EDITOR: Dan Grois, grois@ieee.org

DEPARTMENT: INDUSTRY AND STANDARDS

The Next Frontier For MPEG-5 LCEVC: From HDR and Immersive Video to the Metaverse

Simone Ferrara, Lorenzo Ciccarelli, Amaya Jiménez Moreno, Shiruo Zhao, Yetish Joshi, and Guido Meardi, *V-Nova Ltd., London, GB-W2 6LG, U.K.*

Stefano Battista, *Universita' Politecnica Delle Marche, IT-60100, Ancona, Italy*

FROM THE EDITOR

This article provides a detailed outlook to future video applications of the recently published MPEG-5 LCEVC standard, while especially focusing on HDR, 8K-resolution, Immersive video, XR and Metaverse applications. Since LCEVC is codec-agnostic and is deployed as an "enhancer", it provides a variety of benefits including visual quality improvements, end-to-end complexity reduction, bit-depth extensions, and many others. The LCEVC popularity is significantly increasing, and this article is intended to provide readers with a detailed guidance for efficient and practical LCEVC deployment.

In 2021, the newest MPEG standard was published as MPEG-5 low complexity enhancement video coding (LCEVC). Contrary to typical video codecs, LCEVC is an enhancement codec, meaning it works in combination with other codecs, to produce a more efficiently compressed video. Thanks to its simplified architecture, it is designed to be deployed as a software enhancer, which uses hardware blocks more efficiently. Despite being relatively new, it has already been adopted for a major next-gen television system (TV 3.0 in Brazil) and is being deployed across a full spectrum of applications, from broadcast to broadband. In this article we are focusing on future applications of LCEVC, from high dynamic range, 8K, and immersive video to metaverse, explaining how this new standard can make a positive impact on these applications.

Since the birth of the first digital video coding in the 1970s, many video coding technologies (or "codecs" for short) have been developed starting from H.261 in 1988,[1] passing through AVC/H.264 in 2003[2]—arguably the most successful codec to date—and arriving at the latest generation of video codecs (e.g. AV1[3] and VVC/H.266[4]) in the last few years.

Digital Object Identifier 10.1109/MMUL.2022.3213879
Date of current version 5 January 2023.

This multitude of codecs has been associated by many in the industry with the word "fragmentation" when describing the situation in the market.[5] Nevertheless, the coexistence of different coding formats is a business reality that affects providers, and often the choice of format is constrained by what is supported by consumer devices rather than by what a provider would prefer using.

In such a multicodec world, a format that can work along other formats, rather than trying to replace them, and enhance their performance and feature availability can have an important impact.

MPEG-5 low complexity enhancement video coding (LCEVC) can play that enhancement role for various existing and future use cases.

- LCEVC has already been adopted for the latest next-gen television system—TV 3.0 in Brazil—where it will work alongside VVC to enhance up to 8K resolution for over-the-air transmission, being also available to enhance both AVC and HEVC for over-the-top transmission.[6]
- For hyperscale services, poor quality of experience affects user engagement and viewing times. In addition, traffic management is needed to manage local capacity while minimizing impact on quality. LCEVC can provide higher resolution/quality feeds at lower bitrates while maintaining high resolution even when throttling traffic.
- For video calling and conferencing, low power consumption for encoding on devices and low latency are critical, with encoding having to respond rapidly to bandwidth conditions. LCEVC can provide higher resolution/quality feeds at lower bitrates, native multiscale stream enabling autoadaptation for bandwidth fluctuations with no impact on latency.
- For VoD streaming, lower quality streams drive customer churn and quality of service needs to be maintained with increased consumption over cell networks. LCEVC provides higher resolution at same bitrate as using base codec natively (e.g., FullHD below 1Mbps, UHD from 6Mbps), with fewer buffering incidents on congested networks.
- For live streaming, codec complexity usually limits encoder quality with fast action sports requiring minimal latency, higher resolutions, and higher frame rates. LCEVC can provide higher resolution and frame rates at the same bitrate as using base codec natively, maintaining latency of existing encoding, and reaching users over slow or congested networks.

In this article, we explore how LCEVC can further help on future applications, from high dynamic range (HDR), 8K, and immersive video to metaverse, explaining how this new standard can make a positive impact on those applications.

This article is organized as follows. In "LCEVC: A Brief Overview" section we provide an introduction to LCEVC, with the "Encoder and Decoder Overview" section providing a description of the encoding and decoding processes. The "LCEVC for High Dynamic Range Videos" section to the "LCEVC for XR and the Metaverse" section describe several future applications in which use of LCEVC can significantly make a difference. Specifically, the "LCEVC for High Dynamic Range Videos" section focuses on HDR content, while the "LCEVC for 8K and Immersive Video" section describes use of LCEVC with 8K content and within immersive video. The "LCEVC for XR and the Metaverse" section discusses how LCEVC can be used to enable extended reality (XR) applications in the metaverse. Finally, the "Conclusions" section draws the conclusions and describes some possible exploration work that could extend even further LCEVC applications.

LCEVC: A BRIEF OVERVIEW

MPEG-5 LCEVC—formally known as ISO/IEC 23094-2—has been developed between April 2019 and October 2020, with the final specification published in November 2021.[7]

LCEVC[8,9] consists of a set of tools designed and optimized to encode the differences between a video sequence reconstructed from a sequence coded (and decoded) using a separate coding technology (such as AVC/H.264,[2,10] HEVC/H.265,[11,12] VVC/H.266,[4,13] and EVC[14] or AV1[3,15]) and the uncompressed video sequence. This separate coding technology is usually referred to as "base (layer) codec" since it forms the base layer over which LCEVC reconstructs the video sequence.

The differences—or "residuals" in LCEVC terminology—correspond to two types of errors, namely a) the coding errors introduced by the base layer codec and b) the aliasing errors introduced by the downsampling and upsampling processes, which are typically used in LCEVC. These residuals are encoded using the LCEVC tools into up to two enhancement layers.

