Silicon Image™

White Paper

White Paper Overview

- Introduction
- Video Compression Background
- Aspects for H.264 Video Decoder Implementations
- Implementing H.264 3D, 4K, and Ultra-HD Video Decoders
- Summary

3D and Ultra HD Video

The Next Revolution

December 2010

AUTHORED BY ANTHONY HERNANDEZ - (415)786-2018 - anthony94122@outlook.com

Introduction

The U.S. television industry has undergone a long process of transformation from the first analog color broadcasts in 1953 to the early high-definition TV (HDTV) broadcasts in 1996 and beyond. Since then, HDTV technologies have permeated the marketplace; today, high-definition (HD) video can be sent and received via cable, satellite, DSL, terrestrial broadcast, Blu-ray Disc™ players, and digital TVs (DTVs). Large-screen HDTVs are now ubiquitous in homes, public venues, and commercial/business environments. Even the latest smartphones are capable of recording HD video, decoding HD video streams, and playing HD video on DTVs using an HDMI® (High Definition Multimedia Interface) cable. Major advancements in digital video compression and decompression technologies such as the MPEG-1 and MPEG-2 compression standards developed by the Moving Picture Experts Group have fueled this transformation. MPEG-4 Part 10 (also referred to as MPEG-4 AVC, Advanced Video Coding, and more commonly as H.264) was recently developed in collaboration with the International Telecommunications Union (ITU) and has dramatically improved video storage, communications, and display capabilities. It has taken more than 50 years for video imaging to advance from the low-resolution analog CRT to the HD LCD DTV, but the next digital video imaging revolution will occur far more quickly. To quote InStat[1] analyst Michele Abraham, "Over the next 5 to 10 years, major advancements are planned for both video and audio technologies, where video resolutions in future HDTVs may increase by 4 to 16 times compared to HDTVs on the market today." Continued advancements in displays, communications, storage, algorithms, and semiconductor technology will keep making cutting edge visual experiences available to consumers.

Video imaging is one of the few remaining applications that require the convergence of many new technologies to improve realism. This whitepaper discusses the future of digital image processing with an emphasis on high-resolution video encoding technologies such as H.264, which is a critical component for displaying high-quality video in both consumer and professional marketplaces.

[1] InStat is a multimedia-market research firm

Video Compression Background

History

Vision accounts for 80% of the information humans receive about the world, which explains our ongoing fascination with visual media and means of delivery. Simple cave paintings celebrating successful hunts evolved into complex drawings and paintings that captured significant historical events and beautiful scenery, and illustrated our stories and legends with ever-improving techniques for mimicking 3D perspective on a 2D surface. Artists, engineers, and others leveraged advances in imaging techniques and technologies to bring ever more realism to their creations. Color analog television debuted in the United States in 1953 and in Europe in the 1960s. The Japanese pioneered HD analog TV but quickly discontinued it in favor of digital HDTVs that offered improved realism in smaller, cheaper devices. The ongoing digital revolution has brought digital HDTV to the masses, aided by digital compression technologies such as JPEG, MPEG, and H.264 to deliver ever-improving realism.

Digital Video

It is necessary to both reduce the size of imaging data while preserving video quality when converting from analog to digital imaging. Mathematical algorithms reduce or eliminate unnecessary color information beyond human perception and eliminate repeat or redundant patterns that occur in moving pictures. Compression (also called encoding) is the process of reducing the size of digital information, and decompression (also called decoding) is the process of recreating the original image. Modern algorithms like H.264 can greatly reduce file size while maintaining full image quality.

Lossless vs. Lossy Compression

Compression can be either lossless or lossy. Lossless compression recreates the exact original data throughout the encoding and decoding processes and is required for data files such as text or financial information to preserve the integrity of the data. This accuracy comes at the cost of larger files relative to the original data and is therefore more expensive to store and transport over a network, making it impractical for most still and video imaging applications. Lossy compression sacrifices some accuracy by removing information that the human eye cannot see, which can greatly reduce file size. The difference in file size can be significant: Lossless compression techniques reduce data size by anywhere from 0-3x while lossy algorithms can reduce still image sizes by 10x and video images by up to 300x. For example, a 10MB raw still camera image can be reduced to a 1MB JPEG with very little observable quality loss. A video compressed using H.264 can achieve a compression ratio of over 300. Even so, storage and data transmission limitations may require even higher compression levels, especially as digital cinema resolution and 3D movies being making their way into the home.

Lossy digital image compression uses the following three methods:

- **Vector Quantization:** This method is used by some audio compression standards and has some academic and research potential in signal processing, but has not been practically implemented for video compression with a few past exceptions.

- **Discrete Wavelet Transform (JPEG2000):** This standard is gaining wide acceptance in the professional cinema marketplace because it combines virtually lossless compression with the highest image quality; however, these lower compression levels and very large resulting file sizes make JPEG2000 impractical for consumer applications.

- **Discrete Cosine Transform (DCT) and the similar Integer Transform (used for JPEG, MPEG, and H.264):** These are the workhorses of the consumer-level video compression techniques. These methods convert image data from the spatial to the frequency domain by using DCT transformation techniques. DCT is used by digital cameras (JPEG) and DVD players and satellite broadcasts (MPEG). Integer Transform produces better quality algorithms while also less processing power than DCT, and is the method used by H.264. H.264 currently produces the highest compression levels without significant image quality degradation.

DCT and Integer Transform are adopted by block-based compression methods that:

- Sample an image at regular intervals and disregard certain image information that cannot be perceived by the human eye, and

- Search for repeating patterns from image to image to eliminate redundant information.

