



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

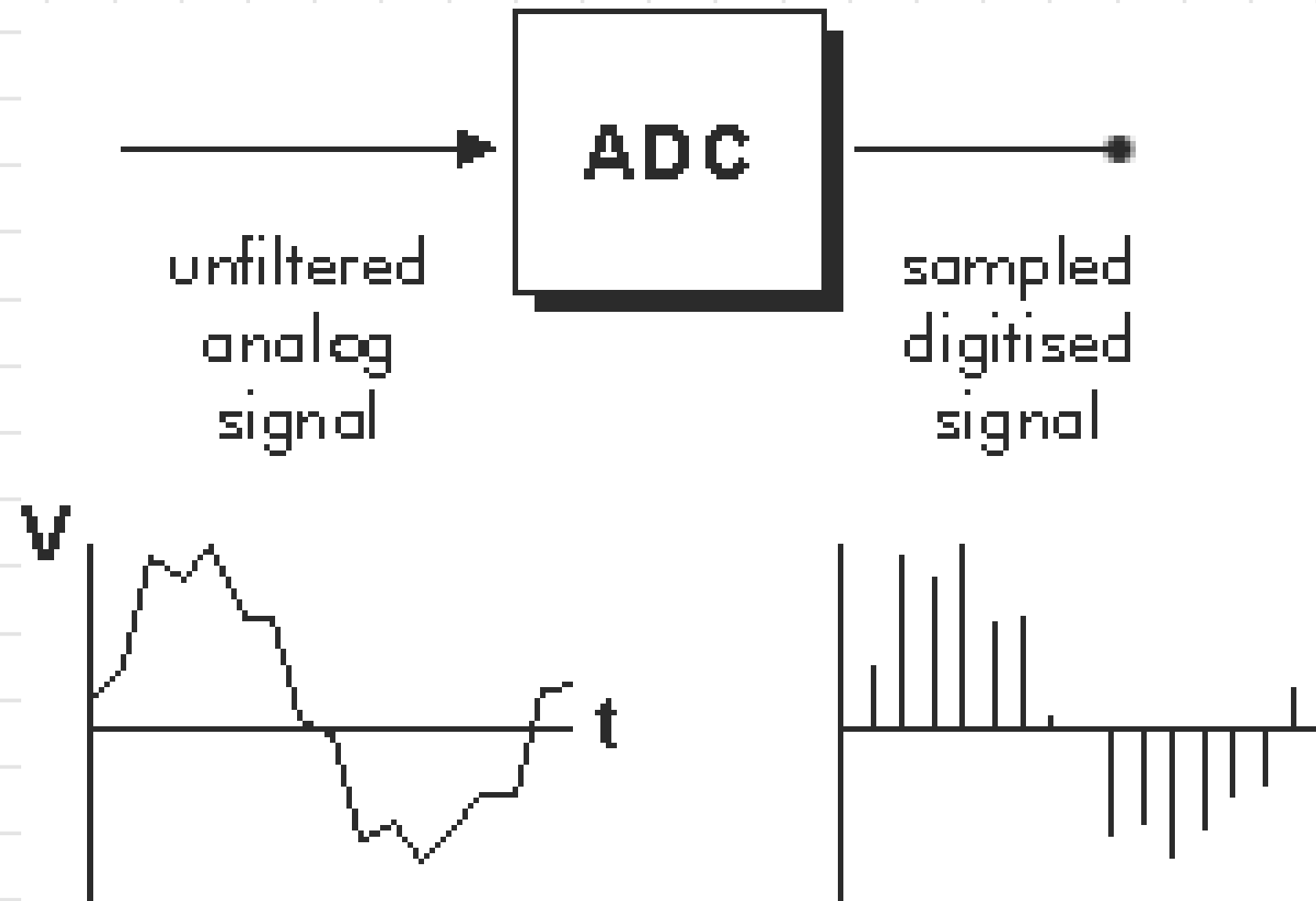
Standard Digital Preservation techniques are needed, plus extra work to deal with AV complexities



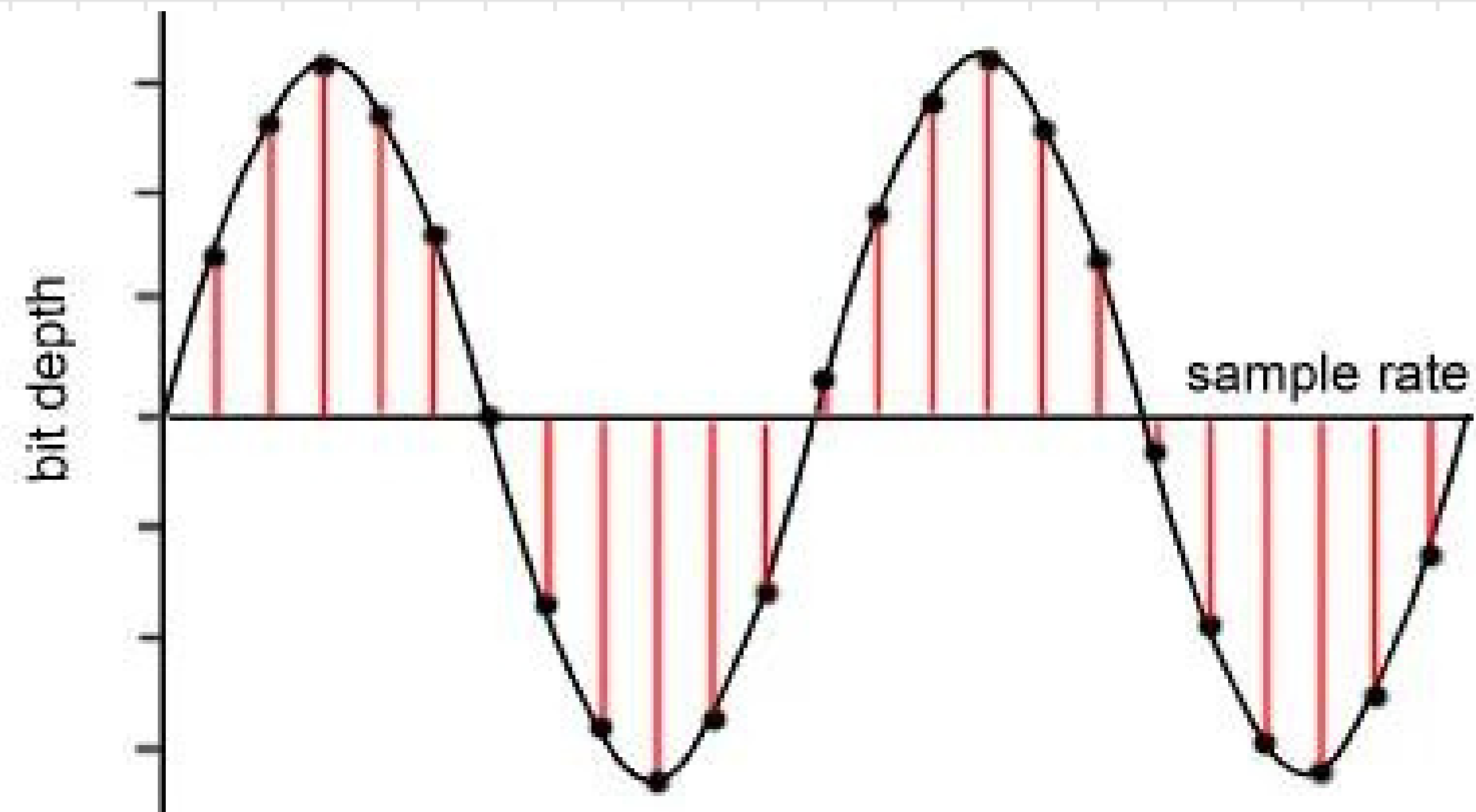
What Preservation Formats need to be able to do

