SCTE ISBE

JOURNAL OF DIGITAL VIDEO

Volume 1, Number 2
September 2016

www.scte.org | www.isbe.org

Society of Cable
Telecommunications
Engineers

International
Society of Broadband
Experts

JOURNAL OF DIGITAL VIDEO

VOLUME 1, NUMBER 2
September 2016

Society of Cable Telecommunications Engineers, Inc.
International Society of Broadband Experts™
140 Philips Road, Exton, PA 19341-1318

Table of Contents

Technical Papers

Letter to the Editor

Editorial Correspondence: If there are errors or omissions to the information provided in this journal, corrections may be sent to our editorial department. Address to: SCTE Journals, SCTE/ISBE, 140 Philips Road, Exton, PA 19341-1318 or email journals@scte.org.

Submissions: If you have ideas or topics for future journal articles, please let us know. Topics must be relevant to our membership and fall under the technologies covered by each respective journal. All submissions will be peer reviewed and published at the discretion of SCTE/ISBE. Electronic submissions are preferred, and should be submitted to SCTE Journals, SCTE/ISBE, 140 Philips Road, Exton, PA 19341-1318 or email journals@scte.org.

Subscriptions: Access to technical journals is a benefit of SCTE/ISBE Membership. Nonmembers can join at www.scte.org/join.

From the Editors

Welcome to the second issue of the *Journal of Digital Video*, a publication of collected papers by the Society of Cable Telecommunications Engineers (SCTE) and its global arm, the International Society of Broadband Experts (ISBE). This September 2016 issue focuses on topics such as advanced video standards and delivery practices, including in particular high dynamic range (HDR), wide color gamut (WCG), dynamic adaptive streaming over HTTP (DASH), virtual reality (VR), the move towards all-IP video transport, and video quality testing and assurance (including closed captioning).

The first four papers meet the theme of SCTE/ISBE Cable-Tec Expo 2016's "Innovation Meets Here". In particular, the article "Methods to Reduce Visual Sickness in Design of New VR Services" explores factors contributing to visually-induced motion sickness and visual discomfort and the understanding reasons why these continue to be obstacles for widespread acceptance of head-mounted displays. The "Ultra-High Definition for Cable Applications" creates anticipation that soon a standard that will define how cable operators may acquire ultra-high definition (UHD) signals from broadcasters or network providers. The article "Minimizing Delivery Latency of Linear Video Service via Adaptive Transport Stream and HTTP Chunked Transfer Encoding" discusses major shortcomings facing HTTP-based transport for live streaming or video, examines existing or under-developing protocol and standards and proposes a solution that will minimize the delivery latency for live/linear video service without sacrificing video encoding efficiency. The article "End-To-End IP Video Services Over 10g EPON Access Network Architectures", in addition to being true to its title, presents architectures and use cases to investigate migration options along with SDN/NFV solutions and provides insights on next steps, with focus on QoS functions in a more programmable network with dynamic service aware resource allocation through centralized control and service assurance.

The remaining two articles further enrich the SCTE/ISBE "Essential Knowledge for Cable Professionals™" promise. In particular, the technical paper "Video Quality STB Evaluation through the Application of a MOS Scoring Framework" skillfully re-introduces Golden Eye experts and effortlessly weaves in the mean opinion score (MOS) techniques to augment the video quality evaluations of settop box devices. The last article, "Efficiently Troubleshoot Closed Captions – and Keep the FCC at Bay", written as a Letter to the Editor, readdresses how important and critical it is for content distributors and video service providers to have tools on hand to look at captions deeply enough to identify the root cause of problems.

We would like to thank the individuals who contributed to this issue of *the Journal of Digital Video*, including the authors, peer reviewers, and the SCTE/ISBE publications and marketing staff. We hope the readers enjoy this issue and that the selected papers spark innovative ideas and further cement essential knowledge in digital video.

In closing, if there is any editorial information or topics that you would like us to consider for the third issue of SCTE/ISBE Journal of Digital Video, please refer to the "editorial correspondence" and "submissions" sections at the bottom of the table of contents for further instructions.

SCTE/ISBE Journal of Digital Video Senior Editors,

Paul Hearty
Vice President- Technology Standards Office, Sony Electronics, Inc.

Niem Dang
Director- Engineering and Standards

JOURNAL OF
DIGITAL VIDEO

Methods to Reduce Visual Sickness in Design of New VR Services

A Technical Paper prepared for SCTE/ISBE by

Khizer R. Khaderi MD, MPH, CEO and Founder, Vizzario Inc.,
Adjunct Associate Professor University of Utah Moran Eye Center
1701 Penmar Ave Venice, CA 90291
Khizer@vizzario.com
801-949-8708

Mohan K. Reddy, CTO Vizzario Inc.
1701 Penmar Ave Venice, CA 90291
mohan@vizzario.com
832-651-2986

Yasser F. Syed, Comcast Distinguished Engineer,
T&P-Video Architecture Group/Comcast Cable, SCTE/ISBE Member
1701 JFK Boulevard Philadelphia, PA 19803
yasser_syed@cable.comcast.com
303-246-8413

Table of Contents

List of Figures

1. Introduction

Although the origins of head-mounted visual displays (HMDs) trace back to the 1960s[1], widespread adoption of this technology in the form of virtual reality (VR) HMDs has only recently taken place. VR is a non-invasive simulation technology that provides an immersive, realistic, three-dimensional (3D) computer-simulated environment in which people perform tasks and experience activities as if they were in the real world. HMDs are goggles, helmets and glasses comprised of lens display systems. These display systems are monocular (one eye), binocular (both eyes, one screen) or dichoptic in nature (both eyes, different screens or image/eye), the latter allowing for stereopsis (depth cues). Early commercially available HMDs, such as the Virtual Research Flight Helmet, Virtual I/O I-Glasses™, and Dynovisor, had limited applications due to their narrow field-of-view (FOV) and form factors including weight, design, and system parameters. Recent advancements have been directed toward making HMDs more comfortable for longer duration of use, with HMD products including Mobile VR (Google Glass, Samsung Gear) and Hardware-dedicated VR (Oculus Rift, HTC Vive) becoming commercially available.

These technical advancements include higher resolution and frame rates, lower persistence and lower latency from improved positional tracking. While advancements have made VR systems more commonplace, an inherent problem with these devices, visually induced motion sickness (VIMS) or simulation sickness, remains an obstacle to the widespread adoption and commercial development of technologies associated with VR based HMDs.[2, 3] VIMS which is related, in part, to visual-vestibular mismatch, has been attributed to significant systemic and perceptual problems with HMDs not commonly experienced with traditional displays. The manifestation of these systemic and perceptual problems include nausea, stomach discomfort, disorientation, postural instability and visual discomfort. In addressing VIMS, one must understand it is a multifactorial process, which allows for multiple solutions to minimize, if not eliminate negative VR experiences. We have broken the potential inciting factors into four categories: Hardware, User, Task and Service factors. Finding method(s) to optimize all four factors may bring us one step closer towards wide spread VR adoption, as well as providing guidance for both VR hardware and software developers.

2. Visually Induced Motion Sickness (VIMS)

Sensory input conflicts are notable in virtual environments (VEs), due to the variability in cues, including position and movement, leading to a disharmonious effect on the visual and vestibular systems, including symptoms of nausea and postural instability.[4, 5] Specific types of HMDs may have mismatch problems with the user's visual system due to improper optical design, resulting in convergence-accommodation conflict and visual discomfort or fatigue.[6-12] Continuous conflict between convergence-accommodation, the user's inter-pupillary distance (IPD), and/or the systems' inter-optical distance (IOD) may induce fixation disparity and heterophoria.[13, 17, 22-24] Moreover, visual symptoms experienced are not time-dependent in terms of virtual environment (VE) immersion; rather, symptoms of visual fatigue, reduced visual acuity and heterophoria, may continue after terminating exposure to HMD-based VE .[20, 25-27]

The aforementioned sensory input conflicts should be viewed in context with other factors that impact HMD-induced VIMS. These factors include the characteristics of the hardware systems, the users, and the content the user is exposed to in the VE.

3. Hardware Factors

3.1. Field of View (FOV)

Hardware device system variables impacting VIMS include viewing mode (e.g. monocular, binocular or dichoptic), headset design (e.g. fit, weight), optics (e.g. misalignment in the optics; contrast, luminance), field of view (FOV), and time lag (i.e. transport delay). For example, field of view (FOV), may be implicated in producing visual discomfort symptoms.[27, 34] The FOV studies show that narrow FOV (<50 degrees) reduces the perception of self-motion and wide FOV (>100 degrees) may increase the presence and level of simulator sickness. Patterson et al. recommend at least 60° FOV for a full immersion experience.[60] The sense of immersion can be provided by parsing horizontal and vertical FOVs, which allows for flexibility in the content presentation. For example, in flight simulation applications, segmenting object presentation within a horizontal FOV of 40° by a vertical FOV of 30° improves the ergonomics and improves pilot performance.

3.2. Resolution, Framerate

Aside from influencing overall image quality, resolution may also affect a user's experience of VIMS. It may be uncomfortable to view low-quality images that are noisy or blurry. The visual resolution in humans is 1 minute of arc and is a technological limitation to many HMD systems. Depending on perceived distance in the VR environment, increased resolution mitigates "pixel perception" as objects gets closer. Refresh or Framerate is another factor affecting visual comfort.

3.3. Time Lag or Latency

Time lag between the individual and the system's action and reaction potentially could influence a user's experience of VIMS symptoms, as it affects human perception of visual and vestibular cues.[26, 34, 36, 37] Therefore, reducing the sensor error of HMD systems may minimize the VIMS experience. HMD optical characteristics, such as eye relief (a fixed distance from the eyepiece lens to its exit pupil), convergence demand, horizontal disparity, vertical misalignment of displays, inter-ocular rotation difference, vertical-horizontal magnification differences, luminance, focus differences, temporal asynchrony, focal distance, field curvature difference and inter-pupillary distance (IPD), are all potential factors that can induce visual discomfort and headache when they are misaligned or not optimally calibrated.[28-30, 38, 39, 41-43, 60-64]

4. User Factors

4.1. Age & Gender

Another factor related to the impact of VIMS is user characteristics. Individuals differ in their susceptibility to VIMS.[65] Age has been shown to have a significant relationship with HMD-related eyestrain symptoms.[40] Children 2-12 years of age have immature visual systems and binocular function that is worse than that of adults; this makes children more susceptible to both visual discomfort caused by HMDs and oculomotor side effects including reduced visual acuity, amblyopia, or strabismus.[18, 24, 66, 67] Adults with limited fusional ranges experienced more visual discomfort, specifically with convergent eye movement in response to stimuli in VEs (Karpicka E, unpublished data).[79] Therefore, age effect on HMDs needs to be further studied and incorporated into the design of future HMDs. In regards to gender, female participants reported more simulator sickness and more often withdrew from HMD-based VEs when compared to male participants.[14, 16, 26, 31, 32, 40] One possible explanation is the gender difference in FOV, with female having a wider FOV, increasing the risk for flicker perception, vection and motion sickness susceptibility.[69]

4.2. Visual Deficits

People with visual deficits may have an increased susceptibility to oculomotor side effects compared to those without such deficits, though more studies are needed in this area. A past history of motion sickness or conditions that preclude these symptoms (migraines), are also notable for predicting susceptibility to motion sickness in HMD-based VEs.[32] Individuals may habituate or adapt to HMD-based VEs (i.e. plasticity), with improvement in symptoms after repeated exposure to VE environments. However, this habituation is variable among groups, with certain individuals adapting much more readily than others following repeated exposures to stimuli.[33, 80-51, 53]

4.3. Plasticity

Individuals with more plasticity may be less likely to experience VIMS, though there may be variability in the amount of time needed to adapt to the VE. Greater plasticity does not translate to reduction or lack of initial symptoms, but rather the ability to improve susceptibility to symptoms quicker, typical following repeated exposures to VEs.[80-51] Thus far, the characteristics of individuals with greater levels of plasticity have not been identified, and this will require further study.

4.4. Posture

Based on the postural instability theory, an individual's posture may also contribute to VIMS. Postural instability has a unique role in VIMS, as it can be a cause for and a result of VIMS. Studies have noted individuals with stable posture are less susceptible to VIMS, suggestive of an inverse relationship between postural stability and VIMS.[40, 44-47] Postural stability is a confluence of sensory inputs that are visual, vestibular and somatosensory in nature. Two neural reflexes involved in postural stability include the vestibular-ocular reflex (stabilizes objects on the retina) and the vestibular-spinal reflex (stabilizes posture while body in motion). When visual and vestibular inputs are not in synchrony, the result is postural instability, VIMS, or both.[46] Postural instability may last for several hours after exposure.[40, 44-47, 52, 54] Special considerations for HMD user safety, as related to the risk of postural instability, must be kept in mind. A recommendation that HMD users allow for a re-orientation/recovery time prior to engaging in potentially dangerous activities such as driving or sports may be in order.

5. Task Factors

Task characteristics have been also identified as potentially affecting VIMS. Among these tasks, duration of time in VEs is most notable. Longer exposure to VE increases the incidence of VIMS. These symptoms may persist up to 60 minutes after exposure.[15, 18, 19-21, 25, 27, 30, 31, 40, 51, 57, 58] Another important factor shown to influence VIMS is vection (i.e. an illusion of self-motion), with faster vection resulting in greater sickness symptoms.[15, 44, 55, 56]

Viewing HMD-based VR in a sitting position may reduce symptoms, as sitting reduces the demands on postural control.[22, 53, 47, 70] More complicated tasks, such as reading, may induce total symptom severity scores and oculomotor-related symptom scores that are significantly higher than those observed with movies or games (Figure 1).[30, 52, 36] These findings imply that more demanding tasks probably will create some degree of eyestrain. Increased reading sensitivity, when compared to watching a movie or playing a game, might be due to activation of different areas of the brain, which may make reading more complex than other tasks. Alternatively, reading can affect attention and blink rate, which may also contribute to an increase in VIMS. Moreover, inappropriate vertical gaze angle may cause increased oculomotor changes and visual discomfort. [23, 59]

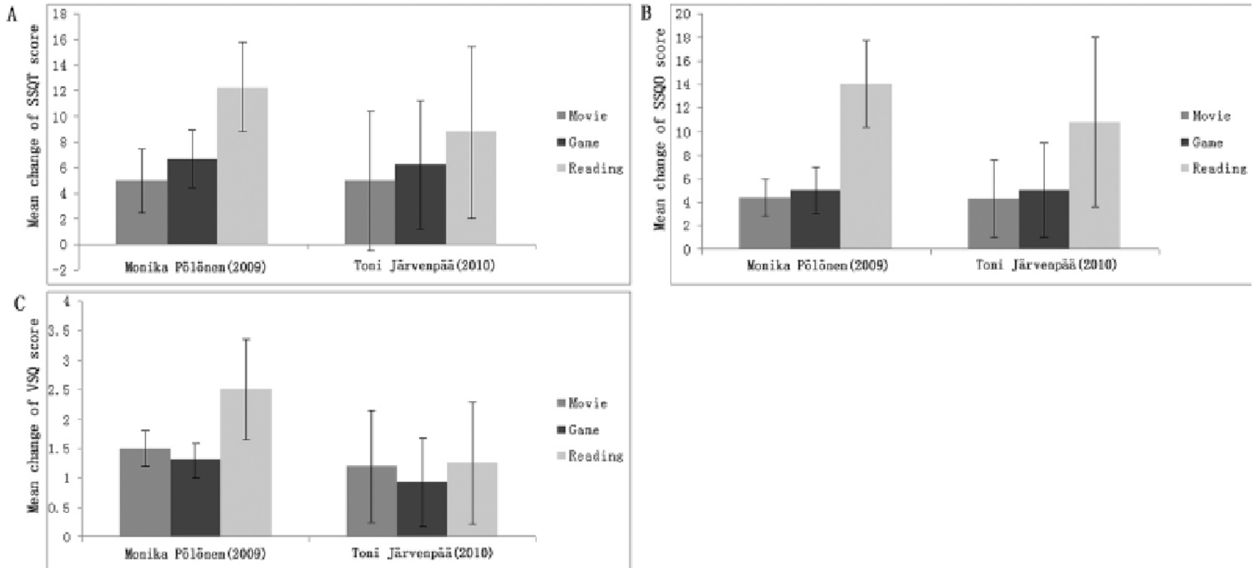


Figure 1 - Different VIMS Score Comparisons of Movie, Gaming, and Reading VR Content

6. Service Factors

Technical advances in hardware components will reduce physiologic ergonomic issues including HMD weight, system time delay, and luminance. However, given the multifactorial nature of VIMS, the conflict between visual and vestibular input remains a significant problem. Further reduction in VIMS needs to consider how content is created and how it influences services factors. Service factors need to consider the VIMS effect on the viewer.

6.1. Audience

Services can be created intended for a wide audience shared experience (e.g. Broadcast) or for narrow niche audience experience (e.g. Longtail content in VoD). For large scale audiences, further VIMS reduction will be highly dependent on content creation. Figure 1 indicates that the type of content can correlate to VIMS scores where content that requires reading over just watching (movies, gaming) can lead to increased VIMS. An increased FOV creates an immersive environment where watching a movie or playing a game would be more preferred rather than reading. User factors could also mitigate VIMS by reducing visual cues that create sensory input conflicts. This could be done by making the viewer more of a detached fixed observer (i.e. reduced vection) and making the content more of scenery type of content which has already been made popular in viewing UHD/HDR material.

6.2. Transmission: Linear Broadcast, On-Demand, Download

Service factors for VR should be cognizant of the bits delivered to HMDs. The flow of bits to HMDs may impact time lag or latency.[71] Potential ways to create a service around this dimension is to permit progressive download, on-demand streaming, or real-time linear streaming deliveries. This progression tracks like the evolution of Video streaming where progressive download was initially done until codec compression efficiencies and bandwidth expansion advanced enough to allow for on-line streaming and ultimately real-time encoding/streaming to occur.

At first glance this would require another generation of video coding to occur beyond HEVC and may become a driving use case for the work being done in JVET (Joint Video Exploration Team).[75] [76] Region Centric approaches such as Splice based or Tile based encoding along with adaptive streaming delivery can reduce the amount of bits needed to be transmitted, but these approaches need to make use of head-tracking information on HMDs in order to avoid VIMS effects.[77] Alternate approaches to get beyond the transmission bit rate restrictions is to move away from video frame based encoding to more computervision-based models. In this case creating high resolution 360 degree objects with complex textures, but low bit rate movement information can allow for tolerable transmission rates. To better mesh with true-life environments, a framework can be created that mimics actual scene but made up of computervision object models (e.g. avatars). Some of this work is being investigated in the MPEG group. [72] [73]

A quickly deployable VR solution for a shared watching experience would be to limit the choices of alternate viewpoints. This would also reduce the amount of network bandwidth needed to support a shared VR experience while allowing for the viewer to have an ability to reduce VIMS to a tolerable level. With limited alternate viewpoints, more of the data can be shared in a broadcast manner. An example of this would be a sporting event with alternate camera angles (one as part of the audience, one with ringside seats, one as the boxer).[74] In these cases, the level of visual discomfort is under the control of the viewer who can switch to a more tolerable camera view. At the same time, the VR content creators can overall limit the maximum VIMS impact of an alternate camera angle by limiting the viewing duration of it. The network transmission burden in this case is spread out among several viewers and is conducive to live event broadcasts while still accommodating different levels of VIMS tolerances.

6.1. Interactivity

VR environments hinge on some level of interactivity between the viewer and the observed environment. The accepted amount of interactivity can be designed into the service and should consider a reduction of VIMS effects.

A key component is reducing time-lag or latency between a decision of the viewer and the reaction of the environment. This need is similar to video conferencing services where the round trip restrictions of 100-150 ms were needed to avoid miscues in the service. VR services would need a much more restrictive limit approaching 50-20 ms to avoid visual miscues in the service.[71] Video Conferencing methods to deal with this were to maintain low-latency real-time video conferencing often at the sacrifice of framerate and resolution. Since VR uses constant framerate, the latency is restricted by the time between two frames (e.g. 30 fps indicates about 33 ms between frames, 90 fps indicates about 11ms between frames). Increased framerate gives the potential of reducing VIMS since the next frame is the quickest interactivity possible. Applying video conferencing latency approaches in VR services would actually increase VIMS effects since video conferencing round trip latency is not even approachable to framerate which provides a gating factor to timelag with for example 90 fps being around a 11ms factor. [71]

Alternate approaches would need to consider if the interactivity would be on-network or off-network. If on-network, the transmission constraints would currently prohibit full interaction and be more amenable to limited anticipatory interactions. An example of this may be switching video feeds by reutilizing alternate camera angles as previously described in the previous section. For off-network cases, the viewer does not react with the transmitted video feed, but rather with the background environment. For instance, a VR “skin” could be developed around a video feed(s) where the reaction is limited to the constructed scene around the feed(s).[74] An example of this would be one or multiple video screens in a simulated environment where local controls for all the video screens is available in a virtually integrated environment with only a small amount of latency. With further refinements of head tracking technology, latency could be further reduced with the application of gaze control where video and audio feed could be activated simply with knowledge of which video screen one is looking at.[74] These types of off-network interactions could be done with minimal VIMS effects.

6.2. Equipment

Services can reduce VIMS effects by creating and making aware to users and the content community a set of guidelines of optimized visual ergonomics for HMD use. These guidelines should also be used as a way to help in service definition and may include the following recommendations:

1. Hardware manufacturers would benefit from standardization of systems/component features;
2. Users should be advised children, women, users with visual field defects, postural instability, or past history of motion sickness are more susceptible to VIMS;
3. Guidelines for HMD usage, as inexperienced users are especially susceptible to developing VIMS and users are different in their adaptation to HMDs;
4. Users should be warned to not use HMDs for extended periods of time, to take frequent breaks, and to avoid driving and operating heavy machinery until VIMS and postural instability resolve;
5. Guidelines for content development designers to minimize vection, eyestrain and VIMS

Beyond this, services still need to carefully consider equipment decisions to make sure the service definition is viable. Should the equipment be leased or are certified COAMed (Customer owned and managed) equipment allowed? Should all VR services use the same type of equipment or can advanced VR services merit more advanced equipment. Lastly how does one evolve VR services for the future?

VR services that are on-demand can help evolve service definition. It provides a way to test out new VR types of features to an audience that may have more “plasticity” than the regular general audience. It also provides a way to test out more advance VR equipment at the same time. The information collected from this service can then be used to help evolve VR services that are part of a shared wide audience experience (e.g. live events).

7. Conclusions

VIMS and visual discomfort continue to be obstacles for widespread acceptance of HMDs; this increases the importance of further research into VIMS. More research is needed to resolve visual-vestibular mismatch, and to develop objective methods of evaluating and quantifying VIMS symptoms such as visual/ocular changes (e.g. ocular movements), physiological changes (e.g. changes in heart rate, blink rate, EEG [electro-encephalography]), and vestibular changes (e.g. perceived spatial velocity).

This issue does not preclude the establishment of VR services, but does limit the definition of it. Still it is possible to create a VR service that addresses a wide audience shared experience though transmission and interactivity restrictions that will limit the type of VR experience as explained in the previous section.

This issue does not preclude the establishment of VR services, but does limit the definition of it. Still it is possible to create a VR service that addresses a wide audience shared experience though transmission and interactivity restrictions that will limit the type of VR experience as explained in the previous section. Creating an on-demand VR service at the same time can address a narrower audience experience and provide more of a VR experience. At the same time, this also provides a pathway to help evolve all VR services and reduce VIMS effects.

8. Abbreviations and Definitions

8.1. Abbreviations

HMD	head-mounted display
VE	virtual environment
VR	virtual reality
IOD	inter-optical distance
IPD	inter-pupillary distance
FOV	field of view
VIMS	visually-induced motion sickness
3D	three dimensional

8.1. Definitions

Accommodation	The process of which the eye changes optical power to maintain a focus on an object
Amblyopia	maldevelopment of the visual system occurring in early childhood, resulting in poor visual function in the affected eye
Convergence	Simultaneous inward movement of both eyes
Fixation Disparity	The difference between the target vergence angle and the ocular convergence angle
Heterophoria	Mis-aligned eyes at rest position
Oculomotor	the third cranial nerve responsible for most eye movements and raising the eyelid
Plasticity	the changeability of the brain throughout an individual's life
Somatosensory	relating to or denoting a sensation (such as pressure, pain, or warmth) that can occur anywhere in the body
Strabismus	misalignment of the eyes, most notable when fixating on an object, resulting in disruption of binocular vision
Vection	following stimulation of a large part of the visual field, a viewer feels to have moved, though he/she is stationary
Visual-Vestibular Mismatch	Sensory conflict between the inner ear and eye

9. Bibliography and References

1. Sutherland IE. The Ultimate Display. In: Proceedings of International Federation of Information Processing (IFIP) Congress; 1965:506-508.
2. Kennedy RS, Hettinger LJ, Lilienthal MG. Simulator Sickness. In: Crampton GH, editor. Motion and Space Sickness. Boca Raton, FL: CRC Press, 1990: 317–341.
3. Kennedy RS, Drexler J, Kennedy RC. Research in visually induced motion sickness. *Appl Ergon* 2010;41(4):494-503.
4. Ukai K, Kibe A. Counterroll torsional eye movement in users of head-mounted displays. *Displays* 2003;24(2):59–63.
5. Bles W, Wertheim AH. Appropriate use of virtual environments to minimise motion sickness. In: Werkhoven P, Breaux R, eds. What is essential for virtual reality systems to meet military human performance goals? Soesterberg, Netherlands): Human Factors Research Institute, 2001.
6. Peli E. Visual issues in the use of a head-mounted monocular display. *Opt Eng* 1990;29(8):883-892.
7. Kooi FL, Toet A. Visual comfort of binocular and 3D displays. *Displays* 2004;25(2-3):99-108.

8. Ukai K, Howarth PA. Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations. *Displays* 2008;29(2):106-116.
9. Menozzi M. Visual ergonomics of head-mounted displays. *Japanese Psychological Research* 2000;42(4):213-221.
10. Urvoy M, Barkowsky M, Le Callet P. How visual fatigue and discomfort impact 3D-TV quality of experience: a comprehensive review of technological, psychophysical, and psychological factors. *Ann Telecommun* 2013;68(11):641–655.
11. Howarth PA. Potential hazards of viewing 3-D stereoscopic
12. television, cinema and computer games: a review. *Ophthalmic and Physiological Optics* 2011;31(2):111-122.
13. Bando T, Iijima A, S. Y. Visual fatigue caused by stereoscopic images and the search for the requirement to prevent them: A review. *Displays* 2012;33(2):76-83.
14. Mon-Williams M, Wann JP, Rushton S. Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display. *Ophthalmic and Physiological Optics* 1993;13(4):387-391.
15. Lampton DR, Rodriguez ME, Cotton JE. Simulator sickness symptoms during team training in immersive virtual environments. In: *Human Factors and Ergonomics Society (HFES) Annual Meeting Proc* 2000;44(5):530-533.
16. Kuze J, Ukai K. Subjective evaluation of visual fatigue caused by motion images. *Displays* 2008;29(2):159–166.
17. Howarth PA, Costello PJ. The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system. *Displays* 1997;18(2):107-116.
18. Howarth PA. Oculomotor changes within virtual environments. *Appl Ergon* 1999;30(1):59-67.
19. Kozulin P, Ames SL, McBrien NA. Effects of a head-mounted display on the oculomotor system of children. *Optom Vis Sci* 2009;86(7):845-856.
20. Rushton S, Mon-Williams M, Wann JP. Binocular vision in a bi-ocular world: new-generation head-mounted displays avoid causing visual deficit. *Displays* 1994;15(4):255-260.
21. Ames SL, Wolffsohn JS, McBrien NA. The development of a symptom questionnaire for assessing virtual reality viewing using a head-mounted display. *Optom Vis Sci* 2005;82(3):168-176.
22. Aaltonen V, Pölönen M. The effect of viewing duration on visual comfort with near-to-eye displays. *Society for Information Display (SID) Symposium Digest of Technical Papers* 2009;40(1):812-814.
23. Costello PJ, Howarth PA. The effects on the visual system of wearing a Head-Mounted Display for an hour. *VISERG Report* 9605;1996.
24. Mon-Williams M, Plooy A, Burgess-Limerick R, Wann J. Gaze angle: a possible mechanism of visual stress in virtual reality headsets. *Ergonomics* 1998;41(3):280-285.
25. Rushton SK, Riddell PM. Developing visual systems and exposure to virtual reality and stereo displays: some concerns and speculations about the demands on accommodation and vergence. *Appl Ergon* 1999;30(1):69-78.
26. Stanney KM, Kennedy RS. Aftereffects from virtual environment exposure: how long do they last? In: *HFES Annual Meeting Proc* 1998;42(21):1476-1480.
27. Stanney KM, Kennedy RS, Drexler JM, Harm DL. Motion sickness and proprioceptive aftereffects following virtual environment exposure. *Appl Ergon* 1999;30(1):27-38.
28. Moss JD, Muth ER. Characteristics of head-mounted displays and their effects on simulator sickness. *Hum Factors* 2011;53(3):308-319.
29. Pölönen M, Järvenpää T, Häkkinen J. Comparison of Near-to-Eye Displays: Subjective Experience and Comfort. *J Disp Technol* 2010;6(1):27-35.
30. Järvenpää T, Pölönen M. Advances in Near-to-Eye Display Optical Characterization. *Society for Information Display (SID) International Symposium Digest of Technical Papers* 2009;40(1):507–510.
31. Järvenpää T, Pölönen M. Optical characterization and ergonomical factors of near-to-eye displays. *J Soc Inf Disp* 2010;18(4):285-292.
32. Jaeger BK, Mourant RR. Comparison of simulator sickness using static and dynamic walking simulators. *HFES Annual Meeting Proc* 2001; 45(27):1896-1900.
33. Kolasinski EM, Gilson RD. Simulator sickness and related findings in a virtual environment. *HFES Annual Meeting Proc* 1998;42(21):1511–1515.
34. Ehrlich JA, Kolasinski EM. A comparison of sickness symptoms between dropout and finishing participants in virtual environment studies. *HFES Annual Meeting Proc* 1998;42(21):1466-1470.
35. Draper MH, Viire ES, Furness TA, Gawron VJ. Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Hum Factors* 2001;43(1):129-146.

