Delivering a Real-Time HD AVC Platform

Introduction

Video compression products based on MPEG-4 AVC [1] are now ready for deployment that enable operators to deliver standard-definition (SD) video at half of today's required bit rates. This paper will outline how AVC evolved from familiar MPEG-2 [2] technology and describe the technology platforms now becoming available to deliver high-definition (HD) AVC.

The introduction of MPEG-2 was a considerable challenge for vendors and operators. Much of the technology was brand new and pioneers faced difficulties at every step, including the encoder, multiplexing, transport, modulation and decoder elements.

In contrast, the new AVC standard impacts only the decoder and encoder functions. Jointly developed by the ITU and ISO/MPEG standards committee, AVC has been designed for carriage over existing MPEG-2 transport and modulation infrastructures. The result is a standard that includes powerful efficiency improvement from:

- ¼ pixel motion prediction
- Segmented motion prediction (smaller block size options)
- Extended reference frame options
- New intra-prediction modes
- New block transform
- In loop de-blocking filter
- CABAC entropy encoding

MPEG-4 AVC effectively extends many MPEG-2 principles, leveraging the substantial additional processing and memory support available today. Early solutions will therefore cost incrementally more than legacy production MPEG-2 solutions, however, with strong open standards and the emergence of powerful integrated silicon, competitive forces will quickly ensure that the new technology is efficient, practical and affordable.

A realistic early target is for AVC-compressed material to require 50% less bit rate than equivalent quality MPEG-2. AVC technology demonstrations have clearly shown that the technology provides bit rate advantages and AVC will only improve over time. The advanced compression methods inherent in AVC have already proven to do a substantially better job with many sequences that present a challenge to MPEG-2.

AVC substantially enhances the business case for delivering bandwidth-hungry HD services. In fact, AVC offers the greatest value to those organizations wanting to deliver HD content. While efficiency gains for HD and SD will be similar, the lack of set top box legacy for HD will make deployment much easier and affordable.

HD compression currently poses a substantial technical challenge, necessitating at least six times more picture area to process than SD. A real-time platform to compress HD AVC with all the primary features and tools faces the impressive task of handling one thousand billion processing operations per second.

Despite these challenges, the first real-time HD AVC solutions have overcome the myriad of technical hurdles and are now being demonstrated.

What is AVC?

The new AVC standard represents the advanced evolution of familiar MPEG-2 technology. Jointly developed by the ITU and ISO/MPEG standards committee, AVC has been designed for carriage over existing MPEG-2 transport and modulation infrastructures. The encoding and decoding elements of AVC are themselves technically challenging, but by pushing the envelope in the early stages, the advanced standards ensure that the technology is very flexible and will not quickly become obsolete.

The standard is also designed to be applicable to a wide array of applications, scaling from the delivery of video to handheld devices all the way up to digital cinema applications. Therefore, there are many profiles and levels associated with the AVC standard. Main Profile at Level three (MP@L3) is the appropriate one for the delivery of full resolution 4:2:0 interlaced SD video while Level four (MP@L4) is appropriate for HD. At the July 2004 MPEG committee meeting in Seattle, a number of important High Profile amendments were approved; known as Fidelity Rate Extensions (FRExt) [3]. These extensions enhance the scope of AVC by offering higher fidelity options, such as 4:2:2 coding, greater sampling range and enhanced quantization tools to further extend the scope and value of AVC.

What Makes AVC Efficient?

AVC is an extension of established MPEG-2 foundation principles. It can be represented with very little change to the simple block diagram that describes MPEG-2.

