

Delivering Video at Half the Bit Rate with AVC

Introduction

MPEG-2 [1] video compression has been around for around 10 years but a compelling new alternative known as MPEG-4 AVC [2] is now available. The advantage of AVC over the older MPEG-2 is increased efficiency. AVC (also known as MPEG-4 Part 10 and H.264) delivers video at half the bit rate of MPEG-2.

AVC was jointly developed by the ITU-T and ISO/MPEG standard committees, based on MPEG-2 foundations it is designed to carry signals over the existing MPEG-2 transport and modulation infrastructure. The encoding and decoding elements of AVC are complex and technically challenging, ensuring that AVC technology will improve over a substantial period of time before the technology matures.

Products based on MPEG-4 AVC are now ready for deployment, enabling operators to deliver standard-definition (SD) video at significantly lower bit rates. This paper will outline how AVC has evolved from the familiar MPEG-2 methods and describe the practical technology required to deliver both SD and high-definition (HD) AVC content.

MPEG-AVC Video Compression

Industry standards are designed to ensure interoperability between encoders and decoders but they also serve to stimulate open competition in the marketplace. This competition benefits operators by driving performance up and costs down over time and stimulating the proliferation of peripheral products and applications.

AVC follows the highly successful MPEG-2 principle of a tightly defined standardized decoder model. The standard does not define the encoding process, only the bit stream syntax and decoding method.

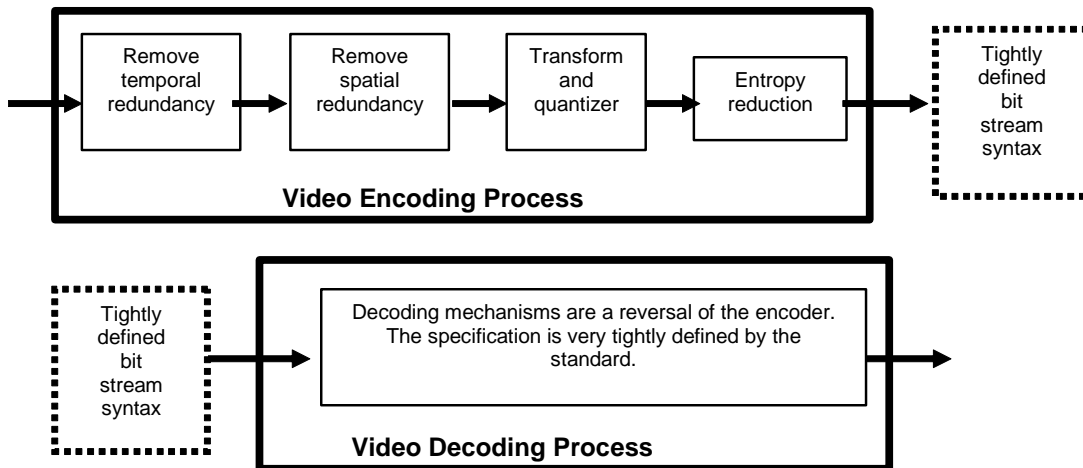


Figure 1 MPEG Video Compression

The result is unlimited flexibility for this robust compression technology to be developed and refined over time. This decoder-centric model is designed for low cost decoders which makes economic sense given the prolific numbers of decoders relative to the number of encoders.

In the mid-90's the introduction of MPEG-2 proved to be tough for vendors and operators. Much of the technology was brand new and the technical challenges were numerous. Pioneers faced

difficulties integrating the compression, multiplexing, transport, modulation and decoder elements. In contrast, the new AVC standard is an advanced evolution of familiar techniques and primarily impacts the decoder and encoder functions. Jointly developed by the ITU and ISO/MPEG standards committee, AVC was ratified in May 2003.

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 video while Level four (MP@L4) is designed for HD. At a meeting in Seattle, in July 2004, the MPEG committee approved a number of important High Profile amendments; 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, and greater sampling range to further extend the scope and value of AVC.

There has been some confusion regarding the various references used to describe the technology. AVC is also known as MPEG-4 part 10 and is quite different from MPEG-4 Part 2 [4], a technology that failed to deliver adequate gains over MPEG-2. The object-based coding attributes of MPEG-4 Part 2 were excellent for computer-generated synthetic content, but not practical for the delivery of real-time video, and MPEG-4 AVC does not include this feature. Complex licensing issues also limited the deployment MPEG-4 Part 2. AVC is also sometimes referred to as the ITU-T H.264 video compression standard.

