Research Article

A Low Delay Scalable HEVC Encoder with Multi-chip Configuration for HFR Video

L. Balaji[1,*] and A Dhanalakshmi[2]

[1]Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr Sagunthala R & D Institute of Science and Technology, Chennai, India
[2]Department of Electronics and Communication Engineering, Panimalar Engineering College, Chennai, India

_____

Abstract

This paper talks about real time 4096-pixel HEVC with 120 fps (frames-per-second) based scalable encoder for the purpose of HFR (high-frame-rate) surveillance videos that undergo encoding, transmission, and further processing. Motion depiction issues, for example, motion-blur and jerkiness might happen in videos that include speedy camera panning or quick-moving objects. A high-frame-rate (HFR) tackles such issues and gives an immersive viewing experience, which will precise even the quick-moving objects without uneasiness. Encoding real-time HFR videos with temporal scalability and minimal latency is needed to offer for HFR video services. A scalable multi-chip configuration is implemented in the proposed encoder that uses two large scale integration encoders of two 4096-pixel 60-fps to accomplish a complete 4096-pixel 120-fps. Moreover, coding efficiency is maintained by means of exchanging the reference picture information closer to the spatial slice boundaries that offers cross-chip motion estimation. The scalable encoder holds-up the bit-stream output with temporal-scalability, which is delivered across a single or double transmission channel. With the usage of motion vector restrictions, the complexity of the scalable encoder is reduced extensively, and achieves a low delay of 21.1 ms.

Keywords: HEVC Scalable Extension; HFR Video; Scalable Multi-chip configuration; motion vector restriction; low delay.

_____

1 Introduction

ISO/IEC H.265/HEVC video coding has been widely used for 4K video-broadcasting, delivery, and required for accurate image representations. Such images are identified with its HFR (high frame rate) with 4096-pixel to be more realistic due to significant improvements in the compressions of video. H.264 (also so-called AVC), the industry norm for compressions of videos that will allow contents of digital-formats required to be recorded, stored and delivered, is more useful for a long period. Though things change rapidly, the term HFR refers to high-frame-rate. HFR correlates to each single topic which is recorded at a higher frame-rate than the usual 24 frames-per-second (fps), which is utilized for the high rate of films. It additionally produces soft-motion expression as well as it encourages subjective motion image quality, preventing motion picture issues such as motion distortion and bounciness. By utilizing the performance of HEVC coding efficiency, not only 4096-pixel and / or 8192-pixel formats, but also offers up to 120-fps for HFR formats remain aided through UHDTV (ultra-high-definition-television). For the quick reach of HFR in videos like 120-fps, the HFR encoders offered higher coding capability and regressive compatibility to the existing 60-fps video systems. This denotes the video streams that are encoded with HFR must be decrypted, processed using bequest non-HFR video appliances. Such temporal scalability featured coding is stated in HEVC scalable extension [1] and makes it feasible to stop the necessity to allocate several bit-streams for similar videos with variable frame-rates. For instance in Japan, where 4096-pixel 60-fps audio-visual delivery was

initiated previously and its usage of set-top-boxes with 60-fps are getting popular. ARIB (Association of Radio Industries and Businesses) stated 120-fps setups to 60-fps with temporal-scalability for HEVC digital broadcasting [2] with HFR coding norms. ARIB stipulates an enhanced satellite communication model such that the transponder contains peak capability of 100-Mbps [3]. It is typically believed a single transponder supplies to more than two channels of 4096-pixel video bit-stream and bit-rate of new 4096-pixel 60-fps is 30-Mbps; hence the bit-rate rises from 60-fps to 120-fps is supposed to be less than 2/3 of 60-fps bit-rate [4].

Numerous HEVC coded LSI chips have been established to assist real-time coding of UHDTV pictures [5] [6]. High-compression, better quality, and higher resolution coding of 4096-pixel video in real-time with no instabilities is accomplished by LSI establishing the leading edge H.265/HEVC. HEVC 4096-pixel 120-fps HFR real-time encoders have three most important constraints, 1) Multi-chip design: By utilizing multi-chip structural design, complete 4096-pixel/120-fps real-time encoding can be accomplished. Multi-chip design retaining the coding efficiency is much needed. 2) Temporal-scalability: 60-fps in the present standard decoder is essential for a 120-fps encoder. 3) Low delay (reduced computation): Low delay is important for an encoder that deals in video contribution with original scenes. In the distant process, the expectation on the usage of HFR video-services, requires an encoder with coding capability of minimal delay.

The HEVC coding algorithm uses temporal-correlation as well as coding-correlation among the frames on the coding-level-unit; hence, it is initially hard to perform the specification changes associated to sequential aspects like frame-rate (temporal-scalability) for the present coders. Here

_____

begins three key constraints engaged in creating 4096-pixel 120-fps HFR HEVC scalable extension real-time encoder together adopting temporal-scalability, 1) Adaptable encoder-control operability to assist customizable reference image formats and double-stream bit-rate regulates both 120-fps, 60-fps systems at same time. 2) Parallel procedure operability of several encoders to allow 120-fps high resolution video coding in real-time. 3) Increase low bit-rate from 60-fps to 120-fps that allows a single transmission channel to deliver several HFR bit-streams.

