An MMT-Based Hierarchical Transmission Module for 4K/120fps Temporally Scalable Video

Yasuhiro MOCHIDA[†a)], Takayuki NAKACHI[†], *and* Takahiro YAMAGUCHI[†], *Members*

SUMMARY   High frame rate (HFR) video is attracting strong interest since it is considered as a next step toward providing Ultra-High Definition video service. For instance, the Association of Radio Industries and Businesses (ARIB) standard, the latest broadcasting standard in Japan, defines a 120 fps broadcasting format. The standard stipulates temporally scalable coding and hierarchical transmission by MPEG Media Transport (MMT), in which the base layer and the enhancement layer are transmitted over different paths for flexible distribution. We have developed the first ever MMT transmitter/receiver module for 4K/120fps temporally scalable video. The module is equipped with a newly proposed encapsulation method of temporally scalable bitstreams with correct boundaries. It is also designed to be tolerant to severe network constraints, including packet loss, arrival timing offset, and delay jitter. We conducted a hierarchical transmission experiment for 4K/120fps temporally scalable video. The experiment demonstrated that the MMT module was successfully fabricated and capable of dealing with severe network constraints. Consequently, the module has excellent potential as a means to support HFR video distribution in various network situations.
key words: MPEG Media Transport (MMT), high frame rate (HFR), temporally scalable coding, hierarchical transmission

1.   Introduction

High Efficiency Video Coding (HEVC) has achieved high compression performance and made it possible to distribute Ultra-High Definition (UHD) video with less bandwidth consumption [1]–[4]. High spatial resolution (4K or more) video distribution services are rapidly spreading; for instance, 4K and 8K satellite broadcasting has been launched in Japan since 2018 [5]. However, temporal resolution as well as spatial resolution should be considered for realistic image representation [6]. High frame rate (HFR) video provides smoothness and sharpness even in fast-moving scenes and improves subjective quality [7]. HFR is attracting strong interest from UHD service providers since it is especially beneficial for sports programs, the major content of UHD services. The UHD television system parameters specified in ITU-R BT.2020 support HFR video formats of up to 120 frames per second (fps) [8].

For broadcasting in particular, it is important to provide not only realistic image representation but also backward compatibility with non-HFR decoders. For instance, in Japan, the Association of Radio Industries and Businesses (ARIB) standard has specified a 120 fps broadcasting service format and stipulates temporally scalable coding [9].

Although 120 fps broadcasting is not included in current 4K and 8K satellite broadcasting, it is considered as a next step [10]. Temporally scalable coding is one of the scalable extensions of HEVC, and generates a hierarchical bitstream composed of a base layer (60 fps sub-bitstream) and an enhancement layer (120 fps subset) [11]. The base layer can be decoded as 60 fps video even by legacy 60 fps decoders since it is independent of the enhancement layer. In 120 fps decoding, the enhancement layer is used in addition to the base layer. In this manner, multiple frame rate videos can be derived from one bitstream, which releases service providers from distributing multiple bitstreams of the same video with different frame rates. Omori et al. developed a real-time 4K/120fps HEVC encoder system with temporally scalable coding [12].

For flexible distribution, the ARIB standard also stipulates hierarchical transmission by MPEG Media Transport (MMT), in which the base layer and the enhancement layer can be transmitted over different paths [13]–[15]. MMT is a container format designed for both broadcasting and broadband network media distribution to succeed MPEG-2 Transport Stream (TS) [16]. The latest broadcasting standards, ARIB STD-B60 and ATSC 3.0, have adopted MMT [17], [18]. In existing broadcasting systems, the MPEG-2 TS format is widely used since it is suitable for uniform media distribution. It multiplexes media data, signaling messages and clock information into one stream; however, it has difficulty in utilizing multiple transmission paths. Thus, it is unsuitable for transmitting hierarchical media data, including temporal scalable video, over multiple paths separately. On the other hand, MMT allows for the transmission path control of each media data, e.g., the base layer over satellite broadcasting and the enhancement layer over broadband networks. It is also possible to deliver the base layer to all decoders for 60 fps decoding and the enhancement layer only to HFR decoders, saving the bandwidth consumption. However, no existing MMT systems have hierarchical transmission capability.

In this paper, we propose a system design of the first ever MMT transmitter/receiver module for 4K/120fps temporally scalable video in accordance with the ARIB STD-B60 [19], [20]. The transmitter encapsulates temporally scalable bitstreams derived from MPEG-2 TS into Media Processing Units (MPUs). The base layer and the enhancement layer can be streamed onto the different paths with the MMT Protocol (MMTP). The receiver reconstructs the 4K/120fps bitstream from the MMTP packets and outputs

it to the subsequent decoder in MPEG-2 TS format. Furthermore, the MMT module is designed to be tolerant to severe network constraints. The packet loss can be recovered by Application Layer Forward Error Correction (AL-FEC). The arrival timing offset between the layers attributable to the different paths and the delay jitter can be absorbed by the receiver. An evaluation experiment demonstrated the module functionality of transmitting 4K/120fps temporally scalable bitstreams over different paths successfully even under packet loss, arrival timing offset and delay jitter. Our work makes the following contributions.

