Configuring Versatile Video Coding: technical guidelines for broadcast and streaming applications.

Łukasz Litwic
Ericsson
Gdańsk, Poland
lukasz.litwic@ericsson.com
Dmytro Rusanovskyy
Qualcomm Technologies Inc
San Diego, CA, USA
dmytror@qti.qualcomm.com
Sean McCarthy
Dolby Laboratories, Inc.
San Francisco, CA, USA
sean.mccarthy@dolby.com
Alan Stein
InterDigital Communications, Inc.
Princeton, NJ, USA
alan.stein@interdigital.com

Abstract – Versatile Video Coding (VVC or H.266) is the latest video coding standard jointly developed by ISO/IEC MPEG and ITU-T VCEG. With best-in-class compression performance, VVC can enhance existing applications and enable new services. As the first VVC implementations enter the market, several application-oriented standards developing organizations and industry fora are defining VVC-based profiles and corresponding receiver capabilities. However, these specifications don't typically prescribe how a service is realized and the impact of the codec's operational parameters on delivered compression performance. To this end, the Media Coding Industry Forum has initiated the development of VVC technical guidelines. These guidelines will serve as a reference for VVC configuration choices to address operational, interoperability, and regulatory needs while achieving optimal compression performance. This paper presents an overview of the guidelines' scope and development process, followed by a discussion of VVC configuration aspects, with focus on new features which are of utmost relevance to broadcast and streaming applications and concludes with analysis of performance for the presented scenarios.

Introduction

Versatile Video Coding (VVC) was standardized by ITU-T as Recommendation H.266 and in ISO and IEC as International Standard 23090-3 (MPEG-I Part 3) [1]. VVC is the latest generation of open video coding standards developed by the Joint Video Experts Team (JVET) of ISO/IEC JTC 1/SC 29 (known as Moving Picture Experts Group MPEG) and by ITU-T Question 6/16 of Study Group 16 (known as Video Coding Experts Group VCEG). The first release of the standard, published in July 2020, delivered a substantial increase in compression efficiency over its predecessor, HEVC/H.265, achieving around 50% bitrate reduction for the same video quality. VVC also improves on HEVC by including a combination of efficient coding tools designed from the start to support a wider range of video content properties including High Dynamic Range (HDR), Wide Color Gamut (WGC), and computer-generated imagery for gaming and remote screen content sharing. VVC's efficient coding tools and rich functionality target efficient delivery of established and emerging video formats such as Ultra-High Definition (UHD) with 4K and 8K resolutions, VR 360º video, and includes built-in scalability support

This paper is excerpted from the Proceedings of the 2023 NAB Broadcast Engineering and Information Technology (BEIT) Conference, © 2023, National Association of Broadcasters, 1 M Street SE, Washington, DC 20003 USA.

already at a single layer bitstream level. Additional versatility can be achieved by metadata signaled in supplemental enhancement information (SEI) messages specified in VVC itself and in VSEI (Versatile SEI messages for coded video bitstreams) (ITU-T Recommendation H.274 | ISO/IEC 23002-7 [2]). As a whole, VVC's coding tools, functionality, and metadata signaling capabilities make it the leading state-of-the-art, feature-complete video codec available.

Unlike HEVC, where new functionality and specialized coding tools were added incrementally over time (e.g., 4:2:2 and 4:4:4 chroma formats were added in HEVC edition 2, scalability support in HEVC edition 3, and screen-content coding in HEVC edition 4), VVC includes support for all this functionality already from the first release. Three sets of profiles were included to support a host of video applications:

- **Main 10** and **Main 10 4:4:4** profiles, where **Main 10** is the flagship profile expected to support a wide range of broadcast and streaming applications and includes support for 8- and 10-bit video, temporal sub-layer functionality and spatial scalability through resolution change. **Main 10 4:4:4** includes support for additional chroma sampling formats: 4:2:2 YCbCr chroma typically used in video production and contribution applications and 4:4:4 RGB format for applications operating on content out of graphics frame buffers, e.g., remote desktop
- **Multilayer Main 10** and **Multilayer Main 10 4:4:4** extend **Main 10** and **Main 10 4:4:4** profiles, respectively, with support for multi-layer bitstreams.
- **Main 10 Still Picture** and **Main 10 4:4:4 Still Picture** specialize **Main 10** and **Main 10 4:4:4** profiles, respectively, to support bitstreams comprising a single intra-coded picture.

The second edition of the VVC standard published in April 2022 added new operation range extension profiles supporting bit depths up to 12 bits for YCbCr chroma formats, intra only profiles including up to 16 bits for RGB formats targeting high fidelity content acquisition and studio production applications.

Following VVC finalization, VVC support was enabled in all relevant media transport and systems standards. Encapsulation in MPEG-2 TS for digital television broadcast services is specified in [3]. Encapsulation of VVC bitstreams in MPEG ISO base media file format (ISOBMFF) is specified in [4]. VVC was also added to MPEG Common Media Application Format (CMAF) including baseline VVC media profile and multilayer VVC media profile [5]. For most typical broadcast and streaming applications, Main 10 profile is expected to provide the requisite compression performance and functionality. Such streamlining of available profiles compared with previous generations of codecs like AVC/H.264 and HEVC/H.265 is expected to have a positive impact on the cost of deployment and interoperability across the ecosystem. At the same time, since VVC includes a much richer set of coding tools and functionality than previous codecs, it would be beneficial to have an industry reference for VVC to guide configuration choices that address relevant user profiles and operating points such as those defined in industry leading application specifications. Such a reference for VVC could facilitate cross-industry harmonization and enhance interoperability across the broadcast and streaming ecosystem.

The Media Coding Industry Forum is an open industry forum with the purpose of furthering adoption of MPEG standards initially focusing on VVC, by establishing them as well-accepted and widely used standards for the benefit of consumers and industry [6]. With the strong industry VVC adoption, the MC-IF Interoperability Working Group has launched work on VVC technical guidelines with the focus on broadcast and streaming applications. These VVC technical guidelines aim to provide information to facilitate incorporation of VVC into application-oriented standards, media production workflows, and media distribution to consumers. In addition to covering best practices for VVC configuration, in terms of both functionality and compression performance, the guidelines aim to include up-to-date information on VVC operating bitrate ranges and provide information on the usage of VVC and SEI messages.

This paper attempts to review the status of VVC deployment and adoption in application standards, introduces MC-IF VVC technical guidelines work, and provides in-depth discussion on performance or functionality of selected VVC configuration aspects. The authors of this paper represent companies who are members of the Media Coding Industry Forum and involved in work on MC-IF guideline.

VVC deployment status

Thanks to strong industry interest, VVC has enjoyed a rapid roll-out across different parts of the entire end-to-end video ecosystem with the first reports of optimized decoder implementation made already during VVC standardization [7][8].

Software decoding

One of the key enablers of such a speedy implementation cycle is the relatively low complexity of the VVC decoding process which is only about 1.5-2x the complexity of an equivalent HEVC decoder [9]. Several publicly available software decoders and players have been released for various platforms: HD playback on Android [12][11][14] and iOS [10][12] mobile platforms, 4K decoding on laptop/desktop Apple M1 [13] , Intel i7 and Intel i9 processors [9] and 8K decoding with AMD EPYC [9] and Intel Xeon including support for Multilayer Main 10 profile [16]. VVC was also reported to playback HD content in Edge, Firefox and Chrome browsers using a WebAssembly build [14][15].