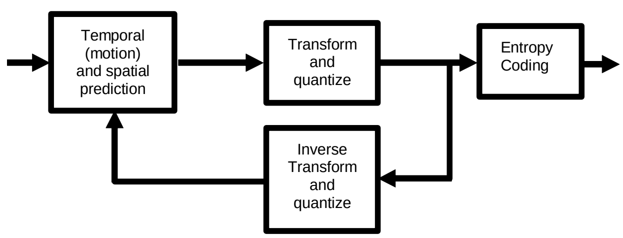

Fig. 1 Basic MPEG-2 Encoder Model

The stream details a sequence of compressed pictures—consisting of slices—which hold macro blocks that contain blocks with a transform description of the pixel structure. Both technologies rely on the principle that occasional spatially compressed pictures are interleaved with motion-compensated prediction pictures to reduce the vast amount of redundancy between adjacent pictures.

Maximum compression efficiency comes from the ability to describe the motion with the least number of bits across a long group of video frames. Motion is tracked on a block-by-block basis, in simplistic terms conveying with motion vectors that "this block was translated this amount in that direction." The imperfections with this technique will leave a residual error signal that needs to be conveyed along with the motion vectors. The AVC standard has several techniques that improve the accuracy of the motion-compensated prediction and dramatically reduce the resultant residual error signal that has to be carried. In practice, the motion prediction cannot solely convey the signal. To account for inaccuracies and identify a place to start decoding, reference pictures are inserted.

Substantial Gains Through Enhanced Prediction Modes

AVC achieves its most significant gains over MPEG-2 through substantial improvements to the motion-compensated prediction process. Step one is to double the accuracy of the motion prediction (AVC offers interpolated quarter pixel prediction). The second step is to use smaller block sizes to allow objects to be tracked more accurately. A third step is to use many more reference frames to search for a good motion predictive match. Lastly, efficient bi-directional predicted pictures, also know as B frames, are used more extensively. There are also numerous other subtle enhancements to reinforce the new technology and maximize its effectiveness.

In detail, the reference frames are known as I pictures; "I" relates to intra or spatial prediction. These frames act as a reference and a place to ensure clean channel changes. They also prevent any prediction errors from accumulating, which can cause further harm The AVC standard offers new modes that allow the intra references to be carried more efficiently. As with MPEG-2, accompanying the "I" pictures are two types of motion predictive pictures: "P"'s and "B"'s. The B frames have the greatest capacity to recognize good predictive reference from other frames and thus are the most efficient. They also outnumber the other pictures in the compressed structure. MPEG-2 uses the familiar IBBPBBPBBP…..structure, AVC follows this basic principle, but dramatically extends it by adding more B frames and offering greater freedom to reference the B frame to other frames.

The primary goal of the motion-prediction mechanisms is to minimize the residual error. This is especially true with the most challenging content. Therefore, the ability to convey the most difficult scenes sets the benchmark of usability—the lowest bit rate for acceptable video quality. Scenes that challenge MPEG-2 are handled more effectively with AVC. For example, scenes with repetitive motion, such as camera flashes, strobes, and conceal/reveal overlaps, are problematic for MPEG-2, but are handled well in AVC with the additional reference frame matches.

Fixed block sizing for motion-prediction processing represents a significant compromise. Large blocks are good for immense flat areas, while smaller blocks allow compact objects to be tracked more efficiently. The disadvantage of smaller blocks is the increased overhead required to carry additional motion vectors. Addressing this, the AVC standard offers improved prediction techniques that allow the motion vectors to be handled more effectively. MPEG-2 relies on motion prediction based on 16 x 16 blocks. AVC solves the compromise dilemma by offering an adaptive hierarchical scheme with block size options down to size 4 x 4.

AVC also extends the adaptive field or frame encoding mechanisms. MPEG-2 uses picture adaptive field or frame coding. AVC adds the tools to allow the field or frame coding to be adapted on a macro-block basis. This helps in applications where parts of the frame are better encoded in field mode and other areas in frame mode.

Improved Intra Prediction

The biggest gains with AVC are derived from updates to the motion-prediction tools. Smaller, but still useful benefits result from new intra-prediction modes that allow the spatially predicted frames and intra-coded macro blocks to be carried with fewer bits. In addition to requiring less bandwidth for equivalent quality, the new methods often provide visible enhancements through more consistent processing of flat areas.