- Hierarchical MPU encapsulation of temporally scalable bitstreams with correct boundaries
- Clock synchronization mechanism
- Module hardware implementation tolerant to severe network constraints

2. Background

2.1 Media Transport Technologies

MPEG-2 TS is a widely-used container format suitable for distributing the same content to many receivers and has been adopted in broadcasting systems. Video and audio are encapsulated into Packetized Elementary Stream (PES), which are divided into TS packets and multiplexed into one stream. On the other hand, it is not possible to combine video/audio components transmitted by separate streams. TS packets can be transmitted over broadband networks using Real-time Transport Protocol (RTP) [21]. However, TS packets are not fit for IP transmission where variable packet length is allowed because the length of a TS packet is fixed at 188 bytes.

MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH) is an adaptive bitrate streaming technique using an HTTP server [22]. An MPEG-DASH client selects the video quality to be acquired according to the available network bandwidth [23]. MPEG-DASH is adopted by over-the-top video distribution services since it works well with content delivery network. However, MPEG-DASH cannot be used in broadcasting because it is based on TCP.

MMT was standardized to satisfy the increasingly diverse requirements for media distribution. Video and audio components can be transmitted over different paths including broadcasting and broadband networks; they can be presented synchronously. For example, as illustrated in Fig. 1, 4K video, 5.1ch audio and multi-lingual subtitle are presented in a home theater while multi-angle video and auxiliary information are presented in a tablet. Besides, MMT is more efficient than MPEG-2 TS in terms of packetizing overhead since MMTP allows for variable packet length.

2.2 Overview of MMT
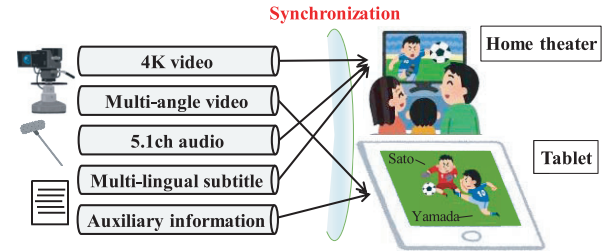
Figure 2 shows the MMT architecture. The major features
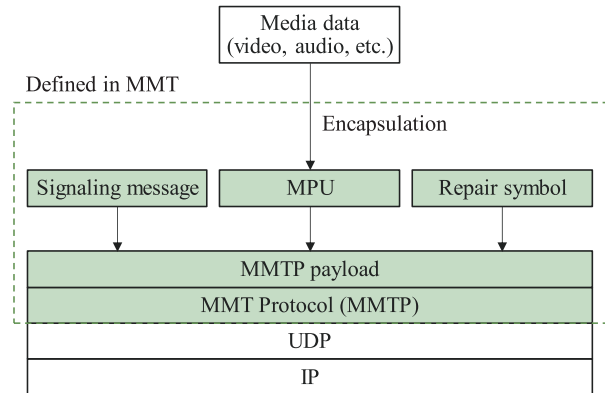

Fig. 1 Media distribution using MMT.

Fig. 2 MMT architecture.

of MMT are a media encapsulation framework, a synchronized presentation framework using Coordinated Universal Time (UTC) timestamps, and an AL-FEC framework. Media data (video, audio, etc.) is split and encapsulated into MPUs. For example, HEVC video bitstreams are split at the Group of Pictures (GOP) boundaries. Each MPU has a sequence number (MPU sequence number) to indicate the processing order. MPUs are transmitted by MMTP packets; this allows for variable packet length suitable for IP networks. Each media data stream identified by a packet ID can be assigned to an arbitrary path. For presentation synchronization, signaling messages deliver UTC timestamps to indicate presentation time of MPUs. Signaling messages also deliver media lists, video parameters, AL-FEC parameters, etc.

For reliable transmission, MMTP packets can be protected from packet loss by AL-FEC. MMT defines a common framework and six coding algorithms [24]. The AL-FEC framework specifies the repair symbol generation process independent of coding algorithms. One of six algorithms can be selected considering the recovery performance and computation complexity.

