ADAPTING ADVANCED VIDEO CODING TO CABLE NETWORKS

Mukta Kar, Ph.D. and Yasser Syed, Ph.D.
Cable Television Laboratories, Inc.

Abstract

This paper describes the current cable environment and what demands it has on development of a cable-friendly AVC video coding stream. It considers what the further restrictions on AVC are for recreating the same services already provided by MPEG-2 plus coexistence in the same network with MPEG-2 video streams and other digital services. The paper also looks at what are the requirements for multiplexing, random access to support existing broadcast services, and how this affects the coding design. Furthermore it looks at additional emerging needs to address services like Personal Video Recorders (PVRs), Video-on-Demand (VOD), and Digital Program Insertion (DPI) and how this can constrain the design on the video stream and the video coding layer. This paper concludes by saying it is not enough to just have a highly efficient video coding standard developed, the cable environment (or any application environment) imposes its own set of demands that trades off the advanced video coding design and some compression efficiency for expected viewer experience.

INTRODUCTION

In the last century television contents (video and audio) have been delivered to consumers mostly in analog format. It was realized then that the fidelity of analog reception is good in noiseless or less noisy environments, and also the analog system is either inefficient or inflexible in the important areas such as service security, bandwidth usage, interactivity, etc. Around the 80's and 90's of the last century, researchers understood the power and flexibility of migrating to digital technology, and it started being practical to do so.

However, the delivery of uncompressed digital audio/video to consumer homes is not an acceptable solution as it takes a huge amount of bandwidth to deliver. To tackle this problem, researchers had been working on how to compress digital audio/video effectively for consumer-oriented applications. Towards the end of the last century, audio-video compression technology matured to the level that it can be practically used for consumer applications, and MPEG-2 standards were created in 1994 to broadcast digital audio/video to consumer homes and offices. MPEG-2 technology takes much less bandwidth than analog delivery (today approximately 10:1 bandwidth efficiency) while providing same or better fidelity (quality of audio/video) even in a noisy environment.

AVC: EMERGING DEMANDS FOR VIDEO CODING

Over the last few years MPEG-2 technology has become very popular and received global acceptance which has led to many new innovations such as HDTV, PVR (DVR), VOD, streaming media, interactive TV, etc. These applications were unthinkable in the analog format and did not exist as MPEG-2 standards were created. While these new applications can help generate more revenues for the broadcast industries (cable, satellite and terrestrial), they also require more digital bandwidth to deliver. For example, an HDTV channel requires a bandwidth 15-20 Mbits/sec whereas a SDTV channel takes only 2-3 Mbits/sec using the MPEG-2 video standard [1,2].

Besides broadcast industries, other video industries such as DVD, streaming media, etc. also needed an advanced video codec that has significantly better compression than MPEG-2. For example, an HD movie cannot be stored on a current DVD disc. As MPEG-2 video needs more bandwidth, today's streaming media technologies use proprietary codecs to deliver their content but encounter storage and software challenges due to a lack of convergence on a compression standard. In another industry, conversational video applications (video conferencing and video telephony) use another video codec standard, but are also facing similar challenges to streaming media. This shows that each of the major video industries today use diverse video standards and different hardware platforms for delivery and storage which impede consumer acceptance of multiple video applications.

One of the answers to the problems above is to compress digital audio/video even more efficiently than that in MPEG-2 in order to be useable across different industries. A better answer will be to create a new video standard that will not only provide significantly better compression than MPEG-2 but also a convergence in hardware platforms across major video applications.

To have better compression efficiency as well as to eliminate divergent requirements in hardware, the video experts in ITU and ISO/MPEG formed a joint video team (JVT) to investigate a solution. Their collaborative work over two years resulted in the creation of an advanced video coding (AVC) in ISO/MPEG (MPEG-4 part 10) and H.264 in ITU in May, 2003 [4]. AVC may be considered as an aggregation of tools, some of which are enhanced MPEG-2 tools and a few additional new tools that have been added to increase compression efficiency while providing the same or better picture quality. In brief, AVC achieved two important objectives - better compression efficiency than MPEG-2 and convergence in hardware across major consumer video applications - but sacrificed backward compatibility with the MPEG-2 video compression standard. However, backward compatibility has been addressed by the silicon designers at the AVC codec chip level such that it becomes transparent to users.