One of the key aspects of LCEVC is that it is designed to be low complexity as all the coding tools are designed and optimized to achieve the best compromise between low computational complexity, memory requirements, and high rate-distortion performance. Importantly, the low complexity aspect is not limited to the coding of the residuals, but it extends to the computational complexity of the overall encoding process, i.e., base codec and LCEVC. This is due to having the base codec typically executed at a lower resolution and LCEVC working on the decoded output of the base codec, thus not requiring exchange of information between base codec and LCEVC. This latter has important consequences, as it enables LCEVC to significantly improve the tradeoff between complexity and coding efficiency even for next-gen codecs, such as AV1, as recently shown in Cobianchi et al.'s work,[16] enabling software implementations both at encoder and decoder.

In terms of compression performances, the verification tests performed at the end of the standardization process over standard dynamic range (SDR) content confirm that significant bitrate savings can be achieved when using a single-layer base codec at quarter

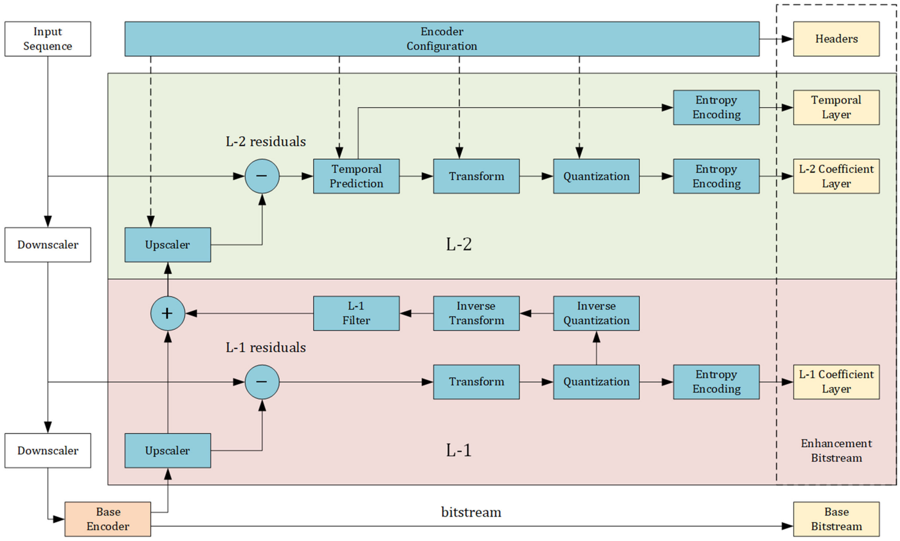

FIGURE 1. High-level block diagram of an LCEVC encoder.

resolution in conjunction with LCEVC at full resolution, when comparing to the same base coding technology at full resolution. The subjective test results reported in ISO/IEC JTC 1/SC 29/WG 04 N0076[17] indicate bitrate savings of approximately 46% over AVC, 31% over HEVC, and 15% over EVC and VVC, at the operational points used for these tests.

ENCODER AND DECODER OVERVIEW

At the encoder (see Figure 1), an input sequence is first downscaled using a two-dimensional nonnormative filter. In typical applications, the downscaling is to one quarter (half width and half height) of the original resolution. However, it is to note that based on the selected configuration, the downscaling can be done only across the vertical dimension, or even skipped completely. Further, again depending on the selected configuration, the downscaled sequence can be again downscaled to an even lower resolution. For simplicity, in this description we assume that there is only one downscaling process.

The downscaled video is compressed with the base codec (e.g., AVC, HEVC, VVC, or AV1) to generate a base bitstream conformant with the respective base codec standard. The compressed video is then decompressed to generate a base reconstruction.

The base reconstruction is used as the input for a first stage of LCEVC enhancement: the sublayer 1 (L-1) residuals are computed subtracting from the downscaled sequence the base reconstruction. These L-1 residuals are then transformed, quantized, and entropy encoded resulting in an L-1 coefficients layer, which will then form part of the LCEVC bitstream. These L-1 residuals are the entropy decoded, dequantized, and inverse transformed to be added back to the base reconstruction to generate an L-1 reconstruction.

Using a normative upscaling filter, chosen among four predefined kernels, or specified as a custom kernel with 4 filter coefficients, the L-1 reconstruction is upsampled and used as input for the second stage of LCEVC enhancement: the sublayer 2 (L-2) residuals are computed subtracting from the original sequence the upscaled L-1 reconstruction, and finally transformed, quantized, and entropy encoded in an L-2 coefficients layer, which will then form part of the LCEVC bitstream.

It should be emphasized that while L-1 process can be selected or not, the L-2 processing is always performed. In other words, LCEVC can operate with one or two enhancement layers, according to the selected configuration.

At the decoder (see Figure 2), the L-1 coefficients layer (if present) is entropy decoded, inverse quantized, and inverse transformed, to compute the L-1 residuals that are added to the base reconstruction to generate the L-1 reconstruction.

The L-1 reconstruction is then upscaled with the same normative filter used at the encoder, and then the L-2 coefficients layer is entropy decoded, inverse quantized, and inverse transformed, to compute the L-2 residuals that are added to the upscaled L-1 reconstruction.

LCEVC FOR HDR VIDEOS

HDR is often coupled with wide color gamut (WCG) to represent a higher range of luminance and chrominance planes in video content, offering the customers

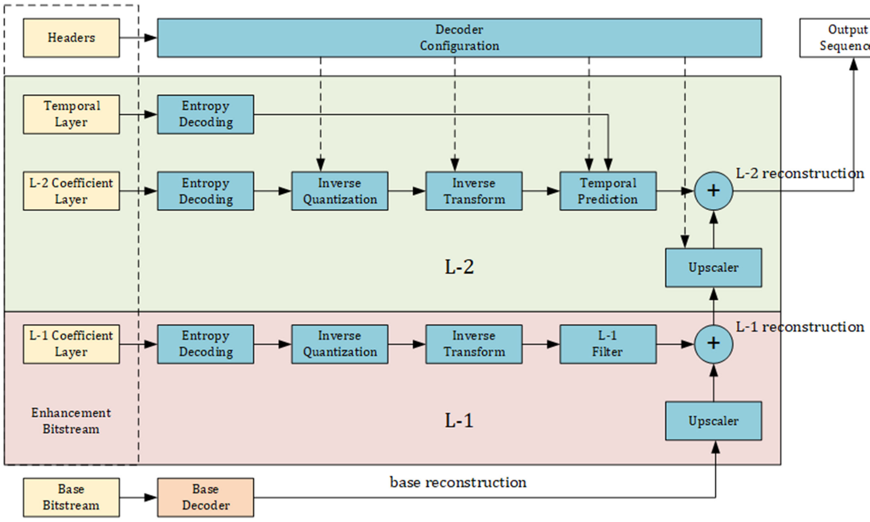

FIGURE 2. High-level block diagram of an LCEVC decoder'.

a high-quality image with levels of brightness (and darkness) and a range of colors closer to those perceived by the human visual system. It is worth noting that HDR and WCG are not independent, but rather they form a color volume,[18] which requires an accurate bit-depth representation to be correctly encoded.