Aoki et al. developed an MMT-based broadcasting system and showed that multiple videos delivered over different paths can be presented synchronously [25]. Nakachi et al. developed an MMT-based video streaming system with FireFort-LDGM (FF-LDGM), one of the coding algorithms, and demonstrated its robustness in an intercontinental streaming experiment over unreliable IP networks [26]. However, no existing systems have achieved MMT-based
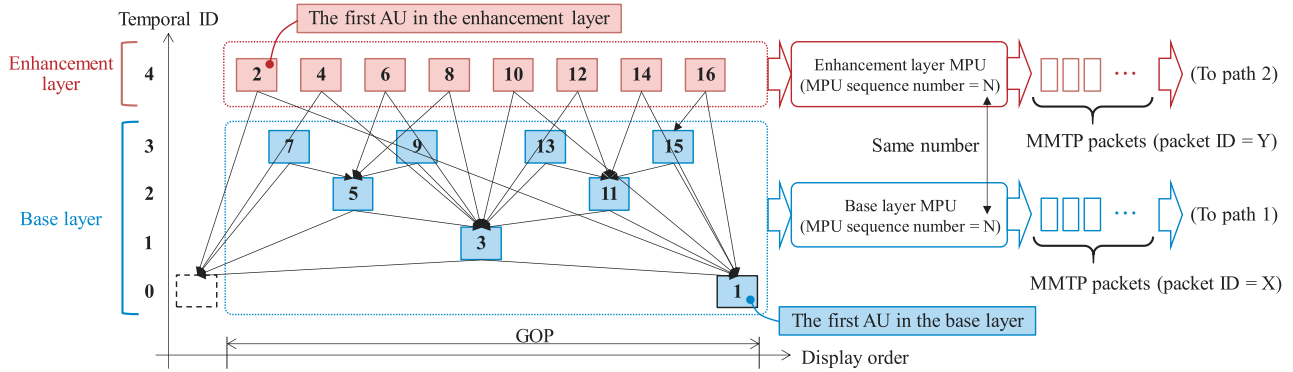
Fig. 3 An example encoding structure and MMT-based hierarchical transmission. Note that each square represents an AU and the number in it is the decoding order.
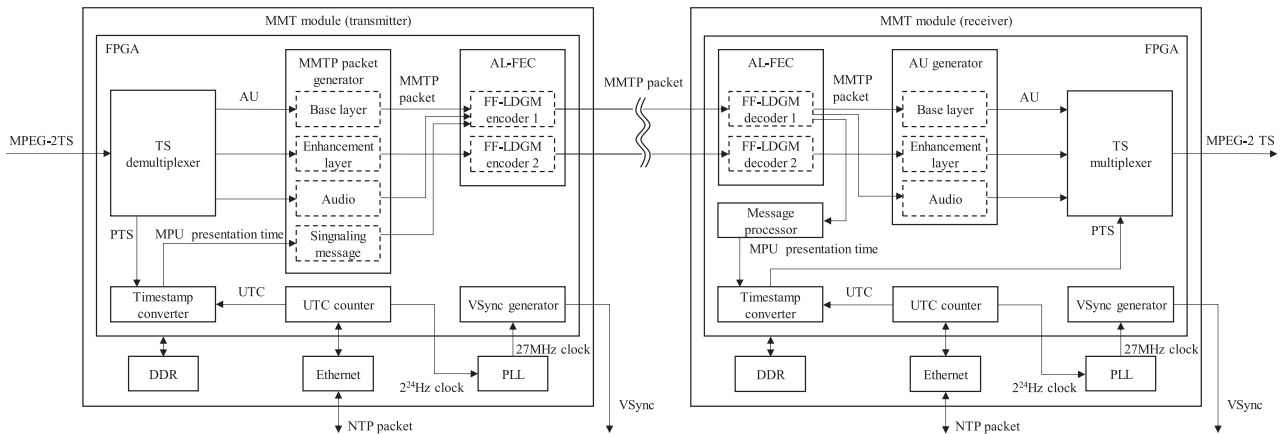

Fig. 4 Block diagrams of a transmitter and a receiver.

hierarchical transmission of temporally scalable bitstreams.

2.3 MMT-Based Hierarchical Transmission

Figure 3 illustrates an example encoding structure of temporally scalable coding and the hierarchical transmission specified in the ARIB STD-B60. Note that each square represents an access unit (AU) containing a coded picture. In a temporally scalable bitstream, the base layer AUs and the enhancement layer AUs appear alternately. Each AU has a temporal ID to indicate its dependency hierarchy. The enhancement layer is composed of the AUs with the highest temporal IDs and the base layer is composed of the other AUs. Each AU is independent of the AUs with higher temporal IDs, and thus the base layer is independently decodable.

A GOP of a temporally scalable bitstream is encapsulated into the base layer MPU and the enhancement layer MPU separately. They shall be given the same MPU sequence number so that the decoder can identify that these MPUs involve AUs that belong to the same GOP. Then, they are transmitted by MMTP packets with different packet IDs, which can be assigned to different paths. The 120 fps receiver receives both layers and reconstructs a 120 fps bitstream from MPUs with the same MPU sequence number

while the 60 fps receiver reconstructs a 60 fps bitstream using only the base layer.

3. System Design

In this section, we propose the system design of the MMT transmitter/receiver. Since most of the existing encoders/decoders are utilizing MPEG-2 TS, the module bidirectionally converts MMT and MPEG-2 TS. Block diagrams of the transmitter and the receiver are shown in Fig. 4.

3.1 Hierarchical MPU Encapsulation

The transmitter input from a preceding encoder is in MPEG-2 TS format. A 4K/120fps video bitstream is encapsulated into hierarchical MPUs and transmitted by MMTP packets over different paths. The receiver reconstructs the 4K/120fps video bitstream from MMTP packets delivered over two paths and outputs it to the subsequent decoder in MPEG-2 TS format.

