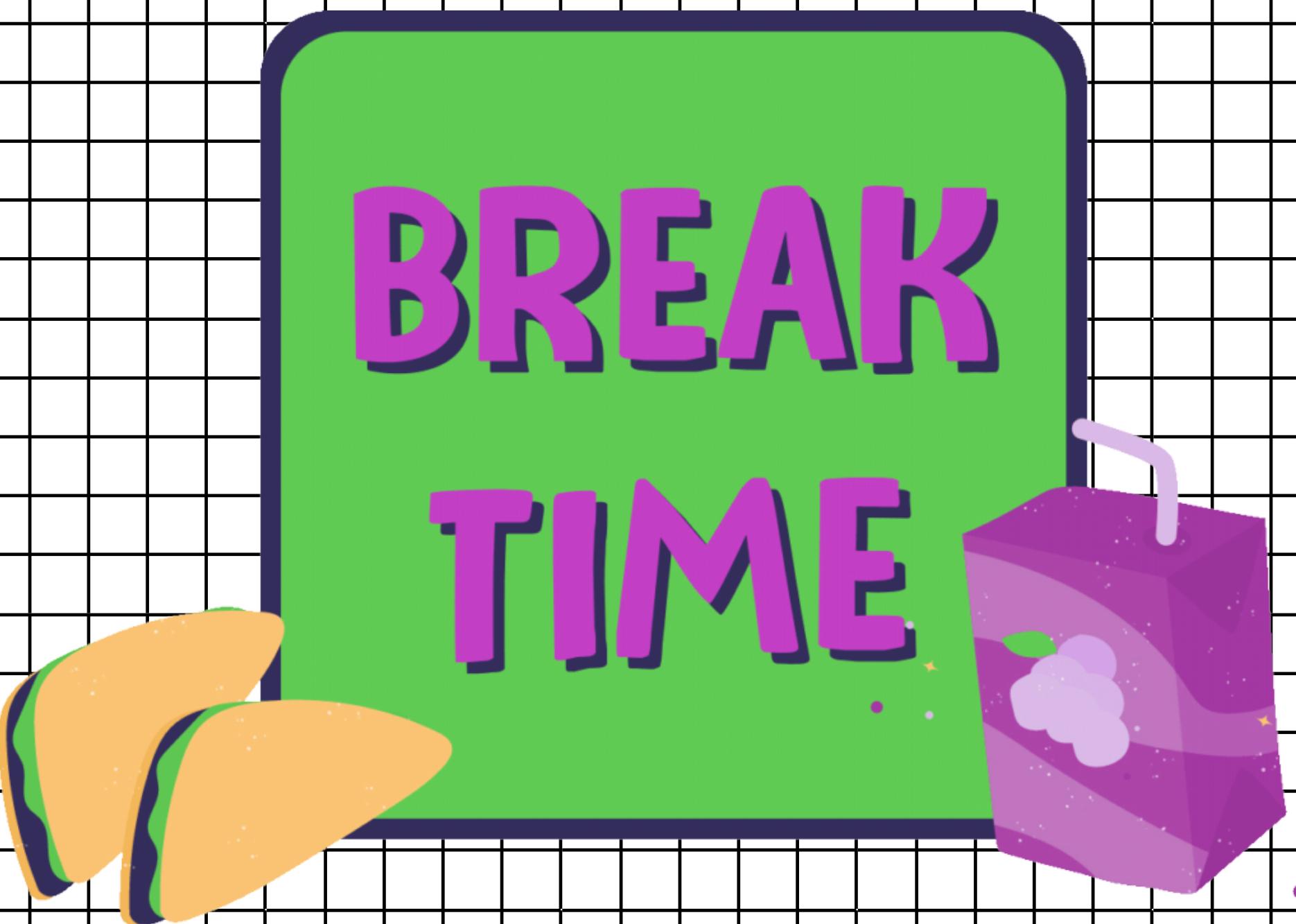
Exercise 03: Let's Break an FFV1 File

Q&A

Exercise 04: Round-Trip Transcode

Exercise 04: Round-Trip Transcode

Q&A



Digitization Station Qualification

Does the Digitization Station work properly?

- Is the analog signal properly converted to digital?
- Is the digital signal properly recorded onto the storage media?
- Does the recorded format match the desired specifications



Is the
analog
signal
properly
converted
to digital?

We need to test that the Analog to Digital converter is working properly. High quality converters are more stable and sound better. Some converters like the Prism Sound Lyra have built in checks to make sure that the converter is working properly



We need to test that all digitized samples are being handled and saved properly. Sometimes when computers or hard drives are overloaded they drop or repeat samples.

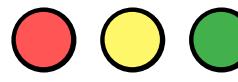
**Is the
digital
signal
properly
recorded?**



Does the recorded format match the desired specs

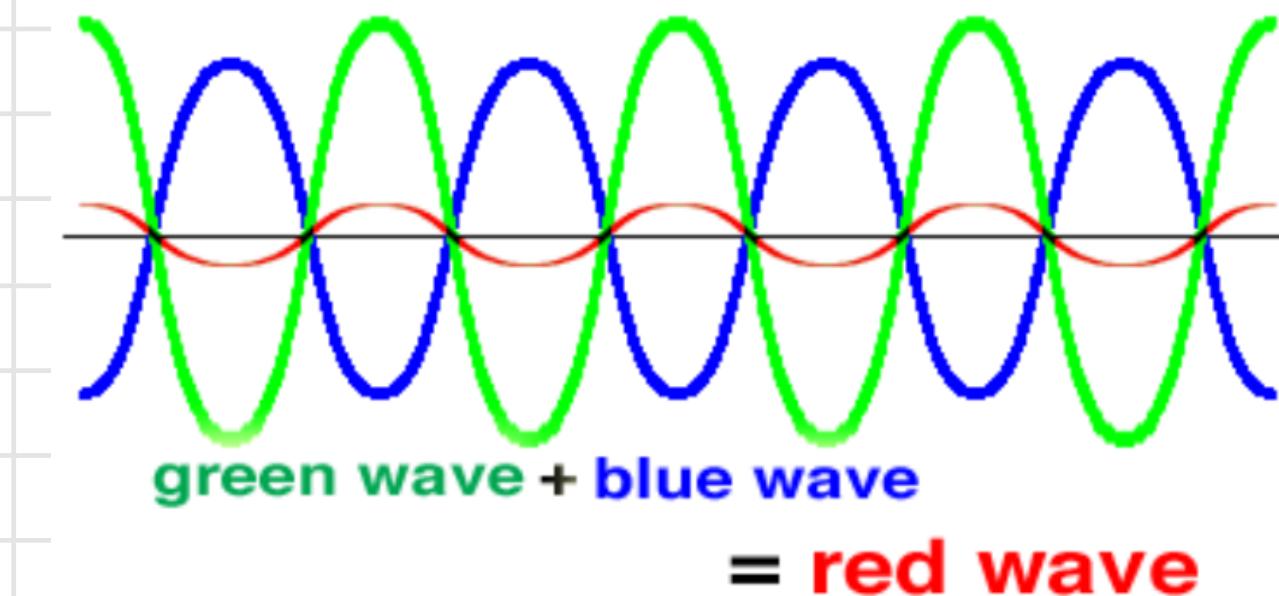
We need to test that the final file matches the desired specifications.