Among the tools and configurations provided by LCEVC, it is possible to select an asymmetric bit-depth configuration between the base codec and LCEVC. This enables two modes of operation: *full HDR* and *extended HDR*.

In the full HDR mode, the bit-depth of the base layer and that of the enhancement layer are the same—today this is typically specified as 10-bit to reflect commercially available HDR formats, but the bit-depth can extend to up to 14 bit.

In the extended HDR mode, the bit-depth of the base layer is *lower* than that of the enhancement layer, meaning for example that an HDR source can be encoded using an AVC/H.264 or HEVC/H.265 8-bit base codec to generate an SDR base layer, and then adding on top an LCEVC 10-bit enhancement layer to bring the full-resolution video to HDR quality, allowing for example HDR support to devices and workflows, which are not designed to deliver HDR quality video. In this way, both devices, which are HDR capable and devices that are non-HDR capable can receive the same bitstream and display HDR quality.

Full HDR Mode

LCEVC can manage both HLG[19] and PQ[20] input sequences. Moreover, it is possible to include HDR metadata (like the ones defined in HDR10,[20] HDR10+,[21] SL-HDR,[22] or Dolby Vision[23]) directly into the SEI messages of the LCEVC bitstream to ensure correct display adaptation in the device display.

Ciccarelli et al.[24] reported several tests performed using LCEVC in full HDR mode. Specifically, LCEVC was tested across a range of input 4K HDR sequences (both PQ and HLG) using 10-bit base codecs. VVC/H.266 was used for the base layer and HEVC/H.265 was used as anchor.

Table 1 reports the results of the comparison between LCEVC enhancing VVC and HEVC. In this test the following implementations were used: 1) for LCEVC, a proprietary implementation provided by V-Nova (V-Nova LCEVC SDK[25]); 2) for VVC, an open-source software implementation provided by HHI Fraunhofer (VVenC 1.2.0[26]); and 3) for HEVC, the software test model implementation (HM 16.22[27]).[a] Two configurations were used. In Configuration A, the proportion of bitrate allocated to base layer and enhancement layer has been set so that the proportion of the overall bitrate allocated to the enhancement layer was between 5% and 20% depending on the operating point. Although not necessarily optimal, the proportions specified in Configuration A are more aligned with those used in the MPEG verification test[17] and are closer to the appropriate proportions used by

[a]Note that for both VVC and LCEVC there are available test model implementations—VTM[28] and LTM,[29] respectively. However, since in this specific test the target was performance for delivery of real-time live content, the test was run using fast and/or commercially available implementations for both VVC and LCEVC.

TABLE 1. HDR results.

	Average BD-Rate (MOS) (LCEVC enhancing VVC versus HEVC)	
	PQ	HLG
Configuration A	–52.58%	–54.60%
Configuration B	–36.63%	–49.17%

LCEVC when encoding the enhancement layers. On the other hand, in Configuration B the proportion of bitrate allocated to base layer and enhancement layer has been set so that the proportion of the overall bitrate allocated to the enhancement layer was kept at 50% for each operating point. Results in Table 1 are reported in terms of average BD-rate using mean opinion score (MOS) as the quality metric. BD-rate[30] represents the bitrate savings provided by a target codec with respect to a reference codec (in the present case, LCEVC-enhancing VVC corresponds to the target codec while HEVC corresponds to the reference codec). A negative BD-rate value corresponds to a saving in bitrate by the same amount. The MOS corresponds to the average score given during subjective tests run according to formal subjective assessments, as defined in Recommendation ITU-R BT.500.[31]

As it may be seen, LCEVC enhancing H.266/VVC provides an average bitrate saving of between 52% and 54% for both PQ and HLG sequences when Configuration A is used, and an average bitrate saving of almost 50% for HLG sequences and 36% for PQ sequences when Configuration B is used instead. Figure 3 shows an example of rate distortion curves for a couple of the 4k HDR sequences used in Ciccarelli et al.'s work.[24]

From a performance perspective, it is possible to see that LCEVC provides coding efficiency comparable to those already shown with SDR content during the MPEG verification tests.[17]

Extended HDR Mode

Jiménez-Moreno et al.[32] discussed the results of using LCEVC using a 10-bit enhancement layer over an 8-bit base layer encoded using AVC. In that study, both a nonnormative preprocessing module and postprocessing dithering module were introduced to reduce the banding artifact that commonly appears when changing from higher to lower bit-depths. As reported, the tests show that LCEVC can produce an HDR quality video starting from an 8-bit x264 base layer, thus providing HDR video also in those scenarios where only x264 can be supported in end-user devices (e.g., streaming applications). A preliminary subjective evaluation validated those results.

However, a better approach would be to use specific tone-mappers for preserving the quality of the base layer. In this context, there are two possible approaches that could be used: *out-of-loop* and *in-loop* approaches.

These two approaches are schematically represented in Figure 4. In the *out-of-loop* approach, the tone mapper is used as a preprocessor at the encoder and as a postprocessor at the decoder. Alternatively, in the *in-loop* approach, the tone mapper is used in the reconstruction path at the encoder (i.e., between the base codec and the calculation of the residuals) and in a similar fashion at the decoder. Note that in both schemes the base can be encoded at the same or lower resolution than the resolution used by the LCEVC encoder—in which case a downscaler and upscaler similar to those shown in Figure 1 are added to the scheme (not shown in Figure 4).