AVC tools consist of enhanced MPEG-2 tools, some new tools, and error resiliency tools. MPEG-AVC or AVC is a large tool box which can meet the needs of most major video applications such as broadcast video, conversational applications, and streaming video. The AVC system has created an opportunity to change the paradigm of coding in AVC. The new tool box allows multiple reference frames to be used for predictions as opposed to the two reference frames allowed in MPEG-2. It also introduces spatial prediction in intra-coded frame and adaptive motion compensation. Motion compensation resolutions' block size may be 16x16, 8x16, 16x8, 8x8, 8x4, 4x8 or 4x4 pixels. It supports block transforms of 8x8 and 4x4 pixels for luma which may be used adaptively. It also introduces in-loop filtering to smooth the sharp edges which are an artifact of using block transforms. AVC also added a few error-resilient tools (not present in MPEG-2) to support video delivery over non-QOS networks such as public IP networks and wireless networks.

CURRENT STATE OF CABLE: THE NEED FOR AVC IN CABLE

About ten years ago, the cable industry was delivering premium TV programs primarily in analog only format. Over the last ten years evolutionary changes have been taking place in cable. Today cable delivers not only analog video but also a few digital enabled services such as digital TV, high speed data, and telephony. Cable is also in the process of using non-linear video services in addition to scheduled types of services. At present cable delivers linear applications like broadcast television but also is adding non-linear services such as VOD, PVR, VOIP. Also we are shifting from the broadcast television model to a personal television one. As a result cable faces a real bandwidth crunch as it has to support the legacy analog and digital broadcast television services due to regulatory and logistical reasons, while adding an array of the new digital services to actively engage with new competition. VOD, MOD, and PVR services require a significant amount of storage at the headend and in client devices. So we need better compression of the digital content than MPEG-2 can provide. AVC will certainly help in delivering more channels over the same digital bandwidth as well as in storing the digital content at the headend or in PVRs.

It was mentioned earlier that MPEG-2 technology has been accepted widely across the global broadcast industry. It provides tremendous business continuity in terms of content, equipment (transmission, storage, and test & measurement), etc. It is expected that the general video industry will migrate to AVC in a next few years, especially in the area of HDTV delivery and storage. AVC may also provide opportunity for adding new services, such as video conferencing and video telephony, using the same AVC hardware platform. In a way, to future-proof our industry with respect to delivery of multiple applications, the cable industry needs to consider all these factors mentioned above.

ADAPTING THE AVC STANDARD TO CABLE

Although AVC created profiles and levels geared towards major video applications, still it is not well tailored towards the cable environment. The video stream may have a much better compression efficiency using all tools in the standard, but the expected viewer experience for cable TV has not been designed in the AVC standard. The adaptation of MPEG-2 to cable was an easier process than what is now being considered because the standard was optimized more towards a broadcast application, while the AVC standard tries to encompass more than that. Additionally cable has also been changing at the same time to address other applications than just broadcast.

Constraints need to be written on the AVC standard (or any future advanced coding standard) to maintain current expected services and viewer experience while improving the plant bandwidth usage efficiencies and expanding on these services. This constraint document will be similar to the standard document SCTE 43 2005 [6] that is written for MPEG-2 video. The network infrastructure and legacy equipment need to continue operating without disruption. Existing services such as the Electronic Program Guide, Emergency Alert Messages, and audio-video synchronization need to be maintained and be agnostic to any new video coding standard. Viewer experiences like channel switching/surfing, PVR recording and interactive video services still need to be maintained transparent to different video coding technologies. At the same time, the splicing of video streams to perform local commercial insertion should be visibly seamless as done in MPEG-2. The cable viewer should not be aware of any changes in video coding technologies, but should benefit from the increased services. Designing constraints may have a minor negative effect on compression efficiency, but will allow the viewer experience to remain intact.

DELIVERY OF AVC CODED VIDEO

It is important to note that currently all digital services are delivered over MPEG-2 transport over the HFC cable network. To avoid expensive changes to the existing MPEG-2 transport-based infrastructure, any new digital services that will be added to cable should also be delivered using existing MPEG-2 multiplexing capability with other services without causing any side effects. This means that the existing services based on analog, MPEG-2 video, and high speed data services will remain unaffected. So it is probably the most important requirement that any new service based on AVC coded video (or any advanced coded video that may be added to cable) should be deliverable over the MPEG-2 transport infrastructure that exists in the cable network today.

It may be noted that AVC is just an advanced video coding standard. Unlike MPEG-2, no specific transport standard has been created for delivery of AVC coded video. The MPEG committee realized that a matured MPEG-2 infrastructure should not be reinvented. Keeping that in mind, MPEG created an amendment to MPEG-2 Systems [3] to deliver AVC-coded video over existing MPEG-2 transport. Figure 1 shows how AVC coded video can be packetized in MPEG-2 transport packets which will enable multiplexing with other elementary streams and other cable digital services. Transport over MPEG-2 will ensure audio-video synchronization (no lip-synch error). Additionally, it will support existing cable applications such as channel change, trick-mode, and DPI. Some constraints will be necessary at the video level (SCTE 43 [6] for MPEG-2 video). The digital transport / multiplex standard SCTE 54 2004 [7] has to be amended. These are discussed below.

