Development of high-performance HDV™ devices based on the HDV 1080i specification for acquisition and production applications

WHITE PAPER

SONY®

HDV™ Devices

INTRODUCTION

Juan Martinez, Senior Manager Technology – Alliances, Sony Electronics Inc.

Abstract
We are rapidly transitioning into a high definition world. Driven by this trend, the demand for HD content creation is increasing. Although there had been a plethora of high-end HD creation tools, until the advent of the HDV™ format, no cost-effective production tools existed.

The HDV format was established on September 30, 2003 by four companies: Canon Inc., Sharp Corporation, Sony Corporation, and Victor Company of Japan, Limited. The concept of the HDV format was to develop a high definition standard capable of inexpensively recording high quality HD video using conventional DV recording media. By using the mechanisms of a DV camcorder, mitigation of development costs and development efficiency would be realized.

Efficient bit-rate reduction while retaining the high quality of HD images is made possible by means of the MPEG-2 compression scheme. In order to use MPEG-2 encoding to compress the large quantity of HD image data, complex signal processing and small silicon area for portable recorders are required. Advancements in semiconductors and signal processing technology have now made possible the use of the HDV format as a standard for low cost content creation.

To meet this end, Sony introduced a compact-sized digital HD video camcorder for professional use, the HVR-Z1 series, which was put on the market in the beginning of 2005. Since the launch of the HDR-FX1 (consumer camcorder), the HVR-Z1 series and the HVR-M10 (compact deck) series, the HDV format has become the most popular HD recording format with about 37,000 HDV 1080i units shipped during the first 6 months of availability.

This paper will explain the HDV format and the technologies employed in the development of these professional devices.

Two spatial resolutions

The HDV format specifies two recording systems: the 720p (progressive) method and the 1080i (interlaced) method. The specifications for both systems are shown below.

Name of Specifications	HDV 720p Specification	HDV 1080i Specification
Media	Same as DV format (DV and/or Mini DV cassette tape)	
Video		
Video Signal (Video Signal Optional)	720/60p, 720/30p, 720/50p, 720/25p (720/24p)	1080/60i, 1080/50i (1080/30p, 1080/25p, 1080/24p)
Number of Pixels (Horizontal X Vertical)	1280 x 720	1440 x 1080
Aspect Ratio	16:9	
Compression	MPEG-2 Video Profile & Level : MP@H-14 (In case of 720/60p and some 720/50p:MP@HL)	
Sampling Frequency for Luminance	74.25MHz	55.7MHz
Sampling Format	4:2:0	
Quantization	8 bits (both luminance and chrominance)	
Bit rate after Compression	Approximately 19Mbps	Approximately 25Mbps
Audio		
Compression	MPEG-1 Audio Layer II	
Sampling Frequency	48kHz	
Quantization	16 bits	
Bit rate after Compression	384kbps	
Audio Mode	Stereo (2 channels)	
Optional Audio Recording	PCM (2 channels or 4 channels, can be simultaneously recorded with MPEG-1 Audio Layer II 2 channels)	MPEG-2 Audio Layer II (4 channels)
System		
Data Format	MPEG-2 Systems	
Stream Type	Transport Stream	Packetized Elementary System
Stream Interface	IEEE1394 (MPEG-2-TS)	

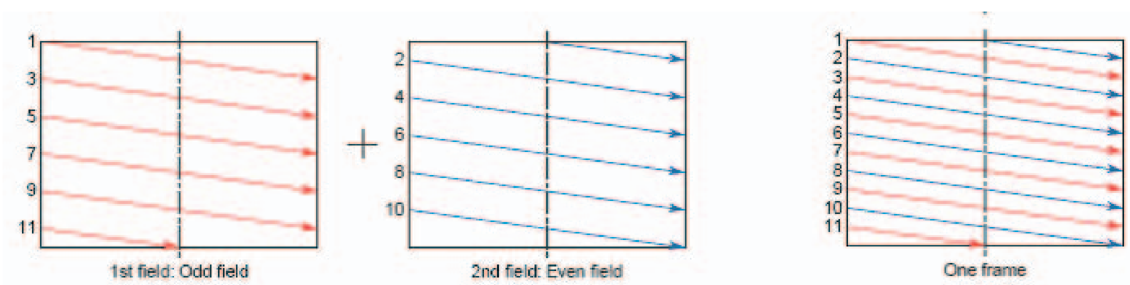
Picture format notation
When describing the picture format, it may be written in a form such as “1080/60i.”
Here is meaning of this notation

1080/60i

- “i” indicates interlace scanning
  “p” indicates progressive scanning
- Frame / field frequency
- Number of effective scanning lines

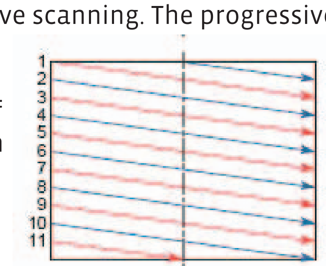
Frame / field frequency basically indicates how many images are produced in one second. If this frequency is 60, then that means 60 images are produced in one second.