The key technical challenge in MMT-based hierarchical transmission is MPU encapsulation with correct GOP boundaries. An MPU is expected to involve AUs that belong to the same GOP; otherwise the MMT receiver cannot reconstruct 4K/120fps bitstreams. Figure 5 illustrates the
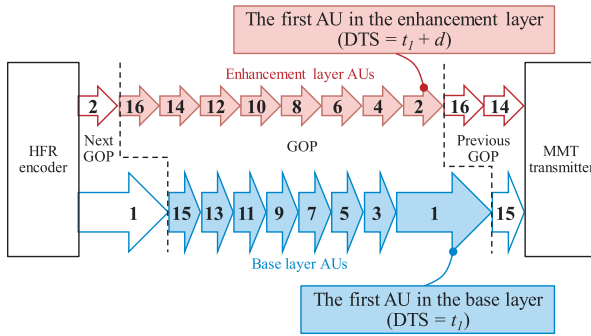
Fig. 5 MPEG-2 TS input for the MMT transmitter. Note that each arrow represents a coded picture and the number in it is the decoding order.

MPEG-2 TS input for the transmitter, which conveys the base layer and the enhancement layer with different identifiers (PIDs). There might be some timing offset between the layers. The TS demultiplexer derives AUs from the TS packets; however, the GOP boundaries are not explicit. In the base layer, the first AU in decoding order in a GOP can be recognized from Network Abstraction Layer (NAL) unit types constituting the AU since only the first AU includes VPS/SPS NAL units. On the other hand, in the enhancement layer, the first AU cannot be recognized in the same way and deeper analysis of NAL payloads is required, which increases implementation complexity.

We propose a GOP boundary detection method utilizing decoding timestamps (DTSs) in MPEG-2 TS instead of NAL payload analysis. First, the demultiplexer detects the first AU from NAL unit types and keeps its DTS as t_1. Then, the DTS of the first AU in the enhancement layer, t_2, can be calculated as $t_2 = t_1 + d$, where d is the DTS difference between consecutive AUs. If t_2 is larger than the enhancement layer DTS at the moment, the demultiplexer waits for the AU with DTS t_2. In case that the enhancement layer DTS is larger than t_2, the expected AU has already passed. Then the demultiplexer waits for the first AU in the next GOP whose DTS is $t_1 + d * (L + 1)$, where L is the number of AUs in a GOP. Our system design has the premise that L is always fixed. Thus, once a correct GOP boundary is determined, the following GOP boundaries can be calculated by d and L. Therefore, this operation is required only once at the start of transmission. In this manner, the demultiplexer is able to determine the GOP boundaries of both layers and give the same MPU sequence number to the corresponding MPUs without delaying the AUs in the buffer.

3.2 Clock Synchronization Mechanism

Another technical challenge is alternating clock synchronization mechanism of MPEG-2 TS. MPEG-2 TS has a clock synchronization mechanism between an encoder and a decoder on the basis of the encoder clock counter in 27MHz, which is called Program Clock Reference (PCR). If there is a clock speed offset between the encoder and the decoder, a continuous operation may cause buffer overflow or buffer

underflow. Thus, in an MPEG-2 TS system, the decoder adjusts its clock speed using PCR embedded in TS packets. Therefore, an alternative clock synchronization mechanism is required in bidirectional conversion between MPEG-2 TS and MMT.

We propose a UTC-based clock synchronization mechanism in MMT modules to prevent buffer overflow and underflow. Each MMT module acquires UTC using the Network Time Protocol (NTP), adjusts 2^{24}Hz clock speed, and individually generates a synchronized 27MHz clock signal using a PLL circuit. The transmitter and the receiver provide a Vertical Synchronization (VSync) signal generated from the clock signal respectively for the preceding encoder and the subsequent decoder, which enables the encoder and the decoder to work synchronously.

Under the clock synchronization, presentation timestamp in MPEG-2 TS (*PTS*) and that of MMT (*mpu_presentation_time*) are converted in the following manner. The timestamp converter periodically preserves a coincident pair of UTC (T_{UTC}) and PCR (T_{PCR}). Then, *mpu_presentation_time* can be calculated by Eq. (1).

$$mpu\_presentation\_time$$
$$= T_{UTC} + (PTS - T_{PCR}) \times 2^{24}/(2.7 \times 10^7) \qquad (1)$$

The MMTP packet generator embeds *mpu_presentation_time* in a Package Access (PA) message to indicate presentation time, attaches the

MMTP header, and sends the packet. In the receiver, *mpu_presentation_time* is again converted to PTS using Eq. (1) inversely and provide it for the TS multiplexer, which gives PTS to MPEG-2 TS.

3.3 Tolerance to Network Constraints

The MMT module is designed to be tolerant to packet loss, arrival timing offset, and delay jitter. In order to protect MMTP packets from packet loss, the MMT module is equipped with FF-LDGM, an AL-FEC algorithm with high efficiency and low complexity. In hierarchical transmission, the base layer and the enhancement layer can be protected by different AL-FEC configurations in accordance with the video layer characteristics or the transmission path characteristics. Figure 6 depicts two example use-cases. In pattern 1, the base layer is protected with 10% redundancy while the enhancement layer is protected with 5% redundancy. The base layer, which is indispensable for decoding, is protected more strongly than the enhancement layer. 60 fps video can be decoded from the base layer even if the enhancement layer cannot be recovered completely. Pattern 2 is a hybrid transmission where the base layer and the enhancement layer are transmitted over satellite broadcasting and IP networks, respectively. Only the enhancement layer is protected by AL-FEC since the satellite broadcasting channel has a loss recovery function.