The *out-of-loop* approach has the advantage of being immediately compatible with any of the existing HDR schemes, provided the same HDR scheme is used at the encoder and decoder. On the other hand, the *in-loop* approach "ties" the LCEVC encoding scheme with a specific HDR scheme. However, it has the advantage of being able to correct any artefacts,

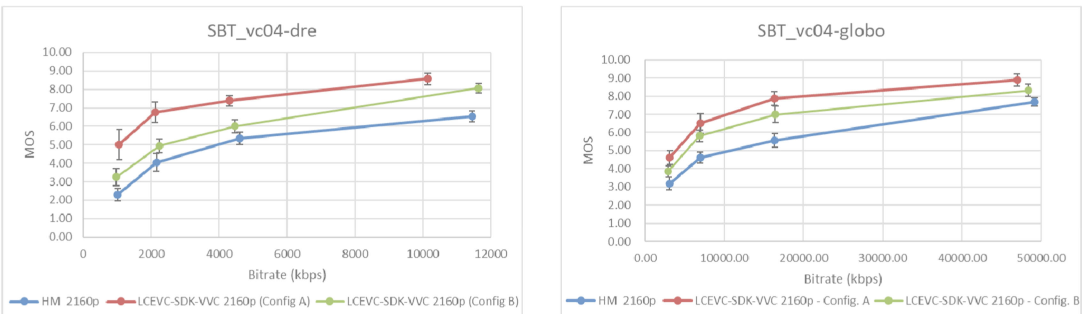

FIGURE 3. Subjective test results for LCEVC enhancing VVC using Configuration A and Configuration B versus HEVC for 4K HDR sequences. These tests were conducted using the same software implementations and configurations used in Table 1.

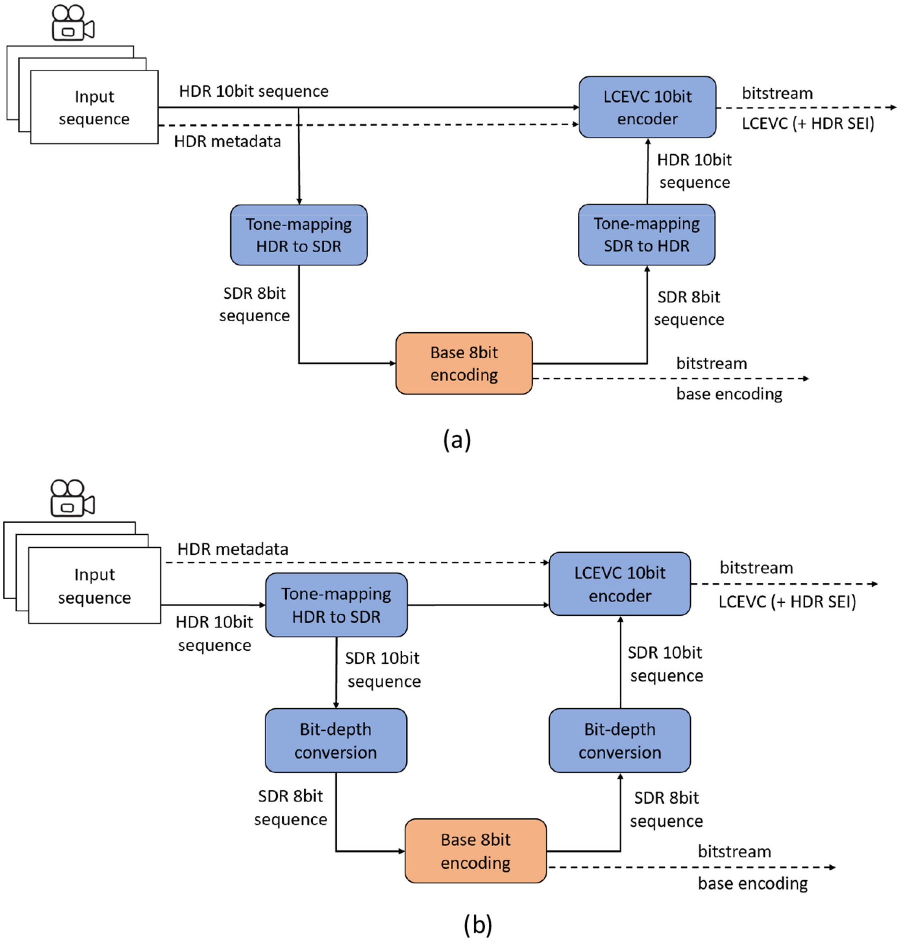

FIGURE 4. In-loop and out-of-loop configurations. (a) *In-loop* configuration. (b) *Out-of-loop* configuration.

which the HDR tone mapper may introduce and/or being unable to correct (this is particularly relevant when considering the color volume space).

LCEVC FOR 8K AND IMMERSIVE VIDEO

With the advent of larger screens and the need to provide immersive experiences to customers, providing higher resolution is an obvious requirement, and with that a higher level of compression efficiency is also needed.

However, the challenge does not end at compression efficiency alone. With the extensive use of portable devices, from mobile phones to VR headsets, power management is also an essential requirement to deliver high-quality interactive videos to consumers.

Thanks to its design principles, LCEVC is well placed to enable delivery of high-quality video while balancing the computational burden on encoding and decoding process. For example, the study presented in Cobianchi et al.'s work[16] demonstrates that LCEVC is a valuable tool to improve the quality-cycles tradeoffs across the full complexity range for three tested standards—AVC, HEVC, and AV1—while enlarging the set of mobile devices capable of playing back high quality and high frame-rate content encoded with AV1, and to extend mobile battery life by up to 50% with respect to the state-of-the-art AV1 software decoding.

8K Video Delivery

8K is the highest resolution defined in the Rec. 2020 (UHDTV) standard.[33] As the natural successor of 4K resolution, TV manufacturers have started selling 8K TV sets, since 2019. Sales of 8K TV sets are expected to reach over 4.4 million units per annum by 2025.[34]

Interestingly, recent scientific research has shown that cognitive and synesthetic factors as well as perceptual factors, and color expression were improved in 8K image quality. The research findings suggested that both

TABLE 2. 8k test results.