It is mentioned earlier that the AVC standard will not only benefit broadcast applications, it will also be used in other major video services such as conversational and streaming media applications. ITU H.320 and Internet Protocol (IP) are used today to deliver conversational and streaming media applications, and these transports can be utilized for AVC as well. Referring back to Figure 1, AVC coded video allows for supporting these 3 major transport protocols (MPEG-2, H.320 & IP), using a network abstraction layer (NAL). First, the coded video and related information (PS and SEI) are packetized in NAL packets which can be variable in size. Each NAL packet will have a type value that tells the type of payload it is carrying (VCL, PS or SEI). Those NAL packets are then re-packetized in MPEG-2, IP and H.320.

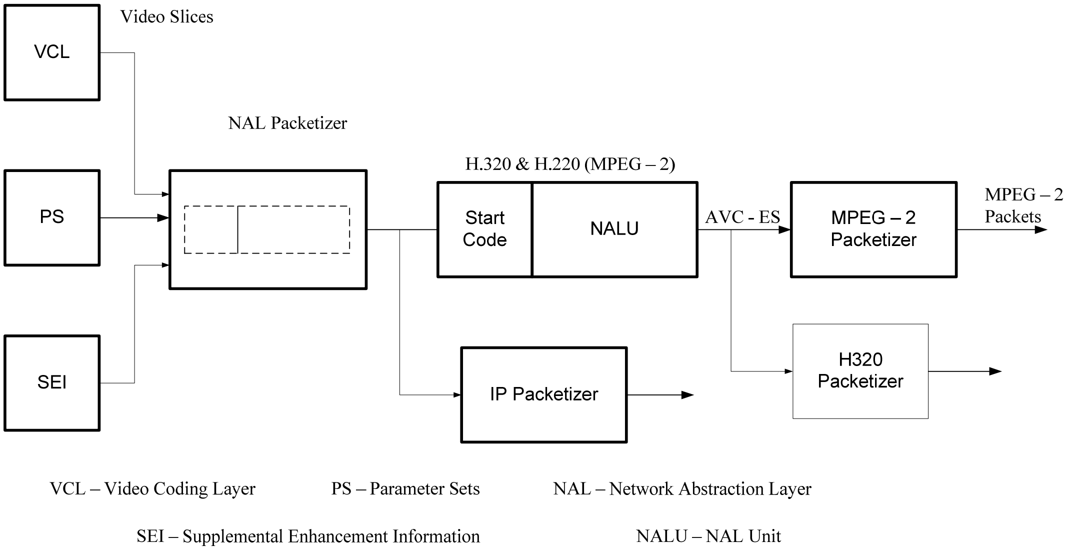


Figure 1. Carriage of AVC Compressed Video

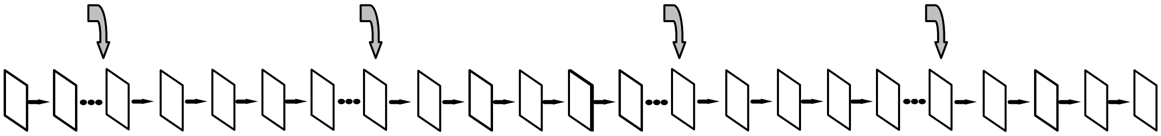


Figure 2. Random Access Points into Video Stream

RANDOM ACCESS REQUIREMENTS FOR CABLE

A critical feature needed for cable is the ability to randomly access the video stream at regular intervals [see Figure 2]. This is an essential need for fast consistent channel changing/ channel acquisition in the broadcast environment, but it also can be quite useful for DVR and commercial (ad) insertion applications. As will be pointed out in the following sections, constraints to design the video stream and transport headers will affect how a video compression standard is used and will tradeoff some potentially very high coding efficiency gains in exchange for supporting a viewer expected experience.

Channel Change requires random access points to allow switching into an existing broadcast channel without an undue amount of delay in receiving the video picture. The viewer after he switches the channel will normally tolerate 1-2 seconds at most of not seeing video (because that is what a viewer already expects from existing analog change times) before assuming something is wrong with the channel and switch away. The channel change time is actually due to a combination of factors such as physical RF tuning, conditional access, parsing streams, and decoding the video stream. Decoding of the video is dependent upon the design of the video stream to allow for regular periods of random access while the rest of the channel change contributing factors are outside factors. A good strategy will allow channel changing to be fast, frequent, and output undisrupted clean video once the decoding process starts or in other words a graceful transition from one channel to the other. Additionally to work in an existing cable environment, a viewer should be able to switch to channels that have alternate video coding technologies (MPEG-2, analog) for the channel with again a graceful transition.