- Linear PCM
- 96kHz
- 24bit
- etc

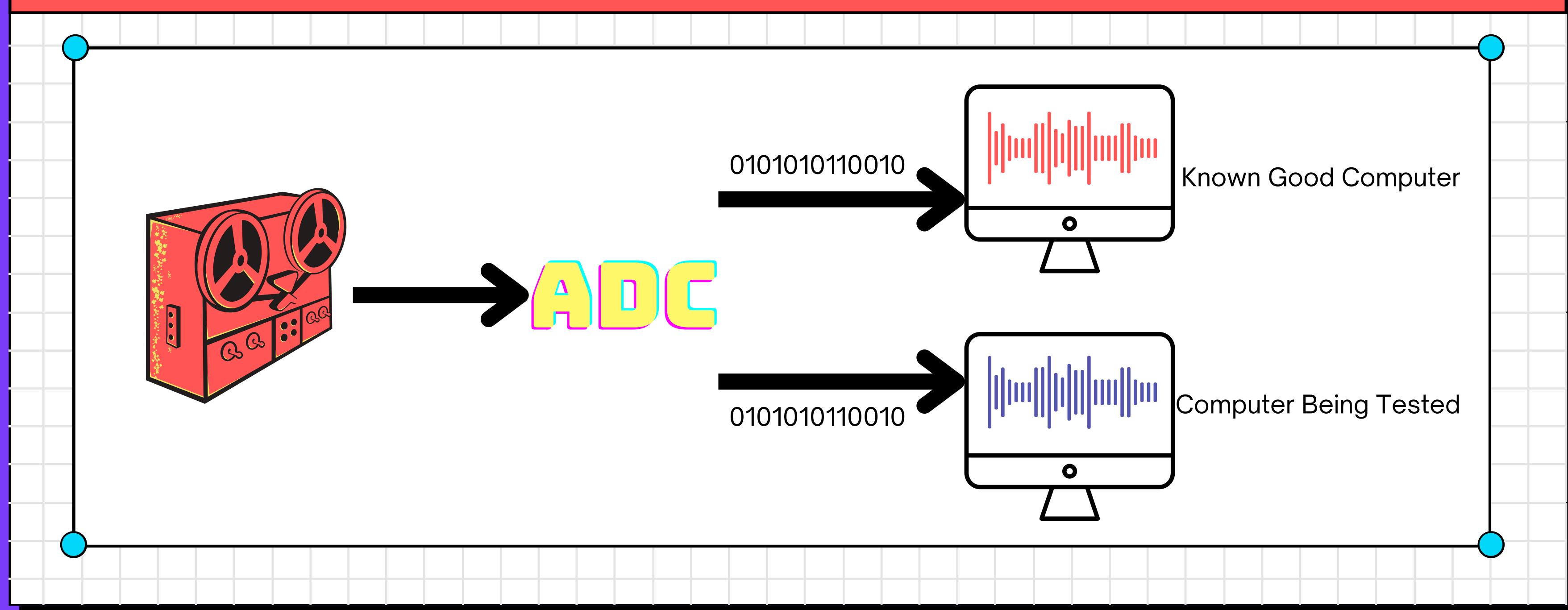


Audio Null Test for Station Validation

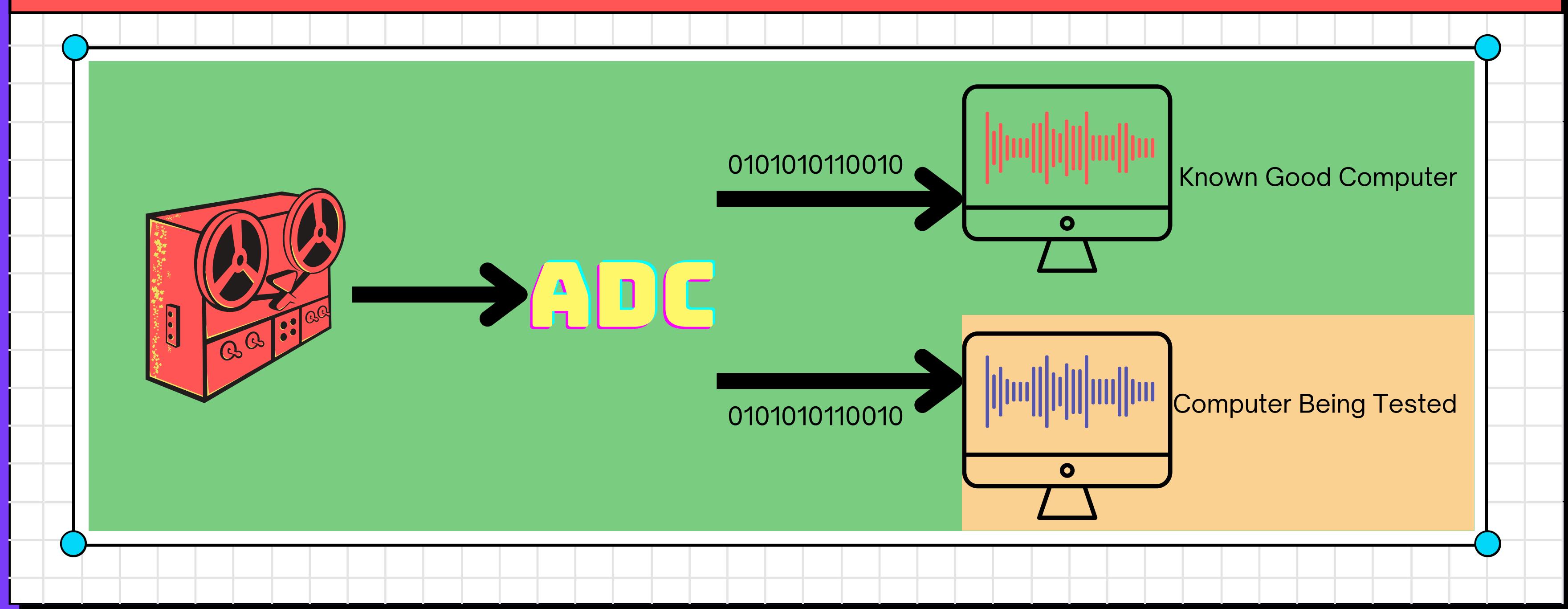
- Null tests ensure that the digitized signal is properly recorded onto the storage media
- Null tests work using the principal of wave addition and interference