36. Pölönen M, Häkkinen J. Near-to-eye display—An accessory for handheld multimedia devices: Subjective studies. *J DisplayTechnol* 2009;5(9):358-367.
37. Nelson WT, Roe MM, Bolia RS, Morley RM. Assessing Simulator Sickness in a See-Through HMD: Effects of Time Delay, Time on Task, and Task Complexity. Air Force Research Laboratory Wright-Patterson AFB, OH;2000.
38. St. Pierre ME, Banerjee S, Hoover AW, Muth ER. The effects of 0.2 Hz varying latency with 20–100 ms varying amplitude on simulator sickness in a helmet mounted display. *Displays* 2015;36:1-8.
39. Aykent B, Zhao Y, Merienne F, Kemeny A. Simulation sickness comparison between a limited field of view virtual reality head mounted display (Oculus) and a medium range field of view static ecological driving simulator (Eco2). *Proceedings of the Driving Simulation Conference Europe* 2014:31-31 - 31-37.
40. Moss J, Scisco JL, Muth E. Simulator Sickness during Head Mounted Display (HMD) of Real World Video Captured Scenes. *HFES Annual Meeting Proc* 2008;52(19):1631-1634.
41. Häkkinen J, Vuori T, Puhakka M. Postural Stability and Sickness Symptoms after HMD Use. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* 2002;1:147-152.
42. Kim K, Rosenthal MZ, Zielinski D, Brady R. Comparison of Desktop, Head Mounted Display, and Six Wall Fully Immersive Systems using a Stressful Task. *IEEE Virtual Reality (VRW)* 2012:143-144.
43. Pölönen M, Järvenpää T, Häkkinen J. Reading e-books on a near-to-eye display: Comparison between a small-sized multimedia display and a hard copy. *Displays* 2012;33(3):157–167.
44. Vlad R, Nahorna O, Ladret P, Gu´erin A. The influence of the visualization task on the simulator sickness symptoms - a comparative SSQ study on 3DTV and 3D immersive glasses. *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)* 2013:1-4.
45. Smart LJ, Jr., Stoffregen TA, Bardy BG. Visually induced motion sickness predicted by postural instability. *Hum Factors* 2002;44(3):451-465.
46. Owen N, Leadbetter AG, Yardley L. Relationship between postural control and motion sickness in healthy subjects. *Brain Res Bull* 1998;47(5):471-474.
47. Cobb SV. Measurement of postural stability before and after immersion in a virtual environment. *Appl Ergon* 1999;30(1):47-57.
48. Stoffregen TA, Faugloire E, Yoshida K, Flanagan MB, Merhi O. Motion sickness and postural sway in console video games. *Hum Factors* 2008;50(2):322-331.
49. Regan EC. Some evidence of adaptation to immersion in virtual reality. *Displays* 1995;16(3):135-139.
50. Hill KJ, Howarth PA. Habituation to the side effects of immersion in a virtual environment. *Displays* 2000;21(1):25–30.
51. Howarth PA, Hodder SG. Characteristics of habituation to motion in a virtual environment. *Displays* 2008;29(2):117-123.
52. Kennedy RS, Stanney KM, Dunlap WP. Duration and exposure to virtual environments: Sickness curves during and across sessions. *Presence: Teleoperators and Virtual Environments* 2000;9(5):463-472.
53. Takada H, Matsuura Y, Takada M, Miyao M. Comparison in Degree of the Motion Sickness Induced by a 3-D Movie on an LCD and an HMD. In: Shumaker R, editor. Virtual and Mixed Reality - New Trends. Berlin: Springer-Verlag, 2011:371–379.
54. Ehrlich JA, Singer MJ, Allen RC. Relationships between head-shoulder divergences and sickness in a virtual environment. *HFES Annual Meeting Proc* 1998;42(21):1471-1475.
55. Kutsuna K, Matsuura Y, Fujikake K, Miyao M, Takada H. Comparison of form in potential functions while maintaining upright posture during exposure to stereoscopic video clips. *Conference proceedings, Engineering in Medicine and Biology Society (EMBC) 35th Annual International Conference of the IEEE* 2013:7200-7203.
56. Jang DP, Kim IY, Nam SW, Wiederhold BK, Wiederhold MD, Kim SI. Analysis of physiological response to two virtual environments: driving and flying simulation. *Cyberpsychol Behav* 2002;5(1):11-18.
57. So RHY, Lo WT. Cybersickness with Virtual Reality Training Applications: a claustrophobia phenomenon with headmounted displays? *Proceedings of the First World Congress on Ergonomics for Global Quality and*
58. Lampton DR, Kolasinski EM, Knerr BW, Bliss JP, Bailey JH, Witmer BG. Side effects and aftereffects of immersion in virtual environments. *HFES Annual Meeting Proc* 1994;38(18):1154-1157.

59. Steinicke F, Bruder G. A Self-Experimentation Report about Long-Term Use of Fully-Immersive Technology. *Proceedings of the Association for Computing Machinery (ACM) Symposium on Spatial User Interaction (SUI)* 2014: 66-69.
60. Pölönen M, Aaltonen V, Nyman G, Häkkinen J. Influence of the monocular near-to-eye display position on the visual system during dual-task performance. *Journal of the Society for Information Display* 2008;16(1
61. Patterson R, Winterbottom MD, Pierce BJ. Perceptual issues in the use of head-mounted visual displays. *Hum Factors* 2006;48(3):555-573.
62. Melzer JE, Brozoski FT, Letowski TR, Harding TH, Rash CE. Guidelines for HMD design In: Rash CE, editor. Helmet-Mounted Displays: Sensation, Perception and Cognition Issues. Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory, 2009: 805–848.
63. Moffitt K. Designing HMDs for viewing comfort. In: Melzer JE, Moffitt K, editors. Head-Mounted Displays: Designing for the User. North Charleston, S.C.: Melzer JE, 2011: 117-145.
64. Cakmakci O, Rolland J. Head-Worn Displays: A Review. *Journal of Display Technology* 2006(3);2:199-216.
65. Task HL. HMD image source, optics, and the visual interface. In: Melzer JE, Moffitt K, editors. Head-Mounted Displays: Designing for the User. North Charleston, S.C.: Melzer JE, 2011: 55-82.
66. Turner M, Love S. Individual differences in user experience of head mounted displays: personality and visuo-spatial attention effects. In: de Waard D, Brookhuis KA, Sommer SM, Verwey WB, editors. Human Factors in the Age of Virtual Reality. Maastricht, the Netherlands: Shaker Publishing, 2003:31-42.
67. Ciuffreda KJ, Levi DM, Selenow A. Amblyopia: Basic and Clinical Aspects. Boston: Butterworth-Heinemann, 1991.
68. Reason JT, Brand JJ. Motion Sickness. London, England: Academic Press, 1975.
69. Kennedy RS, Lanham SD, Massey CJ, Drexler JM. Gender differences in simulator sickness incidence: implications for military virtual reality systems. *Safe Journal* 1995;25(1):69-76.
70. La Viola JJ Jr. A Discussion of Cybersickness in Virtual Environments. *SIGCHI Bulletin* 2000;32(1):47-56.
71. Regan EC, Price KR. Some side-effects of immersion virtual reality: The effects of increasing head movements, of rapid interaction, and of seating subjects. Report 93R022. Farnborough, UK: Army Personnel Research Establishment, 1993.
72. Raaen K, Kjellmo I. Measuring latency in Virtual Reality Systems. ICEC 2015, pp.457-462.
73. Technical report of the joint ad hoc group for digital representations of light/sound fields for immersive media applications, ISO/IEC JTC1/SC29/WG1N16352, Geneva Switzerland, June 2016.
74. Tulvan C, Mekuria R, Li Z. Use Cases for Point Cloud Compression, ISO/IEC JTC1/SC29/WG11 MPEG 2016/N16331, Geneva Switzerland, June 2016.
75. Gower A, Content Provider Trials- VR Production Insights. Presentations from DVB VR Workshop, ISO/IEC JTC1/SC29/WG11 MPEG 2016/N38818, Geneva Switzerland, May 28 2016
76. HEVC, JCT-VC, http://phenix.int-evry.fr/jct/
77. JVET, http://phenix.int-evry.fr/jvet/
78. Wang Y, Hendry, Karczewicz M. Tile based VR video encoding and decoding schemes, ISO/IEC JTC1/SC29/WG11 MPEG 2016 & ITU-T SG16 JCTVC-X0077, Geneva Switzerland, June 2 2016
79. Karpicka E, unpublished data 2016

JOURNAL OF
DIGITAL VIDEO

SCTE ISBE

Ultra-High Definition for Cable Applications

A Technical Paper prepared for SCTE/ISBE by

Thierry Fautier, Harmonic Inc., SCTE/ISBE Member
San Jose, California
Thierry.Fautier@harmonicinc.com
408.542.2726

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

List of Tables

1. Introduction

Cable operators will soon have a standard that defines how they may acquire ultra-high definition (UHD) signals from broadcasters. Depending upon where they are based geographically, operators might encounter digital video broadcast (DVB) or ATSC 3.0 standards for the specification of signal reception. Final distribution will be defined by SCTE.

The ATSC 3.0 suite of standards that defines UHD delivery to UHD TV sets is still being defined, including, in particular, the interface between broadcasters and multi system operators (MSOs). DVB is specifying a final distribution system that can be applied to cable distribution. This paper will focus on the DVB approach, as it is today the only distribution system being specified for a cable network until SCTE defines one.

The DVB UHD-1 Phase 2 standard is still being defined, and an introduction to the standard can be found in Table 1. It will consist of several different elements, as the table indicates.

Table 1 - DVB UHD-1 Phase 2 at a Glance

Part	Details	Reference	status
System	Transport stream	Update from [0]	CRs closed, technical work due to be complete October 2016 (limited changes expected)
	DVB DASH	Update from [1]	Commercial module closed
Video	Video codec with HDR	Update from [0]	CRs closed, technical work due to be complete October 2016
Audio	NGA codec	Update from [0]	CRs closed, technical work due to be complete October 2016
Subtitling	Subtitling carriage with HDR	Update from [2]	CR for text and bitmap closed, technical work due to be complete October 2016

This paper will present the different video options that will be used for broadcast or unicast over HTTP delivery. The discussions around high dynamic range (HDR) are speculative, but they aim to provide insight into the HDR decision-making processes within DVB. Audio aspects will not be included in the discussion, as they imply different techniques. The focus will be on video signals, with p60 as the maximum frame rate, as some DVB countries are using 60 Hz and thus imposing constraints on processing power.

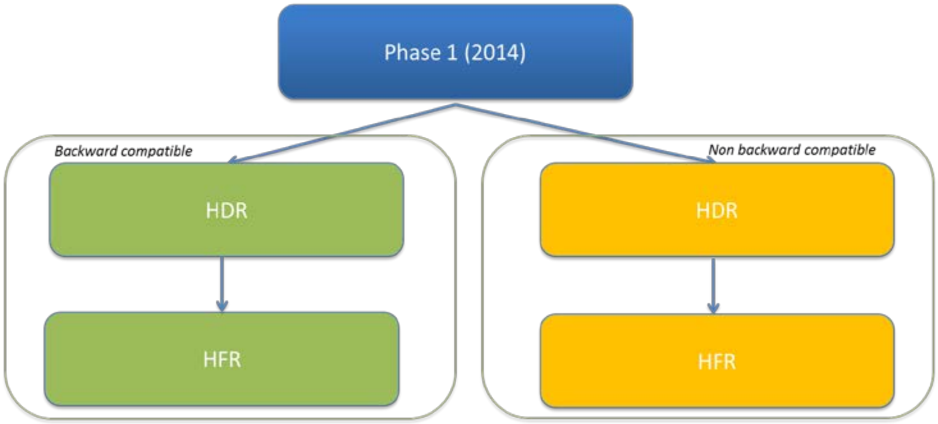


Figure 1 - Backward Compatibility Schemes

2. DVB Backward Compatibility

DVB has defined a backward compatibility scheme against the Phase 1 IRD compliance requirement with UHD-1 Phase 1-bit stream specification [0] and [1]. A Phase 2 decoder will have to be backward compatible to enable delivery of video with HDR and HFR. Possible combinations are described in Figure 1.

By definition, a backward compatible scheme must support HDR and HFR when decoded by a Phase 1 decoder. A non-backward compatible scheme will not impose any backward compatibility constraints on Phase 1 decoders. Phase1 key parameters are summarized in Table 2.

Table 2 - DVB UHD-1 Phase1

Item	Details	Note
Deployment	2014 - 2015	
Codec	HEVC Main 10	
Maximum resolution	3840 x 2160 x 60	
Maximum MPEG level	5.1	
WCG	BT 709/2020	Decoder needed to ingest BT 2020
HDR	No	
HFR	No	
Backward compatibility	N/A	
NGA	No	5.1 HD codecs

3. Market Backward Compatibility

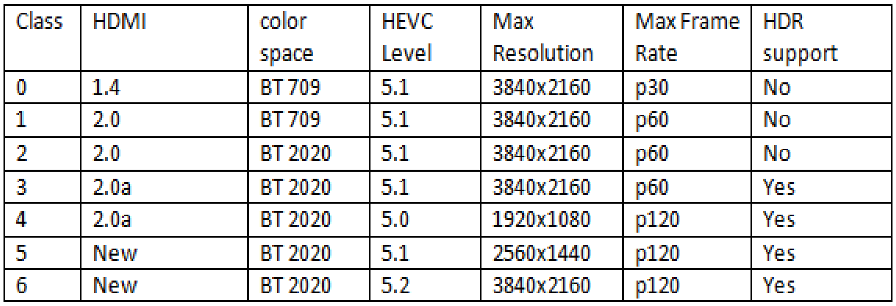
This section will discuss backward compatibility not only from a stream perspective but also from a display angle (i.e., how a Phase 1 display will need to be addressed by a Phase 2 STB). Because the good majority of TV sets sold in the coming years will be SDR, HDR backward compatibility with SDR sets is a requirement. The following section will look at the impact of WCG, HDR and HFR on the backward compatibility of a DVB UHD-1 Phase 2 system.

4. UHD TV types

It is important to understand what types of UHD displays have been deployed so far and what displays will come in the future. Seven classes of UHD TV types have been identified based on the criteria described in Table 3.

Class 0 represents the first generation of UHD TV sets sold. As it can only support p30, it cannot be used for a DVB UHD system, unless a spatial scalable scheme is implemented. This scenario is not highly probable because it would sacrifice QoE.

Table 3 - UHD TV Classes



Class	HDMI	color space	HEVC Level	Max Resolution	Max Frame Rate	HDR support
0	1.4	BT 709	5.1	3840x2160	p30	No
1	2.0	BT 709	5.1	3840x2160	p60	No
2	2.0	BT 2020	5.1	3840x2160	p60	No
3	2.0a	BT 2020	5.1	3840x2160	p60	Yes
4	2.0a	BT 2020	5.0	1920x1080	p120	Yes
5	New	BT 2020	5.1	2560x1440	p120	Yes
6	New	BT 2020	5.2	3840x2160	p120	Yes

Class 1 is the second generation of UHD TV sets deployed with full-frame-rate support and BT 709 color space support, but no HDR. This class imposes the most constraints on backward compatibility. To address this type of set, the industry needs an HDR scheme that can create a BT 709 signal. The HDR section, below, will discuss the possible HDR options.

Class 2 is the third generation of UHD TV sets deployed with full-frame-rate support and BT 2020 color space support, but no HDR. This class imposes fewer constraints on backward compatibility, as the color space is wider than for Class 2. To address this type of set, the industry needs an HDR scheme that can create a BT 2020 signal.

Class 3 is the first generation of UHD TV sets that supports HDR. The frame rate is still limited to p60.

Class 4 is the first generation of UHD TV sets that supports HDR and HFR in p120 but with an HD resolution (1920 x 1080). It is important to note that the signal can be carried in HDMI 2.0a.

Class 5 is the first generation of UHD TV sets that supports HDR and HFR in p120 but with improved resolution compared with Class 4 (4 x 720p resolution of 2560 x 1440). While the signal needs a new HDMI specification, the pixel rate is within the boundaries of MPEG HEVC Main 10 Level 5.1, meaning that a Main 10 decoder, in principle, has the capability to decode this HFR profile.

Class 6 is the ultimate goal of 3840 x 2160 x 120 with HDR. As this requires a new decoder and encoder infrastructure (two times the pixel rate versus Class 3), it is expected to come at a later stage when technology is available at a reasonable cost.

5. WCG and HDR

MPEG, SMPTE, ITU-R, CTA, DVB and ATSC are all working on standardizing HDR for broadcast applications. Because no standard yet has been defined internationally for broadcast, this paper will list all the documented options, referring to the SMPTE HDR paper [4] for details. Table 4 provides a list of all the different HDR options to be discussed further.

Table 4 - HDR Options

HDR	Codec	Codec layers	EOTF	Metadata	Backward compatibility	Standard reference
HDR 10	HEVC Main 10	Single	ST 2084	ST 2086	Decoder based (1)	SMPTE CTA
Dolby Vision	HEVC Main 10	Dual	ST 2084	ST 2086/ETSI	Stream based	BDA ETSI
		Single	ST 2084	ST 2086/Prop	Decoder based (2)	
Technicolor /Philips	HEVC Main 10	Single	ST 2084	Prop	Stream based and decoder based (2)	BDA
HLG	HEVC Main 10	Single	Hybrid log Gamma	No	Native	ITU-R ARIB
SHVC	SHVC	Dual	ST 2084	Being defined	Decoder based (1)	ETSI

(1) Decoder based without metadata, does not guarantee the quality of the restitution of SDR

(2) *Decoder based with metadata, does guarantee the quality of the restitution of SDR*

6. Cable network architectures for UHD

A UHD signal can be distributed in three different ways via a cable network:

- The first option involves using classic quadrature amplitude modulation (QAM) delivery. This is not very efficient, especially when the number of subscribers is low from a network efficiency point of view.
- An alternative is using IP multicast (single rate) delivery over DOCSIS 3.x. The drawback of this option is that the signal cannot be carried to connected TVs, which expect unicast adaptive bit rate protocol.
- Using a unicast adaptive bit rate protocol, such as DASH or HLS over DOCSIS 3.x, is another approach. The drawback of this this scheme is that as soon as the number of subscribers grows, more caching is required, although CableLabs has defined a standard for ABR multicast delivery

[5]. Note that it is possible to have a single bitrate transmission if the QoS is guaranteed by the MSO.

Table 5 outlines the different possible combinations for UHD cable delivery.

Table 5 - Network Delivery Options

Delivery method	Protocol	Note
QAM	TS	Deployed
IP multicast (legacy)	TS	Needs DOCSIS 3.x or FTTH
IP multicast per [5]	DASH	Needs DOCSIS 3.x
HTTP	DASH or HLS	Needs DOCSIS 3.x or FTTH

Moving forward, "cable network" will refer to a combination of the three listed network architectures: QAM, IP multicast and HTTP.

7. HDR options

7.1. HDR 10

Figure 2 shows the HDR solution that has been specified by BDA [3] for the encoding format, as well as by CTA [4] for HDMI transmission. Static metadata is passed via ST 2086, while ST 2094 will define dynamic metadata. As this is the only published HDR standard today, all chips (i.e., decoders and TVs) support HDR 10. HDR 10 was designed as a non-backward compatible scheme at the beginning, but there are implementations in Blu-ray players that will be able to tone map it to an SDR BT 709 color space on the receiver side. Some TV manufacturers and STB chipsets demonstrated similar capabilities at CES 2016. The first broadcast demonstration of HDR 10 was made by Sky Germany, which conducted live trials of HDR 10 in August 2015 [6]. At IBC 2015, Harmonic showed a joint HDR 10 demonstration of a live end-to-end workflow with the Ultra HD Forum.

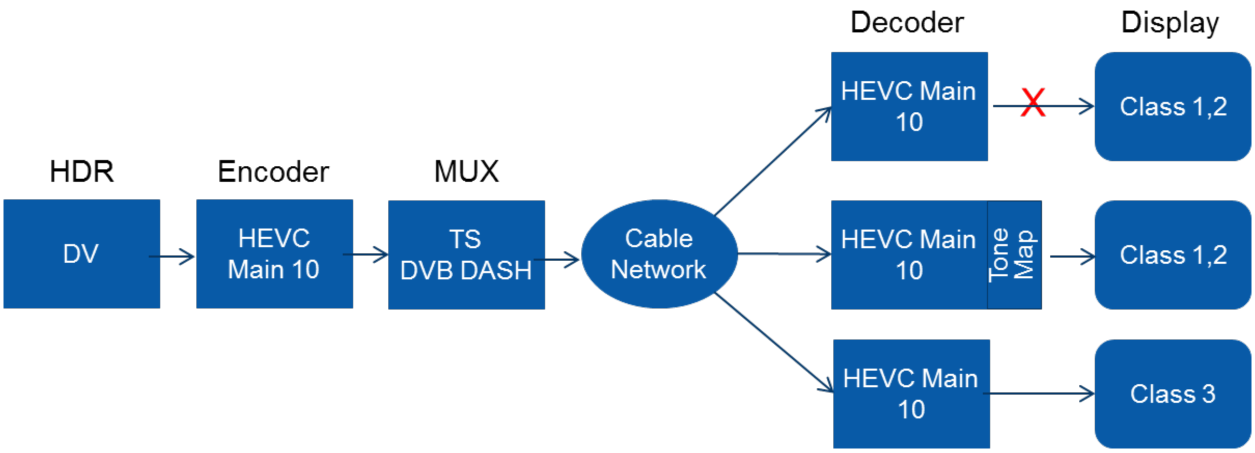


Figure 2 - HDR 10

7.2. Dolby Vision dual layer

Dolby Vision dual layer is solution that will transmit BT 709 to Class 1 TVs and can also address Class 2 TVs with BT 2020. Through an enhancement layer, it can provide an HDR experience on a Dolby Vision consumer playback solution that will display to a Class 3 and above TV. Note that in the BDA specification, the base layer is not BT 709 but HDR 10, so a different specification is required for broadcast application. In that respect, Dolby has started a standardization effort with ETSI to specify a broadcast dual-layer scheme. Metadata is defined in SMPTE ST2086, ST2094 and ETSI ISG CCM. Dolby Vision has already been endorsed by several silicon TV decoder chipsets, including Mediatek, MStar, Sigma, Realtek, and HiSilicon.

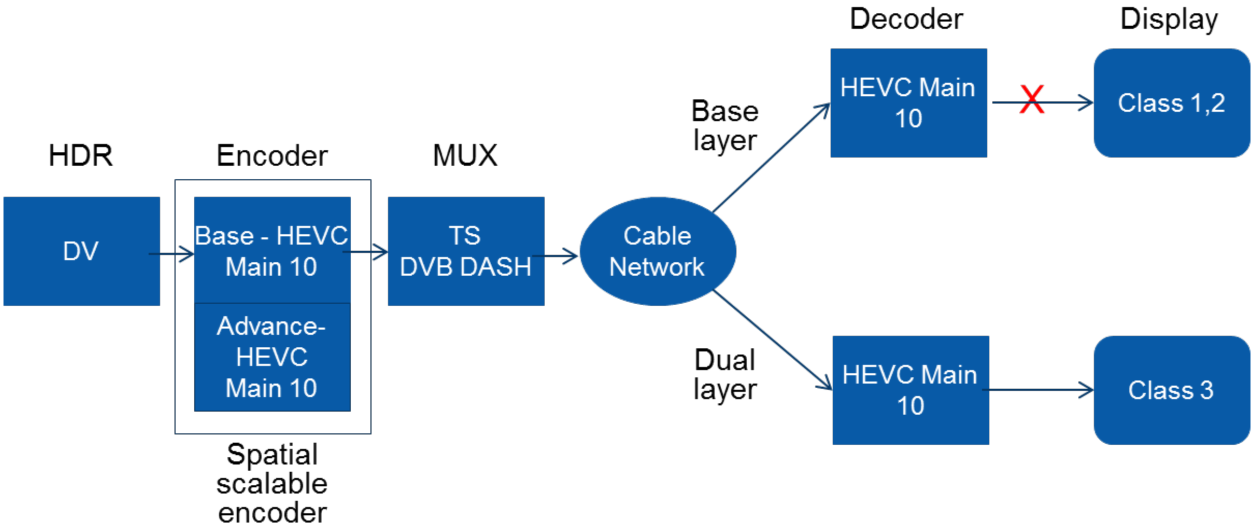


Figure 3 - Dolby Vision Dual Layer

7.3. Dolby Vision single layer

Dolby Vision single layer is proposed for standardization in DVB. It is a non-backward compatible solution that will only address BT 2020 HDR sets of Class 3 TVs and above. The Dolby Vision consumer playback solution can accept a BT 2020 HDR signal and map it to a BT 709 color space. Dolby demonstrated a preview of this solution with Harmonic at IBC2015 [7].

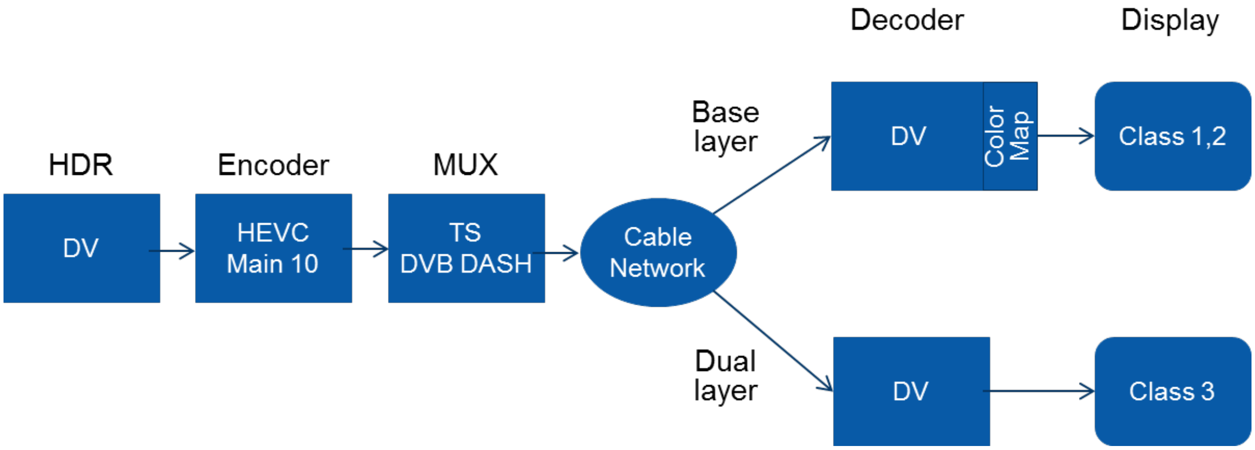


Figure 4 - Dolby Vision Single Layer

7.4. Technicolor/Philips

Technicolor and Philips announced at CES 2016 they would merge their solutions. A single-layer codec solution that is natively compatible with BT 709 UHD decoders is expected. The additional metadata will be used by the decoder to reconstruct an HDR signal. The main characteristic of this solution is that the base layer is compatible with legacy UHD decoders (e.g., DVB Phase 1 decoders) and therefore Class 1 and 2 displays. The drawback is that a new decoder is required to decode the metadata needed to reconstruct the HDR signal. At CES 2016, STM, Marvell, MStar and Sigma Designs announced their next-generation silicon would support Technicolor.

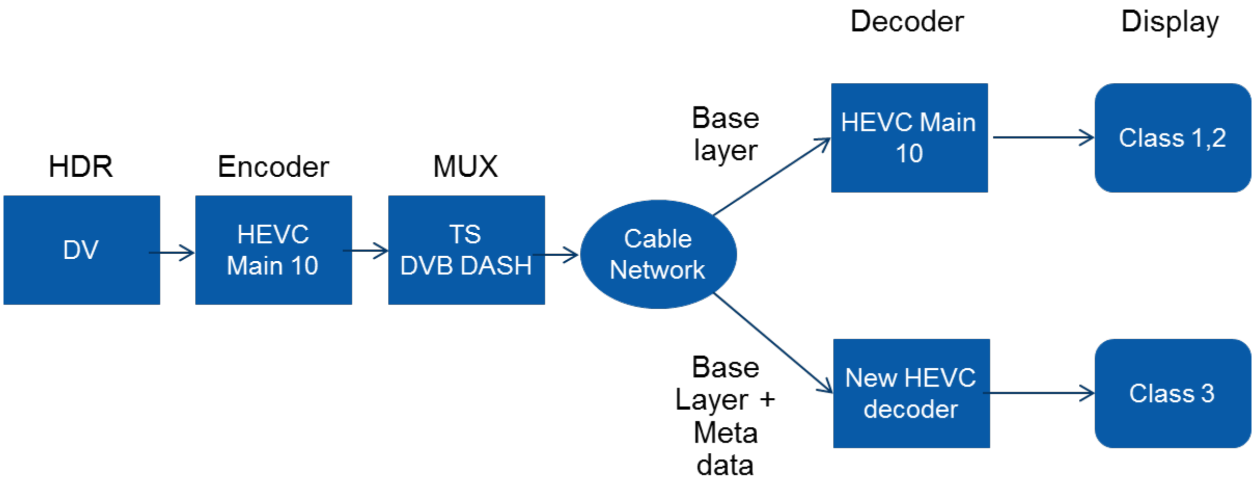


Figure 5 - Technicolor/Philips Merged Solution

7.5. HLG (Hybrid Log Gamma)

HLG is a solution jointly promoted by BBC and NHK. They have agreed on the OETF (Optical Electrical Transfer Function) within the ITU-R [8], now the BT 2100 international standard. The main feature of HLG is that the same stream goes to legacy TVs (Class 1 and 2) and to HDR TVs (Class 3 and above). A key advantage is that HLG can work on existing infrastructure via a software upgrade of the decoder, the HDMI link and the TV set. There is no certainty that HLG will be standardized for Class 1 TVs, as tests are still underway at the time of writing of this paper. Live HLG demonstrations have already been done by SES [9] in May 2016, Sky Germany in August 2015 [6] and Sky PerfecTV [10] in November 2015. As HLG is now being standardized by ITU-R, it is anticipated that there will be more traction for this format for broadcast applications.

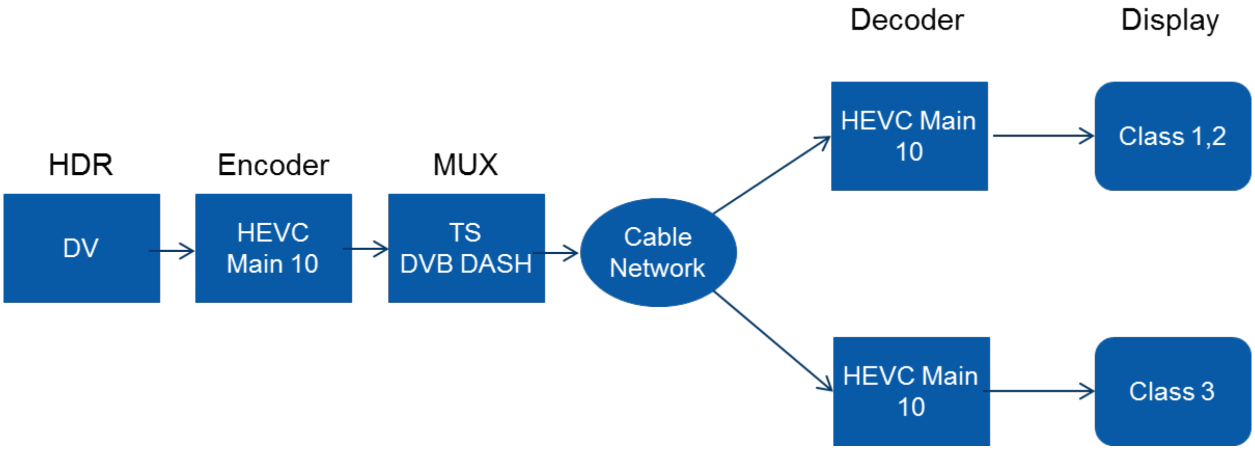


Figure 6 - HLG

7.6. SHVC (Scalable High Efficiency Video Codec)

SHVC is a solution that will transmit BT 709 to Class 1 TVs and can also address Class 2 TVs with BT 2020. Through the SHVC enhancement layer, it can provide an HDR experience on a SHVC playback solution that will display to a Class 3 and above TV. This solution will require an SHVC-capable decoder. Thus far, the solution has not been announced by any semiconductor company, so its chances to be adopted by the market are low, although it is fully standards-based. Metadata is defined in SMPTE ST2086, ST2094.