- 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

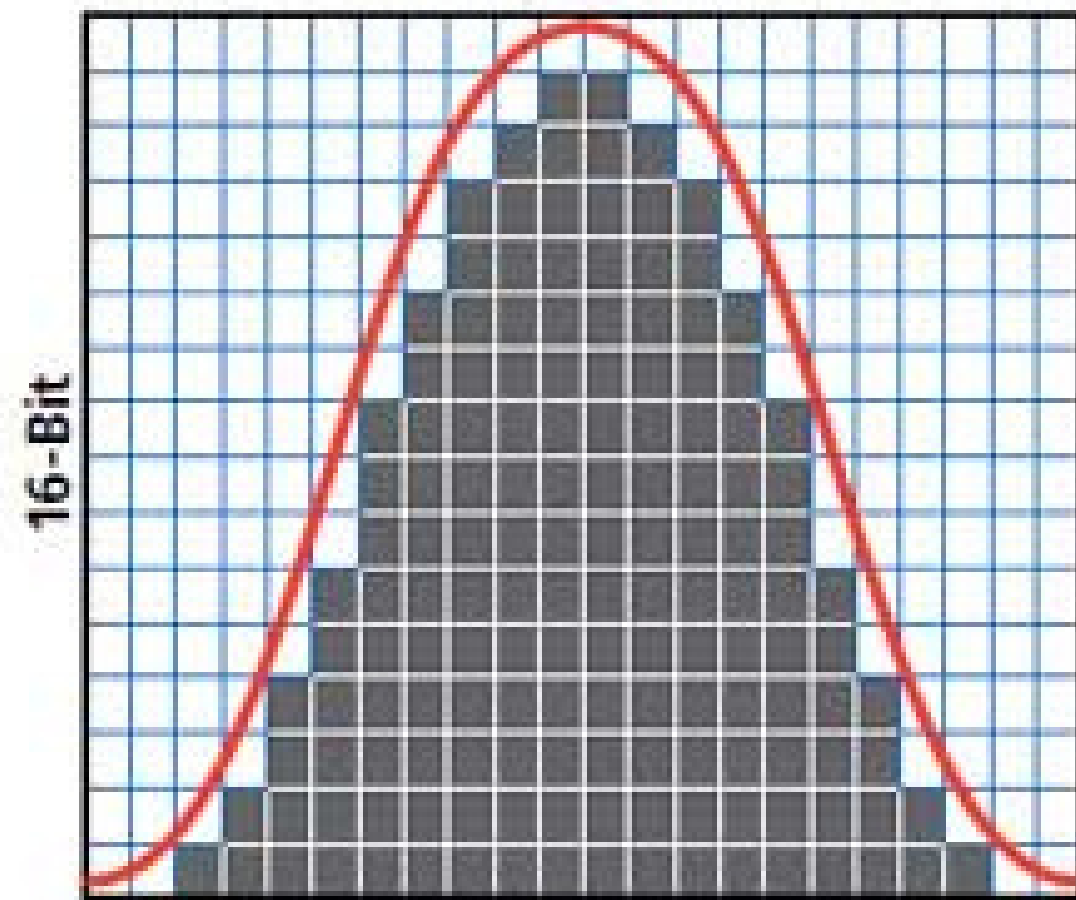
Storing Data Digitally: Conversion



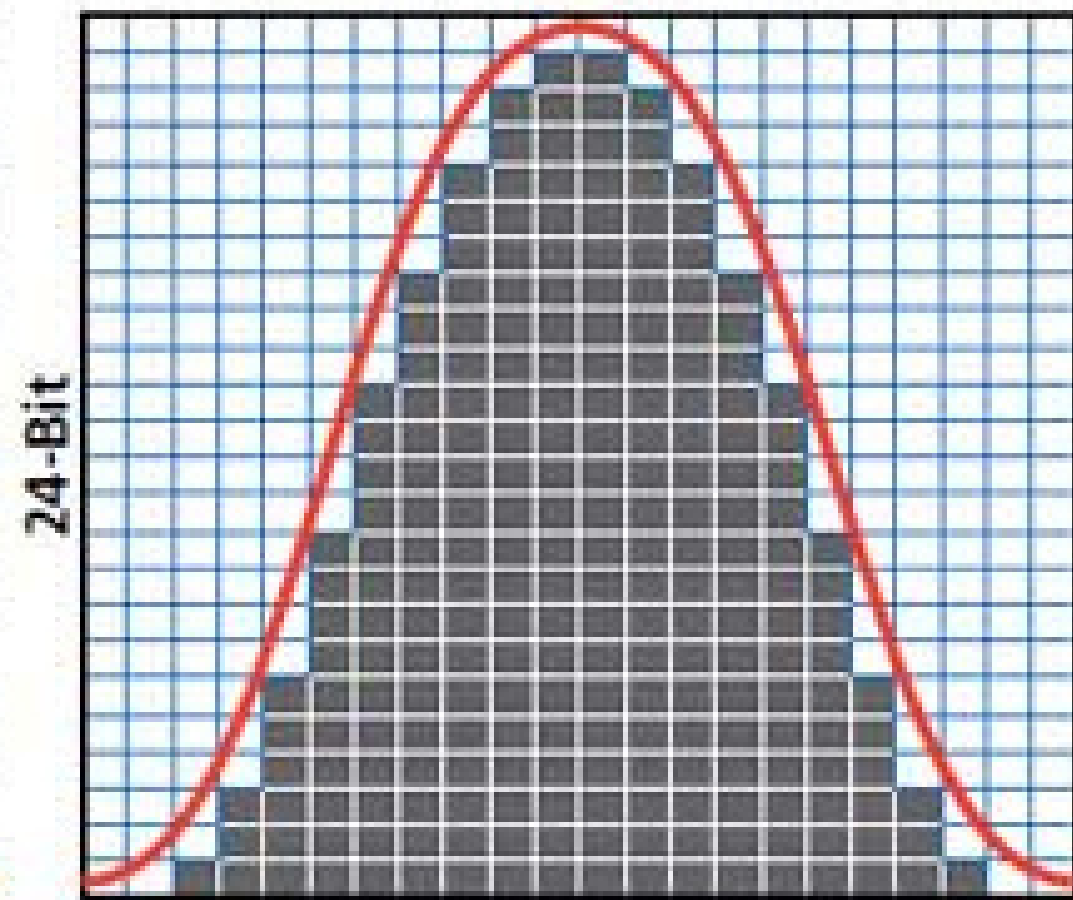
Storing Data Digitally: Conversion



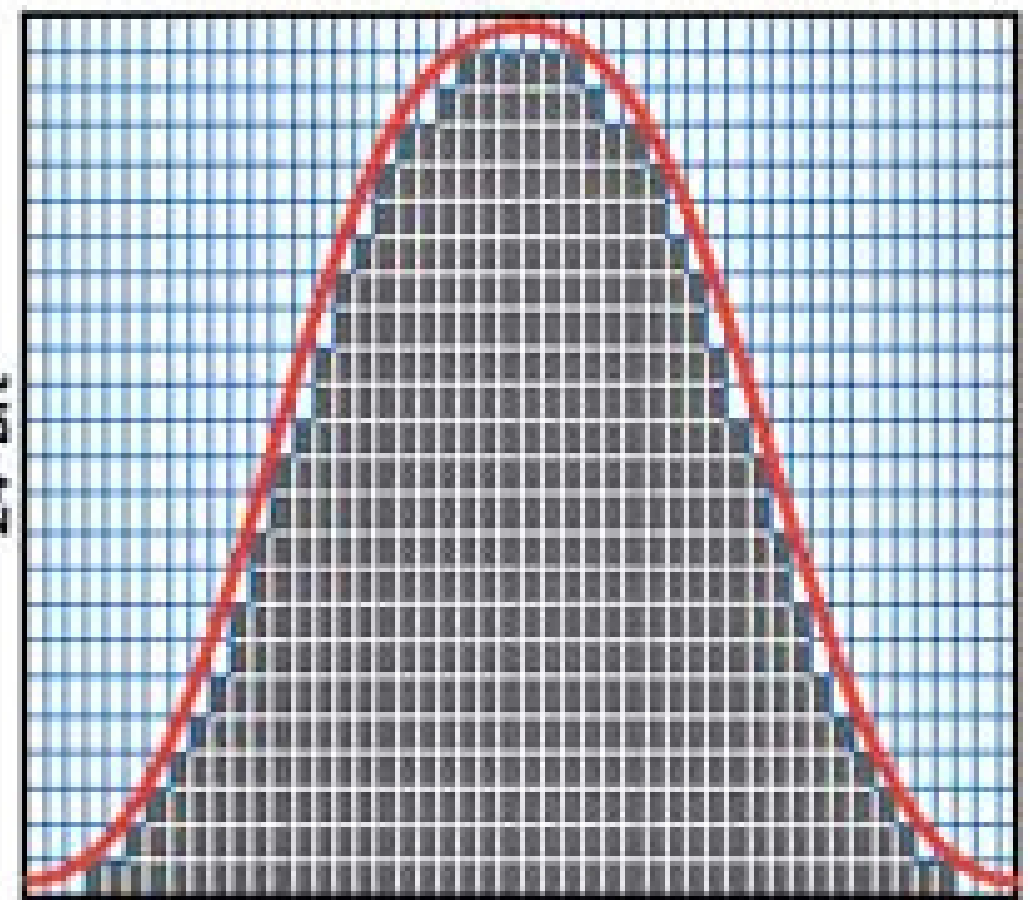
Storing Data Digitally: Conversion



44kHz

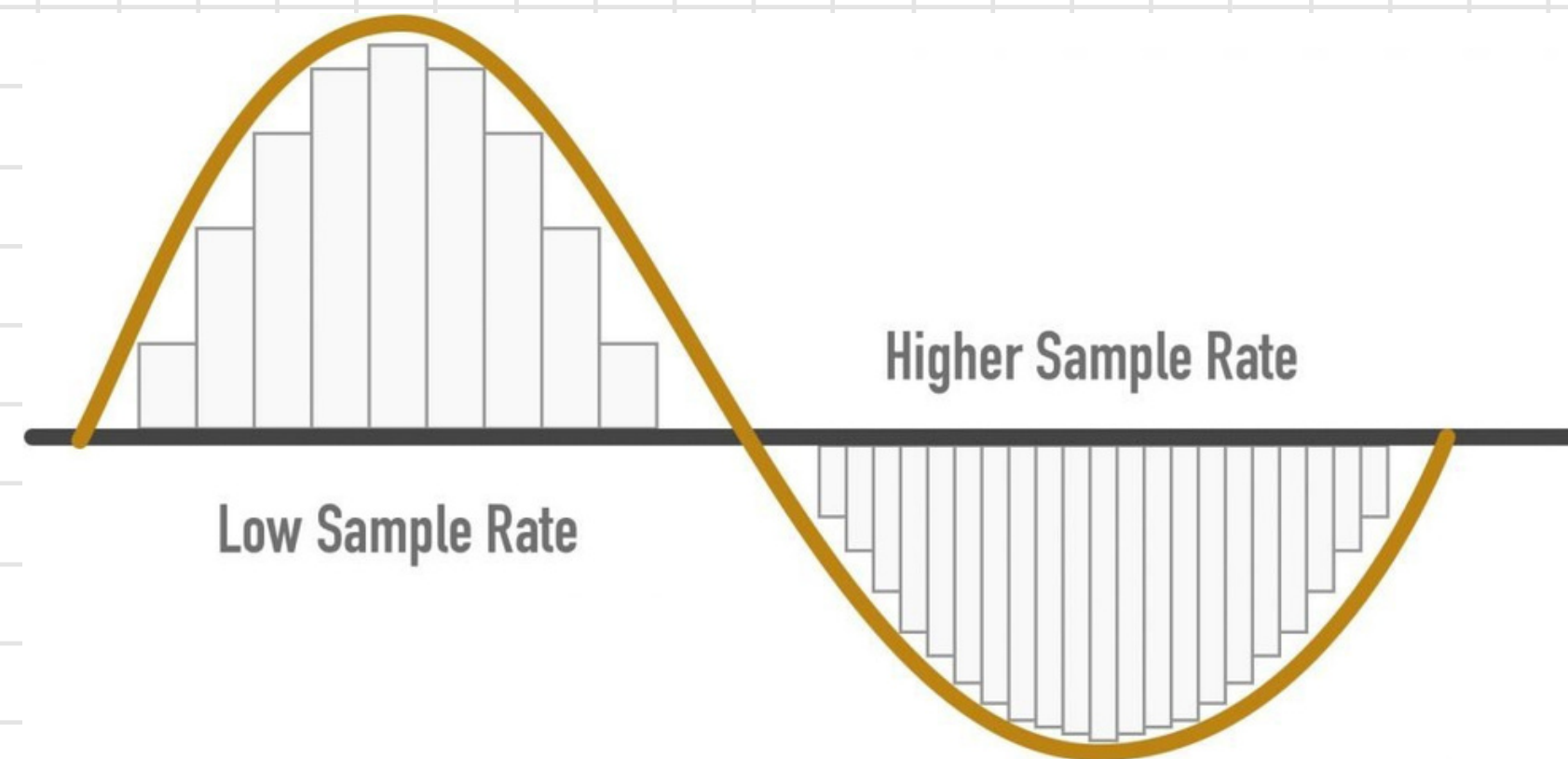


44kHz

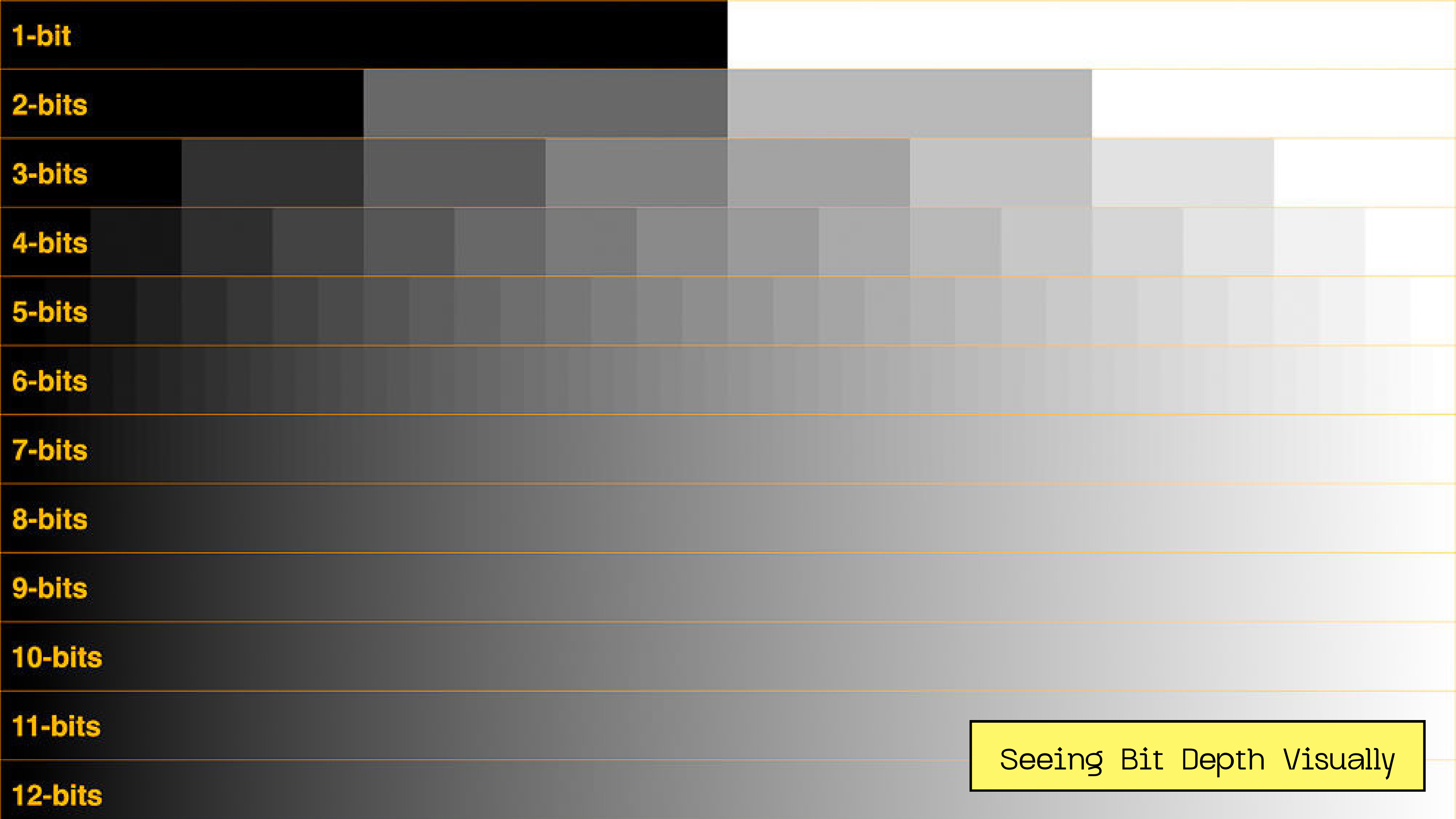


96kHz

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 40kHz to capture the upper limits of human hearing



1-bit

2-bits

3-bits

4-bits

5-bits

6-bits

7-bits

8-bits

9-bits

10-bits

11-bits

12-bits

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

Binary Values

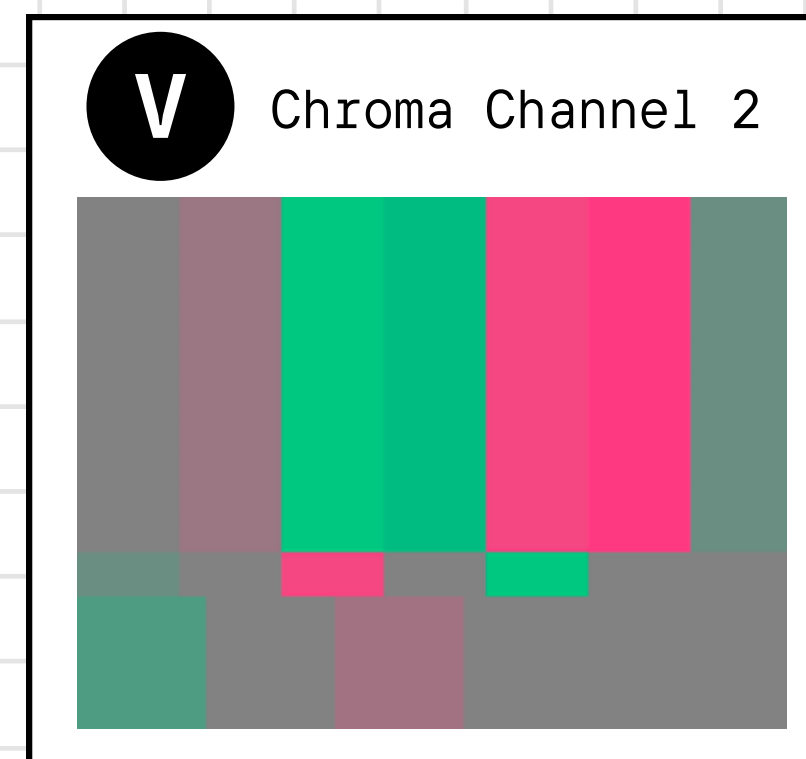
