



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**

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

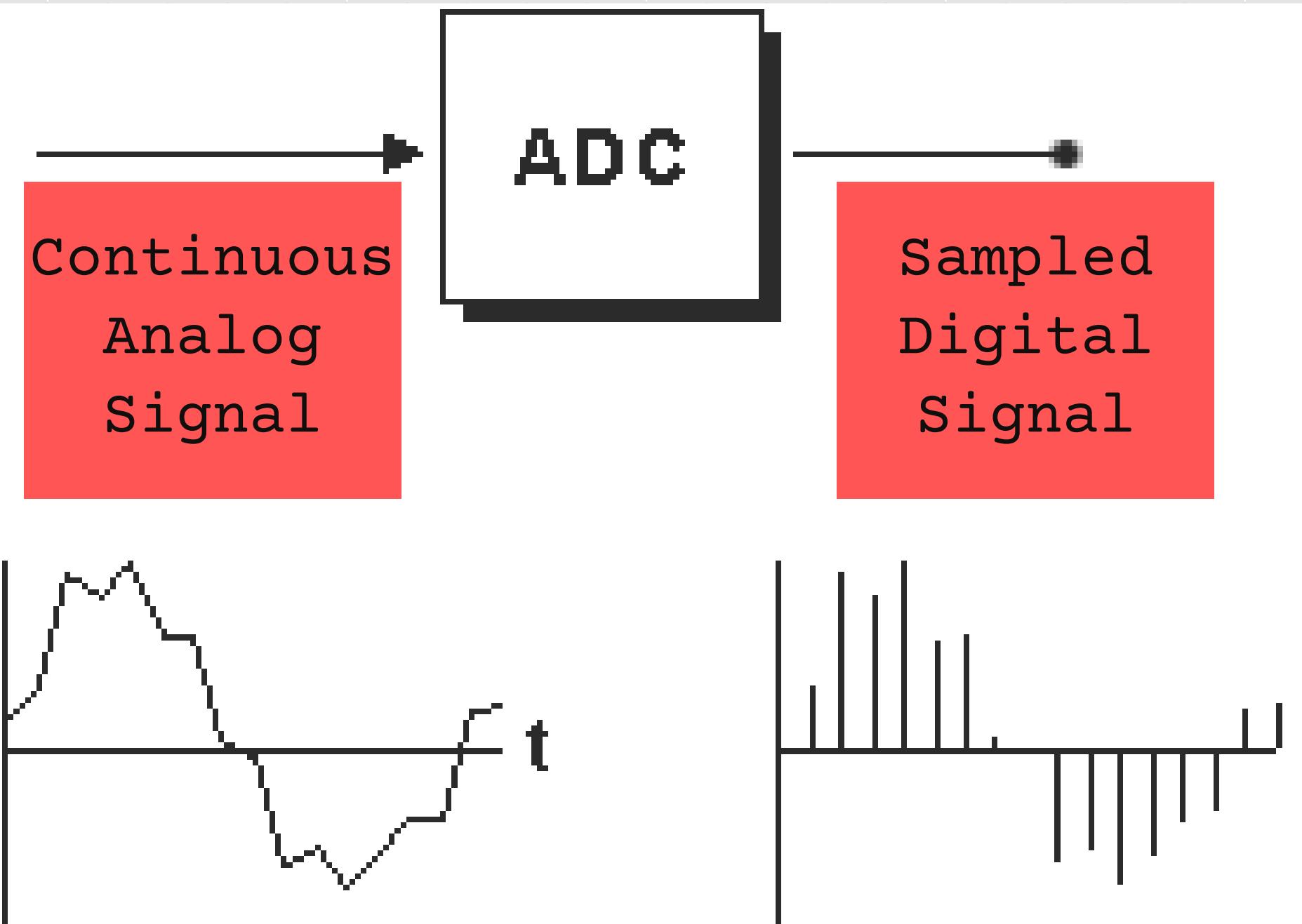
Standard Digital Preservation techniques are needed, plus extra work to deal with AV complexities



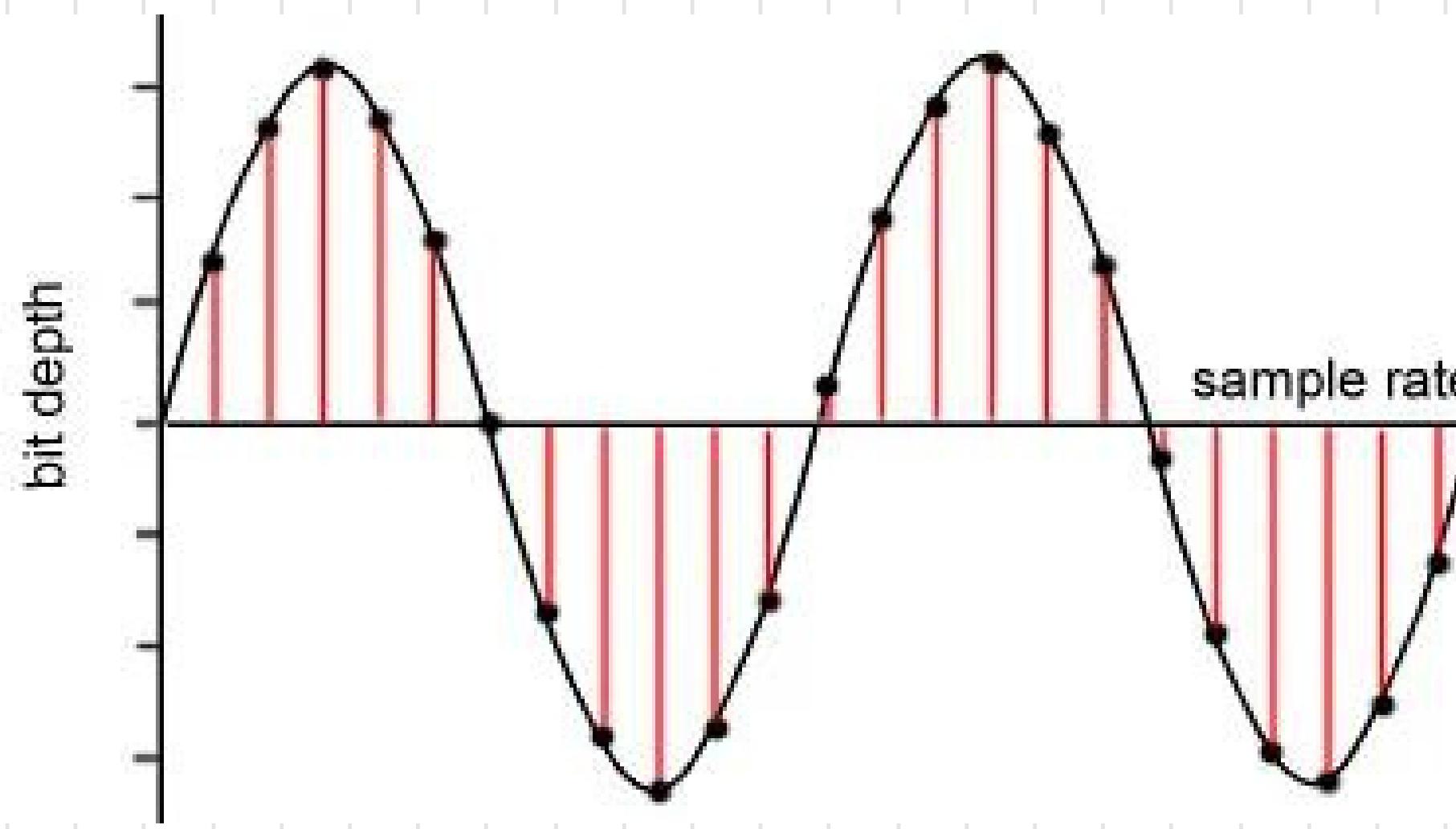
What Preservation Formats need to be able to do