Progressive and interlace scanning
Interlace is the scanning method adopted for television broadcast systems such as NTSC and PAL. In brief, interlace scanning refers to the scanning of every other line of an image as the first field, and then scanning the lines in between as the second field. The benefit of interlace scanning is that it displays motion pictures on a video monitor with negligible flicker. After thoroughly investigating the human eye characteristics, tests have demonstrated that at least 60 images must be displayed within one second to avoid flicker. The result was to use a 1/60-second rate for each scan, but in the first scan, from the top to bottom of the image, only the odd-numbered lines (1, 3, 5...1079) are scanned, and for the next scan, only the even-numbered lines (2, 4, 6...1080) are scanned.



Interlace Scanning method

Progressive scanning
The DTV broadcast transmission infrastructures give the broadcaster the option of using an interlaced system or a non-interlaced system - the latter known as progressive scanning. The progressive scanning system was adopted initially in computer displays, which do not require considerable transmission bandwidth. In progressive scanning, each line is scanned in sequential order from the top to bottom of the picture. The entire 720 lines (or 1080 lines for 1080p) are displayed in one scanning operation.



Recording HD images on DV tapes

The HDV tape transport mechanisms are based on the DV helical scan format. Therefore, videotapes used for DV recording can also be used for HD recording, and the recording time is equivalent. Because the DV tape specification does not spell out the exact chemical formulations or manufacturing methods, but only the physical characteristics, there are many possible formulations that arrive at the target specifications. Some tape formulations and manufacturing methods yield better quality, robustness and stability than others, so the actual "in the field use" experience will vary significantly from one tape brand or model to another. Although most people think of videotapes in terms of the magnetic materials or lubricants used, few realize that the plastic substrate employed contributes significantly to the tracking stability. High quality "premium" tapes such as Sony's "Digital Master™" are preferred not only for HDV but also for DV recordings as its dimensional stability is maintained over a wide temperature range. The benefit is that tracking will be maintained even when recording in extremely high temperatures and then editing in a cool environment.

Helical scan

When videotape recording was being developed in the late 70s, engineers faced a serious problem. Slow tape speed was desirable for longer recording time, but the venerable stationary head tape recording scheme could not yield the high writing speeds required for high-quality video recording. A means of achieving high writing speeds and long record time needed to be invented. The breakthrough solution, "helical scanning," became the basis for all analog and digital videotape recoding to date.

In the helical scan process, the tape is pulled at slow speed, and the high writing speed is achieved by helically wrapping the tape around a rapidly rotating drum with four (in the case of DV and HDV) small embedded record heads. This recording scheme produces recorded tracks that run diagonally across the tape from one edge to the other. In other words, the recorded tracks are parallel to each other but are at an angle to the edge of the tape. Analog formats stored one video field or frame every drum revolution. Digital recording schemes continuously quantify and store the instantaneous signal level as a numerical value, producing considerable amounts of data. To handle the greater data-rates generated by digital recording formats, a segmented recording scheme is utilized where multiple tracks are used to record a single video frame (for example, ten tracks for DV and HDV 720p). For DV and HDV recording, the tracks are subdivided into sectors that carry specific types of payload or operation data, such as ITI (Insert Track Information), which carries tracking information.

DV tape recording

HDV1080i and HDV 720p devices are downward compatible with DV from which both HDV tape transports mechanisms were derived. The DV format records the digital signal following a segmented recording scheme with ten tracks per frame for NTSC, (480/60i) and twelve tracks per frame for PAL, (576/50i). The video, audio, and subcode payloads are recorded into individual sectors within each track.

HDV 720p

The HDV 720p specification simplified and mitigated product design and manufacturing cost by adopting the same track and sector structure as DV. The sector's ITI (tracking signal), subcode and overwrite areas are also used for HDV recording. HDV 720p devices store the entire HDV payload, which encompasses MPEG video (18.3 Mbps), audio (384Kbps), error correction and visual search signal into the sector used exclusively for video by the DV format. Ten TS tracks combined contain one error correction unit.

HDV 1080i

The HDV 1080i specification does not follow the DV tape footprint as HDV 720p does. In order to accommodate a higher bit-rate, the HDV 1080i specification adopted a unique track and sector structure that maximizes the length of the track for MPEG-2 payload recording. The video payload bit-rate is 25.4Mbps, which is 28% higher than what is possible by strictly adhering to the DV track and sector structure. The remaining area contains audio at 384Kbps, error correction data and two visual search signals.