Efficiency from New Tools

Simply put, AVC is a significant evolution of familiar MPEG-2 principles and can be represented with a block diagram very similar to that of MPEG-2. The stream describes a sequence of compressed pictures—consisting of slices—that contain macro blocks with a transform description of the pixel structure. Both AVC and MPEG-2 rely on the principle that occasional pictures compressed spatially are interleaved with pictures that predict and describe the motion. This combination offers a practical way to extract the huge amount of redundancy that exists between adjacent pictures.

Maximum efficiency in video compression comes from finding ways to describe the motion with the least number of bits across a long group of video pictures (GOP). Motion is tracked on a block-by-block basis, conveying with motion vectors that "this block was translated this amount in that direction." The imperfections with this technique result in a residual error signal that is conveyed along with the motion vectors. The AVC standard has several techniques that improve the accuracy of the motion-compensated prediction to dramatically reduce this resultant residual error signal that must be carried. In practice, the motion prediction algorithms cannot solely convey the signal. To account for inaccuracies and to offer a place to start decoding, reference pictures are regularly inserted. The following sections describe in more detail the technology and enhancements in each functional block. At a high level the only difference between the AVC model and the MPEG-2 model is the addition of a loop filter.

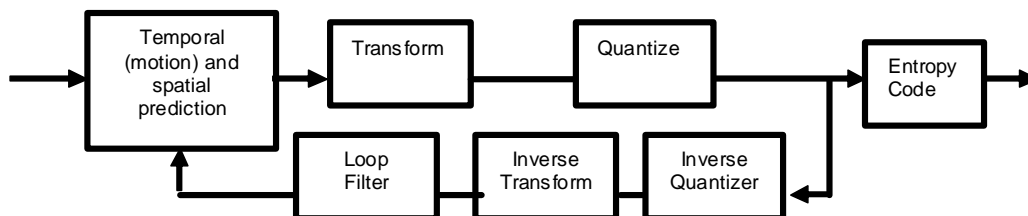


Figure 2 Simple AVC Functional Model

Enhanced Prediction Modes

The superiority of AVC over MPEG-2 is largely due to substantial improvement of the motion compensated prediction. First, AVC doubles the accuracy of the motion prediction by

implementing quarter pixel interpolation. Secondly, objects can be tracked more accurately by using the option to use smaller block sizes. And finally better matches can be found across frames by using the option to search across many more frames. These are simply the highlights; there are numerous other subtle enhancements that further enhance AVC over MPEG-2.

In technical terms, the reference frames are known as I pictures; "I" refers to intra-frame or spatial prediction. These frames do not reference preceding or subsequent frames and act as a place to ensure clean channel changes and as a reset point that prevents prediction errors from accumulating. The AVC standard offers new modes that allow the intra references to be carried more efficiently. As with MPEG-2, the "I" pictures are accompanied by two types (forward and bi-directional prediction) of motion predictive pictures: known as "P" and "B" frames. The B frames have the greatest capability to find good predictive reference from other frames and, because they are the most efficient, outnumber the other pictures in the compressed structure. MPEG-2 used the familiar IBBPBBPBBP picture structure, AVC adds to this principle by allowing an additional B frame to support IBBBP structures and extend the number of reference options.

The primary goal of the motion-prediction mechanisms is to generate accurate vectors that result in minimal residual error. This is especially true with the most challenging content, such as fast-moving sports scenes. The ability to convey the complex scenes with adequate video quality sets the benchmark bit rate for usability. The scenes that most challenge MPEG-2 are handled more effectively with AVC. For example, the advanced AVC new tools generate lower residual error than MPEG-2 for sequences such as repetitive motion, fades/dissolves, camera flashes, strobes and conceal/reveal overlaps benefit most noticeably from the advanced new tools that AVC has to offer.

Fixed block sizing for motion-prediction processing is fundamentally a compromise. Large blocks are good for flat areas, while smaller blocks allow compact objects to be tracked more efficiently. The tradeoff with smaller blocks is the increased overhead required to carry additional motion vectors. MPEG-2 relied on motion prediction based on a fixed 16x16 blocks. AVC improves on MPEG-2 by supporting an adaptive hierarchical scheme with block size options down to 4x4.