Audio Null Test



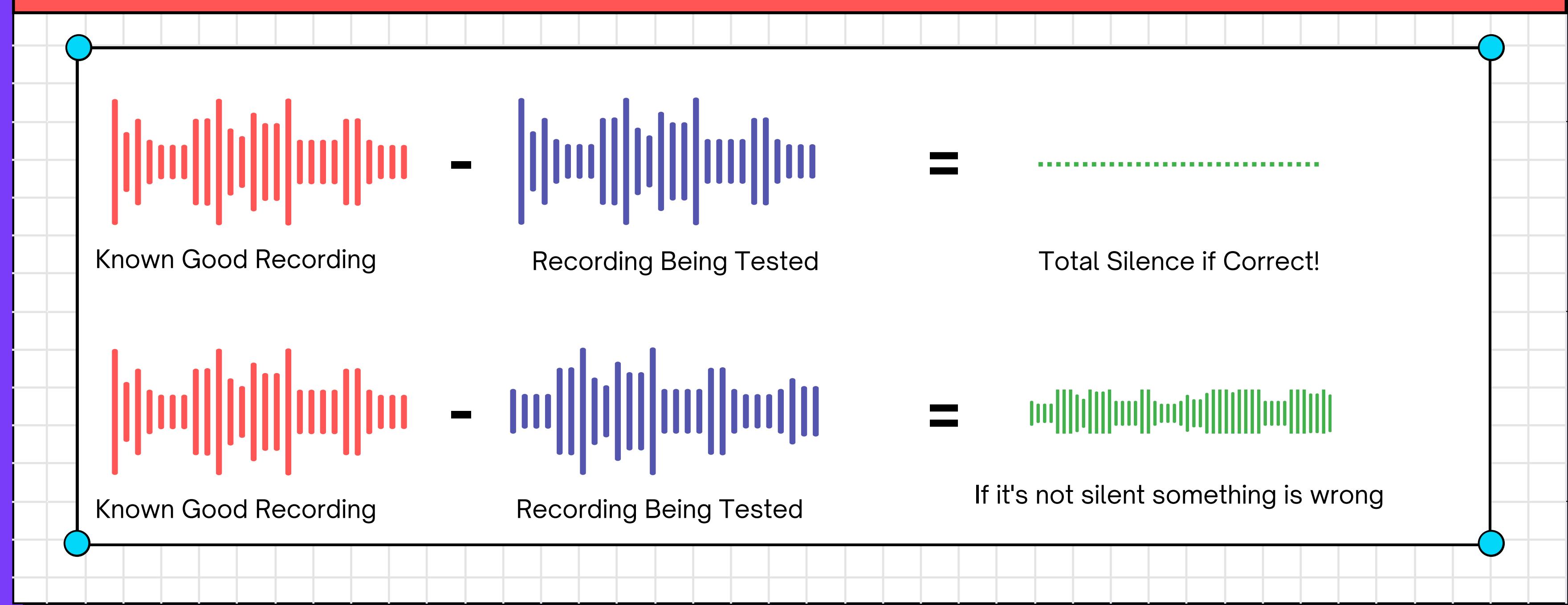
Audio Null Test

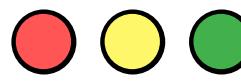


Assumed to be
working properly

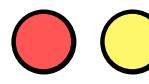
The item we are
testing

Audio Null Test



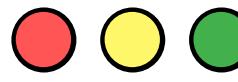


Exercise 05: Audio Station Qualification



Exercise 05: Audio Station Qualification

Q&A



Video Station Qualification

Currently no public guidelines or best practices exist.

But we can rely on the same basic principals

- Define parameters: what exactly are we checking
- Determine test points
- Create tests for these parameters using available tools



Determine Test Points

- Are you checking the entire signal path?
 - Deck -> TBC -> Converter -> Computer
- Just conversion and capture?
 - Analog Signal Generator -> Converter -> Computer
- Just capture?
 - Digital Signal Generator -> Computer
- **Best Practice**
 - Isolate what you are testing for
 - When troubleshooting only change one variable at a time



Define parameters

- Is the analog signal properly converted to digital?
- Is the digital signal properly recorded onto the storage media?
- Does the recorded format match the desired specifications

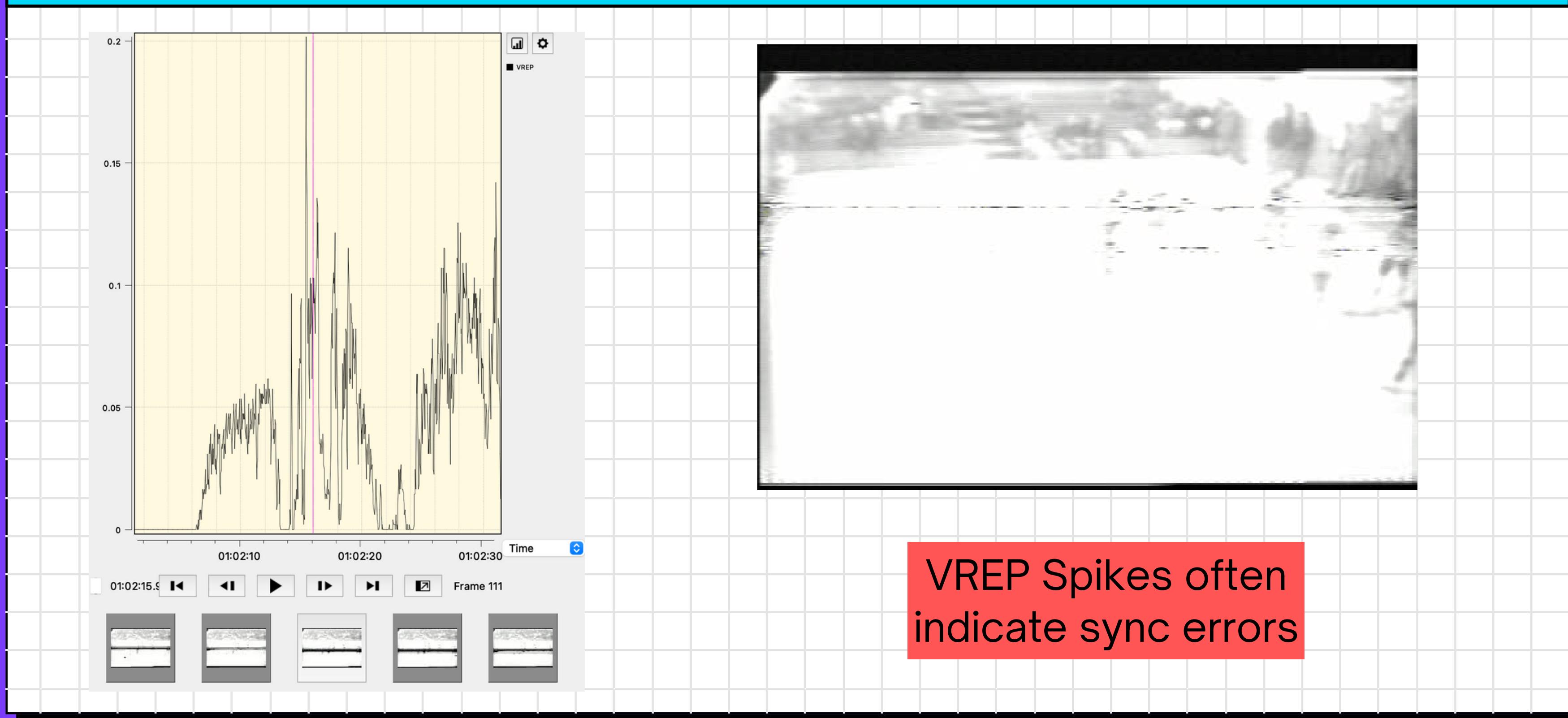


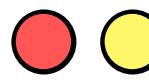
Is the analog signal properly converted to digital?