Combining these techniques greatly reduces video file sizes. Block-based compression standards include JPEG, MPEG-1, MPEG-2, H.264, and others. JPEG focuses exclusively on compressing single static images and is very functional for digital still cameras; it does not look for redundant information between frames in a moving picture. Moving pictures require compression capabilities in the time domain such as MPEG-1, MPEG-2, and H.264, which detect and take advantage of repeating patterns from frame to frame, thereby dramatically improving compression.

MPEG-1 debuted in 1992 followed by MPEG-2 in 1995, which manages video compression more efficiently and supports both standard-definition (SD) and high-definition (HD) formats and is primarily used by satellite, cable, and terrestrial broadcasts. MPEG-4 was launched in 1998 and became popular for use in personal computers and Internet broadcast standards such as DivX. Today, H.264 (also known as MPEG-4 Part 10 or MPEG-4 AVC) offers improved compression levels that more efficiently handle high-resolution video and frame rates.

Block-based video compression typically uses the following three types of video frames:

- **I-Frames (Intra-coded picture):** I-Frames are also used to begin a GOP (see below) or a new scene, or used as a point to restart a video stream when a transmission error occurs and is therefore sometimes referred to as the anchor frame. I-Frames are compressed using algorithms similar to those used in the JPEG digital still camera images because each I-Frame is a self-contained picture and can be independently recreated during the decoding step without reference to any other P- or B-Frames. They are also used in trick modes (fast forward/rewind) that skip P- and B-Frames. An I-Frame may be very large in size compared to a P- or B-frame because each I-Frame contains enough information for the complete image.

- **P-Frames (Predictive-coded frame): P**-Frames reference parts of earlier I-and/or P-Frames to create the current P-Frame. This makes P-Frames much smaller in size than I Frames but also makes them very sensitive to transmission or storage errors because their content depends on the integrity of previous P- and I-Frames. A single error in one P Frame could damage numerous future P Frames and B-Frames, possibly corrupting the entire GOP. Error concealment techniques can mitigate the damage in different ways such as by displaying the last good frame until the next I-Frame is displayed.

- **B-Frames (Bi-directional coded or Bi-predictive frame):** B-Frames look at the difference between the current frame and

both the previous and future frames to construct the image changes.

In general, I-Frames contain the complete image while P- and B-Frames only contain the changes in the current frame compared to surrounding video frames, which means that the majority of video compression takes place in the P- and B-Frames. Frames are typically grouped into sets called Groups of Pictures (GOP). Each GOP consists of one I-Frame or anchor frame and numerous P- and B-Frames; a typical GOP will contain 12-15 frames that provide ½ second of video. For example, a 15-frame GOP will have one I-frame for every 15 frames. Film is shot at 24 frames per second (fps); HD video is therefore often compressed and stored at 24 fps as well. In this case, a GOP could contain 24 frames or 1 second of video.

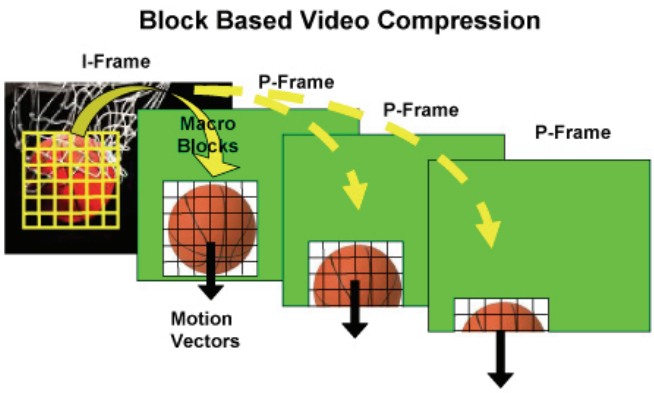

Figure 1: Block-based video compression using I- and P-frames

There are some significant differences between the different block-based approaches; however, most MPEG-1, MPEG-2, MPEG-4, and H.264 digital video compression uses these techniques to compress and decompress digital video.

The compression techniques used in MPEG and H.264 require increasingly complex mathematical algorithms to produce higher levels of compression while also maintaining image quality because of the market need to support ever-increasing video resolution and frame rates. For example:

- An HDTV has 4 times the resolution of video provided by a standard-definition DVD player
- 3D video requires twice the frame rate as 2D video.

The rapidly growing customer demand for still higher resolutions and frame rates to provide additional realism is making the algorithms progressively more complex and beyond the reach of software-only implementations. HD video compression and decompression now uses both semiconductor devices and software but improved compression techniques are also required to keep video transmission bandwidth and file size within the capabilities of communications systems and storage devices. For example, the H.264 compression techniques compress video to between 1/3 and ½ the size of MPEG-2 while maintaining similar image quality but also requires 3 to 5 times the processing power to handle the complex compression algorithm. Moving from SD MPEG-2 to HD H.264 therefore requires about 12 to 20 times the processing power for video decoding since HD video has about 4 times the resolution of SD video. So far, rapidly expanding semiconductor device capabilities have helped compensate for the increased processing demands.

Today, most leading edge broadcast or storage applications use H.264 compression techniques for HD and higher resolutions because it is the most efficient compression algorithms on the market. The latest satellite and cable systems use H.264 video compression and decompression because they can store and transmit more movies through their existing communications infrastructure than they can with MPEG-2 or MPEG-4.

