



# APOGEE LABS



*Simple to understand  
and simple to use, the  
video encoder and  
decoder offer  
flexibility for a wide  
range of applications.*



## VIDEO TUTOR

### HOW VIDEO SYSTEMS WORK

- Ed Snyder

#### INTRODUCTION

Apogee Labs has spent considerable effort and time developing one of the best video compression systems in the industry. In order to accomplish this, an in depth understanding of how video systems work needed to be absorbed. During the course of dealing with this subject and having discussions with potential users of our product, questions have arisen that needed answering. This document contains an attempt to bring together all of the information that has been needed while answering the myriad of curious inquiries. For some, there is more here than needed. For others there is not enough. However, we hope that there is information here that can be used by all.

Apogee Labs is capable of providing standard products to satisfy a large number of applications involving the gathering of video and reliably transporting the image information to a remote site. We have accomplished this by developing a clear understanding and taking advantage of the characteristics of video from its analog source through the digital transport and reconstruction. If you are interested in the details of video, read on from here. If your interest is primarily in our MPEG system, skip to the last section of this document. In either case, use this document for reference to learn some of the finer details.

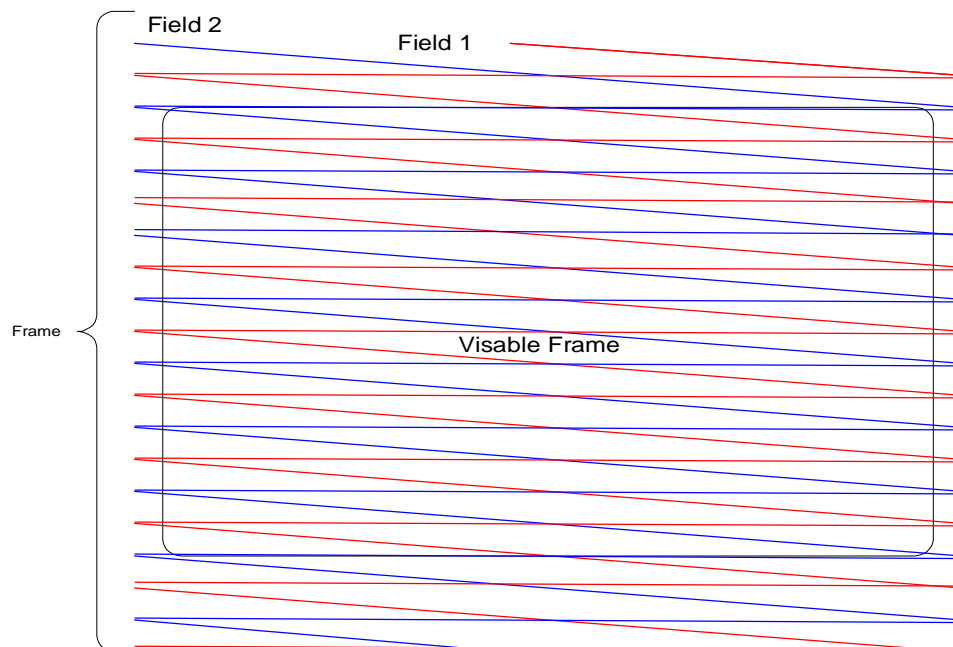
#### ANALOG VIDEO

Any discussion of digital video would be incomplete without a review of analog television principles. It is useful to remember the underlying principles because digital video systems reside in the center between the analog input and output devices, the camera and the cathode ray tube (CRT). Although analog devices are rapidly being replaced by their digital equivalent, the system remains the same as

before this advent so as to maintain backward compatibility. It is significant to remember that a moving picture is in reality an image sequence composed of a series of two-dimensional images that are sequentially ordered in time.

To produce an electronic version of a video image, a transducer is necessary to convert the optical image to an electrical signal. The total image is called a “frame”. In a color television camera, light enters the lens and an image is produced on each of three sensor matrices. Each matrix is sensitive to one of the three primary colors, red, green and blue. Any color in the image can be reproduced by a combination of these three colors. The image on each matrix is scanned into thin horizontal stripes by the scanning process. Television being produced for broadcast currently divides the image into 525 lines in the NTSC standard or 625 lines in the European PAL standard. In the early days of television, the United States standardized on dividing the image into 525 lines. The European community chose 625 lines; thus two incompatible systems developed and have remained to this day. This repeated scanning allows the image to appear to portray motion in objects being shown.