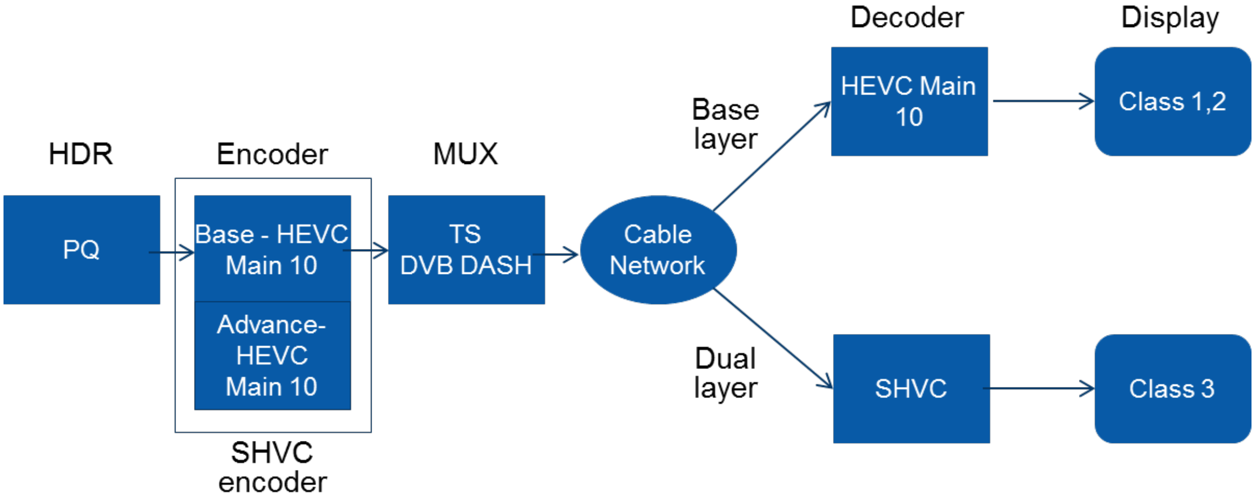


Figure 7 - SHVC

7.6.1. Standardization outlook

In terms of standardization, DVB is currently evaluating all of the technologies discussed throughout this paper and should reach a decision in the summer 2016. One important element in the discussion is ITU-R standardization for HDR. So far PQ (that include HDR10) and HLG (single EOTF) are the only ones selected for standardization by ITU-R.

7.7. HFR solution

The sports community is demanding an increase in the frame rate, namely to 120p. In order to achieve that, a new 120fps workflow needs to be created. As none of this exists today, even for HD resolution, it will take some time to be deployed. On the receiver side, HDMI 2.0 is only capable of supporting 1980 x 1080 x 120 10-bit 4:2:2, so if broadcasters want to reuse existing infrastructure, they will have to limit the resolution. A common perception is that the existing Main 10 Level 5.1 decoder should be able to decode 2540 x 1440 x 120 10-bit 4:2:0, as this is a pixel rate included in Level 5.1. The problem is that most of the existing chips were designed for a p60 output, and moving to p120 will require a redesign of the silicon.

Of course, there is always the option to transmit 3840 x 2160 x 60 and have the TV do the frame-rate conversion in the TV. Harmonic and Sigma Designs demonstrated this at IBC2014, and the audience was impressed by the result. Orange, a partner of the 4EVER Project, has also compared this method to a native 120p transmission and came to the conclusion this FRC was acceptable in terms of video quality for most content.

DVB is looking at HFR as a long-term deployment and plans to have it available in the 2019 timeframe. The proposed scheme for backward compatibility should be based on a temporal scalable approach with a base layer being SDR. Another option would be to have a base layer that is HDR. Table 6 lists all the options for HFR.

Table 6 - HFR Options

Option	HDMI	Layer Coding	HEVC Level	Max Resolution	max frame rate
1	2.0	single	4.1	1080p120	p120
2	New	single	5.1	1440p120	p120
3	New	single	5.2	3840x2160	p120
4	2.0/ New	dual	5.2	3840x2160	p120

Option 1 exists today and can work on some decoders and TVs that support HDMI 2.0 (See Figure 8).

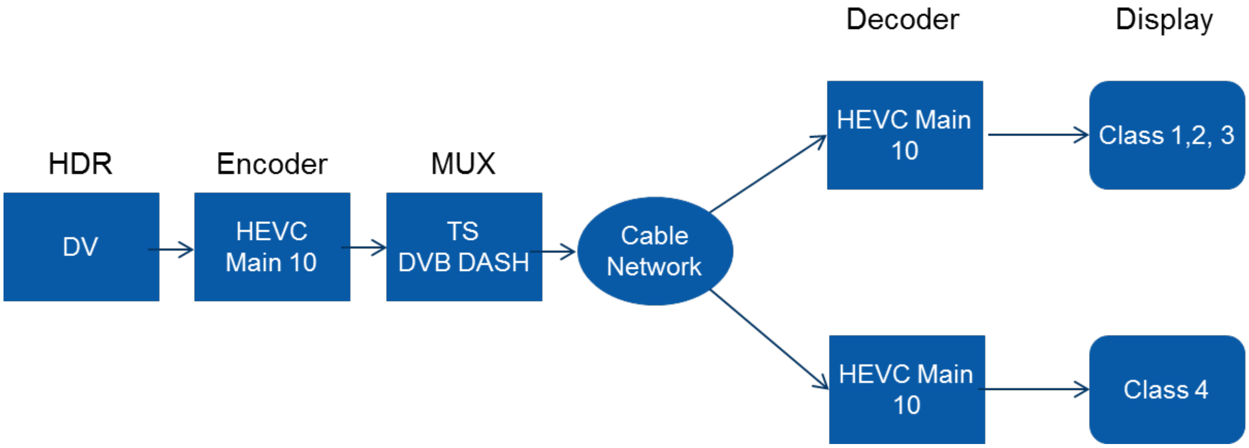


Figure 8 - HFR 1080p120

Option 2 is based on existing decoder technology, the biggest problem being the display refresh rate at 120 Hz for the TV. This is a solution that could find some traction as it has a nice trade-off between resolution and frame rate (See Figure 9).

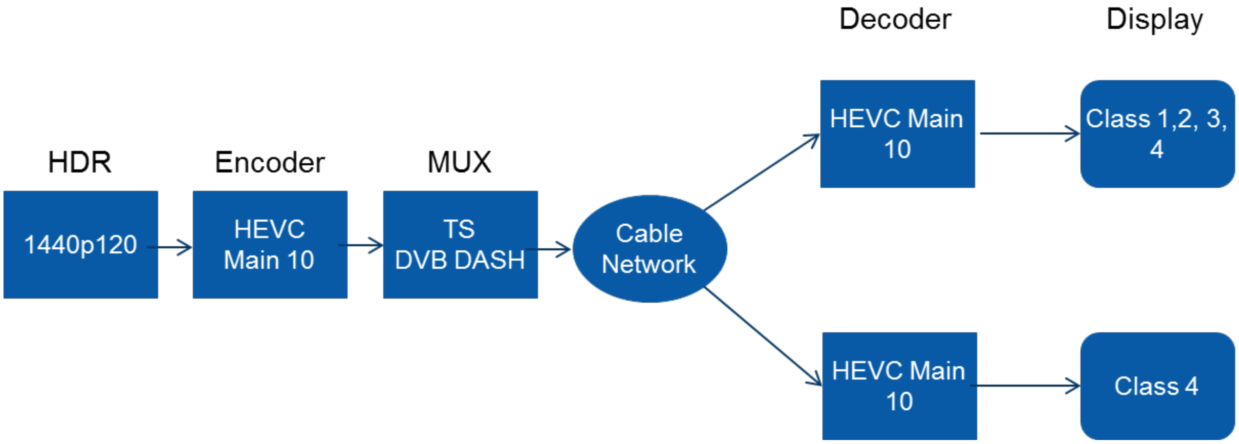


Figure 9 - HFR 1440p120

Option 3 requires a brand-new design for the decoder, display and encoder. This type of solution will likely be more expensive, and the adoption will be gated by the technology maturity. At this point, 2019 seems a reasonable target to launch such a service. (See Figure 10.)

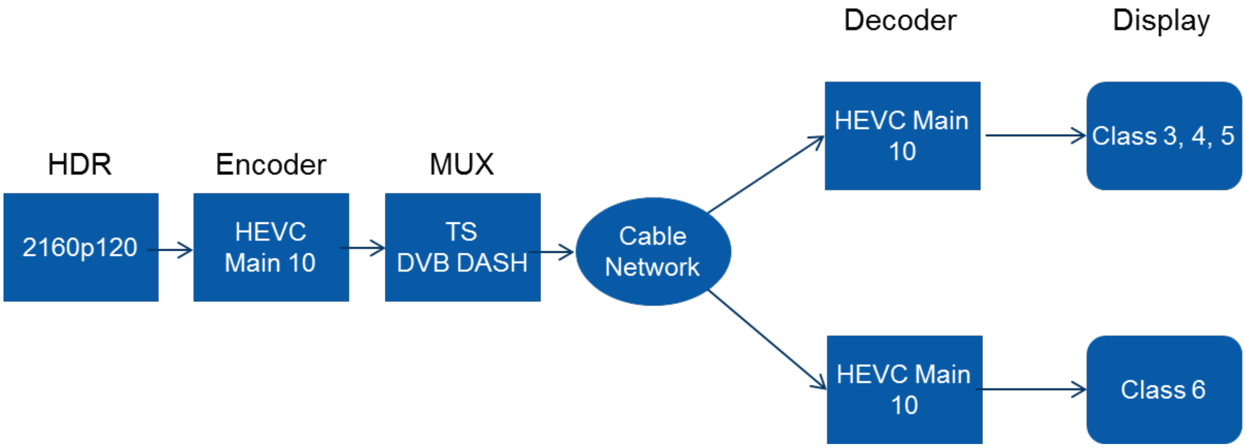


Figure 10 - HFR 2160p120

Option 4 is a temporal scalable solution that will have a base layer either in SDR (Class 1 and 2) or HDR (Class 3). While this solution is complex, the fact that it offers backward compatibility could be appealing. (See Figure 11.)

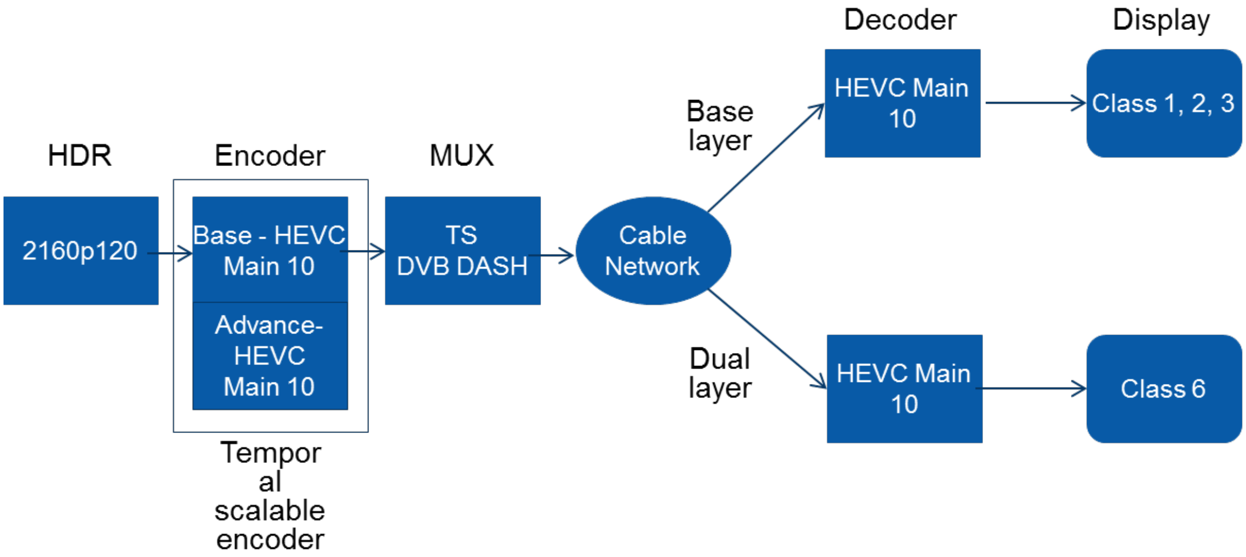


Figure 11 - Temporal Scalable

In conclusion, multiple technology options are possible for HFR. On the production side, p120 is not yet ready, and 2160p120 displays are not mainstream; therefore, HFR is expected to come in a second phase. As many 2160p60 UHD TVs will be deployed, a temporal scalable scheme makes a lot of sense.

8. Recommendations And Guidelines

In addition to the SDO defining standards for Ultra HD, there are other industry groups providing recommendations for the deployment of UHD. Those recommendations will have an impact on the definition of DVB UHD-1 Phase 2.

At CES, the UHD Alliance presented its guidelines [11], which will be accompanied by an Ultra HD Premium certification program. Table 7 provides a high-level summary of the UHD Alliance recommendations.

Table 7 - UHD Alliance Recommendations

Item	UHD Alliance
Publication	CES’16
Application	Disc / VoD Push & OTT
Deployment	2016
Codec	HEVC Main 10
Minimum resolution	3840 x 2160 x 24
Maximum resolution	3840 x 2160 x 60
WCG	BT 2020
Display	90% of P3 colors
HDR EOTF	ST 2084*
HDR metadata	Not specified
Peak brightness /black levels (nits)	1000/0.05 or 540/ 0.0005
Backward compatibility	Receiver based
Audio	NGA recommended

At CES 2016, Samsung, LG and Panasonic all announced they had certified UHD premium TVs. Sony announced a 4000 nits (vs 1000 nits for the Premium certification) TV and did not make any product commitment.

At NAB 2016, the Ultra HD Forum released its Phase A Guidelines for service providers looking at deploying live applications either in a broadcast or unicast environment [12]. Those guidelines are described in Table 8.

Table 8 - Ultra HD Forum Guidelines Preview for Live Phase A Deployments

Item	Ultra HD Forum
Publication	NAB 2016
Application	Live broadcast, multicast and unicast
Deployment	2016
Codec	HEVC Main 10 Level 5.1
Minimum resolution	1920 x 1080 x 50
Maximum resolution	3840 x 2160 x 60
Frame rate	24, 25, 50, 60 +
WCG	BT 2020 / BT 709
HDR EOTF	SDR / PQ10 * /HLG10 **
HDR metadata	Optional
Audio channels	Stereo or 5.1 or channel-based immersive audio
Audio codec	AC-3, EAC-3, HE-ACC, AAC-LC
Subtitling	CTA-608/708, ETSI 300 743, ETSI 300 472, SCTE-27, IMSC1
Transport	TS (broadcast and multicast) DASH ISO BMFF (ABR multicast and unicast)
Backward compatibility	PQ10: simulcast or decoder based HLG 10: built in

+ Fractional frame rate accepted
* PQ10 is HDR10 without metadata
** HLG10 is HLG with BT2020 Main 10

It is important to note that no metadata is used in the broadcast workflow, and there is no possibility to insert metadata in a live production from an economic and operational standpoint.

The Phase A guidelines have been reviewed by various SDOs (e.g., SMPTE, CTA, DVB, EBU, ATSC, CableLabs, SCTE) involved in the standardization of Ultra HD. Based on the feedback, a new version was released [12]. Following that, interoperability plug fests will be performed, during which time those guidelines will be tested, and new iterations of the guidelines created.

Phase B will cover all types of broadcast and unicast deployments for 2017, based on new technologies (such as SoCs), extending beyond the current defined HDR (PQ10 and HLG10). Additionally, it will introduce new features such as high fidelity (as proposed by Dolby Vision), advanced backward compatibility (as proposed by Dolby Vision and Technicolor), HFR and NGA.

9. ABR Scenario

In an ABR deployment scenario, there are fewer constraints on backward compatibility, and it serves different devices with different HDR and HFR requirements. Indeed, as the transmission is expected to be unicast, various implementations of HDR will be used for different devices. That is already happening today with VOD OTT services using diverse HDR for different devices.

10. Conclusion

While there are still a lot of moving parts, ITU-R and Ultra HD Forum guidelines provide some ground for UHD services to be deployed on existing infrastructure (i.e., encoders, STBs, TVs) in 2016. Cable operators will have to work closely with their content providers' partners and assess, from a deployment perspective, the most advantageous technologies to deploy.

11. Abbreviations

11.1. Abbreviations

ATSC	Advance Television System Committee
ARIB	Association of Radio Industries and Businesses
ABR	adaptive bit rate
BDA	Blu-Ray Disc Association
CES	Consumer Electronic Show
CTA	Consumer Technology Association
DOCSIS	Data-Over-Cable service interface specification
DASH	dynamic adaptive streaming over HTTP
DVB	digital video broadcast
EOTF	electrical optical transfer function
ETSI	European Technical Standard Institute
ETSI ISG CCM	industry specification group on intelligent compound content management
FTTH	fiber to the home
FRC	frame rate conversion
HDR	high dynamic range
HFR	high frame Rate
HLG	hybrid log gamma
HLS	HTTP live streaming
ITU	international telecommunication union
IBC	international broadcast conference
MPEG	motion picture expert group
MSO	multi system operator
NGA	next generation audio
OETF	optical electrical transfer function
PQ	perceptual quantizer
QoS	quality of service
QAM	quadrature amplitude modulation
SDR	Standard Dynamic range

SDO	standard defining organization
SoC	system on chip
SMPTE	Society Of Motion Pictures And Television Engineers
SHVC	scalable high efficiency video codec
STB	set top box
UHD	ultra high definition
WCG	wide color gamut

12. References

[0] TS 101 154 v2.2.1:
http://www.etsi.org/deliver/etsi_ts/101100_101199/101154/02.02.01_60/ts_101154v020201p.pdf

[1] DVB DASH: https://www.dvb.org/resources/public/standards/a168_dvb-dash.pdf

[2] DVB Subtitling Systems:
http://www.etsi.org/deliver/etsi_en/300700_300799/300743/01.03.01_60/en_300743v010301p.pdf

[3] Coding Constrains for UHD Blu-ray: http://www.blu-raydisc.com/Assets/Downloadablefile/BD-ROM-AV-WhitePaper_HEVC_150608a-clean.pdf

[4] CTA Defines “HDR Compatible” Displays: https://www.cta.tech/News/News-Releases/Press-Releases/2015-Press-Releases/CEA-Defines-%E2%80%98HDR-Compatible%E2%80%99-Displays.aspx

[5] Video Specifications IP Multicast: http://www.cablelabs.com/wp-content/uploads/specdocs/OC-SP-MS-EMCI-I01-150528.pdf

[6] Sky Germany HDR trial: http://www.insideci.co.uk/news/sky-looking-to-score-with-hdr-4k-uhd-sports-coverage.aspx

[7] Harmonic demonstration of Dolby Vision at IBC 2015: http://harmonicinc.com/news/harmonic-demonstrate-innovation-ip-video-workflows-ultra-hd-and-virtualization-ibc2015

[8] ITU-T BT 2100: http://www.itu.int/rec/R-REC-BT.2100

[9] SES HLG demonstration with LG: http://www.lg.com/sg/press-release/lg-to-demonstrate-high-hdr-capabilities-of-oled-tv-at-ses-astra-industry-day

[10] Sky PerfecTV launches HDR: http://www.inter-bee.com/en/magazine/delivery/detail.php?magazine_id=2872

[11] UHD Alliance recommandations (CES 2016):
http://www.businesswire.com/news/home/20160104006605/en/UHD-Alliance-Defines-Premium-Home-Entertainment-Experience

[12] Ultra HD Forum Guidelines (July 2016): http://ultrahdforum.org/resources/phasea-guidelines-description/

JOURNAL OF
DIGITAL VIDEO

SCTE ISBE

Minimizing Delivery Latency of Linear Video Service via Adaptive Transport Stream and HTTP Chunked Transfer Encoding

A Technical Paper prepared for SCTE/ISBE by

Wendell Sun, Principal System Engineer, ARRIS, SCTE/ISBE Member
6450 Sequence Dr.
San Diego, CA 92121
(858) 404 2774
wendell.sun@arris.com

Table of Contents

List of Figures

1. Introduction

Over-the-Top (OTT) video streaming services are gaining more popularity. Besides a video on demand (VOD) type of service, such as Netflix, more and more linear types of streaming services are also available, such as DISH Network's Sling TV, Sony's Vue, and AT&T's DirectTV Now etc. The OTT video streaming uses Hyper Text Transfer Protocol (HTTP) for content delivery, while the media content is well-prepared and conditioned into multiple bitrate sources, which allow the client player to dynamically select the appropriate bit rate source and seamlessly switch between different bit rate sources per broadband network status. With this approach, it can reach almost any type of terminal device, provide transport resiliency to a network's condition and gives a much better user viewing experience. For the VOD type of service, this is probably perfect since VOD service is delivered to each end device individually in non-real-time fashion, e.g. IP unicast, and there is no concern for delivery latency and no demand to coordinate viewing experience among end users. The initial buffering delay most likely is acceptable, especially when there is a pre-roll ad play. The commercial application of this type of OTT service is very successful, such as Netflix. However, the HTTP-based transport has major shortcomings when it is applied for live or linear video delivery.

The generic live or linear TV service has the following characteristics:

- There is a real time timeline that is referenced by all viewing users at the same time when the content is captured, processed, delivered and consumed. For example, a sport event is being broadcasted while the event is happening
- Compared with an audience on site, normally there is a constant viewing delay for broadcast viewers. The viewing delay is counted from event happening to being viewed remotely and usually is caused by video content acquisition, editing, processing and delivering, as well as mandatory regulation delay request. The smaller the viewing delay is, the better the user viewing experience will be. See the Figure 1 below for reference
- Content is delivered to a large number of viewers simultaneously
- Constant viewing delay is same to all viewers

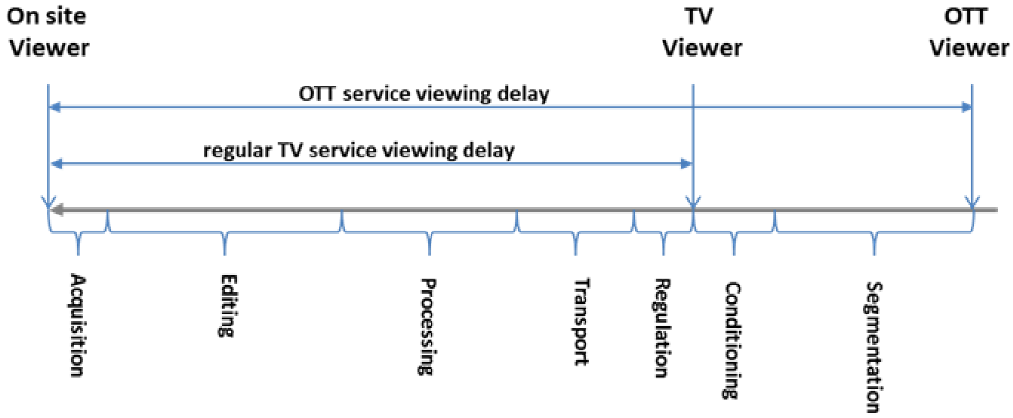


Figure 1 - TV Service Viewing Delays

In the OTT case, HTTP is used for content delivery. In general, HTTP is a transaction based protocol designed for file download. The current adaptive bit-rate streaming uses small segments to compromise HTTP file transfer request. Instead of bit by bit streaming, content is transported segment by segment.

Normally an HTTP transfer will not start until the whole segment is ready. This will add, at minimum, one segment length of extra time to delivery delay besides other processing delays. This is shown in Figure 1 as segmentation delay. The bigger the segment is, the longer the extra latency is. This may force small segments to be defined if low latency live TV service is desired. But for seamless switching purposes, a segment is bounded with an IDR (instantaneous decoding refresh) frame, e.g. closed group-of-pictures (GOP), which requires more coding bandwidth. The smaller the segment, the worse the video encoding efficiency will be. In addition, each segment delivery corresponds with one HTTP get/reply transaction. The smaller the segment is; more HTTP protocol traffic will be in network.

Before OTT service was available, Internet based streaming TV service, such as IPTV service, had a long history. IPTV services have been offered by telco operators to compete with cable operators for more than a decade. IPTV services provide subscribers with similar TV viewing experience, such as fast channel change time and low end to end transport latency, just like traditional broadcast or linear pay-TV service offers. It uses IP multicast as its primary protocol, which provides true data streaming, minimizes delivery latency, and supports content sharing among multiple clients. However, IPTV services require guaranteed bitrate match bandwidth for smooth service delivery. As a result, an IPTV service is only provided in a managed IP network.

This paper briefly reviews IP multicast streaming and examines existing or under-developing protocol and standard, such as the HTTP Chunked Transfer [9] and the CableLabs/SCTE Adaptive Transport Stream [4][5][6]. This paper proposes a solution, which combines IP multicast in backbone network and HTTP in edge access network, to minimize the delivery latency for live/linear video service without sacrificing video encoding efficiency.

2. TV Services via Satellite Broadcast and IP Multicast Streaming

Traditional TV services are all in broadcast mode. The content is acquired from source, edited, and processed for transportation over satellite. Satellite is used for large area content distribution, though fiber optical network may also be used by large MSO in their backbone distribution. Depending on the business model, there are a couple of ways of receiving content for consumption:

- For pay-TV cable operators, an integrated receiver/decoder (IRD) on multichannel video programming distributor (MVPD) headend receives, decrypts, and decodes the content, and then the content is re-distributed to subscribers via cable plant network
- For pay-TV satellite operators, a satellite set up box on subscriber’s home receives, decodes, and renders the content
- For over-the-air operation, it is similar to the pay-TV cable operation in the front line of receiving and decoding, but the last leg of content re-distribution is done by over-the-air radio transmission.

All of these are broadcast one-way from the content source to many content consumption terminals. This usually carries a constant viewing delay, as shown in Figure 1, as regular TV service viewing delay.

Another type of TV service, IPTV, uses broadband Internet as its distribution network. The Internet can support different types of content distributions, such as one-to-one unicast, one-to-many multicast, and one-to-any broadcast. However, IP multicast has significant advantages over IP unicast and broadcast for TV-like services.

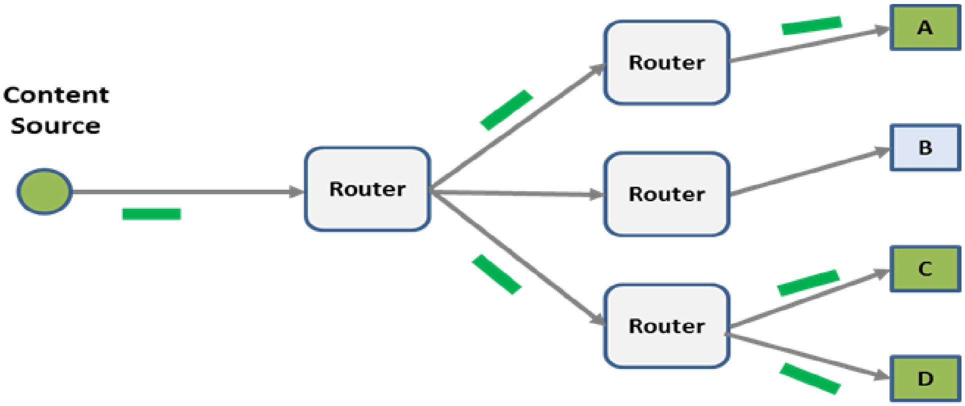


Figure 2 - IP Multicast Illustration

IP Multicast offers a one-to-many distribution model, similar to traditional broadcast TV services that are mentioned above but with one difference. In the world of Internet, IP multicast supports content delivery on per request or per registration basis. To receive content, the end user's device needs to send Internet Group Management Protocol (IGMP) requests to its router to ask for delivery. The router then relays the request all the way back to the origin; if the requested content is not available on it. To stop receiving content, the end user device can send IGMP quit to its router to stop the sending. The user B in Figure 2 is exempted from receiving content. The routers that support IP multicast service only send one copy of same content to the duplicated requests from the next router. For example, both user C and D ask for the same content, the upstream router just needs to send one copy of content to their router. So in this case, even if there is a single user, such as user A, or there are thousand users under the same router, the overall network traffic will look same, thus it scales very well and is much efficient for live or linear type of content distribution. Robinson's paper [3] has a lot of more details on IP multicast discussion.

More importantly, IP multicast is a true IP streaming protocol. Although content is delivered by IP packet, the protocol is designed for content flowing from point A to point B as long as point B joins the multicast group. IP packets are relatively, very small and do not introduce much extra transportation delay compared with other forms of bitstream content delivery.

IP multicast has been used by telco operators to provide IPTV service for a long time. It allows the IPTV service to provide a similar user experience as traditional broadcast TV service does. Even for some cable operators, IP multicast may also be used in their backbone for content distribution due to the popularity of Internet and IP network integration. To supply a satisfied IPTV service or a reliable backbone for content delivery, the only constraint of using IP multicast is the requirement of guaranteed bandwidth – that the bandwidth should be equal or higher than the bitrate of the content stream. For this reason, IP multicast is usually applied in a managed IP network with well-engineered bandwidth allocation.

Whether in a traditional broadcast TV service or in a IPTV service, MPEG-2 Transport Stream (MPEG-2 TS) is used for carrying video and audio content. MPEG-2 TS is the dominant container format for content distribution of all TV services. When an HTTP-based OTT TV service is developed, can it leverage this legacy content distribution system and only build a client facing edge side to support adaptive bit rate (ABR) streaming to multiple types of end user devices?

3. Adaptive Transport Stream

Adaptive transport stream (ATS) is a fully compliant MPEG-2 TS with an embedded segment boundary indicator. The ATS is designed as a mezzanine container format for adaptive streaming, which requires encoding/transcoding of multiple bit rate (MBR) streams, segmentation, and alignment to support seamless bit rate switching. Originally the MBR encoding/transcoding is done by encoder or transcoder, while the segmentation and alignment is processed by a packager. The purpose of ATS is to further exploit encoding/transcoding process to label the segment boundary, thus to relieve the effort of parsing and aligning segment in adaptive streaming packager.

The ATS definition started from CableLabs encoding boundary point (EBP) specification [4] and Adaptive Transport Stream Specification [5], in which an EBP structure is defined in the private data of adaptation field in MPEG-2 TS header to indicate the beginning of each segment. The work was promoted to ANSI/SCTE standardization process, in which SCTE DVS 1196 [6], has been created. Furthermore, it has been contributed to MPEG and an amendment to MPEG-2 Systems specification [7] has been generated to define a set of adaptation field (AF) descriptors to serve the purpose.

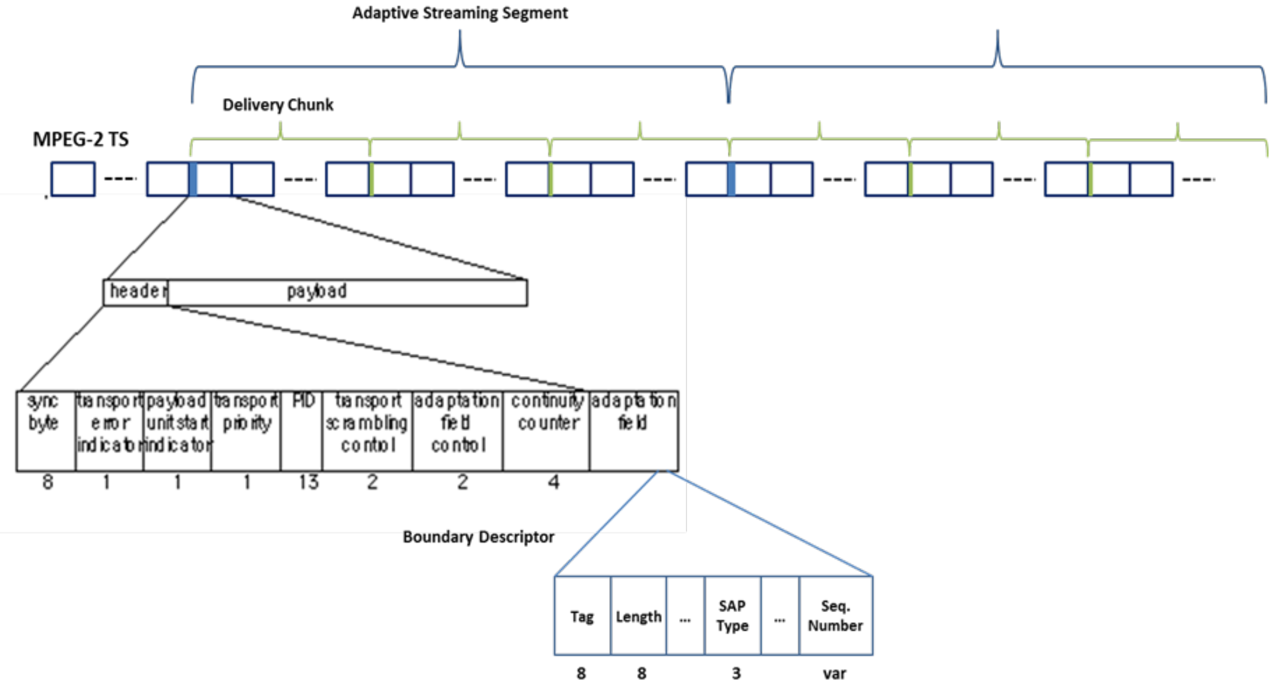


Figure 3 - Structure of Adaptive Transport Stream

As it is shown in Figure 3, the Boundary Descriptor is introduced into the adaptation field of MPEG-2 TS header. The two major parts are the “SAP type” and “Sequence Number” field. The SAP stands for stream access point and is defined by ISO/IEC 14496-12. A SAP type is a definition of media decoding attribute in that stream point. For example, a SAP type 1 means the media sample at the point can be fully decoded without referring other samples and all samples following it can also be correctly decoded. The Sequence

Number field has length of 2, 4, or 8 bytes depending on application. It provides a unique identifier for a segment within its context.