	Compression Standard	Horizontal Resolution (Pixels per Line)	Vertical Resolution (Lines/Frame)	Typical Application
SD	**MPEG-1/2**	**720**	**480/576**	**VCD, SVCD, DVD, STB**
HD	**MPEG-2 /H.264**	**1,920**	**1,080**	**STB, Blu-ray Disc, HD Camcorder**
4K	**H.264**	**4096**	**2304**	**Professional Camcorder, Cinema**
8K	**H.264**	**7680**	**4320**	**Professional Studio Devices, Cinema**

Table 1: Typical video resolutions, compression methods, and applications

Video Compression and Decompression

The colloquial term "video compression" actually refers to both compression (encoding) and decompression (decoding). Compression uses various algorithms to reduce the size of the original digital video stream created by HD digital camcorders or a telecine process that scans 35mm film at high resolution to produce digital video that is then compressed. Compressed video is stored on a device such as a hard drive, DVD, or Blu-ray Disc or it can be transmitted to the home via the Internet, satellite, DSL, or terrestrial broadcast for further processing. Finally the DTV accepts the decompressed and restored video via an HDMI cable and displays the original image.

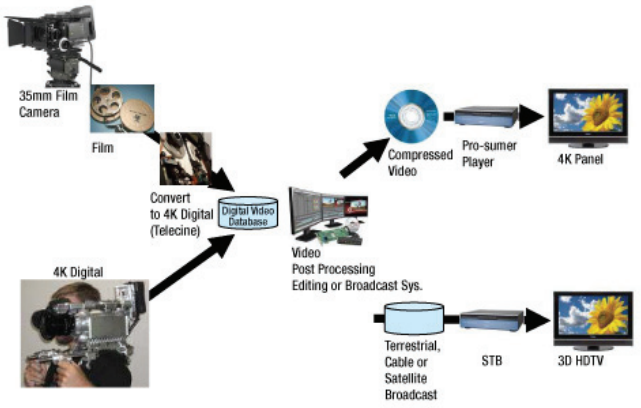

Figure 2: Typical video compression and decompression workflow

Bandwidth

The storage size of video files is important but managing transmission bandwidth is far more critical when transporting content through a network. For example, the ability to transmit HD video at 10Mbps instead of 20Mbps means that the same network can now transmit twice as many movies at once, which can mean significant cost and infrastructure savings. It is therefore critical to find the "sweet spot" that best blends a low bit rate with high-quality video and low storage/transmission cost because consumers are demanding increased performance (resolution, frame rate, etc.) and lower cost. The following four factors determine the balance between high performance and image quality:

- **Resolution:** Image resolutions are trending inexorably upward.
- **Frame Rate:** Frame rates are increasing to provide additional realism and/or support 3D imaging.
- **Transmission Rate:** Network, cable, satellite, and home cable (HDMI) speeds must keep increasing to handle higher resolutions and frame rates.
- **Compression Efficiency:** Increasingly complex algorithms are necessary to minimize file size and transmission bandwidth growth.

The combination of higher resolution and higher frame rates will place high demands on both bandwidth and storage. The efficiency of H.264 helps overcome many of these issues and is capable of supporting higher resolutions such as 4K images, frame rates in excess of the current 24 or 30 fps, and 3D imaging that typically requires up to double the frame rate compared to 2D video.

A Vision for DTV and Video

The current technological capabilities of displays, high-resolution cameras, storage systems, communications infrastructure, and compression techniques already significantly exceed the functionality of today's HDTVs, a discontinuity that makes significant imaging improvements over the next decade possible. This revolution is starting with the replacement of 35mm film with high-resolution "digital" film.

35mm film has dominated the movie industry for over 50 years but high-quality professional digital video cameras are increasingly being used for both TV and cinema productions. Entire movies are now being shot with digital movie cameras, eliminating the need for expensive, cumbersome, and toxic 35mm film processing and editing. Most current professional digital camcorders can capture 4K video (see Table 1, above) at 24, 30, or 60 fps. 4K DTVs are already in production and it is reasonable to expect that 4K video will eventually be available in homes on standard consumer electronics (CE) products. 3D digital cameras that utilize two lenses and imaging pipelines are also becoming available to the prosumer and consumer markets and it will not take long for this capability to reach smartphones.

Supported DTV frame rates are also increasing; many LCD displays can support 60, 120, and 240 fps even though no original video content comes anywhere close to these rates. Frame rate converters convert the 24 or 25 fps video content to higher frame rates and use special algorithms to smooth out high-motion artifacts. Support for high frame rates is one thing; support for high-resolution images is another. Digital still cameras and high-end smartphones routinely support 8 megapixels or more, which is at least 4 times the resolution of current HDTVs. A leading-edge 4K DTV could display an 8-megapixel image. Consumers are beginning to be able to shoot their own 3D and 4K home videos and watch them on 4K displays. HDMI Specification Version 1.4 supports both 3D movies and 4K resolution for CE products, continuing its usefulness as the industry standard interface for HD video and audio content.

Ongoing demand for higher resolutions and frame rates will continue to feed the burgeoning imaging revolution that will require higher video bandwidth combined with efficient compression technology such as H.264 to get the most out of costly storage and transmission infrastructure.

Aspects for H.264 Video Decoder Implementations

This section discusses the different algorithmic, architectural, and design techniques required to efficiently implement an H.264 decoder. The Multiview Video Coding (MVC) extension to the H.264 standard (intended to support 3D TV) is also described in more detail.

H.264 - A Big Step Forward for Higher Resolution Video

H.264 is considered to be the most advanced and complex video compression/decompression standard available for the CE marketplace. Unlike MPEG-2, H.264 does not claim backward compatibility and uses the latest algorithms to improve compression performance. Both coding efficiency and processing complexity have been increased by roughly three times over the most efficient MPEG-2 implementations, which requires additional semiconductor resources.