Hardware decoding

While software decoding enables early technology integration and launch of services, in the long term, deployments in broadcast and streaming ecosystem are expected to rely on hardware decoder capabilities especially for 4K and 8K based video formats including high frame rate (HFR) support (100/120 Hz frame rates). Availability of VVC decoder IP core for 8K120 [17] , SoC decoder for set-top boxed [18] and TV chipsets with VVC support for 4K and up to 8K120 were reported [19] while the first news of consumer TV sets supporting VVC were reported for 2023 [20].

Encoding

On the encoding side, several reports have been made since the launch of VVC. Already in October 2020, a publicly available optimized offline encoder was reported offering performance on par with the VVC reference software encoder (VTM) [8]. Due to inclusion of special encoder features such as adaptive QP, the implementation was demonstrated to outperform the reference VTM encoder in subjective verification tests against HEVC [21] while operating with 100x encoding speed that of the VTM. Several other offline encoders have subsequently been reported offering over 30% performance gains over HEVC [22][23] and have been integrated into cloud-based encoding [24] , transcoding [25] and mobile OTT services [26][27]. VVC encoding software with capabilities up to 8K resolution [28] including capabilities of real-time 8Kp30 encoding were reported [29]. Encoder vendors also reported compression performance gains obtained with VVC encoders using comparable computing resources to currently deployed HEVC encoders. In terms of real-time UHD encoding deployed on a public cloud, VVC was reported to achieve between 15%-30% lower bitrate at the same quality as a commercial HEVC solution at 36% compute cost overhead [30]. Data from another vendor reported VVC real-time implementation to provide between 10%-15% bitrate reduction over HEVC while utilizing the same core coding engine as HEVC encoder [37] for 4K content and on average 20% bitrate reduction for 8K content thus reducing 8K delivery bitrates to fit DVB-T2 or ATSC-3.0 transmission systems at around 35 Mbps [31]. While these performance gains are reported lower than 50% usually expected from new generation of ITU-T and ISO/IEC coding standards, it is important to note that these came from early implementations and encoder complexity in these cases was reported to be between 1x-1.5x of that of equivalent HEVC encoder.

Interoperability

VVC interoperability with several key technologies used in the broadcast and streaming ecosystem has also been demonstrated in end-to-end trials. Over-the-air (OTA) and Over-the-top (OTT) trials of UHD content with VVC demonstrating encapsulation with MPEG-2 over DVB-S2 and ISOBMFF delivered with DASH were reported in [37]. Live 4K VVC streaming with low latency CMAF packaging was also reported in [37]. Interoperability with CMAF was also reported for Multilayer Main 10 decoder on 8K/4K/HD content [33].

Conformance testing

JVET has developed a normative specification *Conformance testing for VVC* as ITU-T Recommendation H.266.1 and ISO/IEC 23090-15 [34]. The specification includes a conformance bitstream test set comprising around 280 bitstreams in over 100 categories to cover all VVC profiles in version 1. DVB released VVC test content for verification and validation of DVB VVC profiles (see also **DVB**) [35]. Commercial conformance and test streams for VVC were also developed already during standardization phased and released shortly after standard completion [36]

In summary, the deployment of VVC seems to have a certain success for such a short term. The annual video developer survey report by Bitmovin, conducted with 424 respondents from over 80 countries [38], indicated that 19% of developer were using VVC in production for live coding and 15% for video on-demand coding, making VVC the third most reported codec in the survey with H.264/AVC indicated by 78% and 85% respondents and HEVC/H.265 indicated by 40% and 42% respondents, respectively. A comprehensive review of reported VVC deployments can be found in [39].

VVC adoptions in applications standards

Video traffic has been growing exponentially. In mobile networks the share of video traffic is estimated at 70% and is projected to increase to around 80 percent share of the traffic by the end of the decade [40]. This growth is driven in part by the increasing consumer demand for content delivered in high quality formats, which can either mean video delivery in new enhanced and immersive formats (e.g., with 8K resolution) or a larger availability of UHD content delivered to customers. This has led several standards developing organizations (SDO) across broadcast and streaming ecosystem to investigate the benefits of VVC for inclusion into their specification and systems. The Advanced Television Systems Committee (ATSC) has noted their intention to include VVC in the ATSC 3.0 standard. A recent report on ATSC 3.0 and Global Convergence mentions explicitly that *“ATSC is currently specifying Versatile Video Coding (VVC) for inclusion in the ATSC 3.0 suite of standards”* [41]. Streaming industry fora like DASH-IF and CTA-Wave added VVC profiles into their guidelines and specifications [42][43]. Digital Video Broadcasting (DVB) and Sistema Brasileiro de Televisão Digital (Fórum SBTVD) already adopted VVC involving a formal process of verification of compression performance claims against commercial requirements. 3GPP SA43GPP SA4, while has not started a formal adoption of new codecs into its standards, concluded a feasibility study on codecs performance, where initial performance analysis of new codecs such as VVC was investigated. The next paragraphs give a brief overview based on publicly available reports of VVC verification and adoption status in these organizations.

DVB

Digital Video Broadcasting (DVB) started investigations into new commercial requirements for new video codecs in 2020. DVB set out a number of performance related commercial requirements to be met by new codecs, including: the ability to delivery 8K video over legacy broadcast multiplexes at excellent quality; the ability to enable five 4K services in a 40 Mbps DVB-T2 multiplex (compared with three 4K services expected with HEVC); provide at least 27% more efficient live broadcast encoding than HEVC, and over 30% performance gains for 4K streaming use cases while maintaining performance gains for sub 4K resolutions[44]. As reported in [46], DVB did not perform internal tests but relied on testing done externally. Within just a year DVB verified VVC meeting its commercial requirements and adopted VVC into its codec toolbox and released the Bluebook specification in February 2022 [45]. Four VVC-based operating points for MPEG-2 transport stream and DVB-DASH delivery were defined. All four operating points include UHD with 4K resolution and frame rates up to

60Hz. All operating points are also HDR-capable and progressive only. Maximum capability is defined for 8K resolution with frame rates up to 120Hz. VVC support in DVB specification also includes a new functionality to allow dynamic changes of bitstream resolution within a single-layer Transport Stream programme. VVC also provides support for bitstream accessibility such as composition of additional sign-language video within a main video bitstream [46]

Fórum SBTVD

Sistema Brasileiro de Televisão Digital (aka Forum SBTVD) is the organization charged with creation of broadcast and hybrid broadcast-broadband television standards in Brazil. SBTVD has been working toward a new standard for Brazil known as **TV 3.0**. The process consisted of a call for proposals published in July 2020, [47], candidates' responses, and testing and evaluation, followed by selection. Use cases and requirements in the CFP include delivery of 4K OTA and 8K via OTT, native HDR/WCG support, HFR and a reduced-resolution portrait-mode closed second video service. for sign language purposes. For video, VVC was selected as the Video Base Layer Codec, for both OTA and OTT delivery.