Channel Change requires a combination of signaling at the transport layer to identify the pertinent video transport packets and designing of the video coding pattern in the video layer to create a clean continuously displayable output video only using data after the access point. Some compression efficiency in the video stream will be sacrificed in order to allow for a consistent and fast channel change between video streams.

HOW MPEG-2 DOES RANDOM ACCESS

An MPEG-2 stream uses the Random Access Indicator (RAI) bit in the MPEG-2 transport (4byte) header adaptation field shown in Figure 3 to identify that this is the transport packet (188 bytes) that identifies an access point in the video stream. Once that transport packet is decoded, sequence header and picture header information is given to identify the dimensions and frame rate of the video. This also starts the process of identifying the next transport packets in the sequence.

At the video coding layer, the random access point identifies the start of a GOP (group of pictures) that consist of a pattern of I, P, and B coded pictures [see Figure 4]. The next random access point will be at the beginning of the next GOP which is normally encoded at intervals of 15 to 30 frames (1/2 second or 1 second) with some leeway for making some occurrences of the interval smaller to adapt to stream changes or stream conditioning situations. The types of pictures indicate the different type of prediction that is used in MPEG-2 video:

I-Pictures - These pictures use intra-frame compression and are not predicted from other pictures (no temporal compression). This is also a

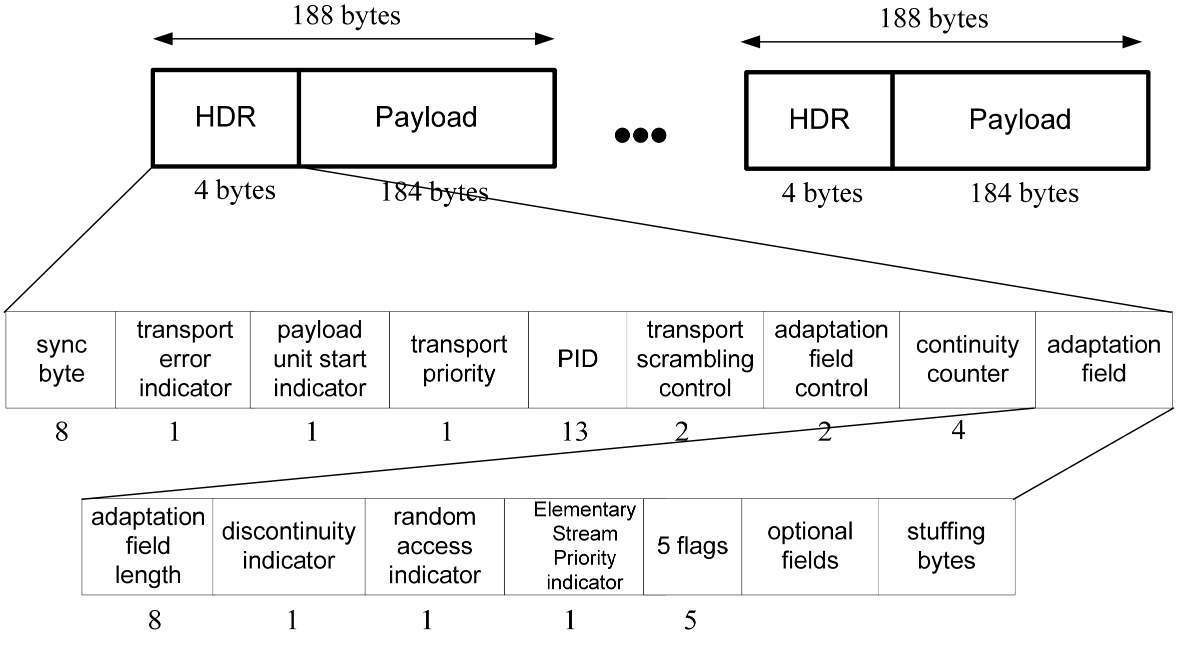


Figure 3. MPEG-2 Transport Stream Packet with Header Information

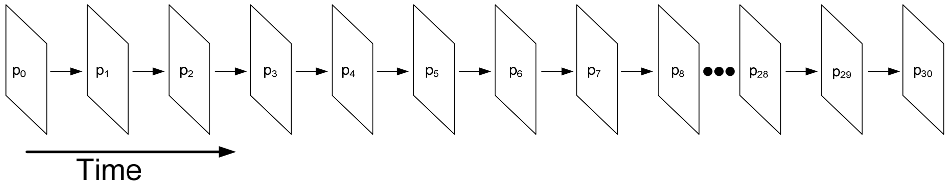
type of anchor picture (in AVC this could also be called a reference picture) which are pictures that can be used for prediction. The I-picture can be the start of a random access point in the video stream.