- 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



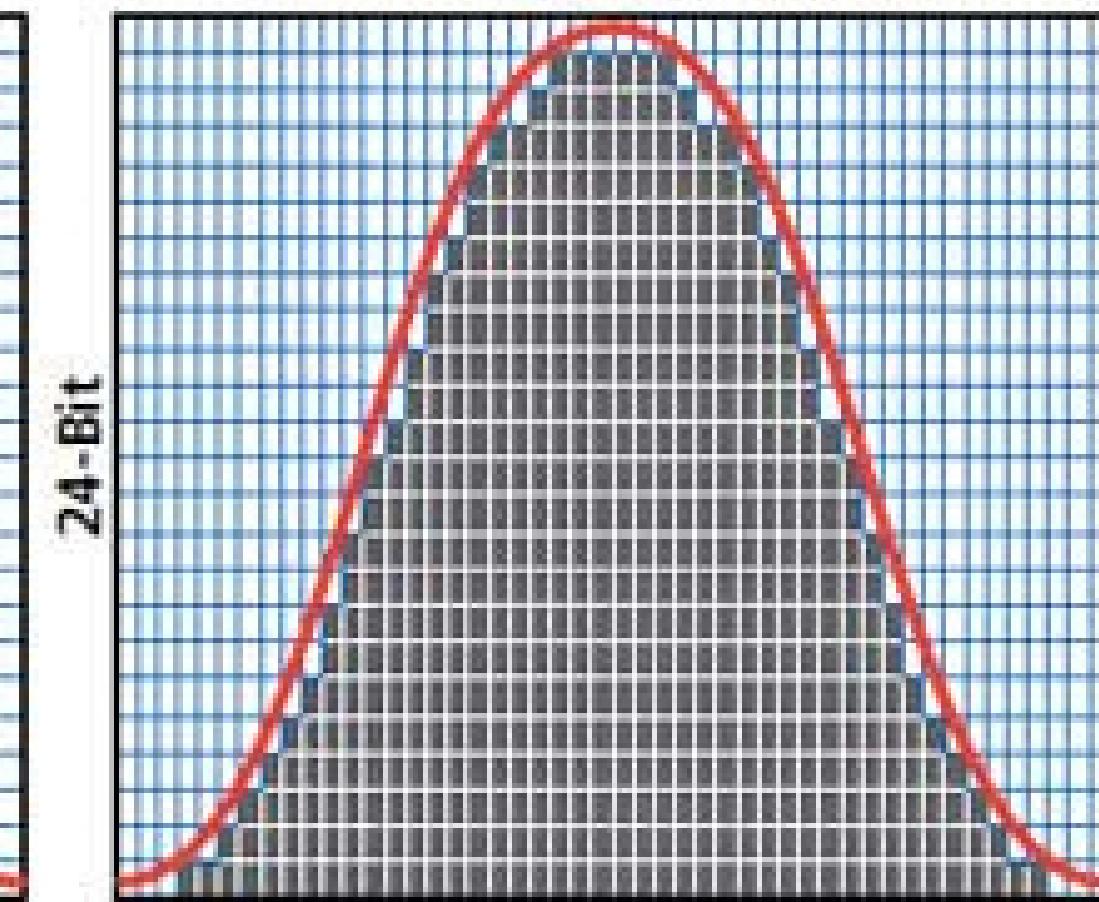
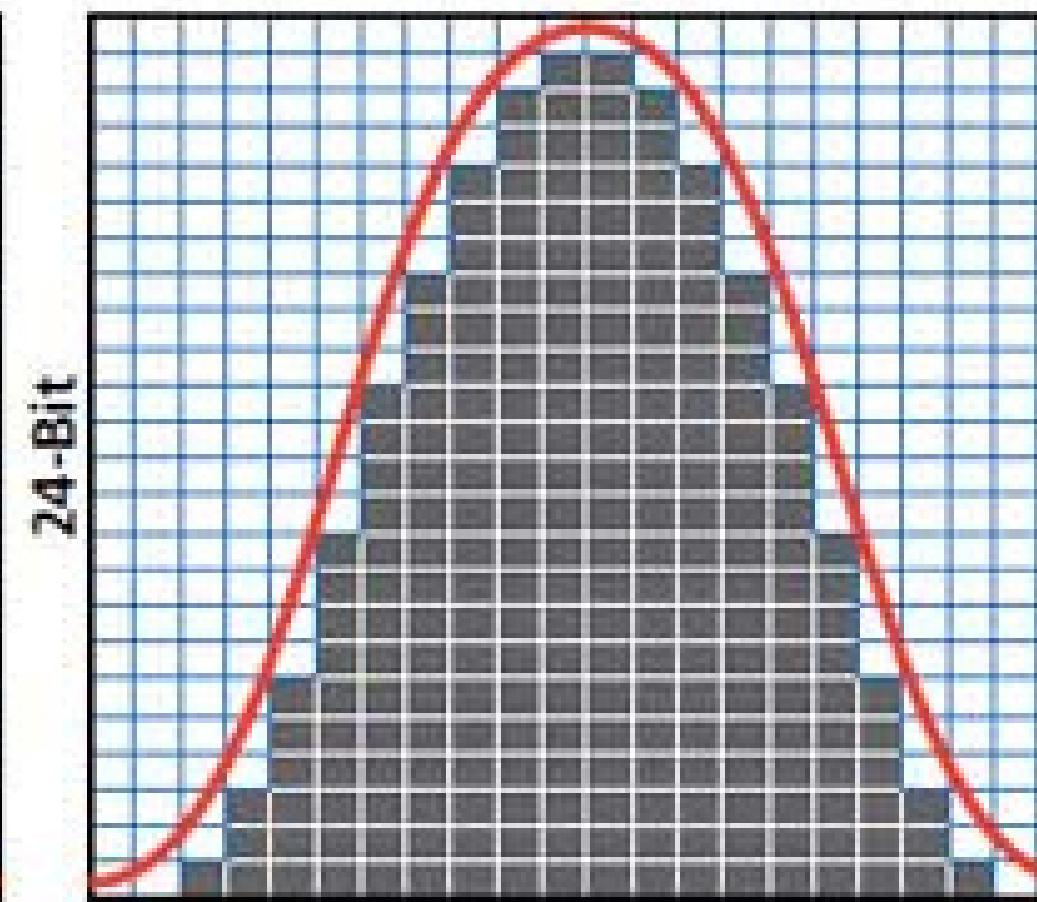
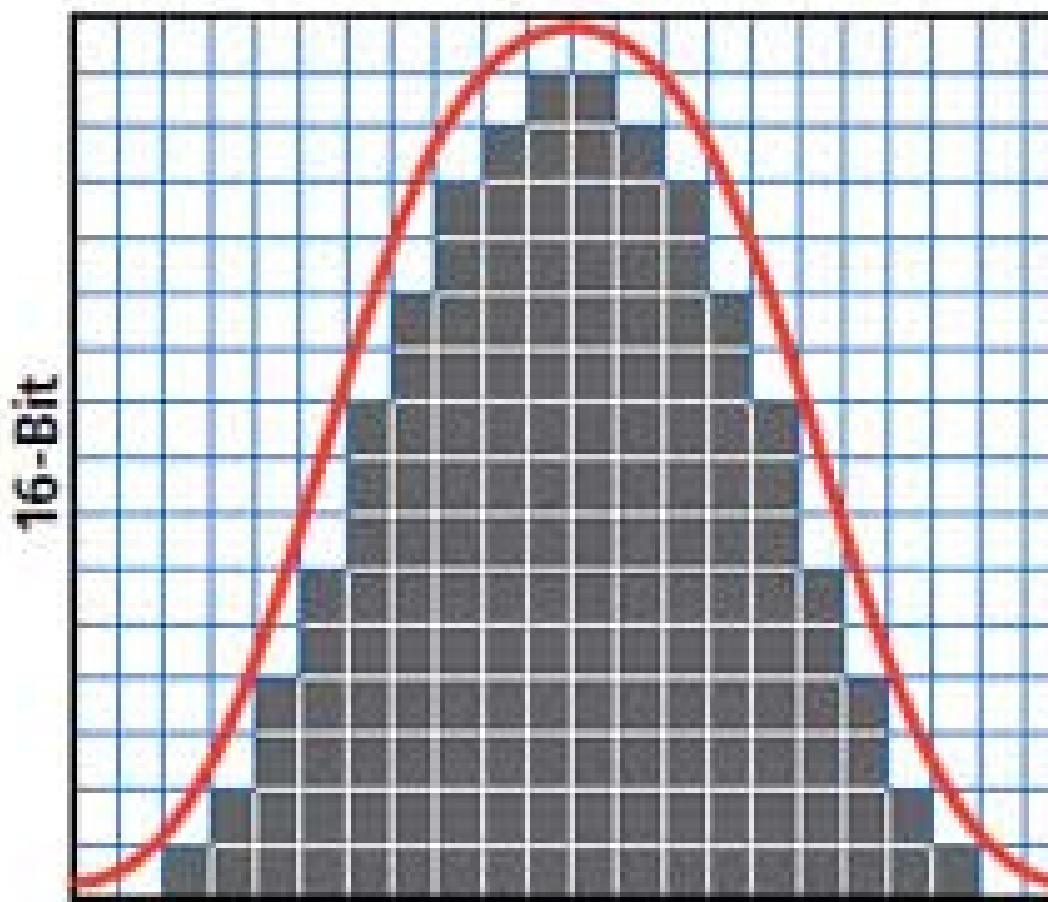
Storing Data Digitally: Conversion