Remaining work (on the application layer) involves drafting specification for audio and video coding and determining operating points based on subjective assessment of video coding quality. End-to-end demonstration of TV 3.0, which would include additional system components, i.e., physical and transport layer, is expected in August 2024 [48].

3GPP SA4

Assessing a potential dominance of video traffic in 5G networks reaching from 65% to 90% of total traffic, 3GPP SA4 working group developed a comprehensive analysis of the 3GPP-defined video compression technologies in terms of their suitability for existing and emerging 5G services. Based on the extensive framework SA4 working group also evaluated the performance of the new video compression technologies, including VVC. Results of this study were published in 3GPP TR 26.955 [59]. In this document, the working group reviewed requirements of its relevant profiles defined in 3GPP specifications, including TV, VR, 360º video VR and streaming video profiles and defined 5 test scenarios in which widely adopted video coding standards, such as H.264/AVC, H.265/HEVC as well as new generation of video coding standards, including H.266/VVC were evaluated. The test scenarios included HD and UHD 4K Streaming (SDR/HDR), Screen Content, Messaging and Social Sharing and Online Gaming. Scenarios definition included typical parameters including resolution, frame rates, target bitrates and recommended encoder configuration suitable for the application. Performance of the codecs under study was assessed at the target bitrate against the set of defined objective quality metrics using BD-rate measurements where BD stands for Bjontegaard delta [54]. Test simulations were conducted over more than 50 test sequences of different characteristics. The results reported in [59] demonstrated superior characteristics of the VVC over other video coding technologies under the categorization in SA4. Comparing to HEVC, VVC demonstrated around 37% and 39% of BD-rate reduction in PSNR (wPSNR for HDR) metrics on average for HD and UHD streaming scenarios, respectively. For Screen Content coding scenario, average BD-Rate gain around 57% in PSNR metric over HEVC was observed. For test content in Messaging and Social Sharing scenario and Online Gaming scenarios, average BD-rate gain of 27% and 32%, respectively, have been reported. With these results, VVC outperformed any other video codecs under study in all scenarios and all test configurations.

MC-IF VVC technical guidelines

The reported scale of investigations and evaluation processes conducted in applications standards SDOs, points to the importance of assessment true capabilities new video coding technologies with respect to legacy codecs as well as alternative coding technologies. Once the new industry profiles are standardized, the assessment of these new profiles against business needs falls on operators and service providers. A comprehensive set of advanced compression tools, rich functionality, flexibility of high-level bitstream operations, and interoperability with associated metadata, contribute to the versatility of VVC and provide very flexible codec configurations options. In order to facilitate the industry adoption of VVC into standards, workflows, and services, MC-IF has launched work on VVC technical guidelines for broadcast and streaming applications. Figure 1 shows a high-level diagram of end-to-end broadcast and streaming ecosystem. The initial scope of the guidelines work is outlined and concerns primarily final emission to end-users as well as primary distribution. For this initial scope, focus is on the use of VVC Main 10 profile. However, other profiles may also be of interest e.g., Multilayer Main 10 which was reported to be used in 8K delivery scenarios and Main 10 4:4:4 in case of primary distribution of video content with YCbCr 4:2:2 chroma format. The first release of the guidelines is expected in 2023 and will be preceded by an open community review process. Further details on the process and all updates regarding guidelines work will be available through MC-IF website [6].

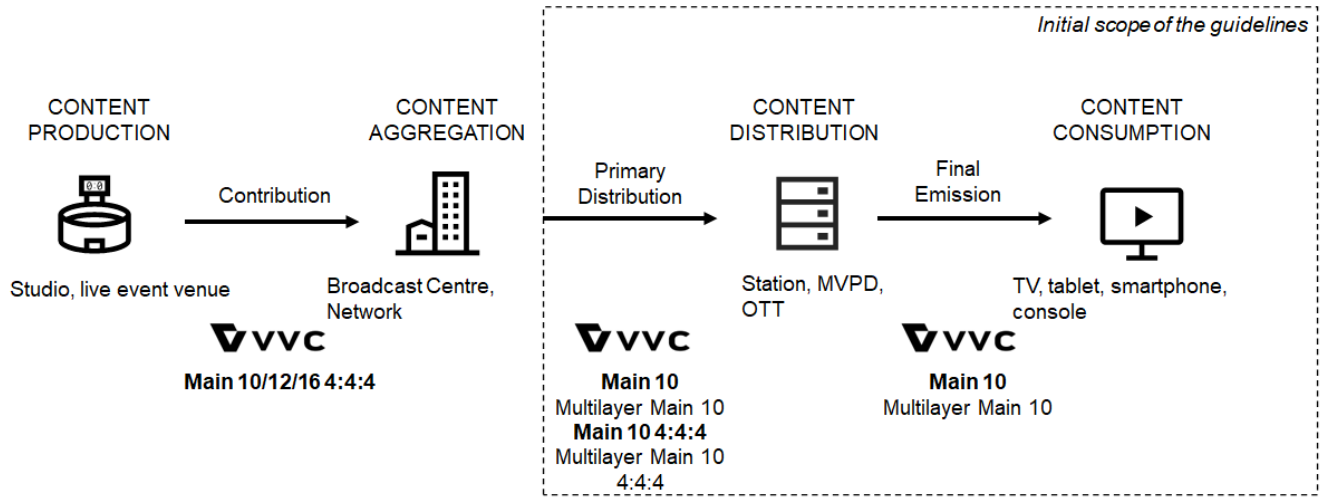

FIGURE 1 SIMPLIFIED END-TO-END BROADCAST AND STREAMING ECOSYSTEM WITH CONTENT CONTRIBUTION, DISTRIBUTION AND FINAL EMISSION STAGES. DASHED PART HIGHLIGHTS THE INITIAL SCOPE OF THE GUIDELINES. VVC PROFILES RELEVANT FOR RESPECTIVE PARTS OF THE ECOSYSTEM ARE HIGHLIGHTED.

The guidelines aim to describe relevant VVC configuration options, in reference to relevant industry VVC profiles, which span through the following areas:

- examples of application of VVC coding tools and VVC performance, including references to industry best practices, relevant external performance results and studies.
- examples of application of VVC functional features including discussion of their benefits and interoperability aspects.
- examples of VVC high-level bitstream operations and signalling.
- examples of VVC usage with associated metadata.

In the following sections we discuss selected examples across these four categories which are investigated for the inclusion in the guidelines. The presented examples do not represent the full set of VVC options and capabilities. For a comprehensive description of VVC please refer to the excellent overview in [49].

VVC coding tools and performance: HDR video

High Dynamic Range (HDR) video became a widespread video format in recent years. Multiple online streaming services such as Netflix, Amazon, Disney+ and others are offering HDR content with hundreds of titles. HDR video services are also defined in specifications of the TV broadcasts SDOs, such as ATSC, DVB and others. Rapid deployment of HDR video services was enabled by development of HDR video formats in SMPTE, ITU-R and HEVC. (HEVC Main 10 profile is capable of compressing 10-bit data and encapsulating HDR metadata into the coded bitstream. Although originally designed without HDR-specific coding tools, HEVC encoders can still be configured for efficient compression of HDR video data as well as signaling HDR-specific metadata. Operational practices of handling HDR video data and tools configurations in HEVC are documented in ITU-T Technical Reports (Supplements) [50] and [51].