To support regular MBR adaptive streaming, the signaling of a Boundary Descriptor in the segment level is good enough. The packager takes ATS as input stream and needs only to parse the bytes in the transport stream packet headers in order to obtain the boundary information. It does not need to parse any bytes in the packet payload. However, to resolve the extra viewing delay caused by segmentation as it is discussed in the introduction section for live/linear TV service, a smaller segment is desired; and coding efficiency should not be impacted. The segment is designed for bit rate adaptive switching, thus it requires closed GOP at segment boundary. While the segment structure is maintained, a smaller delivery unit can be designed for low latency delivery.

The MPEG DASH specification [8] introduces a concept of Delivery Unit Media Segment to support low latency application. As shown in the Figure 3 - Structure of Adaptive Transport Stream, the Delivery Chunk is a smaller delivery unit within the Adaptive Streaming Segment. It can be made up by any group of meaningful coding samples, such as a GOP or even a frame, while making it small enough not to cause transportation delay. Similar to each segment, the delivery chunk can also be identified by the Boundary Descriptor embedded in MPEG-2 TS header.

The insertion of Boundary Descriptor into MPEG-2 TS header can be easily achieved as part of content encoding/transcoding process. It does not add much extra processing to encoder/transcoder, yet it is a big savings for the packager to not need to look up coding payload. It is even more essential when the segment durations are not fixed interval - whether that’s due to flexible GOP structures in the encoded video, or a content-related change, such as a frame-accurate demarcation in program or advertisement.

4. HTTP Chunked Transfer Encoding

The HTTP chunked transfer encoding (CTE) is defined by HTTP/1.1: Message Syntax and Routing [9]. It is a data transfer mechanism in HTTP 1.1, in which data can be sent in a series of "chunks" in responding a single HTTP data request. It uses the Transfer-Encoding header, instead of the Content-Length header. The HTTP sender does not need to wait for the total size of content being available and can start sending data as small chunks with any amount available while still receiving the content. When the chunk is sent, its size is indicated in the Transfer-Encoding header. At the end of content, it sends the last chunk with its size set to be zero. For example, the following is a HTTP transaction with Chunked Transfer Encoding:

```
GET /cte-example.html HTTP/1.1
Host: doc.micrium.com
User-Agent: Mozilla/4.61 [en] (Windows 7 6.1)
TE: chunked
```

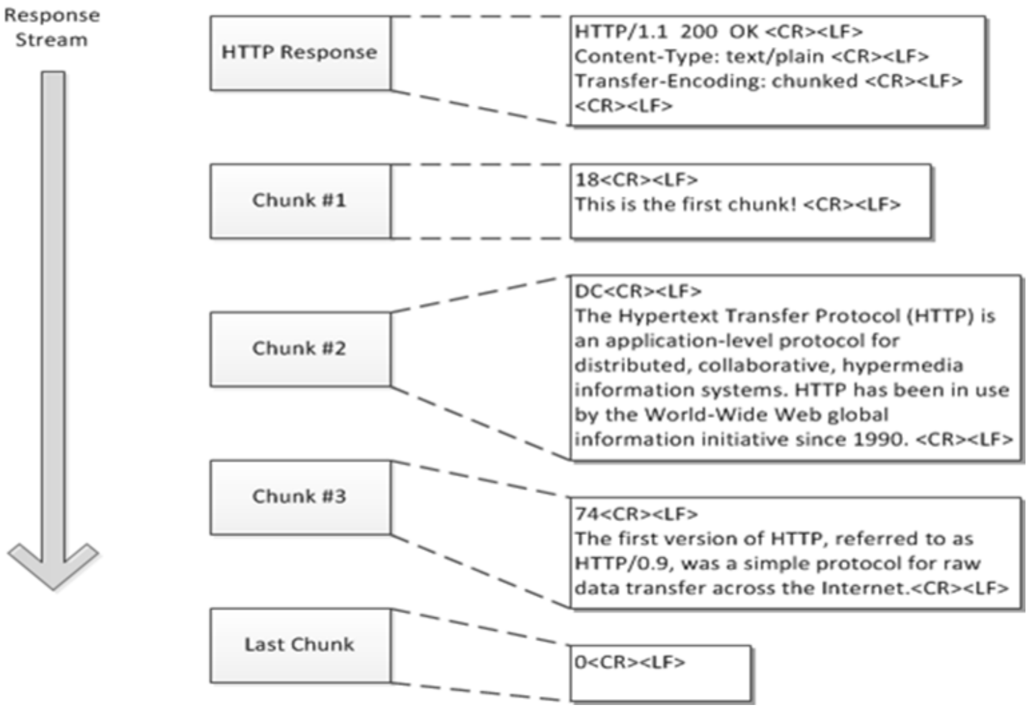


Figure 4 - Example of HTTP Chunked Transfer Encoding Response [10]

5. HTTP Streaming Server

The existing TV services, either cable TV service via satellite based distribution or IPTV service using IP multicast network, have built their business models and have been accepted by millions of subscribers. The OTT TV service is a new model using HTTP protocol. Can these two types of content distribution be combined, especially if it can leverage the existing content distribution model as a backbone network and serve all types of client devices of OTT TV service and provides the same or similar quality of user experience?

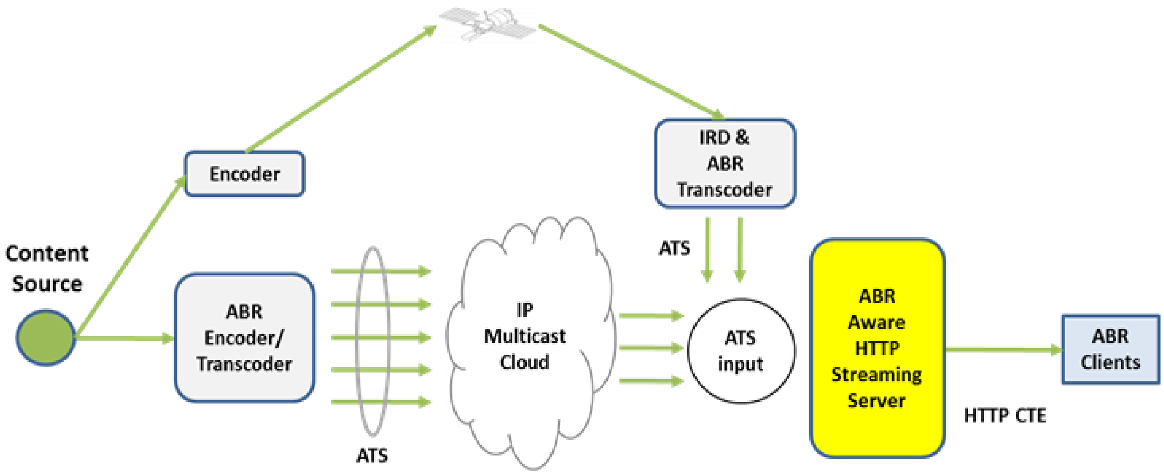


Figure 5 - Hybrid Transport Network for Adaptive Streaming

Figure 5 shows a hybrid transport network. Both satellite based content distribution and IP multicast based network cloud support stable and satisfied TV services. The proposed hybrid transport network uses one of them as a backbone content delivery network, whichever is available. In the satellite distribution case, MBR transcoding and ATS creation are done in MVPD headend, while in IP multicast case, they are done before entering IP multicast cloud. There could be a third case, which is not shown in the diagram, in which content may be first distributed via satellite and then further distributed via IP multicast cloud. What is common in all cases is the creation of ATS. As we have discussed in the Adaptive Transport Stream section, ATS are MPEG-2 TS streams with embedded boundary descriptors to virtually divide them into delivery chunks, and further into adaptive streaming segments.

The client side remains with HTTP based ABR streaming, which allows content to pass most access networks and reach almost all types of client devices.

The ABR Aware HTTP Streaming (A2HS) server is a new component that bridges the two sides. It is an off-the-shelf HTTP server with additional functions of ATS parsing, optional ABR format conversion, and HTTP CTE support.

In usual ABR streaming, a content segment is not made available in manifest to an ABR client until the segment is ready for delivery. This is exactly the cause of segmentation delay. In the proposed A2HS server solution, the content segment(s) can be published in manifest even before the input ATS is pulled, e.g. IP multicast join. When the A2HS server receives content segment requests from an ABR client and the corresponding IP multicast ATS stream is available (otherwise it initiates IGMP to join to the IP multicast group to get it), the server starts parsing the MPEG-2 TS header of the input ATS via seeking the Boundary Descriptor of delivery chunks and sends the delivery chunks via the HTTP CTE approach as it is discussed in the HTTP Chunked Transfer Encoding section while it continues to receive the IP multicast ATS input.

At the end of content segment, which is indicated by an adaptive streaming segment Boundary Descriptor, it just needs to send a zero sized chunk to finish the segment transfer. The ABR client sends next segment request and repeats, and the A2HS server continues serving content with incrementally delivered chunks of segment. If bit rate switching is required, the ABR client sends the next segment request of the switch-to ATS representation. ABR clients can only switch on an adaptive streaming segment boundary. At the same time, the content delivery is in the delivery chunk interval. If the delivery chunk is designed small enough, the HTTP CTE based transport provides a similar function of true streaming.

6. Conclusions

Compared with existing TV services, OTT TV service – especially live and linear TV service – adds extra viewing delay caused by ABR segmentation. This paper reviewed the advantages of satellite based broadcast TV service and IP multicast based IPTV service, as well as under-developed ANSI/SCTE ATS standard and HTTP CTE protocol. The paper proposed a hybrid transport network that leverages the existing broadcast/multicast distribution network as a backbone and bridges it with HTTP CTE on ABR clients. This approach can minimize the extra viewing delay caused by OTT segmentation and improve an OTT TV service's quality of experience.

7. Abbreviations

ABR	adaptive bit rate
AF	adaptation field
ATS	adaptive transport stream

CTE	chunked transfer encoding
EBP	encoding boundary point
GOP	group of pictures
HTTP	Hyper Text Transfer Protocol
IDR	instantaneous decoding refresh
IGMP	Internet Group Management Protocol
IRD	integrated receiver/decoder
MBR	multiple bit rate
MVPD	multichannel video programming distributor
OTT	over-the-top
SAP	stream access point
SCTE	Society of Cable Telecommunications Engineers
VOD	video on demand

8. Bibliography and References

[1] Extending Cable Workflows to OTT: Decoupling Content Generation from Delivery Technologies, Y. Syed, A. Giladi, N. Kipp, W. Sun, SMPTE 2015 Annual Technical Conference and Exhibition

[2] An Architectural Approach towards Achieving Bandwidth Efficient Content Delivery, D. Bhat, N. Kannan, W. Sun, INTX 2016 Spring Technical Forum

[3] The Return of Multicast: Why it Succeeds in a Live Linear World, D. Robinson, http://www.streamingmedia.com/Articles/Editorial/Featured-Articles/The-Return-of-Multicast-Why-it-Succeeds-in-a-Live-Linear-World-108621.aspx

[4] CableLabs OC-SP-EBP-I01-130118, Encoder Boundary Point Specification, www.cablelabs.com

[5] CableLabs OC-SP-ATS-I01-140214, Adaptive Transport Stream Specification, www.cablelabs.com

[6] ANSI/SCTE DVS1196, Adaptive Transport Stream, www.scte.org

[7] ISO/IEC 13818-1:2015 PDAM 7 Virtual Segmentation

[8] ISO/IEC 23009-1:2014 DAM 4, Segment Independent SAP Signaling (SISSI)

[9] IETF RFC7230, Section 4.1, Chunked Transfer Encoding, http://tools.ietf.org/html/rfc7230#section-4.1

[10] HTTP Chunked Transfer Encoding, https://doc.micrium.com/display/httpref/Chunked+Transfer+Encoding

JOURNAL OF
DIGITAL VIDEO

End-To-End IP Video Services Over 10g EPON Access Network Architectures

A Technical Paper prepared for SCTE/ISBE by

Sebnem Zorlu Ozer, Ph.D., Distinguished Systems Engineer, ARRIS, SCTE/ISBE Member
101 Tournament Dr.
Horsham, PA 19044
sebnem.zorluozer@arris.com
215-323-1154

Mike Emmendorfer, VP, Systems Engineering and Architecture, ARRIS, SCTE/ISBE Member
mike.emmendorfer@arris.com
314-616-1152

Table of Contents

List of Figures

List of Tables

1. Introduction

IP Video has been a major force driving cable operators to migrate towards software-driven, cloud-based, all-IP systems. Ethernet Passive Optical Network (EPON) is one of the preferred access network technologies to accelerate the migration to this new network and service architecture while providing a future-proof platform.

While MSOs' end goal is to have elastic service architectures, low-cost agile operations, and scalable and reliable platforms, the migration options and strategies may not be always clear. This is especially true for a new architecture that changes the traditional concepts of vendor products, customer services, and business models. Today's vendor specific product, with well-known interfaces and deployment locations, may now involve multiple physical and virtual entity options from different vendor and open architecture groups. A headend purpose-built EPON OLT (EPON Optical Line Terminal) with integrated Physical (PHY) layer, Media Access Control (MAC) layer, Traffic Management (TM), Layer 2 (L2), and Layer 3 (L3) forwarding, and control/management architecture may now consist of abstracted and disaggregated functions. Strategies may require new partnerships, processes, and integration methodologies.

Depending on the cable operator's current conditions and long-term targets, a centralized EPON OLT architecture with Software-Defined Networking (SDN) support and backward compatibility with current Operations and Business Support Systems (OSS/BSS) may be a first choice. In another case, the preferred approach may be to deploy a distributed EPON OLT with a subset of network functionalities and integrated to virtualized network and subscriber management blocks in the core and cloud architecture. In addition, both control/management and data planes of previously separated product groups may now interact with each other. Video control and access network control planes may be orchestrated to provide better resource utilization and enhanced Quality of Experience (QoE) as a step to provide end-to-end service QoE and Quality of Service QoS, network resource visibility, and troubleshooting. This in turn may help enable service mobility and chaining two or more service functions that can be managed per subscriber and flow policies, and mapped to available network resources. Mapping QoS, an objective measure by quantitative assessment of service and network conditions to QoE, a subjective measure by assessment of end user's level of satisfaction for the services is required for this purpose [9].

In this paper, we will first present a brief overview of Internet Protocol Television (IPTV) service architectures and EPON technologies by focusing on a centralized headend OLT and a distributed remote OLT option. Using test results and analysis for a use case, we will provide throughput benchmarks for both options. Given achievable throughput levels at the access network, we will then discuss IPTV services, QoE requirements and traffic scales. We will present some example architectures and use cases to investigate migration options and provide insights on next steps. SDN and Network Function Virtualization (NFV) concepts will be discussed for both centralized and distributed options to offer a more efficient system architecture that can expedite the integration of new IP video services and technologies while minimizing capital and operational costs. We will focus on the QoS functions in a more programmable network with dynamic service aware resource allocation through centralized control and service assurance.

2. EPON Architectures

There are two main architectures for Ethernet Passive Optical Network (EPON) access systems in the cable networking space, mainly centralized and distributed EPON access architectures [1]. The centralized access architecture (CAA) corresponds to an EPON OLT with the network media access control (MAC) and physical (PHY) layers placed in the service provider's facility. The distributed access architecture (DAA) places the access layer MAC and PHY layers or just the PHY layer in the outside plant or Multi Dwelling Unit (MDU) location. Both architectures may be designed to have separate control/management and data planes (Software-Defined Networking - SDN) and Virtualized Network Functions (VNF) deployed with cloud applications. Different architecture options are possible as described in [1] per the distribution of the following main functions:

- **Router/Multilayer Switch Router (MLSR)** with the following sub-functions: L2 switching, IPv4/IPv6 static and dynamic routing, Dynamic Host Configuration Protocol (DHCP) relay agent, access control lists, link aggregation, control plane redundancy, data fabric interconnecting line cards to the network interface ports.
- **Control plane functions** with sub-functions such as configurations of interfaces, modules and links, statistics collections and analytics, programming EPON MAC and Upstream (US) and Downstream (DS) Traffic Management functionality and control plane for multicast forwarding.
- **DOCSIS Provisioning of EPON (DPoE) Management and Control Layer** supporting fault, configuration, accounting, performance, security (FCAPS) for EPON technology using existing Data Over Cable Service Interface Specification (DOCSIS®[1]) based back office OSS and Network Management Systems (NMS) and processes via Simple Network Management Protocol/Command Line Interface (SNMP/CLI) interfaces and protocols. The virtual Cable Modem (vCM) handles all the operations, administration, maintenance, and provisioning (OAMP) functions for DOCSIS by proxying requests, signaling, and messages to the registered Optical Network Unit (ONU) using EPON OAM messages to provision all required services.
- **DS Traffic Management** with the following subsections: forwarding of data towards EPON MAC per provisioned rules such as classification to service flows, subscriber management filtering, prioritized and parameterized scheduling and QoS algorithms (i.e. per DOCSIS QoS and MESP bandwidth profile), Service Flow/Aggregated Service Flow (SF/ASF) Virtual LAN (VLAN) encapsulation and Differentiated Services Code Point/Type of Service/Tag Protocol Identifier/Class of Service (DSCP/ToS/TPID/CoS) mapping and marking. DS TM also supports multicast forwarding and replication, protocol throttling, packet statistics (Internet Protocol Detail Record (IPDR)) and Communications Assistance for Law Enforcement Act/Lawful Intercept CALEA/LI.
- **US Upper Traffic Management** with sub-functions such as forwarding data from EPON MAC towards the Network to Network Interface (NNI) by supporting classification and policing, cable source verify, protocol throttling, MAC learning, subscriber management filtering, DSCP/ToS/TPID/CoS mapping and marking, packet statistics (IPDR) and CALEA/LI.
- **US Lower Traffic Management (LTM)** supporting Dynamic Bandwidth Allocation (DBA) solicited scheduling with prioritized and parameterized QoS shaping algorithms and unsolicited scheduling and Multi-Point Control Protocol (MPCP) processing. It is common to have US LTM integrated in the EPON MAC subsystem.
- **EPON MAC Layer** based on 802.3 (802.3ah and 802.3av) standards supporting Multipoint Control Protocol (MPCP), encryption, LLID (Logical Link Identifier) tunneling and OAM functionality.
- **EPON PHY Layer** supporting PON optics (PR-type Physical Medium Dependent) and FEC.

In this paper, we focus on two main architectures shown in Figure 1 and Figure 2 to analyze IPTV services over 10G EPON networks. In an SDN enabled EPON access network, data, control and

---

[1] DOCSIS is a trademark of CableLabs

management planes are decoupled and physically separated so that each plane can be implemented on different hardware and software platforms controlled by a centralized SDN controller. As shown in the figures, the DPoE System and vCM may be deployed on a server or in a cloud. Router functionality may be controlled with a separate controller. SDN enabled EPON architectures provide a software programmable network with open standard based protocols and interfaces. Centralized control and orchestration increase video QoE by controlling different network and service elements based on end-to-end service assurance and resource management. Cable operators can extend their IP video services and business models with dynamic creation of new services.

Router function may be virtualized to decouple hardware and software. Different compute, storage and network functions may run on commodity hardware and can be deployed and scaled dynamically depending on service and network changes by a service and network orchestration.

Applying both SDN and NFV with efficient disaggregation of EPON and video functions will enable access I/O functions to be distributed and deployed in the cloud. Micro-services deployed on a common cloud infrastructure will enable customer care, operational and new business applications (e.g. Over the Top (OTT) providers). Same video architecture and SDN/NFV platforms may be deployed for different access technologies.

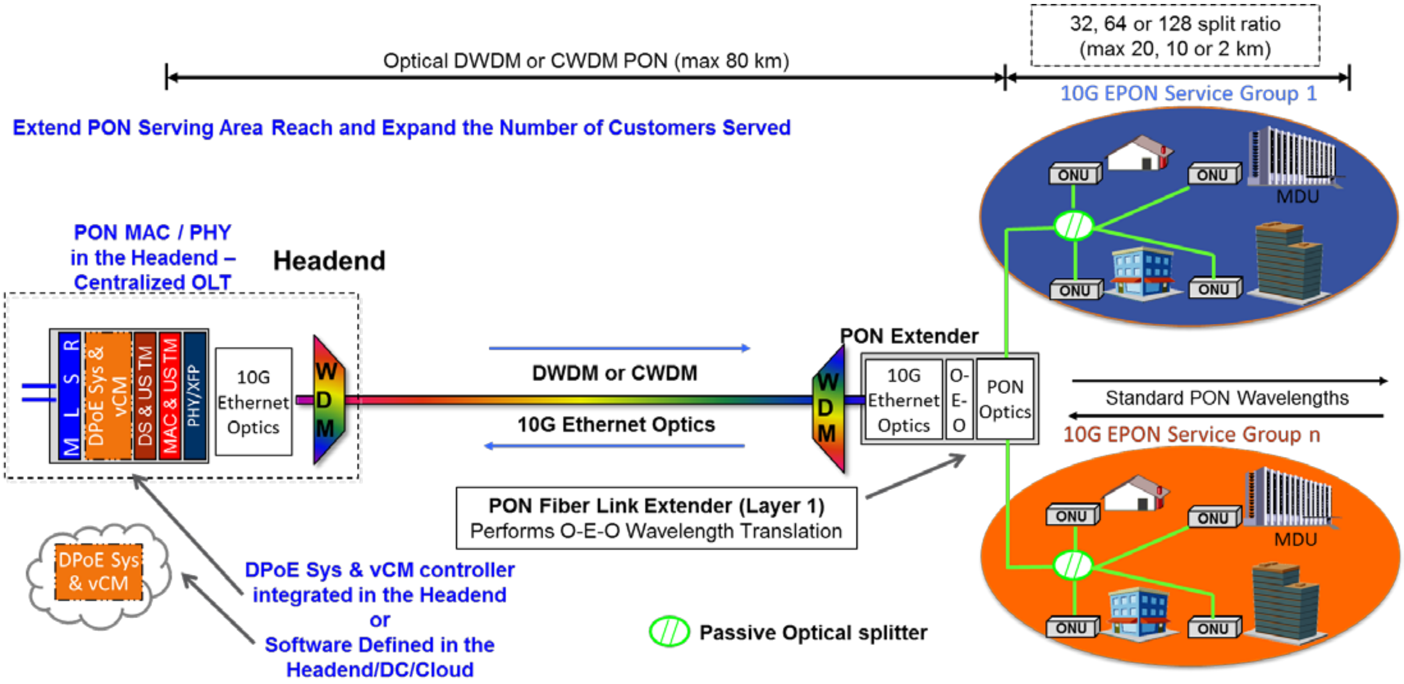


Figure 1 - CAA EPON Architecture with PON Extender

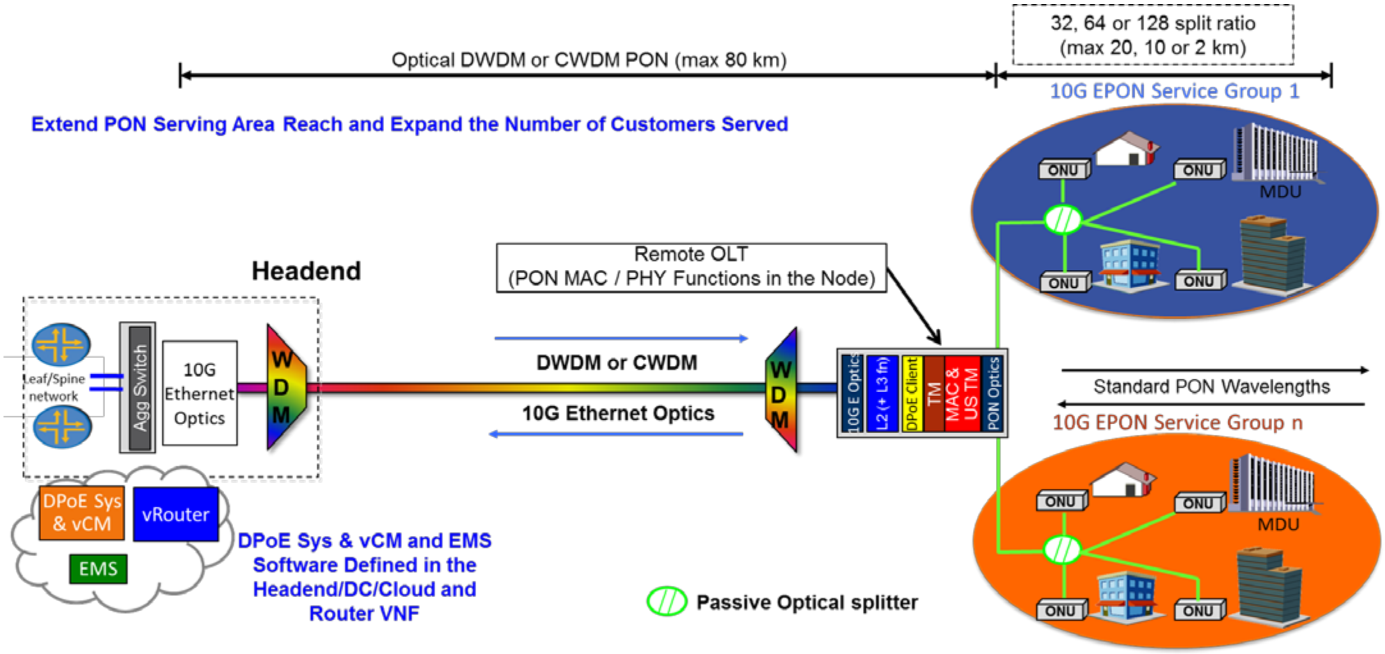


Figure 2 - DAA EPON Architecture with R-OLT

Depending on the network and service architectures, and requirements, some service providers may prefer a centralized EPON access network with a passive Optical Distribution Network (ODN). However, placing the OLTs in the headend/central office (HE/CO) facility may restrict the reach between the HE/CO and end users and/or reduce the split ratio and number of subscribers served per OLT port. PON Extender, a low power active in the node or cabinet, may overcome these issues and maximize fiber utilization between the serving area and the hub/headend while keeping the PON MAC in the facility. Dense Wavelength Division Multiplexing (DWDM) or Coarse Wavelength Division Multiplexing (CWDM) optics may be used for the trunk link between the OLT in the hub/headend and the PON Extender that may perform an optical-to-electrical-to-optical (O-E-O) conversion. Standard EPON optics for short reach may be used between the PON Extender and ONUs. In this architecture, the DPoE System and vCM management and control functions may be integrated in the headend OLT or may be placed in the data center (DC) or cloud. More information on the SDN and NFV concepts for access networks may be found at [7].

The remote OLT (R-OLT) access architecture places the entire OLT functions in the remote device, including the PON MAC as well as most of the PON upstream (US) and downstream (DS) traffic management functions. The DPoE System and vCM management and control functions are placed in the headend, DC or cloud. This architecture also introduces an Element Management System (EMS) for the R-OLT device as well. The R-OLT improves the split ratio and maximizes fiber utilization between the serving area and the hub/headend similar to the PON Extender case. The R-OLT may be connected to an aggregate switch in the headend using the same wavelength multiplexing architecture as the PON Extender. In addition, having MAC and TM in the node improves the US throughput without requiring a higher buffer size and improves the delay as well. As we will discuss in the next sections, the architecture may also help to reduce the delay of upstream video service requests. This solution introduces relatively higher power active node and requires integration of functions between the node and headend/DC/cloud entities.

As analyzed in [1] and some research papers, long reach EPON networks with centralized OLT architecture introduce additional delay in the DBA grant-report architecture, which may affect service delay and throughput requirements for some use cases (e.g. mobile backhaul applications and single LLID/single traffic flow use case with very high bandwidth demand (e.g. >5Gbps) depending on the OLT/ONU features). Note that unsolicited scheduling algorithms (such as unsolicited grant scheduling type instead of solicited DBA scheduler) available in today's EPON MAC chips would be more suitable for periodic and low latency services such as mobile backhaul applications. The other use case of single high-bandwidth flow may be applicable, for example, when a high tier user over a very long reach EPON network performs an upstream bandwidth test. In this case, if a R-OLT and headend OLT are compared, extra optical trunk link delay in the REPORT-GATE messaging (between ~200us to 800 us for 20km to 80km reach) may cause packet drops for the corresponding traffic flow if the ONU link queue size is not adequate.

We tested the CAA system with a PON Extender by using the test setup displayed in Figure 3. The aim of these tests is two-fold: 1) Test results show that PON Extender itself does not introduce any delay or jitter due to the O-E-O conversion designed for EPON networks. The only delay is due to the long reach trunk fiber in the US direction. 2) Test results show that the maximum throughput may be reached for high tier ONUs at 60 km reach which covers the most MSOs' deployments. New ONUs have higher buffer size and the delay for this range may be acceptable for many services. Depending on the service and scale requirements PON Extender or Remote OLT architectures may be preferred by cable operators.

A single UDP upstream flow is created via a traffic generator and ONU link buffer size is configured to observe the maximum achievable throughput. These tests aim to provide a benchmark for the maximum US and DS bandwidth/link rates by using available DBA algorithms with optimal configuration. In other use cases, additional inefficiencies inherent to EPON scheduling may exist in both headend and remote OLT architectures. Note that current DPoE specifications define maximum 255 x 4 kB link queue size. A vendor specific OAM is used to increase this value for the test results shown in Figure 4 for optical trunk length of 60 km and ~3 km between the PON Extender and ONUs. It can be shown that for this use case, 1800 kB of link queue size enables the same maximum achievable throughput (~8.67 Gbps) as the R-OLT case. Some newer ONUs have higher buffer sizes. In the future, they may be adaptively configured per traffic characteristics using SDN/NFV features. Downstream UDP throughput values for data from the network side and buffer requirements are the same as R-OLT case.