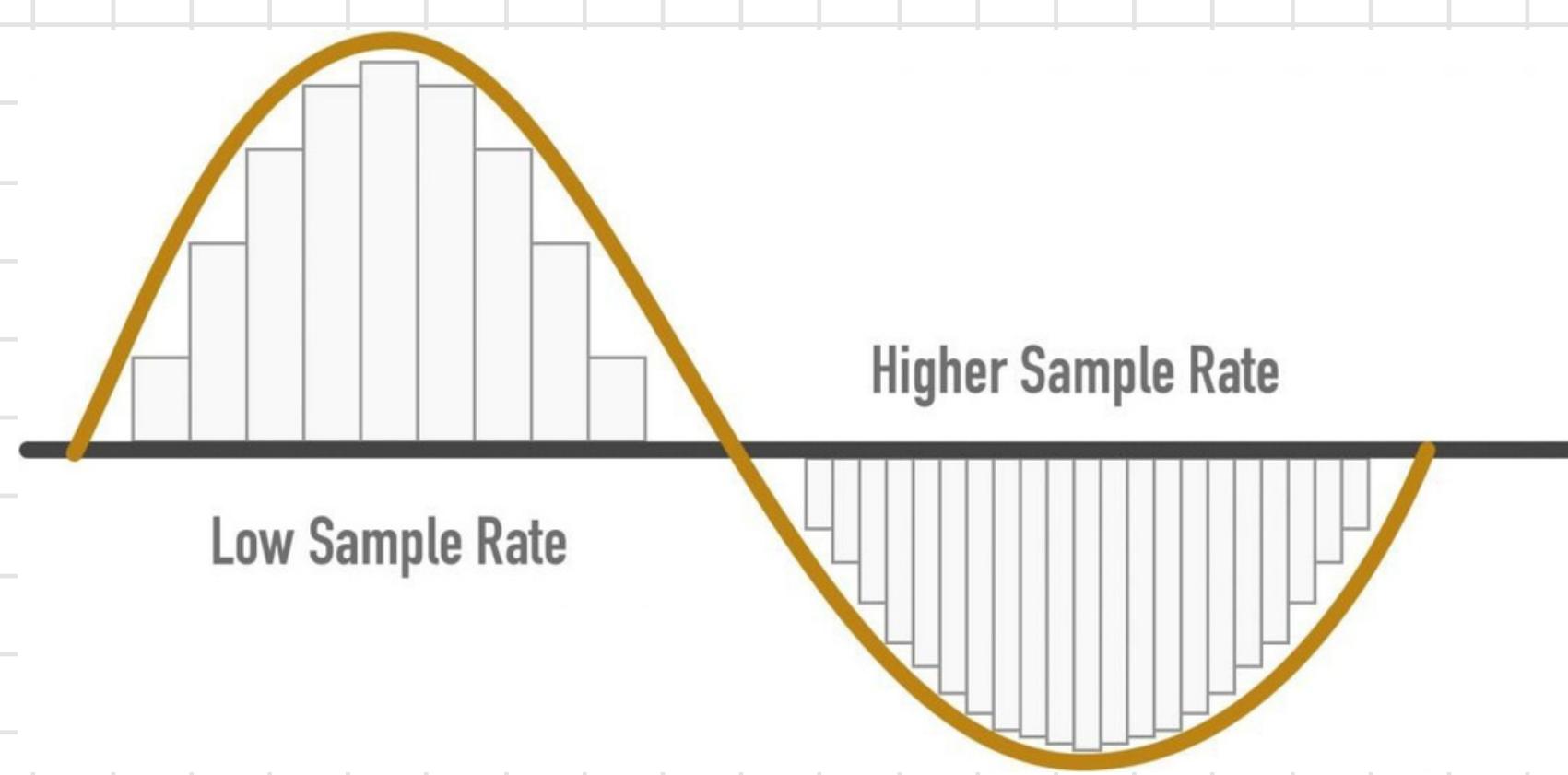
Sample Rate: Time Axis Resolution

Bit Depth: Amplitude Axis Resolution

Storing Data Digitally: Conversion



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



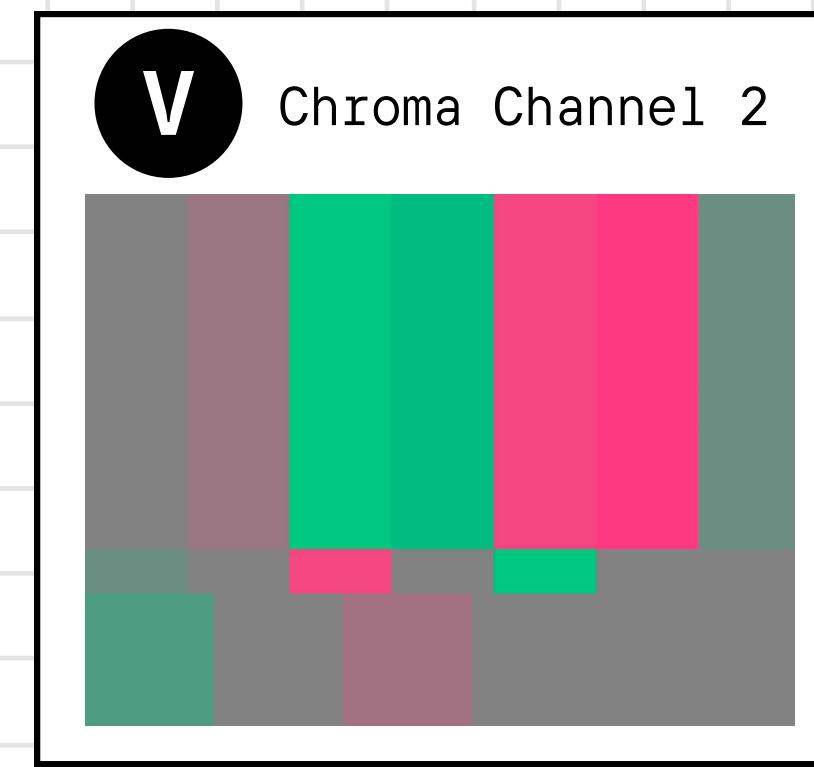
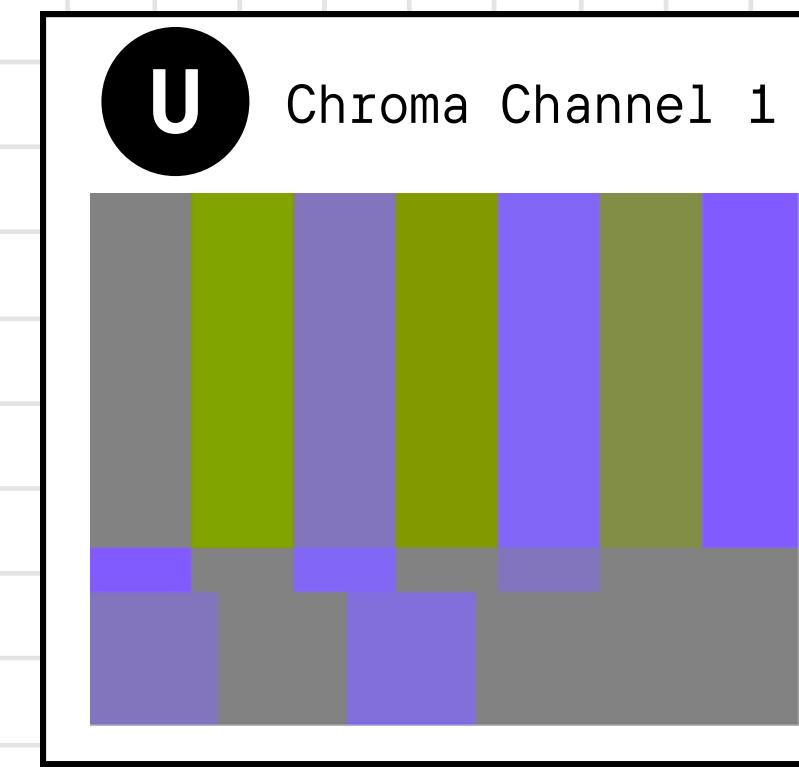
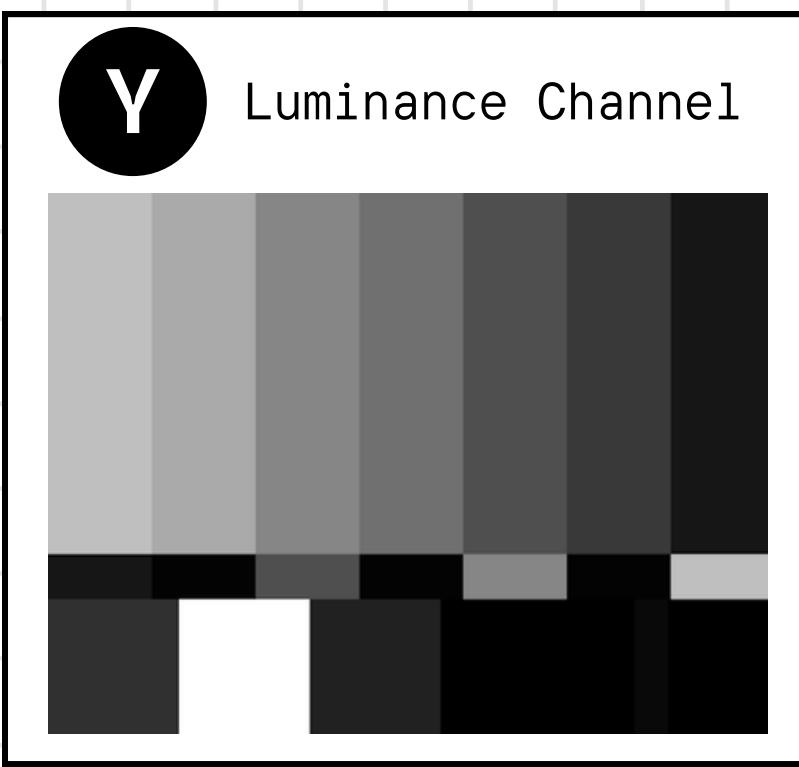
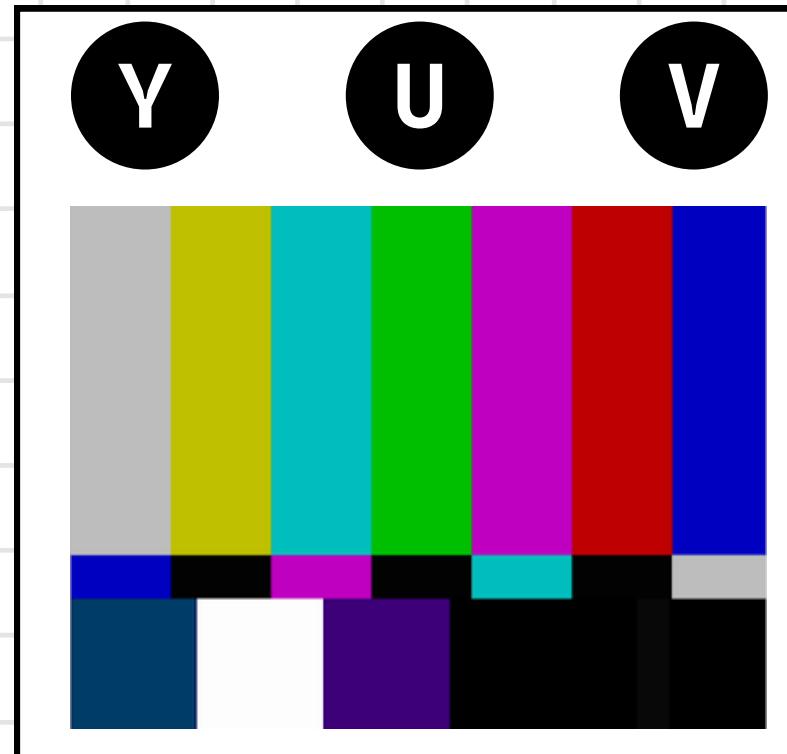
Seeing Bit Depth Visually