Powerful error correction capability

With the MPEG-2 format, which uses inter-frame compression, the impact of missing data is much greater than for the DV format where intra-frame compression is used. For this reason, the amount of data for error correction coding used by the HDV format was greatly increased relative to the DV format. Moreover, by abandoning the DV error correction method that operates within a single track, for a correction method spanning multiple tracks (10 for HDV 720p and 16 for HDV 1080i), error correction capability was drastically enhanced. The benefit of the HDV 1080i sixteen-track error correction scheme has empirically demonstrated an excellent performance record.

HD CONTENT STORAGE AND DISTRIBUTION

HDV is one more acquisition tool available to production crews. Content originated by HDV 1080i products is routinely used as "contribution" for high end productions, inter-cut with other high definition formats and eventually aired on network broadcasts. The HDV 1080i specification compresses HD video at 25 Mbps (same as DV). The 25Mbps data stream consumes only 2GB for every ten minutes of recording.

- **Consumer HD media**

  Optical disk media suitable for long-form high definition program distribution requires high data capacity media. Blu-ray Disc™ optical media is specifically designed for this purpose. Blu-ray Disc recorders can be connected directly to compatible HDV devices via their i.Link® ports for an efficient and cost effective distribution of high quality prerecorded HD content. One single layer Blu-ray Disc media provides 115 minutes of native HDV 1080i recording. It is not necessary to resort to high compression schemes or to use expensive, time-consuming multi-pass encoders for long form program distribution.

- **Storage and distribution as "data"**

  The MPEG-2 tape transport signal used by HDV devices may be directly stored on data tapes, hard disk drives, CD–ROM or DVD-R (as data), solid state memory or even transmitted over networks. The recipient can view the native HDV files using a media player with a suitable software or hardware decoder. One DVD-R disk that can cost less than one dollar can store about 23 minutes of native HDV 1080i content.

- **Shooting HDV for SD distribution**

  Shooting HDV original footage and editing in HD produces standard definition DVDs which are visually superior to those using DV content. Since the MPEG-2 compression used by HDV utilizes the same 4:2:0 color space as the DVD standard, the color sampling remains consistent throughout the production, editing, and encoding workflow. In comparison, content acquired in DV is sampled at 4:1:1, then re-sampled at 4:2:0 during the DVD encoding process, yielding a visually inferior 4:1:0 image on the DVD. Furthermore, the HDV high definition content may be recompressed using MPEG-4 (H.264) or Windows® Media Video HD for distribution via "on demand," or distributed using commonly available optical disks.

- **Film-out**

  HDV 1080i has a very high vertical resolution of 1080 lines. It is possible to deliver 1.85:1 or 2.35:1 film formats with excellent visual quality. The high spatial and temporal resolution of HDV 1080i produces very detailed 35mm film out.

MPEG-2 is used as the compression format

Until the HDV format was launched, no cost-effective HD production tools existed. The HDV consortium's goal was to develop a high definition format that could inexpensively record high quality HD video using conventional DV tape as the recording medium. MPEG-2 was chosen due to its ability to provide high video quality at lower bit-rates when compared to using an intra frame compression codec such as DV. The MPEG-2 standard is not a single rigid specification. It is a flexible toolkit providing four levels and six profiles to choose from, for selecting the most suitable compression scheme for any application.

MPEG-2 LEVEL/ PROFILE TABLE

Level

	Profile					
Frame types	**Simple**	**Main**	**SNR-Scalable**	**Spatially-Scalable**	**High**	**Studio**
PICTURES	I&P	I,P & B	I,P & B	I,P & B	I,P & B	I,P & B
CHROMA SAMPLING	4:2:0	4:2:0	4:2:0	4:2:0	4:2:2~ 4:2:0	4:2:2~ 4:2:0
HIGH						
Samples per Line		1920			1920	1920
Lines per Frame		1152			1152	1080
Frames per Second		60			60	60
Max. Bit-Rate (Mbps)		80			100	300
HIGH 1440						
Samples per Line		1440		1440	1440	
Lines per Frame		1152		1152	1152	
Frames per Second		60		60	60	
Max. Bit-Rate (Mbps)		60		60	80	
MAIN						
Samples per Line	720	720	720	720	720	720
Lines per Frame	576	576	576	576	576	576
Frames per Second	30	30	30	30	30	30
Max. Bit-Rate (Mbps)	15	15	15	15	20	50
LOW						
Samples per Line		352				
Lines per Frame		288				
Frames per Second		30				
Max. Bit-Rate (Mbps)		4				