The motivation behind the work is to develop a 4096-pixel 120-fps HEVC encoder with multi-chip configuration that can achieve low latency and high coding efficiency for high-motion video content. The existing video encoding technologies struggle to maintain high quality and low latency for high-motion scenes, such as those found in sports broadcasts and virtual reality applications. The proposed encoder aims to address these challenges and provide a more immersive viewing experience for users.

2 Background and Related Work

Currently, systems with a wide field-of-view (FOV) and higher spatial resolution than conventional systems are discussed [7] to provide a viewing experience. Though, some study has tested the effect of raising the time resolution or frame rate of the TV system. In the future, on home TVs with a broad FOV, the effect of increasing frame rate and viewing distance by subjectively evaluating the quality of moving images will be quantified. When the frame rate is increased from 60-fps to 120-fps, the quality of moving pictures improves further, and when increasing from 120-fps to 240-fps (3H) (three times the image height), the quality of moving pictures has been improved further.

The quality of moving pictures is not statistically significant, and the degree of improvement depends on the content of the image. Due to the reduced visibility of motion artifacts, video frame rates higher than normal usage have shown improved perceptual quality. Especially motion blur and temporary anti-aliasing. Regardless, the frame rate used in TV and movies has remained the same for many years. Although not widely used today, the new video standard for UHDTV [8]. It defines a broader dynamic range and higher spatial resolution than its previous one; besides supports 120-fps. However, due to anti-blurring and variation in frame-rate, it does not offer the best film quality. In order to avoid these shortcomings, a psycho physical testing was performed [9]. The results show that 120-fps is a good improvement over 60-fps, and the best improvement of saturation estimation is 240-fps.

Using HEVC single-chip video function (4096-pixel 60-fps 4:2:2 high-efficiency video encoding) and the ultra-high scalability of 8K TV, a professional LSI video encoder with image quality for content transmission and distribution has been developed [10]. Statistical adaptive motion estimation (ME) with multiple blocks effectively reduces the high computational complexity of HEVC for real-time processing, and the highly concentrated decision framework in this mode supports high compression performance. The distributed motion vector search chip imaging and high-speed data bus communication can also be used to establish 8K encoding with multiple chips. As a subjective evaluation, it is found that the discussed LSI encoder maintains the same visual quality as the LSI encoder conforming to the above-mentioned standards and reduces the bit rate. The chip is designed and manufactured using 28nm CMOS technology and has been used to create industry-leading 4K and 8K transmission systems.

Future video codec processors must take over the lately standardized H.265, since H.264 is limited in terms of many aspects for UHD images, providing a less complex H.265 video encoder (codec processor). This type of codec processor is in the developing stage under Samsung's 28nm CMOS process. The size of this low-complexity codec is still limited to the traditional H.264 chips. The single-core processor is optimized for RDO mode and bypass mode, and has low power consumption. Compared with HM-13.0, the speed loss is 35 percent and the power consumption when switching to bypass mode is less than 250 mW. The size of the chip is $7.3 \times 5mm^2$ and its internal SRAM is 300KB. The maximum operating frequency is 600MHz at 30-fps encoding in 4K UHD [11].

Japan Broadcasting Corporation (NHK) [12] is undergoing research and development on Super Hi-Vision, which is a 7680×4320 pixel version of Ultra High Definition TV (UHDTV2) for future broadcasting. Previously, NHK and Shizuoka University jointly developed an optional 120-fps, 33-megapixel metal oxide semiconductor image sensor (CMOS) and corresponding 12-bit A/D converter. It provided the highest resolution and frame rate according to the recommendations of the International Telecommunication Union's 2020 Broadcast (TV) Series. NHK has also developed a 3-chip color camera with 120-fps equipped with this image sensor. In this document, NHK recently developed a 33-megapixel 120-fps single-chip color CMOS image sensor and a compact UHDTV2 camera equipped with this image sensor. The width, height, and depth are approximately $5 \times 5 \times 6$ inches, respectively. It weighs only 4.5-pounds, which is much smaller and lighter than the traditional UHDTV2 camera. The optical size of the image sensor is 25-mm in diameter, and the camera is compatible with Super 35-mm Positive Mounting lenses suitable for various digital cinema cameras.

The technology of digital terrestrial broadcasting involves transmission of digital video signals and its coding systems from the broadcasting stations of different categories such as HDTV and SDTV broadcasting [13]. Satellite broadcasting stations use frequency bands more than 11.7 GHz and less than or equal to 12 GHz for television broadcasting, VHF broadcasting and data transmission. 2 GHz (referred to as BS digital broadcast transmission), SDTV, HDTV, VHF radio, and data on broadband transmission systems on controlled satellite broadcasting stations. Broadcasting is called CS Broadband Digital Broadcasting, which corresponds to the "standard digital broadcasting system between standard television broadcasting, etc" [13].

It is the new 4K real-time, low-latency, 120-fps HEVC decoder [14] designed with a parallel processing architecture, which complies with the HEVC Main 4:2:2 10 profile. With the bit-rate of the decoded streaming video, the decoding process can be parallelized and aligned at the frame, slice, and row levels of the coding tree unit. The discussed decoder [14] is implemented in three Arria 10 series FPGAs running at 133 and 150 MHz, and uses 4096-pixel encoder (120-fps).