- Are there sync errors?
- Are there repeated frames?
- Is the digitized signal 10 bits?
- Is the video waveform clipped at all?

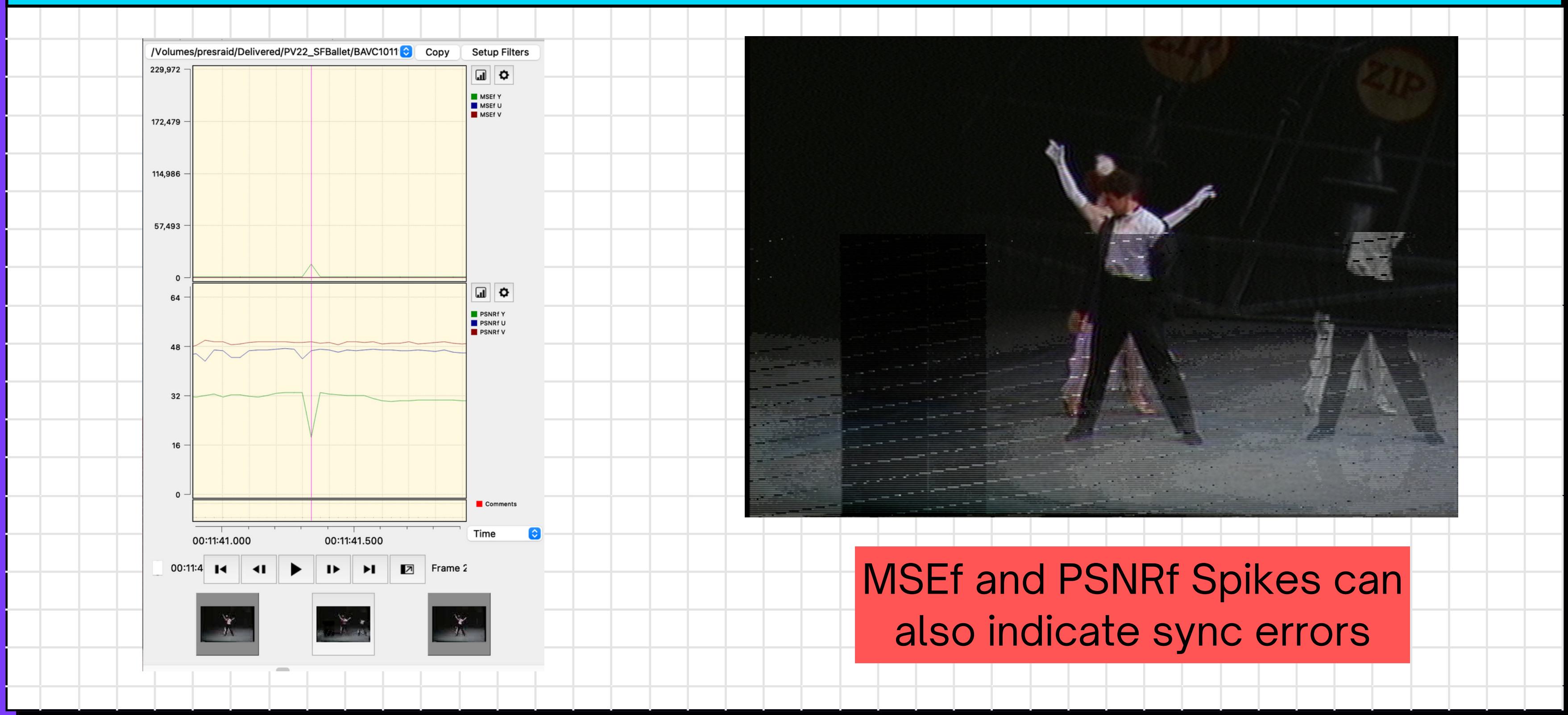
Create tests
for these
parameters
using
available
tools

