**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 /N17944**

**October 2018, Macao SAR, CN**

|  |  |
| --- | --- |
| **Source** | **Requirements** |
| **Title** | **Call for Proposals for Low Complexity Video Coding Enhancements** |
| **Status** | **Approved** |

|  |  |  |  |
| --- | --- | --- | --- |
| *Author(s):* | Vittorio Baroncini  Simone Ferrara  Yan Ye | Email:  Email:  Email: | [baroncini@gmx.com](mailto:baroncini@gmx.com)  [simone.ferrara@v-nova.com](mailto:simone.ferrara@v-nova.com)  [yan.ye@interdigital.com](mailto:yan.ye@interdigital.com) |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Introduction

This document is a Call for Proposals on video coding technology defining a data stream structure defined by two component streams, a base stream decodable by a hardware decoder, and an enhancement stream suitable for software processing implementation with sustainable power consumption. The enhancement stream will provide new features such as compression capability extension to existing codecs, lower encoding and decoding complexity, for live and on demand streaming applications.

# Purpose and procedure

Companies and organizations that have developed video compression technology that they believe to be complying with the requirements set out in document [1] are invited to submit proposals in response to this Call.

There will be two key criteria to evaluate the proposed video compression technology:

1. the first criterion will be compression performance, which will be evaluated using formal subjective tests and BD PSNR and rate [2][3] criteria. In all evaluations, it is anticipated that subjective evaluation will have primary importance. PSNR results are provided for the main purpose of ensuring that the enhancement data is designed to produce a reconstructed signal that approximates the original source as closely as possible.

Results of these tests will be made public (although no direct identification of the proponents will be made in the report of the results unless it is specifically requested or authorized by a proponent to be explicitly identified and a consensus is established to do so). Prior to having evaluated the results of the tests, no commitment to any course of action regarding the proposed technology can be made; and

1. the second criterion will be complexity of the video compression technology, and this will be assessed by evaluating encoding and decoding complexity in accordance with conventional MPEG methodologies such as encoding and decoding time, memory requirements, degree of capability for parallel processing, and sufficient description of the coding complexity of the proposals.

It is anticipated that both criteria will have equal importance.

Descriptions of proposals shall be registered as input documents to the proposal evaluation meeting MPEG 126 in Geneva (see the timeline in section 3). Proponents also need to attend this meeting to present their proposals. Further information about logistical steps to attend the meeting can be obtained from the listed contact persons (see section 8).

# Timeline

The timeline of the Call for Proposals is as follows:

* 2018/10/12: Final Call for Proposals issued (published)
* 2018/11/02 Anchors made available
* 2018/11/02: Formal registration[[1]](#footnote-1) period opens
* 2018/12/20: Formal registration period ends
* 2019/01/10: Test fee paid
* 2019/02/01: Coded test material shall be received by the Test Coordinator
* 2019/02/18: Subjective test begins
* 2019/03/15: Subjective test ends
* 2019/03/18: Subjective test results available
* 2019/03/23: Evaluation of proposals at MPEG meeting

Anticipated tentative timeline after Call for Proposals:

* 2019/03/25-29: Woking draft
* 2019/07/08-12: CD
* 2020/01/13-17: DIS
* 2020/04/20-24: FDIS

# Coding conditions and anchors

The following test sequences and frame rates shall be used:

UHD sequences: 3840×2160p 60 fps 10 bits: "FoodMarket4”[[2]](#footnote-2)  
3840×2160p 50 fps 10 bits: "ParkRunning3"  
3840×2160p 30 fps 10 bits: "Campfire"

HD sequences: 1920×1080p 50fps 8 bits: "BasketballDrive"  
1920×1080p 50fps 8 bits: "Cactus"  
1920×1080p 60fps 10 bits: "RitualDance”

The test sequences are described in greater detail in Annex 3.