H.264 is being adopted for an increasing range of applications, including:

- Blu-ray Discs
- HD TV broadcasting (EU)
- Terrestrial SD and HD TV broadcasting
- PCs
- Portable devices including HD camcorders, smartphones, and Portable Multimedia Players (PMPs)
- Mobile TV broadcasting
- Security systems
- Internet video
- Video conferencing

H.264 refines and optimizes transmission of video content using the increased compression performance offered by the following new and extended algorithms:

Smaller Block Size: While MPEG-2 has a fixed block size of 8 pixels on a side (referred to as 8 x 8), H.264 permits the simultaneous mixing of different block sizes down to a standard size of 4 x 4 pixels. Integer Transform is used instead of the less-powerful DCT. This reduces both the hardware implementation effort and an imaging artifact called "ringing". Unlike MPEG, no truncation errors occur. The macro block size is flexible to allow more efficient coding of small moving frame areas.

Entropy Encoding: Entropy encoding reduces the size of the data by examining the frequency of patterns therein and encoding them in a smaller form. H.264 supports a variety of entropy encoding schemes as opposed to the single technique used by MPEG-2. For example, the new Context-based Adaptive Binary Arithmetic Coding (CABAC) scheme improves compression efficiency by 10% to 20% but is much more computationally complex than MPEG-2 entropy encoding. CABAC is used for both Main and High Profile classes of H.264 compression.

Internal De-blocking: Compressing high-motion video scenes into low-bit-rate video streams can cause block-based compression techniques such as MPEG-2 to break down because the actual borders of the blocks and macro blocks can be visibly seen in the video image, significantly damaging picture quality. With MPEG-2, de-blocking was an optional post-processing step performed after decoding that reduced or eliminated block artifacts but often made the picture appear washed out with greatly reduced detail. The H.264 standard adds an integral de-blocking filter, which significantly improves the subjective

appearance of picture quality (especially at block borders) while retaining image details.

Intra-frame Prediction: H.264 establishes an additional prediction type that uses information from the pixel neighborhood in the current frame for compression. Intra-frame prediction assumes some image areas contain pixels within the same frame that can be precisely calculated based on other nearby pixels. For example, intra-frame prediction can more efficiently compress a video frame with regular structures than a standard integer transform. Intra-frame prediction was not defined in MPEG-1 and MPEG-2.

Extended number of reference pictures: MPEG-2 defines two reference pictures (I or P frames). H.264 extends this number, which is only limited to the maximum size of a reference frame buffer. This increases flexibility; for example, the content of reference pictures can be used very efficiently for special scenarios such as fading scenes.

How H.264 Decoding Works

As with previous video decoding algorithms, H.264 consists of the following three steps:

- **Video Stream Evaluation:** This step extracts and prepares data such as block information and motion vectors from the video stream for use later in the video decoding and frame reconstruction processes.
- **Video Decoding:** During this step, macro block and block data processing generates the information that will be used when reconstructing the video frame.
- **Frame Reconstruction:** Once a macro block is processed, this step adds final information from associated reference frames by using corresponding motion vectors and also removes block artifacts to reconstruct the new video frame.

The next several sections provide a more detailed description of H.264 video decoder operation.

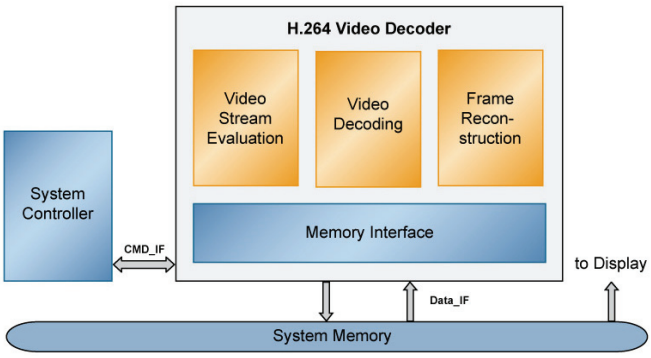

Figure 3: General block diagram of an H.264 video decoder

Video Stream Evaluation

The Video Stream Evaluation step extracts information from the compressed video stream such as the frame size, frame rate, or 3D/MVC capability. This information is stored as parameters that are used during subsequent decoding steps. Other information such as the size and direction of motion vectors are encoded and must be decoded; and some processing of previously stored information is required to reconstruct these values. Similar processing will be performed on pixel data available within the video stream.

The stream evaluation and parameter recovery functions may be encapsulated or linked with the subsequent Video Decoding and Frame Reconstruction tasks. Decoupling this activity from the video and frame reconstruction tasks allows a more flexible and efficient video decoder implementation. Processing these tasks in parallel also increases performance and allows access to different video streams in order to facilitate simultaneous multi-stream decoding. Some video decoder hardware implementations can easily decode numerous video streams with little dependency on software.

Decoding entropy information can present a real challenge for a high-resolution H.264 video decoder. While Context-based Adaptive Variable Length Coding (CAVLC) information decoding can be implemented using typical design techniques, implementing an efficient Context-based Adaptive Binary Arithmetic Coding (CABAC) decoder requires both much engineering expertise and significant processing power that usually requires the CABAC function to be implemented in hardware. A properly implemented CABAC algorithm delivers brilliant decompression results while a low-performing CABAC unit can be a very limiting factor for overall H.264 decoder performance.

In addition to CAVLC and CABAC decoding, the Video Stream Evaluation task also extracts macro block information such as motion vectors, parameters for intra-frame prediction, and loop filtering data (both new in H.264) for use during the Frame Reconstruction task.