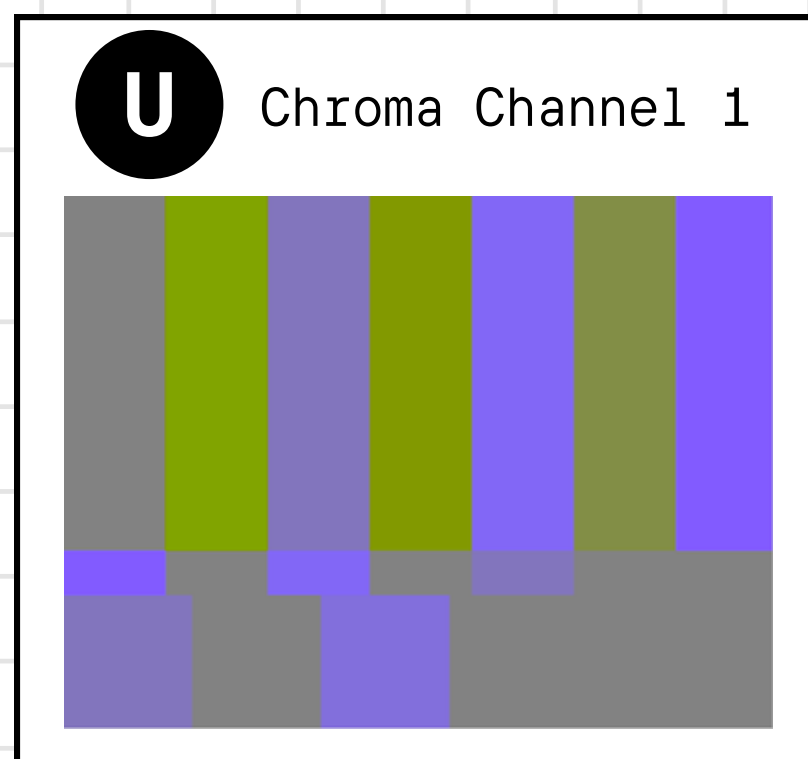
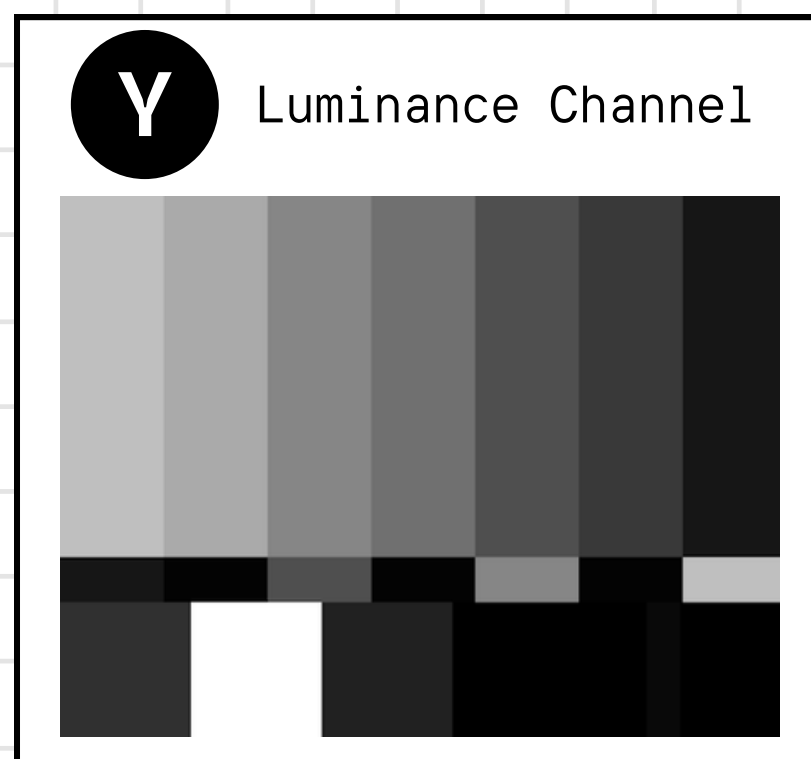
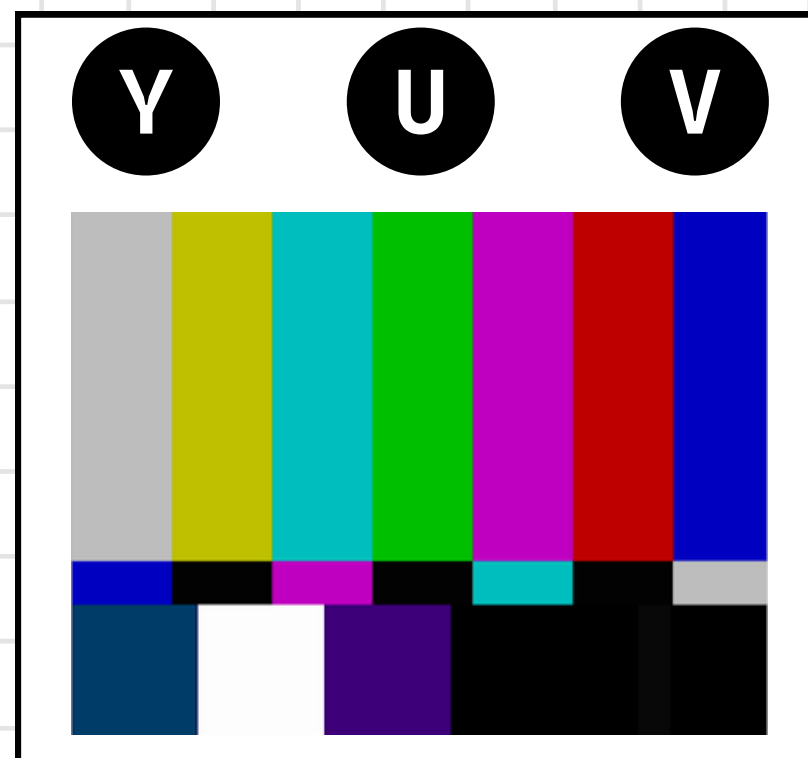
1-bit	0							1																											
2-bits	0			1				2					3																						
3-bits	0	1		2		3		4		5		6		7																					
4-bits	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15																				
5-bits	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31					
6-bits																																			
7-bits																																			
8-bits																																			
9-bits																																			
10-bits																																			
11-bits																																			
12-bits																																			