The target bit rates selected for the sequences are as follows:

Table 1 – Target bit rate points UHD sequences

|  | Target bit rates [kbit/s] | | | |
| --- | --- | --- | --- | --- |
| Sequences | Rate 1 | Rate 2 | Rate 3 | Rate 4 |
| FoodMarket4 | 2200 | 3800 | 6500 | 11000 |
| ParkRunning3 | 6000 | 10000 | 16500 | 22000 |
| Campfire | 6000 | 10000 | 16500 | 22000 |

Table 2 – Target bit rate points for HD sequences

|  | Target bit rates [kbit/s] | | | |
| --- | --- | --- | --- | --- |
| Sequences | Rate 1 | Rate 2 | Rate 3 | Rate 4 |
| RitualDance | 2300 | 3800 | 5500 | 8000 |
| BasketballDrive | 2000 | 3500 | 5000 | 7300 |
| Cactus | 1200 | 2000 | 3200 | 4800 |

The sequences are encoded using the standard Constraint Set 1 as defined below:

* Constraint Set 1 (“Random Access”): for HM-based sequences, not more than 16 frames of structural delay, e.g. 16 pictures “group of pictures (GOP)”, and random-access intervals of 1.1 seconds or less. For JM-based sequences not more than 8 frames of structural delay and random-access intervals of 1.1 seconds or less. A random-access interval of 1.1 seconds or less shall be defined as 32 pictures or less for a video sequence with a frame rate of 24, 25 or 30 frames per second, 48 pictures or less for a video sequence with a frame rate of 50 frames per second, 64 pictures or less for a video sequence with a frame rate of 60 frames per second, and 96 pictures or less for a video sequence with a frame rate of 100 frames per second.

AVC and HEVC anchors will be provided using JM (v. 19.0) and HM (v.16.16), respectively[[3]](#footnote-3). For the HD sequences, proponents shall provide a data stream per sequence enhancing AVC only. For the UHD sequences, proponents shall provide two data streams per sequence, one enhancing AVC and the other enhancing HEVC. Proponents shall use JM (v. 19.0) for AVC and HM (v.16.16) for HEVC. The data streams shall comprise a portion compliant with either HM or JM, respectively (depending on the enhanced codec) and a portion comprising ancillary data as illustrated in the figure in Annex 4. The proponents can use a down-sampler of their choice in their proposal, provided that the down-sampler is described with sufficient details. Proponents are free to propose any processing for the enhancement layer as well as for the up-sampler. The target bit rates in Table 1 and Table 2 refer to the overall bitrate of a video stream, including both the standard and the ancillary data portions.

The evaluation for video quality shall be performed by means of a formal subjective assessment process at the full resolution.

The test shall be done using the DSIS (Double Stimulus Impairment Scale) and adopting an 11 grades impairment scale. The DSIS method is described in Annex 2. The scoring scale is the same suggested by the ITU-R Recommendation BT.2095.

For each test point (combination of test source sequence and compression bit rate) an MOS value will be provided together with the Confidence Interval.

Submissions to the call shall obey the following additional constraints:

* Down-sampling and up-sampling methods shall be described and should be the same for all sequences and bitrates.
* Pre-processing of all input video data except down-sampling (e.g. denoising, color correction, filtering etc.) should not be done.
* Only use post-processing if it is a normative part of the decoding process. Any such post-processing must be documented.
* Quantization settings should be kept static except once per coded bit stream. What change of quantization is used shall be described.
* Proponents shall not optimize encoding parameters or any processing steps using non-automatic means.
* The test video sequences shall not be used as the training set for training large entropy coding tables, VQ codebooks, etc.
* Usage of multi-pass encoding and look-ahead is not done.

# Requirements on Submissions

More information about file formats can be found in Annex 1. Files of decoded sequences and bitstreams shall follow the naming conventions as specified in Annex 1.