The capability to evaluate multiple video streams in parallel allows the decoder to efficiently process the parallel streams with relatively low overhead and significantly increases decoder performance. The evaluation of two parallel video streams combined with the capability to store relevant stream data allows multi-stream decoding of an unlimited number of streams.

For multi-standard video decoders, the Video Stream Evaluation task also has to process the syntax for other standards such as MPEG-2 or VC-1. The general structure of these video formats is similar to H.264 and easier to handle; for example MPEG-2 does not require CABAC functionality.

Video error detection also occurs during the Video Stream Evaluation task. Single bits of video data may be damaged during data storage and/or transmission, which can cause serious image artifacts. A sophisticated channel error correction procedure can correct these errors; however, a worst-case scenario with burst errors could result in damaging or losing full packets of the video stream. Well-balanced error detection and concealment techniques are required to handle these different error scenarios. Error detection and concealment can be performed by a hardware decoder with or without system processor software support depending on the type of error. Users will not normally notice image quality issues when errors are isolated to a single block or macro block and are properly detected and corrected. Poor macro block error concealment can result in seemingly random "green blocks" observed in the video, which is a clear sign of poor video decoder design. A worst-case scenario where an entire GOP is destroyed by a burst error often causes the most current good image frame to be repeated for a full GOP time period or possibly longer, which may appear to the viewer as a "freeze" frame or jerky motion in a movie. This may be the only way to correct such catastrophic errors.

Video Decoding

The Video Decoding step of the H.264 decoding process manages computations at the macro block level and below using data prepared during the Video Stream Evaluation step. Macro block processing includes AC/DC prediction, inverse quantization, inverse integer transform and intra-frame prediction. Intra-frame prediction is a self-contained function including a huge number of modes optimized for different page content that does not require information from other frames, making this a very flexible function that allows further optimization of H.264 compression levels. The output from the Video Decoding step prepares pixel data for the subsequent Frame Reconstruction step.

The H.264 video decoder can utilize an architectural approach where all calculation tasks are autonomously executed step-by-step in a hardwired pipeline for optimal pixel-processing performance. This reduces clock

speed requirements because multiple macro blocks are processed simultaneously. Performance can be further increased by using more than one parallel processing pipeline. As in the Video Stream Evaluation step, an implementation that shares processing functions allows the decoder to support multiple video compression standards such as MPEG-2, VC-1 and H.264 with only small extensions.

Frame Reconstruction

The primary function of the Frame Reconstruction step is to provide motion compensation and de-blocking filter processing. Motion compensation enables reading of data for the macro block pixels from the proper location within the reference frame defined by the associated motion vectors. These motion vectors may have an accuracy of a quarter pixel for the H.264 video compression standard; intensive interpolation and filtering is therefore required at this stage. The de-blocking filter removes the block edge artifacts that are typically introduced when there is not enough image data available for the video decoder to reconstruct a clear image. Precise execution of the de-blocking algorithm is necessary for a high quality frame reconstruction. This step needs information supplied by both the foregoing Video Stream Evaluation and Video Decoding steps of the H.264 video decoding process.

This step requires very intensive read/write system memory access to reference frame data based on motion vector information. Memory access is also required to write the reconstructed frame back into memory after processing and to display this image on the DTV. It is extremely important that the memory controller optimizes memory access to reduce access latencies, especially in high-resolution and/or multi-channel video decoders.

Decoding 3D video streams such as H.264 MVC approximately doubles the required processing performance for the Video Decoding and Frame Reconstruction steps. Access to the system memory is required for two parallel frames, one each for the right and left views. A properly architected H.264 decoder can support both 2D and 3D video streams using the same implementation.

Memory Interface

Video decoding applications used in products such as Blu-ray Disc players, set-top-boxes (STBs), and DTVs often use unified memory systems that give many clients access to system memory. These clients have different bandwidth and maximum tolerable latency requirements. Memory bandwidth use can vary significantly depending on the memory access pattern. Optimized memory systems on modern DDR2 and DDR3 SDRAMs can achieve up to 80-90% bandwidth utilization with a properly architected memory controller. Poor memory controller architectures that do not take the specific memory requirements of an H.264 video decoder into account can reduce memory bandwidth utilization to as low as 15% to 20% of the theoretical maximum throughput. For example, if a video decoder implementation needed 1GB/s per second of actual read/write memory access to decode a video stream, an 80% efficient memory controller would need a memory system with a theoretical maximum read/write bandwidth of 1.25 GB/s; a 20% efficient memory system would need to provide 5 GB/s, making it more expensive and less practical to implement. It is therefore very important to optimize concurrent access to the memory system in decoder system implementations. Highly efficient mechanisms may include intelligent pipelined architectures that pre-fetch data and decouple process execution to improve memory access efficiency. Memory implementations such as these become increasingly important as video streams move to higher resolutions and/or use more complex encoding techniques such as multiple reference frames.

Multiview Video, 3D, and H.264 MVC

Multiview video has recently gained broad market interest as evidenced by the growing popularity of 3D cinema movies and 3D TVs. Multiview video combines multiple video streams from different viewing angles to represent a single scene with the visual perception of a 3D image. Multiview techniques are most commonly used for 3D imaging but could potentially be applied also to security applications. Storage and transmission bandwidth requirements could be doubled unless the potential image redundancy in multiview applications can be taken advantage of.

Multiview Coding (MVC) is an extension of the H.264 standard that includes new techniques for improving coding efficiency and adding new multiview-specific functions. MVC takes advantage of some of the interfaces and transport mechanisms introduced for the H.264 Scalable Video Coding (SVC) extension; however, system-level MVC integration is conceptually more challenging because the decoder output may contain more than one view that can consist of any combination of the views with any temporal level. Generating all of the output views for MVC video streams also requires careful consideration and control of the available decoder resources. The following MVC features are useful for enabling applications such as 3D video.

