MIGRATING DIGITAL AD-INSERTION APPLICATIONS FROM MPEG-2 TO AVC (H.264)

Mukta Kar, Ph.D., Cable Television Laboratories, Inc.
Sam Narasimhan, Ph.D., Motorola, Inc.

Abstract

Splicing is the fundamental technique used to insert commercials or short programs in channels, for editing audio/video content in post-production houses, and for channel switching in headends and other broadcast stations. Splicing is currently used in US Cable networks for digital ad-insertion based on MPEG-2 video [3], SCTE and ITU-T standards [4] and there are plans to migrate these applications and develop associated standards based on the emerging H.264/AVC video [1] in the near future.

In these new applications, the splicing equipment (or function) combines two independently encoded AVC streams and is expected to produce a stream for receiving equipment that conforms to both AVC Video [1] and MPEG-2 Systems [2]. To achieve significantly higher compression efficiency than that of MPEG-2 video while providing same or better quality video, AVC compression standard has introduced several new tools, reference picture structures and enhanced MPEG-2 tools all of which can be used adaptively based on the nature of the content. All these make AVC more complex compared to prior compression schemes in addition to being not backward compatible with MPEG-2 video.

Many of the Standards Development Organizations (SDOs) such as DVB/ETSI, SCTE, DVD and ARIB have completed the specifications related to the adoption of AVC in broadcast, VOD and PVR applications. This paper outlines issues related to splicing between two independently coded AVC streams for local Ad-insertion applications and proposes schemes for generating an AVC Video [1] and MPEG-2 Systems [2] conformant output by such splicing equipment so that a seamless or near-seamless splicing can be achieved.

INTRODUCTION

The opportunity for Local commercial insertion has been created to benefit the communities and its businesses on a local, zonal or regional basis since the days of Analog Television. Local commercial / advertisement (Ad) opportunity provides the US broadcast television industry over 35 billion dollars in revenue. The revenue from this opportunity for the cable industry has grown from a few million to a few billion dollars with 5-6 billion in revenue expected in 2008. As it provides a significant cash flow for our MSOs, increasing this revenue further has a prime importance to the industry. In the days of analog television, most local Ad-insertion equipment was proprietary in nature and hence non-interoperable. In moving from the era of analog television to a digital one, the cable industry understood the problems in using proprietary equipment and the advantages in using inter-operable equipment from a multi-vendor market place. To create such a competitive multi-vendor market place, the cable industry took initiative in standardization efforts in both the international (ITU/ISO) and national levels (SCTE) that covered not only Audio-Video but other areas such as cable modem, VOIP, etc. One such application area is the local program/commercial insertion.

Figure 1 displays a block schematic of a typical local ad-insertion system. A timing signal known as Cue-tone (in analog) or Cue-message [4] (in digital) is embedded with a program and then distributed to the headends or broadcast affiliates via satellite. MSOs receive such a program using an IRD (Integrated Receiver Decoder) at their headend. Then the headend equipment separates Cue-tone/Cue-message from the program. Based on the timing signal in the cue-tone/cue-message, a splicer and Ad-server replaces the national ad with a local ad. This process is known as local ad insertion.