P-Pictures - These pictures can either use intra-frame compression or prediction from the nearest past anchor frame (forward predicted). The P-picture is also another type of anchor frame.

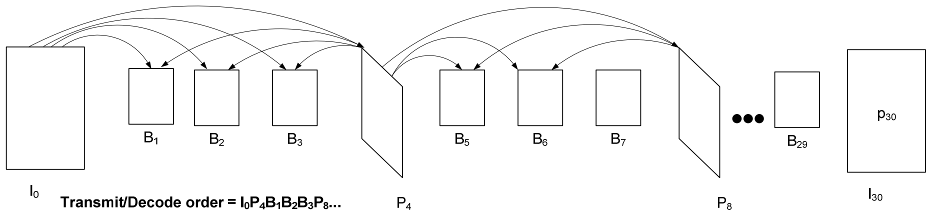
B-Pictures - In addition to intra-frame compression and forward prediction, these pictures can also predict from the nearest anchor frame in the future (backward predicted) as well as predict both forward and backward at the same time (bi-directionally predicted). The B-picture cannot be used in turn for prediction and is not an anchor frame.

There are a couple of things to note. Prediction is only dependent on the nearest anchor frames. This means that to decode a picture, at a maximum, one needs to store is a segment of the video stream that contains that particular picture and all pictures up to and including the nearest anchor pictures in both directions (so long as the anchor pictures are already decoded). In the reference picture buffer this allows for a natural “bumping” process to occur such that when an anchor picture gets stored and decoded then the oldest anchor picture can be discarded. Also of note, is that any picture acting as an anchor picture must come earlier in the transmit/decode order than the picture that requires that prediction (even though in cases of backward prediction that anchor picture is actually displayed later). Lastly, exiting (switching channels) before any anchor picture in transmit/decode order allows for the video displayed before the point to have no jumps in time (temporal continuity).

(a). Video Stream Segment in display order



(b). Typical MPEG – 2 Predictions with reference to the above picture segment in display order



(c). Possible Predictions in AVC Coded Video

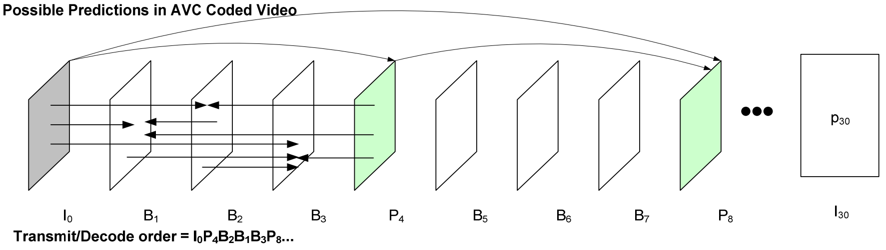


Figure 4. Possible MPEG-2 and AVC Prediction and Coded Pictures of Video Stream

The MPEG-2 GOP structure length and pattern indicates the video stream segment guarantees that all anchor pictures contained within are decodable and which allows for any pictures in smaller segments (mini-GOPS) predicted from the anchors pictures to be decodable. Having a random access point start at a beginning of a GOP allows for continuously displayable video from that point on. This in turn directly affects channel change time and the video quality in the channel change.

ADAPTING RANDOM ACCESS TO AVC

The MPEG-2 GOP structure is a very powerful mechanism for random access and the AVC encoding standard can accommodate this GOP structure, but this substantially limits temporal compression improvements that AVC can offer. Even though the concept of I, P, B are called the same as in MPEG-2, AVC has changed what they meant and also changed prediction rules between I,P, and Bs [see Figure 4]. These changes are:

Pictures – These are made up of combination of slices which can be of the same type or a combination of different types. Prediction from a slice requires storage of the entire picture. Pictures are differentiated between referenced (anchor) pictures and non- referenced pictures which are not used as predicted pictures. Using MPEG-2 Terminology (not really correct for AVC but often used), an “I” picture can have only I slices, “P” Picture can have I slices and P slices, a “B” picture can have I, P, and B slices.