HDV

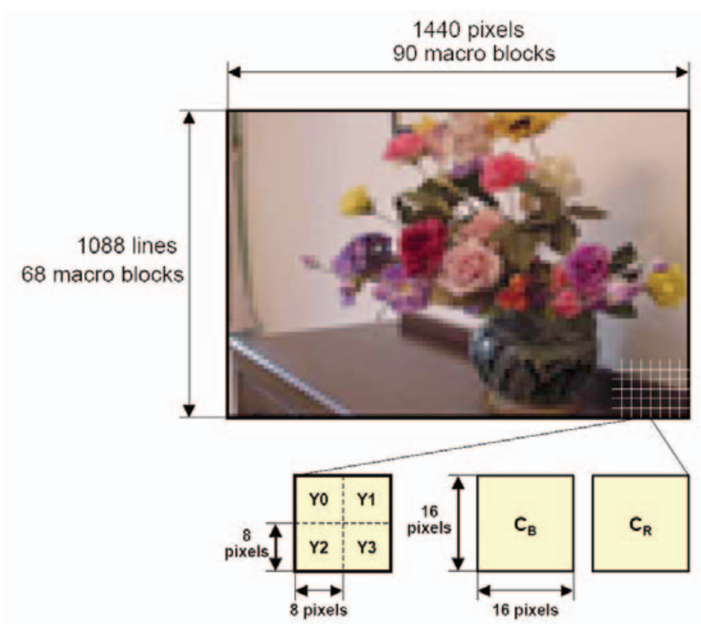
In the case of the HDV format, efficient bit-rate reduction while retaining the high quality of HD images is possible by means of the MPEG-2 MP@HL-1440 (main profile at high level) compression scheme highlighted above.

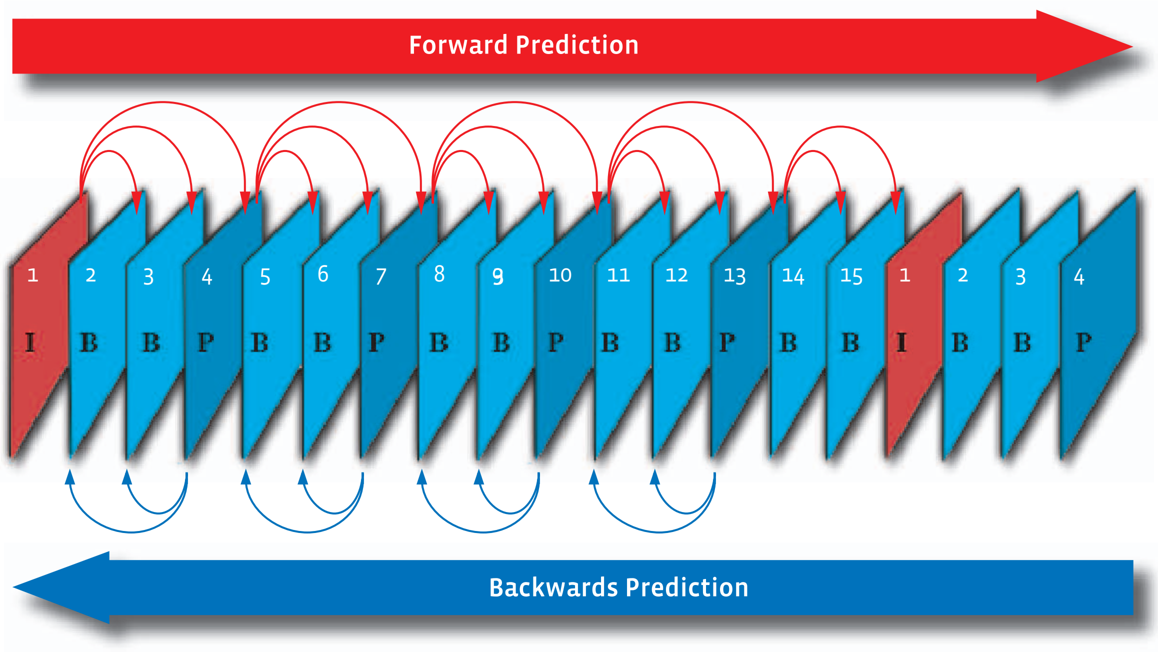
MPEG-2 SIMPLE EXPLANATION

The MPEG (Moving Picture Experts Group) created the MPEG-2 standard as a "compression toolkit" that could accommodate a wide range of picture sizes, from standard definition to high definition, at a higher picture quality for a given bit-rate. MPEG-2 was approved in 1994 as a standard intended for delivery of high quality digital video. It is the compression scheme used for DVD disks, direct digital broadcast satellites, terrestrial and cable high-definition TV (HDTV), digital standard definition broadcast (SDTV), and cable TV (CATV).

The Moving Pictures Expert Group showed its wisdom by not rigidly specifying the compression algorithms. Instead, they merely specified the syntax for storing and transmitting the compressed data, as well as the decoder. This approach freed the encoder manufacturers to continue to refine the encoding algorithm. The only constraint is that it must produce valid MPEG-2 streams that can be decompressed by any MPEG-2 compliant decoder.