So far the above discussed codec finds shortcomings on providing bit-stream output of 300 Mbps and with 37 ms and or 21.8 ms end-to-end delay. Also these codec has not been implemented for the scalable extension of HEVC.

The development of a low delay scalable HEVC encoder for high frame rate (HFR) video is important because it can provide a more immersive viewing experience with its high frame rate capabilities. HFR video encoding and transmission solve the problems created by fast-moving scenes, such as motion blur and jerkiness, and improve subjective moving picture quality. Additionally, a low delay scalable HEVC encoder can reduce the latency between the input and output of the video, which is important for real-time applications such as live streaming and video conferencing. This can improve the overall user experience and make the technology more practical for a wider range of applications.

This suggests that the technology could be useful in industries that require high-quality video, such as entertainment, sports, and gaming. Additionally, the low delay aspect of the technology could be useful in real-time applications such as live streaming and video conferencing, which are used in a variety of industries including education, healthcare, and business. Development of a low delay scalable HEVC encoder for HFR video is important for providing a more immersive viewing experience and reducing latency in real-time applications. Next Section describes the proposed framework of the scalable Multi-chip configuration integrated with LSI.

3 Proposed Framework

The proposed encoder system achieves 120-fps temporally scalable HEVC encoding for existing 60-fps based systems through modification in the customizable software architecture of encoder LSIs. The encoder also achieves 4096-pixel 120-fps video encoding in real-time through the synchronized operation of multiple 2000-pixel 120-fps encoders working in parallel. Also the proposed encoder system provides a simple programming interface as custom functions of the top layer, which eases the complexity of controlling the hardware and handling HEVC common basic functions. These design choices are significant because they enable the proposed encoder system to achieve high frame rate video encoding with improved efficiency and reduced latency, while also providing a user-friendly interface for customization and control.

A multi-chip configured low delay HEVC scalable encoder for HFR video service is proposed. Initially, a 120-fps HFR scalable HEVC encoder developed with multi-chip configuration. It constitutes two 60-fps encoder LSIs with 4096-pixel. Cross-chip motion estimation is provided across spatial-slice boundaries for information exchange about the reference pictures. This development is used to perform temporal scalability with minimal delay. A temporal scalability stream will be transmitted over a single or double transmission channel that attains minimal delay using motion vector restriction that will be described in the following subsections.

3.1 Scalable Multi-chip Configuration
A full 4096-pixel 120-fps real-time encoder is obtained through 2 chips of 4096-pixel 60-fps encoder LSI. Fig. 1 depicts various blocks of LSI with PCI Express (PCIe) data I/O interface. This interface allows both LSIs to be linked and shares different data that are necessary for encoding multi-chip picture information, quantized data in all the slices.
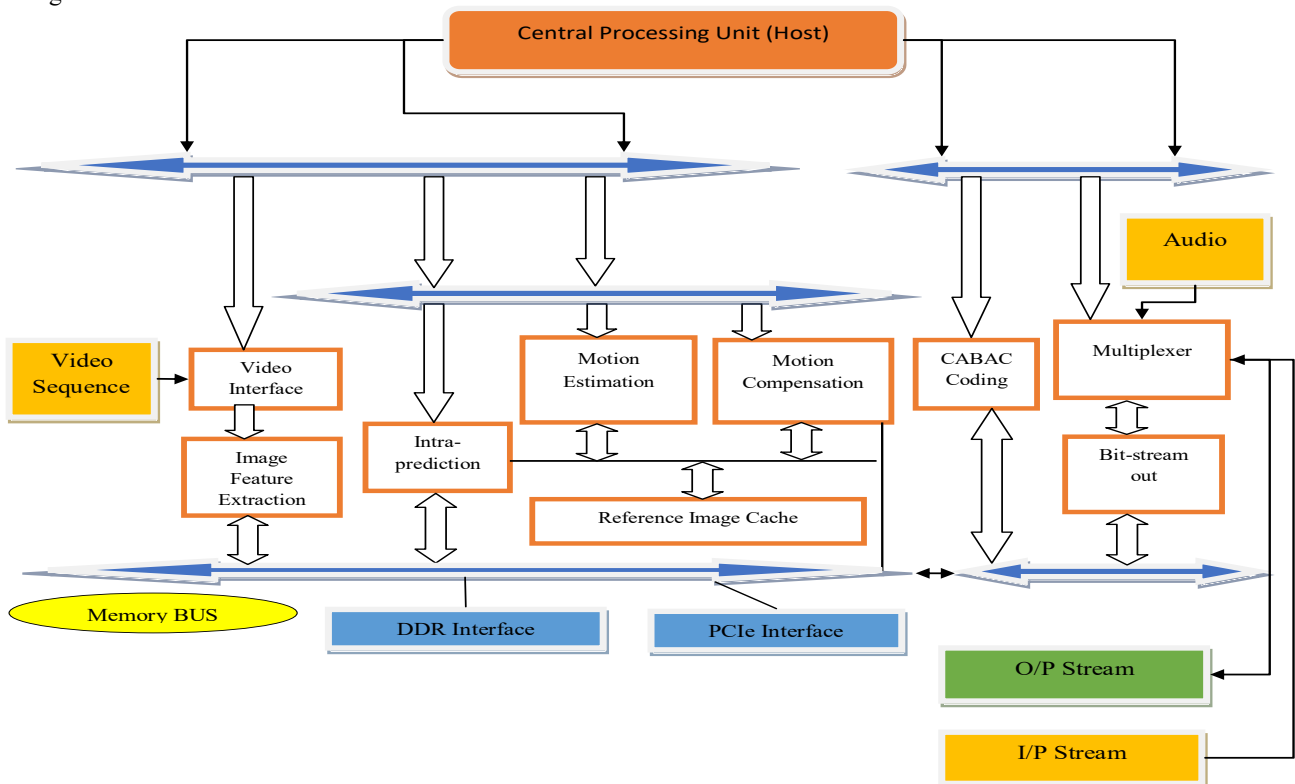