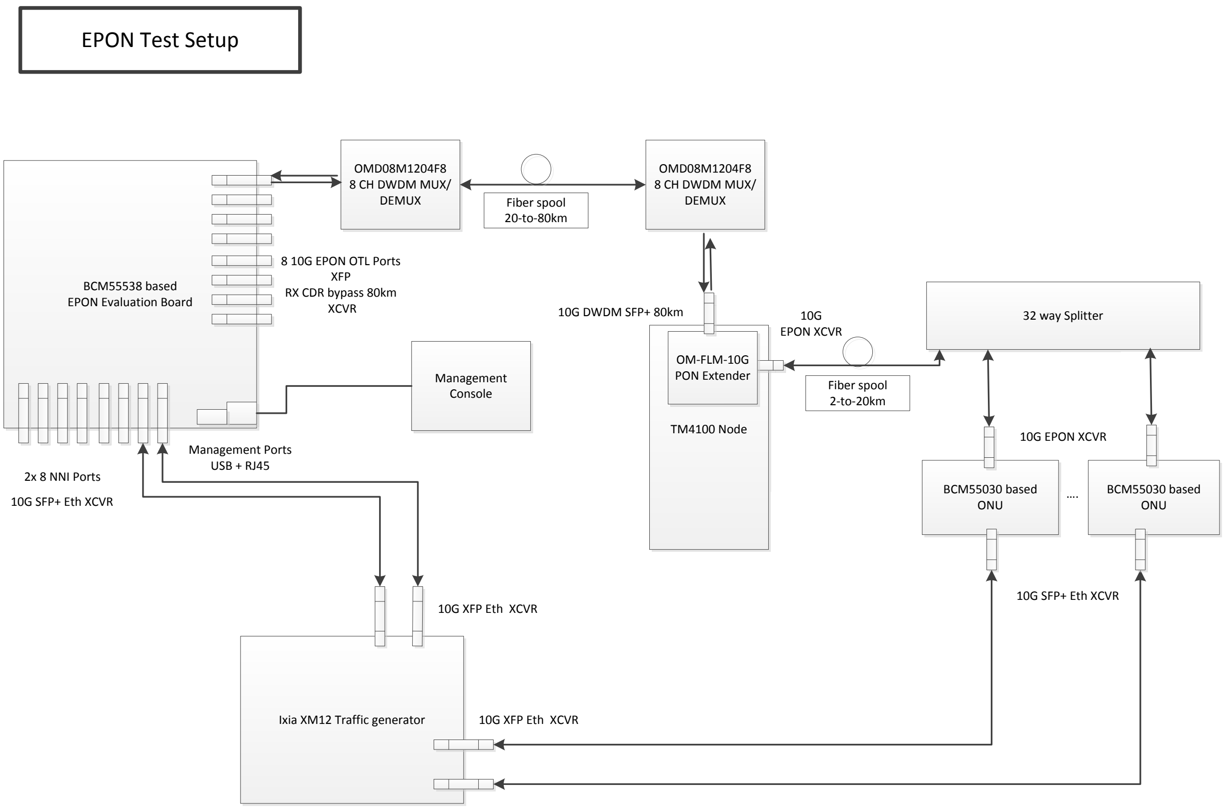


Figure 3 - 10G EPON Test Setup

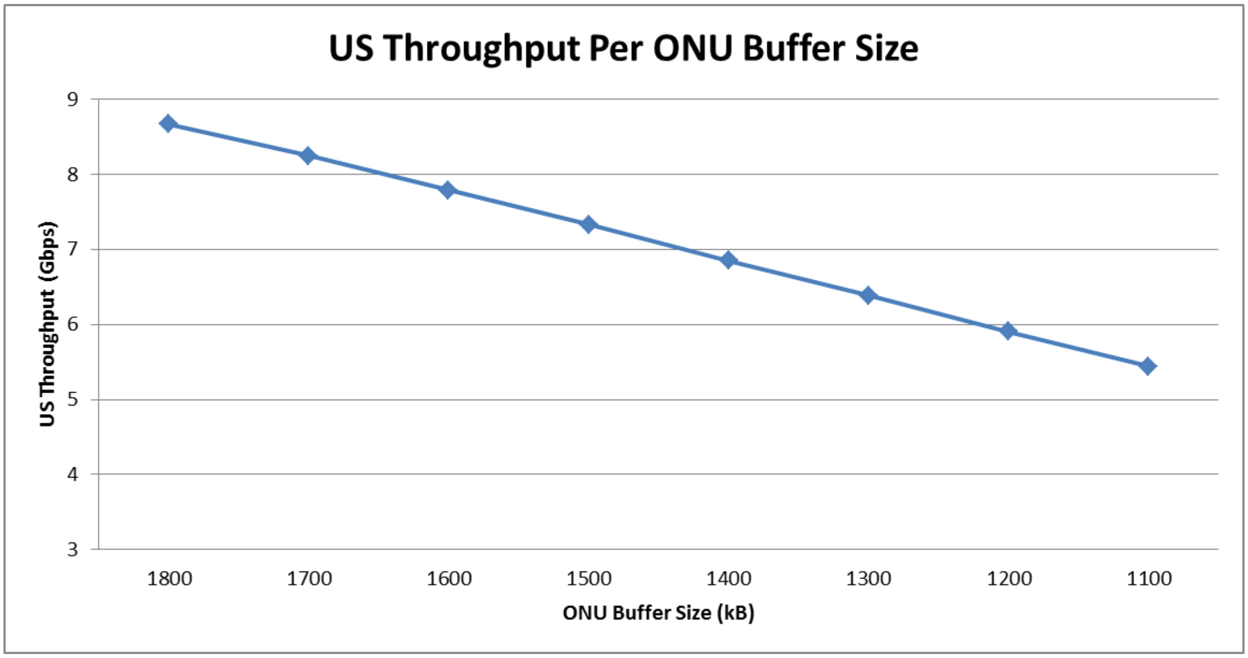


Figure 4 - Upstream Throughput of a Single Flow per ONU Link Buffer Size at Optical Trunk Length of 60 km

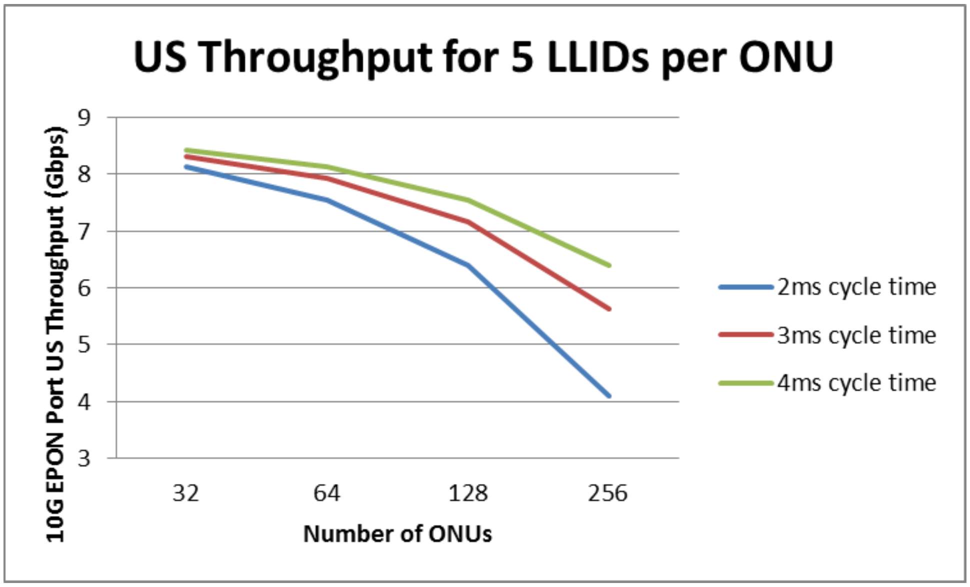


Figure 5 - 10G EPON Port Upstream Throughput per Number of ONUs and Cycle Time for 5 LLIDs per ONU Based on EPON Overhead Computations

Figure 5 shows analysis of US throughput per port based on the network overhead, cycle time, and number of ONUs for 5 LLIDs per ONU. Note that for multiple ONU and LLID cases the difference of maximum achievable throughput per ONU link buffer size for CAA and DAA architecture is not significant as the single flow case due to the dominant burst and scheduling overhead of multiple LLIDs. Most service providers aim the same number of ONUs for the PON Extender and Remote-OLT architectures, e.g. 64 or 128 ONUs per 10G EPON port. Higher numbers, e.g. 256 ONUs per port may be mostly applicable to MDU cases, but would be deployed for relatively lower average traffic growth and average billboard bandwidth in the US direction. Figure 6 shows analysis of DS throughput per port based on the network overhead, cycle time, and number of ONUs for 5 LLIDs per ONU.

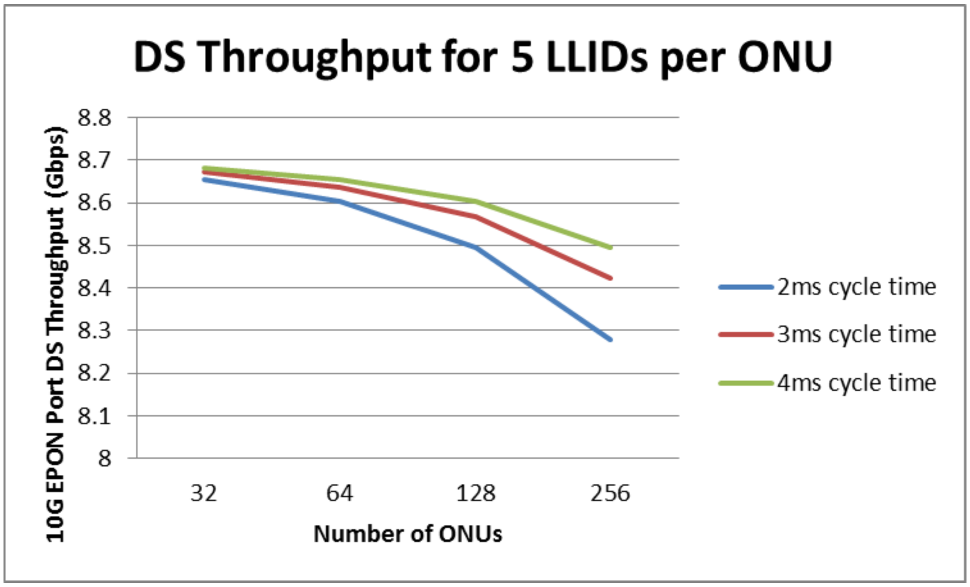


Figure 6 - 10G EPON Port Downstream Throughput per Number of ONUs and Cycle Time for 5 LLIDs per ONU Based on EPON Overhead Computations

These values are used to have a coarse throughput benchmark for the two EPON architectures and to compare IPTV traffic and scaling requirements with the available access network achievable throughput values. EPON access networks with the DPoE support have different MAC, traffic management, and forwarding features that affect the throughput, latency, jitter, availability, and reliability of IP video services. These will be discussed after first providing an overview of the current and future IPTV architectures and the end-to-end video system over EPON access networks.

3. IPTV Networks over EPON

Cable operators' traditional video architecture consists of purpose built hardware products with integrated control and management software designed per the serving groups' physical locations. The video core network includes encoders, a Video-on-Demand (VOD) library, and conditional access servers communicating with the edge network over a video transport network. The edge network consists of VOD and ad insertion servers, switched digital video, and Aloha interactive network with edge Quadrature Amplitude Modulation (QAM) to transmit UDP based MPEG video to the subscriber homes over traditional Hybrid Fiber-Coax (HFC) networks. The home architecture is based on vendor specific cable modems, video gateways, and set-top boxes designed for TV screens.

Increasing competition from telco, satellite operators, and OTT providers, new consumer electronics and novel video technologies with enhanced quality offer new options for the cable subscribers. End-users want to reach any content (managed and unmanaged video from different content providers), anytime (live, on demand, recorded, offline, and online), anywhere (at home and away from home), and from any device (multiscreen, multi format, and different QoS levels). This in turn, brings new requirements for cable operators such as offering new services and flexible service plans, providing higher capacity and elastic resources, and adopting new business models. In the meantime, cable operators need to reduce OPEX (operational expenditure) and CAPEX (capital expenditure) and decrease time to market services to create new revenue areas. Figure 7 displays an example of future end-to-end IP video architecture with EPON access network where both services and networks support SDN, NFV, and cloud applications.

IP video over EPON systems consist of VOD, linear/live, and network digital video recorder/cloud DVR (nDVR/cDVR) services. VOD and nDVR services are unicast services. Future video architectures converge to Hypertext Transfer Protocol (HTTP)-Adaptive Bit Rate (ABR)/IP networks [2,6]. In this architecture HTTP/ABR based unicast and multicast video is delivered over an IP network. Multicast ABR (M-ABR) [3,4] and Smart ABR (S-ABR) [5] provide an adaptive delivery system managed and controlled by the network and service components.

In this example architecture, the home network consists of an IP home gateway with integrated EPON ONU and an embedded multicast client that communicates with the multicast controller in the M-ABR system. IP set-top boxes, smart TVs, tablets, mobile phones, and game consoles are video clients in the home while virtual Customer Premise Equipments (CPEs) and virtual subscriber gateways may be deployed for abstracted service chain applications (including parental control, firewall, and mobility). Although not shown in the diagram, Wi-Fi networks and services may be also deployed with SDN wireless controller, virtual services, and network components to support both home and away from home Wi-Fi access. Some of the video functionality may still need to be embedded in hardware as Physical Network Function (PNF) due to latency and processing performance requirements. Other video functions may be deployed as VNFs.

The access network may consist of R-OLT or a headend OLT (with or without a PON Extender) as described in the previous section. In this architecture, a purpose built router or a vRouter may be the first hop router to the OLT and may be configured and managed by a router SDN controller. A R-OLT SDN controller manages and configures the remote access device. DPoE System and vCM SDN controllers manage and configure the OLT DPoE and vCM for ONUs that can communicate with the R-OLT/OLT over DPoE extended OAM (eOAM). R-OLT/OLT may be connected to the routers/switches, virtual functions, and controllers over a physical leaf-spine fabric network and overlay network, which are controlled and managed by an SDN controller and orchestrated by a resource orchestrator. Resource and service orchestrators may span in and over hubs/headends/DCs/clouds for coordinating VNFs, PNFs, other elements, servers, and service chains based on operator directives from the service and network applications, and OSS/BSS back office managing policies, system topology, and subscriber database, as well as other business models.

Both network and service statistics are connected at different layers for analytics, business intelligence, accounting, proactive resource management, troubleshooting, security, and service assurance. This enables end-to-end network and service visibility, monitoring, and control for enhanced system performance.

VOD services may be delivered over purpose built physical elements and virtual functions. Transcoded and packaged VOD streams are delivered over origin, intermediate, edge caches, and IP Content Distribution Networks (CDNs) to the home gateway over virtual router (vRouter) and OLT/R-OLT as a unicast service flow. nDVR streams are delivered as unicast similar to VOD. Smart ABR systems enhance the OTT performance by incorporating network control for the bit rate adaptation [5]. CDNs may be virtual elements for an elastic and adaptive delivery network.

Linear video may be delivered using M-ABR architecture with NACK Oriented Reliable Multicast (NORM) as the multicast transport protocol [3]. NORM provides flexible FEC and data repair mechanism. IP gateways deliver the stream as unicast in the home, directly from the cache or by requesting a unicast HTTP stream from the edge cache/CDN that may in turn get the stream from a higher layer cache/CDN. Depending on the MSO architecture, a live session may be always multicast or only a set of sessions may be multicast (depending on the viewer numbers and analysis based on the corresponding network gains). This is controlled by the Multicast (M-ABR) controller that informs an IP gateway which in turn joins the multicast session over the EPON network and receives the stream from the Multicast (M-ABR) server. Until the gateway receives the multicast stream, the session is delivered as an unicast service flow.

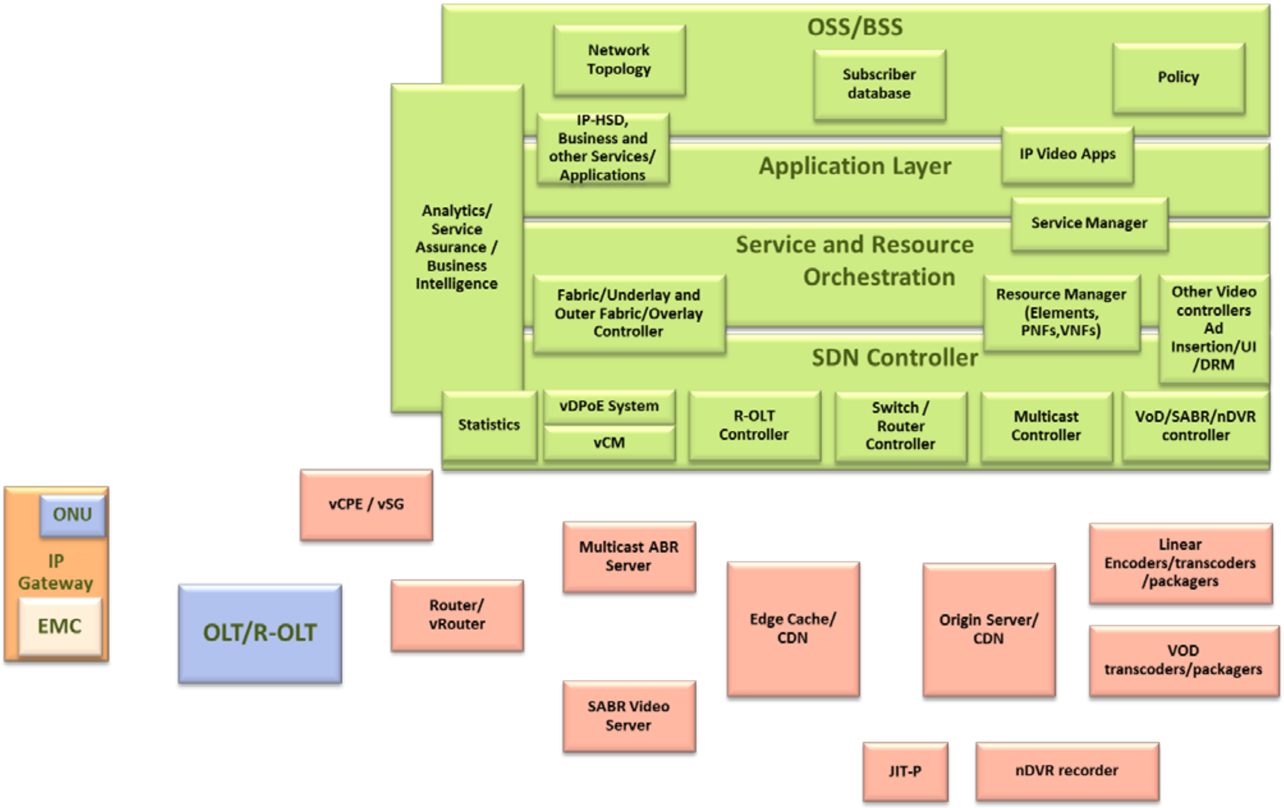


Figure 7 - SDN/NFV/Cloud Enabled End-to-End IP Video Architecture with EPON Access Network: Green blocks represent control and management components while pink blocks represent data plane elements.

3.1. IP Video Scaling Considerations

In this section, we assume a scenario for IP video over EPON based on an estimated subscriber serving area and IP video viewership numbers (as shown in Table 1) to provide guidelines on the scaling considerations.

Table 1 - Example IPTV Scaling Scenario

Parameters	Values per port	Values per R-OLT (4 ports)	Values per line card (16 ports)	Values per chassis (6 line cards)
HHP per serving area	64	64	64	64
Subscriber penetration	100%	85%	75%	65%
IP Video services take rate	100%	100%	100%	100%
Avg # of Potential viewers per home	2.63	2.63	2.63	2.63
% IP VOD/nDVR viewers	25%	%20	%16	12%
IP VOD/nDVR Mbps	6	6	6	6
% IP Linear Viewers	75%	80%	84%	88%
MC weighted Bit rate, Mbps	6	6	6	6
Number of available programs	500	500	500	500

For a deployment with 64 subscribers per port and 100% penetration level, we can estimate 168 viewers per 10G EPON port. As discussed in [4], we can assume close to 90% of all potential viewers are prime time viewers and linear program viewers are around 75% of all viewers at prime time. Per these assumptions, 38 VOD/nDVR and 113 IP linear video out of 151 peak viewers' programs will be delivered in a PON port. We assume this trend will hold for upcoming years even though new business models may bring more customized services. For example, per stream ad insertions may be applied at the home gateway. Therefore, we assume these numbers will change only by 1% for the upcoming years.

The current video programs consist of mostly Standard Definition (SD) and High Definition (HD) options while Ultra High Definition (UHD) programs are now available. Current average video rate may be assumed as 4-6 Mbps. 4k/8k UHD video enhancements increase the average bit rate while new efficient codecs like HEVC will reduce it. We assume although AVC will be used to explore UHD services, HEVC and other new codecs will dominate in the future [10]. We may expect a slight increase in the average bit rate in the upcoming years with more UHD video being available.

If multicast is not used, all IP video will be transmitted as unicast. If the cable operator's policy is based on viewership or program popularity, the number of unique multicast programs to a port will define the gain. A Zipf distribution may be assumed to find unique programs watched by the active viewers given a maximum number of available linear programs [4]. The results depend on the program popularity curves per ad-zone/serving area. Per the graphs in [4], we can derive that at peak times 43 unique multicast streams will be transmitted to 113 viewers per port at average out of 500 available programs. Figure 8[2] provides the number of unique multicast streams for 113 viewers. As seen in Table 2, 486 Mbps total IP video is expected at peak times when multicast is used as opposed to 906 Mbps when multicast is not deployed, providing 50% reduction on the port bandwidth consumption. The bandwidth efficiency increases at the aggregation points as shown in Table 2.

---

[2] This chart was developed using the ARRIS Multicast model. Early work using this model was described in [4].

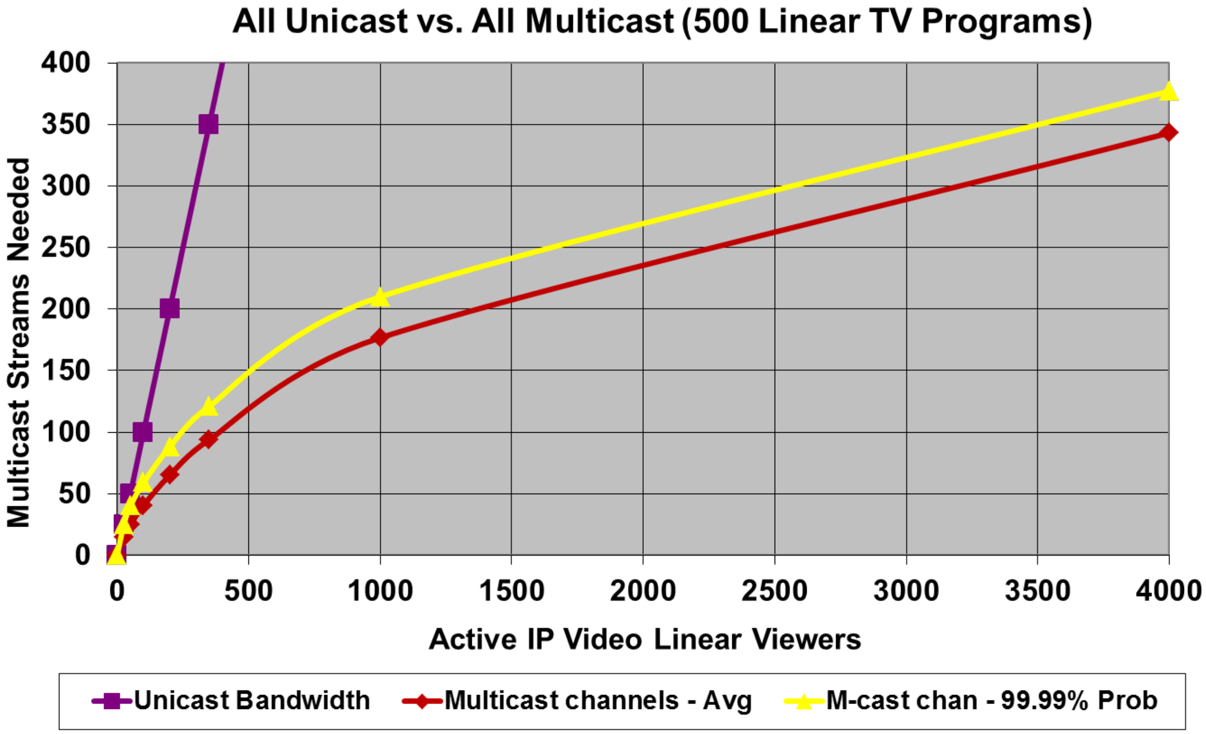


Figure 8 - Unique Multicast Streams Needed Per Active IP Video Linear Viewers

Compared to 10 Gbps EPON port bandwidth, these numbers may be supported by both CAA and DAA EPON architectures. Note that long trunk link does not introduce any extra delay in the downstream transmission (impact on the TCP delay is negligible compared to overall end-to-end delay).

The total efficiency is based on the total bandwidth consumption from the CDN/cache to the home gateway. From the CDN/cache to the multicast server the linear program will be unicast. The efficiency may be increased by having the edge cache closer to the serving area for popular programs. Furthermore virtual CDNs may be moved based on the demand. From the multicast server to the vRouter/router (and over the fabric network), the multicast delivery will improve the efficiency. Table 2 shows the gain per R-OLT with 4 EPON ports, per a 16 EPON port line card for a headend OLT and per an integrated router or vRouter serving 96 EPON ports as examples of aggregation points. Figure 8 is used to estimate the multicast programs per number of active linear video viewers in a multicast zone to compute the bandwidth consumed from the backend to the corresponding multicast zone. For example, for a R-OLT with 4 ports, multicast streams needed for R-OLT viewers are derived from Figure 8 and bandwidth required to send the channel sessions to the R-OLT is computed. For the chassis level, two multicast zones are assumed.

Table 2 - Example IPTV Scaling Scenario Results

Bandwidth Consumption	Values per port	Values per R-OLT (4 ports)	Values per line card (16 ports)	Values per chassis (6 line cards)
All Unicast Bandwidth (Mbps)	906	2736	8448	37440
Unicast + Multicast Bandwidth (Mbps)	486	1134	2424	7848

The main tradeoff is between the bandwidth consumption and control plane complexity. The vRouter/router or other MLSR entities and R-OLT may process the Internet Group Management Protocol (IGMP)/ Multicast Listener Discovery (MLD) membership requests and keep track of membership changes. In the M-ABR architecture, since a video stream in the multicast list is transmitted as unicast until the multicast join is processed, the unicast bandwidth consumption will temporarily increase especially during half-hours depending on the control processing rate. The QoE of each end user will not be affected as long as bandwidth requirements are met since channel change rates will be based on unicast delivery. Other interactions between the gateway to the multicast controller and among controllers must be also taken into account in the overall bandwidth and delay requirements. In addition, the scaling for statistics collection and end-to-end analysis in a SDN/NFV architecture will require real-time big data analysis [8].

It is important to analyze both the data and control plane resource consumption. In SDN and NFV enabled EPON networks, the processing requirements will depend based on distributed functionalities. We will provide two examples in “Use Case Analysis” Section below to discuss the data, management and control plane processing for M-ABR architectures. Based on the estimated number of subscribers and programs per port, the physical and virtual elements may be designed to support a certain number of ports or R-OLTs based on the atomic units they serve.

3.2. IP Video Performance Considerations

IP video Quality of Experience (QoE) assesses the end user’s level of satisfaction for the video services and products. QoE is a subjective measure that may differ from subscriber to subscriber even for the same service and network conditions. 5 point scale Mean Opinion Score (MOS) is the arithmetic mean of all the individual scores that depends on the human perception, expectations and past experience to name a few.

Quality of Service (QoS) is an objective measure of the service by quantitative assessment of service and network conditions such as error rates, throughput, frame delay and jitter, availability, reliability and service processing delay (guidance on these values may be found at standards such as CableLabs PKT-SP-CODEC-MEDIA-C01-140314, ITU G.1080). Mapping QoS to QoE has been an active research area as new multimedia services emerge and requires a multidisciplinary approach including engineering, cognitive neuroscience, economics, statistics and social psychology [9]. Extensive research has been done on video streaming to show that the multimedia device, resolution, bit rate, codec type, buffer size, video content all affect the quality. In addition to QoE, network utilization and fairness are parameters for measuring system performance. MSOs may use different techniques for QoE monitoring including automatic algorithms to detect network conditions and analyzing customer feedback via data mining algorithms and correlating both.

IP video services require end-to-end performance management. Traditional cable systems had well defined boundaries of core, headend/hub, edge/access, and home networks serving managed services with legacy business models. Therefore, a service's data, control, and management plane had a known starting and ending system entity and a common path for all the planes. The definition of "end-to-end" is changing with SDN/NFV systems since data and control planes are separated and system entities are abstracted and located not by physical directives, but by architecture, resource, and service requirements. The end goal is to have an end-to-end system management and control between applications. In this new architecture, service assurance and resource management functions may change and adapt resource usage and service paths. Although agile operations and adaptive resource management are aimed to reduce traditional systems limitations and complexity, migration strategies will define how to achieve this end goal while overcoming integration, interoperability issues, and complexity of a new disaggregated system. For instance, availability and reliability of a system must not be dependent on the availability and reliability of individual disaggregated functions. Resource and service orchestration is therefore crucial in these architectures.

As analyzed in [2,4,5,6], HTTP/ABR based video delivery over all-IP networks with SDN/NFV and cloud application support will enhance the network and resource utilization, end-user QoE, and create new business models for the cable operators. Adaptive bit rate architecture is improved with smart ABR and multicast ABR with NORM (NACK oriented reliable multicast- RFC 5740) functionalities for unicast and multicast video respectively. IP based CDNs and on-demand cache functionality improves QoE by reducing network delay and jitter while increasing resource utilization. Service and network controllers may interact with each other to improve system performance. For example, multicast controllers can communicate with OLT controllers and Wi-Fi controllers while resource and service orchestration coordinates the underlay, overlay networks and both physical and virtual elements. New overlay techniques and distribution of multicast functionality may also improve multicast control plane performance.

In the EPON access networks, unicast video flows may be classified based on upstream and downstream classifiers to apply corresponding QoS operations. QoS support for multicast IPTV flows is implemented per DPoEv2 specifications. Downstream multicast QoS is based on the Group Configuration (GC) and Group QoS Configuration (GQC) with corresponding Service Class (SC) parameters. A single session for GQC instantiating a Group Service Flow (GSF) is defined by a unique combination of Source, Group (S,G) or (*,G) IP addresses matching a GC entry. Admission control may be supported over EPON channels as well as over fabric and virtual networks between video applications. EPON systems support forwarding based on DOCSIS QoS parameters both in the upstream and downstream directions using priority queuing (such as weighted round robin/weighted fair queuing (WRR/WFQ) shapers). Upstream links carry mostly video signaling, however with the proliferation of social media and applications such as citizen journalism IP video is expected to increase in the US as well. Hierarchical priority levels may be applied per class and per flow. DPoEv2 specifications define Best Effort (BE) and non-Real Time Polling (nRTP) scheduling options. The schedulers are in general implemented to support service flow SLAs through configurable minimum reserved rate, maximum sustained rate and maximum burst parameters and weights. A minimum reserved rate may be set based on video (SD/HD/UHD) rate characteristics and the number of streams per ONU. In addition to video data, video signaling (like unicast requests and IGMP/MLD membership messages as well as queries) may be prioritized to reduce the delay. Fair scheduling and Weighted Random Early Detection (WRED) or active queue management with flow control ensure SLA based sharing of resources during congestion.

Since controllers (e.g. multicast, OLT/DPoE System/vCM, and Wi-Fi controllers) may interact and a service and resource orchestration may coordinate them, once SDN/NFV networks are defined and deployed, new end-to-end QoS functionalities may be adopted such as end-to-end admission control and path adaptation, congestion control, and QoS value mapping.