MPEG-2 realizes very high compression efficiencies while maintaining high video quality by taking advantage of temporal redundancies within a sequence of images. The MPEG-2 codec works on two stages. In the first step, all the video frame images are divided into 8-pixel luminance blocks and 16-pixel color blocks. One macro block contains four luminance blocks and two chrominance blocks. The blocks are compressed using DCT-based intra-frame compression techniques similar to that used by DV. Then, using the first compressed image as a reference frame, (called an I-frame), the second stage eliminates redundant information, keeping only those parts of the following images (B- and P-frames) that differ from the reference image. During playback, the decoder will then reconstruct all images based on the reference image and the "difference data" contained in the B- and P-frames. This combination of I, B and P frames is known as a Group of Pictures (GOP).





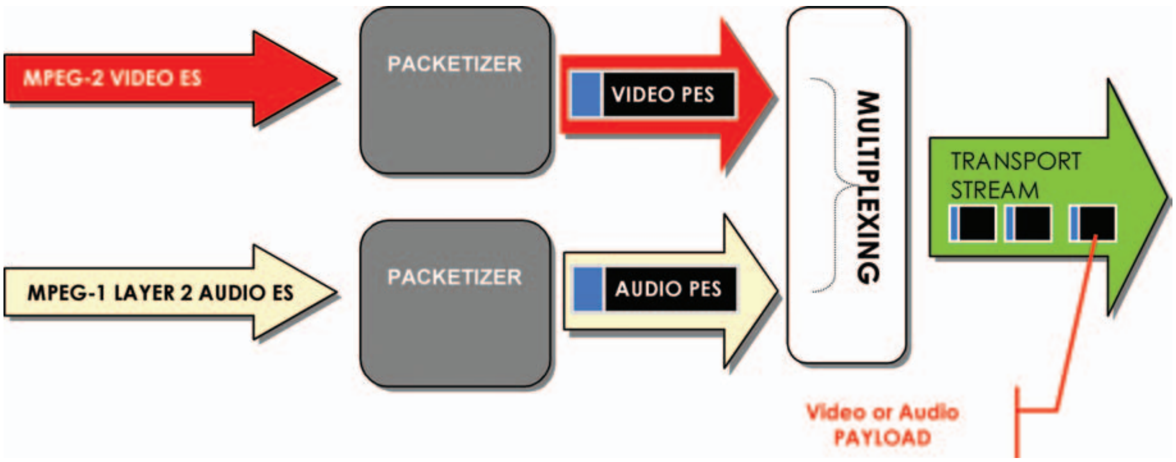
Elementary Streams (ES)
The output from the MPEG-2 video and audio encoders are elementary streams. Elementary streams are continuous and do not stop until the source ends. Each ES contains only one type of data (audio or video) from a single audio or video encoder.

Video ES
The raw output of an MPEG-2 encoder is called a video elementary stream or video ES. The data rate of the HDV 1080i video elementary stream is 25Mbps.

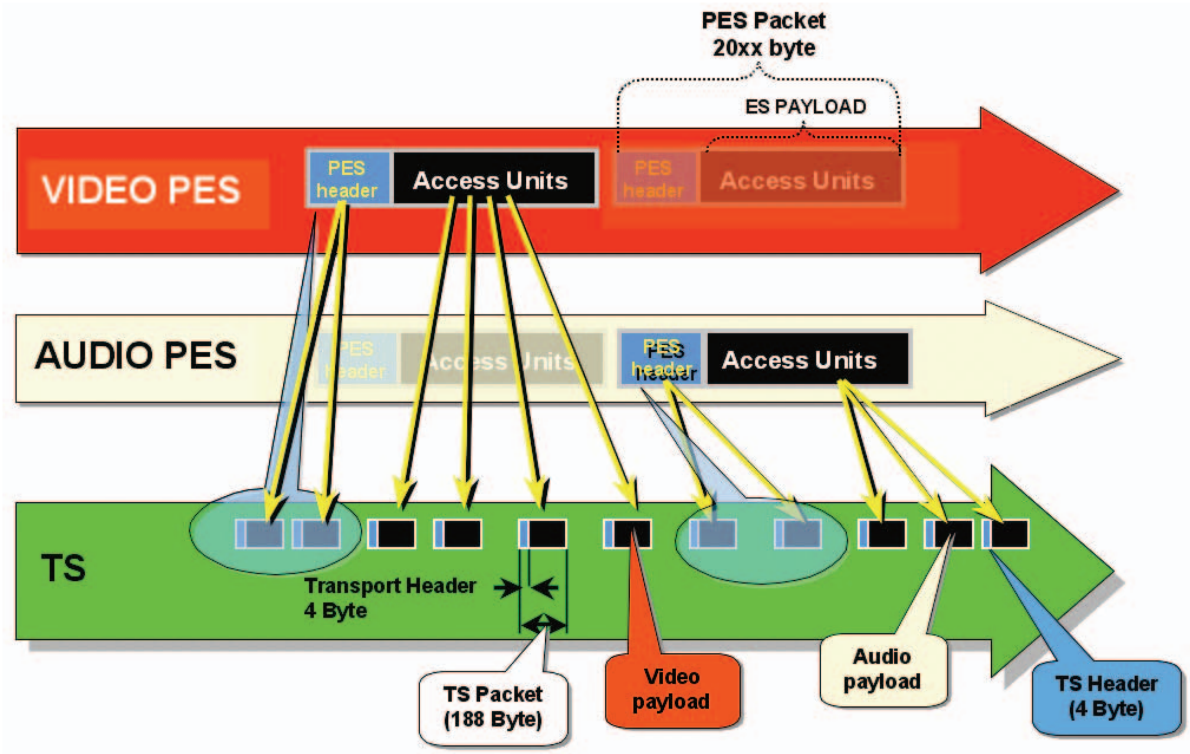
Audio ES
The HDV audio is also compressed using the MPEG-2 compatible MPEG-1 Audio Layer II audio codec. The data rate of the compressed audio elementary stream is 384Kbps, the highest data rate permissible, providing good compression efficiencies while maintaining high audio quality.