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 Modes

Improvements of AVC over MPEG-2 go beyond motion prediction. AVC also supports new intra-prediction modes that allow the spatially predicted frames and intra-coded macro blocks to be described with fewer bits, improving compression and providing more consistent processing, especially in flat areas.

AVC Prediction Options - Huge Development Potential

AVC deserves the attention it has been getting but it is not immune to the usual hype that surrounds a new technology. In the near term it is realistic to expect AVC to match high quality MPEG-2 at half the bit rate. However, AVC is poised at the start of its technology lifecycle. AVC tools offer an array of permutations and development possibilities for improving video compression that are only beginning to be explored and exploited. The AVC algorithms are flexible with tremendous potential to be refined and enhanced. This favors a programmable approach to encoder design. Even though processing support for transform and rate control are well served by general purpose CPU's the block-based processing and search mechanisms are not. DSP chips can be used, but a more powerful and flexible approach is to employ

programmable hardware such as an array of Field Programmable Gate Arrays (FPGA's). Implementations that combine FPGA's and general purpose CPU's offer the ideal combination of processing efficiency and power to fully support AVC encoding.

New Transforms and Quantization

After determining the motion vectors, 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. AVC accomplishes this with enhanced versions 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 array. The result is a hierarchy of coefficients that enables precise control of the final bit rate by using a quantization technique to reduce the resolution.

MPEG-2 used a discrete cosine transform (DCT) based on an 8x8 pixel block. This was effective, but mathematical imperfections could introduce additional error between the encoder and decoder. AVC overcomes this weakness with 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. This new transform is explicitly defined in the standard and will provide significantly better decoded video.

Even though the AVC transform function is fully defined there are many opportunities to continually improve encoding efficiency. The AVC standard does not define the process for managing quantization and optimizing the efficient use of available bits. These areas provide rich opportunity for achieving significant performance improvement over time.

New De-blocking Filter

During moments of extreme compression stress, when the block edge artifacts become visually prominent, the de-blocking filter (also known as "loop filter") processes the block edges to help the edges merge with adjacent blocks. AVC takes advantage of this technique from the ITU-T H.263 [5] standard to implement a filter in the inverse loop of the encoding process. The benefits are most significant in flat area such as background sky scenes as well as facial features. General purpose CPU's can also provide the processing cycles to support the de-blocking filter function.

Enhanced Entropy Encoding

Entropy coding is the final stage in the compression process. The principle is similar to the ZIP application on a personal computer. You input a data stream to get a smaller stream out. Unlike quantization, the entropy coding process is 100 percent reversible, with zero loss. AVC Main Profile adds powerful technology options, known as Context Adaptive Binary Arithmetic Coding (CABAC) and Context Adaptive Variable Length Coding (CAVLC). The entropy coding modes are fully defined encoder functions, so there is little potential for improvement. However, support for the most efficient CABAC mode is not mandatory and only the most advanced products are likely to offer support for this feature.

In principle the CAVLC process is a more refined version of the MPEG-2 Huffman entropy coding. The difference is that the tables are context adaptive. The CABAC process offers substantial gains of about 15 percent. It is, however, computationally intensive for both the encoder and decoder. CAVLC is simpler and operates on discrete integer boundaries, but the tradeoff is less bandwidth efficiency gain. The vast amount of processing cycles to support the CABAC function is well served by powerful general purpose CPU's.

Alternative Technologies

The emergence of AVC will make it difficult for proprietary, next-generation alternatives to prosper. Operators typically assume that an open standard technology will create a competitive marketplace that improves price/performance faster and generates a more vibrant ecosystem of peripherals and applications than proprietary technologies.

One alternative from the PC world that has survived and has received plenty of attention is Microsoft Windows Media technology. Microsoft has heavily promoted its Windows Media technology as the new compression format for video across all devices, not just PC's. Microsoft is in the process of standardizing this format, now known as VC1 through SMPTE. As a result of the standards effort Microsoft has revealed that the underlying technology is quite similar to AVC. VC1 differentiates itself by adopting a toolset that is heavily optimized for decoding on PC's but the cost of omitting some tools such as CABAC that would enable greater bandwidth efficiency.

With full support from the highest levels, Microsoft's VC1 is likely to establish itself in some market areas. For example both AVC and VC1 have been included as advanced compression standards for the next generation of DVD's that will support high-definition content.