A technique called “interlace” is used to reduce the flicker effect that would be caused by showing as few as 30 frames each second. Figure 1 shows the principle of interlaced scanning.



**Figure 1 Interlaced Video Scan**

The image is scanned beginning in the upper left-hand corner, scanning across to the right. The imaging signal is then turned off and the scan position is returned rapidly to the left-hand side of the image to a position that is the width of one line lower than the previous scan line. Thus, an unscanned line is left between two successive scans. Every other line in the image is scanned--a process that takes one half of the frame time or, in the NTSC system, 1/60th of a second.

Scanning recommences at the top of the screen and scans the previously unscanned image areas. This fills in the missing image stripes and after the additional 1/60th of a second, the second field of the frame is complete.

The first 20 lines of each frame are not used to convey picture information, but are part the vertical blanking interval. The vertical blanking interval (VBI) is the time allotted for retrace when the scanner returns to the top of the screen and includes non-imaging lines. The non-imaging lines do not contain picture information. Some of these lines are used for specific purposes. Line 10 is commonly designated as the point at which switching between two synchronized video sources occurs.

As the lines are scanned, the variations in light are converted into varying voltages. Thus during the time period of the scan line, 63.5  $\mu$ s, there is a continuously varying voltage produced by the scanning device, typically a camera. During the retrace period, the output is blanked, or set to zero volts. During the scanning process three signals, one each for each of the primary colors, are produced with signal amplitudes equivalent to the detected magnitude of the color intensity. At the end of each horizontal line and at the end of the vertical scan synchronization pulses are produced. The time period between the end of one scan line and the beginning of the next is the horizontal blanking interval.

At the receiver, the same three (RGB) varying voltage signals are developed and used to modulate the electron beams in the CRT. By sweeping the beams across the screen in synchronization with the original imaging device, the image that was originally encoded into an electronic signal is reproduced.

The bandwidth necessary to convey reasonable picture detail of a monochrome (black and white) image and the accompanying audio is approximately 6 MHz. Directly transmitting the three signals of an RGB color signal would occupy the equivalent bandwidth of two or three normal black and white television channels. This, of course, is unacceptable in a real-world situation. However, the fine details of an image are contained in the luminance, or brightness portion of the image. Color information can be bandwidth limited without loss of detail in the picture.

High frequencies in the luminance component of an image are due to abrupt changes of brightness level. An abrupt change in color between two parts of an image is almost always accompanied by an abrupt change in brightness level as well. If the edge of an object is well defined by a clean high frequency edge in luminance, the color information can be allowed to vary more slowly without any loss of apparent picture quality.

The system that was developed in the early days of television combined the red, green and blue signals in a matrix to produce a luminance signal, called the "Y" signal and two color difference signals. The Y signal is produced by adding the RGB signals in the proportions of  $0.299 ER + 0.587 EG + 0.114 EB$ . The composite video signal is composed of luminance (Y) and two color difference signals (R-Y and B-T). All three original RGB signals are redeveloped at the receiver from these three signals.

Separating the color information from the luminance information has important consequences for transmission of the signal. Since the color difference signals carry less high-frequency information, they can be bandwidth limited and transmitted in a narrower bandwidth than the luminance signal. By using the largest part of a 6 MHz bandwidth for the Y signal, while modulating the two color difference signals onto a subcarrier, the color signal, along with audio, can be transmitted in a 6 MHz bandwidth.

## DIGITAL VIDEO

If a video stream is to be incorporated into a modern data link it must be in a digital format. Digitized video information is the precursor to image compression. Most of the video sources that we use today produce an analog output. Usually conforming to either the NTSC or PAL standard, this signal is digitized by applying it to an analog to digital converter that samples the analog signal at a regular interval.

The Nyquist theorem states that the sample rate must be at least twice the highest frequency to be converted. If unexpected high frequency components are present the resultant digital conversion becomes aliased. Prior to sampling the signal it is passed through a low-pass filter. Extensive investigation has resulted in the standard sampling rate of 13.5 Msps. This rate was chosen based on two criteria. It is sufficiently fast to satisfy the Nyquist criteria for dealing with the luminance (the highest frequency component) and is also satisfactory for sampling the chrominance information. The two chrominance signals are each sampled at one half the luminance rate. The 13.5 Msps rate also satisfies the 625-line PAL system.

Digital sampling rates for NTSC and PAL are synchronized at two, three, or four times the subcarrier frequency. The shorthand for these rates is 2fsc, 3fsc, and 4fsc, respectively.