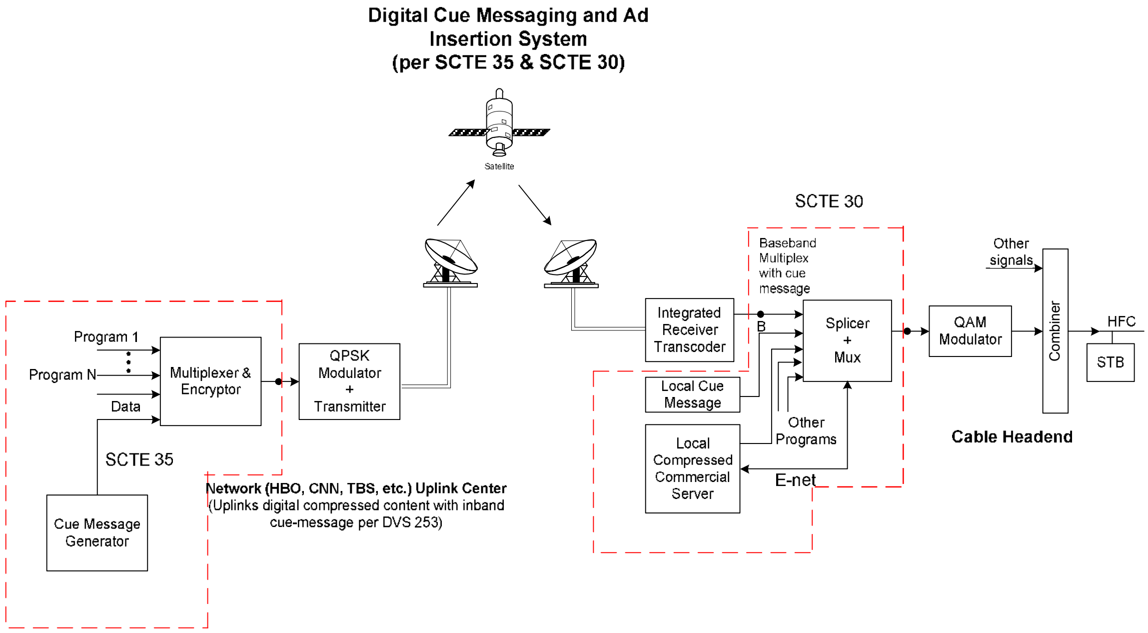


Figure 1. Schematic Block Diagram of Digital Program/Ad Insertion System

As shown in Figure 2, Local Ad Insertion technology in the analog video domain was simple in nature as the transmit order of frames is same as the display order. Splicing the digital uncompressed video is also simple for the same reasons. The process involves frame accurate timing signals indicating the beginning and end of a national advertisement in a program that need to be replaced with a local ad or perhaps with an updated ad. Delivery of analog video or uncompressed video to consumer homes is very inefficient in the usage of bandwidth in addition to many other limitations. Digital video compression technology coupled with digital modulation provides significant efficiency, flexibility and other benefits in delivering digital video and audio to consumer homes. Also digital technology provides superior video quality compared to analog as the former is less prone to noise.