Furthermore, the receiver is designed to absorb arrival timing offset between the layers due to the different transmission paths and delay jitter. If the offset propagates to
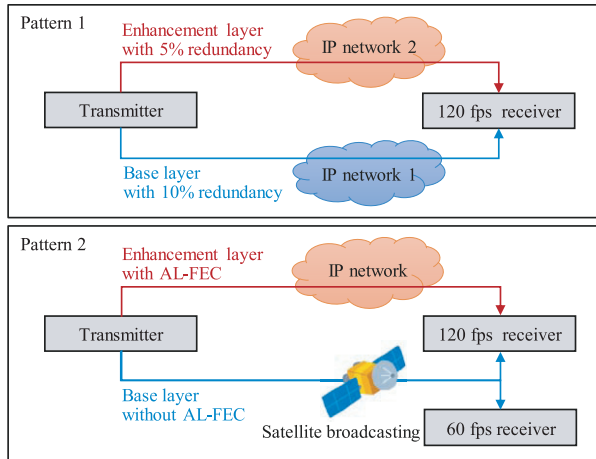
Fig. 6 AL-FEC use-cases.

Table 1 MMT module specification.

IP	version	IPv4 and IPv6
	protocol	MMTP over UDP
	bitrate	Max. 216Mbps
Video	format	3840x2160
		119.88 fps
		Progressive
	coding	HEVC main10 profile
Audio	format	2ch and 5.1ch
	coding	MPEG-4 AAC-LC
AL-FEC	block size	Max. 8000 packets
	redundancy	Max. 20%
Transmitter	FPGA type	Artix7@100MHz
	slice	28482
	block RAM	215
	DSP	47
	power	4.2W
Receiver	FPGA type	Artix7@100MHz
	slice	27024
	block RAM	223
	DSP	5
	power	4.3W

the MPEG-2 TS output of the receiver, it can cause buffer overflow in the subsequent decoder. Therefore, the receiver buffers MMTP packets in an external memory and suppresses the timing offset between the layers.

4. Implementation

We implemented the MMT module described in the previous section using an FPGA. Detailed module specifications are given in Table 1 and the fabricated module is shown in Fig. 7. The module's transmitter and receiver modes can be switched by reconfiguring its FPGA. The module deals with 4K/120fps HEVC video and 5.1ch MPEG-4 AAC-LC audio. The MMTP packets can be transmitted over IPv4 and IPv6. The maximum bitrate is over 200 Mbps, which is applicable not only to distribution but also to concentration. The module is equipped with two AL-FEC coding algorithms, FF-LDGM and Pro-MPEG. The maximum block size of FF-LDGM is large enough to restore burst loss. The receiver is capable of absorbing the timing offset up to 2000 ms, the maximum possible offset between hybrid transmission utilizing satellite broadcasting and broadband networks.

5. Evaluation

We conducted a hierarchical transmission experiment of 4K/120fps video in which the base layer and the enhancement layer were transmitted over different paths.

5.1 Experiment

The experimental system is depicted in Fig. 8. The streaming system was composed of an HFR encoder and an MMT transmitter while the viewing system was composed of an HFR decoder and an MMT receiver. The video input interface of the streaming system was eight 3G-SDI links. 4K/120fps video was provided as two 4K/60fps videos. One video was composed of odd frames while the other was
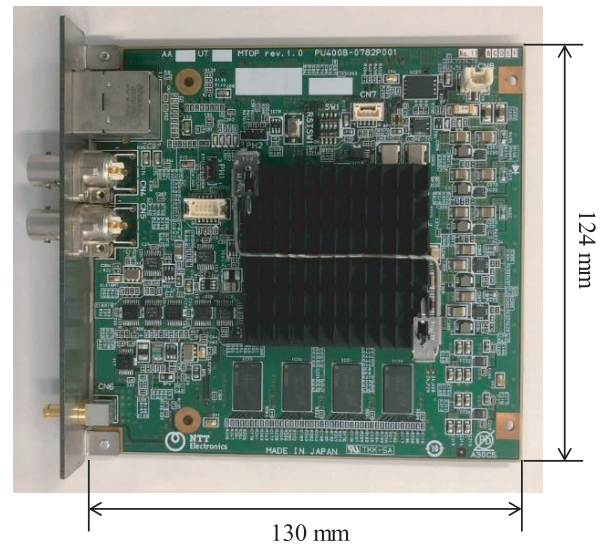

Fig. 7 Fabricated MMT module.