Decimal Values

Decimal Values

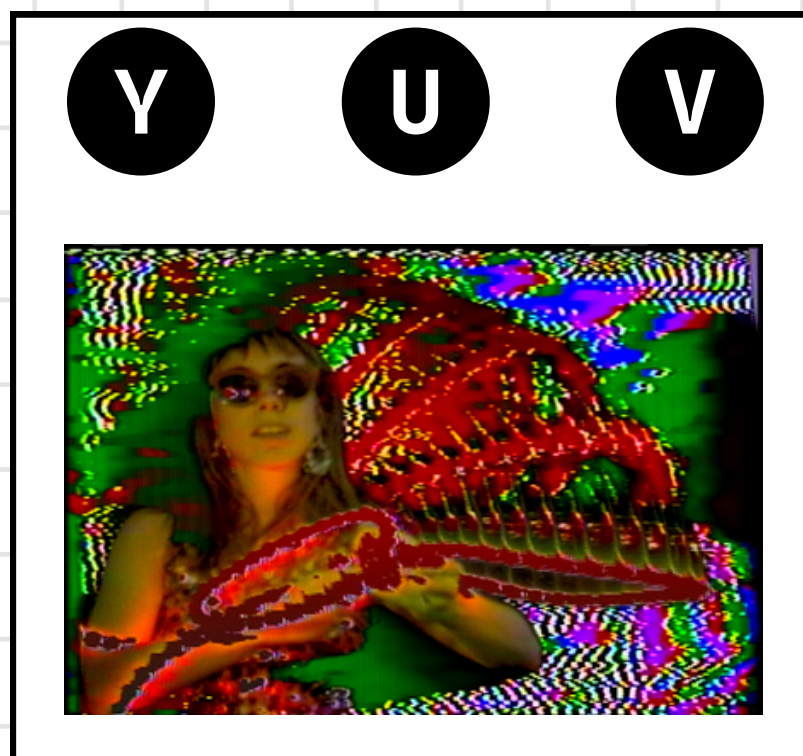


YUV Color Encoding





YUV Color Encoding



Y Luminance Channel



U Chroma Channel 1



V Chroma Channel 2



10-Bit Video Data Scope

	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464
249	0187	016D	0162	017A	01B8	0206	0237	021D	0207	01F4	01DE	01E9	022F	026B	028A	02B6	02D8	02A5	023C	01C5	0141	00CE	009D	0093	0091
	018B	018B	01A9	01A9	0192	0192	018E	018E	01A3	01A3	0198	0198	0184	0184	0196	0196	019B	019B	019D	019D	01A8	01A8	0199	0199	0191
	0245	0245	0245	0245	025A	025A	026B	026B	0278	0278	0264	0264	0285	0285	0292	0292	0285	0285	0286	0286	0273	0273	0282	0282	0281
250	019B	0197	017C	0169	0189	01BD	01EA	0206	020D	0209	0212	0222	0207	01E8	0204	024F	0269	0262	0242	01DB	0138	00C3	0091	0117	0117
	018B	018B	01C2	01C2	01B2	01B2	01AE	01AE	019B	019B	0183	0183	0196	0196	01C6	01C6	01C2	01C2	018C	018C	017E	017E	01DC	01BC	01A7
	024A	024A	024B	024B	0248	0248	0256	0256	0254	0254	026A	026A	028F	028F	0294	0294	0278	0278	026D	026D	0293	0293	0283	0283	0281
251	0191	0175	0172	0197	01EA	0226	0238	0218	01DB	01CA	01CE	01DF	01F1	01EE	0201	022D	0248	023D	0234	0215	01C2	0168	0137	014D	0181
	01BC	01BC	01C8	01C8	01CA	01CA	01BE	01BE	01A3	01A3	0197	0197	0198	0198	01BC	01BC	01B9	01B9	01A5	01A5	019C	019C	01BC	01BC	0191
	023A	023A	0248	0248	024C	024C	024C	024C	0258	0258	0261	0261	0272	0272	0287	0287	026F	026F	0271	0271	0285	0285	0273	0273	0251
252	018B	0168	0169	01C4	0235	0273	0282	023C	01E1	019F	019C	01CD	0208	023C	026E	026B	0256	0252	026F	0297	027A	028D	019A	0158	0161
	01F3	01F3	01EA	01EA	01B6	01B6	01C2	01C2	01DF	01DF	01B4	01B4	01BA	01BA	01C8	01C8	01C9	01C9	01B2	01B2	0198	0198	01AC	01AC	01A7
	0244	0244	0215	0215	022A	022A	024A	024A	0245	0245	0245	0245	026D	026D	0278	0278	026D	026D	0261	0261	0274	0274	0283	0283	0281
253	01AD	019F	01A7	01E3	0237	025D	0258	021C	01BC	0187	01A6	0207	026C	02BC	02E1	02B1	0268	025A	0282	02BE	02BB	023C	0182	018D	0091
	01E3	01E3	01D8	01D8	01C2	01C2	01D2	01D2	0208	0208	01D3	01D3	01B9	01B9	01D5	01D5	01E5	01E5	01B4	01B4	0198	0198	01B6	01B6	0281
	0236	0236	021B	021B	022A	022A	0245	0245	0233	0233	021B	021B	0256	0256	026D	026D	024F	024F	023F	023F	025E	025E	028C	028C	0281
254	0152	015E	0183	01C7	0222	0258	024A	01FF	01B9	01A8	01C8	0218	0298	02D8	02EF	02CA	0298	0273	028E	02DA	02D6	024E	0188	00C8	0091
	01EB	01EB	01F3	01F3	0207	0207	01F4	01F4	01BE	01BE	01D8	01D8	01E5	01E5	01D1	01D1	01C8	01C8	01DF	01DF	028D	028D	01BA	01BA	0281
	0237	0237	0248	0248	0235	0235	0211	0211	021D	021D	024A	024A	023B	023B	023F	023F	025B	025B	026F	026F	0242	0242	028E	028E	0281
255	00E7	00E4	0107	0166	01D2	021D	023B	0232	028A	01E7	01FC	0236	028F	02D1	02E6	02CA	0297	0287	029F	02CB	02B5	024E	0196	008A	0091
	01EB	01EB	01E1	01E1	01FB	01FB	0208	0208	01EF	01EF	01ED	01ED	01D9	01D9	01CA	01CA	01C8	01C8	01CE	01CE	0208	0208	019F	019F	0281
	0287	0287	028F	022F	0231	0231	021B	021B	0228	0228	0225	0225	0223	0223	0231	0231	0248	0248	024C	024C	023A	023A	021C	021C	0281
256	0088	00C8	00E6	0133	01A7	01F9	020C	01EE	01C5	0199	01B3	0208	0272	02BD	02CB	0294	023F	0214	0223	0258	0274	0228	019A	018D	0091
	01EB	01EB	01F3	01F3	01D7	01D7	01D5	01D5	01E6	01E6	01CC	01CC	01B4	01B4	01CE	01CE	01DB	01DB	01C2	01C2	01AD	01AD	01F1	01F1	0281
	028B	028B	0287	0287	028C	028C	028F	028F	0212	0212	028B	028B	0224	0224	0241	0241	0231	0231	022E	022E	0263	0263	027E	027E	0091
257	00DC	0088	0083	0185	01AD	0245	029A	027D	01F3	0153	0138	018C	0211	027B	0287	0235	01D4	01A8	01B4	01DB	01F7	01DD	017D	0116	0091
	0288	0288	0182	01DE	01A7	01A7	01FB	01FB	0219	0219	01AE	01AE	0186	0186	01CA	01CA	01E7	01E7	01BC	01BC	01A8	01A8	01D7	01D7	0181
	0218	0218	0178	0178	028F	028F	024A	024A	0282	0282	01E7	01E7	022C	022C	024F	024F	022E	022E	022B	022B	024D	024D	024C	024C	0281
258	02C8	02CD	02C8	02C9	02E8	030E	033F	0328	0299	01DF	0178	0148	016E	01CD	0208	01D1	017B	015E	0184	01C8	01EC	01CB	016E	0119	0091
	01CE	01CE	01F9	01F9	01EE	01EE	01DF	01DF	01E5	01E5	0208	0208	01CC	01CC	019D	019D	01B4	01B4	01E1	01E1	01FB	01FB	01CE	01CE	0281
	0216	0216	028E	028E	0288	0288	0288	0288	0215	0215	028D	028D	01FE	01FE	028E	028E	0242	0242	024A	024A	022E	022E	0222	0222	0281
259	0317	02FB	0314	0311	0314	032B	0331	02FB	0266	019F	0182	00A8	0083	0119	0173	0173	0148	0133	0167	01C9	021F	0225	01E7	0191	0181
	01F9	01F9	01EC	01EC	01D8	01D8	01B8	01B8	01E1	01E1	0208	0208	0188	0188	01A7	01A7	01BC	01BC	01ED	01ED	01F3	01F3	01C8	01C8	0181
	0282	0282	01F2	01F2	01FE	01FE	01FF	01FF	021C	021C	0285	0285	0288	0288	0234	0234	0258	0258	024A	024A	0217	0217	028A	028A	0281
260	02D5	02DC	02DE	02E4	02F1	0308	032A	02EE	0214	0188	0083	0083	011E	0157	0184	00E8	012C	0183	01F8	0243	0258	022F	022A	0281	
	01D5	01D5	01EF	01EF	01FD	01FD	0283	0283	0288	0288	0288	0288	0288	01DF	01FB	0178	0288	0288	01EB	01EB	01C8	01C8	01D7	01D7	0281
	0287	0287	0286	0286	0285	0285	0286	0286	0288	0288	0288	0288	0288	0238	0218	0218	0288	0288	0286	0286	0229	0229	024D	024D	0281