With three-channel, component signals, the sampling shorthand becomes a ratio. The first number refers to the sampling rate used for the luminance signal, while the second and third numbers refer to the rates for the red and blue color-difference signals. A 14:7:7 system is one in which the luminance signal is sampled at 14 MHz and the color-difference signals are each sampled at 7 MHz.

Sampling schemes based on multiples of NTSC or PAL subcarrier frequency have been abandoned in favor of a single sampling standard for both 525- and 625-line component systems. Nevertheless, the shorthand remains. In current usage, the shorthand 4:2:2 is as follows: The "4" usually represents the internationally agreed upon sampling frequency of 13.5 MHz. Other numbers represent corresponding fractions of that frequency. A 4:1:1 ratio describes a system with luminance sampled at 13.5 MHz and color-difference signals sampled at 3.375 MHz.

The shorthand continues to evolve. Contrary to what one might expect from the discussion above, the 4:2:0 ratio frequently seen in discussions on MPEG compression does not indicate a system without a blue color-difference component. Here, the shorthand describes a video stream in which there are only two color difference samples (one red, one blue) for every four luminance samples.

Each digital sample results in a numeric value represented by the number of bits in the converter that is proportional to the analog signal. More bits in the converted sample results in greater image resolution. A single bit could be used to indicate fully saturated white and black. By adding bits, or quantization levels, the gradation of the image becomes finer. Both gray-scale and color are affected. Another way of thinking of the advantage of adding bits to the quantizer is signal to noise ratio. Each additional bit added to the quantizer increases the SNR by 6 dB.

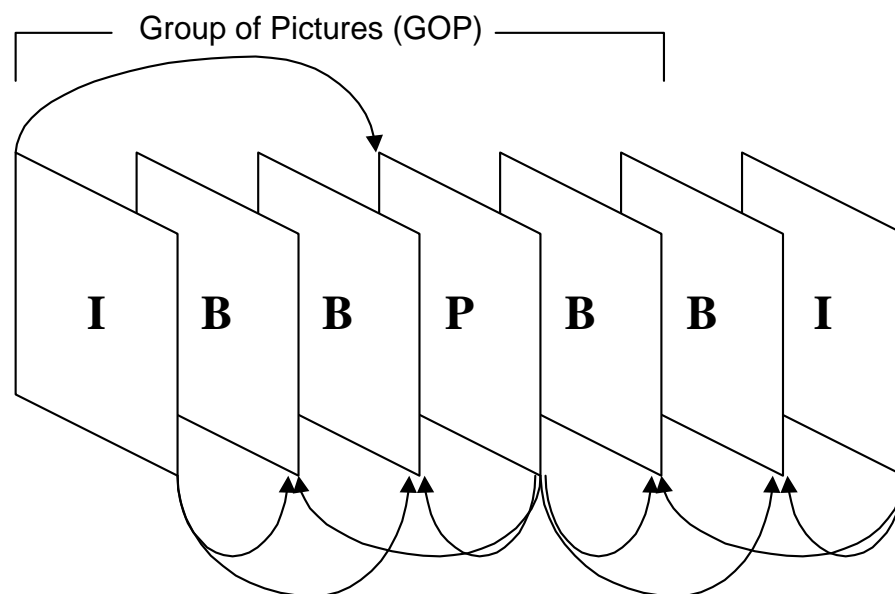
Obviously, for the best picture quality one would choose the highest sample rate and the largest number of bits for the quantizer. The problem is that there is probably not enough bandwidth in the data link to accommodate the serial transmission rate that results. Remember that the luminance is sampled at 13.5 Msps and each of the chrominance difference signals are sampled at 6.75 Msps. If each sample is quantized at eight bits, the resultant serial bit rate is 215 Mbps. A ten-bit system requires the transmission rate of 270 Mbps.

Digitizing an NTSC composite signal is usually accomplished by sampling at four times the frequency of the chrominance subcarrier. The frequency used is  $4 \times 3.58 \text{ MHz} = 14.3 \text{ MHz}$ . This sampling rate produces 910 samples per line. Contrasting with component video where only the visible portion of the video is digitized, when digitizing composite video the entire signal is digitized. The analog synchronization information translates directly to usable digital values. If the digitized video stream is to be passed on a data link, the rate is very high. The only solution is to compress the information.