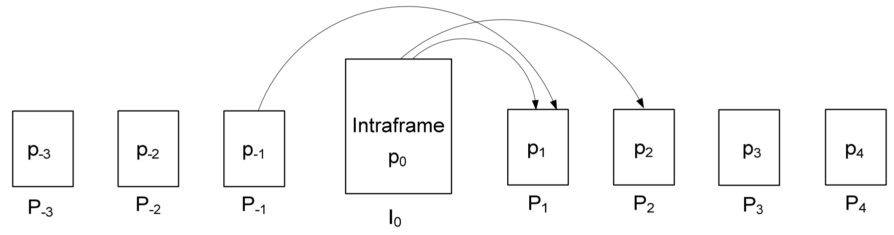
I, P, B – These refer to actual slice types in the pictures rather than the pictures themselves. A single picture can contain different combination of these slices depending on the number of slices allocated to pictures. An MPEG-2 like picture can be emulated by limiting the combination of slices in the picture.

Predicted Pictures – These pictures whether forward or backward predicted do not need to predict from the nearest reference (anchor) frame. They can refer to any reference frame stored in the Decoded Picture Buffer (DPB) and are limited to maximum number of reference frames storable at any one time. Though one may only need to predict only from one slice of the picture, the entire picture must be stored. To emulate MPEG-2 behavior simply limit prediction to nearest anchor neighbors.

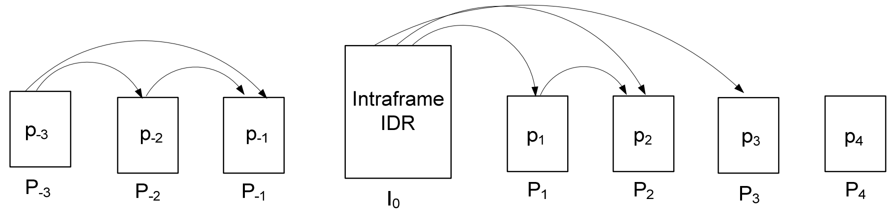
IDR – This is an intra-frame picture using intra compression techniques that also use spatial prediction between macroblocks. An IDR (Instantaneous Decoder Refresh) is a special intra-frame picture that indicates that any pictures subsequent to an IDR cannot reference (predict from) pictures before the IDR in transmission order (see Figure 5). Without an IDR tag, an intra-frame picture allows for pictures after it to reference or predicted from one that came before the intra-frame picture in transmission order. An MPEG-2 I picture can be emulated by an IDR. Some differences between using an IDR tag or not is that the bit rate spikes up around an IDR than if the tag was not used.

P - These slices can now be both forward or backward predicted. For MPEG-2 behavior, this would need to be limited to forward prediction to only I or P slices.

B - These slices can now be reference (anchor) frames and predicted in either direction from any two slices and can be weighted. To emulate MPEG-2 behavior, B’s should not be able to predict from a B and forward and backward dual prediction should be used without weighting.



(a) Prediction around I frame in AVC



(b) Prediction rule in AVC with IDR frame

Figure 5. Difference of Intra-frame with and without IDR Signaling

The MPEG-2 GOP structure can be contained in the design of a cable-friendly AVC video stream, but many of the temporal compression improvement techniques that the AVC standard provides would be lost.

Providing for random access in cable should not make the AVC stream conform to an MPEG-2 structure but rather accommodate it and other approaches. In order to do this, rules need to be developed (not implementations) that allow for fast channel change times and good video quality during the change while not preventing too much loss in compression improvements. In order to do this, a concept of a segment still needs to exist and it is important that the transmit/decode order and the display order do not get too far out of alignment.

Some possible things to consider are creating something similar to a GOP length in AVC which would bound the channel change time. Another thing would be to limit the time segment/ stream segment that contain reference pictures that could be used for decoding a particular picture. The use of backward prediction in AVC can be extended over longer windows which can give more compression efficiency, but may cause problems in some cable applications. Backward prediction needs to be constrained such that it does not affect existing cable applications or such that stream conditioning can handle infrequent trouble spots that might occur. To keep the decode and display order from getting too far out of alignment, bounding the presentation time from the decode time of the random accessed picture could be used. Additionally setting a bound between the random accessed picture and the first clean picture that allows for continuously displayable video from that point on would also be useful in this process. Lastly providing some rules to allow for points of time continuity in the stream would be helpful.