In summary, IP video performance is measured by end user QoE, resource assignment, utilization, operational complexity and cost. All these performance criteria must be monitored and analyzed in an end-to-end fashion. Although new architectures enable end-to-end visibility and control, the amount and variety of real-time and offline data generated from the ABR based video eco-system needs a big data and real-time analytics system [8]. Such an analytics system will facilitate to have an elastic resource management, agile operations, and increased QoE and also enable automatic integration of new services.

4. A Use Case Analysis: IPTV Multicast over EPON Systems

In this section, we will provide two example EPON network architectures to discuss distribution of video and access network functionalities for M-ABR use case: a CAA EPON architecture with PON Extender and a DAA EPON architecture with Remote OLT.

4.1. IPTV Multicast over CAA PON Architecture with PON Extender

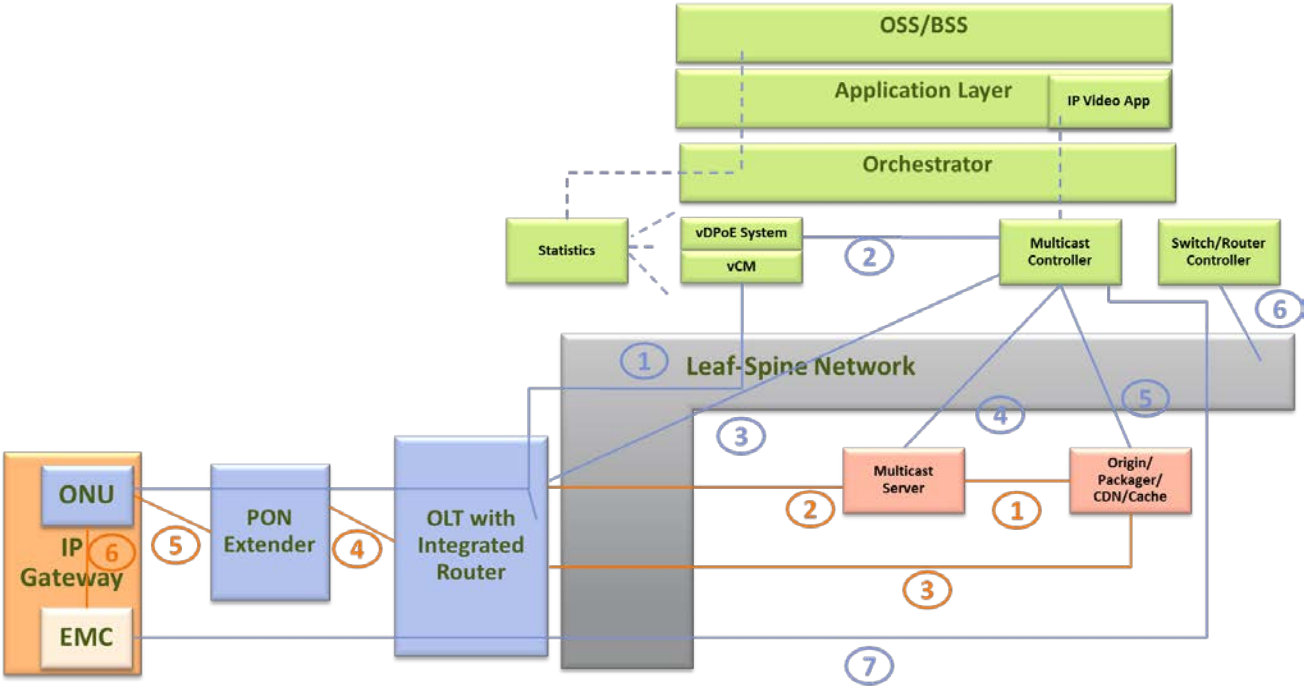


Figure 9 - Simplified M-ABR Architecture Over A CAA PON Access Network with a PON Extender

Figure 9 shows a more simplified architecture example with multicast ABR [3,4, and 6] delivery for linear video over a CAA PON architecture with a PON extender. The blue lines represent the control/management plane while the orange lines represent the data plane. Cable operators may first choose to separate the OLT/DPoE system and vCM functionality from a purpose built OLT with integrated router per criteria discussed in [1]. The video functions such as multicast server and CDN may be purpose built appliances or virtual functions. Such a system may help the cable operator to apply SDN to the EPON access network while operating the access data plane as today's equivalent HFC network. This evolutionary step may help to deploy EPON access networks faster without intelligent active components in the field while a step is taken for software defined back-office management/control.

When a home CPE requests a linear program, if it is cached in the home gateway, the gateway delivers the program as unicast. Otherwise, the gateway sends a unicast HTTP request to get the program that is delivered as unicast ABR through cdn-gw interface[3] (data interfaces 3,4,5, and 6) between CDN/cache and gateway. Multicast controllers may communicate with the packager to determine the content and bit rates for multicast (control interface 5). As described in [3], some operators have their multicast servers pull directly from their packager. Just-in-time packagers may be located both centrally in the architecture and at the CDN edges.

In this example, the unicast request may be sent from the ONU as an upstream unicast EPON service flow assigned for video signaling. The downstream unicast video may be sent from the OLT as a unicast EPON service flow assigned for all unicast video. Both service flows may be assigned corresponding DOCSIS QoS parameters as discussed in Section 3.2.

The cable operator's multicast content selection algorithm may be based on viewership (e.g. multicast when there are more than one concurrent viewer) or on a policy (e.g. multicast n most popular programs). This selection algorithm, along with other policies, may be directed by the IP video application to the multicast controller using a northbound interface (bssi interface) of the SDN controller. Service and resource orchestration may assign video functions that will serve a certain serving area. The servers and elements may be connected over a leaf-spine fabric network. An overlay tunnel may be defined for the data delivery between the CDN and the OLT router (data interfaces 1 and 2). A controller may manage the fabric forwarding (control interface 6).

Per its communication with the multicast controller, the Embedded Multicast Client (EMC) in the gateway learns the multicast channel map via mc-emc interface (control interface 7). If it is a new program, the multicast controller (MC) sends the Uniform Resource Identifier (URI) to the multicast server via mc-ms interface (control interface 4). The multicast server (MS) gets video segments from the CDN/edge cache via cdn-ms interface (data interface 1) acting as an HTTP client and sends multicast video via NORM to the gateway proxy cache over ms-emc interface (data interfaces 2, 4, 5, and 6).

The EMC sends an IGMP/MLD join to the ONU (data interface 6). If a PON extender is used, the ONU is connected to the PON extender over EPON wavelengths (data interface 5). The PON extender communicates with the OLT over DWDM/CWDM wavelengths (data interface 4). PON extender is a simple O-E-O device and may be seen as a PHY device. The OLT may apply admission control, join authorization, and access control lists to accept or deny the join. If the join request is accepted, the OLT adds the CPE and ONU to the multicast table. If it is a new multicast stream, the OLT will request the stream from the routers on path to the multicast server (e.g. using Protocol Independent Multicast – Sparse Mode (PIM-SM)). As previously explained, the DPoE system adopts the IP multicast model defined in DOCSIS MULPIv3.0 specifications. The replication of IP multicast traffic onto mLLIDs is done per downstream channel in the MAC domain over mulpi interface (data interfaces 4 and 5). Each GSF has unique multicast Logical Link Identifier (mLLID) within the MAC domain while multiple IP multicast sessions may have the same GSF and mLLID. The service class name for the corresponding GQC may use DOCSIS QoS parameters. The multicast controller may have interface to the OLT or

---

[3] Interfaces are explained for an EPON access network based on the definitions in [3] described for an HFC network.

vDPoE System/vCM to control multicast QoS or other functions using mc-ccap interface (control interfaces 2 and 3). The OLT/vCM communicates mLLID and filtering and forwarding rules to the ONU via eOAM (control interface 1 where eOAM may be tunneled to the OLT). The OLT also supports multicast query. Per the DpoEv2 specifications, IP serving groups may be defined with control GSF to deliver general queries.

The EMC sends an IGMP/MLD leave request to the OLT when the CPE leaves the program or when the session is not in the multicast program list anymore. Note that, The OLT will remove the mLLID/GSF if this is the last leave in the MAC domain and will send a leave request to the router towards multicast server if this is the last leave in the OLT. Accordingly, the multicast server will cease to receive the multicast session when the program is not in the multicast list.

The OLT and vDPoE system/vCM may report viewership and network related statistics to the EMS/NMS via ossi interface. These reports along with other information from the multicast server and controller may be used for orchestration of resources and services as well.

Southbound controller interfaces may be based on legacy protocols such as SNMP/CLI and/or may use new protocols and data models such as Network Configuration Protocol (Netconf)/YANG models, OpenFlow interfaces. Northbound controller interfaces may use similar interfaces and/or HTTP based Application Program Interfaces (APIs) to the application layer.

4.2. IPTV Multicast over DAA PON Architecture with Remote OLT

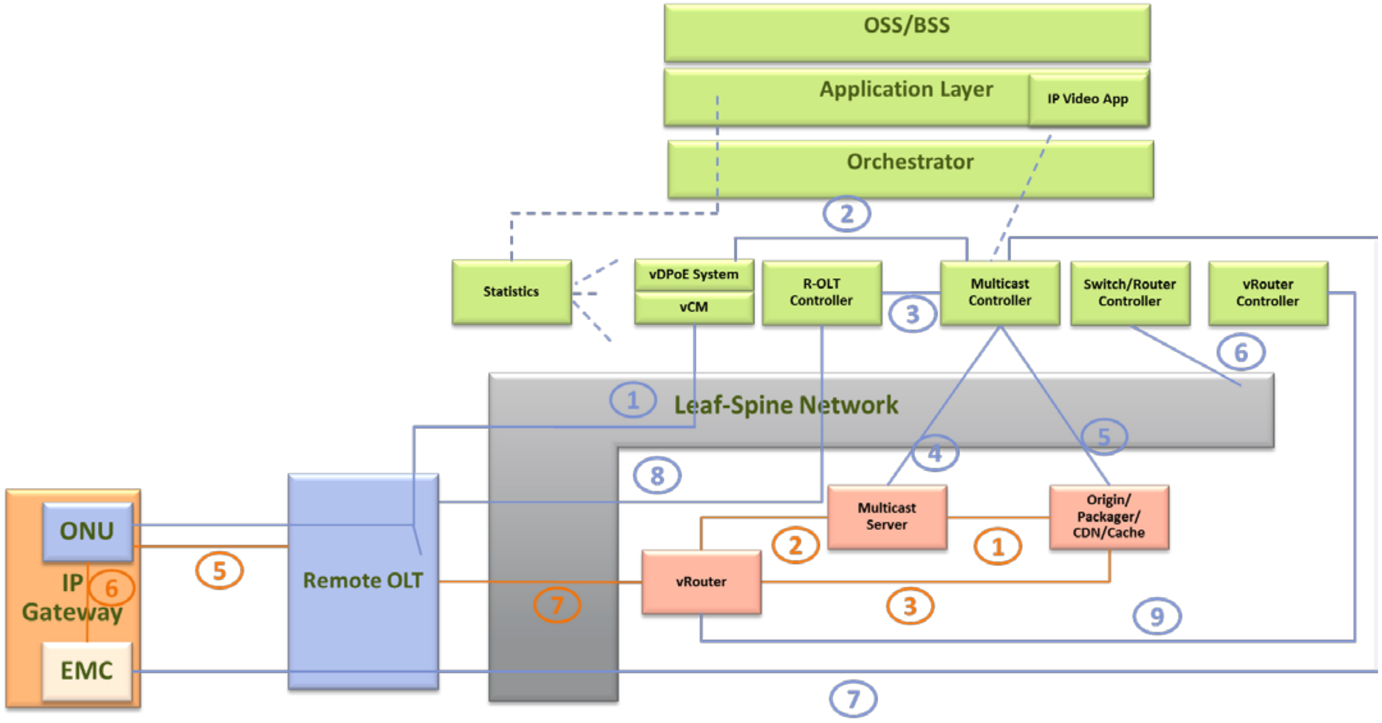


Figure 10 - Simplified M-ABR Architecture over EPON R-OLT Access Network

Figure 10 shows another example with multicast ABR delivery for linear video over DAA PON architecture with remote OLT. Some cable operators may opt to deploy a remote OLT connected to a virtual router per criteria discussed in [1]. In this case, R-OLT may be a Layer 2 device connected to a virtual router (data interface 7) that forwards unicast video delivered from CDN/edge cache (data interface 3) or multicast video delivered from multicast server (data interface 2). A vRouter controller configures and manages vRouter virtual function (control interface 9). A R-OLT controller configures and manages the R-OLT for initialization and run-time non-DPoE operations (control interface 8). Depending on the functionality distribution, the multicast controller may communicate with R-OLT controller and vDPoE System/vCM (control interfaces 3,2) to control multicast QoS or other functions.

The data and control plane flows are similar to the architecture described in the previous section. Since the router is not integrated in the R-OLT, IGMP/MLD membership and replication functionalities may be distributed among these components based on physical and virtual element capabilities and resource utilization efficiency. For example, R-OLT can process the dynamic multicast membership, and request the session from vRouter only when this is the first request in the R-OLT using IGMP/MLD proxy functionality.

5. Conclusions

Cable operators have been considering new architecture options in access networks, video services, and back office systems. EPON is one of the preferred access network technologies to accelerate the migration to an SDN/NFV enabled IP video architecture while providing a future-proof platform. Both centralized and distributed 10G EPON access network options may support new IP video architectures. Depending on the cable operator’s current and future requirements, headend or remote OLT architectures may be integrated with IP video systems under an umbrella SDN/NFV architecture. This requires cross-product groups (e.g. video delivery and access network groups) to cooperate more closely to reuse common platforms and interfaces, maximize the interaction among network and service elements, and enable an integrated service assurance and resource management. Therefore migration strategies will require not only adopting new network technologies but also new partnerships, processes, and integration methodologies.

Although access agnostic systems may be future targets for cable operators, the migration strategy will include end-to-end management and control with backward compatibility. Some integration functionality may be first deployed with vendor specific elements and interfaces for the existing components while standards and open architecture based elements and interfaces may be defined for new components.

We provided some example architectures and defined functionalities for an EPON based IP video architecture to investigate migration options and provide insights on IPTV multicast design. Having the control plane separated from access and video delivery elements and virtualizing data plane will enable service and resource orchestration to optimize the system performance and to create, modify, and delete service chains. Although silos are broken in such an end-to-end IP video architecture, a more dynamic system that co-exists with legacy networks will increase the complexity. Vendors that have extensive experience on existing operations may help to find solutions to decrease this complexity during the migration.

6. Abbreviations and Definitions

6.1. Abbreviations

ABR	adaptive bit rate
ASF	aggregated service flow
BSS	business support system
CAA	centralized access architecture
CALEA	communications assistance for law enforcement act
CDN	content distribution network
CIR	committed information rate
CLI	command-line interface
CoS	class of service
CPE	customer premise equipment
CWDM	coarse wavelength division multiplexing
DAA	distributed access architecture
DBA	dynamic bandwidth allocation
DC	data center
DHCP	dynamic host configuration protocol
DOCSIS	Data-Over-Cable Service Interface Specification
DPoE	DOCSIS provisioning of EPON
DS	Downstream
DSCP	differentiated services code point
DVR	digital video recorder
DWDM	dense wavelength division multiplexing
EMC	embedded multicast client
EMS	element management system
eOAM	extended operations, administration, and maintenance
EPON	Ethernet passive optical network
FCAPS	fault, configuration, accounting, performance, security
GC	group configuration
GQC	group QoS configuration
GSF	group service flow
HFC	hybrid fiber-coax
HD	high definition
HTTP	hypertext transfer protocol
Hz	Hertz
IGMP	Internet group management protocol
IPDR	Internet protocol detail record
IPTV	Internet protocol television
ISBE	international society of broadband experts
LLID	logical link identifier
LI	lawful intercept
M-ABR	multicast adaptive bit rate
MAC	media access control
MC	multicast controller

MDU	multi dwelling unit
mLLID	multicast logical link identifier
MLD	multicast listener discovery
MLSR	multi-layer switch router
MPCP	multi-point control protocol
MS	multicast server
nDVR	network digital video recorder
NFV	network function virtualization
NMS	network management system
NNI	network to network interface
NORM	NACK-oriented reliable multicast
OAMP	operations, administration, maintenance, and provisioning
ODN	optical distribution network
OLT	optical line terminal
ONU	optical network unit
OSS	operations support system
OTT	over the top
PHY	Physical
PNF	physical network function
PON	passive optical network
R-OLT	Remote optical line terminal
QoE	quality of experience
QoS	quality of service
S-ABR	smart adaptive bit rate
SCTE	Society of Cable Telecommunications Engineers
SD	standard definition
SDN	software-defined networking
SF	service flow
SNMP	simple network management protocol
TM	traffic management
TPID	tag protocol identifier
ToS	type of service
UHD	ultra high definition
US	Upstream
vCM	virtual cable modem
VNF	virtual network function
VLAN	virtual local area network
VOD	video on demand
vRouter	virtual router
WFQ	weighted fair queueing
WRR	weighted round robin
WRED	weighted random early detection

6.2. Definitions

10G EPON	EPON as defined in [802.3ah], now part of [802.3].

Adaptive Bit Rate	A streaming video technique where Players select between multiple bit rate encodings of the same video stream.
Access Network	The PON network between the Gateway/ONU and the OLT.
Content Distribution Network	A network designed to minimizing latency by distributing network objects onto geographically diverse servers
Downstream	Information flowing from the hub to the user
Embedded Multicast Client	The function embedded in the Gateway which joins multicast groups and receives multicast content.
EPON Operations and Maintenance Messaging	EPON OAM messaging as defined in [802.3] and [DPoE-OAMv2.0]
Gateway	A customer premises device which facilitates delivery of video, data and other services
Headend	The central location on the cable network that is responsible for injecting broadcast video and other signals in the downstream direction
IP Multicast	A delivery mechanism whereby IP packets can be transmitted to/received from devices that have explicitly joined a multicast group.
Linear TV	A continuous content stream from a provider, e.g., a broadcast television network.
Multicast Controller	A device which controls what channels are provided via multicast.
Multicast Server	A device which delivers content via multicast
Orchestrator	Coordinates one or more controllers to provide end-to-end service.
SDN Controller	Device manager that implements some or all of the device control plane as well as manage device configuration.
Service flow	A unidirectional flow of packets from the upper layer service entity to the RF/Optical PHY with pre-defined QoS traffic parameters
Upstream	Information flowing from the user to the hub
YANG	YANG is a data modeling language used to model configuration and state data manipulated by the Network Configuration Protocol (NETCONF).

7. Bibliography and References

[1] *A Comparison of Centralized vs. Distributed Access Architectures for PON*, Mike Emmendorfer and Sebnem Zorlu Ozer, INTX 2016, Spring Tech Forum

[2] *An Architectural Approach towards Achieving Bandwidth Efficient Content Delivery*, Dinkar Bhat, Navneeth Kannan, and Wendell Sun, INTX 2016, Spring Tech Forum

[3] CableLabs IP Multicast Adaptive Bit Rate Architecture Technical Report, OC-TR-IP-MULTI-ARCH-V01-141112

[4] *Challenges Delivering Multiscreen Linear TV Services: Multicast-Assisted ABR to the Rescue*, John Ulm, ARRIS White Paper, 2014

[5] *Smart ABR: The Future Of Managed Ip Video Services*, John Ulm, Ajay Luthra, Praveen Moorthy, Mark Schmidt, Haifeng Xu, NCTA 2013, Spring Tech Forum

[6] *ABR Delivery Architectures and Virtualization*, Santhana Chari and Niranjan Samant, ARRIS White Paper, 2014.

[7] CableLabs SDN Architecture for Cable Access Networks Technical Report, VNE-TR-SDN-ARCH-V01-150625

[8] *Effective Utilization of M-ABR (Multicast Assisted ABR) using big data and Real-Time Analytics*, Sridhar Kunisetty, Jeffrey Tyre, and Robert Myers, INTX 2016, Spring Tech Forum

[9] From OBI and SNR to Ookla and Netflix: How Network Impairments affect Customer Perceptions: The role of Leading and Lagging indicators as we evolve HFC to FTTP, Sebnem Zorlu Ozer, Amarildo Vieira, John Chrostowski, Venk Mutalik, INTX 2015, Spring Tech Forum

[10] *Ultra HD – Connecting the Present to the Future*, Sean McCarthy, ARRIS White Paper, 2014

Acknowledgments

The authors would like to thank the following individuals: Nagesh Nandiraju (Comcast), Jean-Christophe Marion (Broadcom) and John Ulm, Dan Torbet, Janet Bean, Carl Daun, Dwain Frieh, Bill Hanks, Dave Hartrum (ARRIS)

JOURNAL OF
DIGITAL VIDEO

SCTE ISBE

Video Quality STB Evaluation through the Application of a MOS Scoring Framework

A Technical Paper prepared for SCTE/ISBE by

Joseph Frank, Technical Manager/Video Product Engineering, Comcast Cable,
SCTE/ISBE Member
1701 JFK Blvd
Philadelphia, PA 19103
Joseph_Frank2@cable.comcast.com
215-286-2269

JOURNAL OF
DIGITAL VIDEO

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

JOURNAL OF
DIGITAL VIDEO

1. Abstract

Video quality is a key differentiator of video encoder and decoder devices. Video quality evaluations usually require hours of viewing various video test streams and test patterns per a designated device. These evaluations are performed by skilled Golden Eye video experts. The results are highly subjective and biased to the particular Golden Eye's perspective. This paper describes the integration of Mean Opinion Score (MOS) techniques to augment the video quality evaluations of settop box (STB) devices. A framework for these evaluations is discussed along with enhancements to improve the overall video quality process. The process is explained through the evaluation of a WI-FI STB. This framework integrates any MOS device or devices of choice providing a method of objective automation which improves both the efficiency and consistency of the evaluations.

2. Introduction

Automated methods to evaluate visual quality of digital video as perceived by the human eye have gained importance in the industry due to the large number of digital video applications consumed by viewers. Digital video is often edited and passes through several decompression and compression stages before it reaches the viewer. These processing stages usually degrade the video quality. Furthermore, the application of Internet Protocol (IP) video along with WI-FI distribution can impose additional degradations and artifacts. Quality of service (QoS) and quality of experience (QoE) play a key role in determining the overall quality of video. Together, they measure the quality of the service delivery and the service's overall user experience. The term QoS with regards to this paper should not be confused with Class of Service that assigns a given priority to a service within a network. QoS in this context measures the transmission of bits from the source to the destination. QoE quantifies the user experience as consumers view video and listen to audio. For the purposes of this paper, QoE will focus on video viewing [1].

QoS metrics are a fundamental prelude to QoE metrics. With regards to QoS, packet loss, buffer overflow, and buffer underflow produce macroblocking, blackouts, and frozen screens. Such artifacts result in significant impact to the viewer experience. QoS artifacts are usually the result of network (NW) transmission issues such as dropped packets. An important point to note is that video decoder devices vary in their response to defective content. Some decoders successfully employ error correction to minimize transport errors. Similarly, encoders take advantage of the MPEG group of pictures (GOP) structure and sometimes minimize transmission artifacts to a single frame interval.

Video frames are partitioned into GOP's where certain frames contain most of the video detail and serve as reference frames ("I" and "P" frames). The subsequent frames and slices use the reference frames to create "B" frames based on the video changes or deltas. If a reference "I" frame is dropped, it has a greater impact on the viewer's QoE than if a "B" frame is dropped. Dropped "I" or "P" frames can produce artifacts for many frame intervals and can cause packet buffer overflow or buffer underflow conditions which certainly have significant impact on the user experience. A loss of an "I" frame causes all frames until the next "I" frame to be lost. A loss of a "P" frame affects the quality of all frames from the current "P" to the next "I" frame. A loss of a "B" frame affects only the current "B" frame.

In addition to network transmissions errors, many artifacts can be caused by the video encoding or transcoding process. Typically, content is received from the programmer and re-encoded and then groomed during ad insertion by the service provider before broad distribution. Macroblocking and video smearing, for example, can be caused by substandard encoding or transcoding. These artifacts result due to high quantization and deficient motion estimation. QoE tools can inspect the integrity of the MPEG structure and determine the severity of potential video lossiness. During the video transmission, several indicators can be analyzed for QoE such as available bandwidth, bitrate, quantization, video resolution, and GOP structure [1].

While all the bits and bytes can be correct according to the MPEG syntax, the resulting video can still be poor regardless of how well it is transcoded due to the poor quality of the original content. To truly measure video quality, more is required rather than just performing MPEG analysis. As a result, researchers have developed perceptual quality metrics that mathematically model human vision to determine the video quality. Perceptual quality analysis (PQA) instills objective measurement of video quality when compared to the subjective nature of Golden Eye viewing. It utilizes low pass filtering to simulate human vision and is performed on decompressed video pixels as observed by a normal viewer. It is applied after the last stage of the video pipeline following all video post-processing (i.e. scaling, de-interlacing, and graphics). Intermediate variables are sometimes calculated and combined into a relative score which represents the human viewing experience.

3. Video Quality Parameters and Measurement

Golden Eye experts have defined a number of Video Quality parameters to subjectively grade colorfulness, blur, contouring, frozen video, noise, jerkiness, black screens, blockiness, scene complexity, and motion complexity [2]. Many of these parameters are not coupled to any particular video output resolution or input video encode style. A few of these parameters, however, are more useful with regards to evaluating advanced video features. Such metrics are based upon real world experiments where trained video personal have watched numerous videos, scored the quality, and rated the associated artifacts. Various types of video content along with test patterns are studied and viewed for many iterations. Artifacts, for example, are recorded for blur, frozen video, jerkiness, black screens, and blockiness. These parameters are then compiled into a final grade. Similarly, PQA tools can mathematically compute a subset of these parameters and derive a mean opinion score (MOS). The MOS provides a single value which assesses video on a scale (usually 1 to 5). Some PQA tools represent the scores from 0 to 500 or even 0 to 100. The following list of video quality parameters and definitions are itemized below. They have been assembled from several PQA tools.

- ***Spatial Complexity*** is derived from the differences between the edge energy in a video frame. Spatial energy increases proportionally to scene detail. Very low edge energy indicates blurring or smearing, while extremely high edge energy indicates spatial edge noise, block distortion, tiling, or noise. This parameter is particularly useful when evaluating the intricacies of the content regardless of resolution or encode classification (i.e. MPEG2, H.264, or HEVC). An example of spatial complexity is depicted in Figure 1.

Figure 1 - Sample Pictures with low/high Spatial Complexity

- ***Temporal Complexity*** measures the motion energy between video frames. Video with high motion activity (e.g. sports content) or a large number of scene splices have greater temporal complexity. Video with little motion activity (e.g. news, static images) have low temporal complexity [2]. Similar to spatial complexity, this parameter is also valuable when assessing the content. It pertains to all encode types and all video output resolutions from standard through ultra-high definition (UHD).
- ***Blockiness*** is a defect where the edges of blocks or rows of blocks are seen as a checkerboard-like or grid-like pattern. This is the most common artifact which is viewed, and it applies to all encode types and video resolutions. It is most noticeable during action scenes with complicated movements. It also occurs for low bitrate content. Blockiness can result from transmission errors, decoder or encoder errors. An example of blockiness is illustrated in Figure 2.

Figure 2 - Sample Pictures With High/Low Blockiness

- ***Jerkiness*** is a defect where either video stops suddenly and resumes, or motion does not transition smoothly. Inadequate buffer size or poor motion estimation are often the culprits to this artifact. It affects all video resolutions and encode types.
- ***Frozen Video*** is distinguished from jerkiness by repetition of a full frame of video over a given time interval, approximately 5 to 15 seconds. This artifact normally occurs during severe transmission errors which are too difficult to conceal. This artifact occurs across all resolutions and encode types.
- ***Blackout*** is full frames of black video occurring over a time interval of approximately 5 to 15 sec. Blackouts occur naturally in content during scene splices. However, they can also be introduced by video decoder error concealment due to transmission loss, ad insertion, buffer underrun, and buffer overrun conditions.

- ***Blurriness*** is the loss of fine detail and the smearing of edges in the video. It usually arises due to low pass filtering of high frequency content prior to or during the encode process. Blurriness is common when viewing interpolated standard definition (SD) video scaled to an HD television. Similar to blockiness, blurriness applies to all encode types and video output resolutions. An example of blockiness is illustrated in Figure 3.

Figure 3 - Sample Pictures With Low/High Blur

- ***Mosquito Noise*** appears as random speckles on an otherwise smooth surface. High frequency distortions occur in the form of spurious pixels. Mosquito noise is a lossy compression artifact which occurs due to high quantization regardless of video output resolution or encode type. An example of noise is illustrated in Figure 4.

Figure 4 - Sample Picture With Mosquito Noise

- ***Haloing or Ringing*** is typically seen around areas of high contrast such as sharp lines, text edges, and graphics. Sometimes the graphic appears to extend into the background. Haloing can manifest itself as smaller details in graphics appearing to soften the edges. Like mosquito noise, it is a byproduct of the compression process. An example of haloing is illustrated in Figure 5.

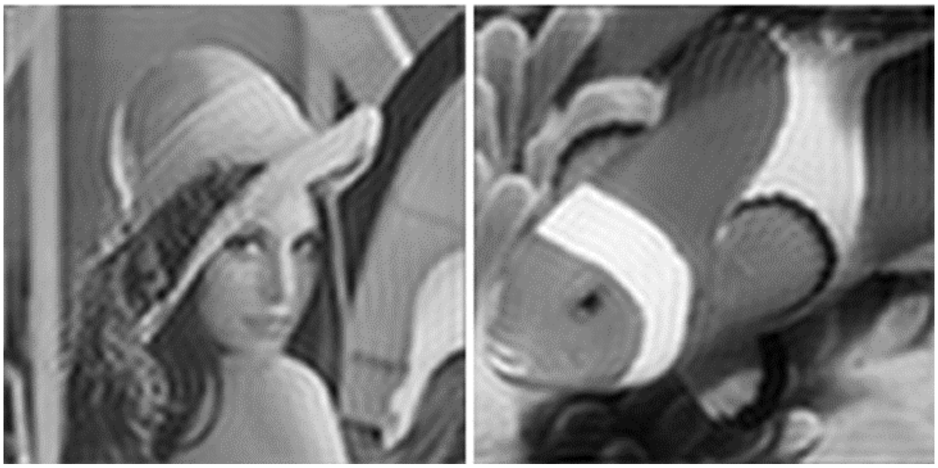


Figure 5 - Sample Pictures With Haloing

- ***Contouring or Banding*** is a defect where abrupt changes between shades of the same color create color bands instead of a gradual change. It results from an implementation of low precision in the encode/ decode process. This can be seen in scenes with high degrees of smooth color with a gradual change in contrast such as sunsets, dawns, or clear blue skies. Contouring is most noticeable for high dynamic range (HDR) content where 10 to 12 pixel values are essential to the video quality. By increasing the pixel precision in both the encode and decode processes, this artifact can be easily managed. An example of contouring is illustrated in Figure 6.

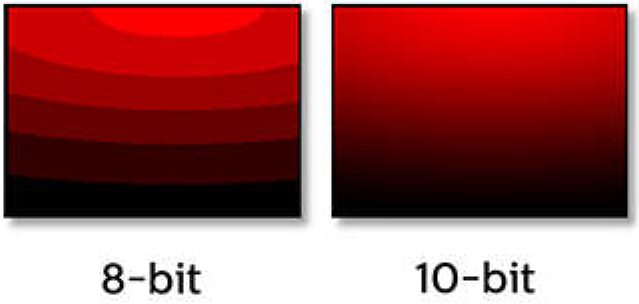


Figure 6 - Sample Pictures With High/ Low Contouring

- ***Pumping*** or ***Breathing*** is a defect where the video or portions of the video appear to pulse at a regular frequency. This is typically seen in large areas of smooth neutral colors. In many scenarios it occurs during the "I" or IDR frames of large GOP's. Subsequent "P" and "B" frames tend to degrade in quality due to the compression process. Like many of the other parameters, this artifact occurs regardless of video output resolution or encode type. When the next "I" or IDR frame occurs at the beginning of the next GOP, a sudden pulse in quality occurs.
- ***Colorfulness*** measures the level of color intensity of the video. It quantifies the range of color and determines if color saturation has occurred. This metric is a valuable tool for assessing wide color gamut (WCG) especially as video transitions from BT.709 to BT.2020.