The practices documented in [50] and [51] are also relevant for VVC, but VVC takes HDR support to a different level. In VVC, all coding tools are applicable to both SDR and HDR video. The distinction between SDR and HDR is made only in high level syntax. Specifically, the colour primaries and optical-electronic transfer characteristics functions that distinguish SDR, hybrid-log gamma HDR (HDR HLG), and perceptual quantizer HDR (HDR PQ) are signaled in the VVC bitstream using video usability information (VUI) metadata specified in VSEI. VSEI also specifies mastering display colour volume (MDCV) and content light level indication (CLLI) SEI messages that provide metadata that are used to optimize display of SDR and HDR coded video on consumer displays. (See the section **Metadata for VVC** for more information.)

Although all VVC coding tools apply to both SDR and HDR, it is worth noting that some new VVC coding tools were originally studied in the context of HDR and later shown to also benefit SDR. In particular, the coding tools called luma mapping with chroma scaling (LMCS) and luma-intensity based deblocking filtering adjustment were aimed at facilitating compression of HDR video data. The luma mapping of the LMCS adjusts the dynamic range of the input video signal by redistributing the codewords across the dynamic range to improve compression efficiency. The chroma residual scaling of LMCS is utilized to compensate for the interaction between the luma signal and its corresponding chroma signals and reduce visual distortions in coded video with WCG representation (BT.2100) [52]. Example encoding algorithms to derive LMCS and luma intensity deblocking parameters are implemented in the VVC reference software (VTM) with specific optimizations for SDR, HDR HLG, and HDR PQ video data. For more details see sections 3.7.2 and 3.7.3 of [53].

Objective quality metrics used during HEVC and VVC development for HDR video data have been documented in [54]. It was determined that calculating the PSNR for the luma and chroma codewords, as is typically done for SDR video, is still appropriate for some HDR HLG video formats. For HDR PQ video, additional metrics were determined to be more suitable. These metrics include: PSNRL100, wPSNR and DE100, as described below and in more detail in [55].

- PSNRL100: This metric is calculated using the luminance values rather than the luma codewords. It is based on the CIELab colour space representation of the input and output sample values of the codec.
- wPSNR: This is a PSNR-like metric calculated from the codewords that attempts to compensate for the more significant distribution of luma codewords to the darker regions by performing a weighting of the codewords before calculating the PSNR.
- DE100: This is a metric based on the CIELab colour space representation of the input and output sample values of the codec. It is specifically targeted at chrominance fidelity.

The performance of VVC for coding of HDR video data compared to HEVC and other codecs has been assessed on multiple occasions. In document [59], VVC performance assessment was conducted by the 3GPP SA4 working group. The coding performance of VVC for HDR PQ material was compared to the coding performance of HEVC, with both encoders using reference software implementations of JVET. For the set target bitrates, the resulting quality was assessed using objective quality metrics, namely wPSNR, PSNRL100 and DE100 as recommended in [54]. It is reported that VVC provides 39% coding gain on average for the wPSNR metric, 35% for PSNRL100, and around 57% for the DE100 metric.

While objective results provide a good indication of performance, subjective assessment of quality of coding performance is very valuable not only to understand coding performance in relation to other codecs but also in relation to determining commercially relevant operating points. MPEG leads the industry best practice in this area and produces high quality subjective verification tests for its codecs, performed by independent and qualified laboratories with the participation of naïve test subjects. Such tests follow established assessment protocols specified in ITU recommendations ITU-R BT.500 [56] and ITU-T P.910 [57]. MPEG subjective quality verification test results are reported in [32]. The assessment included 4k UHD test sequences with HDR HLG content and HDR PQ content. Reported MOS-over-bitrate results demonstrated overall gain of about 49% for the HLG test sequences and an overall gain of about 52% for the PQ test sequences. Figure 2 shows pooled MOS quality assessment data as a function of bitrate for tested VVC and HEVC codecs. The 11-grade MOS scale used in the test represents a full range of video quality impairments present in tested bitstream, from “0” score corresponding to “severely annoying (everywhere)” to “10” corresponding to “imperceptible”. “8” corresponds to “slightly perceptible (everywhere)” and can be associated with a broadcast quality. Plots in Figure 2 show that for VVC, averaged bitrates for UHD HDR content range from around 7Mbps for PQ content used in the test to around 12.5 Mbps for HLG content used in the test. This constitutes 50% bitrate reduction compared with HEVC at the same MOS quality achieved with average bitrates from around 15Mbps for PQ content used in the test to around 27Mbps for HLG content used in the test.

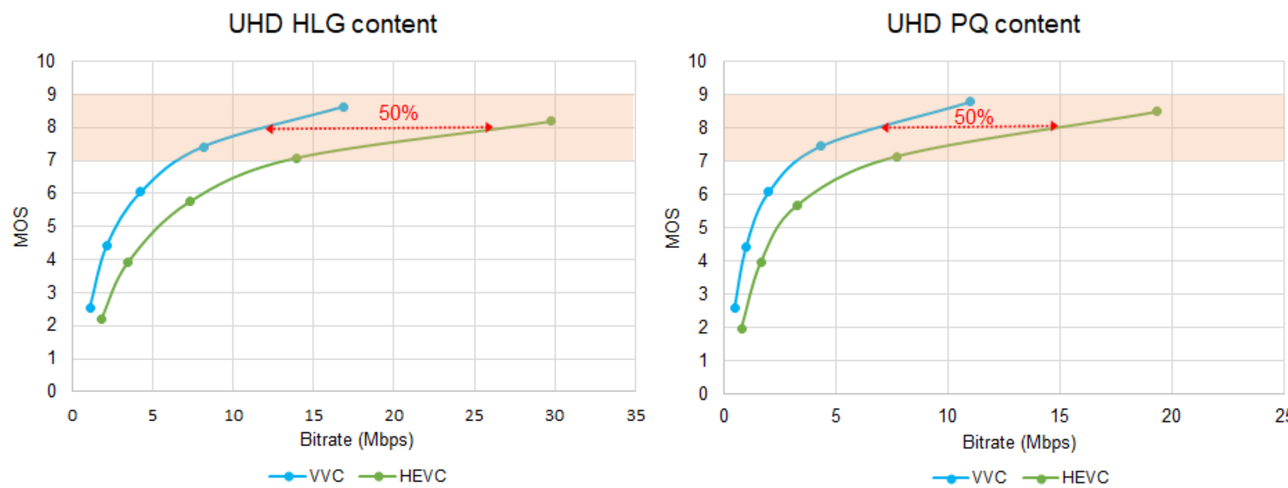

FIGURE 2: MOS QUALITY OVER BITRATE FOR UHD HLG AND PQ CONTENT POOLED OVER 5 TEST SEQUENCES IN EACH CATEGORY. DATA TO CREATE THESE PLOTS WAS USED FROM [32].