Compression of 3D Video

Multiview video sequences are captured by different cameras in different positions that are recording different angles of the same scene, which creates view redundancy in 3D video streams. The MVC standard utilizes prediction between the different views to exploit this image redundancy and improve the compression ratio for 3D video streams.

Scalability and Adaptation: The decoder implementation must be scalable and adaptable in order to meet the requirements of MVC video streams. For example, a 3D TV displaying multiple simultaneous views may be decoding more views than stereoscopic two-view displays. The MVC standard defines efficient ways to easily separate any subset of the views from the entire bit stream while maintaining backward compatibility for standard two-dimensional TV applications The MVC extension of the H.264 standard achieves backward compatibility by defining the bit stream so that an H.264/AVC-compliant decoder can decode a single 2D view and discard the rest of the 3D data while an MVC-compliant decoder can decode all the views and generate the 3D video. Backward compatibility is also supported by the related communication protocols for broadcast and storage applications. Any device capable of receiving a standard H.264/AVC stream over the MPEG-2 Transport Stream or the Real-Time Transport Protocol over IP is also capable of receiving an MVC stream over these protocols.

Computational Complexity: The amount of information processed for 3D video is significantly higher than for regular 2D video. Also, the additional dependency between views could make the implementation overly complex. The MVC standard includes efficient methods to buffering pictures used for prediction and enabling the parallel processing of separate views in order to reduce the complexity of 3D decoders. Also, the MVC standard does not change the underlying coding tools used in the H.264/AVC standard, which allows the reuse of widely deployed hardware accelerators and optimized software implementations.

MVC Standard: The MVC standard offers high-quality video compression and was added as Annex H to the H.264/AVC codecs. It allows efficient synchronous encoding of information from multiple cameras using a single video data stream. This standard can compress the 3D video for 3D TV, and Free Viewpoint

Television (FTV) systems that allow viewers to control which view of the scene will appear on the screen. MVC is backward compatible with H.264/AVC codecs, which allows it to be widely used in different 2D and 3D display devices.

Video Decoding Architectures: Hardware vs. Software

Decoding a H.264 video stream as described above can be done using software and/or hardware. Early H.264 video decoders relied primarily on software-only solutions, using high-performance general purpose processors to implement the complex H.264 algorithms. These implementations were very flexible in order to accommodate anticipated future changes to the standard; they were also extremely costly and consumed significant amounts of power.

Next-generation H.264 video decoders implemented certain low-level functions in hardware (such as hardware accelerators for processing intensive tasks including motion compensation or the filtering used in the de-blocking process) but still used software for significant amounts of processing. This hybrid architecture lowered implementation costs but many of these implementations remained both expensive and not readily extensible to support higher resolutions or higher frame rates without additional CPUs running at very high clock speeds. Nevertheless, many such hybrid implementations are currently used for SD and HD products using H.264 or other video compression standards. Decoding high resolution 3D or 4K video streams will become much more difficult and costly to implement using hybrid H.264 video decoders, meaning that future decoder implementations will continue moving toward a more "hardwired" architecture with less dependency on software for processing.

The H.264 video decoding standard is both market proven and very stable. Thus, there is little need for the flexibility provided by software beyond a few special features such as error concealment and audio/video synchronization. A forward-looking and scalable hardwired architecture can provide high-performance support for the highest resolutions and frame rates at very low-cost and with very low power consumption. These implementations consist of a chain of processing units dedicated to each special task that require very little interaction with the CPU and that can be implemented in more compact and cost-efficient silicon die areas than programmable processor architectures. Performing all tasks on macro block level and lower (and especially handling CABAC data with its significant processing requirements) in hardware reduces the overall size of the decoder design while significantly lowering CPU performance requirements. This frees up CPU processing power for use by typical software applications.

A hardwired solution as just described can support future performance extensions such as higher frame rate, larger frame size, parallel decoding of multiple streams for 3D video, and general multi-channel video. Properly architected hardwired solutions that take advantage of the ongoing advances in semiconductor capabilities will enable additional video decoding performance while only requiring relatively simple upgrades.

Implementing H.264 3D, 4K, and Ultra-HD Video Decoders

The latest Silicon Image scalable cineramIC 3D multi-standard video decoder IP is an excellent example of a scalable, high-performance, hardwired H.264 video decoder. This IP core supports resolutions up to 8K x 4K; alternatively, the available video decoding bandwidth can be used to decode up to 16 simultaneous HD streams. The cineramIC H.264 video decoder IP core is built on the market-proven Silicon Image hardwired architecture that is now available for integration into System-on-Chip (SoC) semiconductors. SD, HD, and 4K decoder implementations can also be integrated into FPGAs.

This latest implementation can decode video streams with resolutions up to 8192 x 4320 pixels with frame rates of 30 fps (8K x 4K @ 30 fps), enabling a cutting-edge digital cinema experience. The decoder is also capable of decoding two 4K x 2K @ 60 fps ultra-HD video streams, eight 1080p @ 60 fps in parallel, or up to 16 HD or SD streams with only little software intervention thanks to the hardwired, scalable, high-performance cineramIC video pipeline architecture that produces exceptional performance at low cost and minimal power consumption.