Identifying Sync Errors in QCTools

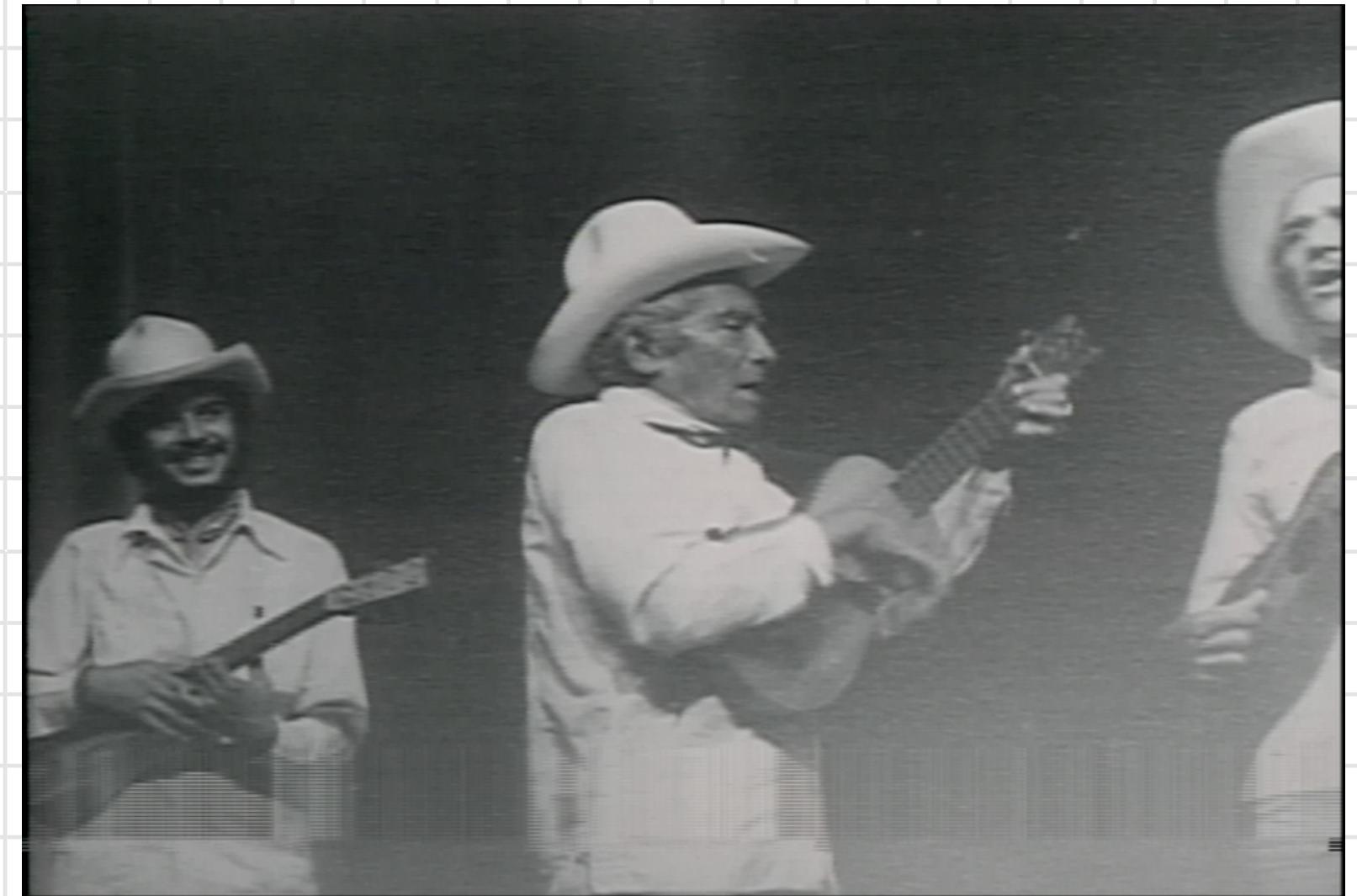




Identifying Sync Errors in QCTools

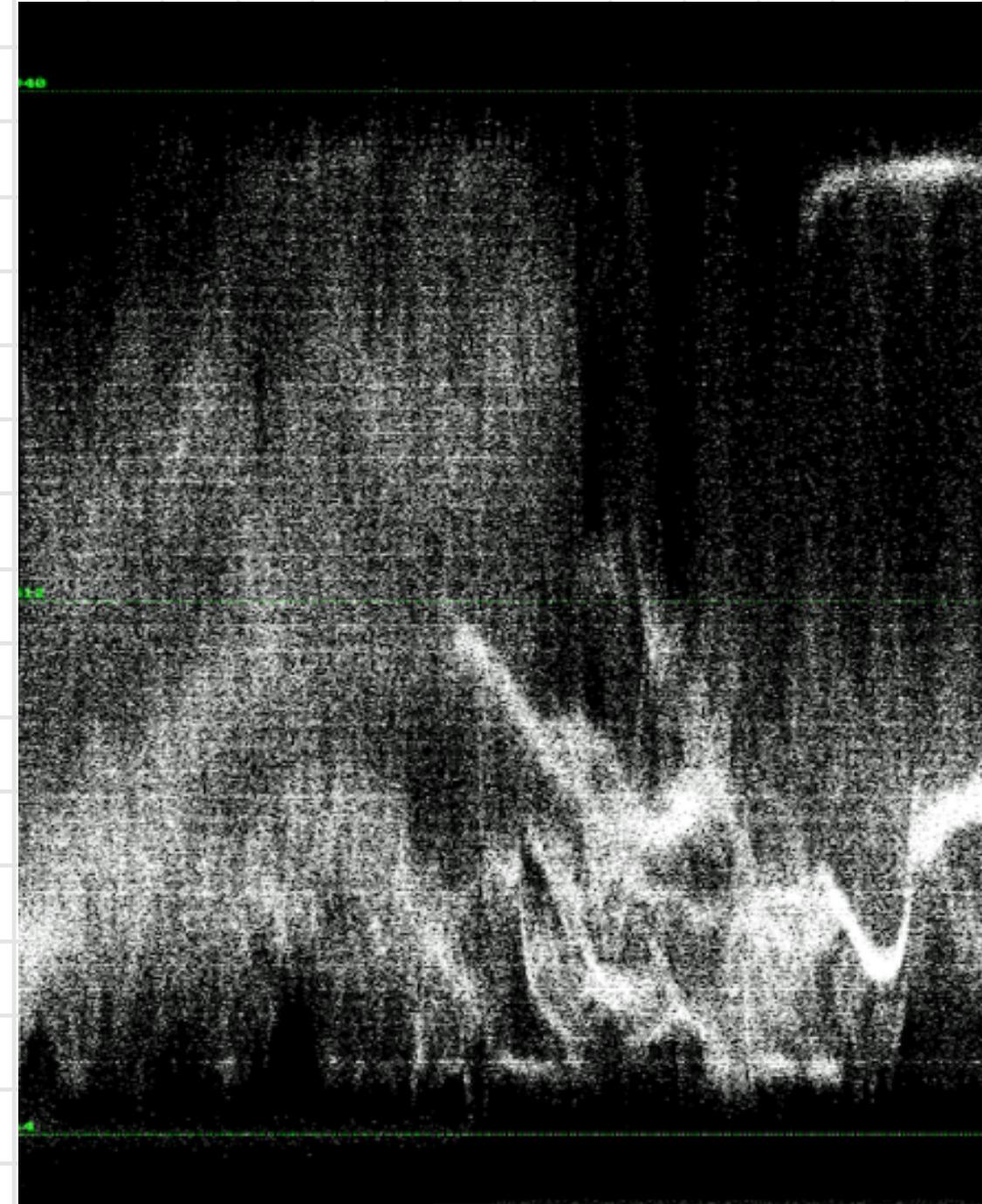


Identifying Repeated Lines in QCTools

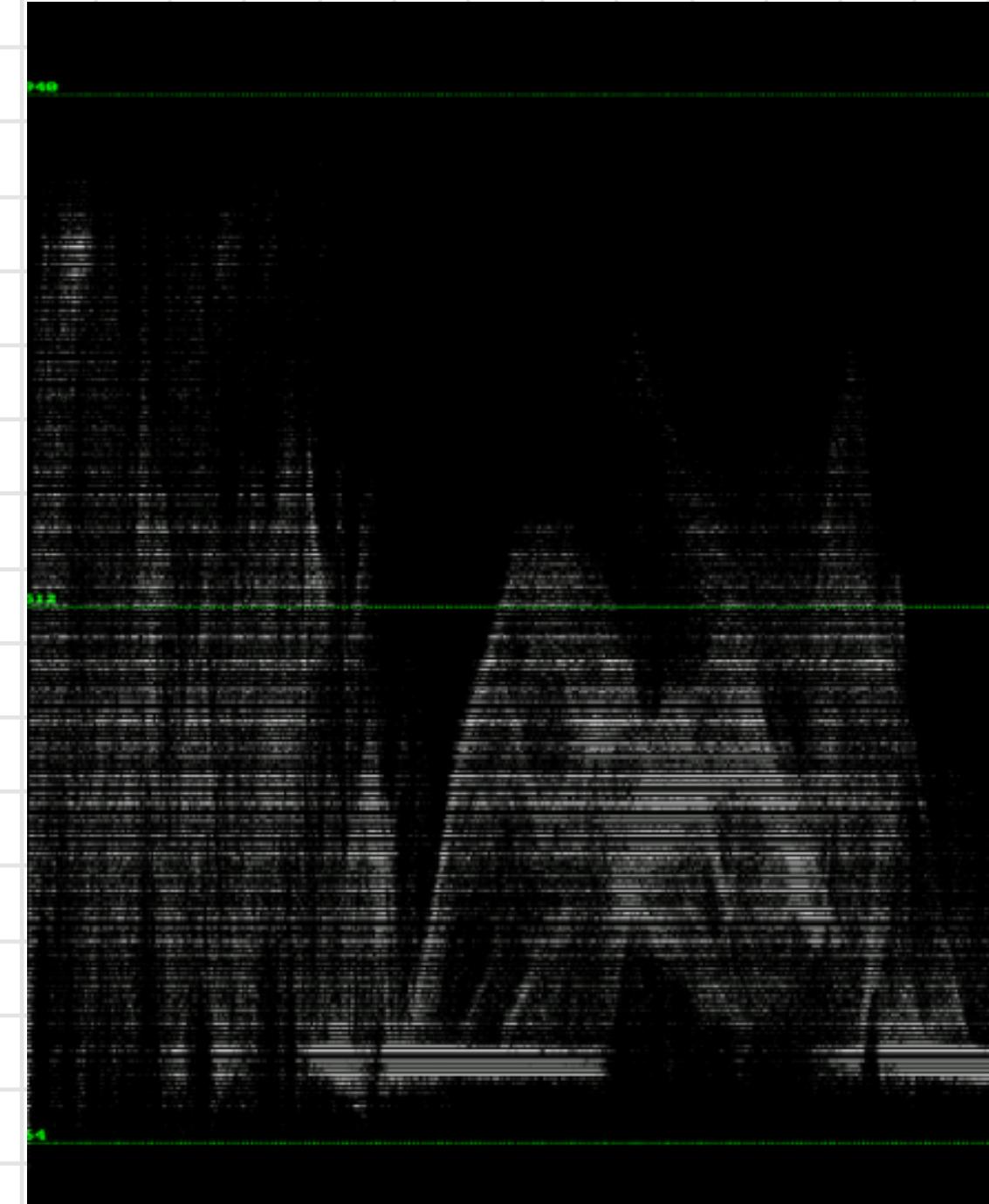


VREP Spikes can help find repeated lines often seen in poor Betacam Transfers