Packetized Elementary Stream (PES)
The continuous elementary bit-stream is then fed into a packetizer, which divides the ES streams into parts of a fixed size in bytes. These packets are known as Packetized Elementary Stream (PES) packets. A PES contains only one type of payload data, video or audio, from a single encoder. Each PES packet begins with a packet header that includes a unique packet ID. The header data also identifies the source of the payload (video or audio) as well as ordering and timing information needed by the decoder to recreate the audio and video and reproduce them in sync.

Transport Stream (TS)
The next stage multiplexes the video and audio PES into a single stream for storage or transmission. The MPEG standard defines two methods for multiplexing video and audio elementary stream data: Program and Transport. The Transport Stream method was selected for HDV recording because it simplifies detection of the start and end of frames as well as facilitating recovery from packet loss or corruption.



Thus, the video-PES and audio-PES streams are multiplexed to form a single Transport Stream. The transport stream packets have a fixed 188 Byte packet size. The PES packet size is variable e.g. 2048 Byte; much longer than a transport packet. Thus, each PES packet is subdivided into multiple TS packets as described below.



As shown in the graphic above, the PES header is placed at the beginning of a transport packet payload, following the transport packet header. The remaining PES packet content fills the payloads of consecutive transport packets until all the PES packet data is used up. The final transport packet may be padded with blank information (digital ones) to make it conform to the specified 188 Byte packet size. The transport stream containing the video and audio packetized elementary streams is transmitted via the i.Link interface to another compatible HDV device for dubbing or for storage and editing on a PC hard drive.

CHALLENGES

Searching and Editing Capabilities

The use of helical scan and MPEG-2 long-GOP encoding makes tasks requiring individual frame access, such as frame accurate cuing, editing, and fast-forward or fast-reverse visual search, extremely challenging. The MPEG-2 inter-frame coding algorithm compresses several frames into one GOP sequence, making it necessary to decode all the pictures in the GOP in order to access any individual picture.

Native MPEG-2 Editing

Editing long-GOP MPEG-2 files is not without challenges. Unlike uncompressed video, or intra-frame compressed video formats such as HDCAM® or DV, the MPEG-2 compression employed in HDV reduces bit-rate and file size by inter-frame compression, where all frames forming an HDV 1080i GOP are compressed together. The GOP contains one "I" reference frame followed by a sequence of B & P partial frames that contain only the fractions of the image that differ from the reference frame. It would be possible to edit at each "I" frame, but this solution does not provide acceptable accuracy. In order to achieve frame accurate "native" MPEG-2 editing, it was necessary to create new tools that allow cutting in the middle of GOPs, removing the unused frames and splicing the partial GOPs together, without visually degrading the image. It is also necessary to repair the bit-rate and GOP sequence for the all the frames contained within the edit transition. In other words, the MPEG-2 compression and PES and TS structure must be restored to the original HDV coding parameters for the duration of the entire edit or effect, so that a standard decoder can seamlessly play it back.