Proponents shall provide the following; incomplete proposals will not be considered:

1. Coded test material submission to be received on hard disk drive by February 1st 2019:
2. Bitstreams for all test cases and all bit rates as specified.
3. Decoded sequences (YUV files) for all test cases as specified and all bit rates as specified.
4. Binary decoder and encoder executable.
5. Evidence that the portion of the bitstream corresponding to the JM or HM bitstream is decodable by a standard JM or HM decoder, respectively.
6. MD5 checksum files.
7. Coded test material to be brought for the MPEG meeting 126 in Geneva:
8. Bitstreams for all test cases as specified and all bit rates as specified.
9. Decoded sequences (YUV files) for all test cases as specified and all bit rates as specified.
10. Binary decoder and encoder executable.
11. Evidence that the portion of the bitstream corresponding to the JM or HM bitstream is decodable by a standard JM or HM decoder, respectively.
12. MD5 checksum files.
13. Document to be submitted before the evaluation meeting 126 in Geneva shall contain the following:
14. A technical description of the proposal sufficient for full conceptual understanding and generation of equivalent performance results by experts and for conveying the degree of optimization required to replicate the performance. This description should include all data processing paths and individual data processing components used to generate the bitstreams. It does not need to include complete bitstream format or implementation details, although as much detail as possible is desired.
15. The technical description shall also contain a statement about the programming language in which the software is written, e.g. C/C++ and platforms on which the binaries were compiled.6
16. The technical description shall state how the proposed technology behaves in terms of random access to any picture within the sequence. For example, a description of the GOP structure and the maximum number of pictures that must be decoded to access any picture could be given.
17. The technical description shall specify the expected encoding and decoding delay characteristics of the technology, including structural delay e.g. due to the amount of picture reordering and buffering, the degree of picture-level multi-pass decisions and the degree by which the delay can be minimized by parallel processing.
18. The technical description shall contain information suitable to assess the complexity of the implementation of the technology, including the following but not limited to:

* Encoding time[[4]](#footnote-4)(for each submitted bitstream) of the software implementation. Proponents shall provide a description of the platform and methodology used to determine the time. To help interpretation, a description of software and algorithm optimizations undertaken, if any, is welcome.
* Decoding time[[5]](#footnote-5) for each bitstream running the software implementation of the proposal, and for the corresponding constraint case anchor bitstream(s)[[6]](#footnote-6) run on the same platform. Proponents shall provide a description of the platform and methodology used to determine the time. To help interpretation, a description of software optimisations undertaken, if any, is encouraged.
* Expected memory usage of encoder and decoder.
* Complexity characteristics of all the tools used.
* Degree of capability for parallel processing.

1. Optional information

Proponents are encouraged (but not required) to allow other committee participants to have access, on a temporary or permanent basis, to their encoded bitstreams and binary executables or source code.

# Subsequent provision of Source Code and IPR consideration

Proponents are advised that, upon acceptance for further evaluation, it will be required that certain parts of any technology proposed be made available in source code format to participants in the core experiments process and for potential inclusion in the prospective standard as reference software. When a particular technology is a candidate for further evaluation, commitment to provide such software is a condition of participation. The software shall produce identical results to those submitted to the test. Additionally, submission of improvements (bug fixes, etc.) is certainly encouraged.

Furthermore, proponents are advised that this Call is being made subject to the common patent policy of ITU-T/ITU-R/ISO/IEC (see http://www.itu.int/en/ITU-T/ipr/Pages/policy.aspx or ISO/IEC Directives Part 1) and the other established policies of the standardization organizations.

# Test sites and delivery of test material

The proposals submission material will be evaluated by means of a formal subjective assessment process. The tests will be conducted by the Test Coordinator and the EVATech Laboratories.

The bitstream and YUV files will have to follow the naming syntax described in Annex 2.

All proponents need to deliver, by the due date of February 1st 2019, a solid state drive (SSD) to the address of the Test Coordinator (see section 9). The disk shall contain the bitstreams, YUV files, and decoder executable[[7]](#footnote-7) used by the proponent to generate the YUV files from the bitstreams.

Reception of the disk will be confirmed by the Test Coordinator. Any inconvenience caused by unexpected delivery delay or a failure of the disk will be under the complete responsibility of the proponents. However, solutions will be negotiated to ensure that the data can still be included in the test if feasible, which means that correct and complete data needs to be available before the beginning of the test at the latest.

The MD5 checksums of all bitstreams, decoder executable files, and YUV files shall be provided in a text file included on the disk in order to verify that the data can be read without corruption.

Further technical details on the delivery of the coded material are provided in Annex 1.

