

## A 120 fps High Frame Rate Real-time Video Encoder

*Yuya Omori, Takayuki Onishi, Hiroe Iwasaki, and Atsushi Shimizu*

### Abstract

This article describes a real-time HEVC (High Efficiency Video Coding) encoder operating at 120 frames per second (fps) that is designed to achieve higher frame rate video services. The encoder achieves 4K/120 fps video encoding in real time through the synchronized operation of multiple 2K/120 fps encoders working in parallel. This encoder also makes it possible to achieve temporal scalable coding and transmission with upward compatibility for existing 60 fps based systems. This temporal scalability is expected to contribute to rapid expansion of the high frame rate video service field. The proposed encoder systems will also open the door to next-generation high frame rate ultra-high definition television services.

*Keywords: high frame rate, encoder, hardware*

### 1. Introduction

The latest video coding standard, H.265/HEVC (High Efficiency Video Coding), achieves double the coding efficiency of H.264/AVC (Advanced Video Coding), making it possible to provide higher definition video services economically. In recent years, 4K video broadcasting and distribution have become increasingly widespread, and realistic image representations are becoming increasingly popular. Such representations demand not only high spatial resolution but also many other factors (**Fig. 1**).

A high frame rate (HFR) improves the moving picture quality and is essential for creating more realistic image representations [1]. The next-generation television system specified in Recommendation ITU-R<sup>\*1</sup> BT.2020—Parameter values for ultra-high definition television (UHDTV) systems for production and international programme exchange—supports HFR formats up to 120 fps. For the spread of HFR video services, temporal scalable coding with upward compatibility for current 60 fps based video services is also necessary for encoders. Several HEVC real-time encoders have been developed to enable HEVC encoding of over-4K images. However, they are only

capable of encoding images up to 60 fps due to the increase in computational complexity and the need for temporal scalable functionality.

This article presents our new 4K/120 fps HEVC encoder, which achieves HFR 120 fps real-time encoding and 120/60 fps temporal scalable functionality by exploiting our 4K/60 fps HEVC codec large-scale integrated circuit (LSI) called NARA, which stands for Next-generation Encoder Architecture for Real-time HEVC Applications [2].

When the flexible customizable software architecture of the NARA LSI is utilized, merely modifying the custom functions layer of the NARA software architecture makes it possible to achieve temporal scalability such as upward compatible reference picture structures and dual-stream bit-rate control functions. An encoder equipped with one NARA LSI has 2K/120 fps encoding capability, and synchronized operation of multiple encoders working in parallel also achieves scalability to larger 4K/120 fps input images. These scalable functionalities of the proposed encoder will contribute to the development of

<sup>\*1</sup> ITU-R: International Telecommunication Union - Radiocommunication Sector

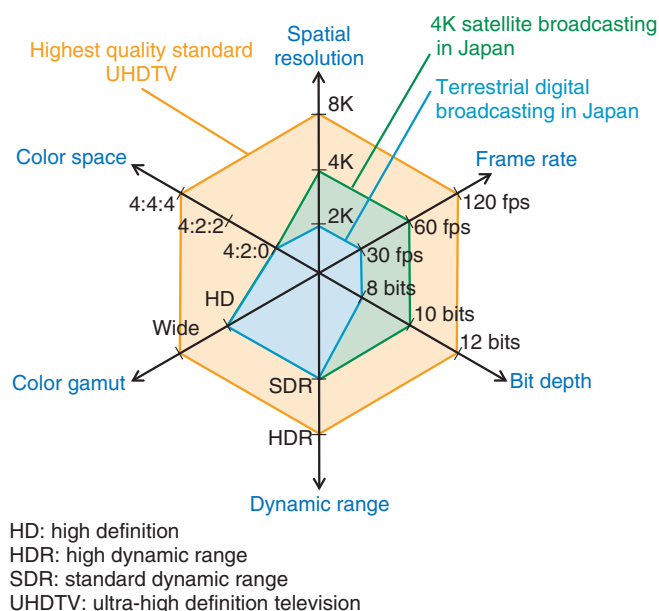


Fig. 1. Factors required for high quality video.

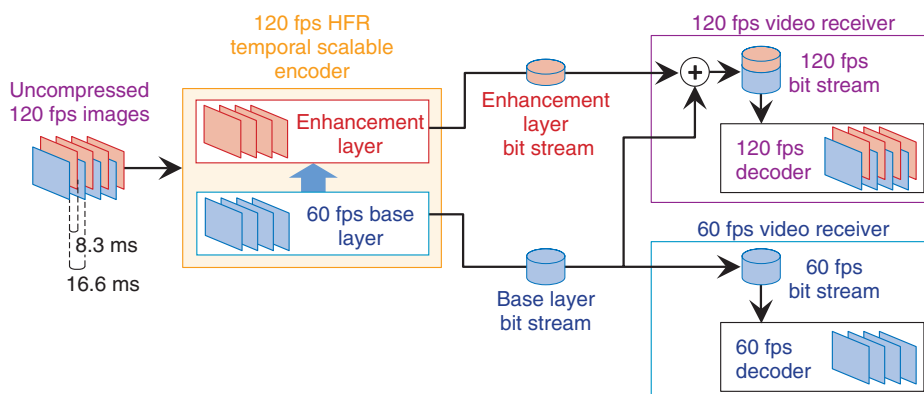


Fig. 2. HFR temporal scalable video distribution.

new broadcasting and distribution systems for the upcoming UHDTV services.