## VIDEO COMPRESSION

Once the video is digitized it may be applied to a digital data compression process. The most common algorithm in use at this time is MPEG. There are still some applications that use motion JPEG or MJPEG. Both the MPEG and JPEG systems take advantage of the fact that some of the information content is expendable and contribute little to the visible image. Both use the discrete cosine transform to separate the high frequencies from the low. Both systems use quantizers to permanently discard information deemed unimportant. Transformation takes place on groups of pixels, either on a line or in a geographic area of the image. The compression process is performed entirely on each image. This Intra frame compression process allows each image to be captured and reproduced at random.

The primary difference between MJPEG and MPEG is image component or pixel movement. The basic MJPEG video stream is composed of a series of JPEG still images. MPEG adds the component of motion estimation. Instead of sequentially encoding and transmitting completely self-described compressed image (I) frames as is done in MJPEG, MPEG encodes an intracoded frame followed by a series of delta pictures. These delta pictures only contain vectors that describe the movement of image items. When a delta frame calculation is based only on previous images, it is called a P (predictive)-frame. If the prediction is based on previous as well as future frames, it is called a B (bi-directional predictive)-frame. B and P frames are computationally more intensive than I frames; but they improve the compression efficiency since they contain far less data than the I-frame. MPEG-2 improves on MPEG-1 due to its ability to calculate motion estimation between interlaced fields, not just frames. MPEG-2 also supports higher frame resolutions than MPEG-1.



The I-frame is the reference or key picture that allows the complete sequence to be processed to develop the sense of motion. Significant picture to picture transitions, such as scene changes, require key pictures. Because vector prediction is the best possible, but not perfect, pixel cluster to cluster match and because a series of delta pictures build on each other, eventually the video quality degrades unless a key picture (I-frame) is inserted. Errors that enter the bit stream during the storage or transmission affect subsequent delta pictures until the next I-frame resets the sequence. Delta pictures are difficult at best and in most cases impossible to edit.

MPEG-4 does not have a constant definition. This relatively new standard contains a series of audio codecs, video codecs, graphics formats and a robust high-level file format derived from the format that Apple uses for the QuickTime system. Various application needs are satisfied by drawing on the various available components to tailor the codec suite.

At extreme compression levels, the MPEG-1 and -2 technique results in visibly blocky artifacts. The wavelet compression technique simultaneously transforms the entire frame of pixels and can sometimes produce more visually pleasing results at comparable compression ratios. However, this methodology has not seen a standard implementation in the industry. An attempt is being made in the form of the JPEG2000 specification. This standard, although identified as JPEG, is wavelet and primarily addresses still images. Any video implementation of this standard results in a similar process as MJPEG but using wavelet technology instead.

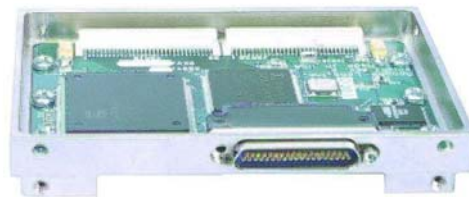
### THE APOGEE LABS MPEG-2 SYSTEM

There are many applications that are completely satisfied by the standard MPEG-2 algorithm while there are others that require a system like the MJPEG system. General purpose applications that have available to them a wide bandwidth data link successfully use the standard MPEG-2 algorithm. To achieve quality video requires a bandwidth exceeding 5 Mbps. Because of the use of predictive information with MPEG, errors in the data become exaggerated between reference I-frames.

Apogee Labs has produced a video digital encoding and decoding system that offers a great deal of flexibility, allowing adjustment to a wide range of measurement and transmission conditions. Simplicity of programming and automatic adjustments always produces the best picture, virtually guaranteeing superior results.



Simple to understand and simple to use, the video encoder and decoder offer flexibility for a wide range of applications.



**Offered as BOTH ground equipment and Airborne packages.**

Simple to understand and simple to use, the video encoder and decoder offer flexibility for a wide range of applications.

Offered as BOTH ground equipment and Airborne packages.

The Apogee Labs Video Encoder (VEM) and Decoder (VDM) Modules utilize standard MPEG II techniques with enhanced functionality and supports:

- Both Inter AND Intra-Frame Compression
- Selectable Frame Rate
- Automatic Decoder Operation
- NTSC and PAL

Today, video surveillance is a basic security requirement in many areas. There are needs to observe empty rooms, highway traffic, street scenes, aircraft landings, views from aircraft in flight, experiments in process and many more. Often the viewing location is separated from the observer by a significant distance. Multiplexing is a requirement when multiple cameras are to be monitored simultaneously.