New Transforms and Quantization

Once the motion vectors have been identified, the next stage in the process is to convey the residual errors. These error pictures need to be carried at a resolution that is appropriate to meet the desired bit rate targets. In AVC this is done using enhancements of proven MPEG-2 mechanisms. The first stage is to convert picture information from square pixel blocks into a stream of coefficients that convey the frequencies within the block. This results in a hierarchy of coefficients. It can be manipulated using a quantization technique to enhance or reduce the resolution and therefore offer precise control of the final bit rate.

MPEG-2 uses a discrete cosine transform (DCT) based on an 8x8 pixel block. This is effective for some applications, but imperfect, since errors in the math could lead to problems between the encoder and decoder. To get around this weakness, the new AVC technology uses a new 4x4 integer transform that is designed for accuracy, ease of processing and can be implemented using 16-bit integer values with addition and bit-shifting operations. The implementation of the new DCT is explicitly defined in the standard and will bring better decoded video as time goes on.

Another area that will yield significant gains over time relates to the bit allocation process. Quantization or bit rate control is the key part of the process, enabling the system to determine how to use bits wisely to attain the desired bit rate. The DCT is fully defined, with no room for improvement. In contrast, the quantization rate control process offers the potential for continued advancement over time. The processing cycles to support DCT and rate control are well served by general purpose CPU computation.

New In-Loop De-Blocking Filter

AVC takes advantage of a technique borrowed from the ITU-T H.263 [4] standard; implementing a filter in the inverse loop of the encoding process. This function improves the perceived video quality during moments of extreme compression stress when the quantization process has been forced to such an aggressive level that the block edge artifacts become visually prominent. The de-blocking filter, also referred to as the “in-loop filter”, processes the block edges and combines them with adjacent blocks. The benefits of this are most significant in the flat areas and during moments of compression overload. The processing cycles to support the de-blocking filter function are also well served by general purpose CPU computation.

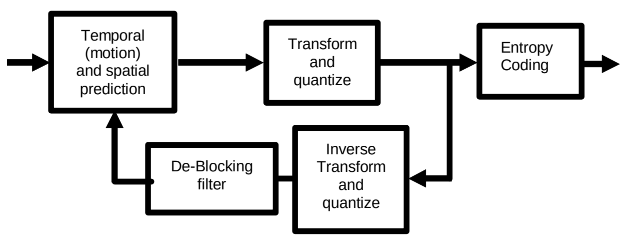

Fig. 2 Basic MPEG-4 AVC Encoder Model showing the De-Blocking filter. This the primary fundamental change to the basic MPEG-2 model.

New Entropy Coding Technology

The final stage in the compression process is entropy coding. Here, AVC Main Profile adds powerful technology features—including Context Adaptive Binary Arithmetic Coding (CABAC) and Context Adaptive Variable Length Coding (CAVLC). The principle is similar to the ZIP application on a personal computer. Payload is reduced with this loss-less process to enable more efficient delivery.

The CAVLC process is a more refined version of the Huffman entropy coding method used with MPEG-2. The difference is that the tables are context adaptive. The CABAC process offers substantial gains of about 20 percent at the expense of intensive computational resource for both the encoder and decoder. CAVLC is simpler and operates on discrete integer boundaries, but the trade-off is less efficiency gain. CABAC is a fully defined function, so there is no potential for improvement. Support for the most efficient CABAC mode is not mandatory, only the most advanced products will have enough processing power to support this capability. The vast amount of serial processing cycles to support the CABAC function can be supported on powerful general purpose CPU’s.