## 2. Encoder system architecture for HFR temporal scalability

The Association of Radio Industries and Businesses (ARIB), an incorporated association promoting the practical application and dissemination of radio systems in Japan, regulates 120/60 fps temporal scalable formats as an HFR scalable coding standard based on HEVC [3]. The ARIB standard specifies that 120 fps

HFR bit streams must have temporal scalability, and that encoded picture data are distributed into a 60 fps base layer stream and an enhancement layer stream, as illustrated in Fig. 2. Dual-stream bit-rate control must be performed for both the base layer and enhancement layer streams to ensure constant bit-rate encoding and distribution for both 60 fps and 120 fps decoders. In addition, base layer images and enhancement layer images should be periodically received and decoded one by one alternately to prevent deviation of the decoding time for 60 fps decoders. This limitation of the decoding time leads to changes in the

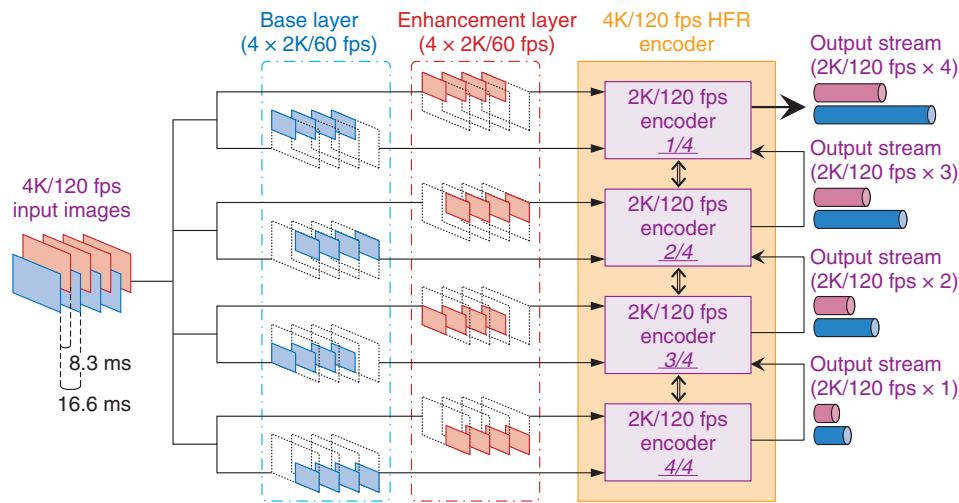


Fig. 3. System configuration.

reference structure of inter-frame prediction.

The temporal scalable encoding function is added to the existing NARA LSI by utilizing the flexible customizable software architecture of a NARA LSI with large motion search capability. The LSI's software architecture consists of three hierarchical layers; the top function layer is the software for handling fundamental HEVC functions and user functions. This software hierarchy not only solves the complexity of tediously controlling the hardware using a low-level interface and the difficulty of handling HEVC common basic functions, but also provides a simple programming interface as custom functions of the top layer. Temporal scalability essentially requires complicated modifications of the LSI's encoding method. However, because of this software architecture, the dual-stream bit-rate control and the reference structure modification are achieved in some of the custom functions, and thus, they can be easily customized with higher level programming.

### 3. System configuration

The system configuration of the 4K/120 fps encoder we have developed is shown in **Fig. 3**. It consists of four 2K/120 fps encoders. Each 2K/120 fps encoder includes one NARA LSI and handles the encoding of one of the 2K images, which is squarely divided from the original 4K input images. The 2K/120 fps encoder receives a 2K/120 fps input video as two sets of 2K/60 fps video sequences, which correspond to the base layer and the enhancement layer, by using a

multi-channel input functionality of the NARA LSI. The encoder rearranges the two 2K/60 fps sequences to form one 2K/120 fps image, and all 2K/120 fps encoders operate cooperatively by exchanging synchronization signals in order to share the common clocking and time stamp values to form one 4K/120 fps encoder.

The output stream is constructed in the MPEG-2 Transport Stream (TS)<sup>\*2</sup> format so that the base layer stream and the enhancement layer stream have different video packet identifiers. Here, the NARA LSI's multi-channel video output function is effectively utilized. Four 2K/120 fps transport streams can be transmitted in parallel, and they can also be transmitted with one multiplexed stream by using the encoder's multi-channel multiplex and output functionality with cascaded transport stream input/output connections, as illustrated in **Fig. 3**.