VVC functional features: reference picture resampling

Delivery of video services by the means of providing multiple renditions of content at varying resolutions and bitrates has been the backbone of adaptive streaming for broadband (DASH or HLS). Initially, designed to target widely varying end-device display formats and variable available bandwidth, techniques such as “Per-title encoding” were introduced to dynamically select optimal resolutions (from a QoE perspective) for each VOD content. Decoding of short segments of video with different bitrates and resolution could be achieved with the use of closed GOP coding, which inserts IDR (Instantaneous Decoder Refresh) at the start of each segment. In terms of compression performance, the closed GOP configuration is less efficient than the open GOP configuration typically used in broadcast. The coding penalty increases with higher frequency of random access points i.e., shorter segments as IDR frames are less efficiently coded which in effect cause bitrate spikes impacting rate control buffers and makes the penalty higher for low latency streaming. Previous attempts with seamless switching of resolutions with HEVC were reported in [60] with the use of the scalable profile of HEVC and more recently in [61] with the use of Main 10 HEVC profile. However, the former approach could be viable only to applications supporting scalable HEVC profiles, while the interoperability tests with HEVC showed issues at switching since such a functionality had not been tested at the time of deployment.

VVC includes a built-in spatial scalability functionality which allows changing resolution at any inter-coded picture. The mechanism uses reference picture resampling (RPR) which effectively adjusts the resolution of reference pictures used for prediction on the fly. This functionality is supported in the Main 10 profile while Multilayer profiles extend this through efficient high level signaling, however, no additional inter-layer signal processing tools are required [49]. In [62] authors investigated VVC open GOP resolution switching in the adaptive streaming use case and identified potential issues with prediction-error drift in random access skipped leading (RASL) pictures due to change of resolution at RAP (Intra picture) which precede RASL pictures in coding order. It was, however, demonstrated that exercising certain constraints in the encoder mitigated most of the issues and resulted in compression gains up to 9% BD-rate over typical closed GOP setting for low latency live streaming. Reference picture resampling functionality also enables changing resolution within a representation. While VVC allows high flexibility with regards to scaling factors and frequency of resolution change at any picture, application specifications may provide some constraints on how often and which resolutions can be used by the codec. For example, the new revision of DVB TS 101 154 [45] , which allows RPR use in VVC profiles, restricts downsampling scaling factors to 2/3 and 1/2 while new coded resolutions can be introduced at minimum interval of 2 seconds. However, different application standards may choose to adopt different sets of constraints for the use of RPR functionality.

In order to study the benefits of VVC RPR in the context of resolution change within a representation (e.g., as would be used in live broadcast), we used the VTM reference software implementation with RPR encoder functionality from [58]. The encoding algorithm implemented in VTM aims to serve as an illustration of RPR functionality and is configured to limit resolution change at high bitrates while more likely to trigger downsampling at lower bitrates. Resolution change decisions can be made for each reordering segment of pictures (hierarchical B pictures structure) and can use scaling factors: 4/5, 2/3 and 1/2. For the test content we used sequences from 3GPP SA4 5G video test set from the study described in section **3GPP SA4**. Sequences used were from the SDR set of 4K-TV scenario, (HDR results were reported in the section above), for which overall VVC performance gains achieved with reference encoder VTM were reported as 37% BD-rate with PSNR-YUV, 33% BD-rate with PSNR-Y, 33% BD-rate with MS-SSIM and 38% BD-rate with VMAF metrics respectively. Details of these metrics as used in the 3GPP 5G codec study are detailed in section 5.5 of [59]. In our test, we used 4 fixed QP settings (22, 27, 32 and 37) for VTM software as in the original study plus QP 42 at the low end of bitrate scale to align VVC and HEVC quality at the lowest bitrates.

Our tests showed substantial coding gains could be achieved for two out of eight sequences from the test set: Soccer [63] and Tunnel Flag [64]. For three of the sequences, either negligible gain or loss were observed and for the remaining three sequences RPR functionality was not triggered at all which is a correct behavior in case encoder does not predict content benefiting from change of resolution. However, this is still an interesting result since the sequences in the test set were not pre-selected towards showcasing RPR functionality and RPR gains are expected to be content dependent.

Results in Table 1 show extra coding gain can be achieved over HEVC with the use of the RPR coding tool in VVC. Coding gain was consistent for all reported metrics, smaller difference reported by PSNR-based metrics and higher with MS-SSIM and VMAF. Looking at the rate-distortion plots in Figure 2, we notice that coding gain is contributed through a higher quality metric value rather than shifted bitrates. Visual inspection showed that VVC is generally cleaner around the players than HEVC while VVC with RPR looks more consistent than VVC and suffers less from visible artefacts in the background and textured areas like the one highlighted in Figure 3 for Soccer sequence. Figure 4 shows RPR operation for Tunnel Flag sequence at 5.2 Mbps. The sequence comprises two parts, first part which is drive through a tunnel which then cross-fades into a waving flag part. As it is shown in Figure 5, RPR encoding algorithm utilizes three different scaling factors: 4/5 and 2/3 for the tunnel part and 1/2 scaling for the flag part, where the change occurs right after the cross-fade. While we did not conduct a detailed analysis of encoder complexity impact (encodings were run on a cluster with a mix of CPUs), ballpark results showed a decrease of about 15%-20% in computing resources when RPR was employed.

	BD-rate gains			
Test	PSNR-Y [%]	PSNR-YUV [%]	MS-SSIM [%]	VMAF [%]
Soccer				
VVC vs. HEVC	40.34%	39.27%	39.26%	40.23%
VVC with RPR vs. HEVC	44.10%	43.02%	43.67%	48.77%
Tunnel Flag				
VVC vs. HEVC	49.57%	52.64%	52.86%	55.19%
VVC with RPR vs. HEVC	51.08%	54.38%	57.21%	61.55%

TABLE 1: BD-RATE GAINS FOR VVC AND VVC WITH RPR FUNCTIONALITY OVER HEVC.

RPR is a new promising functionality offered by VVC which could benefit both resolution switching in ABR video delivery as well as introduce dynamic changes of resolution within a bitstream or representation. In the latter case, it can either provide coding gain or optimize bitrate delivery for the same QoE, e.g., in statistical multiplexing by allowing greater flexibility in allocating bitrates across programs. When employed, RPR can also have an impact on CPU encoder complexity and therefore could contribute to reducing power consumption in the headend. The functionality is supported with some operational constraints in the DVB codec specification [45] and scalability is also considered for SBTVD TV 3.0 specification requirements [47].