	BD-Rate (%)	BD-Distortion (VMAF)
BodeMuseum	−41.99	5.56
OberbaumSpree	−49.74	6.59
NeptuneFountain2	−38.81	7.62
NeptuneFountain3	−36.70	6.87
QuadrigaTree	−36.67	7.47
SubwayTree	−28.92	4.88
TiergartenParkway	−37.91	7.43
Average	−38.68	6.63

perceptual and cognitive factors are reinforced in 8K.[35] Irrespective of the quality improvement, it is undeniable that 8K will require a much higher level of compression to be delivered across existing and future delivery network, whether it is broadcast, broadband, or a combination of them. Furthermore, producing 8K content in a multiresolution world poses itself challenges both from an economical perspective and a sustainability point of view.

In this context, LCEVC can help providing the right tradeoff between compression and complexity as already demonstrated at lower resolutions.[16]

In the following analysis we have tested LCEVC enhancing HEVC against native HEVC. For the tests we have used real-time implementations of LCEVC and HEVC, namely V-Nova LCEVC SDK[25] and an open source HEVC software encoder implementation (x265 3.7[36]). For the test we used the 8K Berlin Test Sequences[37] and used constant rate factor quality configurations for both LCEVC and x265.

The test results are reported in Table 2. As it can be seen, LCEVC enhancing HEVC can provide an average bitrate saving of almost 39% with an average increase of over 6 VMAF[b] points over a corresponding encoding with HEVC.

MPEG Immersive Video

The MPEG immersive video (MIV) standard, formally known as ISO/IEC 23090-12,[39] entered the Final Draft International Standard balloting stage in October 2021. The goal of the MIV standard is to provide efficient coding of immersive, six degrees of freedom (6DoF) volumetric visual scenes. An immersive 6DoF representation, unlike a three degrees of freedom

[b]Video Multimethod Assessment Fusion (VMAF) is an objective full-reference video quality metric used to evaluate the quality of different video codecs, encoders, encoding settings, or transmission variants.[38]

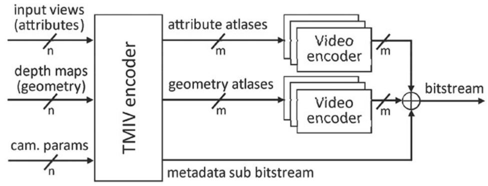

FIGURE 5. High-level block diagram of an MIV encoder.

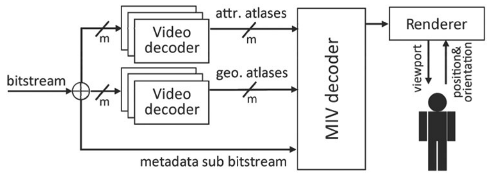

FIGURE 6. High-level block diagram of an MIV decoder.

(3DoF) representation, provides a larger viewing space, where viewers have both translational and rotational freedom of movement at their disposition. 6DoF videos also enable the perception of motion parallax, where the relative positions of scene geometry change with the pose of the viewer. The absence of motion parallax in 3DoF videos is inconsistent with the workings of a normal human visual system and often leads to visual discomfort.

As further detailed in ISO/IEC JTC 1/SC 29/AG 03 N 59,[40] an MIV encoder, illustrated in Figure 5, generates one or more attribute atlases and geometry atlases and metadata that describe the atlases. Atlases are a composition of patches extracted from the source or virtual views. The resulting attribute and geometry atlases are encoded as a video bitstream with a 2-D video encoder, while the metadata is encoded using the MIV standard.

The decoder/renderer shown in Figure 6 first performs video decoding, and then reconstructs views by reversing the atlas packing process. The MIV bitstream contains metadata indicating the packing order, position, rotation, and source view number of each patch in the atlas, which are used in the reconstruction process.

Being a standard that requires implementation on battery powered devices, MIV directly benefits from a reduction in complexity. Moreover, like any other technology that tries to reproduce multidimensional content, precise representation of depth information is critical to ensure that the content is rendered appropriately based on the instantaneous view point the user is looking at.

LCEVC is used in MIV with configurations that differ between geometry and texture surfaces. Specifically, for geometry surfaces LCEVC enhances a base layer encoded using VVC at the same resolution as

TABLE 3. MIV test results with LCEVC enhancing VVC.

	Objective metrics (Low bitrates)			Objective metrics (High bitrates)			Subjective evaluation	Timings	
	IV-PSNR	VMAF	SSIM	IV-PSNR	VMAF	SSIM	MOS	Video encoding	Decoding + Rendering
Museum (B)	4.0%	–3.5%	–4.2%	9.2%	–5.9%	0.7%	–8.6%	91.0%	83.4%
Fan (O)	3.5%	–2.2%	–3.0%	8.1%	–2.2%	–4.8%	–20.2%	78.3%	85.9%
Painter (D)	–5.1%	–11.8%	–9.3%	–1.6%	–12.2%	–7.6%	–3.5%	82.0%	95.1%
Frog (E)	–6.3%	–6.1%	–4.5%	–11.6%	–9.5%	–7.3%	–13.4%	73.1%	86.1%
Chess (N)	1.5%	–2.7%	–5.5%	9.8%	–10.3%	–9.5%	0%	67.9%	88.7%
Fencing (L)	–11.5%	–7.7%	–8.6%	–2.3%	–6.6%	–4.6%	–19.3%	69.8%	88.4%
Average	**–2.3%**	**–5.7%**	**–5.8%**	**1.9%**	**–7.8%**	**–5.5%**	**–10.8%**	**77.0%**	**87.9%**

the enhancement layer. On the other hand, for texture surfaces LCEVC enhances a base layer encoded using VVC at a quarter resolution of the enhancement layer.

We can observe average savings of between 5% and 7% for VMAF (with peaks of 12%) and around 5% for SSIM[c] (with peaks of over 9%). In terms of IV-PSNR,[d] we see average savings at low bitrates and a slight loss at higher bitrates.

After having verified that LCEVC enhancing VVC can improve objective metrics and having confirmed the same by visually inspecting the sequences, a formal subjective evaluation of the pose traces has been conducted to verify if similar visual quality improvements can be observed also by naïve viewers. For each sequence 3 pose-traces per rate have been evaluated using 16 naïve viewers. As shown in Table 3, subjective evaluations confirmed an average gain of almost 11%, with peaks of up to 20% in a couple of sequences.

Table 3 also shows the end-to-end timing analysis. Using the configuration described above, LCEVC can achieve average savings on encoding times of 23% (over 30% in some sequences). In addition, average savings of 12% on decoding times can be achieved (over 16% in some sequences).