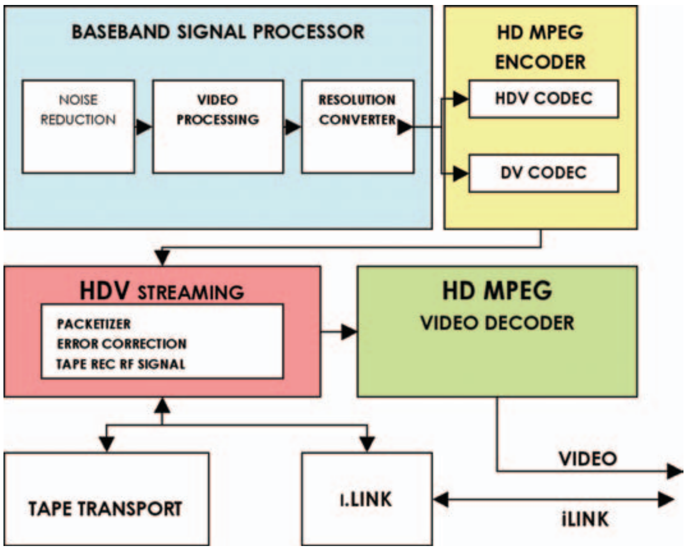
The native MPEG editing process described above is computationally intensive, and initially made "native" long-GOP editing impractical. Thanks to the advent of powerful new microprocessors, cost-effective memory and sophisticated software algorithms, efficient MPEG-2 long-GOP editing is now a reality.

HD CODEC

HD Codec Engine

In order to produce a compact camera with a signal processing engine and MPEG-2 encoder suitable for high quality HDV1080i recording, new small-sized silicon devices capable of complex signal processing were required. The most advanced semiconductor design and manufacturing technologies were used to create the "HD Codec Engine" for Sony's HDR-FX1 and HVR-Z1 series camcorders. The "HD Codec Engine" consists of 4 main LSI groups: the Baseband Signal Processor, the HD-MPEG Video Encoder, the HDV Streaming Processor, and the HD-MPEG Video Decoder. These high-performance LSIs were developed by utilizing manufacturing innovations like the "submicron process rule," an advanced semiconductor technology that enables extreme miniaturization and low power consumption. To put it in perspective, the total amount of logic gates required to implement the complex algorithms used by the processing, encoding, streaming and decoding blocks exceeds 5.4 million transistors.

HD Codec Engine



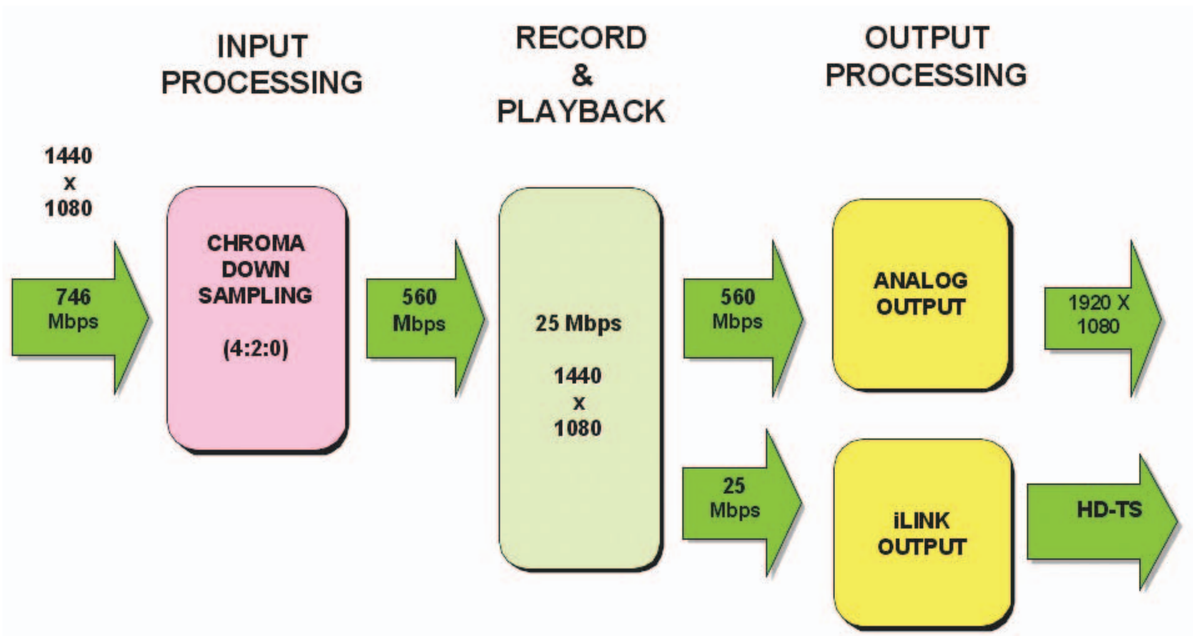
Baseband Signal Processor LSI.
The first building block of the HD Codec Engine is the Baseband Signal Processor. The Advanced Signal Processor is the most sophisticated HD Codec Engine LSI, with over two million logic gates and a total power consumption of 700mW. Following the 180 nanometer process rule, the Baseband Signal Processor receives the signal from the DXP (digital extended processor) after gamma correction and other RGB (4:4:4) camera video processes, and performs advanced image processing functions such as noise reduction and "color correction."