Fig. 1. Encoder LSI block diagram

Two other methods are available to handle in parallel by means of both the LSIs, such as temporal and spatial video divisions respectively, as in Fig. 2. The 120-fps video can be split into two 60-fps and each one can be encoded with a single LSI encoder in temporal video division. As both LSI runs at 4096-pixel 60-fps as normally, this shows the benefit of requiring only slight changes to the existing LSI behavior. However, HEVC estimates motion with regard to other frames, all the encoded 4096-pixel image data must be

moved to other LSI. Within 1/120 secs, 20.7 MB of picture information is required to be transmitted in each direction, but the PCIe data bus with 2.5 GBps data rate makes it clumsy. On the other hand, during spatial video division, the input frames are split into two slices, each of which is encoded using single LSI at 120-fps. Despite the fact that spatial separation necessitates changes to an LSI's actions, similar to changing the size of encoded frame in one chip attains the benefit of data-transmission among the chips. As a result, for two-LSI parallel processing, we choose spatial division. Coding efficiency will not be achieved, if the coding process is enclosed with every spatially separated slice. The key reason for this degradation is that motion estimation and compensation cannot be performed on the motion vector around the slice boundary. On sharing a single portion of reference picture information close to the boundary, this proposed scheme avoids degradation in coding performance with minimal data transfer.
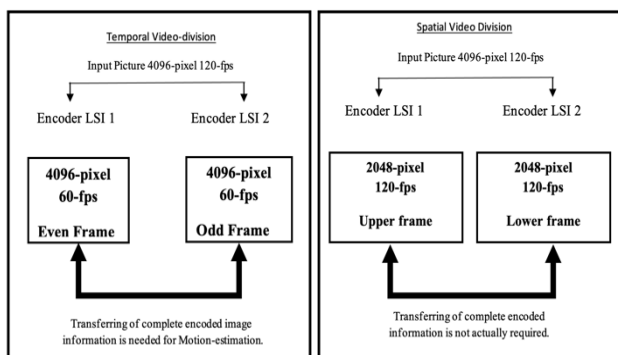

Fig. 2. Temporal and Spatial video-division

3.2 Spatial vs Temporal division

Spatial division was preferred over temporal division for parallel processing with two LSIs because it has several advantages over temporal division. Temporal video division requires a minor customization of the existing LSI behavior, but in the case of scalable coding mode, all the encoded data of odd frames needs to be transferred to the other LSI responsible for encoding the even frames. This can result in a significant amount of data transfer and may increase the complexity of the system. On the other hand, spatial video division can be easily implemented and does not require significant modifications to the existing LSI behavior. However, compared to the non-division case in which all encoding is performed virtually in one LSI, spatial video division may degrade the coding efficiency because motion vectors across the slice boundary cannot be computed in motion estimation and compensation. The proposed system suppresses such coding efficiency degradation with a small amount of data transfer by exchanging only part of the reference picture data near the boundary.

The efficiency of cross-chip motion estimation and compensation in multichip configuration is depicted in Fig. 3. One LSI encodes the split half-size 4096-pixel video. Horizontal motion is more likely to occur than vertical motion, so horizontal division is being used. The data which is encoded are decoded locally within the encoder. This decoded data is saved in external DDRs that are put together to every LSI. The memory bus and the PCIe interface are responsible for shifting the decoded data from one LSI DDRs to another LSI DDRs. Yet, there is no restriction on motion vectors among the LSIs. Major motion vectors

around the slice-boundary fall inside a specific altitude range upright from the boundary. As a result, motion vectors from the encoded block in another slice do not often reach the top portion of the upper-slice or the bottom portion of the lower-slice. And merely a portion closer to the slice boundary, the reference picture data is necessary to maintain each slice's coding quality.
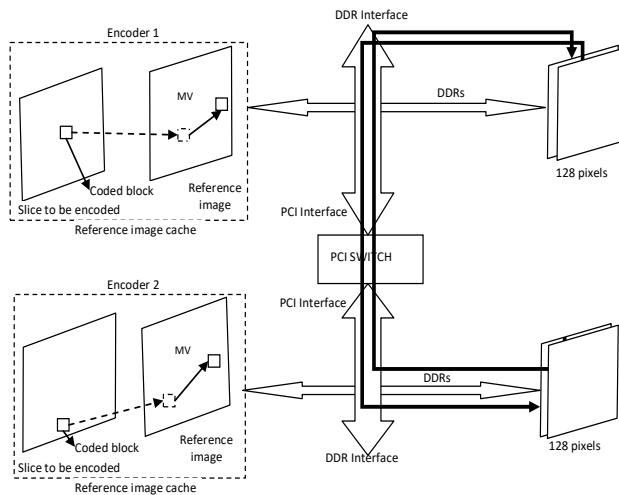