AVC Deployment Scenarios

MPEG-4 AVC has shown significant efficiency gains over MPEG-2, an advantage that is compelling for operations where bandwidth is limited or expensive. The relative economy of AVC relative to MPEG-2 is compelling enough to ensure widespread deployment of the new technology in the near term. AVC effectively extends many MPEG-2 principles but is a big evolutionary step. The new features do require substantial additional processing and memory support that are way beyond the capabilities of MPEG-2 platforms. Early implementations may cost incrementally more than legacy MPEG-2 solutions, but open standards, market competition and the emergence of powerful integrated silicon will quickly take AVC beyond the technology limits constraining MPEG-2. Integrated silicon that supports both HD and SD AVC is now sampling and by next year HD AVC Set Top Boxes (STB) are expected to be available for less than \$150.

The greatest interest in AVC has come from point-to-point, satellite-to-home, and video over DSL, and IP network applications. Encoder and professional decoder products have been launched to deliver AVC over existing MPEG-2 transport infrastructures.

High Definition

After many years of false starts, high-definition television is now overcoming all the obstacles that have hampered its introduction. HD content is now abundant and HD-capable receivers are attaining mass market price points. Most importantly HDTV is finally being heavily promoted by the major cable and DTH operators in the huge U.S. market. HD is also strongly positioned in Australia, Korea and Japan and strong European interest resurfaced again at IBC2004 conference in Amsterdam. Interest in AVC is very timely as it dramatically enhances the practical ability of operators to profitably deliver HD and other bandwidth-intensive services.

AVC Platform Implementation

Practical AVC encoding and decoding solutions must overcome substantial technical challenges. SD AVC compression requires around 10 times more processing power than MPEG-2 and there is now a dearth of ready-made silicon solutions for professional AVC compression applications. There is plenty of motivation for vendors to develop silicon solutions for decoding but not for professional encoders. HD encoding compounds the challenge. A real-

time platform designed to compress HD AVC with all the features and tools could require more than 600 billion operations per second.

Off line encoding is far simpler than real-time encoding and lends itself to software applications running on powerful computers. Real-time AVC compression must have useable latency (e.g. two seconds), be standards compliant, never fall behind and endure all possible content scenarios.

Hardware or Software – Which Works Best?

The goal of video compression is to maximize efficiency and get the best looking video at the lowest bit rates. The big challenge is to develop products that support full-resolution processing and all the tools that make AVC efficient; including CABAC, MBAFF, de-blocking (loop) filter and multiple reference frames. One powerful approach is to combine CPU's with hardware acceleration in scalable and fault-tolerant server platforms. Moore's law has ensured that CPU's today are massively powerful, but their architecture is not naturally suited to video processing. Hardware is suited to the block-based and pixel-level processing tasks that 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 met with a hardware approach designed to process the repeated Sum of Absolute Difference (SAD) calculations to find the best motion prediction. The data comparisons are very repetitive and many of the calculations are re-used. Currently the CPU and DSP 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.

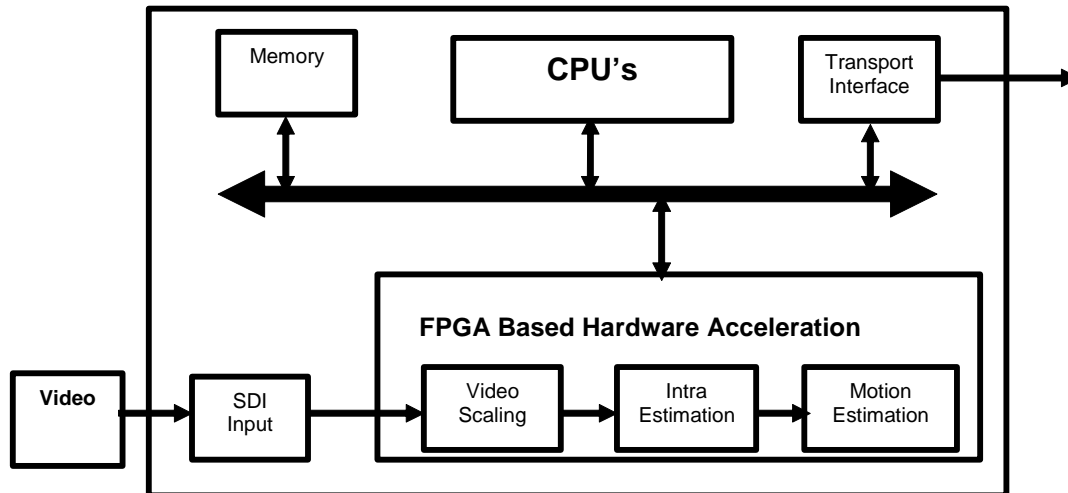


Figure 3. An SD implementation based on the use of powerful server CPU's with FPGA based hardware acceleration

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.