composed of even frames; each can be conveyed by quad 3G-SDI links. The HFR encoder provided the temporally scalable bitstream for the transmitter module in MPEG-2 TS. The base layer and the enhancement layer have different PIDs. The MMT transmitter generated the base layer packets and the enhancement layer packets, which were transmitted to the receiver over different paths. The MMT receiver reconstructed MPEG-2 TS using both layer packets and provide it for the HFR decoder. The bitrates of the base layer and the enhancement layer were respectively 40 Mbps and 20 Mbps. Figure 9 shows the decoded video. Fast-moving scene was presented smoothly without decoding errors. As a
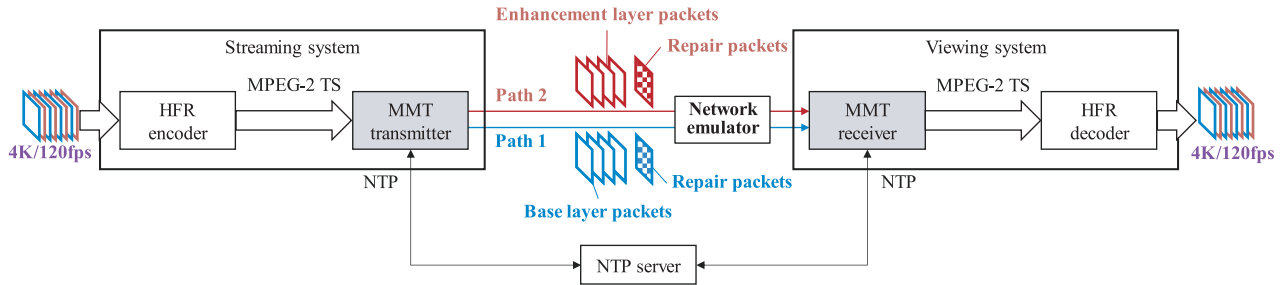
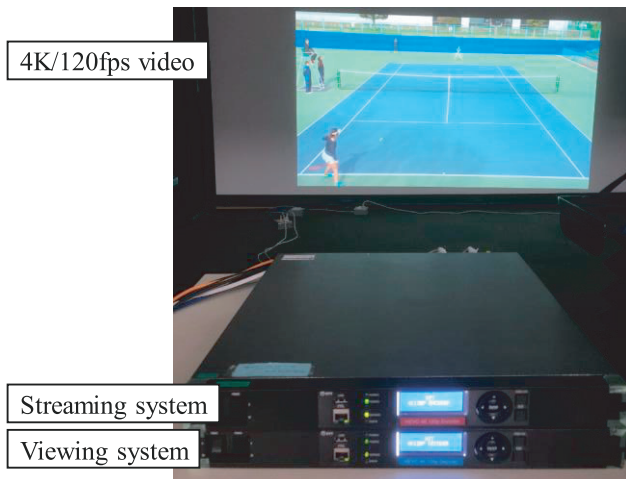**Fig. 8**   Experimental system for hierarchical transmission.

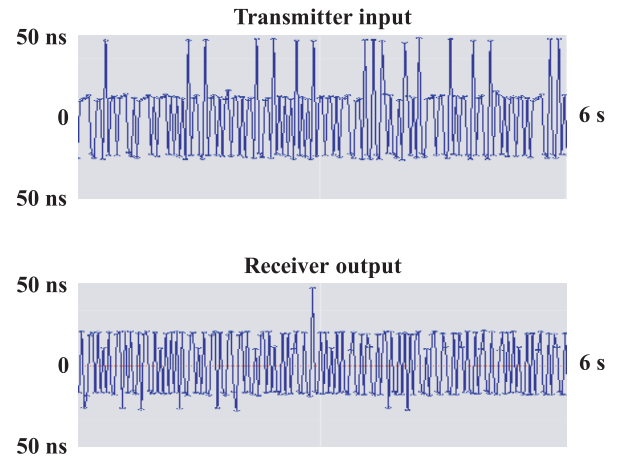**Fig. 9**   Decoded 4K/120fps video.

Fig. 10   PCR jitter.

result, we confirmed that 4K/120fps video was successfully decoded.

In order to evaluate the tolerance to network constraints, we inserted a network emulator in the transmission paths. First, 2% packet loss was introduced into both paths. The MMTP packets were protected by FF-LDGM with block size of 1000 packets and redundancy of 20%. Even under high packet loss ratio, 4K/120fps video was successfully decoded.

Next, in order to evaluate the receiver capability of absorbing the arrival timing offset and jitter, constant delay and delay jitter were inserted in the enhancement layer using the emulator. Then, we analyzed the MPEG-2 TS output of the receiver to investigate the propagation of the offset and jitter. If PCR jitter included in it is too large, the subsequent decoder cannot decode it since the decoder adjusts its clock depending on PCR. If the DTS of both layers are not close enough, the decoder buffer may overflow. Figure 10 shows PCR jitter under constant delay of 2000 ms and delay jitter up to 100 ms. In spite of the significant delay jitter in the transmission path, PCR jitter was under 50 ns, which was as small as that of the transmitter input. Figure 11 shows the DTS with reference to PCR. Although the MMTP packets of the enhancement layer was delayed for 2000 ms, the DTS of both layers in the receiver output were as close as those in the transmitter input. As a result, the MPEG-2 TS output

of the receiver was successfully decoded.