# Testing fee

Proponents will be charged a fee per each category to which an algorithm proposal is submitted (but not multiple responses to a single category). The fee will be EUR 3,500 per submission. Such fee will cover the logistical costs without any profit. The fee is non-refundable after the formal registration is made.

# Contact(s)

Lu Yu (Video Chair)

Director of the Institute of Information and Communication Networks Engineering

Zhejiang University, China

Tel. +86 571 87953107, email [yul@zju.edu.cn](mailto:yul@zju.edu.cn)

Vittorio Baroncini (Test Coordinator)

Technical Director

EVATech

Via Pisanelli, 4, 00196 – Rome, Italy

Tel. +39-3335474643, email [baroncini@gmx.com](mailto:vittorio@fub.it)

# References

[1] “Requirements for Low Complexity Video Coding Enhancements”, N18098, October 2018, Macao, China

[2] Gisle Bjontegaard, “Calculation of Average PSNR Differences between RD curves”, ITU-T SG16/Q6 VCEG 13th meeting, Austin, Texas, USA, April 2001, Doc. VCEG-M33 (available at <http://wftp3.itu.int/av-arch/video-site/0104_Aus/>).

[3] Gisle Bjontegaard, “Improvements of the BD-PSNR model”, ITU-T SG16/Q6 VCEG 35th meeting, Berlin, Germany, 16–18 July, 2008, Doc. VCEG-AI11 (available at <http://wftp3.itu.int/av-arch/video-site/0807_Ber/>).

**Annex 1**

Distribution of original video material files containing test sequences is done in YUV files with extension “.yuv”.

HEVC Anchor bitstreams are provided with extension “.hevc”. AVC Anchor bitstreams are provided with extension “.avc”. Bitstream formats of proposals can be proprietary, but must contain all information necessary to decode the sequences at a given data rate (e.g. no additional parameter files). The file size of the bitstream will be used as a proof that the bit rate limitation has been observed. The file extension of a proposal bitstream shall be “.bit”.

Decoded sequences shall be provided in the same “.yuv” format as originals, with the exception that the colour depth shall be 10 bits per component for the relevant sequences.

All files delivered (bitstreams, decoded sequences and binary decoders) must be accompanied by an MD5 checksum file to enable identification of corrupted files. An MD5 checksum tool that may be used for that purpose is typically available as part of UNIX/LINUX operating systems; if this tool is used, it should be run with option “-b” (binary). For the Windows operating systems, a compatible tool can be obtained from <http://www.pc-tools.net/win32/md5sums/>. This tool should be run with additional option “-u” to generate the same output as under UNIX.

The SSD should be shipped (for handling in customs) with a declaration “used hard disk drive for scientific purposes, to be returned to owner” and low value specification (e.g. EUR 20). The use of a SSD with substantially larger size than needed is discouraged. The hard disk drive should be a portable SSD with a USB C interface. NTFS file format shall be used.

**Annex 2**

**Test**

### Test method: DSIS - Double Stimulus Impairment Scale

The Formal Subjective Assessment of the YUV video files received from the Proponents will be visually assessed for quality using the DSIS (Double Stimulus Impairment Scale) test method as specified in this Annex.

The DSIS method will be used under the schema of evaluation of the impairment, using an 11 grades impairment scale, ranging from "0" (lowest quality) to "10" (highest quality).

The structure of the Basic Test Cell (BTC) of the DSIS method is made by two consecutive presentations of the video clip under test; at first the original version of the video clip (SRC, Source Reference Content) is displayed, announced by the letter “A” on a mid grey screen (half a second), immediately afterwards the coded version of the video clip (PVS, Processed Video Sequence), announced by the letter “B” on a mid grey screen (half a second), is presented; then a message displays for 5 seconds asking the viewers to vote (see Figure C-1)



*Figure 1 – DSIS BTC*

*Voting*

The viewers will be asked to express their vote by putting a number in a box on a scoring sheet.