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

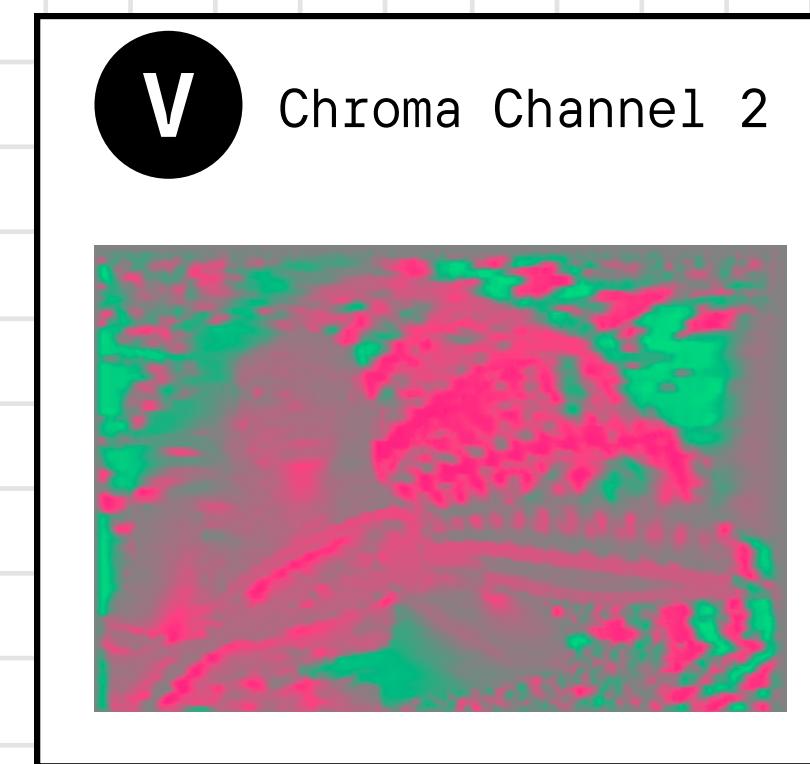
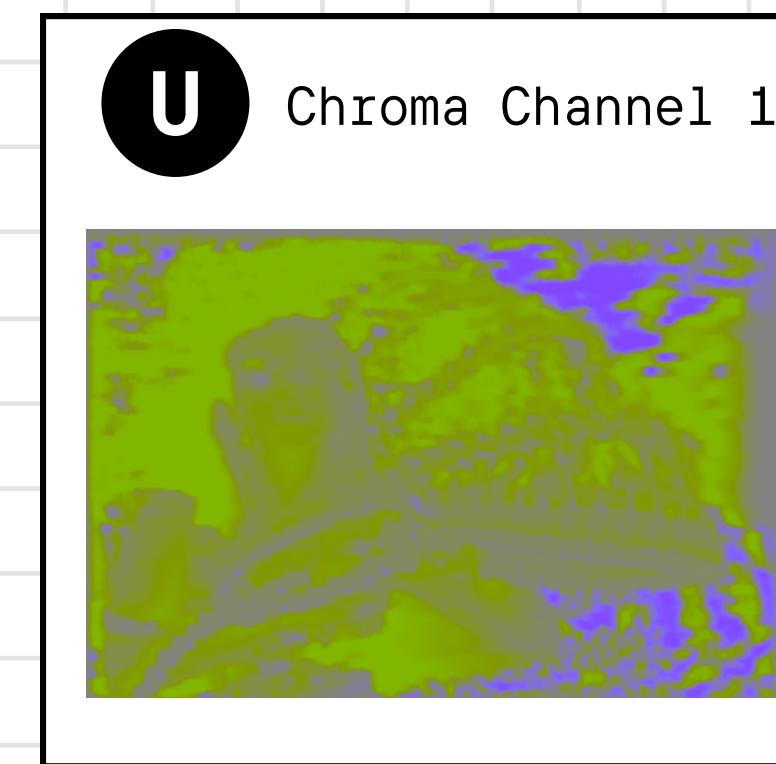
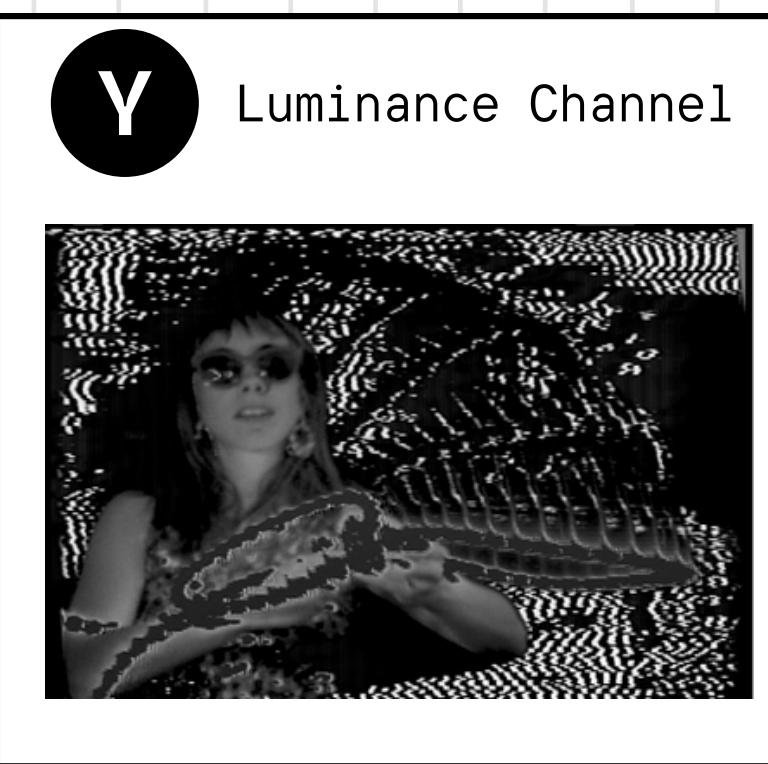
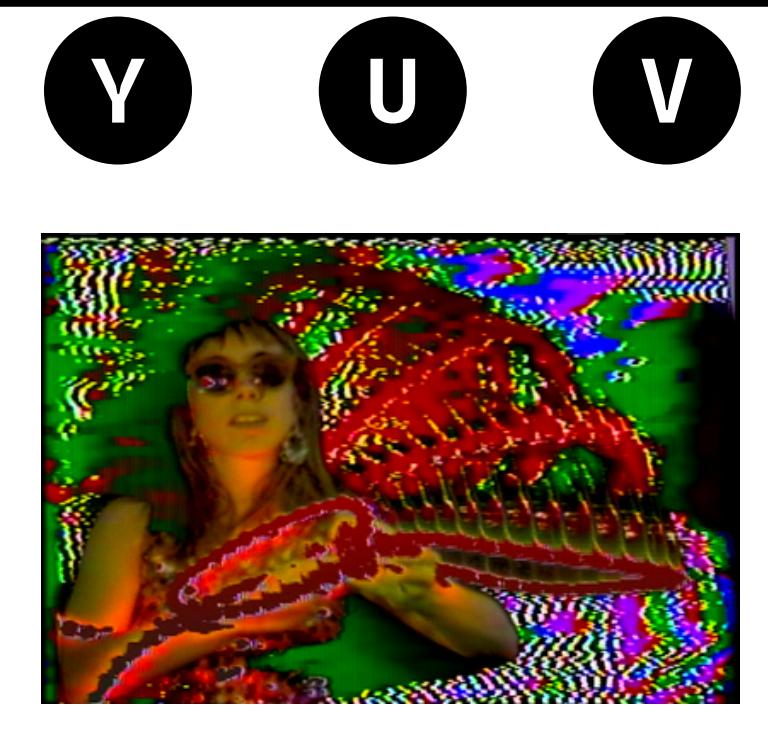


YUV Color Encoding





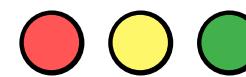
YUV Color Encoding



10-Bit Video Data Scope

10-Bit Video Data Scope

02D6	024E	0188	00DB	0007	00C7
02BD	02BD	01BA	01BA	0151	0161
0242	0242	02BE	02BE	0256	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

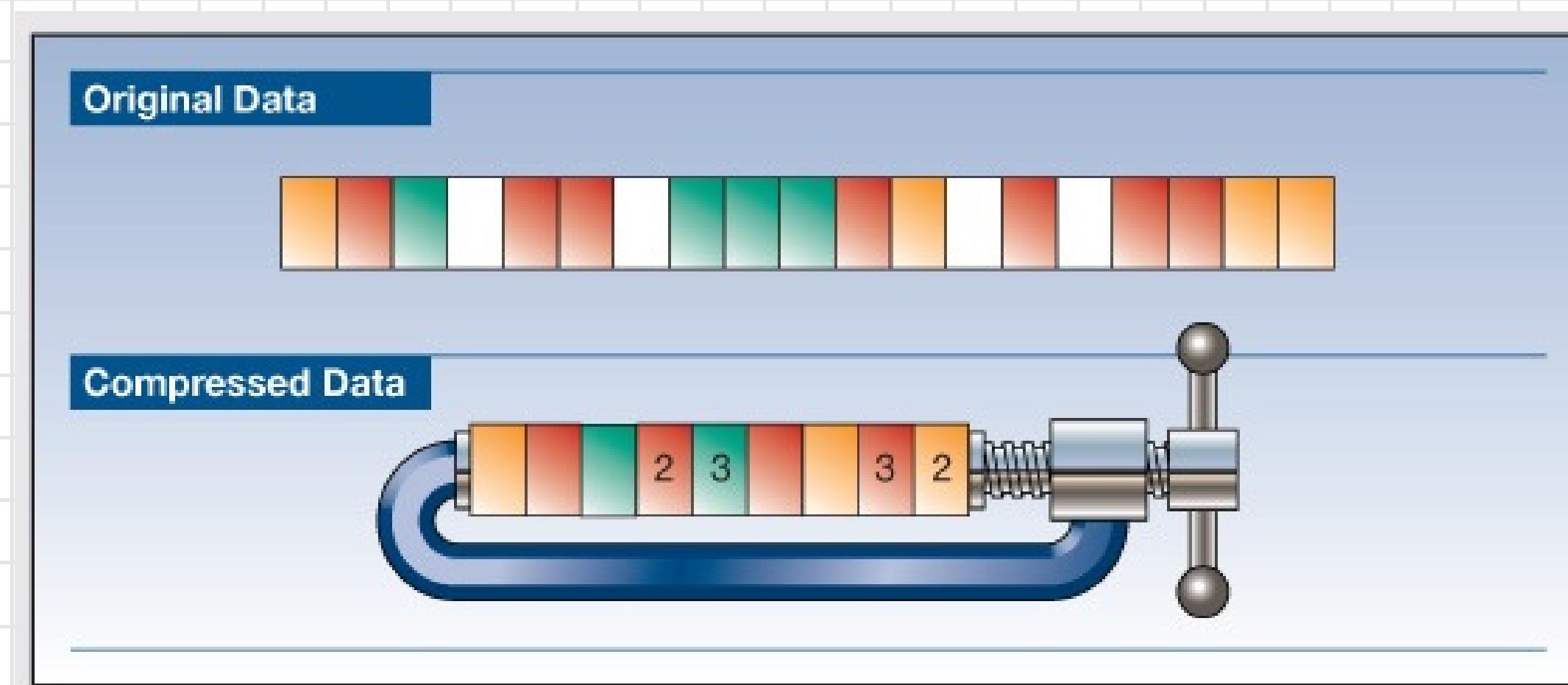


Codec vs Container

- 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

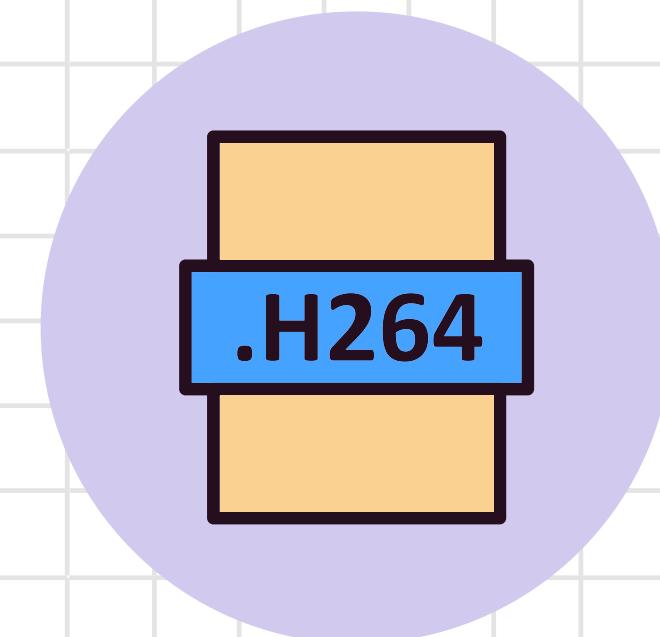


- Lossy
 - Some data is lost during compression and cannot be recovered
 - Common for "access files"
- Lossless
 - No data is lost during compression
 - All original data can be recovered

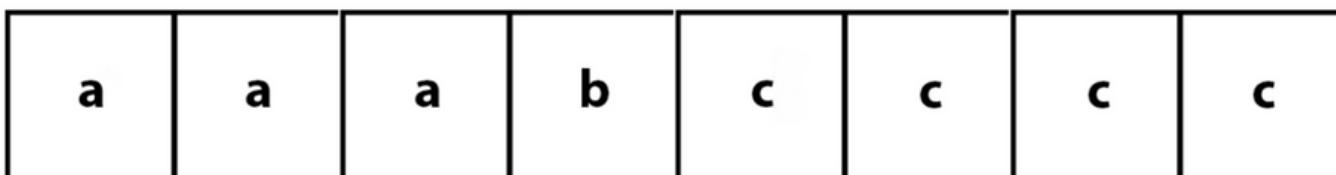
Codecs:
Lossy
vs
Lossless

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

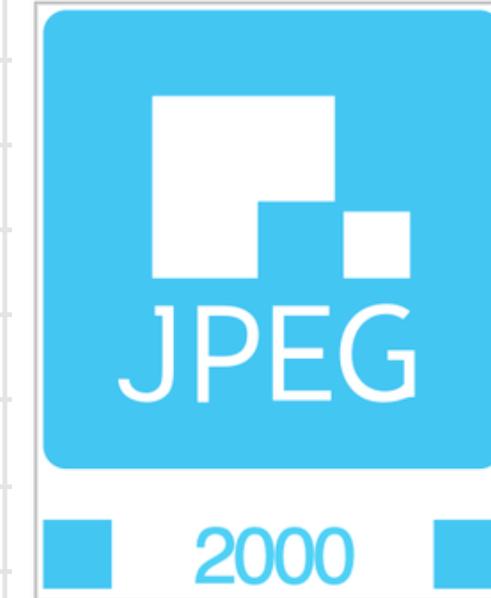


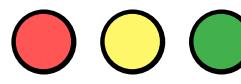
run-length encoding



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)