PQA tools use various subsets of the parameters described above which are pertinent to their algorithms for calculating MOS scores. These tools apply human vision model calculations to interpret the interaction between various artifacts. The parameters or artifacts first are prioritized. The higher prioritized artifacts usually mask those of lower priority. The time interval of each parameter is also taken into consideration. A higher scored artifact over a shorter interval can be commensurate to a lower scored artifact over a longer time interval. Two major categories of PQA tools exist. They will be discussed briefly for background purposes only. The first is categorized as Non-referenced. Most Non-reference based tools incorporate the psychophysical features of the human vison model as parameters described above which perform a bottom up approach to measuring video quality. Non-referenced MOS tools evaluate video without any knowledge of the pre-encoded video or any of the network statistics. They operate on only decoded video pixels with no relevance to the input video.

In contrast, Referenced and Reduced Referenced based MOS tools rely on designated video content usually of pristine quality. Reduced Reference relies on minimal reference to the input video source content whereas Reference fully relies on the input video source content [3]. The designated referenced content can be a high quality raw source video or lightly compressed (high bitrate) MPEG content (ProRes). The referenced based tools most often measure the video quality using either difference mean opinion score (DMOS) or just noticeable differences (JND) algorithms where the output video of a device is compared to its input video. For a successful comparison, the input and output video must be both temporally and spatially aligned. DMOS incorporates a structural similarity model (SSIM) algorithm. SSIM is based on a top down approach where structural information of video scenes from the input and output videos are isolated and compared for structural similarity [4].

SSIM, for example, focuses on luminance differences regarding brightness and contrast. It also performs correlation between the input and output video luminance patterns. Higher degrees of structural similarity provide a good indication of perceived video quality. JND, on the other hand, relies on comparisons between pairs of images within the input and output videos respectively. These image comparisons are quantified into perceptual differences with regards to luma and chroma values. These differences are examined as impairments which have been introduced into the output video. The impairments ultimately produce lower video quality scores as perceived by viewers. In order for both of these algorithms to be successful, prior knowledge of the video content before any subsequent processing is required [5].

4. Video Quality Framework

Non-referenced and Referenced MOS tools each have their advantages and disadvantages. Non-referenced tools usually utilize well understood psychophysical parameters such as blockiness, blur, and noise. Non-referenced tools are also easier to automate. The main disadvantage is that psychophysical parameters are highly non-linear, and the video images are often quite complex. The parameters are usually implemented in a somewhat linear manner since they are based on spatial edge energy and temporal energy between pixels. As a result, the measurements can be inconsistent over multiple executions. The advantage of Referenced tools is that the intricacies of the input video are very well understood. Mathematical calculations are more deterministic and consistent by nature. The major disadvantage of Referenced base tools, however, are that they are more difficult to automate. Referenced based tools are primarily designed for file based analysis. Capture of pixels from an HDMI output and integration to the file based analyzer can be challenging. Although a framework can be implemented to use either Referenced or Non-referenced based tools, this paper proposes a framework which integrates some of the concepts of a Referenced based MOS tool and applies them to a Non-referenced MOS tool in an automated manner.

The overall objective of a Video Quality Framework (VQF) is to augment and reduce the level of Golden Eye visual analysis. Golden eye analysis is very time consuming and subjective in nature. It can be very inconsistent at times due to human error. The integration of MOS tools into the video analysis process impose a level of objectivity to the effort. The VQF must provide a consistent user interface and be tool agnostic as MOS tools evolve in the future. The VQF cannot be tightly coupled to any one tool and must be easy to adapt in the future. It must incorporate automation as the backbone of its utilization in order to minimize the time and level of manual effort. Automation becomes essential to cost reduction of the QoE validation process. The VQF must instill thresholds to establish pass/ fail criteria for validation. Analysis must be persistent across many STB's. Results from different dates and different devices must be easy to access and compare. To achieve adequate comparisons, the results between various tools and devices must be coherent. Finally, the VQF must account for user and tool error conditions. A VQF architecture is illustrated in Figure 7 below.

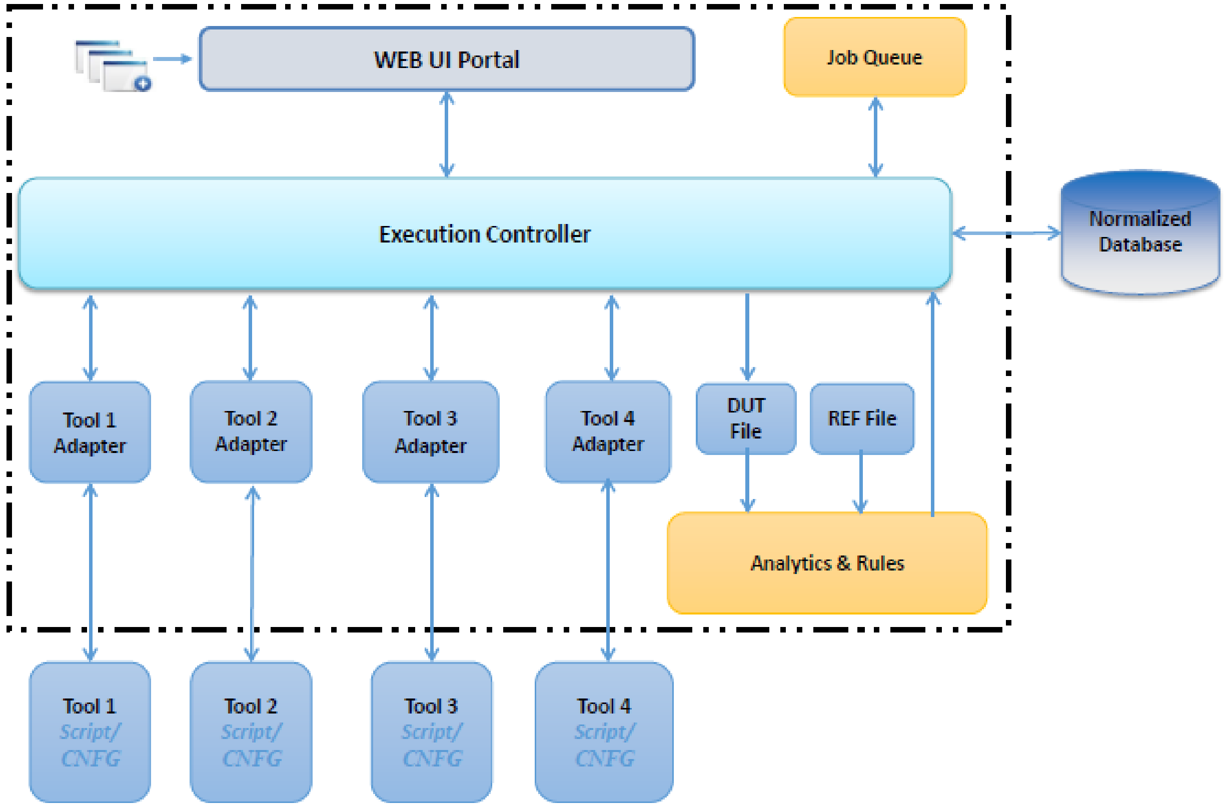


Figure 7 - VQF Firmware Architecture

The VQF FW architecture partitions into the following four modular layers: Presentation, Database (DB), Service, and Control. The Presentation layer consists of a User Interface (UI). The Presentation Layer hosts a web portal where the user can initiate video quality job requests and graphically view execution reports. The user can select to queue jobs or view archived results. While queueing a job, the user enters additional information which is stored in the DB such as STB type, video content criteria, and tool specific parameters. In addition to storing and maintaining job related information, the DB layer also stores PQA parameters such as MOS, Jerkiness, Blockiness, and Blurriness. The DB offers persistence which is vital for archiving purposes. In addition to the Presentation and DB layers, the Service layer provides individual adaptors to communicate with designated MOS tools.

An adaptor acts as an interface to the tool. It synchronizes the tool and works closely with the tool's internal scripts and configuration files to evaluate the selected content. It monitors the tool's execution state and collects all PQA results by job completion. The adaptors simplify the insertion of new MOS tools and the removal of obsolete MOS tools. The Control layer acts the glue between the other layers and manages an execution state machine for the VQF. It interfaces with the Presentation and DB layers to enqueue jobs. It interacts with the Service layer to dequeue jobs, and collect the results. It incorporates an analytics and rules engine to process the results, and eventually archive them into the DB for future viewing as graphical plots via the Presentation layer.

5. Video Signatures

The VQF provides a mechanism to employ automated MOS scoring. The integration of Non-reference PQA tools maintains an overlap and comfort level with Golden Eye analysis since it uses many of the same psychophysical parameters to describe processed artifacts which are normally observed at the device output video. To create a reproducible environment which evaluates the video quality of various STB devices on a continual basis, it is imperative to characterize the input video similar to how it is applied by Referenced based MOS tools. The application of known video content establishes meaningful comparisons of different video decoder devices. While performing MOS analysis, it is also effective to incorporate video genres from various categories such as shopping, news, action films, high spliced, and sports. Finally, in order to further guarantee repeatability and automation, the sampling interval for the collection of PQA data should be determined and fixed.

PQA data can be collected in intervals based on number of frames, number of seconds, or number of minutes. A single frame interval is effective for evaluation of encoder devices when determining the allocation of bits per a frame. For STB evaluations, GOP or fragment intervals provide relevant data with regards to both commercial significance and analysis of the user experience. By establishing a given PQA tool, designated input stream, fixed sampling interval, and selected output resolution, the collected PQA data can be combined to create a unique signature for a designated STB video decoder device. This signature is statistical by nature since PQA data is statistical. The signature becomes a reference or finger print for future comparisons of devices under test. The signature incorporates a similar concept to the reference used by Reference MOS tools. This capability is built into the VQF analytics engine where data collected from a Device under Test (DUT) is compared to the previous baselined signature. As a result, changes to video quality can be objectively quantified. Acceptable thresholds for video quality can also be established. Quality assurance (QA) teams can repeatedly implement this process to validate video quality for either SW or HW upgrades to existing devices and impose a pass/fail criteria.

The VQF offers a structure to automate video quality analysis. The integration of signatures enables repeatability, and the application of thresholds facilitate pass/ fail criteria. However, consistent signature creation and repeatable job executions can be challenging to achieve. First of all, HDMI probes can experience problems synchronizing the output video and capturing data on a consistent basis. The same probe devices used in different environments can also have slightly different calibrations. The sampling interval can also have a significant impact on captured HDMI pixels if the rate is too high. Higher sampling rates create more accurate signatures, but 0.5 and 1 sec sample rates often produce erratic HDMI captures. As a result, the Non-referenced MOS tools will calculate inconsistent results at particular time intervals for certain parameters over multiple executions. The Blockiness parameter, for example, is the most compromised due to its sensitivity to edge gradient deltas. The type of video content can also exasperate the consistency issue especially with regards to high motion sports or action scenes.

Although inconsistency exists with Non-referenced MOS tools due to HDMI capture challenges, the VQF analytic engine can perform operations to improve the signature creation. The analytics engine utilizes the Nyquist sampling rate and appropriately decimates the captured data. The data is decimated according to GOP or IP video fragment time intervals. The analytics engine also correlates results from several iterations and removes outliers between multiple iterations. It assimilates the results of these operations and combines them into a single signature. An example of an analytics engine enhancement process is depicted in Figure 8below.

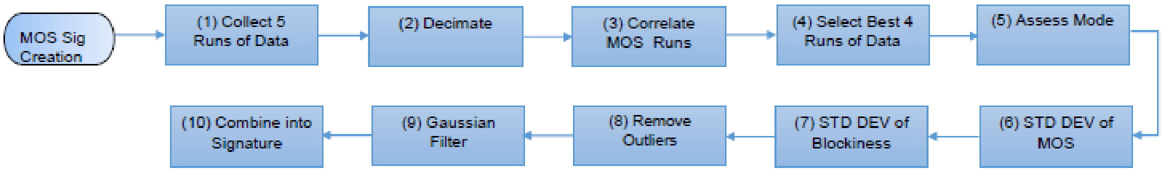


Figure 8 - VQF Analytics Engine Signature Enhancement

The analytics engine steps from Figure 8 are described as follows. This example uses 8 minutes of video content. It collects data on 1 second intervals and decimates it to 2 second samples matching the GOP/ fragment MPEG structure.

1. Collect MOS scores from 5 consecutive iterations using 1 sec samples. Iterations should be from separate probes (same probe type) in different environments using a designated video decoder STB and well known video content.
2. Appropriately low pass filter results from each iteration and decimate to desired interval in seconds. For the 8 minute content, 2 second intervals were chosen.
3. Correlate 5 iterations
4. Select 4 best correlated iterations.
5. Perform Mode computation of 4 iterations and use mode value if the mode count is at least 2.
6. Perform standard deviation of MOS results per a time sample.
7. Perform standard deviation of Blockiness results per a time sample. Blockiness results are the greatest source of MOS inconsistencies.
8. Remove outliers by applying thresholds to MOS and Blockiness standard deviation results.
9. Apply Gaussian filter on intervals where mode values were not selected.
10. Combine filtered results and mode values into a signature.

The process described above is applied to both the baseline and DUT signatures. The baseline signature must be well understood and validated with respect to the input video content. The increases and decreases in the results should be justified. By using a consistent signature process meaningful comparisons can be quantified between the baseline and DUT's.

6. WI-FI STB Evaluation

A credible method for obtaining signatures in relation to designated video content and video decoder devices is a crucial element of the VQF analytics engine. The VQF analytics engine offers flexibility which increases the number of use cases for a VQF automated environment. An example of such an environment is illustrated in Figure 9 below. For this example, two different WI-FI STB's are evaluated and compared for their video quality. The video quality analysis is performed by two similar Non-referenced HDMI probes in separate environments. The assessment is executed on two different STB's simultaneously.

The VQF interfaces to both HDMI probes. It controls the entire evaluation process and extracts the results from each probe. The results are computed by the analytics engine into individual signatures for comparisons. The selected video content for this evaluation is IP and employs Adaptive Bitrate (ABR) which is encoded in H.264 at 3.5 Mb/s, 1.2 Mb/s, 750 Kb/s, and 500 Kb/s. The video consists of shopping, live news network, action movie, and home/ garden segments which are seamlessly spliced together. Each segment is approximately 2 minutes in length. The input resolution of the video content varies according to the ABR, and the output resolution of each STB is assigned to 1080i30. Samples were collected at 1 second intervals and decimated to 2 seconds which represents the video content fragment structure. The WI-FI distribution of video is later degraded by adding iPerf traffic, using an FTP server which is connected to the video gateway via Ethernet.

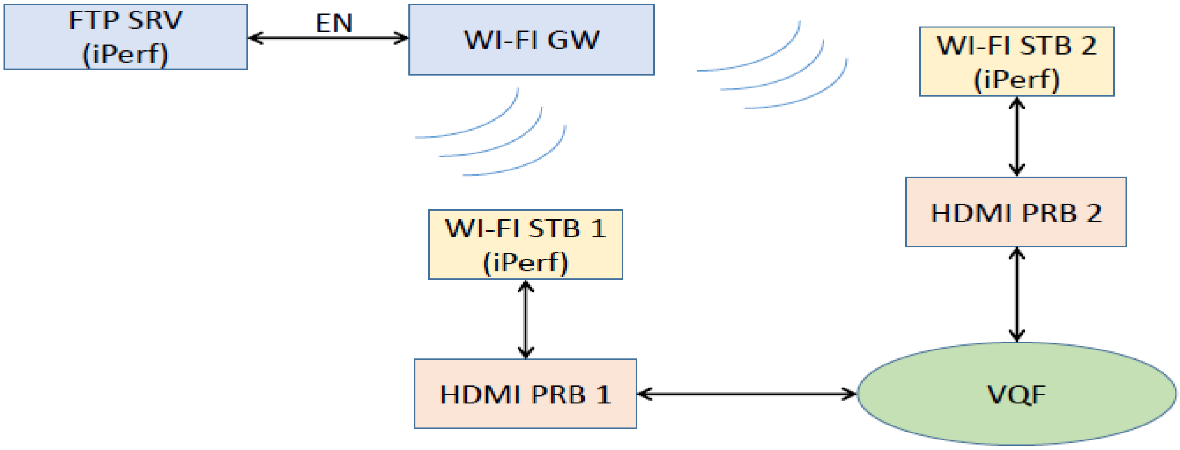
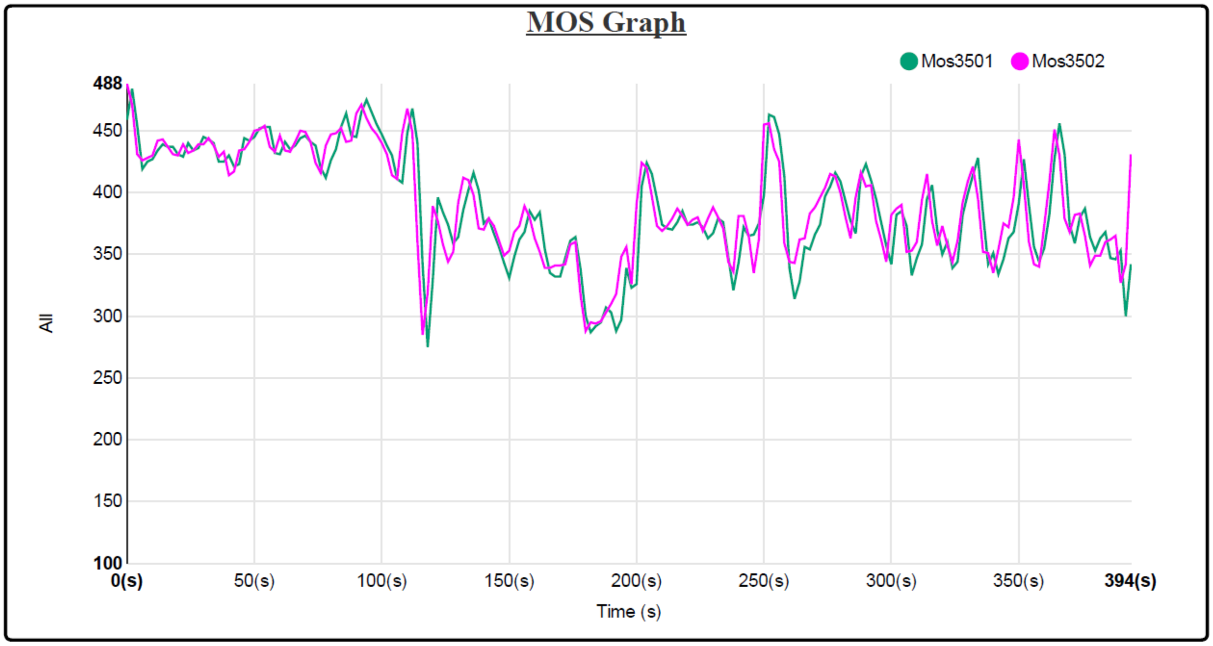


Figure 9 - VQF Evaluation Environment

Using the WI-FI video STB environment above, two separate measurements are performed. The first measurement establishes a baseline where iPerf is adding no traffic data to the WI-FI transmission. The highest ABR input resolution transport stream of 3.5 Mb/s is transmitted. The first STB with an ID of 3501 is located directly underneath the video gateway. The second STB is located 15 ft. to the right of the gateway. A key point to consider is that separate probes with slightly different calibrations are used along with two different STB's which were sampled from different skews. A comparison of the two baselines is depicted in Figure 10 below.



MOS MEAN : STDDEV	MOS CORRELATION	JERK MEAN : STDDEV	BLUR MEAN : STDDEV	BLOCK MEAN : STDDEV
Mos3501 : 388.76 45.21 Mos3502 : 390.77 42.77	Mos3501_3502 : 0.882	Jerk3501 : 0.00 0.00 Jerk3502 : 0.00 0.00	Blur3501 : 16.84 7.95 Blur3502 : 16.49 7.65	Block3501 : 3.04 4.08 Block3502 : 2.93 4.01

Figure 10 - VQF STB Baseline MOS Results

First, a little insight into the relationship between the video content and the baseline signatures needs to be understood. The shopping content occurs over the first 120 samples. A dip occurs where the splice point transitions to the news content. The shopping content has high contrast and small areas of high spatial detail. Motion is basically non-existent so this content is relatively easy to encode and decode, resulting in high MOS scores. The particular segment of news content has many scene splices, intricate detail, and a moving ticker scroll across the bottom. Due to the high spatial and motion complexity, a sharp transition point occurs at sample 120, and the MOS scores are noticeably lower. It is obviously more challenging to encode and decode this content. The next transition point occurs at sample 240 denoting action film content. The motion complexity increases, and the spatial detail decreases. As a result, the splice point is not as sharp and the MOS scores increase a little. Similarly, the splice point is not as sharp around sample 360 where the home and garden content begins. The home and garden content consists of high spatial detail and considerable splice scenes for kitchen content. The MOS scores continue to dip as the scene splices occur in rapid succession.

With regards to this specific content, the signatures are credible. It is observed that the signatures correlate well at 88%. The overall video quality MOS scores are very good with an average in the upper 300’s. Another point to note is that the standard deviations are about 9% of the total scale indicating good overall video quality consistency. No jerkiness exists, and the blockiness is very low which is expected for H.264 encodes which employ a de-blocking algorithm. The impact of the motion during the last three video segments is attributed to blurriness as depicted in Figure 11. Once again, the artifacts between the two STB's correlate well. Differences can be associated to variances in the probe calibration, dithering of the STB's, and orientations of the devices with regards to WI-FI transmission.

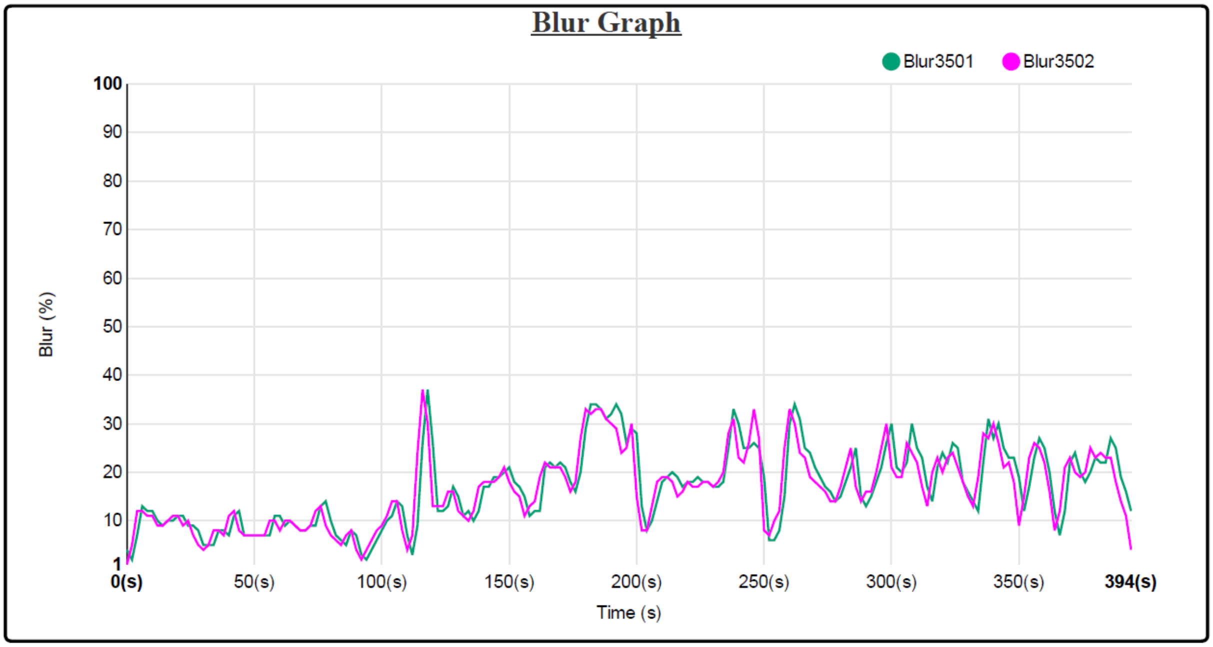
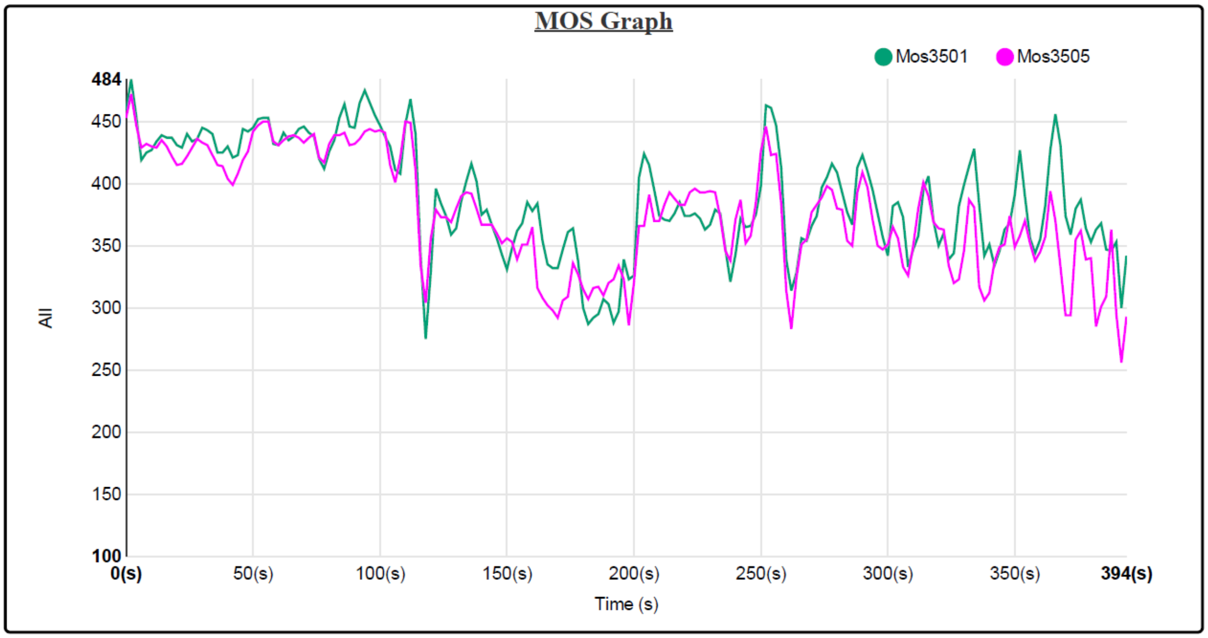
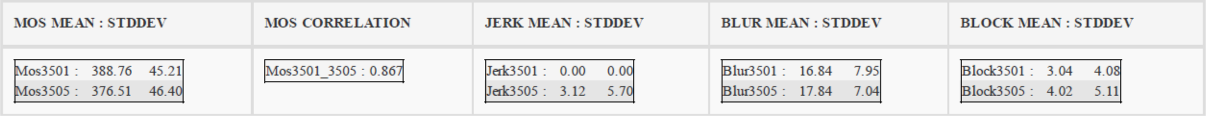


Figure 11 - VQF STB Baseline Blur Results

Once the baselines have been confirmed, the next step is to evaluate the impact of degradations to video due to WI-FI network traffic. To accomplish this, the STB's have been equipped with iPerf. Approximately 141 Mb/s of additional data is transmitted on the same channel as the original H.264 video bitstream. This use case is executed by the VQF and results are collected again. The new DUT signature is generated and compared to the original baseline signature which persists in the DB. The comparison of STB 1 which is located directly below the gateway is illustrated in Figure 12, Figure 13, and Figure 14 below. The original baseline has the ID of 3501, and the DUT has the ID of 3505.

As can be observed from Figure 12, the signatures still correlate well at 86.7%. The video impairments which occur can be categorized as modest. The additional traffic causes a reduction in the ABR input resolution. Differences in video quality start to become obvious with this particular HDMI probe when the MOS decreases by greater than 25. The degradation becomes more perceivable if the decrease continues for more than 2 consecutive samples. The artifacts in the graphs are evident between samples 150 to 200 (during the news segment) and after sample 350 (during the home and garden segment). The shopping and action film segments hold up very well overall. The artifacts can be associated to an increase in jerkiness and blockiness as illustrated in Figure 13 and Figure 14. Jerkiness and blockiness artifacts start to become noticeable when their values experience increases greater than 5 with regards to the respective baselines. It just so happens that these two artifacts have a greater negative affect on the overall user experience especially if they persist over time. The home and garden content is impacted by both of these artifacts whereas the news content is more impacted by jerkiness.