Fig. 3. Motion-estimation and motion-compensation over slice-split boundary

The reference picture information that lies within the altitude range of the slice boundaries will be shared by both the LSIs, once the encoder instructs. The height is fixed to 128-pixel during the regular case, in view of other case data transfer amounts in the PCI interface for the encoder. With just 1.2 MB of picture data is transferred per channel, which is only 5.9 percent of complete 4096-pixel picture transfer. Motion estimation is performed by all LSIs with the reference pictures of another slice that is transferred during slice encoding. This I/O functionality of LSI data allows cross-chip motion prediction that leads to multi-chip, and also increases coding performance.

3.3 Hardware and Software Architecture

The proposed encoder system's hardware architecture includes a dedicated hardwired encoding core, which is controlled by software at different levels of hierarchy. The software architecture consists of three major layers: the hardware layer, the hardware control layer, and the function layer. The hardware control layer includes PRISC (Prediction core RISC) software that controls the encoding core with close proximity to enable accurate picture quality control at HEVC's coding-unit-level, and MRISC (Middle-level RISC) software that controls the other part above the slice-level. The function layer includes TRISC (Top-level RISC) software for handling HEVC fundamental functions and user customizable functions. The software hierarchy eases the complexity of controlling the hardware using a coding-unit-level interface and the difficulty of handling HEVC common basic functions, while also providing a simple programming interface as custom functions of the top layer. Two parameter sets in the custom function are set for temporal scalability, and they are stepwise broken down to the hardware layer controlling the HEVC core. Overall, the proposed encoder system's hardware and software architecture are designed to achieve high frame rate video encoding with improved efficiency and reduced latency, while also providing a user-friendly interface for customization and control.

3.4 Scalable Bit-stream

HEVC scalable extension includes temporal scalable coding, which will be complex to handle in view of increased HFR video services. To compensate for temporal scalability, the HFR video-distribution for temporal scalability is designed.
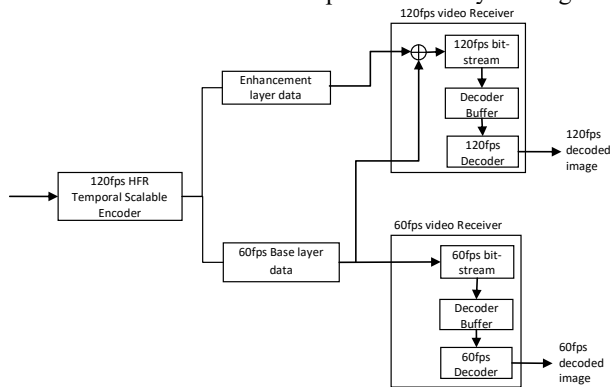

Fig. 4. Temporal Scalable video-distribution for HFR

Fig. 4, shows the video-distribution of HFR for temporal scalability with 120-fps encoded picture data. The picture data is made up of two packet data types: one 60-fps packet data for base layer and 60-fps packet data for enhancement layer. The decimated 60-fps video is played by a regular 60-fps system that only picks the packet data of the base layer and also decodes it as well. Both base and enhancement layer's packet data could be sent as a single stream through the same transmission channel or individually through two other transmission channels. Merely the base layer packet has to be particularly chosen by the 60-fps decoder when data transmission is in the same direction. When data must be transferred over 2 channels, 2 streams are required to be created, with a packet for each individual channel. Henceforth, the encoder accomplishes temporal scalability, thanks to an FPGA implementation that transmits a temporally distinct packet from single or double video output on performing double bit rate control.
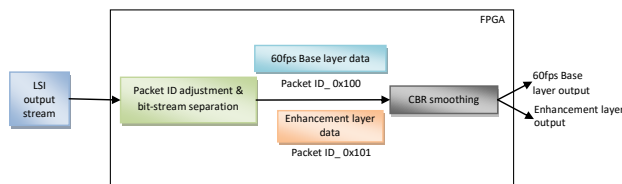

Fig. 5. Temporally-scalable transmission structure

Fig. 5 shows the structure of temporal scalable-transmission with two channel outputs representing one for base layer and another for enhancement layer information. They are interleaved with no difference in LSI output. Fig. 4 illustrates temporal scalable video-distribution for HFR video into enhancement layer data and 60-fps base layer data. Enforcing temporal scalability, a single stream is divided into two streams with the designed FPGA. It divides the arriving stream into each layer by changing the stream's packet Ids. Regardless of whether there is single or double output, the device sets the packet ID to dissimilar values for every layer class in compliance with the common one. It monitors sensitive data on the packet header to decide if the packet belongs to the base or enhancement layer. Once the packet data belongs to the enhancement layer, the ID will be changed to a different value, allowing the 60-fps decoder to categorize between the base layer and enhancement layer packets. The encoder then outputs a video stream with double layered information at each constant bit rate (CBR).