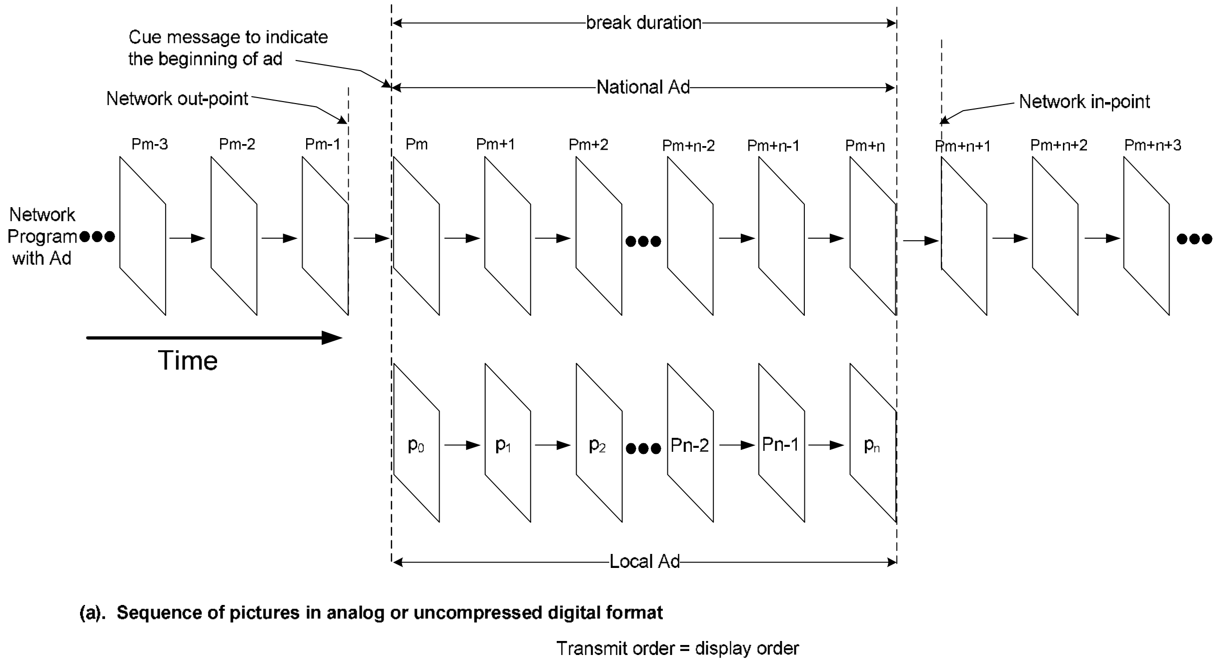


Figure 2. Simplified Diagram of Local Ad Insertion

However, digital video compression technology introduces problems for some of these broadcast applications (e.g. splicing between two MPEG-2 streams or between two AVC streams) primarily due to two reasons – (a) Order of the video frames gets modified in compressed domain (transmission order does not maintain display order) and (b) compressed frames depend on other frame or frames (called as reference frames) for decoding / decompressing them in the decoder.

SPLICING BETWEEN STREAMS WITH MPEG-2 VIDEO

Figure 3(a) illustrates a segment of video where frames are in display order. Figure 3(b) depicts the same segment when compressed in compliance with the MPEG-2 video standard and sent over a transmission channel. One may notice that the transmit/decode order is not same as the display order and hence a splice cannot be done at all picture boundaries. For example if one splices out at any of the B pictures, then this will introduce gaps in the display. For seamless and near seamless splicing between two compressed video streams, MPEG-2 TSTD buffer conformance must also be maintained where decoder buffer (buffer size of 1.8Mbits for MP@ML) must not overflow or underflow as this may result in artifacts. Typical buffer behavior in a MPEG-2 decoder is shown in Figure 4 assuming the decoder receives a constant bitrate channel. MPEG-2 provided tools to achieve seamless or near seamless splicing but it does not tell on how to achieve it. MPEG left it to MPEG-2 product designers for innovation and product differentiation. To achieve seamless or near seamless splicing, some constraints may have to be maintained while creating the streams to be spliced. Such constraints may include GOP structure, an anchor frame at the out-point of the first stream and an I frame with a sequence header and closed GOP at in-point of the second stream.

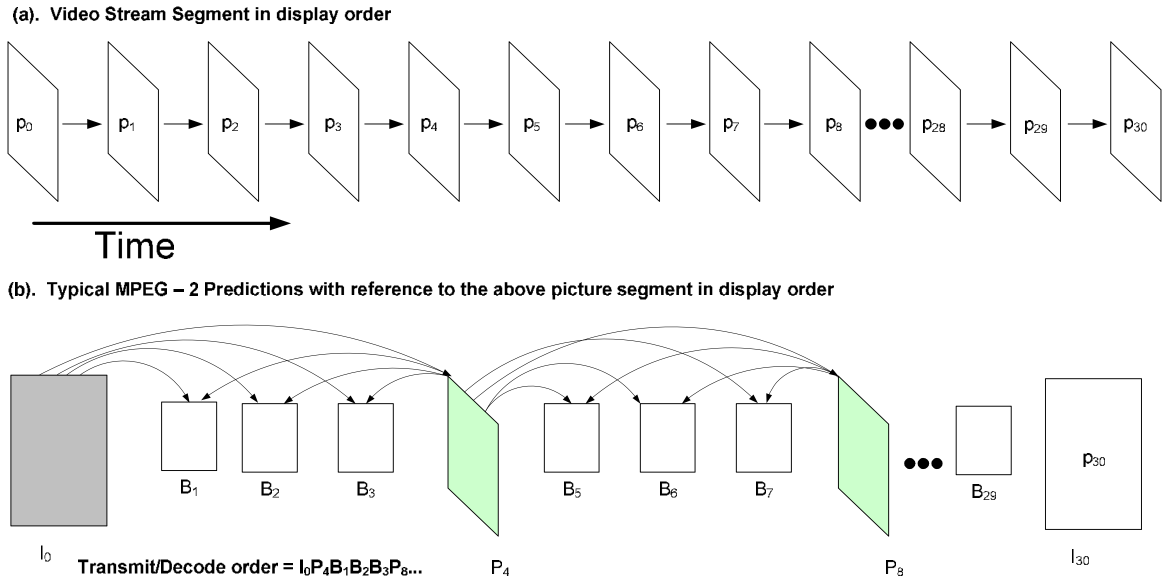
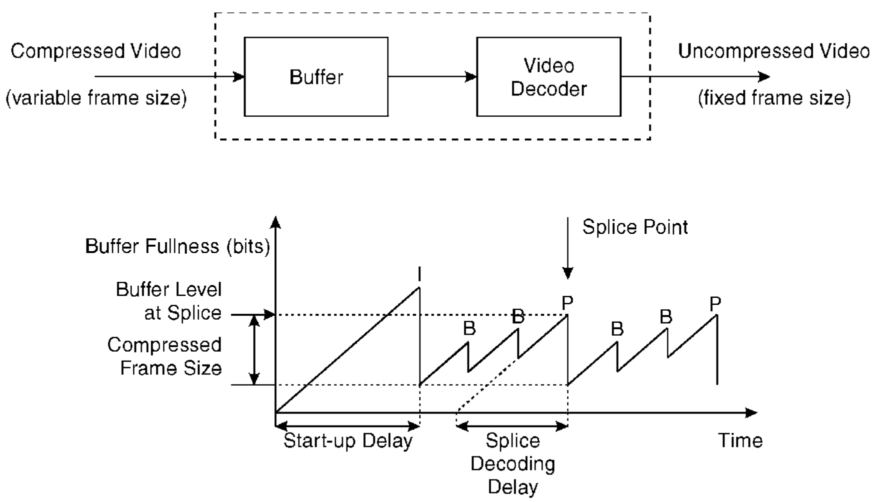