5.2   Discussion

The experiment demonstrated the MMT module functionality of transmitting 4K/120fps temporally scalable bitstreams over different paths successfully. MMT-based hierarchical transmission makes it possible to provide a scalable video distribution service that delivers the base layer to all decoders and the enhancement layer only to HFR decoders, which cannot be provided by MPEG-2 TS. The capability of dealing with severe network constraints was also demonstrated. Under these network conditions, video bitstreams cannot be transmitted by MPEG-2 TS since it is sensitive to the jitter. Consequently, excellent potential to support HFR video distribution service in various network situations was shown.

6.   Conclusion

In this paper, we presented the first ever MPEG Media Transport (MMT) module for 4K/120fps temporally scalable video. In order to transmit the base layer and the enhancement layer over different paths in accordance with the ARIB STD-B60 standard, the module is equipped with a newly proposed GOP boundary detection method of temporally scalable bitstreams. It is also designed to be tolerant to severe network constraints, including packet loss, arrival
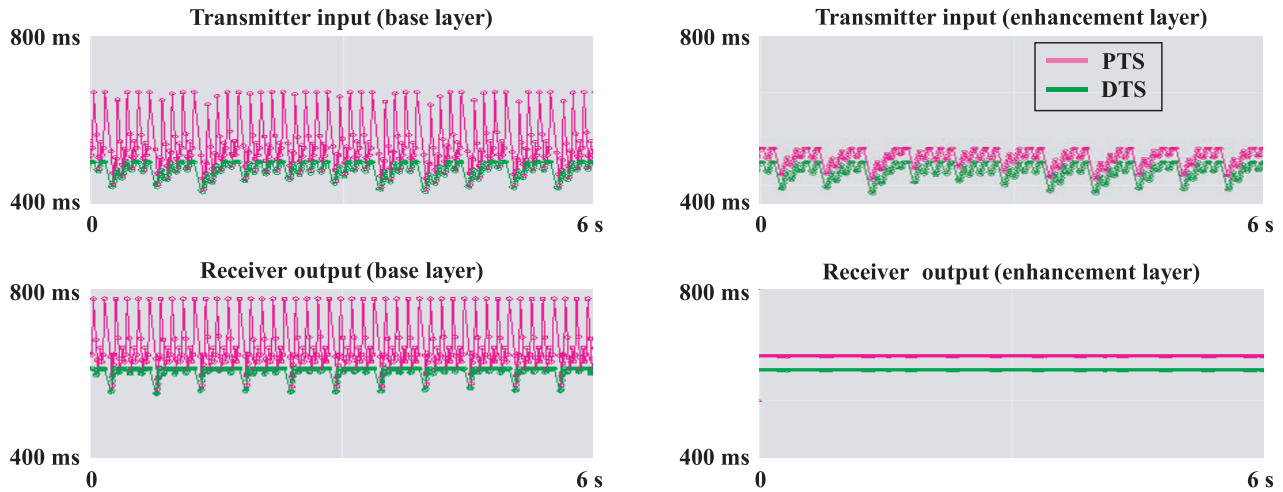
Fig. 11 DTS/PTS with reference to PCR.

timing offset, and delay jitter. An evaluation experiment demonstrated that the MMT module was successfully fabricated and that it has excellent potential as a means to support high frame rate video distribution in various network situations.

References

[1] Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding, ISO/IEC 23008-2:2017, Oct. 2017.

[2] J.-R. Ohm, G.J. Sullivan, H. Schwarz, T.K. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards — Including High Efficiency Video Coding (HEVC)," IEEE Trans. Circuits Syst. Video Technol., vol.22, no.12, pp.1669–1684, 2012.

[3] T.K. Heng, W. Asano, T. Itoh, A. Tanizawa, J. Yamaguchi, T. Matsuo, and T. Kodama, "A highly parallelized H.265/HEVC real-time UHD software encoder," 2014 IEEE International Conference on Image Processing (ICIP), pp.1213–1217, 2014.

[4] T. Onishi, T. Sano, Y. Nishida, K. Yokohari, K. Nakamura, K. Nitta, K. Kawashima, J. Okamoto, N. Ono, A. Sagata, H. Iwasaki, M. Ikeda, and A. Shimizu, "A Single-Chip 4K 60-fps 4:2:2 HEVC Video Encoder LSI Employing Efficient Motion Estimation and Mode Decision Framework With Scalability to 8K," IEEE Trans. VLSI Syst., vol.26, no.10, pp.1930–1938, Oct. 2018.

[5] Y. Miki, K. Takizawa, I. Masuhara, M. Miyazaki, A. Hanada, K. Mitani, M. Abe, S. Mitsuhashi, T. Sakiyama, and K. Ichikawa, "Ready for 8K UHDTV broadcasting in Japan," IBC 2015 Conference, p.9, Amsterdam, 2015.

[6] T. Yamashita, K. Masaoka, K. Ohmura, M. Emoto, Y. Nishida, and M. Sugawara, "Super Hi-Vision Video Parameters for Next-Generation Television," SMPTE Motion Imaging Journal, vol.121, no.4, pp.63–68, 2012.