For complete utilization of the transmission channels in bandwidth-limited conditions, the CBR transmission channel is required, and the decoders receive buffers that are prepared under CBR conditions. Smoothed streams are received as outputs from the encoder at each defined bit-rate for the verification of buffer in each layer, taking into account equally 60-fps and 120-fps decoders. Both streams will be multiplexed into one as a single stream, if only one output is available.

3.5 Motion Vector Restriction and Encoding

The bidirectional process of inter-frame prediction and the buffers (sending and receiving) waiting time are the most common causes of codec delay. Gradual decoder refresh (GDR) is commonly employed to lessen the codec delay. An intra-column shift is adopted to avoid the delay in frame rearrangement by referring only to the past images as per the order. A column of intra-blocks alone is made up without utilizing the intra-picture is set, and the position of the column is circulated via every frame. Since every frame has similar coding conditions, the variations in the amounts of frames coded are restrained. The amounts of block-line code variations are also restrained by employing intra-column. As a result, the time required to fill the send buffer is reduced. A discrepancy in the reference image occurs due to intra-column shift between the codec. It will not occur for the encoded blocks that are located right to the column. For the blocks that are available left to the column offers a pointing vector to the right. The pointing vector needs to be referred, prior to intra-column refresh. Subsequently, an unsuccessful image being decoded is generated partially at the decoder start time, since the decoder processing starts at any time. Thus, the proposed encoder performs a modification on the encoder LSI motion estimation processing to restrict motion vector position.

The motion estimation of LSI contains multiples of hardware cores, and performs stepwise processing using dissimilar search ranges and dissimilar pixel accuracies. As a result, modification is done on each core to fix motion vector restriction. A pre-motion broad-range search core fixes a search area with a limited position to avoid protruding from another column. The search area moves to another column (right), if the block is on the opposite column (left). One more search core that uses fractional pixel positioning accuracy searches the vector position estimated by the pre-motion broad-range core. Although with the restrictions from the pre-motion broad-range core, the pixels on the other column (right) are used. This is because of the length of the tap filter to generate pixel values using fractional accuracy positioning. As a result, motion estimation is not processed by the core for pixel positioning under restricted violation. Including this motion vector restriction to all the cores attains a low delay with no discrepancies of reference images.

4 Results and Discussion

The encoder uses a multichip architecture that spatially divides a 4096-pixel 120-fps input image into two slices, and each slice is processed with a 4096-pixel 60-fps encoder LSI. Transfer of reference picture data near the slice boundary enables a multichip architecture maintaining high encoding efficiency. The encoder supports two modes essential for 120 fps, temporal-scalable coding mode and low-delay coding mode. The proposed system suppresses

coding efficiency degradation with a small amount of data transfer by exchanging only part of the reference picture data near the boundary. The proposed work claims that this technology has the potential to provide a more immersive viewing experience for users and improve the quality of high-motion video content.

The experiment was conducted to execute (implement) and determine the proposed encoder coding quality using SHM reference software [15]. The encoder is also executed and the performance of the encoder is observed based on the timing schedule. To assess the proposed encoder with multi-chip spatial division four 4096-pixel 120-fps test video sequences such as Duckstakeoff, Parkrun, Fourpeople, Kimono, Parkscene were considered as shown in Fig. 6. The encoder performance in terms of BD-BR (bit rate) [16] is compared with the standard SHM.

For encoder execution, the standard is considered in such a way that the motion vector is not allowed through the slice boundaries. Whereas, the proposed multi-chip configuration allows the motion vector to deviate vertically through slice boundaries by 128-pixels. The chosen test video sequences indicate slow and fast motion with small and large spatial details. The configurations include random-access scalable (RA-Scalable), random-access SNR (RA-SNR), low-delay P (LD-P), and low-delay B (LD-B) for the standard and proposed encoder. The obtained results are compared with the standard listed in Table 1.

Table 1. BD-BR representation for different sequences under various configurations

Video Sequence	Ducks takeoff	Park run	Four people	Kimono	Park scene	Average
RA-Scalable	-2.36	-2.07	-2.12	-2.05	-2.15	-2.15
RA-SNR	-2.19	-2.01	-2.03	-2.01	-2.25	-2.09
LD-P	-2.39	-2.16	-2.21	-2.13	-2.35	-2.24
LD-B	-2.42	-2.19	-2.25	-2.20	-2.41	-2.29

Fig. 7 depicts the device organization of the developed full 4096-pixel 120-fps encoder. FPGA #1 performs spatial division with two sets of slices (each 4096-pixel 60-fps video), and each is encoded by a single LSI. The PCIe interface connects the two LSIs. The multiplexer available in one LSI combines the encoded video information from two sets that result into one complete 4096-pixel 120-fps data. This output data will be fed as one channel to FPGA #2 as described in Fig. 5, which performs temporal scalability to produce single or double transmission channels. The proposed encoder is also capable to provide single channel output without using FPGA #2. The processing delay of the proposed encoder was calculated to assess its low delay performance as in Fig. 8. The input video sequence delay of 8 ms is incurred for transforming from a two channel of 4096-pixel 60-fps input. Simultaneous execution of two chip delay incurs an encoding delay of 7.9 ms, which is one half time of a 4096-pixel image using a single chip. Channel bytes appending delay is 4 ms. For smooth CBR, channel data corresponding to a peak of 384-pixels can alone be stored in a send buffer, and the buffering delay incurred is 1.2 ms. The complete delay of the proposed encoder is 21.1 ms. It is observed that the proposed encoder achieves better coding efficiency by simulations and a low delay of 21.1 ms by execution.