In summary, the following benefits are demonstrated.

- LCEVC provides visual quality improvements, particularly at lower bitrates, when used to enhance the anchor. In particular, it can attain a visual quality which is comparable or better than that achieved when using the anchor alone.
- LCEVC provides a significant reduction in the overall complexity of the end-to-end process. In particular, when used to enhance the anchor it can reduce encoding times up to 30% compared to the anchor used alone along with a reduction in decoding time up to 15%.
- LCEVC allows extending the bit-depth of the geometries from 10 to 14 bit, thus adding precisions to the geometry representations. In particular, starting from a geometry reconstructed at 10 bit using the anchors, it adds coded information to bring it back to the 14 bits of the original geometry representation. This feature could also enable the original geometry representation to be processed at a higher resolution than in the current setup.

The abovementioned has been demonstrated by using LCEVC in conjunction with the current anchor (VVenC) used as a base codec. However, it is important to note that LCEVC is an enhancement that can be applied to any codec format and/or implementation.

LCEVC FOR XR AND THE METAVERSE

In recent years there has been an increase interest around the metaverse and use of XR for delivering immersive experiences. To ensure mass adoption, there are at least the following three requirements to meet.

- Quality of experience is critical to meet end-users' expectations for visually stunning, realistic, smooth, and immersive experiences. These require highest audio-visual quality, high-quality 3D objects, realistic lighting, high resolutions, high frame rates, and very low latency.

[c]SSIM is a widely used full-reference metric for assessment of visual quality of images and remote sensing data.[41]
[d]IV-PSNR is a metric that allow for the evaluation of quality loss for typical immersive video distortions: corresponding pixel shift and global component difference.[42]

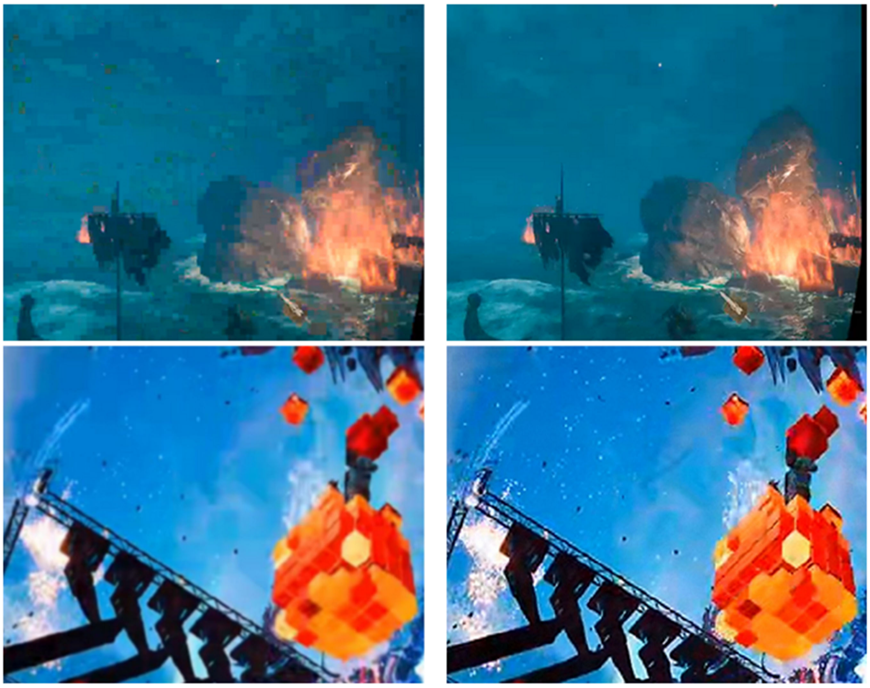

FIGURE 7. Comparison of NVENC versus LCEVC-enhanced NVENC at 25Mbps in ultralow latency setting. Top row: NVENC H264 (left) versus LCEVC-enhancing NVENC H264 (right). Bottom row: NVENC H265 (left) versus LCEVC-enhancing NVENC H265 (right).

- XR devices must be small and light for mass adoption, since consumers will not wear heavy gear for several hours a day. This limits the amount of electronics deployable on those devices as well as the available battery life, meaning that processing high-quality 3-D rendering in real time is very challenging.
- Metaverse applications must be as interoperable as web pages, meaning that very different viewing devices (including both XR headsets/smart glasses and more traditional TVs, mobile phones, tablets, or laptops) must be able to access them. This will ensure the greatest possible user base for their services. In June 2022 a Metaverse Standards Forum was launched[43] with a focus on fostering interoperability in the ecosystem.

Therefore, the most significant technical challenge to mass adoption remains the ability to render in real-time high-quality content on lightweight wearable XR headsets, which can consume a limited amount of power (1–2 Watt) and have significant less GPU computational power than typical gamers PCs. It is likely that this challenge will require a paradigm shift in terms of processing and rendering, with the 3-D rendering performed on a device (nearby or remote) different from the display device, and the resulting rendered frames streamed to the display device.

In other words, an XR device will just have to manage its sensors, process/send data of what the user is doing and decode the received high-framerate high-resolution video. This, in turn, would result in some key requirements: high-quality video within limited bandwidth constraints (30–50 Mbps) and ultralow latency transmission (20–30 ms) with no jitter. In this context, MPEG-5 LCEVC is uniquely positioned to work well for this scenario.

LCEVC can enable high-framerate (e.g., 72fps and beyond) stereoscopic 4K using a 25–50 Mbps overall bandwidth. For example, we have tested LCEVC enhancing H.264/AVC and H.265/HEVC over a GPU using the hardware-based encoders, which are natively available on such platforms (i.e., NVENC[44]). Figure 7 shows some screenshots taken from sequences encoded using NVENC (left-hand side) and LCEVC-enhancing NVENC (right-hand side). In all cases, the sequences were encoded at 25Mbps in an ultralow-latency setting.