[7] M. Emoto, Y. Kusakabe, and M. Sugawara, "High-Frame-Rate Motion Picture Quality and Its Independenceof Viewing Distance," J. Display Technol., vol.10, no.8, pp.635–641, 2014.

[8] Parameter values for ultra-high definition television systems for production and international programme exchange, ITU-R Rec. BT.2020-2, Oct. 2015.

[9] Video coding, audio coding, and multiplexing specifications for digital broadcasting, ARIB STD-B32 Version 3.9, Dec. 2016.

[10] Y. Sugito, S. Iwasaki, K. Chida, K. Iguchi, K. Kanda, X. Lei, H. Miyoshi, and K. Kazui, "A Study on the Required Video Bit-rate for 8K 120-Hz HEVC/H.265 Temporal Scalable Coding," 2018 Picture

Coding Symposium (PCS), pp.106–110, 2018.

[11] J.M. Boyce, Y. Ye, J. Chen, and A.K. Ramasubramonian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard," IEEE Trans. Circuits Syst. Video Technol., vol.26, no.1, pp.20–34, 2016.

[12] Y. Omori, T. Onishi, H. Iwasaki, and A. Shimizu, "A 120 fps High Frame Rate Real-time HEVC Video Encoder with Parallel Configuration Scalable to 4K," 2017 IEEE Symposium in Low-Power and High-Speed Chips (COOL CHIPS), pp.1–3, 2017.

[13] Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 1: MPEG media transport (MMT), ISO/IEC 23008-1:2017, Aug. 2017.

[14] Y. Lim, S. Aoki, I. Bouazizi, and J. Song, "New MPEG Transport Standard for Next Generation Hybrid Broadcasting System With IP," IEEE Trans. Broadcast., vol.60, no.2, pp.160–169, 2014.

[15] T. Nakachi, "A Media Transport Technology MMT in the Post-Ultra High Definition Age," IEICE ESS Fundamentals Review, vol.11, no.4, pp.288–301, 2018.

[16] Information technology – Generic coding of moving pictures and associated audio information – Part 1: Systems, ISO/IEC 13818-1:2018, March 2018.

[17] MMT-Based Media Transport Scheme in Digital Broadcasting Systems, ARIB STD-B60 Version 1.12, April 2018.

[18] ATSC 3.0 SYSTEM, ATSC A/300:2017, Oct. 2017.

[19] Y. Mochida, T. Yamaguchi, and T. Fujii, "MMT Converter for MPEG-2 TS Audio/Video," in ACM TVX2018 INDUSTRY, June 2018.

[20] Y. Mochida, T. Nakachi, T. Yamaguchi, T. Onishi, and K. Nakamura, "An MMT Module for 4K/120fps Temporally Scalable Video," 2019 IEEE International Symposium on Circuits and Systems (ISCAS), pp.1–5, 2019.

[21] RTP Payload Format for MPEG1/MPEG2 Video, IETF RFC 2250, Jan. 1998.

[22] Information technology – Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats, ISO/IEC 23009-1:2019, Dec. 2019.

[23] T.C. Thang, Q.-D. Ho, J.W. Kang, and A.T. Pham, "Adaptive streaming of audiovisual content using MPEG DASH," IEEE Trans. Consum. Electron., vol.58, no.1, pp.78–85, 2012.

[24] Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 10: MPEG media transport forward error correction (FEC) codes, ISO/IEC 23008-10:2015, April 2015.

[25] S. Aoki, Y. Kawamura, K. Otsuki, N. Nakamura, and T. Kimura, "Development of MMT-based broadcasting system for hybrid deliv-

ery," 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp.1–4, 2015.

[26] T. Nakachi, Y. Tonomura, D. Shirai, T. Yamaguchi, and T. Fujii, "Robust UHD video transmission using MMT FireFort-LDGM code," 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp.1–4, 2015.

Yasuhiro Mochida   received his B.E. and M.E. degrees from The University of Tokyo in 2009 and 2011. Since he joined NTT Laboratories in 2011, he has been engaged in research on video transmission over IP networks including transport protocol, presentation synchronization and video conferencing. His current research interest is in low-latency video transmission over high speed optical networks. He is a member of IEICE.

Takayuki Nakachi   received the Ph.D. degree in electrical engineering from Keio University, Tokyo, Japan, in 1997. Since joining NTT in 1997, he has been engaged in research on super-high-definition image/video coding, media transport technologies. From 2006 to 2007, he was a visiting scientist at Stanford University. He also actively participates in MPEG international standardization activities. His current research interests include sparse modeling, communication science, information theory and signal processing. He received the 26th TELECOM System Technology Award, the 6th Paper Award of Journal of Signal Processing and the Best Paper Award of IEEE ISPACS2015. He is a member of IEEE and IEICE of Japan.

Takahiro Yamaguchi   received his B. Eng., M. Eng. and Ph. D degrees in Electronic Engineering from the University of Electro-Communications, Tokyo, Japan, in 1991, 1993, and 1998, respectively. He joined NTT in 1998 and has been researching super high definition image distribution systems. He is currently a group leader of media processing systems research group in NTT Network Innovation Laboratories. He is a member of IEICE and ITE of Japan.