Fig. 6. Test Video Sequences (a) Duckstakeoff, (b) Parkrun, (c) Fourpeople, (d) Kimono, (e) Parkscene
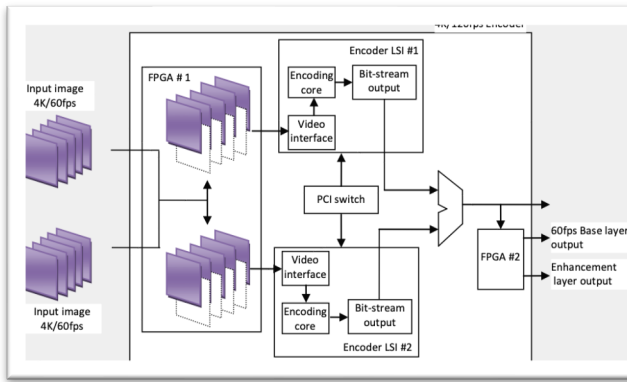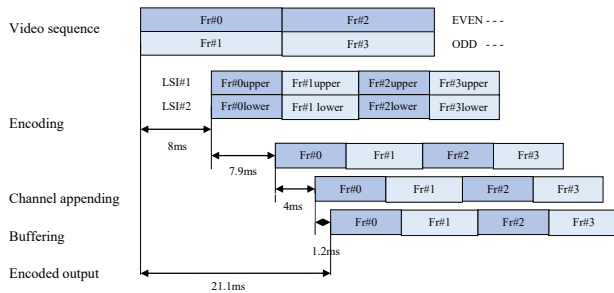
Fig. 7. System Framework

Fig. 8. Timing diagram

It does describe the experimental results of the proposed encoder in terms of coding efficiency and delay. The experiments are conducted to evaluate the coding efficiency of the proposed encoder using four 4096-pixel 120-fps video sequences. The BD-bitrate (Bjøntegaard delta-bitrate) was calculated to perform valid evaluation that did not depend on the objective image quality. The results showed that the proposed encoder achieved high coding efficiency with a small amount of data transfer by exchanging only part of the reference picture data near the boundary. The encoder also supports low-delay coding mode, which is essential for real-time video applications.

In the context of high frame rate video applications, the proposed encoder system has several implications. Firstly, it can improve the viewing experience for users by providing smoother and more realistic motion in videos. This is particularly important for fast-action content such as sports, where high frame rates can help capture the fast movements of athletes and provide a more engaging viewing experience. Secondly, the ability to achieve 4096-pixel 120-fps video encoding in real-time can enable more immersive and high-quality video experiences for users. This can be particularly useful in applications such as virtual reality, where high frame rates are essential for providing a seamless and realistic experience.

Moreover, the proposed encoder system's hardware and software architecture can be customized to meet specific needs, making it a versatile solution for various applications. For example, it can be used in surveillance systems to capture fast-moving objects with high accuracy, or in medical imaging systems to provide high-quality images for diagnosis and treatment.

Overall, the proposed encoder system has significant implications for the future of high frame rate video applications. It provides a scalable and customizable solution for achieving high frame rates, which can improve the viewing experience and enable more immersive and engaging video experiences for users.

The technology can impact real-world use cases in several ways. Firstly, it can improve the quality of video content in various applications such as sports, entertainment, and gaming. The high frame rate capabilities of the encoder can provide smoother and more realistic motion in videos, making the viewing experience more engaging and enjoyable for users.

Secondly, the technology can be used in surveillance systems to capture fast-moving objects with high accuracy. The high frame rate capabilities of the encoder can help detect and track objects more effectively, improving the overall performance of the surveillance system.

Thirdly, the technology can be used in medical imaging systems to provide high-quality images for diagnosis and treatment. The high frame rate capabilities of the encoder can help capture fast-moving objects such as blood flow, improving the accuracy of medical imaging and diagnosis.

Overall, the technology has the potential to impact real-world use cases in various applications, improving the quality and accuracy of video content and enhancing the performance of surveillance and medical imaging systems.

5 Conclusion

A low delay scalable HEVC encoder with complete 4096-pixel 120-fps encoding for HFR video is described. The input video of 4096-pixel 120-fps is split into two equal slices spatially, in which both are processed by an encoder LSI of 4096-pixel 60-fps. A scalable multi-chip configuration is provided by transferring the reference picture information over the slice boundaries that offered better coding efficiency. The vital commitments of this work is the improvement of a temporally-scalable encoder that accomplishes backward-compatibility and real-time encoding demonstration. The discoveries show that the coding bits needed for the enhancement layer is lower than the 60-fps base layer, and the delivery of 120-fps video can be accomplished more productively than that of 60-fps video as far as the coding effectiveness per frame. Predominantly, this work gives a promising answer for high frame rate video encoding and its expected applications in different fields. In addition, the proposed encoder further provides 120-fps temporal scalability with low delay of 21.1ms, which reminds to the future services of UHDTV.