In addition, the LCEVC data can be transmitted in a separate lower priority data channel and dropped on the fly—even in the middle of frame transmission, after encoding—in case of network congestion, without compromising the base layer and with minimal impact on visual quality. This allows to immediately reacting to the frequent oscillations and packet loss inherent to wireless transmission, avoiding piling up tens of milliseconds of latency jitter at every sudden drop of bandwidth. Further, from an overall end-to-end system latency point of view, LCEVC enables to initiate transmission and decoding of the base layer while still encoding the enhancement layer, providing an additional degree of freedom to reduce coding latencies.

As also shown in the "LCEVC for High Dynamic Range Videos" section previously, by leveraging a 16-bit accelerated pipeline, LCEVC can encode HDR video at 12 or 14 bit without any processing and/or bitrate overhead, while still using the available 8-or 10-bit base layer codecs.

Finally, as shown in the "LCEVC for 8K and Immersive Video" section previously, the ability by LCEVC to process higher bit depths can enable sending 14-bit depth maps along with the video, allowing more realistic depth-based reprojections and eye tracking-based varifocal adjustments.

CONCLUSION

In this article, we have presented MPEG-5 LCEVC, the newest MPEG standard, and provided a brief overview. We have specifically explored how LCEVC can provide a significant benefit for future applications. We have shown how LCEVC can carry HDR data in a very efficient way, as well as enhance an SDR data stream with HDR data carried within a single bitstream. We have also shown how LCEVC can efficiently encode higher resolution videos, such as 8k and described the advantages it can bring in terms of coding efficiency, encoding and decoding speed up, and bit-depth precision when used with immersive content within the MIV standard. Finally, we have described how LCEVC can play a pivotal enabling role in the deployment of the metaverse and the split computing paradigm needed for it to work.

REFERENCES

1. "H.261: Video codec for audiovisual services at p x 384 kbit/s," 1988. [Online]. Available: https://www.itu.int/rec/T-REC-H.261-198811-S/en
2. T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003, doi: 10.1109/TCSVT.2003.815165.
3. J. Han et al., "A technical overview of AV1," *Proc. IEEE*, vol. 109, no. 9, pp. 1435–1462, Sep. 2021, doi: 10.1109/JPROC.2021.3058584.
4. B. Bross et al., "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736–3764, Oct. 2021, doi: 10.1109/TCSVT.2021.3101953.
5. J. De Cock, "Navigating a multi-codec world," 2021. [Online]. Available: https://www.streamingmedia.com/Articles/Editorial/Featured-Articles/Navigating-a-Multi-Codec-World-145580.aspx?utm_source=related_articles&utm_medium=gutenberg&utm_campaign=Editors_selection
6. TV 3.0 Project, 2022. [Online]. Available: https://forumsbtvd.org.br/tv3_0/
7. ISO/IEC 23094-2:2021 Information technology – General video coding — Part 2: Low complexity enhancement video coding, 2021. [Online]. Available: https://www.iso.org/standard/79143.html
8. G. Meardi et al., "MPEG-5 part 2: Low complexity enhancement video coding (LCEVC): Overview and performance evaluation," in *Proc. SPIE 11510, Appl. Digit. Image Process. XLIII*, 2020, Art. no. 115101C, doi: 10.1117/12.2569246.
9. S. Battista et al., "Overview of the low complexity enhancement video coding (LCEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jun. 2022.
10. H.264/AVC, ISO/IEC 14496-5/AMD 6, information technology — coding of audio-visual objects — Part 5: Reference software — amendment 6: Advanced video coding (AVC) and high efficiency advanced audio coding (HE AAC) reference software, 2008.
11. G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012, doi: 10.1109/TCSVT.2012.2221191.
12. H.265/HEVC. ISO/IEC 23008-5, information technology — high efficiency coding and media delivery in heterogeneous environments — Part 5: Reference software for high efficiency video coding, 2017.
13. H.266/VVC. ISO/IEC 23090-16, Information technology – coded representation of immersive media – Part 16: Reference software for versatile video coding, 2021.
14. K. Choi, J. Chen, D. Rusanovskyy, K.-P. Choi, and E. S. Jang, "An overview of the MPEG-5 essential video coding standard [Standards in a nutshell]," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 160–167, May 2020, doi: 10.1109/MSP.2020.2971765.
15. AV1, 2018. [Online]. Available: https://aomediacodec.github.io/av1-spec/av1-spec.pdf
16. G. Cobianchi et al., "Enhancing SVT-AV1 with LCEVC to improve quality-cycles trade-offs and enhance sustainability of VOD transcoding," in *Proc. SPIE 12226, Appl. Digit. Image Process. XLV*, 2022, Art. no. 122260S, doi: 10.1117/12.2633882.