### 4. Implementation

A photograph of the overall 4K/120 fps encoder with four 2K/120 fps encoder devices is shown in **Fig. 4**, and the encoder specifications are listed in **Table 1**. This encoder inputs a 4K/120 fps video with eight 3G-SDIs (third-generation serial digital interfaces), and outputs an HEVC temporal scalable stream in the MPEG-2 TS format by using a DVB-ASI

<sup>\*2</sup> MPEG-2 Transport Stream: A Moving Picture Experts Group standard digital container format for transmission and storage of audio, video, and Program and System Information Protocol (PSIP) data.



Fig. 4. Photograph of encoder.

Table 1. System specifications.

Video format	3840 × 2160 pixels / 120 fps
Video coding	HEVC Main 10 profile
Container format	MPEG-2 Transport Stream
Input	3G-SDI × 8
Output	DVB-ASI or 1000BASE-T

(digital video broadcasting - asynchronous serial interface) or by using Internet protocol (IP) connections. We observed that the 4K/120 fps bit streams encoded by our encoder were successfully decoded and played by other HFR systems.

A photograph taken during a demonstration of 4K/120 fps real-time transmission with our HFR encoder is shown in Fig. 5. Uncompressed 4K/120 fps images were encoded at a constant bit rate of 80 Mbit/s, where the bit rate of the base layer was 60 Mbit/s and that of the enhancement layer was 20 Mbit/s and images were distributed over an IP connection. The other real-time 4K/120 fps software decoder received the compressed bit stream, which was demultiplexed from MPEG-2 TS and decoded in real time. Then the 4K/120 fps decoded images were displayed on the screen by the 4K/120 fps-enabled projector.

No 120 fps video services have been available to date due to the enormous video data size and the requirement for compatibility with legacy systems. This has made it difficult to achieve 120 fps video



Fig. 5. Real-time encoding demonstration.

distribution. Our device solves this problem and provides advanced high quality HFR video by combining it with other existing 4K/120 fps-enabled devices such as a camera, decoder, and projector.

## 5. Conclusion

We presented a new 120 fps real-time HEVC encoder for higher frame rate video encoding and transmission exploiting the existing HEVC encoder LSI. This 120 fps encoder makes it possible to achieve temporal scalable HEVC coding and transmission with upward compatibility for 60 fps. We plan to continue developing video coding devices to contribute to the provision of new services.

## References

- [1] M. Emoto, Y. Kusakabe, and M. Sugawara, "High-frame-rate Motion Picture Quality and Its Independence of Viewing Distance," *J. Disp. Technol.*, Vol. 10, No. 8, pp. 635–641, 2014.
- [2] T. Onishi, T. Sano, Y. Nishida, K. Yokohari, J. Su, K. Nakamura, K. Nitta, K. Kawashima, J. Okamoto, N. Ono, R. Kusaba, A. Sagata, H. Iwasaki, M. Ikeda, and A. Shimizu, "Single-chip 4K 60 fps 4:2:2 HEVC Video Encoder LSI with 8K Scalability," *Proc. of 2015 Symposium on VLSI Circuits (VLSI Circuits)*, C54-C55, Kyoto, Japan, June 2015.
- [3] ARIB STD-B32: "Video Coding, Audio Coding, and Multiplexing Specifications for Digital Broadcasting," Version 3.8, Sept. 2016.



**Yuya Omori**

Researcher, Visual Media Coding Group, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in information and communication engineering from the University of Tokyo in 2012 and 2014. He joined NTT Media Intelligence Laboratories in 2014 and has been engaged in research and development (R&D) of parallel processing architecture and a high efficiency video codec algorithm. He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and the Institute of Electronics, Information and Communication Engineers (IEICE).



**Takayuki Onishi**

Senior Research Engineer, Visual Media Coding Group, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in information and communication engineering from the University of Tokyo in 1997 and 1999. He joined NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 1999 and has been engaged in R&D of high quality image coding and transmission. He is a member of IEEE and IEICE.



**Hiroe Iwasaki**

Senior Research Engineer, Supervisor, Visual Media Coding Group, Visual Media Project, NTT Media Intelligence Laboratories.

She received a B.S. and Ph.D. in information science from Tsukuba University, Ibaraki, in 1992 and 2006. She joined NTT Switching System Laboratories in 1992. She has been involved in R&D of an architecture and real-time operating system for multimedia embedded system LSIs. She moved to NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 1997. She is a member of IEEE, IEICE, and the Information Processing Society of Japan.



**Atsushi Shimizu**

Senior Research Engineer, Supervisor, Visual Media Coding Group, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in electronic engineering from Nihon University, Tokyo, in 1990 and 1992. Since joining NTT in 1992, he has been working on video compression algorithm and software development. He is a member of IEICE, the Institute of Image Electronics Engineers of Japan, and the Institute of Image Information and Television Engineers.

---