Unleashing the AVC Potential

The new AVC standard has attracted a lot of attention and deservedly so. It has also endured the initial marketing hype and uncertainties about real capabilities. By implementing all the main tools that the new standard offers, a realistic early target is for HD AVC to require 50% less bit rate than equivalent quality MPEG-2. The expectation is that the first constant bit rate AVC services will be deployed at between 1.5 and 2.0 Mbps for SD content and between 6 and 10 Mbps for HD. The first implementations deployed thus far have only hinted at the potential that the new AVC tools offer. There will be ongoing development to enhance the algorithms and this clearly this favors platforms based on a programmable approach.

Practical AVC encoding and decoding solutions must overcome substantial technical challenges. Fully supported SD AVC compression requires around 10 times more processing power than MPEG-2 and there is currently a dearth of ready-made silicon solutions for professional AVC compression applications. A real-time platform designed to compress HD AVC with all the features and tools could require more than 600 billion operations per second. Real-time AVC compression must have useable latency (e.g. two seconds), be standards compliant, never fall behind and endure all possible content scenarios.
There is plenty of motivation for vendors to develop silicon solutions for decoding, but much less for professional encoders. This is certainly true for real time HD AVC encoding and early solutions need to workaround the lack of silicon support.

Supporting the Rich AVC Toolset

The goal of video compression is to maximize distribution efficiency and get the best looking video at the lowest possible bit rates. The big challenge is to develop products that support full-resolution processing and include all of the practical tools that make AVC efficient: therefore including CABAC, MBAFF, de-blocking (loop) filter and multiple reference frames. The complexity of these new tools means that an existing MPEG-2 design augmented with additional processing is liable to struggle to support a useful set of tools. A more powerful approach is to combine the massive processing resource from CPU's with hardware acceleration in a scalable and fault-tolerant server platform. The CPU's of today are massively powerful, but their architecture is not universally suited to all video processing tasks. Hardware is suited to the block-based and pixel-level processing tasks, which are not efficiently handled by the CPU architecture alone.

The problem is that some of the tasks, like exhaustive search motion-estimation, use many compute cycles but are also very data-flow intensive, thus requiring complex register pipelines with fast memory access. This is best addressed with a hardware-centric approach designed to process the repeated Sum of Absolute Difference (SAD) calculations and identify the best motion prediction. The data comparisons are very repetitive and many of the calculations are re-used. Current processor based implementations tend to struggle to feed the Arithmetic Logic Units (ALU) from cache, but an FPGA approach can be customized to retain all the values in a register pipeline.

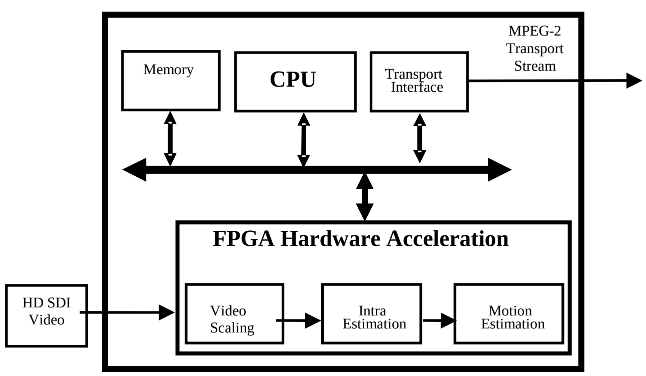

Fig. 3 SD Encoder Architecture

The diagram above shows the outline of an SD AVC implementation based on the use of powerful CPU's augmented with FPGA based hardware acceleration.