The new cineramIC IP core meets or exceeds the requirements of H.264 Main Profile and High Profile up to Level 5.1 including H.264 MVC, MPEG-2 Main Profile at High Level, and SMPTE 421M VC-1 (Simple, Main and Advanced Profile at Level 4). The cineramIC decoder also supports MVC Multiview Video decoding for 3D HD and ultra-HD resolutions up to 4K x 2K. One 3D 4K x 2K @ 60 fps 3D ultra-HD video stream can be decoded in real-time using only a 300MHz clock implemented in a 40nm ASIC process. The decoder also supports decoding multiple 2D or 3D streams in parallel. For example, four 3D 1080p @ 60 fps streams can be decoded simultaneously using the cineramIC IP core's overall decode processing capabilities. High-resolution still images in Exif JPEG format can also be decoded up to 16K x 8K image size (128 megapixel). For example, a still image with 8K x 4K image size (32 megapixel) can be decoded at a rate of 9 images per second and an 8 megapixel image at a rate of 38 images per second.

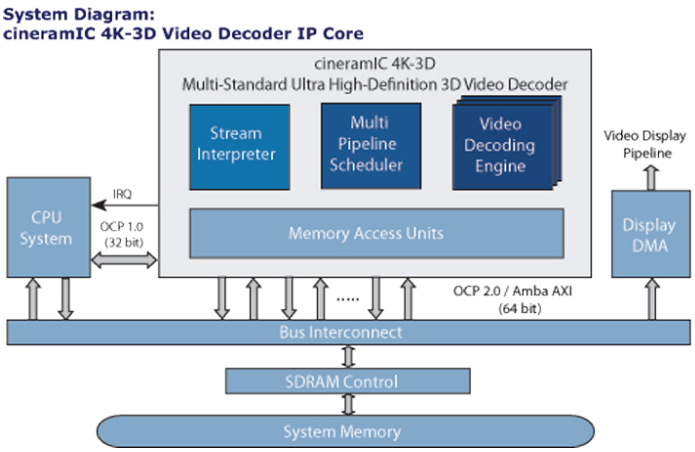

Figure 4: The Silicon Image cineramIC 4K-3D ultra high-resolution multi-standard video decoder

Figure 4 contains a block diagram of the new Silicon Image cineramIC 4K-3D core for frames sizes up to 8K x 4K. The highest performance capabilities of the latest cineramIC core can be achieved by using more than one Video Decoding Engine block in the video decoder.

The scalability of the cineramIC 4K-3D IP Core allows SoC designers to select the design that best meets their specific performance requirements. One to four video decoding pipelines using its own Video Decoding Engine block each work in parallel and are organized and internally controlled by the Multi-Pipeline Scheduler block. This allows the design to be optimized to meet almost any system requirements. Four pipelines can be instantiated in the design for applications such

Image Resolution/ MVC	Frame Rate [frames/sec]	Clock Speed[MHz]	Number of Pipelines	CPU Performance [MIPS]	Notes
8Kx4Kp	30	300	4	2	Maximum performance with 4 pipes
4Kx2Kp / 3D	60	300	4	8	Maximum performance with 4 pipes
1080p / 3D	60	75	4	8	Maximum performance with 4 pipes
1080p / 3D	60	300	1	8	Minimum area/power with 1 pipe
1080p	60	38	4	4	Maximum performance with 4 pipes
1080p	60	75	2	4	Average performance and area/power with 2 pipes
1080p	60	150	1	4	Minimum area/power with 1 pipe
720p	30	8	4	2	Maximum performance with 4 pipes
720p	30	33	1	2	Minimum area/power with 1 pipe

Note: p=progressive frames

Table 2: Performance overview of the scalable cineramIC 4K-3D IP core

as high resolution 3D, 4K or 8K video decoding that require very high performance. Use of four pipelines increases performance at the cost of some additional SoC area. SoC designers developing mobile applications that require only HD resolution can use the same IP core with only one Video Decoding Engine, which reduces SoC area size and power consumption. The driver software does not need to be modified; SoC designers using this IP core for several different applications do not have to adapt the software for each specific application.

The new, scalable Silicon Image H.264 video decoder IP core is available for integration into ASIC SoC designs for the following applications:

- Multi-channel display applications such as surveillance or multi-picture in picture DTVs or IP TVs

- Semi-professional and professional 4K and 8K broadcast and studio applications
- Very high resolution projectors and displays
- 4K professional and "prosumer" video camcorders
- 3D and Multi-view DTV
- Ultra-HD Home Cinema
- High-end Blu-ray Disc Players, STBs, and recorders with 3D Support.

A special real-time IP Core product variant is available for FPGA implementations with reduced performance that supports resolutions up to 3D/4K. This variant makes the technology very appealing for use in lower unit volume applications such as professional video editing, broadcast, and medical, military or surveillance applications.

The scalable cineramIC hardwired architecture makes this technology one of the smallest synthesizable cores on the market running at very low clock rates. The core is substantially smaller than software-only video decoders. It is also smaller and uses less power than most hybrid hardware/software solutions. Few of these hybrid hardware/software decoders approach the high resolution/frame rate decoding capabilities offered by the cineramIC video decoder. This design from Silicon Image can be used for multiple product variants using a single architecture and implementation, making the cineramIC IP core an extremely attractive business proposition.

Table 2 summarizes cineramIC 4K-3D performance for different frame sizes, 2D/3D video modes, and number of video decoding pipelines.

cineramIC 4K-3D IP core decoding capabilities are implemented as an autonomous pipeline of various hardware blocks executing the tasks as described above for H.264 decoding. The block functionality is shared by all three H.264, MPEG-2, and VC-1 video standards. The driver software for the IP core runs on an external CPU. A general purpose 32-bit processor executes the setup and general control tasks. All IP core variants require less than 2 Million Instructions Per Second (MIPS) of CPU time to support 8K x 4K video decoding and less than 8 MIPS for 3D 4Kx2K @ 60fps, making the new scalable cineramIC 4K-3D one of the world's highest performing - and lowest cost - digital video decoders.