Identifying 10bit/8bit Errors in QCTools



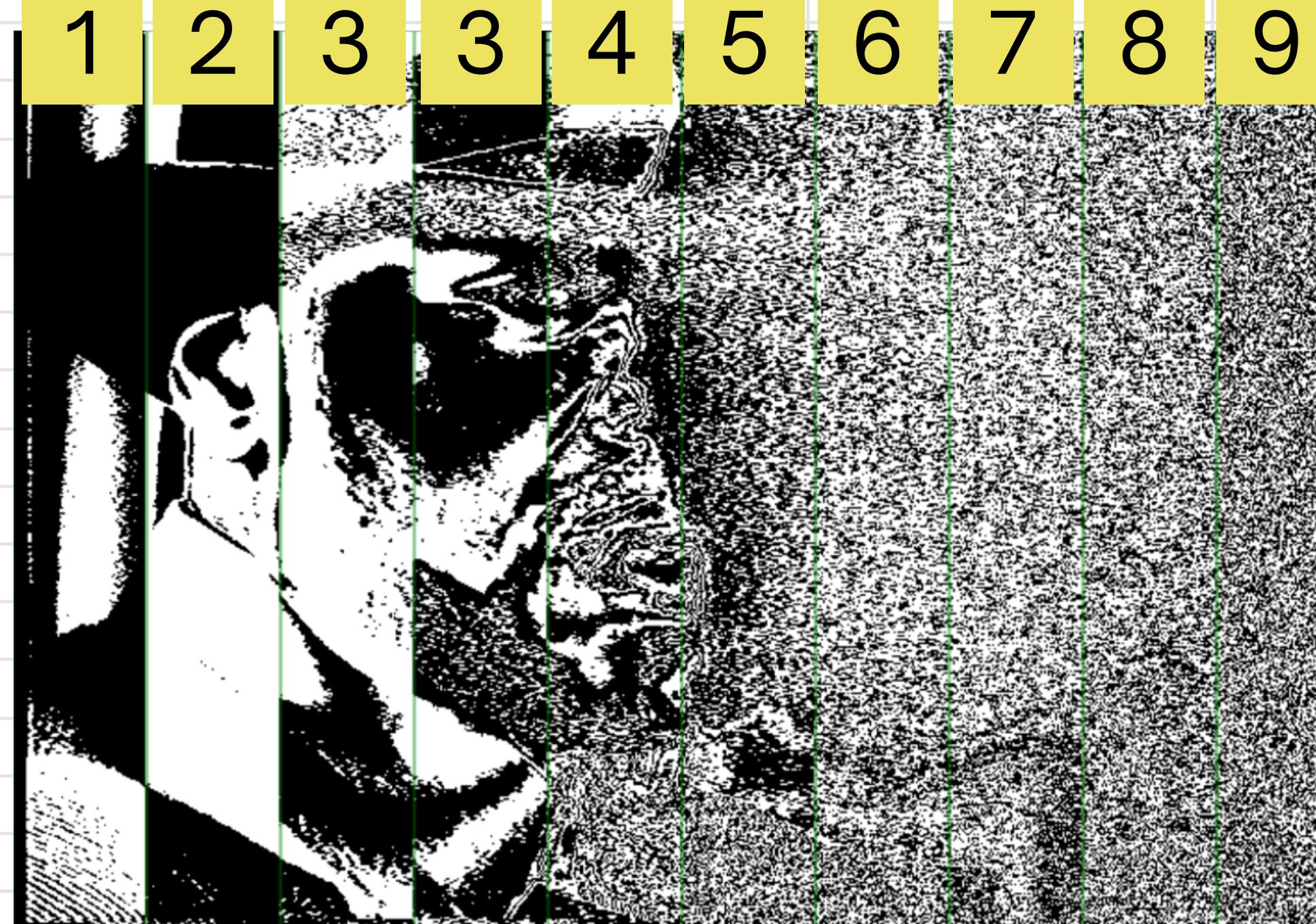
Proper 10 Bit Waveform



8 Bits of Data in 10 Bit Scale

Identifying 10bit/8bit Errors in QCTools

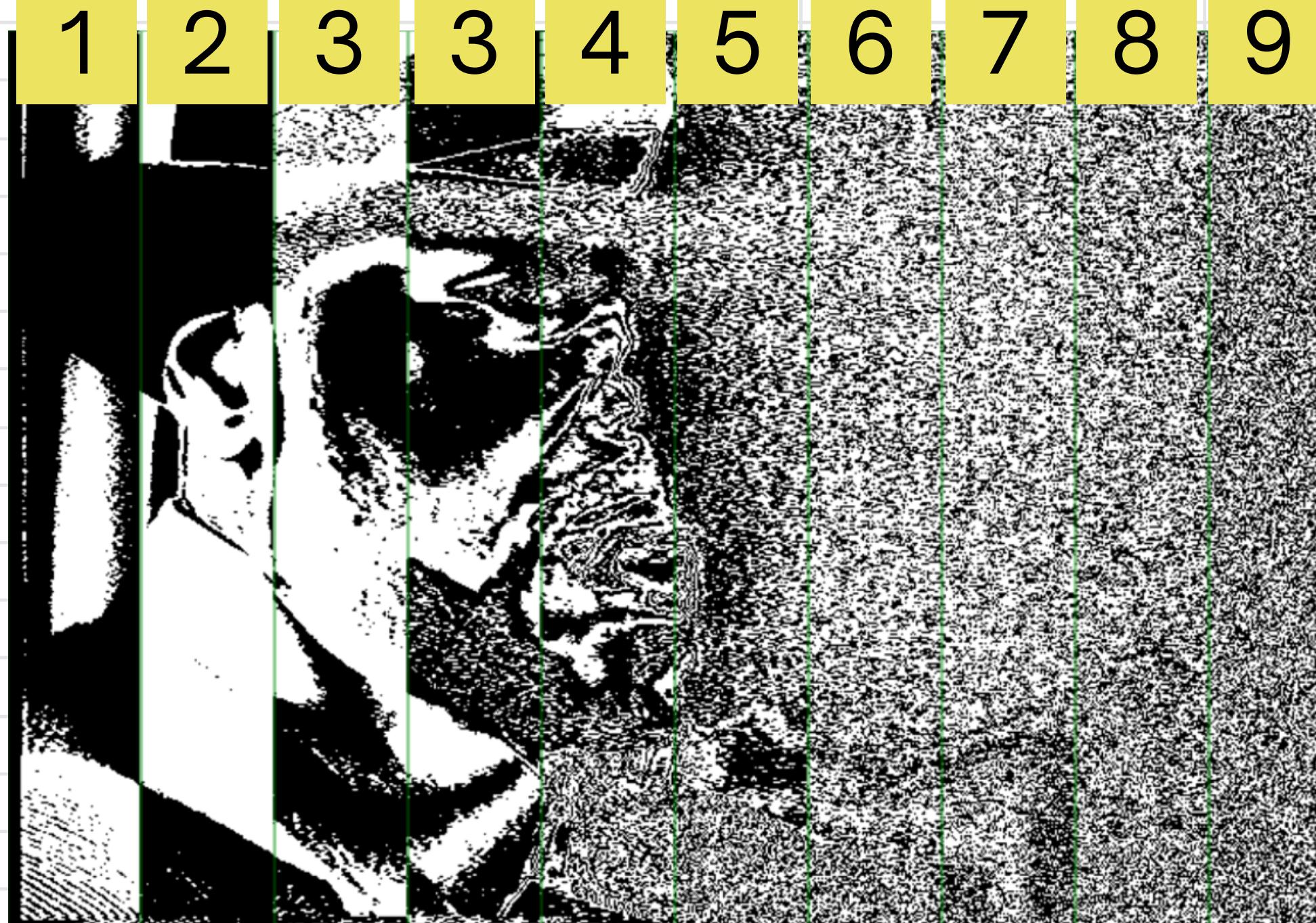
Bit
Position