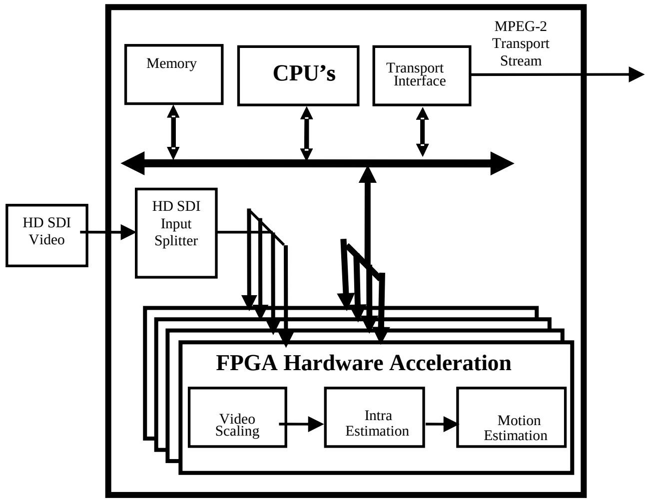

Fig. 4 HD Encoder Architecture

The diagram above shows how the SD architecture can be scaled up to support a variety of HD AVC implementations. HD solutions based on this approach have already been demonstrated. As with MPEG-2, the HD implementations are scaled up versions of the SD platforms. The practical approach is to partition the picture into horizontal stripes and implement a parallel array of processing to simultaneously process each stripe. The trick is to “hide the join” and ensure a perceptually seamless process. This requires that the encoder share reference information across the slice partition boundaries.

An advantage of combining general purpose CPU's with FPGA's is the continual improvement in performance and decreasing cost of the silicon and server platform technology. The equipment costs associated with a compact platform capable of a thousand billion operations per second is now affordable and economically deployable. AVC platforms based on this approach will leverage these economies of scale in new and innovative ways that will provide a fast return on investment while continuing to advance performance parameters significantly.

Deployment

HD AVC promises to dramatically enhance the economics and the practical ability of operators to deliver HD along with other bandwidth-intensive services. Compared to SD there is greater motivation to implement AVC because there is little or no legacy set top box scenario to be overcome.

Early AVC solutions will cost more than legacy production MPEG-2 solutions, but with strong open standards and the planned emergence of powerful integrated silicon, competitive forces will quickly ensure that the new technology, especially when incorporated into receive devices, is efficient, practical and affordable. Integrated decoder silicon that supports both SD and HD AVC is now sampling and by the end of 2005 HD AVC set top boxes should be available for less than $150.

Multi-channel video distributors and other facility operators are anticipating HD AVC deployments at around 10 Mbps, with some looking to deliver content in the 6 Mbps region. As the technology develops, performance increases and cost continue to come down, HD content delivery will become the mainstream. Advanced stream processing functions now available with MPEG-2 will be developed and readily available for AVC, (such as splicing and rate shaping).

After many years of false starts, HDTV is now overcoming all the technical and economic obstacles that have hampered its universal adoption. There is now abundant HD content available while HD-capable receivers are attaining mass market price points. HD is finally being heavily promoted by the major cable and DTH operators in the U.S. market. The technology also strongly positioned in Australia, Korea and Japan and is once again getting attention from European countries.

Summary

Operators around the world are beginning to understand what AVC compression can do for their businesses. After careful consideration, the trend is to eschew other compression methods in favor of implementing AVC into their facilities and distribution platforms.
The challenges are being overcome and prototypes of affordable real-time HD AVC solutions have already been demonstrated. In addition, the availability of set top boxes based on integrated silicon will mean that HD AVC will be a deployable reality in 2005.

We live in a world of pent up HD demand and limited bandwidth. The availability of HD AVC products and systems is very timely.

Acknowledgements

The author wishes to thank his colleagues at Modulus Video for their assistance.

References

[1]. AVC ref. ISO/IEC 14496-10:2003 - Coding of audio-visual objects -- Part 10: Advanced Video Coding
[2]. MPEG-2 ref. ISO/IEC 13818-2:2000 - Generic coding of moving pictures and associated audio information
[3]. FRExt Amendment to ITU-T Rec. H.264 | ISO/IEC 14496-10 Advanced Video Coding, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG
(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), July 2004
[4]. ITU-T-T Recommendation H.263: Video coding for low bit-rate communication Version 1, November 1995; Version 2 (H.263+), January 1998; Version 3 (H.263++), November 2000.