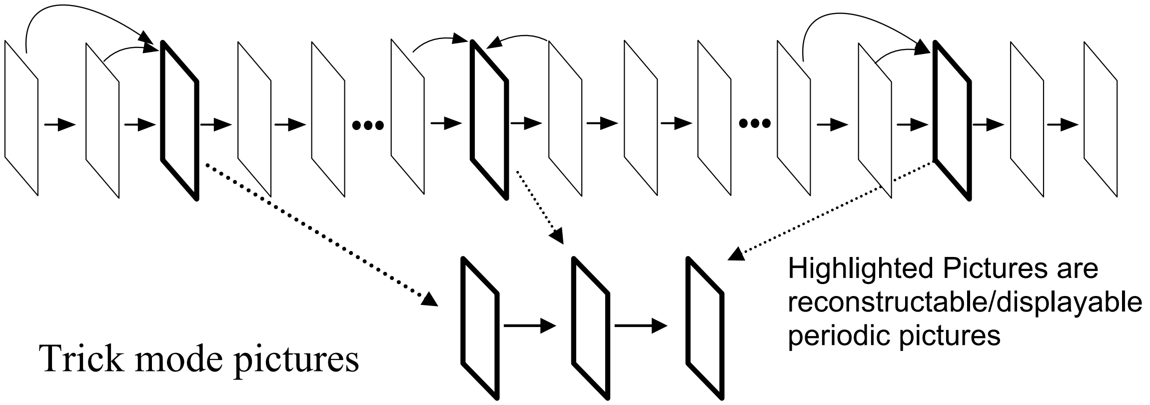


Figure 6. Generation of Trick Modes using Small Subset of Pictures

EMERGING APPLICATIONS: DIGITAL VIDEO RECORDERS

In the present cable environment, cable friendly video streams may also have to address other applications often in the same stream meant for broadcast. This would mean more stringent requirements on the RAP interval for streams that address these additional applications.

A popular application that is used with broadcast streams is the DVR storage and navigation of that stream. This is a service that viewers expect to have because it exists in the analog television and MPEG-2 domains today. In the AVC domain this is not as easy and unless constraints are used to ensure a compatible stream is provided, the DVR design may require a full decode/local encode of the stream prior to storage.

For the stream itself, regular self contained pictures or field pictures (for interlaced) need to be available at less than or equal to 1 second intervals. To avoid running decoder chips at faster than normal decode rates, the displayable pictures for DVR FFWD/RWD need to depend on a smaller subset of pictures in the RAP interval [see Figure 6 above].

Additionally slower FFWD/RWD rates that approach real-time playback need to consider smooth motion techniques between these displayable frames which would warrant a periodicity between sampled frames. On the transport level, several things need to be signaled about the video stream. One of these areas is identifying regular video starting points such as a start of the clean video display in the RAP interval. Another areas is the identification of reference versus non- reference pictures which identifies some pictures that may not be needed for a DVR FFWD/RWD decode. Lastly doing transport signaling in the clear can allow for manipulation of the video stream even though it may be encrypted on the hard drive.

AVC or any other advanced video codec needs to have the ability to do this. This may require consideration of the DVR application while designing the RAP interval. It would also require developing signaling in the transport layer which is a factor usually not considered in a video coding standard.

EMERGING APPLICATIONS: DIGITAL COMMERCIAL INSERTION

The digital program insertion (DPI) or insertion of local commercials or short programs is achieved by splicing from a network program stream to a local advertisement (ad) stream at a specific point in time. At the end of the ad a second splice occurs to get back to the network program. The timing of splice points is signaled by SCTE 35 [5] splice_info_section() messages carried in the transport stream (TS) packets of network program stream. In splicing, the exit point from the network program stream is called the OUT point or out of the network point. The start of the ad or the second stream is called the IN point. Similarly, at the end of the local ad-insertion window, the splicer exits the ad-stream and returns back to the network stream. So one DPI opportunity involves two splicings or two pairs of IN and OUT points [see Figure 7 below].

The requirement of splicing from one compressed stream to another compressed one is to ensure temporal completeness or temporal continuity of the exited stream. This is mainly achieved in MPEG-2 splicing by exiting on an anchor frame such as I or P picture. Sometimes this is called completing the mini-GOP. If one exits a stream in the middle of a group of B-frames, there will be a temporal jump in presentation space between the surviving B-frame(s) and the earlier anchor frame in decode order. When splicing between two AVC coded streams, this temporal continuity at an OUT point must exist as well. The use of IDR pictures can be helpful (but may not be necessary) to create this temporal continuity.