Further to help much higher frame rate and resolution, it can be optimized with its performance and scalability. Another heading is to investigate the capability of this innovation in different fields, for example, sports broadcasting, augmented reality, and clinical imaging. For instance, high frame rate video can give more exact and itemized data for sports, while augmented reality and clinical imaging can gain from the improved visual quality and authenticity. Moreover, the proposed encoder can be incorporated with the rising advancements, for example, 5G communication and cloud computing to empower low-latency, real-time video. In general, this work gives a strong groundwork to future innovative work in the fields of high frame rate video encoding and expected applications.

_____

References

[1] 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, Jan. 2016, doi: 10.1109/TCSVT.2015.2461951.

[2] "Video Coding, Audio Coding, and Multiplexing Specifications for Digital Broadcasting." Association of Radio Industries and Businesses, Tokyo, Japan, Dec. 2016. [Online]. Available: https://www.arib.or.jp/english/html/overview/doc/6-STD-B32v3_9-E1.pdf

[3] "Transmission System for Advanced Wide Band Digital Satellite Broadcasting." Association of Radio Industries and Businesses (ARIB)., Tokyo, Japan, Jul. 2014. [Online]. Available: http://www.arib.or.jp/english/html/overview/doc/6-STD-B44v2_0-E1.pdf

[4] T. Anttalainen and V. Jääskeläinen, *Introduction to communication networks*. Boston: Artech House, 2015.

[5] T.-M. Liu *et al.*, "A 0.76 mm^2 0.22 nJ/Pixel DL-Assisted 4K Video Encoder LSI for Quality-of-Experience Over Smartphones," *IEEE Solid-State Circuits Lett.*, vol. 1, no. 12, pp. 221–224, Dec. 2018, doi: 10.1109/LSSC.2019.2905958.

[6] C. -C. Ju et al., "A 0.5 nJ/Pixel 4 K H.265/HEVC Codec LSI for Multi-Format Smartphone Applications," in IEEE Journal of Solid-State Circuits, vol. 51, no. 1, pp. 56-67, Jan. 2016, doi: 10.1109/JSSC.2015.2465857.

[7] M. Emoto, Y. Kusakabe, and M. Sugawara, "High-Frame-Rate Motion Picture Quality and its Independence of Viewing Distance," *J. Display Technol.*, vol. 10, no. 8, pp. 635–641, Aug. 2014, doi: 10.1109/JDT.2014.2312233.

[8] "Parameter values for ultra-high definition television systems for production and international programme exchange." International Telecommunication Union, Geneva, Switzerland, Oct. 2015. [Online]. Available: http://www.itu.int/publ/R-REC/en

[9] Y. Kuroki, T. Nishi, S. Kobayashi, H. Oyaizu, and S. Yoshimura, "3.4: Improvement of Motion Image Quality by High Frame Rate," *SID Symposium Digest*, vol. 37, no. 1, p. 14, 2006, doi: 10.1889/1.2433276.

[10] T. Onishi *et al.*, "Single-chip 4K 60fps 4:2:2 HEVC video encoder LSI with 8K scalability," in *2015 Symp. VLSI Circ. (VLSI Circuits)*, Kyoto, Japan: IEEE, Jun. 2015, pp. C54–C55. doi: 10.1109/VLSIC.2015.7231325.

[11] S. Lee, H. Kim, and N. Eum, "Reduced complexity single core based HEVC video codec processor for mobile 4K-UHD applications," in *2016 IEEE 6th Int. Conf. Cons. Electr. - Berlin (ICCE-Berlin)*, Berlin, Germany: IEEE, Sep. 2016, pp. 94–95. doi: 10.1109/ICCE-Berlin.2016.7684727.

[12] H. Shimamoto *et al.*, "A Compact 120 Frames/sec UHDTV2 Camera with 35mm PL Mount Lens," *SMPTE Mot. Imag. J*, vol. 123, no. 4, pp. 21–28, May 2014, doi: 10.5594/j18413.

[13] "Video Coding, Audio Coding, and Multiplexing Specifications for Digital Broadcasting." Association of Radio Industries and Businesses, Tokyo, Japan, Sep. 2018. [Online]. Available: https://www.arib.or.jp/english/html/overview/doc/6-STD-B32v3_9-E1.pdf

[14] K. Nakamura *et al.*, "Low Delay 4K 120fps HEVC Decoder with Parallel Processing Architecture," in *2019 IEEE Symp. Low-Pow. High-Speed Chips (COOL CHIPS)*, Yokohama, Japan: IEEE, Apr. 2019, pp. 1–3. doi: 10.1109/CoolChips.2019.8721335.

[15] V. Seregin and H. Yong, "HEVC Scalability Extension (SHVC)." Fraunhofer HH, Berlin, Germany, 2017. [Online]. Available: https://hevc.hhi.fraunhofer.de/shvc

[16] G. Bjøntegaard, "Calculation of Average PSNR Differences between RD curves", *ITU-T SG 16/Q.6 13th VCEG Meeting*, Austin, Texas, USA, document VCEG-M33 (available at http://ftp3.itu.int/av-arch/video-site), Apr. 2001.