Exercise 01: Compression Comparison



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

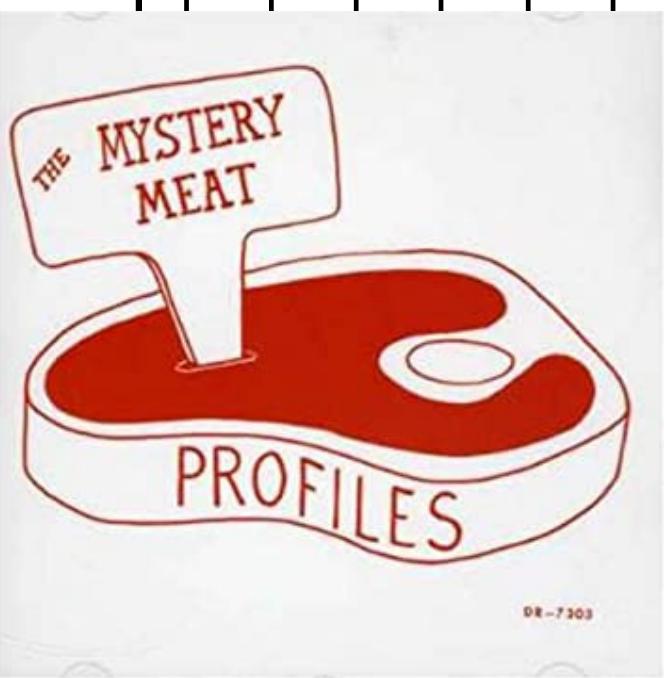


Format Longevity and Adoption

- 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 02: Media Mystery Meat



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 04: Round-Trip Transcode

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.

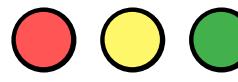
**Is the
digital
signal
properly
recorded?**



Does the recorded format match the desired specs

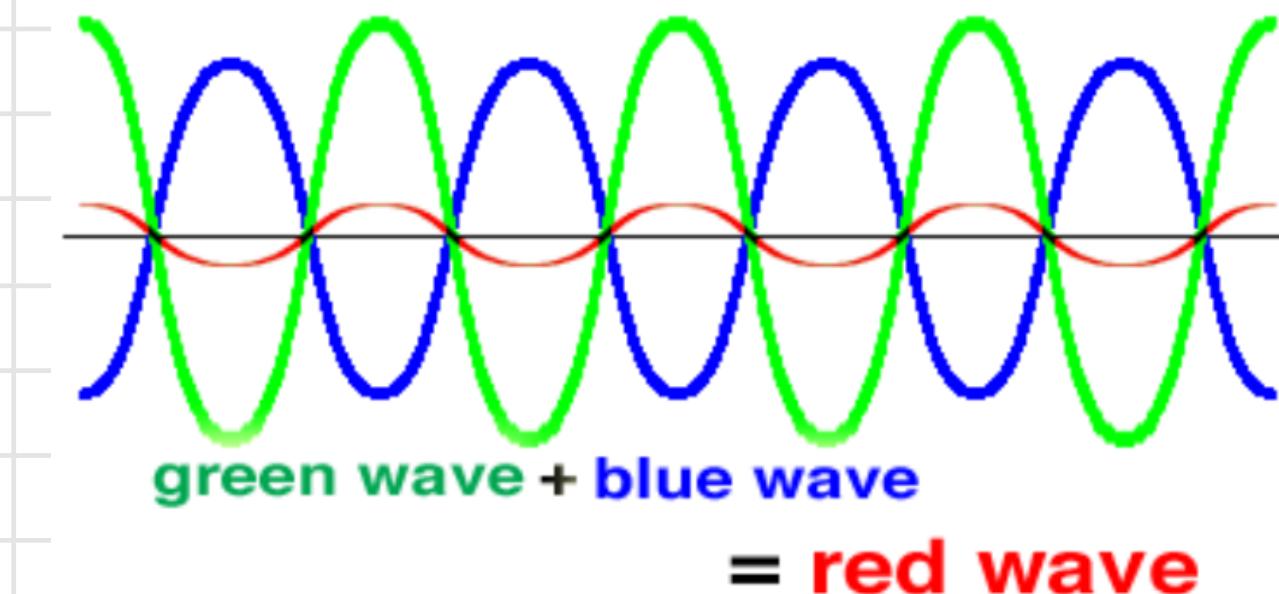
We need to test that the final file matches the desired specifications.

- Linear PCM
- 96kHz
- 24bit
- etc

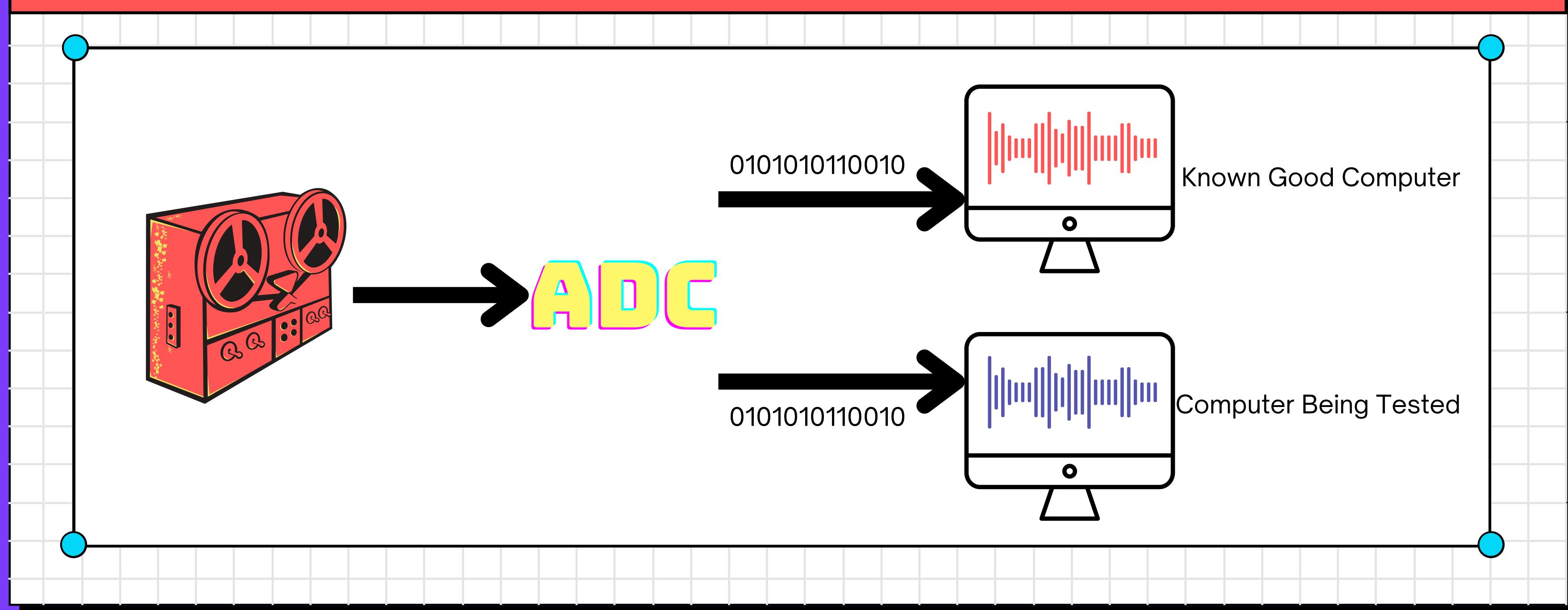


Audio Null Test for Station Validation

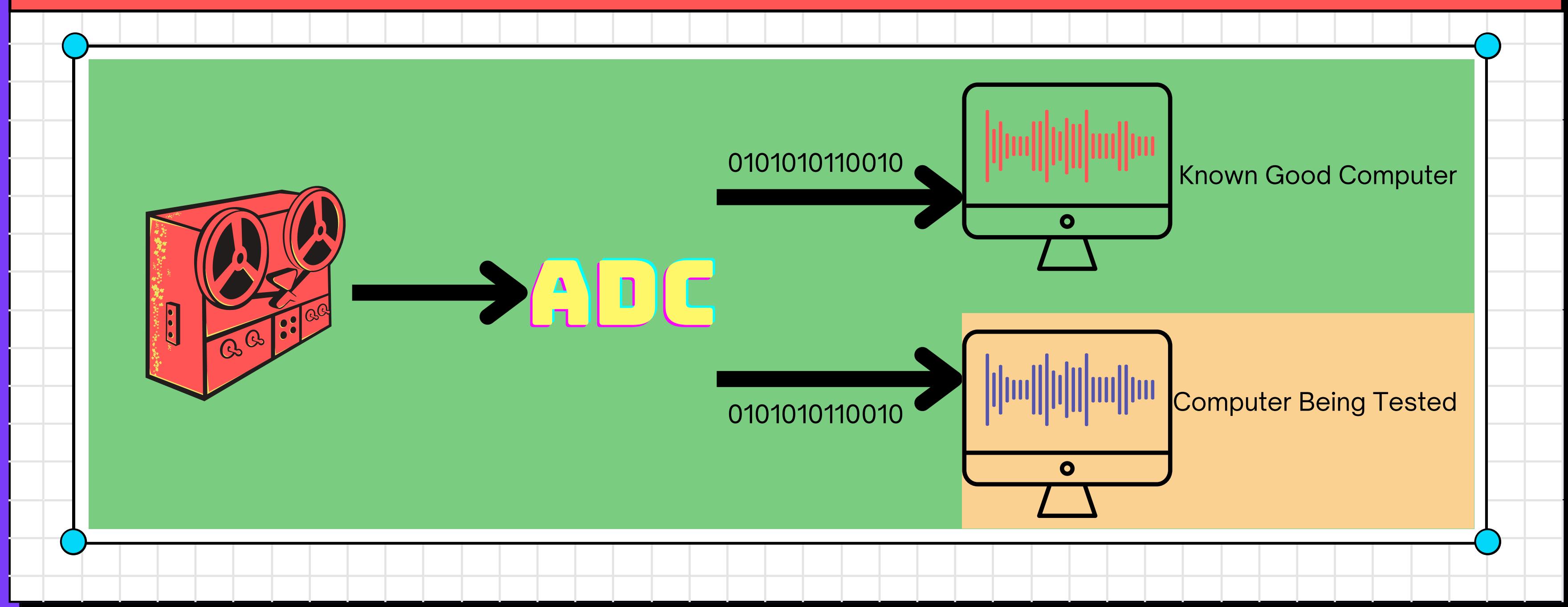
- Null tests ensure that the digitized signal is properly recorded onto the storage media
- Null tests work using the principal of wave addition and interference