10-Bit Video Data Scope

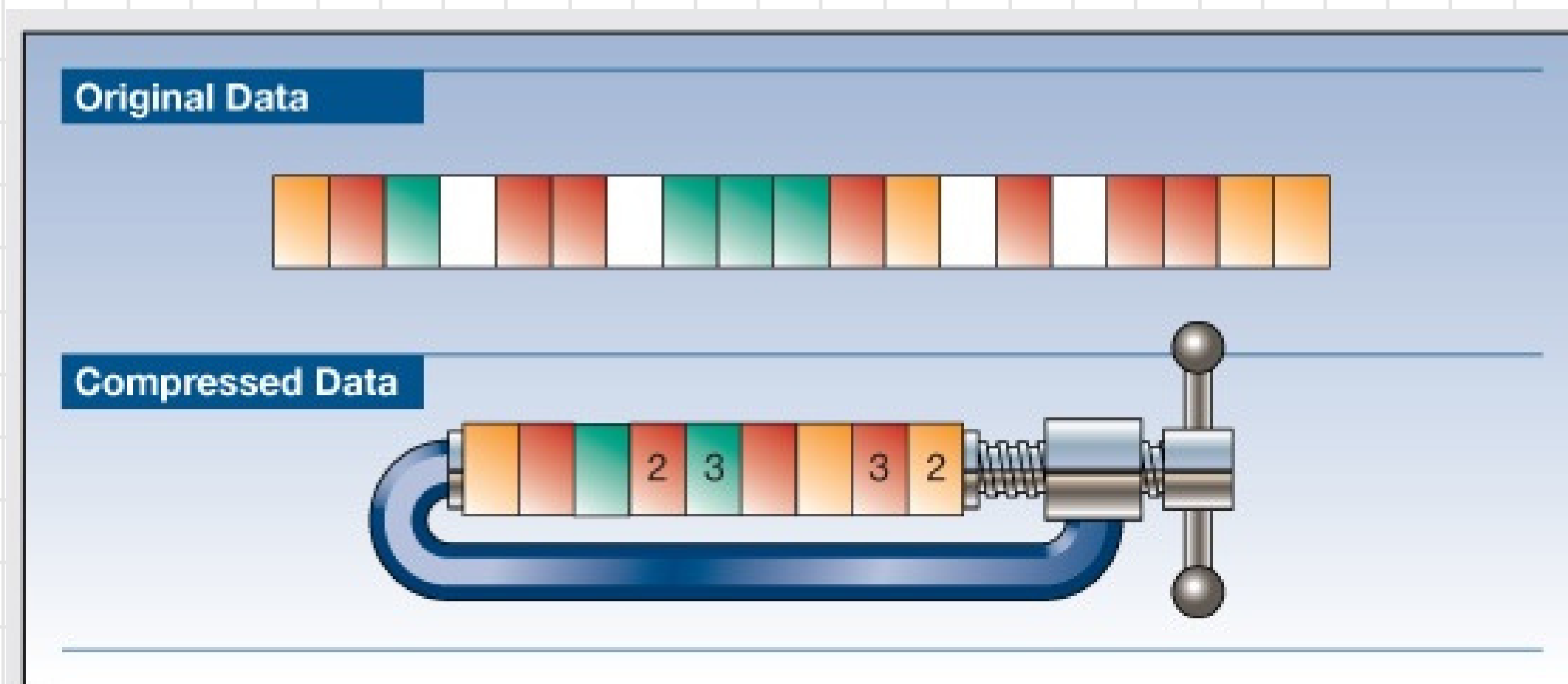
02D6	024E	0188	00DB	0087	00C7
020D	020D	01BA	01BA	0161	0161
0242	0242	020E	020E	0266	0266
02B5	024E	0196	00DA	00B4	009E
0200	0200	019F	019F	0153	0153
023A	023A	021C	021C	027D	027D
0274	0228	019A	010D	00C0	00B6
01AD	01AD	01F1	01F1	021B	021B
0263	0263	027E	027E	0230	0230

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

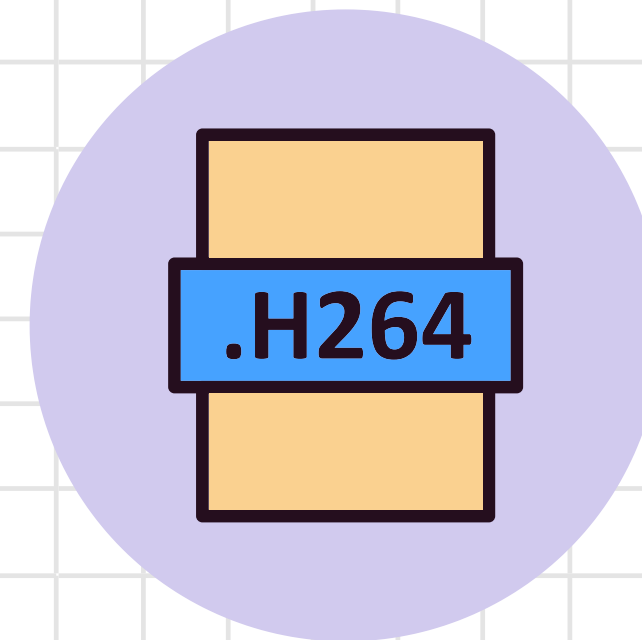
- Lossy
 - Some data is lost during compression and cannot be recovered
 - Common for "access files"
- Lossless
 - No data is lost during compression
 - All original data can be recovered



**Codecs:
Lossy
vs
Lossless**

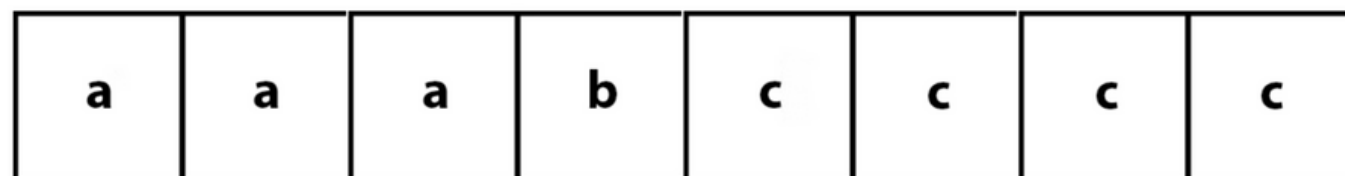
Lossy Compression

- 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!

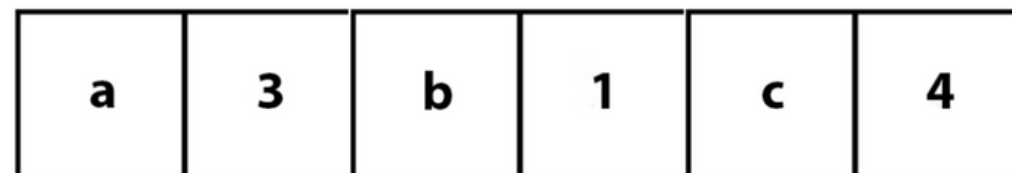


Lossless Compression

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



run-length encoding



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**

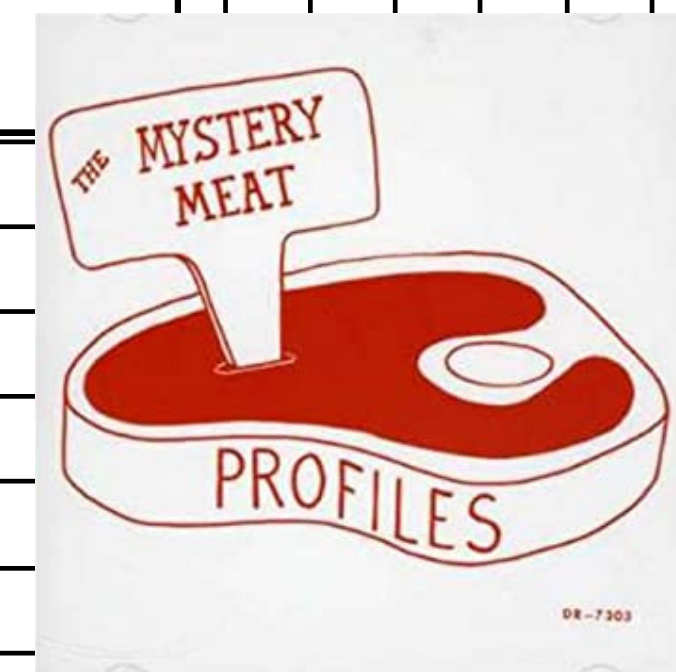


Format Longevity and Adoption

- 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
BREATH



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 ancilliary
data

How do
FFV1 and
MKV Work
Together?