Figure 3. Typical Prediction Methodology used in MPEG-2 Video Compression.



A Typical Buffer Trajectory

Figure 4. Simplified diagram of an MPEG-2 decoder and level of bits in the decoder buffer.

An additional function that needs to be met by digital ad-insertion systems is the matching of decode delays between the network and splice stream as shown in Figure 4.

Even though it may be easy to splice between two ‘well conditioned’ MPEG-2 transport streams (standards such as SMPTE 312M specify this stream conditioning), the stream conditioning and matching impose too much of a burden and constraint on the uplink encoders and ad-servers. Hence SCTE developed specifications such as SCTE 35 [4] (digital cue-message standard) to signal the splice opportunities in the compressed video stream and splicers were developed to perform the tasks outlined above so that splicing can occur without imposing too many constraints on the uplink or ad-servers. Majority of the splicers deployed currently perform the following functions:

- Continuous bitrate transcoding to maintain an average compressed bitrate between the two video streams.
- Time base adjustment to maintain a common PCR, PTS and DTS between the streams without introduction of any discontinuities.
- Matching the vbv-delay between the two streams so that decode and presentation are continuous when the output reaches the settop units at consumer premises.
- Splice between content in film-mode and non-film-mode by maintaining field parity.

In addition, the splicers also maintain conformance to MPEG-2 video and systems standards in their output and make sure that PSI information does not change across the splice so that settop units can present a seamless transition between network program and advertisements.

AVC VIDEO CODING AND HIERARCHICAL GOP STRUCTURES

It has been mentioned earlier that to achieve better compression efficiency than that of MPEG-2 video standard [3], AVC [1] introduced many new tools and enhanced some MPEG-2 tools. One of these tools is in the use of B-pictures as reference and hierarchical use of such B pictures.(MPEG-2 video does not allow B frames to be used as reference). This particular tool introduces an additional complexity for splicing and this will be discussed later. The MPEG/JVT committee also structured AVC in a very flexible way so that it can be implemented in a wide range of applications that includes broadcast, video telephony, video conferencing, and video streaming.