Audio Null Test



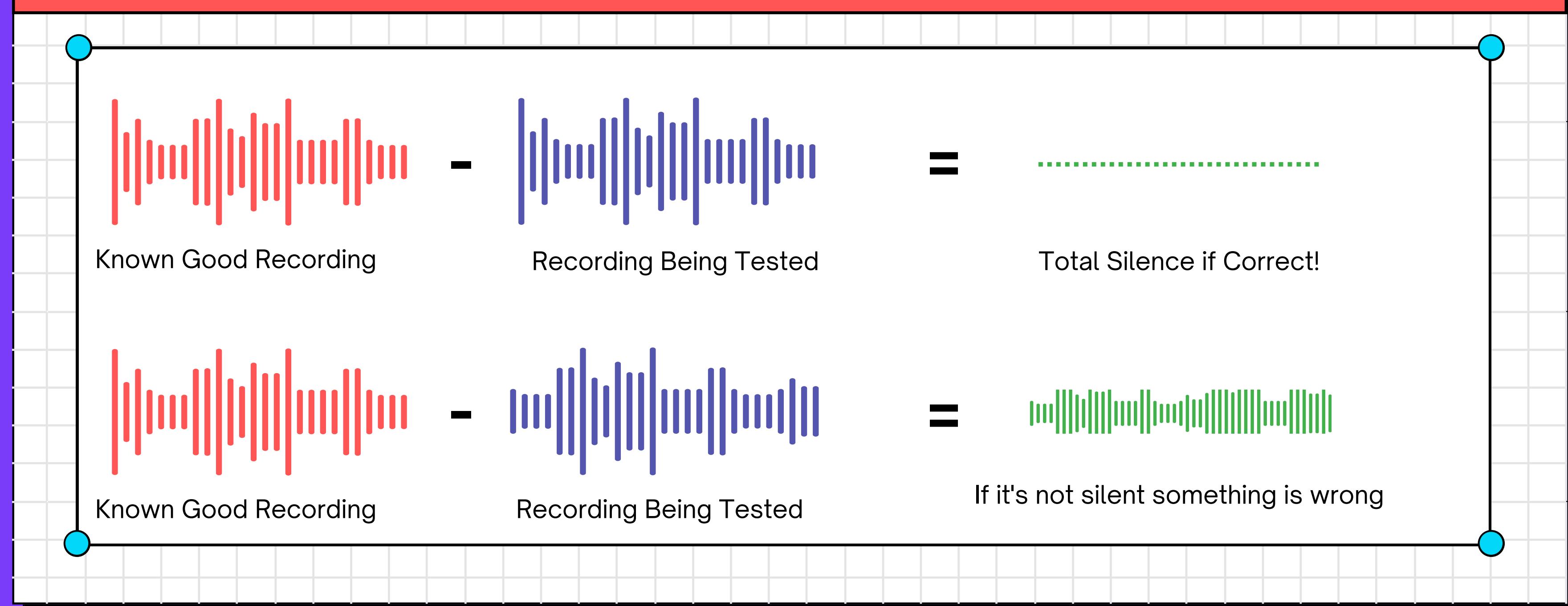
Audio Null Test

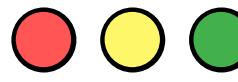


Assumed to be
working properly

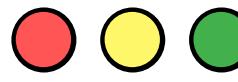
The item we are
testing

Audio Null Test





Exercise 05: Audio Station Qualification



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

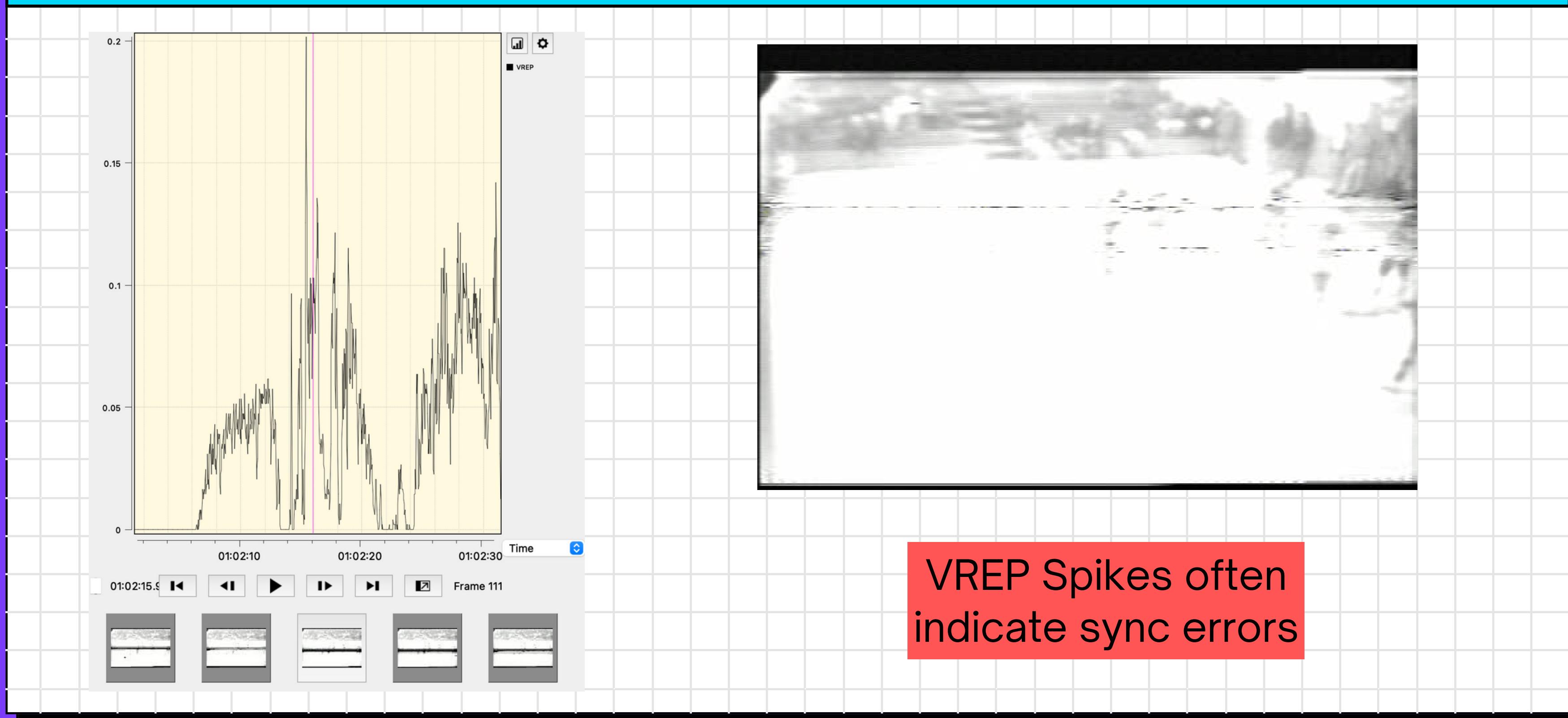


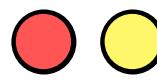
Is the analog signal properly converted to digital?

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

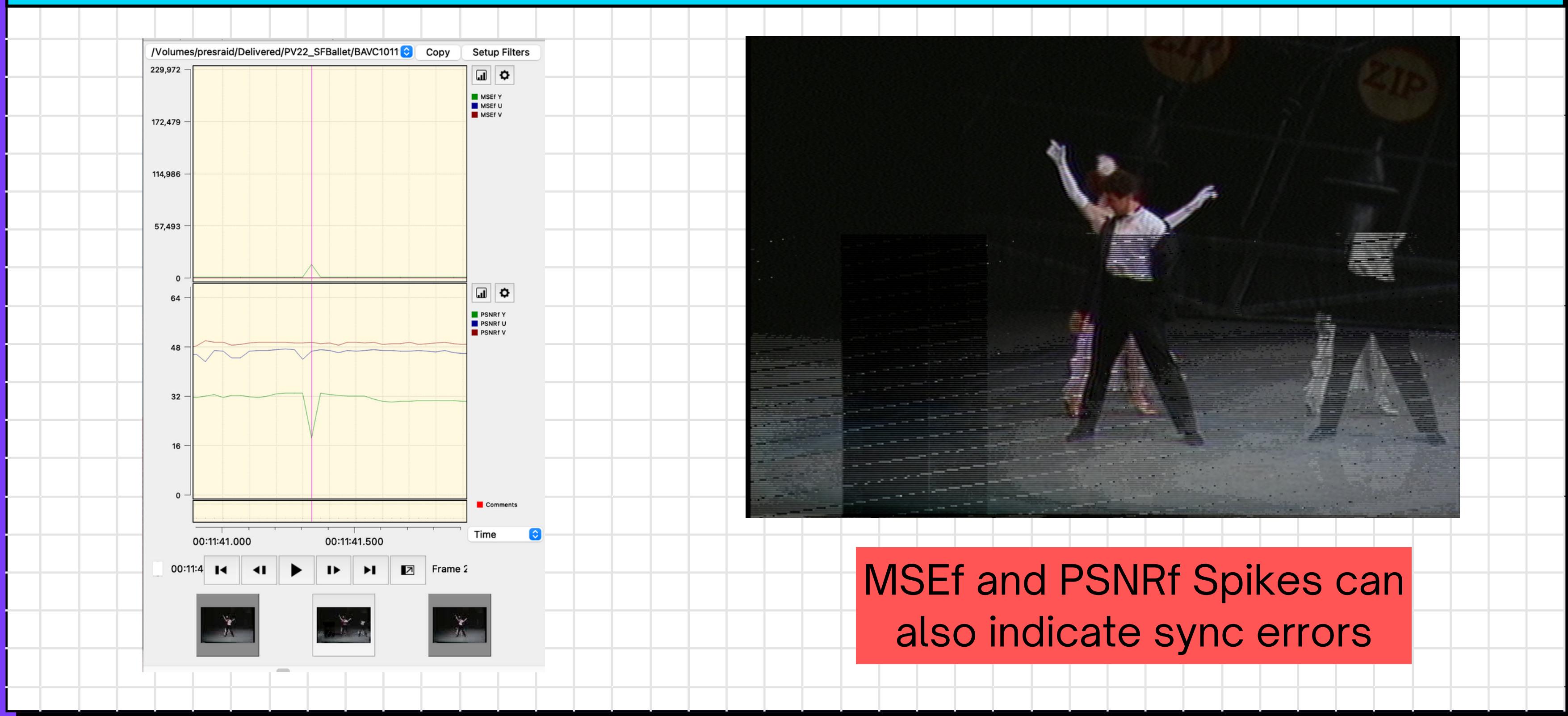
Create tests
for these
parameters
using
available
tools