The cineramIC IP core reads the compressed video stream from a buffer located in the system memory (SDRAM) and generates decoded video in YCbCr 4:2:0 (JPEG also in YCbCr 4:2:2) format. Reconstructed frames are stored in the system memory to be used for later processing or presentation. Dedicated memory access units are implemented for different tasks such as stream reading and reading and writing of reference frames in order to optimize performance and latency requirements for video decoding. The clock in the memory access units of the cineramIC decoder is decoupled from the clock of those other blocks executing stream evaluation, video decoding, and frame reconstruction tasks. Clock domain crossing is included in the video decoder IP. The cineramIC memory access units provide 64-bit OCP 2.0 and AMBA AXI compliant ports to the customer memory system. An IP core user must provide their own system memory controller with the memory bandwidth necessary to support the 64-bit cineramIC memory interface.

The cineramIC core also supports automatic hardware-based video stream context switching that can manage the decoding of up to 16 video streams with no additional software intervention required. Each individual stream can have differing resolutions and use different compression standards. For example, a single decoder can decode multiple SD and HD video streams where some streams are

H.264 and others are VC-1 or MPEG-2. The cineramIC core user can provide additional software to allow processing an unlimited number of streams and still pictures simultaneously as long as the overall decoder bandwidth capacities are not exceeded. The hardware is also capable of detecting and correcting most video stream errors. Only exceptional error scenarios will require software-based error concealment.

The design's gate count depends on the number of Video Decoder Engines (VDEs) and is in the range of 800k gates (for one VDE for minimum area and power) to 2,150k Gates (four VDEs for maximum performance). This includes the complete memory interface, stream reader functionality, and extra logic for context switch support (1 gate = 2 input NAND equivalent). For a one-pipe design, 40 internal SRAM instances are used to provide a total of ~225 Kbit RAM size for storing of intermediate data for process execution. A four-pipe design requires 118 SRAM instances with overall total of ~690 Kbit RAM size.

The cineramIC 4K-3D video decoder is written in Verilog RTL and the package includes synthesis scripts, an extensive verification environment, reference driver software, and detailed user manuals for both hardware and software. Both Silicon Image and our IP core customers have proven the architecture and various implementations both on silicon and in the market. The ASIC variant of the cineramIC 4K-3D video decoder has been proven to run on TSMC 65GP and TSMC 40LP technology with a 350 MHz clock frequency and can also be synthesized for other technologies and processes. Silicon Image engineers have extensively tested and verified the video decoder during development using industry-accepted simulation techniques and FPGA implementations. Silicon Image also uses both the Allegro and Fraunhofer H.264 certified video streams for testing standards conformance as well as the video decoder's performance and error handling characteristics.

All Silicon Image cineramIC 4K-3D video decoder customers also receive driver software that implements an easy-to-use API for multi-stream, multi-standard, and multi-view control. The driver software is delivered as C source code that is portable to any CPU. Design scalability is transparent to the driver software, meaning that a customer can use different numbers of VDEs for different designs without having to modify the software.

Summary

This white paper provided background information about H.264 requirements, operation, trends, tasks, issues, and tradeoffs associated with planning and implementing an H.264 video decoder. Basic performance variations are driven by requirements such as video resolution, frame rate, and the number of simultaneously decoded video streams. An example of an actual scalable, high-performance H.264 video decoder implementation was presented that supports applications including multi-channel video decoding, 3D, and ultra-high definition 8K x 4K decoding. This sample IP Core implementation provide an SoC designer with very high performance and improved time to market while keeping silicon die cost and power requirements to a minimum.

Learn More

For more information about CE standards and related implementations, visit the following web sites:

Standards:

- International Telecommunication Union (ITU, for H.264): http://www.itu.int
- Moving Picture Experts Group (MPEG): http://mpeg.chiariglione.org
- Society of Motion Picture and Television Engineers (SMPTE, for VC-1): http://www.smpte.org
- Joint Photographic Experts Group (JPEG): www.jpeg.org

For more information about Silicon Image semiconductor products and IP offerings, visit:

Silicon Image: www.siliconimage.com

Silicon Image IP Cores: www.siliconimage.com/iplicensing/index.aspx

For any comments or questions, please email: IP_licensing@siliconimage.com

About Silicon Image, Inc.

Silicon Image is a leading provider of advanced, interoperable connectivity solutions that enable the reliable distribution and presentation of high-definition (HD) content for consumer electronics, mobile, and PC markets. The company delivers its technology via semiconductor and intellectual property (IP) products that are compliant with global industry standards and also feature industry leading Silicon Image innovations such as InstaPort™. Silicon Image's products are deployed by the world's leading electronics manufacturers in devices such as desktop and notebook PCs, DTVs, Blu-Ray Disc™ players, audio-video receivers, as well as mobile phones, tablets and digital cameras. Silicon Image has driven the creation of the highly successful HDMI® and DVI™ industry standards, as well as the latest standards for mobile devices - SPMT™ (Serial Port Memory Technology) and MHL™ (Mobile High-Definition Link). Via its wholly-owned subsidiary, Simplay Labs, Silicon Image offers manufacturers comprehensive standards interoperability and compliance testing services.

1060 E. Arques Avenue, Sunnyvale, CA 94085 T 408.616.4000 F 408.830.9530 www.siliconimage.com