FFV1 MKV For Video Preservation

the standardization efforts for both 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?

Null tests work using the principal of wave
addition and interference

- Flksdlkasjd

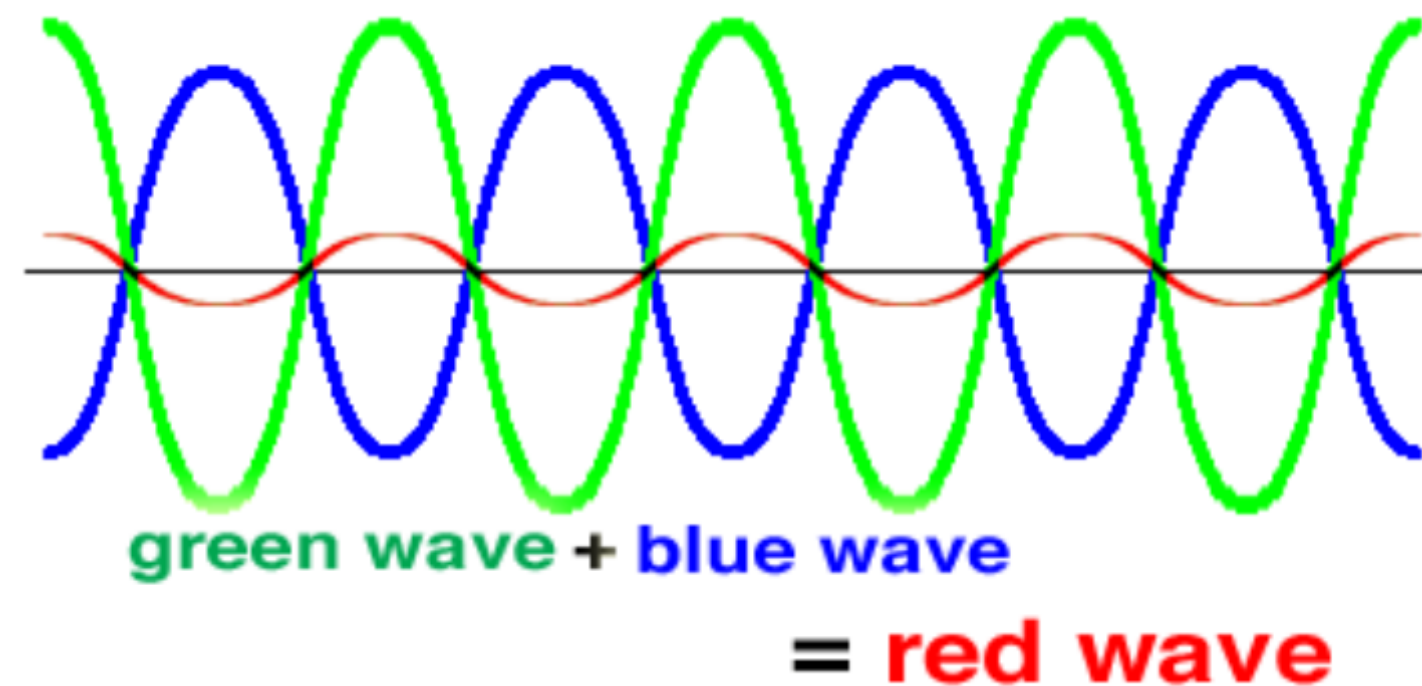
Is the
digital
signal
properly
recorded?

Does the
recorded
format
match the
desired
specs

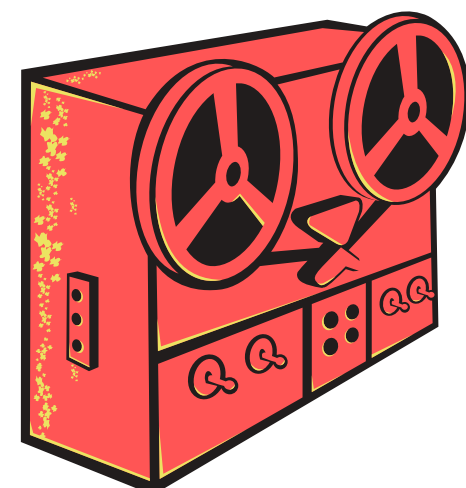
Null tests work using the principal of wave
addition and interference

Null Tests for Audio Station Validation

Null tests work using the principal of wave addition and interference



Audio Null Test



→ ADC

0101010110010



Known Good Computer

0101010110010



Computer Being Tested

Audio Null Test



Known Good Recording

-



Recording Being Tested

=



Total Silence if Correct!



Known Good Recording

-



Recording Being Tested

=



If it's not silent something is wrong

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

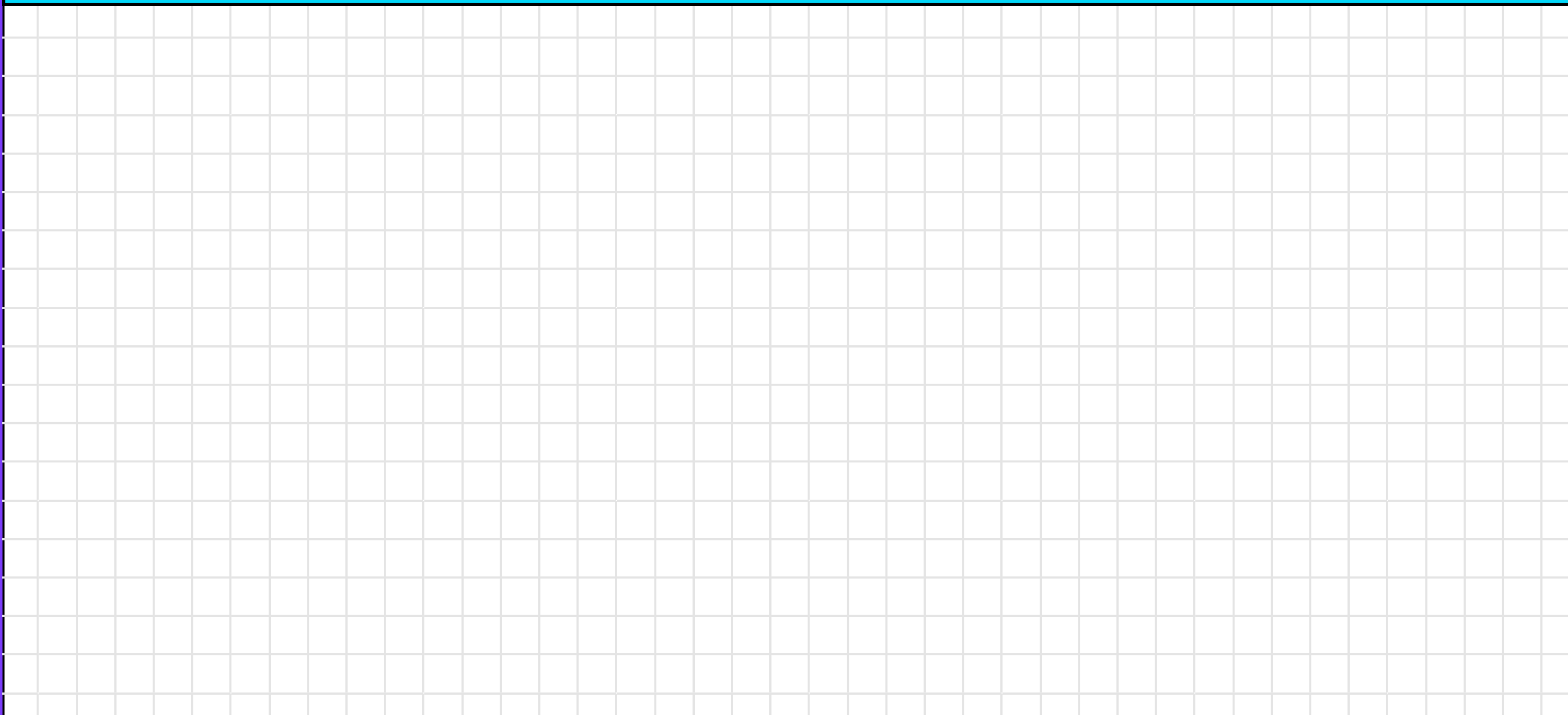
- Is the analog signal properly converted to digital?
- Is the digital signal properly recorded onto the storage media?
- Does the recorded format match the desired specifications

Is the analog signal properly converted to digital?