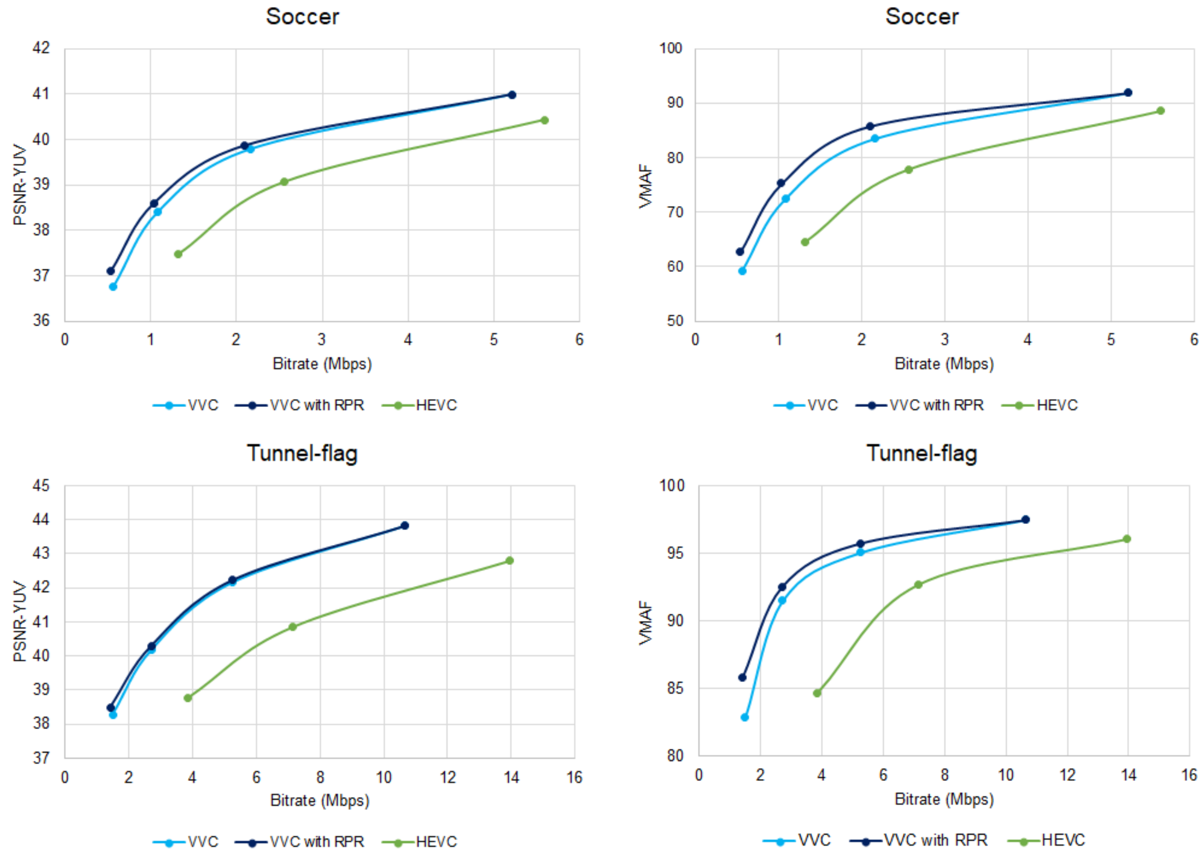

FIGURE 3: RATE-DISTORTION PLOTS FOR SOCCER AND TUNNEL FLAG SEQUENCES FOR PSNR-YUV AND VMAF METRICS.

FIGURE 4: SCREEN SHOT FROM DECODED SEQUENCE FROM HEVC BITSTREAM (LEFT), VVC BITSTREAM (MIDDLE) AND VVC BITSTREAM WITH RPR (RIGHT). VVC WITH RPR DECODED PICTURES WERE UPSCALED TO THE ORIGINAL 4K RESOLUTION. HIGHLIGHTED AREAS POINT TO VISIBLE ARTEFACTS: HALO AROUND PLAYER'S SHOULDER (LEFT) AND SOFT PATCH IN THE ASTRO TURF (MIDDLE) SEQUENCE USED IN THIS TEST IS FROM [63].

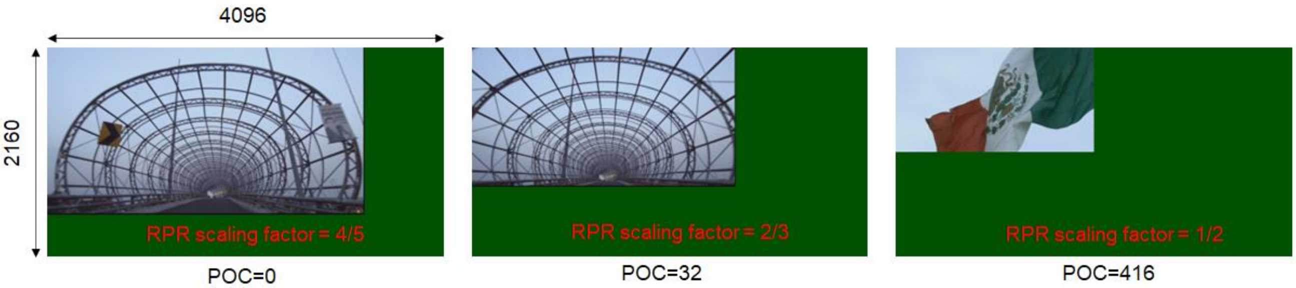

FIGURE 5: VISUALIZATION OF RPR OPERATION FOR TUNNEL FLAG SEQUENCE AT QP 32 (5.2 MBPS). 4096X2160 IS THE ORIGINAL RESOLUTION OF THE SEQUENCE WHICH IS REPRESENTED WITH THE GREEN-PADDED AREA. IMAGES REPRESENT DECODED PICTURES RESOLUTION WITH POC (PICTURE ORDER COUNT) REPRESENTING TIMINGS WHERE CHANGES OF RESOLUTION OCCURRED. SEQUENCE USED IN THIS TEST IS FROM [64].

High-level bitstream operations: VVC subpictures

Improved service accessibility is one of increasingly important requirements for video delivery systems. One example is provision of supplemental video bitstream with sign language interpreter [47], where the supplemental video is displayed on request from the end-user. VVC provides functionality for flexible bitstream extraction and merging operations. This functionality is essential in VVC design and provides significant improvement for coding high resolution immersive video such as in VR 360º video applications. Bitstream extraction and merging can be supported with the use of VVC subpictures which are one of the picture partitioning tools available in VVC. Subpictures may be coded independently of other subpictures and can be flexibly rearranged. A detailed overview of VVC subpictures functionality is available in [65].

One way to use this feature is to provide two video bitstreams, main video bitstream which includes sub-pictures arrangement which corresponds to the desired composition, and a supplemental video at lower resolution which can be composed into the main video bitstream mapping directly into the subpictures arrangement in the main video. Such a bitstream operation is performed before video decoding. Restrictions on ensuring the same video characteristics such as bit depth, frame rate, colour space etc., need to be observed. More information on the usage of VVC subpictures in the context of accessibility and personalization for video services can be found in Annex M of [45].

Metadata for VVC

With a few special exceptions, the VVC standard only specifies the syntax, semantics, and decoding processes required for conforming video decoders. Information about how video is intended to be post-processed, displayed, or otherwise used is specified mostly in the VSEI standard. VSEI specifies two kinds of metadata: video useability information (VUI) and SEI messages.

VUI parameters provide information for the correct display of coded video. For example, VUI parameters indicate that the coded video is progressive or interlaced; SDR, HDR HLG, or HDR PQ; that the source video has a particular aspect ratio; and any other parameters necessary for correct interpretation of the coded video.

SEI messages provide additional information that can assist decoders, displays, and other video receivers perform as desired by the content producer. For example, the display of omnidirectional 360º-video can be assisted by signalling one of the omnidirectional video specific SEI messages in the VVC bitstream. As another example, the annotated regions SEI message can be used to assist video analysis applications by signalling the size and location of objects identified in a video.