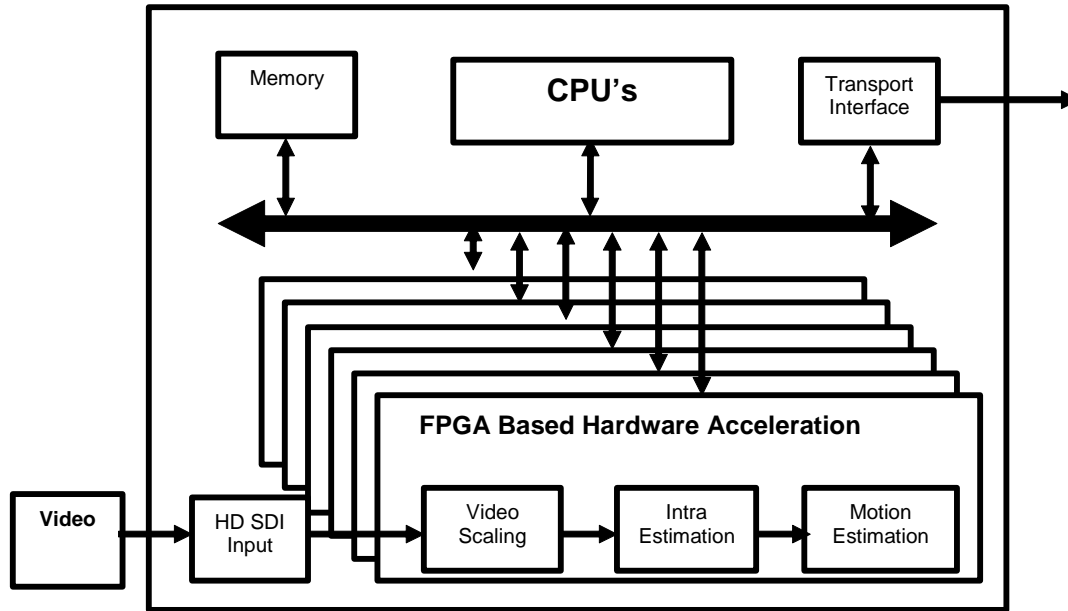


Figure 4. HD architecture implemented as a scaled SD implementation

The diagram above shows how the SD architecture based on FPGA plus CPU's can be scaled to support HD AVC implementation. The trick with the HD implementation is to share resources across the partition boundaries to ensure a seamless process.

Moore's law continues to play an important role in pricing and product advancement, so the above AVC architectures can look forward to even more affordable processing power becoming available to help extract the vast potential from the rich AVC toolset.

Performance

A series of product demonstrations clearly show that AVC can already do a substantially better job with many sequences that present a challenge to MPEG-2. Moreover AVC is at the beginning of its technology lifecycle and will continue to improve significantly over time. The expectation is that the first constant bit rate AVC services will be deployed at around 1.5Mbps for SD and between 6 and 8Mbps for HD.

Summary

MPEG-4 AVC has emerged as the new open standard with the attributes of efficiency and practicality that will ultimately replace MPEG-2 for the delivery of video. MPEG-2 has served operators very well, with performance and costs having improved dramatically over the past decade but has become a mature technology with little scope for continued enhancement. AVC technology is sure to follow a similar technology lifecycle.

Practical AVC solutions are now available to deliver the new standard over existing MPEG-2 transport infrastructures and are now being deployed. HD solutions will follow shortly, with the first prototype products expected in early 2005.

References

- [1]. MPEG-2 ref. ISO/IEC 13818-2:2000 - Generic coding of moving pictures and associated audio information
- [2]. AVC ref. ISO/IEC 14496-10:2003 - Coding of audio-visual objects -- Part 10: Advanced Video Coding
- [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]. MPEG-4 ref. ISO/IEC 14496-2:2001/Amd 3:2003 - New levels and tools for MPEG-4 visual
- [[5]. 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.