Identifying Sync Errors in QCTools

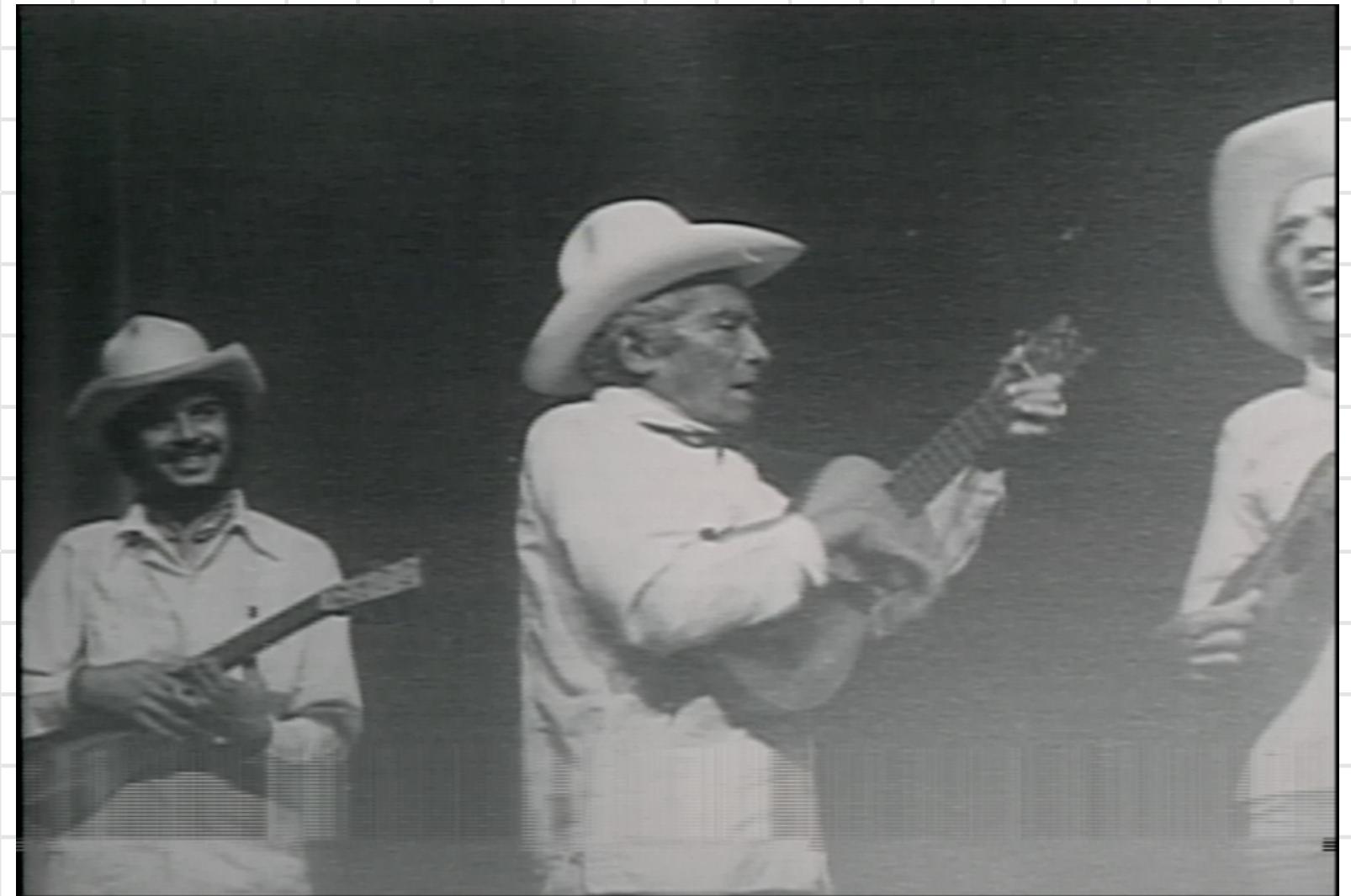




Identifying Sync Errors in QCTools

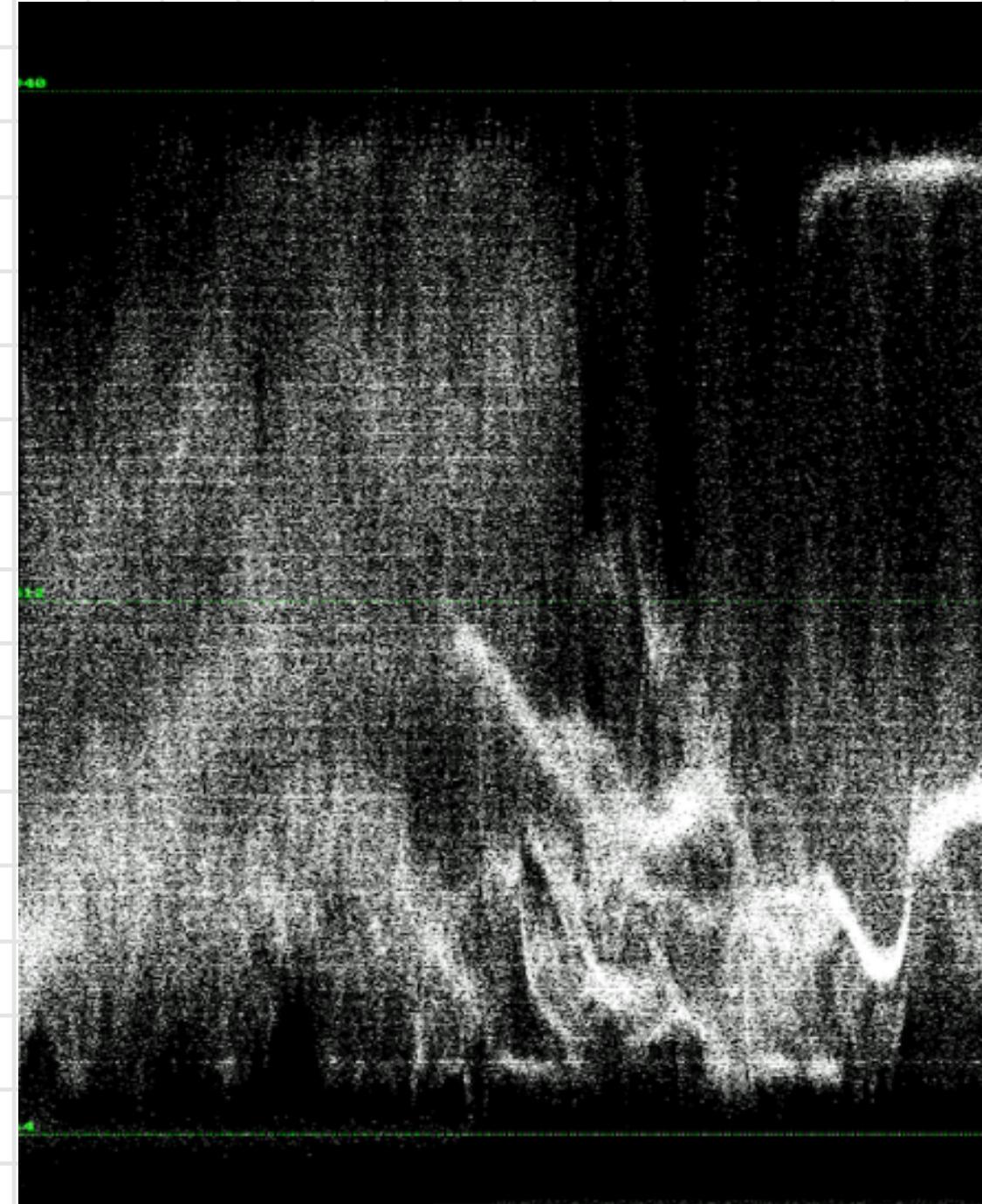


Identifying Repeated Lines in QCTools

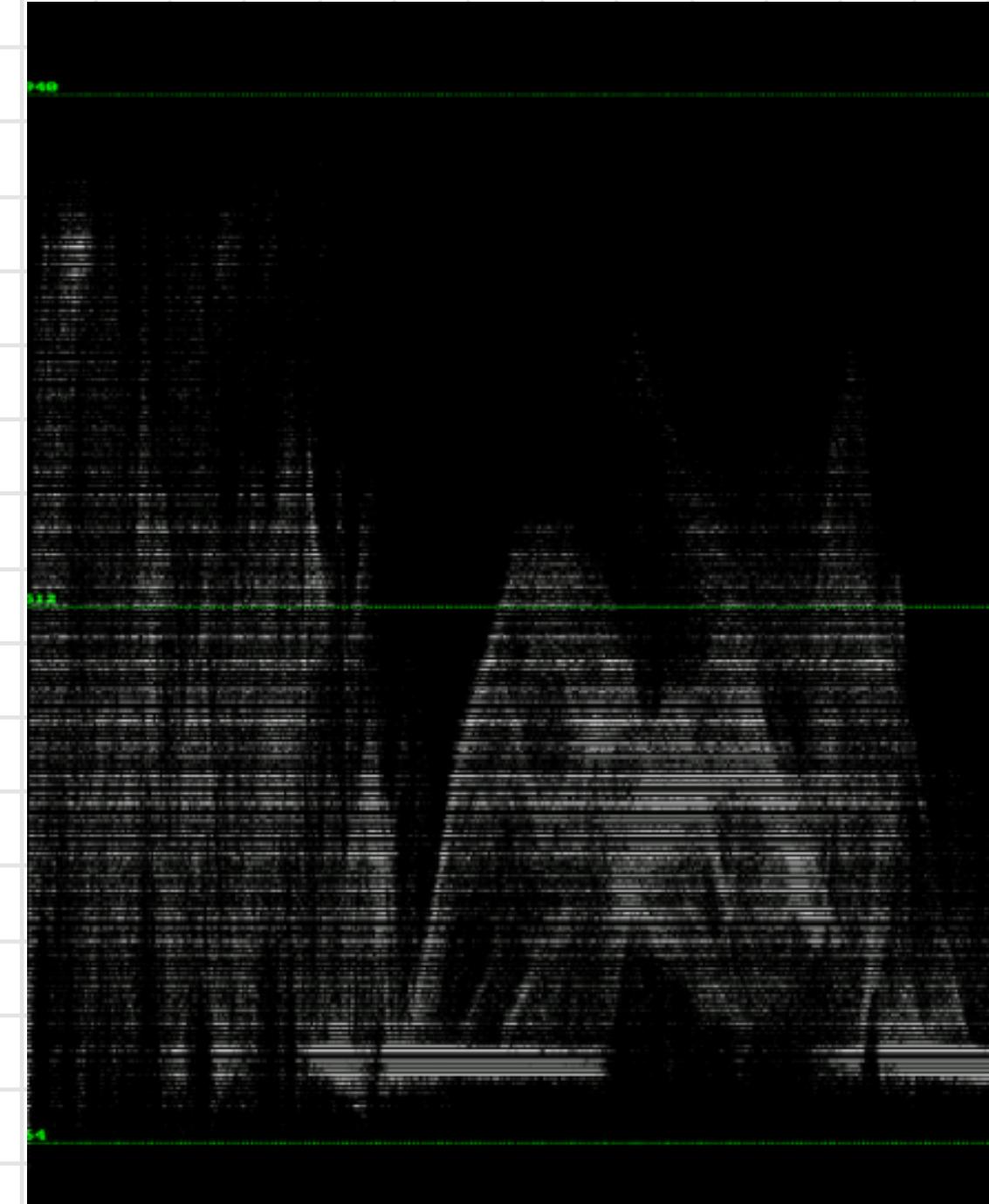


VREP Spikes can help find repeated lines often seen in poor Betacam Transfers

Identifying 10bit/8bit Errors in QCTools



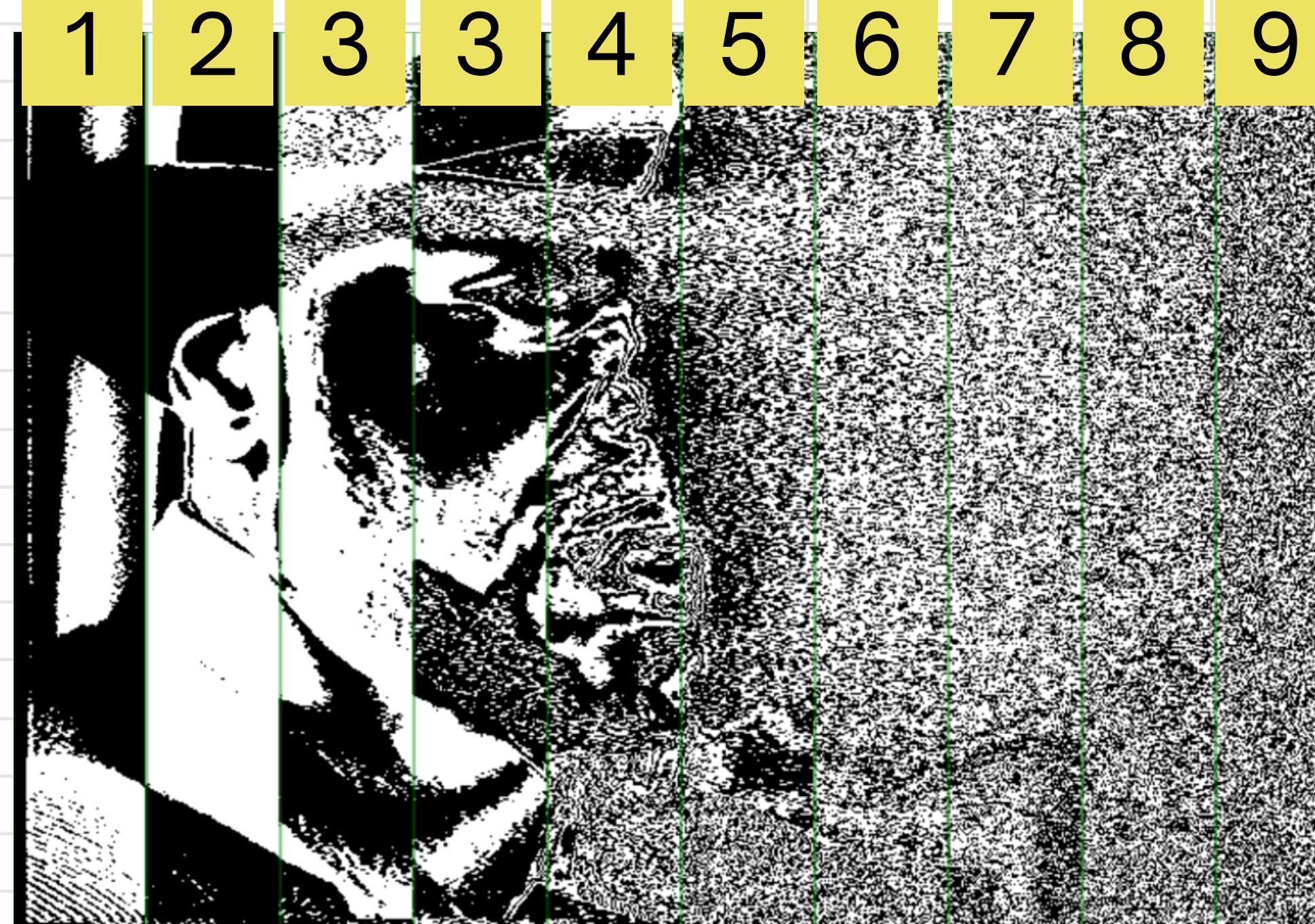
Proper 10 Bit Waveform



8 Bits of Data in 10 Bit Scale

Identifying 10bit/8bit Errors in QCTools

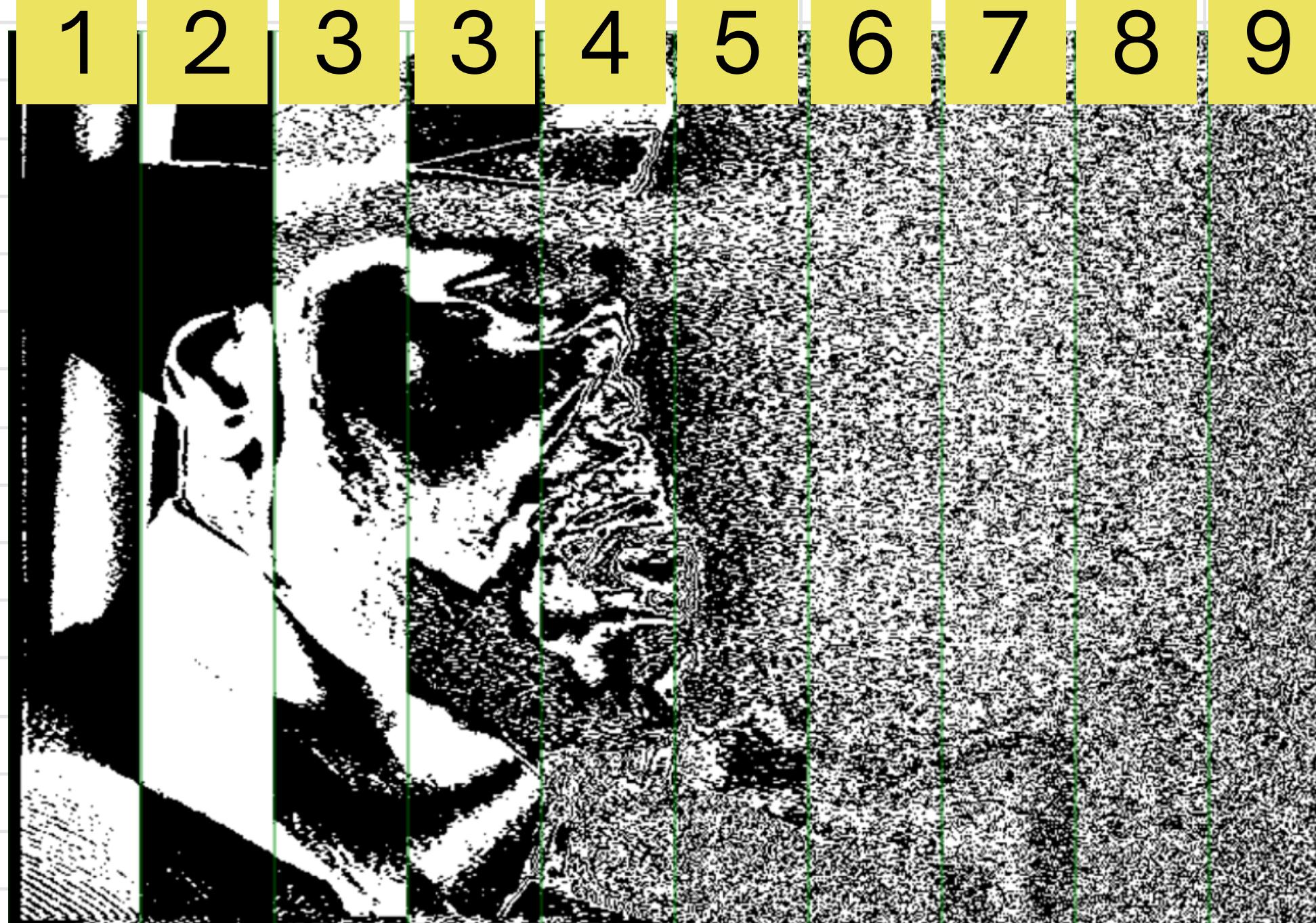
Bit
Position



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

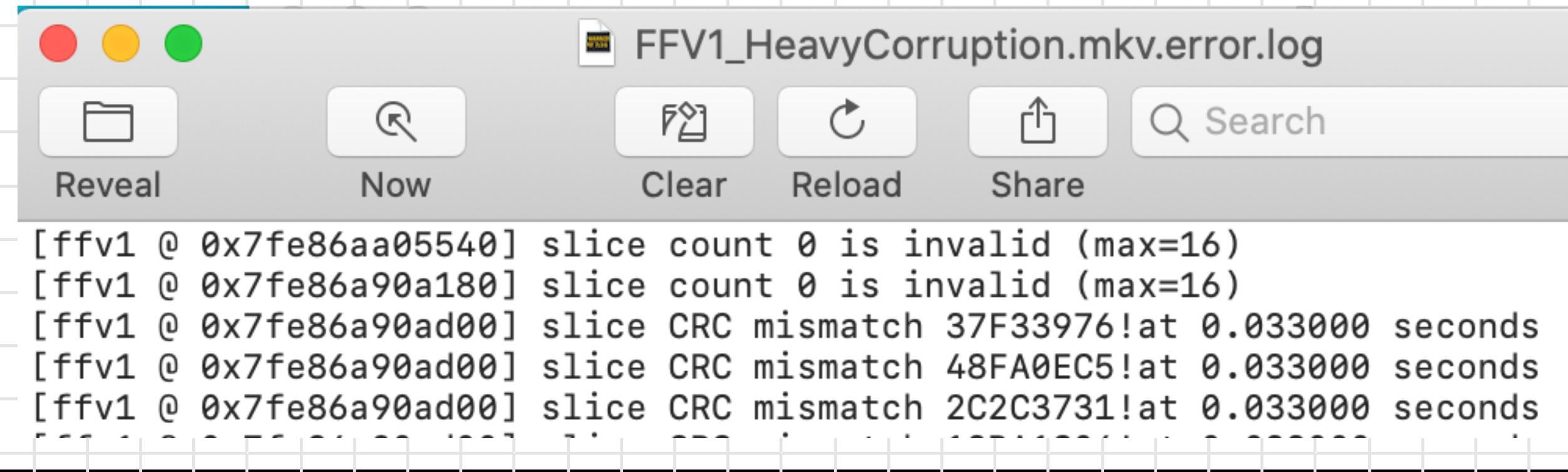
Files	Implen
FFV1.mkv	✓ Valid
FFV1_HeavyCorruption.mkv	✗ Not valid
FFV1_LightCorruption.mkv	✓ Valid
FFV1_MediumCorruption.mkv	✓ Valid

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





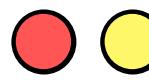
Does the recorded format match the desired specifications

- 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

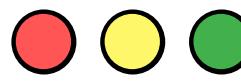
File: /Volumes/SymplyUltra/TransferProjects/PV21_SJSU/04_Loading/Drive07/Files_Mezzanine/BAVC1013138_214_mezzanine.mov

▼ 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 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!