The scoring sheet for a DSIS test is made of a section for each BTC; each section is made of a numbered box (see figure 2). The viewers will write a number in the box that corresponds to the number of the message “***Vote N***” that is shown on the screen. When writing "10", the subject will declare that he/she saw no difference between the SRC and the PVS; when writing "0" the subject will declare that he/she saw a significant and clearly visible amount of difference between the SRC and the PVS across the entire observed image. The vote has to be written when the message "Vote N" appears on the screen. The number "N" is a numerical progressive indication on the screen aiming to help the viewing subjects to use the appropriate column of the scoring sheet.



*Figure 2 – Example of DSIS test method scoring sheet*

## Training and stabilization phase

The viewing subjects shall be trained by means of a short practice (training) session. The video material used for the training session will be carefully selected from those of the test, taking care to represent as much as possible the extent of visual quality and the kind of impairment expected to appear during the test.

A “stabilization phase” made of three BTCs is located at the beginning of each test session; the stabilization BTCs will be selected representing best, mid and worst quality. In this way, the viewing subjects will have an indication of the range of quality they are expected to evaluate during that session.The scores of the stabilization phase are discarded. Consistency of the behaviour of the subjects will be checked inserting in the session a BTC in which original (SRC) is compared to original (SRC).

## Laboratory setup

The laboratory for a subjective assessment is planned to be quiet, far from any external light source and in a room whose walls, ceiling and floor will be made of non reflective material. High quality TV sets will be used as monitors. All internal local post processing TV features will be disabled. The video server and SW will be able to play video, at both HD and UHD video resolutions and at 24, 30, 50 and 60 frames per second, without any limitation, or without introducing any additional temporal or visual artefacts. The MUP player will be used for the tests.

## Viewing distance, seats and monitor size

55” commercially available OLED TV set will be used. The viewing distance will be 2H, where H is equal to the height of the active part of the screen. Three subjects will be seated in front of the 55” display.

## Viewing environment

The test area will be protected from any external visual or audio pollution. Internal general light will be low (just enough to allow the viewing subjects to fill out the scoring sheets) and a uniform light has to be placed behind the monitor; this light will have an intensity of maximum 30 nits. No light source has to be directed to the screen or create reflections; ceiling, floor and walls of the laboratory have to be made of non-reflecting black or dark grey material (e.g. carpet or velvet).

## Statistical analysis and presentation of the results

The data collected from the score sheets, filled out by the viewing subjects, will be stored in an Excel spread-sheet. For each coding condition the Mean Opinion Score (MOS) and associated Confidence Interval (CI) values will be given in the spread-sheets. The MOS and CI values will be used to draw graphs. The graphs will be drawn grouping the results for each video test sequence. No graph grouping results from different video sequences will be considered.

### Proponents identification and file names

Each Proponent will be identified with a two digit code, e.g. Pxx. The binary and decoded video files will be identified by a name as here below:

**PnnSxxCzRy.<filetype>**

where:

* Pnn identifies the Proponent
* Sxx identifies the original video clip used to produce the coded video, as identified in the tables of Annex 3
* C1 identifies the constraint set (only Constrain 1 [no delay]) is considered in this CfP.
* Ry identifies the bit rate y, as identified in Table , Table
* <filetype> identifies the kind of file:
  + .bit = bitstream
  + .yuv = decoded video clip in YUV format

**Annex 3 – Description of test sequences**

UHD test sequences

| Sequence name | Clip ID | Resolution | Frames | Frame rate | Chroma format | Bit depth |
| --- | --- | --- | --- | --- | --- | --- |
| FoodMarket4 | S01 | 3840×2160 | 600 | 60 | 4:2:0 | 10 |
| ParkRunning3 | S04 | 3840×2160 | 500 | 50 | 4:2:0 | 10 |
| Campfire | S05 | 3840×2160 | 300 | 30 | 4:2:0 | 10 |