MOS MEAN : STDDEV	MOS CORRELATION	JERK MEAN : STDDEV	BLUR MEAN : STDDEV	BLOCK MEAN : STDDEV
Mos3501 : 388.76 45.21 Mos3505 : 376.51 46.40	Mos3501_3505 : 0.867	Jerk3501 : 0.00 0.00 Jerk3505 : 3.12 5.70	Blur3501 : 16.84 7.95 Blur3505 : 17.84 7.04	Block3501 : 3.04 4.08 Block3505 : 4.02 5.11

Figure 12 - VQF STB 1 Baseline and DUT MOS Comparison

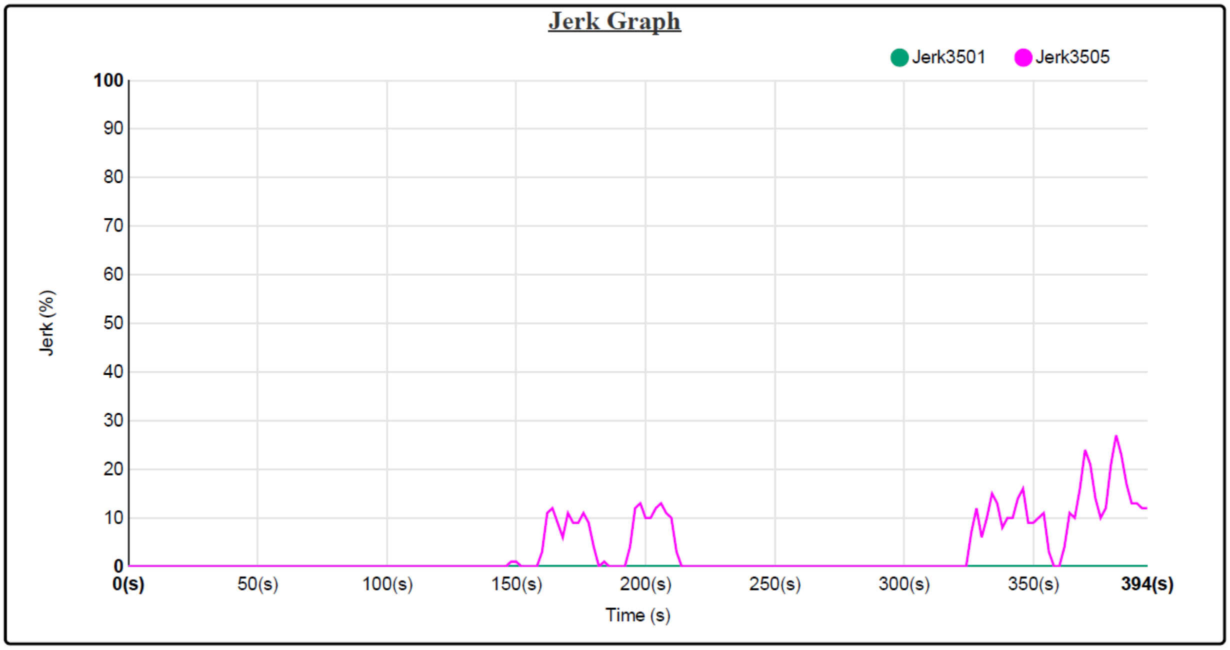


Figure 13 - VQF STB 1 Baseline and DUT Jerkiness Comparison

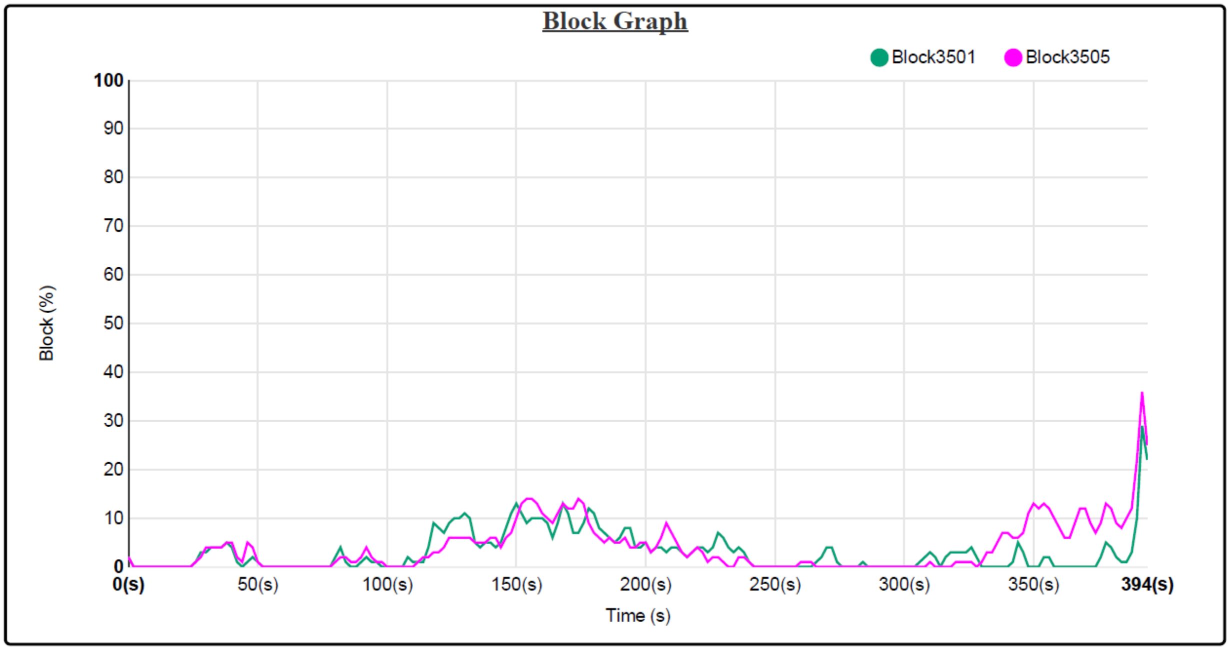
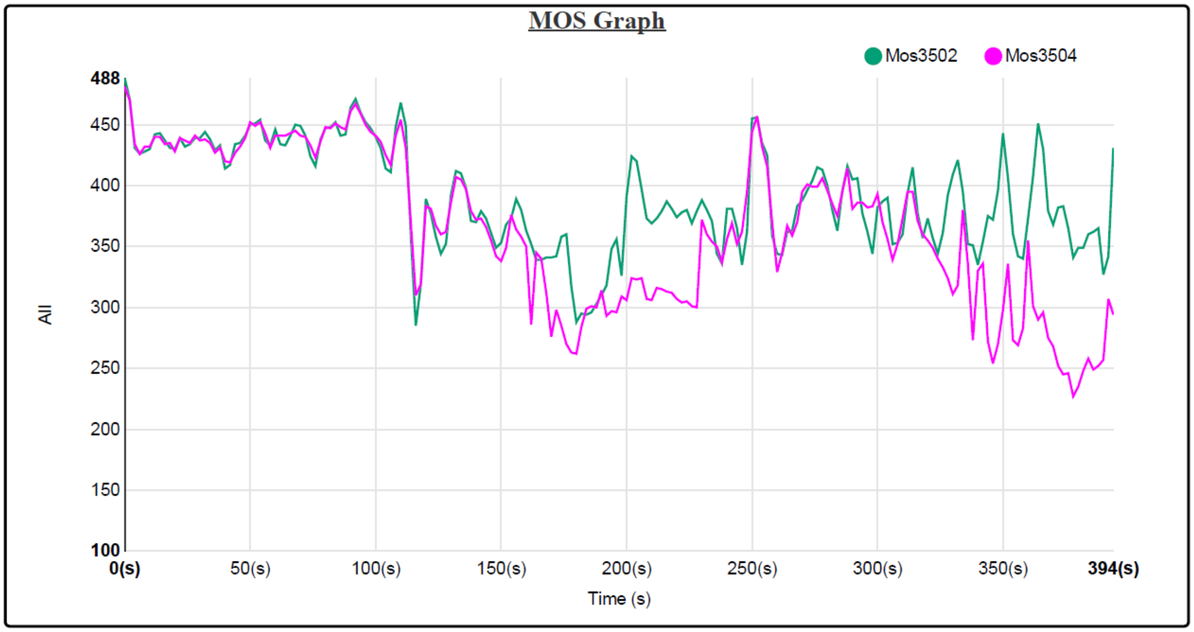


Figure 14 - VQF STB 1 Baseline and DUT Blockiness Comparison

The comparison of STB 2 which is located approximately 15 ft. away from the gateway is depicted in Figure 15, Figure 16, and Figure 17 below. The original baseline has the ID of 3502, and the DUT has the ID of 3504. Once again a decrease in ABR video input can be attributed to the increased WI-FI traffic. As indicated by the statistics, the correlation has a noticeable drop off due to the increased distance from the gateway. The video quality can be categorized as poor in certain segments. The average MOS has decreased by just over 25, and a noticeable increase has occurred in the standard deviation. The jerkiness and blockiness have increased considerably. The overall blur has slightly decreased due to the masking by the blockiness since it has higher priority over blur for this particular HDMI probe. The shopping content holds up well again due to the lack of complicated video characteristics. The action film, however, starts to experience noticeable degradation near its endpoint (between samples 300 and 360). This is not too surprising due to the motion characteristics of this content. Both the news content (between samples 175 to 240) and the home and garden content (after sample 360) also suffer greater degradations. The jerkiness artifacts have great impact as illustrated in Figure 16 during those highlighted sample intervals. Similarly, the blockiness artifacts also cause considerable impact to the user experience as depicted in Figure 17. The distance coupled with the increased WI-FI bandwidth combine to have greater impact on the user experience.



MOS MEAN : STDDEV	MOS CORRELATION	JERK MEAN : STDDEV	BLUR MEAN : STDDEV	BLOCK MEAN : STDDEV
Mos3502 : 390.77 42.77 Mos3504 : 365.28 63.14	Mos3502_3504 : 0.753	Jerk3502 : 0.00 0.00 Jerk3504 : 5.16 10.11	Blur3502 : 16.49 7.65 Blur3504 : 15.44 6.81	Block3502 : 2.93 4.01 Block3504 : 9.22 11.99

Figure 15 - VQF STB 2 Baseline and DUT MOS Comparison

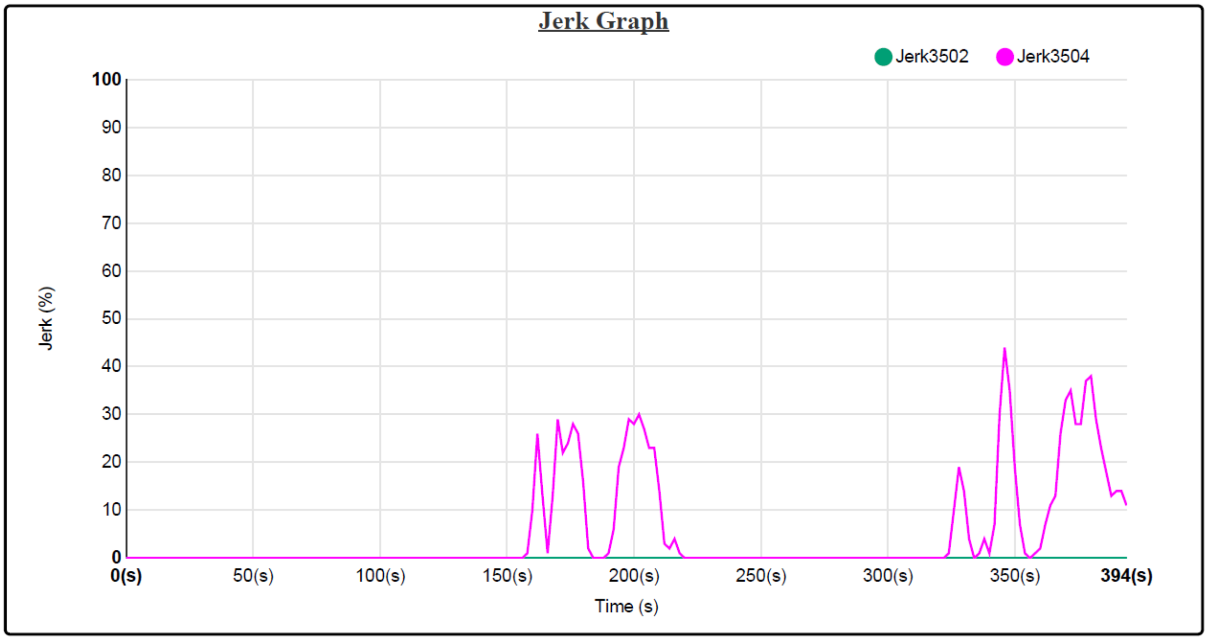


Figure 16 – VQF STB 2 Baseline and DUT Jerkiness Comparison

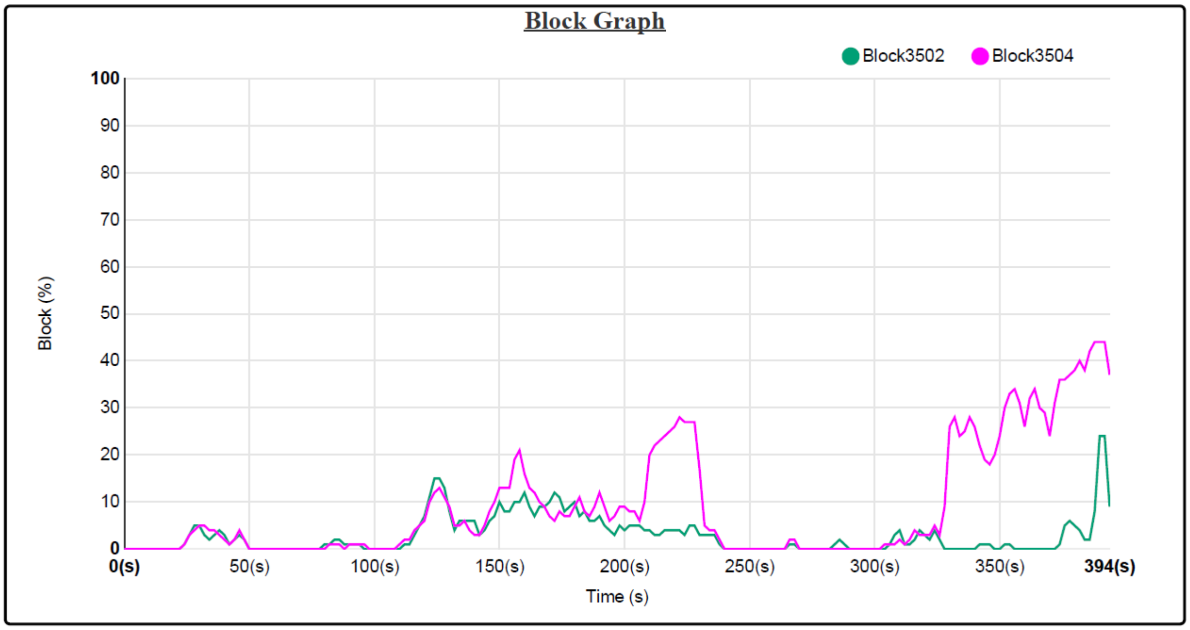
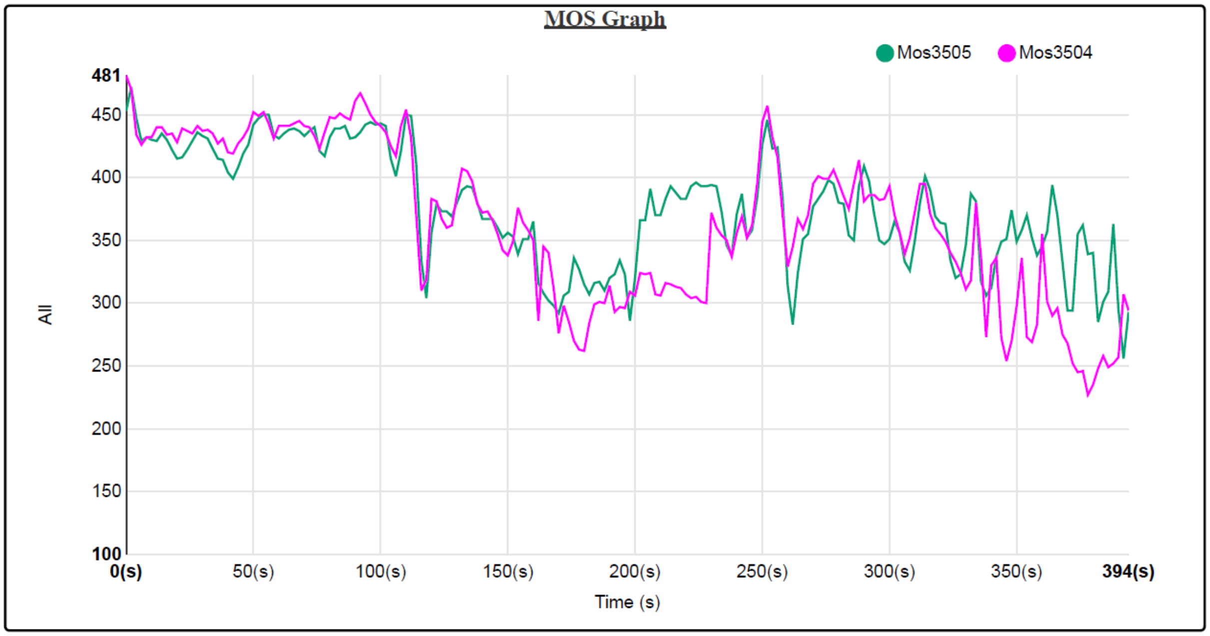


Figure 17 - VQF STB 2 Baseline and DUT Blockiness Comparison

The comparison of the two DUT's is depicted in Figure 18 below. When referring to the baseline comparison in Figure 10, STB2 had a slight advantage in video quality according to the MOS graph. However, when the data traffic was increased by iPerf, the video quality for STB 1 is now better as indicated by the average MOS and standard deviation values. Although the two STB's still correlate well at 82%, the jerkiness artifact is slightly higher and the blockiness artifact is noticeably higher. These artifacts are reflected in the MOS graphs between samples 150 and 240 (news content) and after sample 350 (home and garden). The major difference between the two STB's relates to their orientation and distance from the video gateway device. More use case scenarios can be executed to further characterize the WI-FI device in an ABR environment. In this designated use case, the VQF offers high value when characterizing WI-FI STB video devices. It is particularly powerful when other WI-FI parameters such as WI-FI signal strength, modulation scheme, distance, orientation, and transmitted error packets are characterized in conjunction with these video parameters. This just highlights one example of an application of an automated VQF environment.



MOS MEAN : STDDEV	MOS CORRELATION	JERK MEAN : STDDEV	BLUR MEAN : STDDEV	BLOCK MEAN : STDDEV
Mos3505 : 376.51 46.40 Mos3504 : 365.28 63.14	Mos3505_3504 : 0.821	Jerk3505 : 3.12 5.70 Jerk3504 : 5.16 10.11	Blur3505 : 17.84 7.04 Blur3504 : 15.44 6.81	Block3505 : 4.02 5.11 Block3504 : 9.22 11.99

Figure 18 - VQF STB DUT MOS Comparison

7. Conclusion

Video quality is a major differentiator when characterizing STB devices since it plays a significant role in determining the overall user experience. In coordination with Golden Eye analysis, automated video quality evaluations provide a level of objectivity to the entire process and offer a reduction in both the manual analysis and the cost of validation. A natural transition exists between Golden Eye analysis and many Non-referenced based PQA tools. The Non-referenced PQA tool incorporates many of the same parameters which are diagnosed by Golden Eye experts. These parameters help quantify the impact of the video encode and decode process with regards to the user experience. They correlate the relationship between QoS and QoE and seamlessly incorporate them into the video quality process.

In order to augment Golden Eye analysis, a VQF is proposed which integrates automation into the evaluation methodology. It borrows the concepts of known video input content and improved consistency, which are often associated with Referenced based tools, and applies them to Non-referenced based tools. The VQF offers flexibility to interface both Referenced and Non-referenced PQA tools through its Service layer using various adaptors which control the individual PQA probes. A common user interface is employed regardless of the PQA probe to enqueue jobs and view results. An analytic and rules engine is implemented which assimilates the video parameters. It performs operations on these parameters and generates signatures from designated HDMI probes which can be used concurrently in different environments. The resulting signatures act as a reference for Non-referenced based PQA probes. Because of the addition of MOS PQA tools, the VQF transforms the subjective Golden Eye process into an objective means for characterizing STB's. An example of the characterization for a Wi-Fi STB device is described in the previous section. A few other use cases which have been implemented in the VQF are listed below.

1. Comparisons between competing STB devices.
2. Measure video quality deltas based on established thresholds when new STB software revisions are about to be released.
3. Measure video quality between competing camera products.
4. Establishing encoding parameters for cameras to optimize tradeoffs between storage and video quality.
5. Measure video quality before and after trick play.
6. Compare video quality differences between MPEG2, H.264, and HEVC over various encoding parameters.
7. Comparisons of IP video, linear video, and VOD generated content.
8. UHD/ HDR comparison of decoded output to reference input streams using Referenced based PQA tool

Finally, such a framework is not limited to just video quality. Based on the capabilities of the integrated tools, audio quality, data quality, and voice quality can also be measured. Signatures from the corresponding results can be created, stored in the persistent DB, and viewed for comparison. Thus, the VQF provides an automated solution which incorporates objectivity into the data analysis of many types of customer premise devices.

8. Abbreviations

ABR	adaptive bitrate
DB	database
DMOS	difference mean opinion score
DUT	device under test
GOP	group of pictures
HD	high definition
HDR	high dynamic range
IP	Internet Protocol
JND	just noticeable differences
MOS	mean opinion score
NW	network
PQA	picture quality analysis
QA	quality assurance

QoE	quality of experience
QoS	quality of service
SD	standard definition
SSIM	Structural Similarity Model
STB	settop box
UI	user interface
UHD	ultra-high definition
VQF	video quality framework
WCG	wide color gamut

9. Bibliography and References

[1] IEEE TRANSACTIONS ON IMAGE PROCESSING- 2009: “Study of Subjective and Objective Quality Assessment of Video”, Kalpana Seshadrinathan*, *Member, IEEE*, Rajiv Soundararajan, *Student Member, IEEE*, Alan C. Bovik, *Fellow, IEEE* and Lawrence K. C

[2] “Perceptual Video Metrics, A New vocabulary for QOE”, Cheetah, Nov 14, 2014

[3] “White Paper: Advancing To Multi-Scale SSIM”, Video Clarity

[4] “Understanding MOS, JND, and PSNR”, Video Clarity

[5] “Display Device-Adapted Video Quality-of-Experience Assessment”, Abdul Rehman, Kai Zeng, and Zhou Wang Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Can, Feb 2015

JOURNAL OF
DIGITAL VIDEO

Efficiently Troubleshoot Closed Captions – and Keep the FCC at Bay

A Letter to the Editor prepared for SCTE/ISBE by

Steven R Holmes, Sr Application Engineer, Tektronix
14150 SW Karl Braun Drive
Beaverton, Oregon 97077
steven.r.holmes@tektronix.com
1-800-833-9200

1. Introduction

Closed captions are an important part of TV programming and mandated by FCC rules for such factors as accuracy, synchronicity, completeness and placement. The current CEA-708 standard for closed captions is significantly more complex than the previous 608 captions and much can go wrong. To keep the FCC at bay, it's important for content distributors and video service providers to have tools on hand to look at captions deeply enough to identify the root cause of problems.

2. New Closed Captioning Rules

In 2014 the FCC set new, improved rules for TV closed captioning to ensure that viewers who are deaf and hard of hearing have full access to programming. These rules were intended to solve concerns about captioning quality and provided much-needed guidance to video programming distributors and programmers.

The FCC's closed captioning rules(1,2) apply to all television programming with captions, addressing quality standards for accuracy, synchronicity (timing), program completeness, and placement of closed captions. To summarize, the new rules mandate that captions must be:

– Accurate: Captions must match the spoken words in the dialogue and convey background noises and other sounds to the fullest extent possible.

– Synchronous: Captions must coincide with their corresponding spoken words and sounds to the greatest extent possible and must be displayed on the screen at a speed that can be read by viewers.

– Complete: Captions must run from the beginning to the end of the program to the fullest extent possible.

– Properly placed: Captions should not block other important visual content on the screen, overlap one another, run off the edge of the video screen, or be blocked by other information.

While the new rules apply to how the captions look on the screen, they do nothing to help with troubleshooting captions when there are problems and they start to violate FCC rules. This is where you need a set of tools that lets you look at captions deep enough to analyze the root cause of what is wrong with the captions – and avoid FCC fines.

3. Current Closed Captioning

Today we use CEA-708 captions for any HD program that is transmitted digitally and can be also used for SD programing, although this is not common. The 708 captions use the digital television closed caption or DTVCC channel and a 608 compatibility channel. 608 captions, which are commonly referred to as Line 21 captions, used to be the standard for closed captioning. The FCC rules(1,2) further state that to be counted as captioned DTV programming hours the broadcast DTV signal must include both CEA-708-B and CEA-608 caption data, as show in the waveform view in Figure 1.

I USE 'EM ALL THE TIME
TO AVOID PAYING
LONG-DISTANCE CHARGES.

I USE 'EM ALL THE TIME
TO AVOID PAYING
LONG-DISTANCE CHARGES.

Figure 1 - This shows how a waveform monitor can be used to confirm the presence of 708 and 608 captions on the same DTV signal.

The 608 compatibility channel is required for simple down conversion to SD from an HD program's 708 DTVCC captions. With 708 DTVCC captions being 10 times more complex than 608 captions, the set-top box or other down conversion equipment may not have the computing power needed to down convert 708 captions. In such cases, the device uses the 608 caption data embedded in the CEA708 DTV captions for the down converted video stream.

4. Common Caption Problems

Since 708 DTVCC caption data is much more complex than 608 caption data there is a lot more room for things to go wrong. One of the things that can happen to captions is that there is a splice into a commercial a couple of frames early that cuts off the clear command. In CEA-608 there is a provision for automatic caption erasure so the captions will not normally stay on the screen through the commercial, but in CEA-708 there is no such provision so in many cases the captions will stay on the screen statically through the entire commercial.

CEA-708 captions use a windowed-based captioning system (text boxes on the screen). There are seven predefined window styles, each with 11 options, and seven predefined pen styles also with 11 options each. From eight to 22 colors are supported. Any of the 11 items in the predefined styles can be sent as a command to override the value in the predefined style item. This can lead to some very interesting captions when there is a data hit or some other glitch that changes one of the values in the captions.

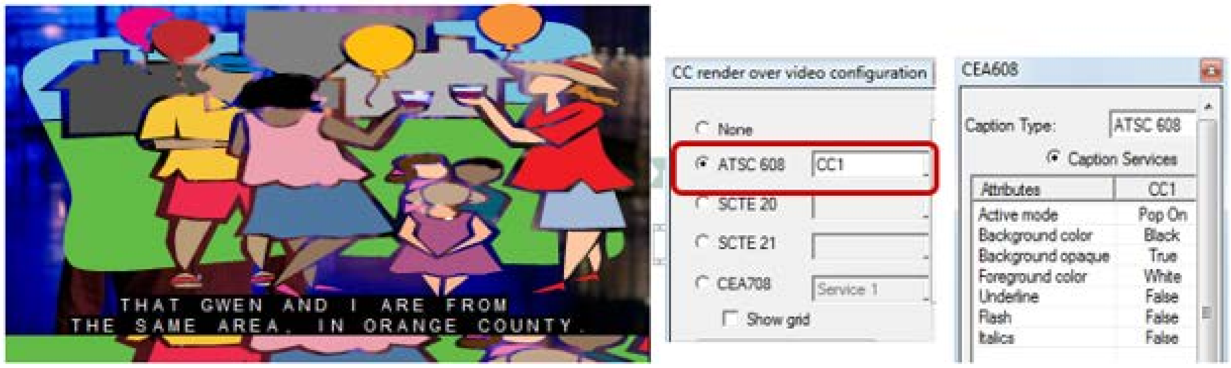
One of the common issues is changing the font color to black on a black background, meaning all you get is a black bar where the captions should be as shown in Figure 2. With the right analysis tool, such as a good waveform monitor, you can see the caption as shown in Figure 3 to quickly determine that there is actually text in the black bar as indicated by the red text to indicate the error and to help find the root cause of the caption not being visible.

Figure 2 - Black Text On A Black Background As Viewed On A Monitor

Figure 3 - The black text that was hidden in the black background is now shown in red using a Tektronix waveform monitor allowing the operator to quickly fix the problem.

Another common problem can be the visible command in the 708 DTVCC data that transfers data from the caption buffer when characters are coming in a few characters at a time. When all of a sentence or paragraph has been sent the visible command will transfer the data from the caption buffer to the active window. Most of the time when the 708 captions are missing but the 608 captions are intact, the visible command is not being sent for some reason as shown in Figure 4 and Figure 5.

Many caption decoders respond differently when the visible command is missing. Some caption decoders will put the caption data on the screen as new caption data comes in while others rely on receipt of the visible command. As a result, it is not uncommon to get different results for the same caption issue based on how the caption decoder responds to the issue. Different models of devices respond differently to the same error condition even within the same manufacturer’s equipment.



CC render over video configuration

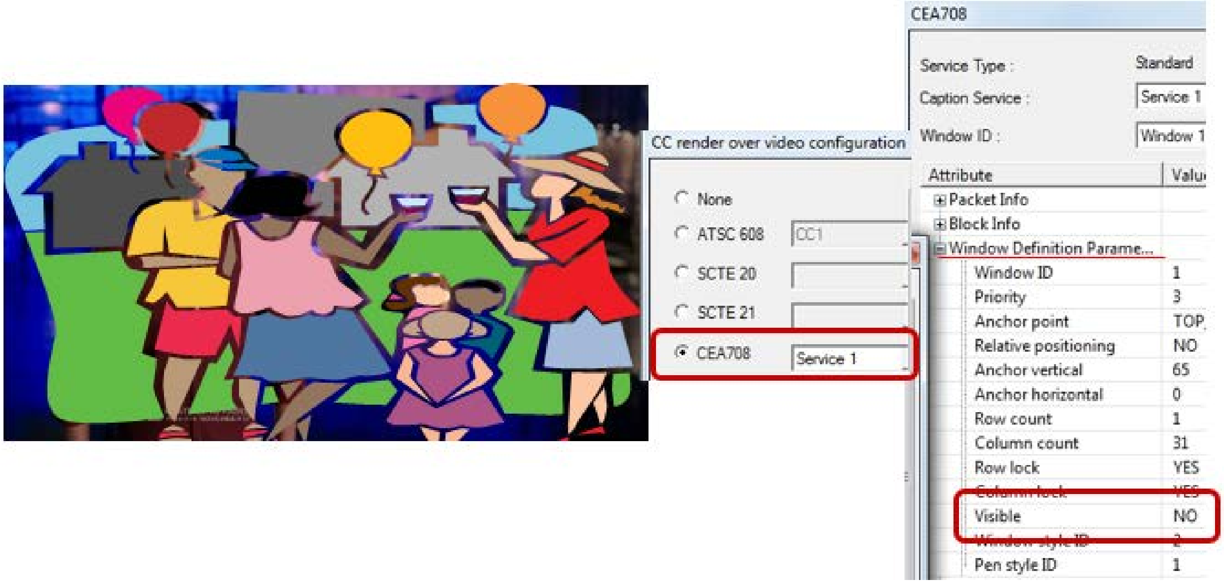
○ None
● ATSC 608 CC1
○ SCTE 20
○ SCTE 21
○ CEA708 Service 1
☐ Show grid

CEA608

Caption Type: ATSC 608
● Caption Services

Attributes	CC1
Active mode	Pop On
Background color	Black
Background opaque	True
Foreground color	White
Underline	False
Flash	False
Italics	False

Figure 4 - Analysis of 608 captions using a caption analyzer, such as the Tektronix MTS4EA


CC render over video configuration

○ None
○ ATSC 608 CC1
○ SCTE 20
○ SCTE 21
● CEA708 Service 1

CEA708

Service Type : Standard
Caption Service : Service 1
Window ID : Window 1

Attribute	Valu
⊞ Packet Info	
⊞ Block Info	
⊟ Window Definition Parame...	
Window ID	1
Priority	3
Anchor point	TOP
Relative positioning	NO
Anchor vertical	65
Anchor horizontal	0
Row count	1
Column count	31
Row lock	YES
Column lock	YES
Visible	NO
Window style ID	2
Pen style ID	1

Figure 5 - Analysis of missing 708 captions using a caption analysis tool

5. Conclusions

Addressing closed caption quality standards for accuracy, synchronicity (timing), program completeness and placement often requires eyes on screen to see when people are talking and that you are not covering up any part of the picture data that is important to context or the storyline. But troubleshooting deeper issues requires a completely different toolset than what is required to create the video itself. This requires working with vendor partners who can offer a wide set of tools for ensuring your files have captions in place, the caption data being transmitted is intact and you can decipher the commands being sent and display the data. Having the proper toolsets will help you troubleshoot captions quickly and help keep you in good graces with the FCC.

6. Notes and References

1] FCC 00-259 – 2000, Report and Order, Closed Captioning Requirements for Digital Television Receivers, Federal Communications Commission

[2] Implementing Closed Captioning for DTV, GRAHAM JONES, National Association of Broadcasters Washington, DC

About the Author

Steve Holmes has over 30 years of video experience. Before joining Tektronix, Steve spent 20 years with GTE/Verizon installing, teaching, engineering, and managing video and CATV systems. During this time he was teaching video transmission and video engineering to non-video engineers. He was involved in one of the first Video on Demand, Near Video on Demand, FTTH (fiber to the home), FTTN (fiber to the node) and HFC (hybrid fiber coax), trials before the development of digital video compression systems. For the past 19 years, Steve has worked as a Sr. Video Applications Engineer for Tektronix consulting and teaching test methodologies for broadcast, studio, cable and manufacturing for customers based in the United States and Western Canada.

Society of Cable Telecommunications Engineers

International Society of Broadband Experts

Society of Cable Telecommunications Engineers, Inc.
International Society of Broadband Experts™
140 Philips Road, Exton, PA 19341-1318
T: 800-542-5040 F: 610-884-7237

www.scte.org | www.isbe.org