17. Verification test report on the compression performance of low complexity enhancement video coding. ISO/IEC jtc 1/SC 29/WG 04 N0076 (20173), 2021. [Online]. Available: https://www.mpegstandards.org/wp-content/uploads/mpeg_meetings/134_OnLine/w20173.zip
18. 2018. [Online]. Available: https://professional.dolby.com/siteassets/pdfs/dolby-vision-measuring-perceptual-color-volume-v7.1.pdf
19. ARIB-STD-B67, 2015. [Online]. Available: http://www.arib.or.jp/english/html/overview/doc/2-STD-B67v1_0.pdf
20. "ST 2084:2014 - SMPTE Standard - High Dynamic range electro-optical transfer function of mastering reference displays," in ST 2084:2014, pp. 1–14, Aug. 29, 2014, doi: 10.5594/SMPTE.ST2084.201. [Online]. Available: https://ieeexplore.ieee.org/document/7291452
21. 2015. [Online]. Available: https://www.bigpicturebigsound.com/What-Makes-a-TV-HDR-Compatible-The-CEA-Sets-Guidelines.shtml
22. "ST 2094-40:2020 - SMPTE Standard - Dynamic Metadata for color volume transform — Application #4," in ST 2094-40:2020, pp. 1–29, May 16, 2020, Citation of metadata for SL-HDR, doi: 10.5594/SMPTE.ST2094-40.202. [Online]. Available: https://ieeexplore.ieee.org/document/9095450
23. 2016. [Online]. Available: https://professional.dolby.com/siteassets/pdfs/dolby-vision-whitepaper_an-introduction-to-dolby-vision_0916.pdf
24. L. Ciccarelli, S. Ferrara, and F. Maurer, "MPEG-5 LCEVC for 3.0 next generation digital TV in Brazil," *Front. Signal Process.*, vol. 2, 2022, Art. no. 884254, doi: 10.3389/frsip.2022.884254.
25. "MPEG specifications, performance evaluations, user guides, product documentation and other LCEVC-related resources," 2021. [Online]. Available: https://www.lcevc.org/lcevc-resources/
26. VVenC GitHub repository, 2021. [Online]. Available: https://github.com/fraunhoferhhi/vvenc
27. HEVC Test Model (HM) GitLab repository, 2010–2022. [Online]. Available: https://vcgit.hhi.fraunhofer.de/jvet/HM
28. VVC Test Model (VTM) repository, 2018–2022. [Online]. Available: https://jvet.hhi.fraunhofer.de/
29. LCEVC Test Model (LTM) repository, 2019–2022. [Online]. Available: http://mpegx.int-evry.fr/software/MPEG/LCEVC/LTM
30. G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG16/Q6 input document VCEG-M33, 2001.
31. Recommendation ITU-R BT.500-14, "Methodologies for the subjective assessment of the quality of television images," 2019. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-14-201910-I!!PDF-E.pdf
32. A. Jiménez-Moreno, L. Ciccarelli, R. Clucas, and S. Ferrara, "HDR video coding with MPEG-5 LCEVC," in *Proc. 1st Mile-High Video Conf.*, 2021, pp. 25–31, doi: 10.1145/3510450.3517307.
33. Rec. ITU-R 2020, "Parameter values for ultra-high definition television systems for production and international programme exchange."
34. 2022. [Online]. Available: https://www.statista.com/statistics/1294400/8k-tv-shipments-worldwide/
35. D. E. Park, Y. J. Kim, and Y. K. Park, "Cognitive effect on image quality in full ultra-high definition(8 K)," *J. Inf. Display*, vol. 21, no. 2, pp. 103–111, 2020, doi: 10.1080/15980316.2019.1704895.
36. x265 (929cd6f8 - Release3.7_RC2), open source HEVC software encoder, 2014–2022. [Online]. Available: https://www.videolan.org/developers/x265.html
37. 8K Berlin Test Sequences, 2022. [Online]. Available: https://www.hhi.fraunhofer.de/en/departments/vca/research-groups/video-coding-systems/8k-sequences.html
38. Video Multimethod Assessment Fusion (VMAF), 2016–2022. [Online]. Available: https://github.com/Netflix/vmaf
39. ISO/IEC 23090-12, "Information technology - Coded representation of immersive media - Part 12: MPEG immersive video," 2021. [Online]. Available: https://www.iso.org/standard/79113.html
40. "White paper on MPEG immersive video," ISO/IEC JTC 1/SC 29/AG 03 N 59, Jan. 2022.
41. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.
42. A. Dziembowski, D. Mieloch, J. Stankowski, and A. Grzelka, "IV-PSNR – the objective quality metric for immersive video applications," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2022.3179575.
43. Metaverse Standards Forum, 2022. [Online]. Available: https://metaverse-standards.org/
44. "NVIDIA NVENC is a hardware-based encoder available on NVIDIA GPU and supporting various codec formats, including H.264/AVC and H.265/HEVC," 2022. [Online]. Available: https://developer.nvidia.com/nvidia-video-codec-sdk

SIMONE FERRARA is currently a senior vice president at V-Nova with responsibility for the technology, IP, and standardization strategy, including driving its execution across the company. Ferrara received his M.Sc. degree in electrical engineering from Washington University in St. Louis, St. Louis, MO, USA and M.Sc. degree in telecommunications engineering from Politecnico di Milano, Milan, Italy. Contact him at simone.ferrara@v-nova.com.

LORENZO CICCARELLI is technical lead for applied R&D with V-Nova, London, GB-W2 6LG, U.K. He was a researcher and team lead on several video compression projects both in the broadcasting and video conferencing industry. He has been involved in the algorithm design and standardization process for MPEG5-Part2 LCEVC. Ciccarelli received the Laurea degree in electronic engineering and telecommunication from the Università Politecnica delle Marche, Ancona, Italy. Contact him at lorenzo.ciccarelli@v-nova.com.

AMAYA JIMÉNEZ-MORENO is currently a video research engineer at V-Nova. She has been assistant professor at Universidad Carlos III de Madrid and Universidad San Pablo CEU de Madrid, Madrid, Spain. She has been part of the Biocomputing Unit, National Center of Biotechnology, Spanish National Research Council (CSIC), Madrid, Spain. Her research interests include video coding and image processing, topics for which she has coauthored several papers in international journals and conferences. Jiménez-Moreno received her Ph.D. degree in telecommunication engineering from the Universidad Carlos III de Madrid. Contact her at amaya.moreno@v-nova.com.

SHIRUO ZHAO is currently working on the standardization of video codecs. She is an R&D researcher at V-Nova focusing on data compression. Zhao received her M.Sc. degree in human–computer interaction from the University of St Andrews, St Andrews, U.K. Contact her at shiruo.zhao@v-nova.com.

YETISH JOSHI is currently a video quality engineer at V-Nova, with particular focus on conducting experiments related to visual quality. He has worked on technology/industry related to 3-D positional audio, railway signaling systems, and automatic license/number plate recognition systems. Joshi received his Ph.D. degree in "Low complexity in-loop perceptual video coding" from Middlesex University, London, U.K. Contact him at yetish.joshi@v-nova.com.

GUIDO MEARDI is a co-founder and CEO of V-Nova, a leading company in data compression and artificial intelligence. He is a keen innovator, entrepreneur, and investor, with relevant business building experience and half a dozen exits. Meardi received his MBA degree from MIT Sloan, Cambridge, MA, USA, where he was a Siebel scholar, and M.Sc. degree in computer engineering from Politecnico di Milano, Milan, Italy, where he was an Intel scholar. Contact him at guido.meardi@v-nova.com.

STEFANO BATTISTA was a researcher with the video and multimedia group of Telecom Italia Lab, Torino, Italy, from 1991 to 1997. His main activities have been on video coding, video analysis, 3-D modeling, multimedia systems, with a focus on standardization for multimedia communications. Battista received his Laurea degree in electronic engineering from the Università Politecnica delle Marche, Ancona, Italy. Contact him at s.battista@univpm.it.