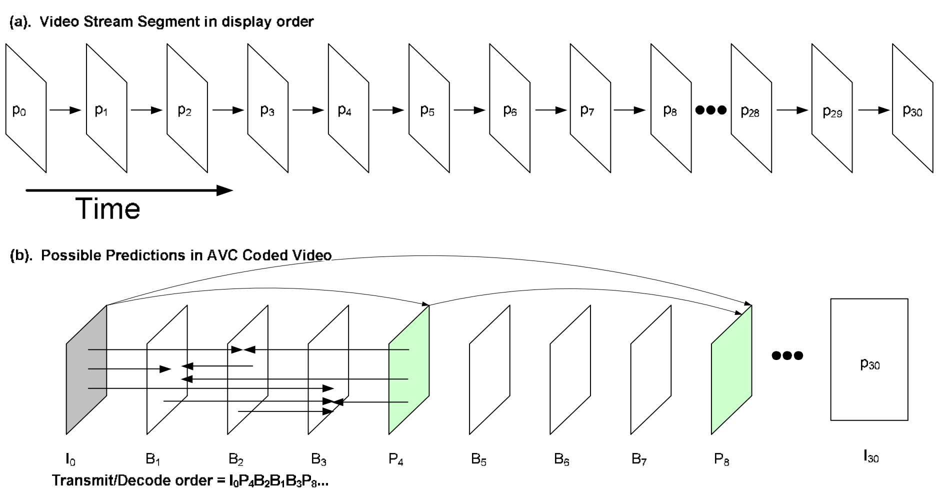


Figure 5. Typical Predication Methodology used in AVC Video Compression.

Use of B frames for reference in AVC and the associated GOP structures that use this are sometimes called 'Hierarchical GOP structures'. Figure 5(b) shows a typical GOP structure in AVC video where I, P and B frames are used. Frames are in display order in Figure 5(a), where as Figure 5(b) shows a typical compression structure of AVC where hierarchical GOP structures are used with B pictures as reference. One may notice that the difference in transmit/decode order of pictures in Figure 3(b) and Figure 5(b). The use of B pictures as reference and GOP hierarchy makes decoding AVC coded stream more complex compared to MPEG-2 coded video with respect to the management of reference pictures in the decoder memory. AVC also introduced IDR picture and changed traditional definition of I picture which is used in MPEG-2 video. These advanced tools and flexibility of GOP structures make seamless or near-seamless splicing using AVC video [1] for local ad-insertion more challenging compared to MPEG-2 video [3].

ADDITIONAL SPLICING ISSUES WITH AVC VIDEO COMPARED TO MPEG-2

AVC splicers have to implement all the functions that are implemented by MPEG-2 video splicers. These include bitrate transcoding, time base adjustment, CPB delay matching and maintaining field parity between film and non-film modes. In addition, AVC splicers need to manage another function to accommodate the 'Hierarchical GOP' structure variations between the streams being spliced. The following illustrates the issue and proposed solutions.

MPEG-2 provides relatively simple GOP structure involving I, P, and B pictures. Depending on a scene the parameter m (number of B picture between two anchor frames) varies. MPEG-2 allows only two reference pictures at any time which are managed in the decoder memory using FIFO method (also known as bumping process) where arrival of a new reference picture pushes out the older reference picture out of the decoder memory. In MPEG-2 (for non-low-delay mode) the delay between the decode time of first access unit in the sequence and the display time of first access unit in the sequence is always 'one frame period' for both closed GOP (I,P,B,B,P,B..) and open GOP (I,B,B,P,B..) structures. Let us call this as the 'display latency'. This display latency is always 'one frame period' for MPEG-2 sequences that use different values of m and hence MPEG-2 splicer's were able to concatenate any two sequences and still maintain conformance to T-STD, VBV and constant display rate at their output without any difficulty.

In AVC this is not true as the 'display latency' for different video sequences vary based on the hierarchical GOP structures which use B pictures as reference pictures. The display latency can range from 'one picture period' (for non-stored-B structures or MPEG-2 like structures) to several picture periods based on the levels of hierarchy. In addition, unlike MPEG-2 video AVC mandates the maintenance of constant display rate (constant DPB output) when coded video sequences are concatenated. This makes concatenation or splicing two such video sequences with different display latencies difficult as the output cannot conform to AVC specifications (I.E; maintain constant delta in the CPB removal time and DPB output time). If the display latencies do not match, then one will see a missing picture at DPB output time or see 2 access units with the same DPB output time. One solution to this is to only combine sequences that have the same GOP structure and this mandate is not attractive in Cable networks for applications that use splicing. Based on inputs from the US Cable community, AVC has agreed to modify the standard allowing the use of a marker called end_of_stream NAL unit to splice between two AVC coded video sequences or streams with different display latency where the requirement to maintain constant DPB output does not apply. The next section covers proposed solutions for splicing between AVC streams based on this action by AVC.