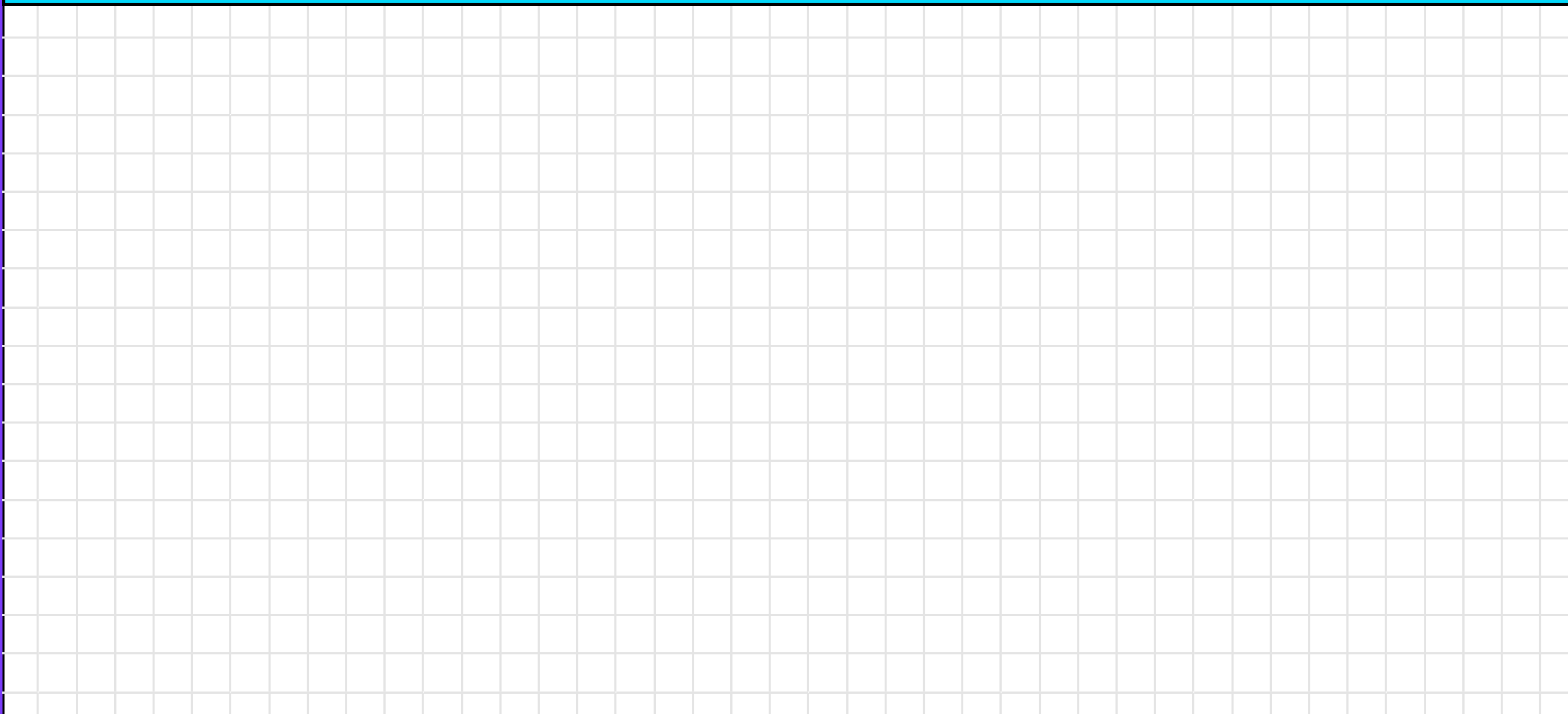
- Are there sync errors?
- Are there repeated frames?
- Is the digitized signal 10 bits?
- Is the video waveform clipped at all?