Compression relies on redundancy within an image, and groups of images in the case of MPEG, to effectively reduce the bit-rate without degrading the image quality. Noise is random, and when noise is mixed with the video image the encoder can not discriminate between the noise and the original signal, causing compression inefficiencies and loss of video quality. There are many existing noise reduction systems available today, but traditional noise reduction algorithms effectively reduce noise at the expense of visual loss of fine detail and/ or subtle color and/or luminance shades. To perform noise reduction without introducing visually destructive artifacts, it is necessary to use complex algorithms capable of discriminating noise from low level signals. These noise reduction algorithms are computationally intensive and not available "off the shelf". The advanced noise reduction and signal processing algorithms executed within the Advanced Signal Processor are proprietary and were specifically designed for this application.

After all the complex video processing and noise reduction have been completed, the video signal is down-sampled by the output stage to 8-bit 4:2:0 as required by the MPEG-2 encoder. The architecture of the Baseband Signal Processor applies high bit-depth, high-bandwidth processing on the front end, then scales the signal after the complex processing is finished, preserving subtle image details while minimizing noise and other artifacts. Furthermore, the Baseband Image Processor LSI provides real time, uncompressed analog video output to the viewfinder and LCD panel, as well as 1920 x 1080 analog component video output directly from the camera head. This LSI also offers real time down-conversion of the high-definition signal to a standard definition signal compatible with conventional television displays.

HD-MPEG video encoder.
The second building block of the HD Codec Engine is a miniature MPEG-2 encoder capable of producing high-quality compressed video. The realization of this LSI was a major breakthrough, as an HDV 1080i encoder suitable for portable devices needs to handle the large 1440 x 1080 frame at a high sixty fields per second, with no sacrifice in quality . New technologies developed for high capacity microprocessors were applied to the encoder LSI in order to achieve the complexity necessary for this demanding application. Designed to the 130 nanometer process rule, the HD-MPEG encoder is very compact, yet it contains one and half million transistors and consumes a mere 200mW of power.

HDV RECORDING PROCESS



The HD MPEG Video Encoder LSI is fed by the last stage of Baseband Processor, which delivers 8-bit 1440 x 1080 4:2:0 as required by the MPEG-2 @ ML/ H1440 spec. Therefore, the data rate of the raw signal fed to the encoder is: (1440 x 1080) @ 4:2:0 = 560Mbps. The compression rate must be 22.4:1to produce 25Mbps. A bit-rate reduction ratio of 22.4:1 with no visual degradation requires highly complex and sophisticated algorithms. The MPEG-2 specification does not constrain the encoder algorithm; only the decoder is rigidly defined. This approach enabled engineers to develop and refine proprietary coding algorithms that produce high quality, efficient bit streams which are fully compatible with standard decoders.

After the MPEG Video Encoder LSI compresses the 560Mbps input signal with MPEG-2 @ ML/ H1440, the resultant 25Mbps MPEG-2 is routed to the HD Streaming Processor which conforms it so that it may it may be recorded onto videotape or output through the i.Link interface for editing or storage.

HDV Streaming Processor
The HDV Streaming Processor LSI creates the PES and TS streams in accordance to the HDV format. The TS signal may be output though the i.Link for editing or dubbing. An alternate path adds the ITI (tracking), sub-code, 8X plus 24X visual search signals and error correction signal which are necessary for tape recording and playback. The tape data is organized following the HDV 1080i helical tape track & sector structure specification and RF modulated for tape recording. This LSI was designed to the 180 nanometer process rule and contains one million two hundred gates. It was developed utilizing C Language with a large capacity 36M Bit DRAM and CPU.

HD-MPEG video decoder

The decoder LSI converts the MPEG-2 signal from tape playback or from the i.Link into a high-definition baseband video signal. Its decoding algorithm has been optimized for producing stable output when operating with still images or frame-by-frame tracking, which are usually poorly handled by MPEG-2. This LSI was designed to the 180 nanometer rule with seven hundred thousand transistors and 320mW current draw.

CONCLUSIONS

Conclusions

The HD Codec Engine enables the practical implementation of small, high performance camcorders and VTRs based on the HDV 1080i specification for acquisition and post production applications. These devices can inexpensively record high quality HD video using conventional DV tape as a recording medium. By taking advantage of the tape mechanisms, interfaces, and media already developed for DV, Sony has helped set the stage for affordable HDV equipment and low media costs, creating a straightforward, cost-effective migration path from DV to HD.