In the case of the IN point the stream is a new one to a decoder, so the stream should have all the parameters such as profiles and levels, picture resolution, etc., at the start that is needed for a decoder before starting to decode any picture. This is insured in MPEG-2 coded video stream starting with the sequence header, picture header, etc., followed by an I-frame. A similar solution in AVC would possibly be to start with a sequence parameter set (SPS), picture

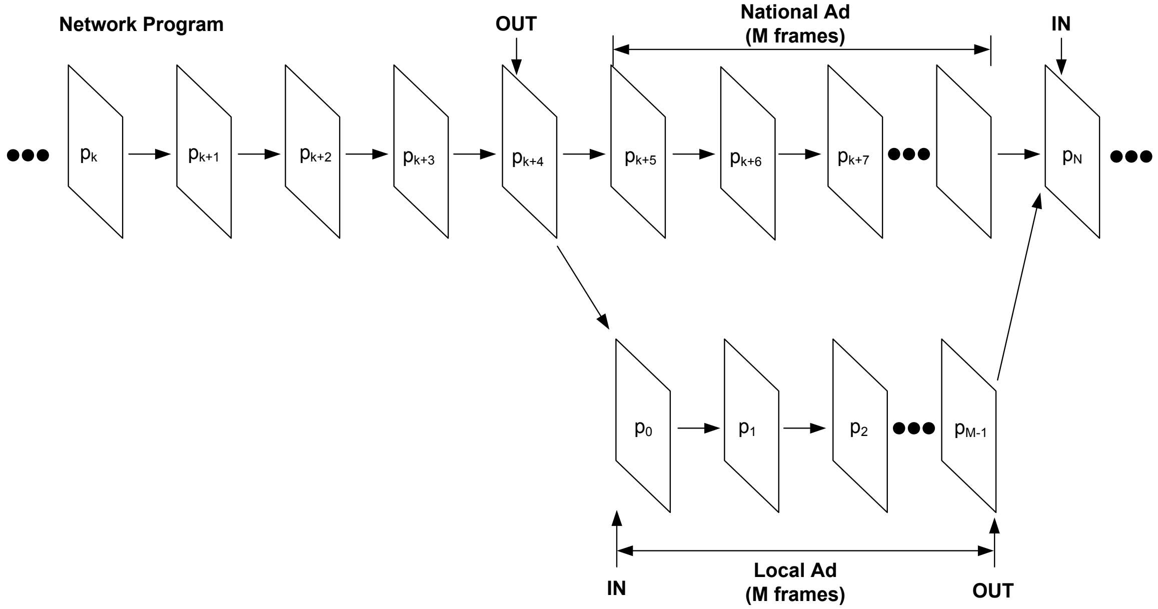


Figure 7. Simplified Diagram of Local Commercial Insertion

parameter set (PPS), etc., and an IDR picture. Also it may be recommended that one or more mini-GOPs may be encoded around any splice points just to avoid any splicing complexity due to delayed signaling for the splice points (ad-insertion). At the end of a commercial when the splicer returns to the network, the exited ad should also have- temporal continuity when returning to the network feed.

CONCLUSIONS

The AVC standardization process is in progress. Issues/requirements related to major broadcast applications such as channel change time, VOD, PVR, DPI have been investigated and defined. It has been apparent that some amount of bindings will be required to make it cable application-friendly. However, constraints will be minimized to achieve maximum compression efficiency while maintaining higher quality video than that produced by other similar standards such as DVB, Blue-ray DVD, HD-DVD, etc. The least amount of constraints will also leave room for encoders to produce a bitstream with potentially higher compression efficiency without compromising quality in the future.

ACKNOWLEDGEMENTS

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

They would also like to thank the members of the advance codec drafting group in SCTE for many helpful discussions.

REFERENCES:

1. ISO/IEC IS 13818-2, International Standard (2000), MPEG-2 Video.

2. ISO/IEC 13818-2: 1996/Cor. 1:2000 (E) MPEG 2 Video Technical Corrigendum 1.

3. ITU-T Rec. H222.0 | ISO/IEC 13818-1 (2nd edition, 2000): “Information Technology – Generic Coding of moving pictures and associated audio: Systems”- and ITU-T Rec. H.222.0 | ISO/IEC 13818-1:2000/AMD.3: (2003): “Amendment 3: Transport of AVC video data over ITU-T Rec. H222.0 | ISO/IEC 13818-1 streams”.

4. ITU-T Rec. 264 | ISO/IEC 14496-10 AVC: “Advanced Video Coding for Generic Audiovisual Services” (May 2003) and Amendment-1 to 14496-10 “Fidelity Range Extensions (2004) ”.

5. ANSI/SCTE 35-2004, Digital Program Insertion Cueing Message for Cable.

6. SCTE 43-2005, Digital Video Systems Characteristics Standard for Cable Television.

7. ANSI/SCTE 54-2004, Digital Video Service Multiplex and Transport System Standard for Cable Television.