UHD sequences example pictures

|  |  |  |
| --- | --- | --- |
| FoodMarket4 | ParkRunning3 | Campfire |

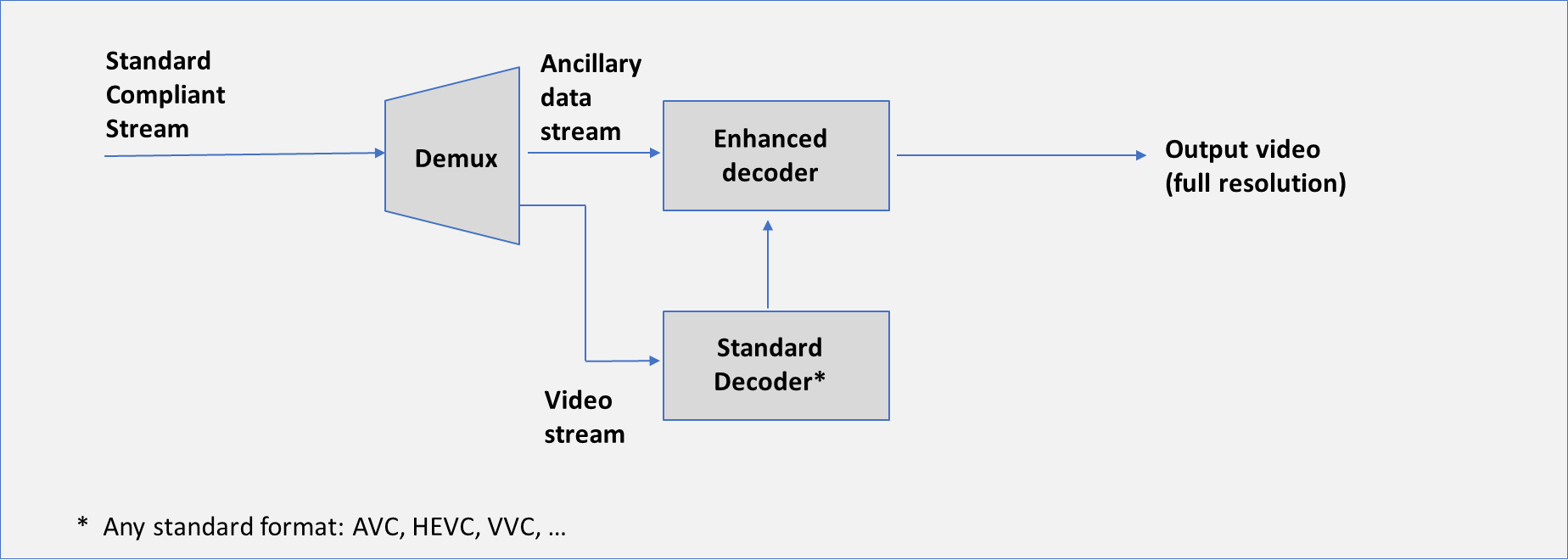
HD test sequences

| Sequence name | Clip ID | Resolution | Frames | Frame rate | Chroma format | Bit depth |
| --- | --- | --- | --- | --- | --- | --- |
| RitualDance | S11 | 1920×1080 | 600 | 60 | 4:2:0 | 10 |
| BasketballDrive | S13 | 1920×1080 | 500 | 50 | 4:2:0 | 8 |
| Cactus | S14 | 1920×1080 | 500 | 50 | 4:2:0 | 8 |

HD sequence example pictures

|  |  |  |
| --- | --- | --- |
| RitualDance | BasketBallDrive | Cactus |

**Annex 4**

****

1. Formal registration intention shall be provided by mail to the Test Chair. The Test Chair will answer providing confirmation of the registration and providing the P-code (Pxx) of the Proponent; the P-code will be used in the file naming process (see Annex 2). [↑](#footnote-ref-1)
2. For the sequence “FoodMarket4” only the first 600 frames shall be used. [↑](#footnote-ref-2)
3. AVC anchors generated at 8-bit. [↑](#footnote-ref-3)
4. For example, using ntimer for Windows systems. [↑](#footnote-ref-4)
5. For example, using ntimer for Windows systems. [↑](#footnote-ref-5)
6. The decoder source code to be used to process the anchor bitstreams will be provided to proponents and must be compiled as-is, without modification of source code, compiler flags, or settings. [↑](#footnote-ref-6)
7. The decoder executable must be able to perform the decoding operation on a computer with a Windows 10 operating system. [↑](#footnote-ref-7)