PROPOSED SOLUTIONS FOR BOTH STREAM CONDITIONING AND SPLICER FUNCTIONS

1. In order to enable seamless splicing between two video sequences at different horizontal

resolutions (at same frame rate and vertical resolution), the application standards such as SCTE 128 [5] are mandating the same number of pictures in the DPB for all the horizontal resolutions (determined by the highest horizontal resolution for the level such as 720 for SD and 1920 for HD). This allows the decoders to keep the same DPB memory management across the resolution change and hence produce seamless output.

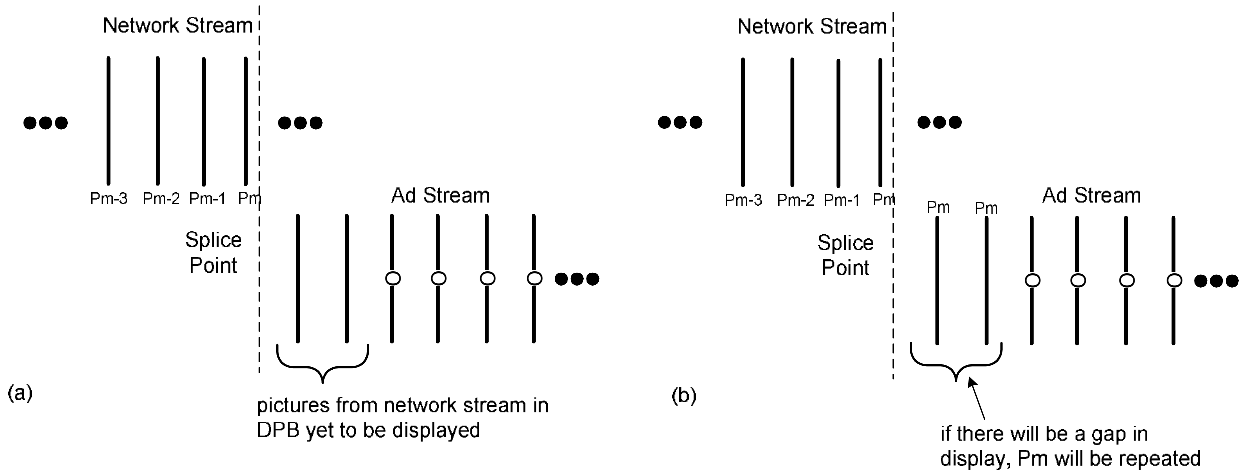


Figure 6. Seamless / near-seamless splicing without the use of end_of_stream NAL unit at the splice point (a) a case where no_output_of_prior_pics_flag=0, (b) a near-seamless case where no_output_of_prior_pics_flag=1.

2. AVC also agreed to loosen the requirement for decoders to infer the DPB management using the no_output_of_prior_pics_flag (I.E., infer this to be '1' and clear the DPB when there is a resolution change). The change allows the application standards to mandate that receivers process this flag correctly so that DPB is managed per the transmission systems intent. Splicers can set this flag correctly at the transition points to achieve 'seamless' splicing as shown in Figure 6(a).

3. The third proposal is the appropriate use of end_of_stream NAL unit at the splice transition points (called Out or In-Point) so that seamless or near-seamless splicing is possible. This is shown in Figure 7. In some combinations, seamless splicing can be achieved without the use of end_of_stream NAL unit. The first example is where the display latencies match between the streams. The second example is where the display latency difference can be adjusted by the use of Picture timing SEI message with an appropriate value for pic_struct. This SEI with the pic_struct value allows repetition of last displayed picture and this can be used to splice a stream with a higher display latency into a stream with a lower display latency. This also requires the first stream to be coded using frame pictures. For all other combinations of streams, the end_of_stream NAL unit should be used with the correct setting of the no_output_of_prior_pics_flag at the transition point to manage the DPB buffer and make sure that two pictures with the same display time are precluded. Seamless splicing with end_of_stream NAL unit can also be achieved by offsetting the decode time of the pictures in Ad Stream appropriately. This mode is not recommended as most receivers expect the decode time to be contiguous between network and ad-streams.