Proper 10 Bit Slice

Identifying 10bit/8bit Errors in QCTools

Bit
Position

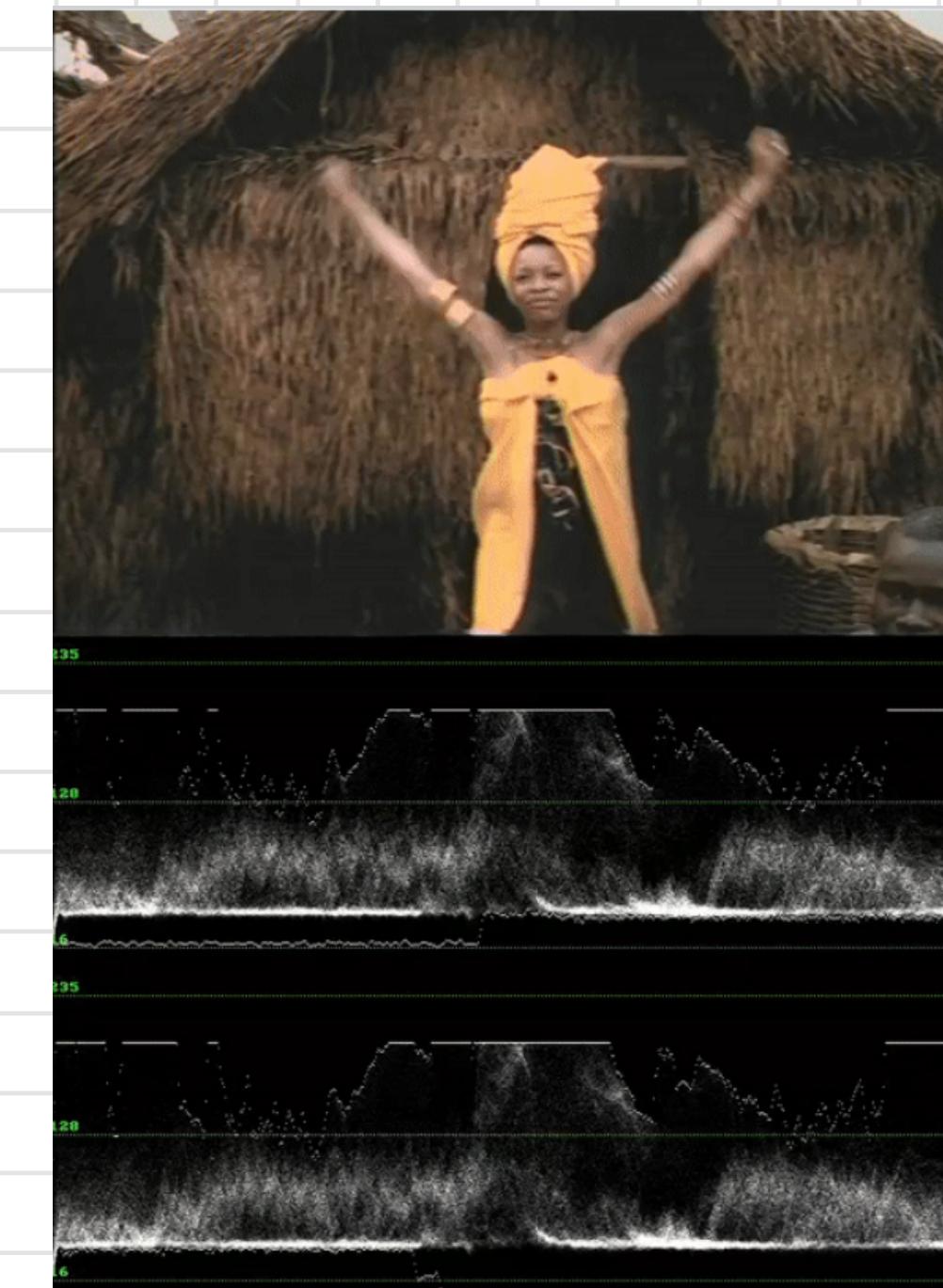


Only 8 Bits of data in 10 Bit Slide

Identifying Clipping in QCTools



No Clipping



Clipping



Is the digital signal properly recorded onto the storage media?