**Create tests
for these
parameters
using
available
tools**

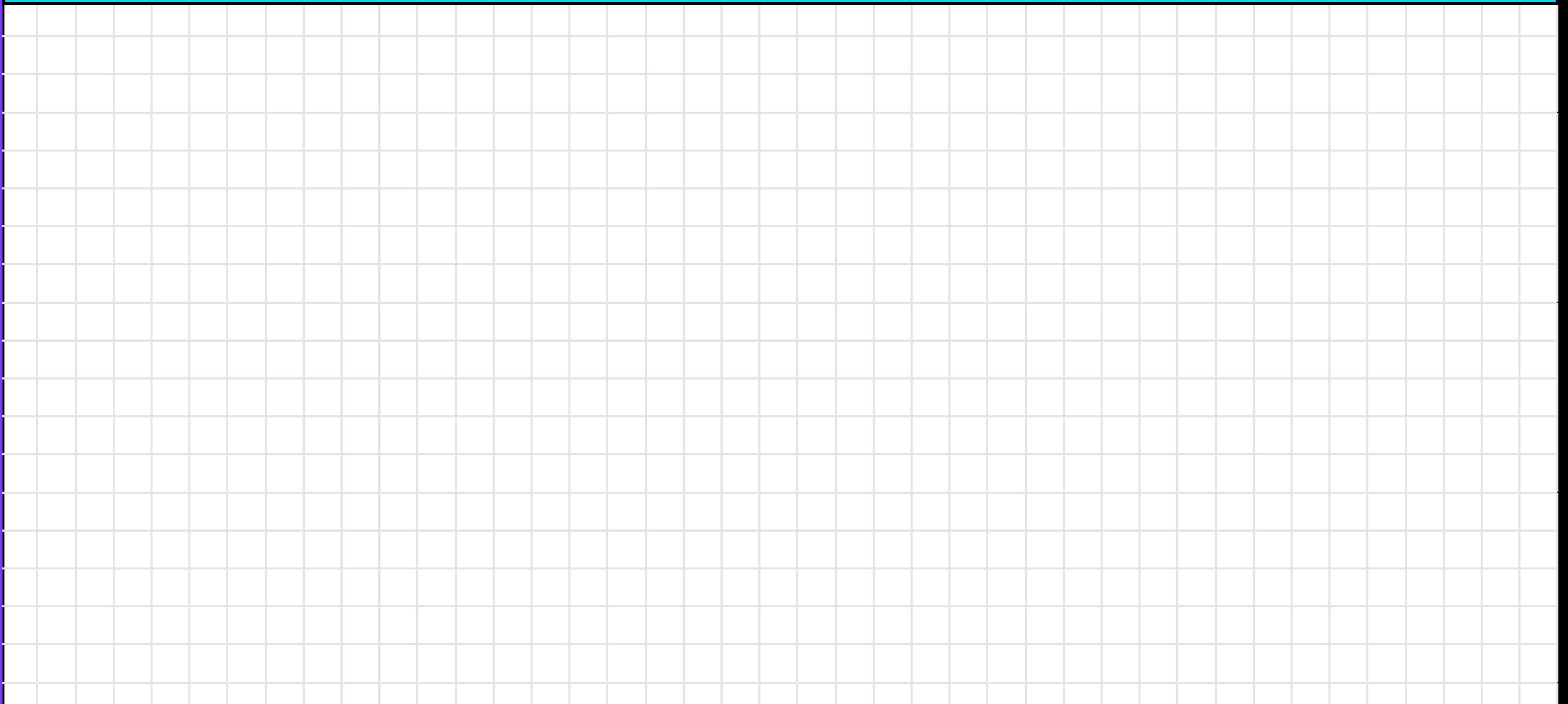
Identifying Sync Errors in QCTools



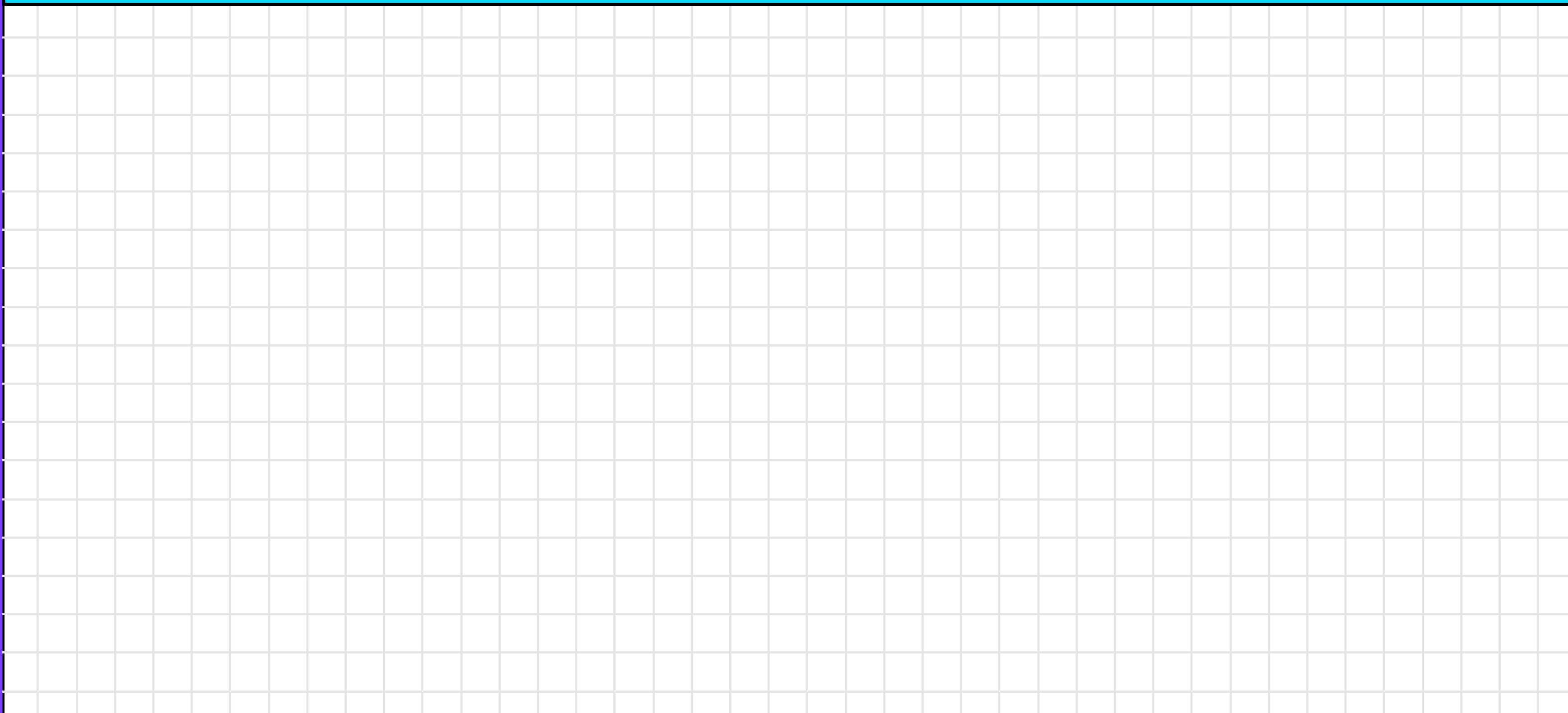
Identifying Repeated Frames in QCTools



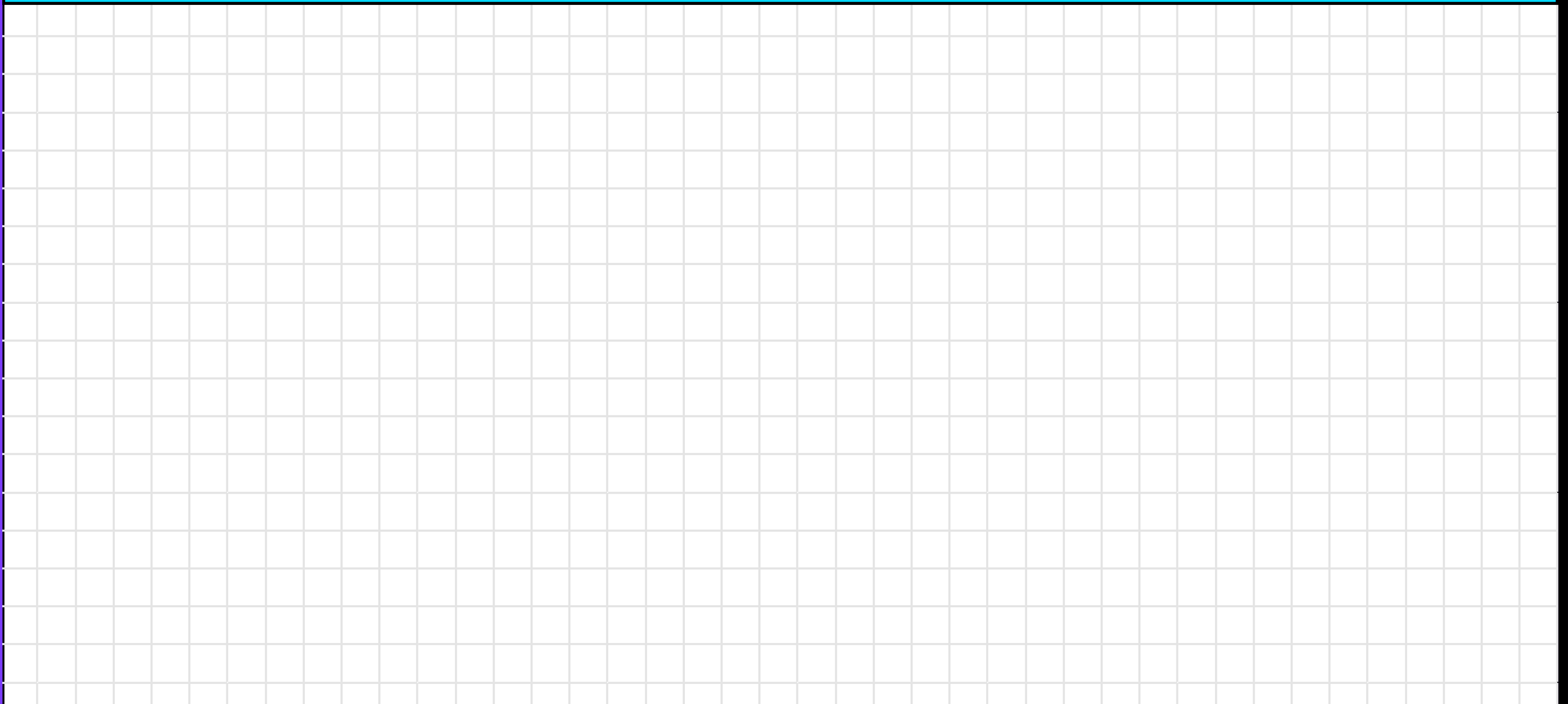
Identifying 10bit/8bit Errors in QCTools



Identifying Clipping in QCTools



Identifying Progressive Video in QCTools



Is the digital signal properly recorded onto the storage media?

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

Create tests for these parameters using available tools

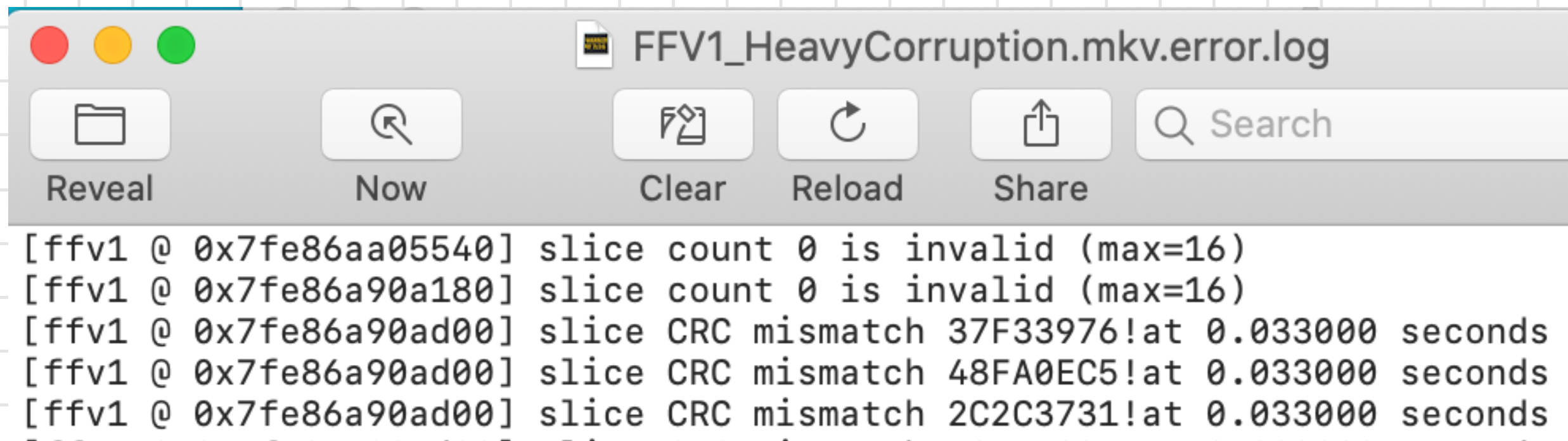
Identifying Bad FFV1 with MediaConch

Files	↓	Implem	↑
FFV1.mkv		✓ Valid	
FFV1_HeavyCorruption.mkv		✗ Not valid	
FFV1_LightCorruption.mkv		✓ Valid	
FFV1_MediumCorruption.mkv		✓ Valid	

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

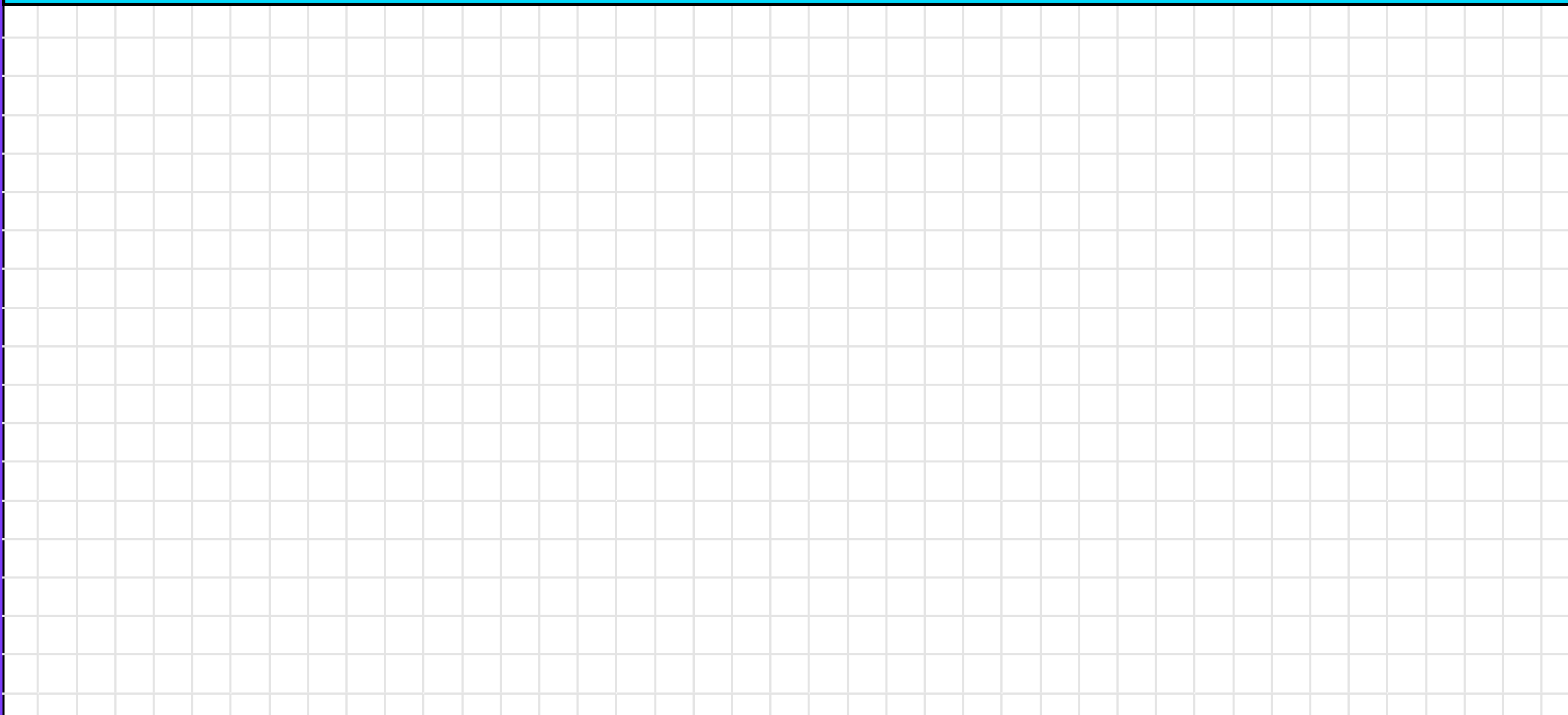


Does the recorded format match the desired specifications

- Does the file match the desired specifications?

**Create tests
for these
parameters
using
available
tools**

Identifying Conformance with MediaConch



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



415.558.2158



Reach Out With Any Questions!