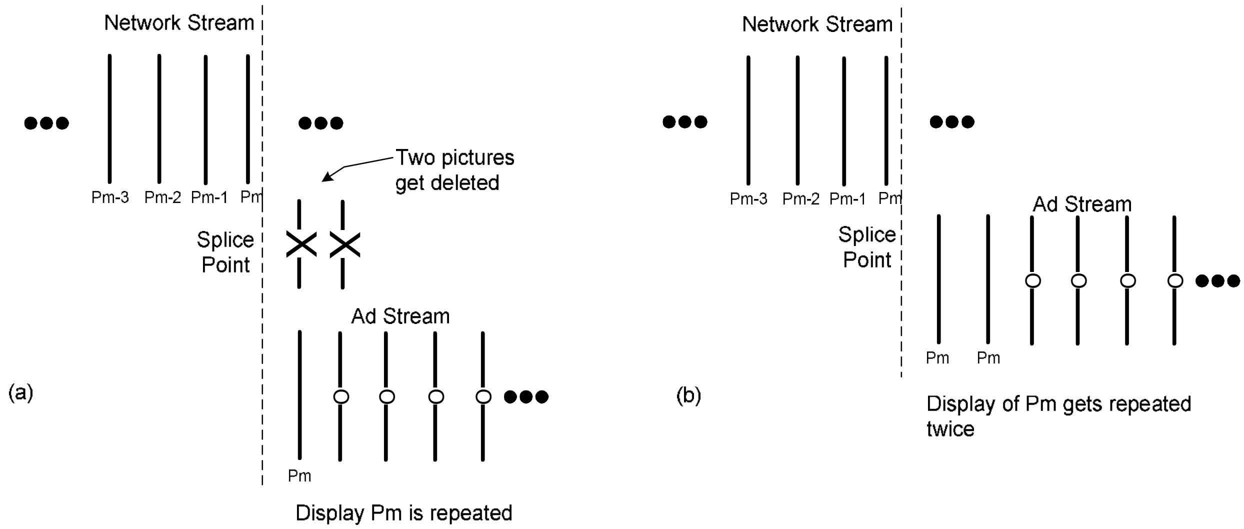


Figure 7. Near-seamless splicing using end_of_stream NAL marker at the splice point (a) a case where no_output_of_prior_pics_flag=1 (b) a case where no_output_of_prior_pics_flag=0.

SUMMARY

In this paper the technology of splicing and its importance to achieve local ad-insertion for the broadcast industry have been discussed. The local ad-insertion application provides significant amount of revenue to the industry, in particular, the cable MSOs. The challenges of splicing two video streams in the compressed domain (e.g. MPEG-2 video) for seamless and near-seamless viewing experience have been discussed. It has been also shown that splicing of two AVC video streams is much more difficult than MPEG-2 video as AVC video coding uses more advanced video tools and complex coding structure to achieve higher compression efficiency. This paper proposes a few methods/solutions to splice AVC streams to achieve seamless or near-seamless ad-insertion as needed in the broadcast industry.

ACKNOWLEDGEMENTS

The authors wish to thank the management of CableLabs and Motorola for their support and encouragement in performing this work.

BIBLIOGRAPHY

1. ITU-T Rec. H.264 | ISO/IEC 14496-10, (2005), “Information Technology – Coding of audio visual objects – Part 10: Advanced Video Coding.”

2. ISO/IEC 13818-1, (2007), “Information Technology – Generic coding of moving pictures and associated audio – Part 1: systems.”

3. ISO/IEC 13818-1, (2000), “Information Technology – Generic coding of moving pictures and associated audio – Part 2: video.”

4. ANSI/SCTE 35-2004 | ITU-T J.181, Digital Program Insertion Cueing Message for Cable.

SCTE 128 (2007), AVC Video Systems and Transport Constraints for Cable Television