As noted previously, several SEI messages were originally developed in the context of HDR and have since become incorporated into application standards. The mastering display colour volume (MDCV)

SEI message provides information about how source video looked to creative professionals when the content was created (i.e., mastered). Consumer displays can use the information in the MDCV SEI message to adjust how coded video is displayed so that it matches the source look as closely as possible. The content light level information (CLLI) SEI message indicates how bright the source video was when mastered. Information in the CLLI SEI message can also be used to optimize the look of displayed video and has the potential to enable consumer displays to improve power management. Other SEI messages including the content colour volume (CCV), ambient viewing environment (AVE), and alternative transfer characteristics (ACT) SEI messages can also enable consumer displays to process coded video content to provide better consumer experiences. It is worth noting that although many of the mentioned SEI messages were developed initially for HDR, they have become increasingly applicable to SDR video data as the brightness of SDR consumer displays has increased.

The film grain characteristics (FGC) SEI message has also become increasingly significant with the increase of film grain synthesis in high-value streaming services [65]. The FGS SEI message signals the statistical characteristics of spatial temporal grain added to decoded video. There are two main use cases – one old and one new – for which adding grain to decoded video can be beneficial. The first and original use case was to replace actual film grain that had been filtered out of source video prior to encoding to save bandwidth. The newer use case is to subtly mask compression artefacts (even for source video that did not have original film grain) to reduce bitrate and improve perceived sharpness. A Technical Report on use of film grain technologies is currently in development in ITU-T and ISO/IEC with publication expected in the 2nd half of 2023. Software illustrating the use of the FGS SEI message are available in the AVC, HEVC, and VVC reference software. Open-source film grain synthesis implementations are available in GitHub repositories [67][68].

Although not yet finalized, a pair of neural-network post-filter (NNPF) SEI messages [69] have attracted significant attention by enabling the use of neural networks for post-processing operations such as super-resolution, frame rate upsampling, chroma format conversion, and colourization. The NNPF characteristics (NNPFC) SEI message signals neural-network parameters/weights (contained in an ISO/IEC 15938-17 bitstream [70] or identified by a universal resource identifier [71]) and additional information needed for a receiver to determine if it can implement the indicated neural network. Several neural networks can be signalled using multiple NNPFC SEI messages to support different receiver capabilities and different post-processing operations. Specific neural networks are invoked from the available neural networks using an NNPF activation (NNPFA) SEI message. It is expected that the NNPFC and NNPFA SEI messages will be finalized in VSEI in the 2nd half of 2023.

Conclusions

Since its finalization in 2020, VVC has enjoyed a steady adoption progress across various implementations spanning the whole end-to-end chain. Also, application layer SDOs are looking into advancements in video codecs in order to further improve penetration of existing services while creating new operating points for delivery of enhanced video services with formats such as 8K. While first adoptions of VVC into application specifications for broadcast and streaming, MC-IF has launched work on VVC technical guidelines. Such guidelines are aimed to cover VVC configuration aspects including its rich combination of coding tools, functionality and associated metadata, to facilitate VVC adoption and facilitate VVC interoperability across products and services. In this paper we provided an overview of VVC’s deployment status across implementation and standards, discussed the scope of the launched work in MC-IF on VVC technical guidelines and provided selected examples of VVC configuration aspects which demonstrate VVC superiority in terms of coding performance and functionality against the needs of broadcast and streaming industry, and together with other aspects of VVC are being investigated for inclusion in the MC-IF VVC technical guidelines.

References

[1] ITU-T | ISO/IEC, "Rec. ITU-T H.266 | ISO/IEC 23090-3 "Information technology Coded representation of immersive media - Part 3: Versatile Video Coding”," 2022.

[2] ITU-T | ISO/IEC, "Rec. ITU-T H.274 | ISO/IEC 23002-7 "Versatile supplemental enhancement information messages for coded video bitstreams"," 2022.

[3] ITU-T | ISO/IEC, “Rec. ITU-T H.222.0 | ISO/IEC 13818-1: "Information technology - Generic Coding of moving pictures and associated audio information: Systems",” 2022

[4] ISO/IEC, “ISO/IEC: 14496-15:2017: "Information technology -- Coding of audio-visual objects -- Part 15: Carriage of network abstraction layer (NAL) unit structured video in ISO base media file format”,” 2022.

[5] ISO/IEC, “ISO/IEC 23000-19:2019/FDIS: "Information technology -- Multimedia application format (MPEG-A) -- Part 19: Common media application format (CMAF) for segmented media (2nd edition)"," 2022.

[6] Media Coding Industry Forum, https://www.mc-if.org/.

[7] Bossen, F., “Performance of a reasonably fast VVC software decoder”, *JVET-S0224 ,* July 2020.

[8] Wieckowski, A., et al., “Open optimized VVC encoder (VVenC) and decoder (VvdeC) implementations,” *JVET-T0099,* October 2020.

[9] Bossen, F., Sühring, K., Wieckowski, A., Liu, S., “VVC Complexity and Software Implementation Analysis,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 31, Issue 10, October 2021, pp. 3765–3778.

[10] Li., Y., et al., “Performance of a VVC Software Decoder on Mobile Platform,” *JVET-U0071,* January 2021.

[11] Yu, L., et al., “VVC software decoder implementation for mobile devices,” *JVET-V0132,* April 2021.

[12] Feng, W.-L., et al., “VVC Software Decoder for Mobile Platforms,” *JVET-V0070,* April 2021.

[13] Li, L., et al., “Update on a VVC software decoder, BVC, for heterogeneous CPU plus GPU systems,” *JVET-Y0054,* January 2022.

[14] Wieckowski, A., et al., “Update on open, optimized VVC implementations VVenC and VVdeC” *JVET-AB0044,* October 2022.

[15] Fraunhofer HHI, “VVdeC Web Player,” https://github.com/fraunhoferhhi/vvdecWebPlayer, accessed February 2023.

[16] Spin Digital, “Spin Digital Announces a VVC 8K Decoder and Media Player”, June 2021

[17] Allegro DVT, https://www.allegrodvt.com/products/al-d320-decoder-ip/, accessed February 2023.

[18] Realtek, “Realtek Launches World’s First 4K UHD Set-top Box SoC (RTD1319D) Supports VVC/H.266 Video Decoding, GPU with 10-bit Graphics, Multiple CAS, and HDMI 2.1a”, 29 August 2022.

[19] MediaTek, MediaTek | Pentonic 2000 | Flagship 8K TV SoC, accessed February 2023.

[20] Larsen, R., “Sony and Philips will be the first TV makers to use MediaTek Pentonic chips”, “*FlatPanelsHD”*, 9 January 2023.

[21] ITU-T | ISO/IEC Joint Video Experts Team (JVET), “VVC verification test report for UHD SDR video content,” *JVET-T2020 ,* October 2020.

[22] Cui, J., et al., “An optimized VVC encoder implementation,” *JVET-V0127*, April 2021.

[23] Dong, X., et al., “Update on the progress of the optimized VVC encoder implementation, Ali266” *JVET-X0104,* October 2021.

[24] Bitmovin, “Bitmovin to Unveil Next-Generation VOD Encoder at the 2022 NAB Show”, press release, 30 March 2022.

[25] Tencent, “Tencent Cloud Becomes First Cloud Service Provider to Support H.266/VVC Standard”, press release, July 16, 2021.