- Are the dropped frames?
- Does the built-in file checksum verification pass?

Create tests for these parameters using available tools

Identifying Bad FFV1 with MediaConch

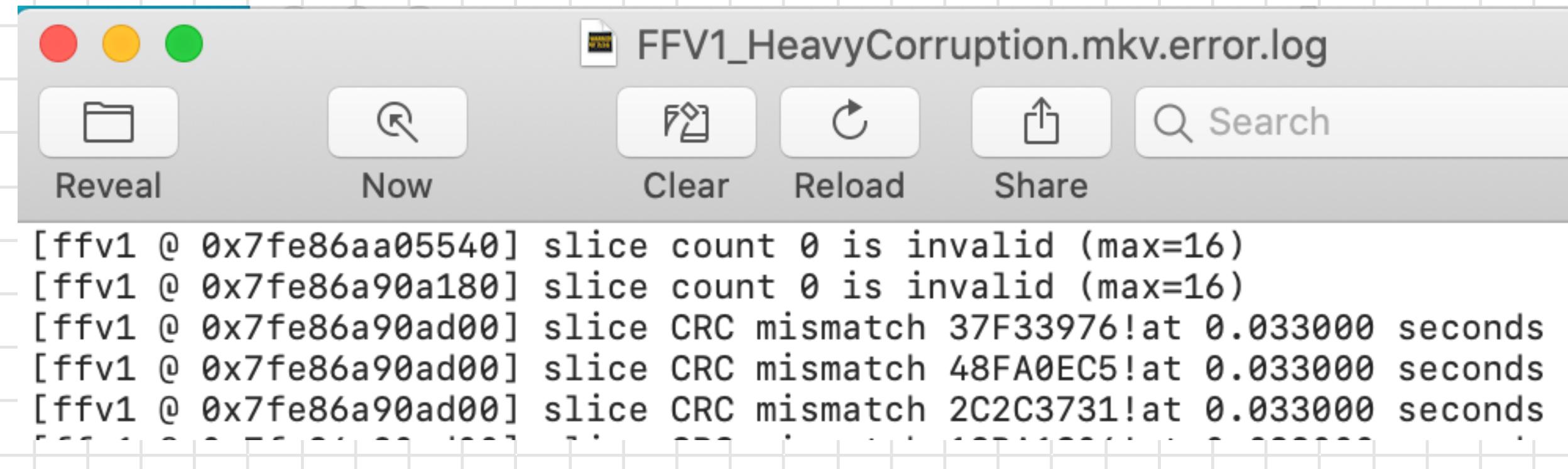
Files	Implen
FFV1.mkv	✓ Valid
FFV1_HeavyCorruption.mkv	✗ Not valid
FFV1_LightCorruption.mkv	✓ Valid
FFV1_MediumCorruption.mkv	✓ Valid

This will only work with files that have bad checksums in the beginning of the file! Beware of false positives.

Identifying invalid checksums and errors with FFmpeg

```
ffmpeg -v error -i file.mkv -f null - 2> file.mkv.error.log
```

This will create a sidecar log file that lists all of the errors that FFmpeg encountered during decoding process





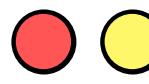
Does the recorded format match the desired specifications

- Does the file match the desired specifications?

Create tests for these parameters using available tools

Identifying Conformance with MediaConch

Files	Implen	Policy
BAVC1013216_239.mov	✓ Valid	✓ Vrecord 10-bit MOV Master
BAVC1013214_237.mov	✓ Valid	✓ Vrecord 10-bit MOV Master
BAVC1013211_234.mov	✓ Valid	✓ Vrecord 10-bit MOV Master
BAVC1013208_231.mov	✓ Valid	✓ Vrecord 10-bit MOV Master
BAVC1013138_214_mezzanine.mo...	✓ Valid	✗ Vrecord 10-bit MOV Master
BAVC1013138_214.mov	👁️ ⬇️	👁️ ⬇️



Identifying Conformance with MediaConch



MediaConch Report

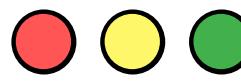
File: /Volumes/SymplyUltra/TransferProjects/PV21_SJSU/04_Loading/Drive07/Files_Mezzanine/BAVC1013138_214_mezzanine.mov

❖ Vrecord 10-bit MOV Master ✗ fail

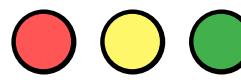
10-bit Uncompressed MOV

Type: and I Rules run: 38 | Fail count: 8 | Pass count: 30

- General/VideoCount is 1 ✓ pass
- General/AudioCount is 1 ✓ pass
- General/FileExtension is mov ✓ pass
- General/Format is MPEG-4 ✓ pass
- General/Format_Profile is QuickTime ✓ pass
- General/CodecID is qt ✓ pass
- General/OverallBitRate_Mode is CBR ✗ fail (Actual: VBR)
- General/FrameRate is 29.970 ✓ pass
- Video/Format is YUV ✗ fail (Actual: ProRes)
- Video/CodecID is v210 ✗ fail (Actual: apch)
- Video/BitRate_Mode is CBR ✗ fail (Actual: VBR)



Exercise 06: Video Station Qualification



Exercise 06: Video Station Qualification

Q&A

Exercise 6 Results

- Introduction
- Scope and Limitation
- Review of Related Literature
- Research Framework
- Methodology
- Results
- Conclusion
- Introduction
- Scope and Limitation
- Review of Related Literature
- Research Framework
- Methodology
- Results
- Conclusion



If you didn't identify these problems go back and see if you can now!



Conclusion

- Compression is good for preservation as long as it's lossless (and you can prove it!)
- There's a few choices of formats for video preservation, choose what works best for you
- You should be able to qualify a digitization tool using open source tools
- Nobody has compiled a series of Video Station Qualification tests yet...

Morgan Oscar Morel



morgan@bavc.org



[@av_morgan](https://twitter.com/av_morgan)



415.558.2158



Reach Out With Any Questions!