[26] Jia, Y., et al., “Ali266 @ Youku: trial deployment of VVC for video streaming,” *JVET-Y0122,* January 2022.

[27] "MX Player becomes first OTT to deploy H.266, cuts down video streaming data consumption into half", *Best Media Info,* 16 June 2021
[28] *TVNewsCheck* "IBC 2022: MainConcept Launches VVC/H.266 SDK Beta And Demos Cloud-Ready 8K Broadcast Technology",*TVNewsCheck*, 31 August 2022.
[29] Sanz-Rodriguez, S., Alvarez-Mesa, M. and Ching Chi, Ch., "A VVC/H.266 real-time software encoder for UHD live video applications," *JVET-AB0043*, October 2022.
[30] Le Tanou, J., "MediaKind enables live UHD OTT VVC streaming", "https://www.mediakind.com/blog/mediakind-enables-live-uhd-ott-vvc-streaming-on-a-public-cloud/", 21 September 2022.
[31] Biatek, T., Abdoli, M., Guionnet, T., Nasrallah, A., and Raulet, M., "Future MPEG standards VVC and EVC: 8K broadcast enabler," *IBC*, 2020.
[32] ITU-T | ISO/IEC Joint Video Experts Team (JVET), "VVC verification test report for high dynamic range video content," *JVET-W2020*, July 2021.
[33] NHK, "CMAF Multilayer-compatible VVC Decoder", press release, https://www.nhk.or.jp/strl/english/news/2022/5.html, 2022.
[34] ITU-T | ISO/IEC, "Rec. ITU-T H.266.1 | ISO/IEC 23090-15 " Information technology — Coded representation of immersive media — Part 15: Conformance testing for versatile video coding"," 2022.
[35] DVB, "VVC test content", https://dvb.org/specifications/verification-validation/vvc-test-content/, 2 February 2023.
[36] Allegro DVT, H266 VVC Standard JVET Syntax Compliance Bitstreams Validation Decoder (allegrodvt.com), accessed February 2023.
[37] W. Hamidouche et al., "Versatile Video Coding Standard: A Review from Coding Tools to Consumers Deployment," *IEEE Consumer Electronics Magazine*, vol. 11, no. 5, September 2022, pp. 10-24.
[38] Bitmovin, "The 6th annual Bitmovin Video Developer Report," December 2022.
[39] Sullivan, G.J., "Deployment status of the VVC standard," *JVET-AC0021*, January 2023.
[40] Ericsson, "Ericsson Mobility Report," November 2022.
[41] ATSC, "ATSC Planning Team Report: ATSC 3.0 and Global Convergence", December 2022
[42] DASH Industry Forum, "DASH-IF Interoperability Points; Part 7: Video," May 2022.
[43] CTA, "CTA Specification, Web Application Video Ecosystem – Content Specification, CTA-5001-D," 2021.
[44] Power, J., "New video codecs for DVB services: fast-forward to 2024…," *DVB Scene,* Issue 58, September 2021.
[45] DVB, "Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications," *DVB Document A001 Rev.21*, November 2022.
[46] Drugeon, V., "New in the DVB codec toolbox: Versatile Video Coding (VVC)," *DVB Scene*, Issue 59, March 2022.
[47] SBTVD Forum, TV3.0 Call for proposals, https://forumsbtvd.org.br/wp-content/uploads/2020/07/SBTVDTV-3-0-CfP.pdf
[48] SBTVD Forum, Fórum SBTVD | TV 3.0 Project (forumsbtvd.org.br), accessed February 2023.
[49] B. Bross et al., "Overview of the Versatile Video Coding (VVC) Standard and its Applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, October 2021 pp. 3736-3764.
[50] ITU-T, "H. Sup15: Conversion and coding practices for HDR/WCG Y'CbCr 4:2:0 video with PQ transfer characteristics," 2017.
[51] ITU-T, "H. Sup18: Signalling, backward compatibility and display adaptation for HDR/WCG video coding," 2017.
[52] ITU-R, "Rec. ITU-R BT.2100: Image parameter values for high dynamic range television for use in production and international programme exchange," 2020.
[53] ITU-T | ISO/IEC Joint Video Experts Team (JVET), "Algorithm description for Versatile Video Coding and Test Model 18 (VTM18)," *JVET-AB2002*, October 2022.

[54] ITU-T, “Technical Paper ITU-T HSTP-VID-WPOM (07/2020): Working practices using objective metrics for evaluation of video coding efficiency experiments," 2020.
[55] ITU-T | ISO/IEC Joint Video Experts Team (JVET), “VTM and HM common test conditions and evaluation procedures for HDR/WCG video,” *JVET-AC2011*, January 2023.
[56] ITU-R, “Rec. ITU-R BT.500: Methodologies for the subjective assessment of the quality of television images,” 2019.
[57] ITU-R, “Rec. ITU-T P.910: Subjective video quality assessment methods for multimedia applications,” July 2022.
[58] Andersson., K., et al., “GOP-based RPR encoder control,”*JVET-AB0080,* October 2022*.*
[59] 3GPP, “TR: 26.955: Video codec characteristics for 5G-based services and applications,” June 2022.
[60] Yan, Y., Hannuksela M. M., Li, H., "Seamless switching of H.265/HEVC-coded dash representations with open GOP prediction structure,” *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 4121-4125.
[61] Ducloux, X., Diascorn, J-L., Fautier, T., “Exploring the benefits of dynamic resolution encoding and support in DVB standards," *IBC*, 2022.
[62] Skupin, R., et al., “Open GOP resolution switching in HTTP adaptive streaming with VVC,” *Proc. 35th Picture Coding Symp. (PCS)*, June 2021, pp. 1–5.
[63] Cable Labs, “Soccer,”, https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo/ReferenceSequences/Soccer/Soccer.json, accessed February 2023.
[64] Netflix, “Tunnel flag”, https://dash-large-files.akamaized.net/WAVE/3GPP/5GVideo/ReferenceSequences/Tunnel-Flag/Tunnel-Flag.json, accessed February 2023.
[65] Wang, Y. -K., et al., "The High-Level Syntax of the Versatile Video Coding (VVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, October 2021, pp. 3779-3800.
[66] Norkin, A., and Birkbeck, N., “Film Grain Synthesis for AV1 Video Codec,” *2018 Data Compression Conference*, March 2018, pp. 3–12.
[67] Ittiam Systems, “Film Grain Synthesis Library,”, https://github.com/ittiam-systems/libfgs, accessed March 2023.
[68] InterDigital Inc., “VersatileFilmGrain,”, https://github.com/InterDigitalInc/VersatileFilmGrain, accessed February 2023.
[69] McCarthy. S., Deshpande, S., Hannuksela, M. M., Hendry, Sullivan, G. J., and Wang, Y.-K. “Improvements under consideration for neural network post filter SEI messages,” *JVET- AC2032,* January 2023.
[70] ISO/IEC, “ISO/IEC: 15938-17, Information technology – Multimedia content description interface – Part 17: Compression of neural networks for multimedia content description and analysis,” 2022.
[71] IETF, “IETF Standard 66, Uniform Resource Identifiers (URI): Generic Syntax,” http://tools.ietf.org/html/rfc3986