To maintain security, the video transmission must be encoded. Digital video provides the solution to these issues. However, given realities, the bandwidth of the transmission is limited.

Video compression resolves a multitude of problems: whereas a typical analog video signal is passable in a 6 MHz bandwidth, the digital equivalent requires upwards of a hundred Mbps. It is obvious that such a rate must be reduced to permit the use of most data links. Furthermore, to record digitized video in its raw form requires vast storage capacities.

The VEM and VDM maintain the Apogee philosophy of providing maximum capability devices with minimum programming requirements.

- Regardless of the setup conditions, the VEM automatically compensates to provide the best possible picture quality.
- Knowledge of how the video was generated is not needed to setup the video decoder.
- The video encoder passes all required information to the decoder so that the reproduce function is completely automatic.

The Apogee Labs Video Encoder Modules (VEM) and Video Decoder Modules (VDM) incorporate and exploit MPEG compression techniques. The VEM and complementary VDM devices permit fine-tuning to a specific application.

The method selected for compression of a video image is dependant upon the application. Two factors effecting compression relate to motion and image resolution. Four possibilities exist.

- High motion, high resolution is needed to track an aircraft or missile in flight.
- High motion, low resolution may be used to observe moving objects with little need for detail.
- Low motion, high resolution is required to monitor an area with the need to clearly identify an intruder.
- Low motion, low resolution can be used to monitor static situations without the need for detail.

However, Apogee Labs recognizes that there is more to encoding and transmitting video than just the full data rate at full frame rate and the standard predictive encoding method. To accommodate constraining conditions two unique modes of encoding are included.

- Full frame rate fully compliant MPEG II is supported for both NTSC and PAL video formats.
- Normal mode permits inter-frame compression.
- I-Frame modes (with or without Interlace) support intra-frame compression (I-Frame modes are similar to motion JPEG.)

If data bandwidth is not a constraint but clarity and the ability to step through recorded video frame-by-frame (starting at any spot in the recording) are important, the VEM provides intra-frame compression in the I-Frame modes. These modes eliminate all predictive estimates that are normally in the MPEG compression technique. Compression is performed on each individual picture without regard to preceding or succeeding images. Digitizing video in the I-Frame modes, at the full frame rate, results in reproduced images of the highest fidelity. Recordings made of digitized I-Frame encoded video may be viewed frame by frame beginning at any position in the recorded file.

To facilitate operation on reduced bandwidth data links, the I-Frame mode of the VEM permits selecting reduced frame rates. Nineteen frame rates, from full rate (NTSC 30 and PAL 25 frames per second) to ¼-frame per second are available to support narrow bandwidth applications. At ¼ frames per second, quality images are transmitted at data rates as low as 72 kbps.

The video frame is made up of two independent fields. Each field is a complete image in its own right. The two fields are presented on a CRT display by interlacing. In rare circumstances, two fields from a single frame can be from two different scenes. This frequently occurs in film converted-to-digital movies. At higher frame rates this is not a problem. However, at low rates this effect can be bothersome. To eliminate this condition the VEM is capable of generating its output in non-interlaced mode (I-FR-NI).

Our I-frame only mode (equivalent to Wavelet or MJPEG) produces full, separately encoded, compressed video frames. We are confident that a natural image at 8 Mbps is effectively as good as the source.

In what we call the Normal MPEG mode the compressor produces a full image compression once every 19 images and produces only change information and motion vector information of image elements for the other 18 image frames. There is some concern by image analysts that extremely fine changes in a scene may be missed or that the motion of objects from field to field may be slightly distorted. However, except for those applications that demand pixel-accurate images for detailed analysis, the Normal Mode of our MPEG system produces an extremely high quality image at a data of 3.7 Mbps that is comparable to the original source. This represents quite a large saving from the raw rate of 216 Mbps.

Unique to the Apogee Labs MPEG 2 system is its ability to reduce frame rates using the I-frame mode. This mode maintains maximum quality images while using minimum data bandwidth. For example, running the system at 10 frames per second permits setting the data transmission bit rate down to 2.7 Mbps.

The Apogee Labs video encoding system incorporates a synchronization method that guarantees that the output video rate is identical to the input video rate. All other known digital video systems typically drop or repeat frames as needed in order to maintain synchronization in streaming

applications. The VEM/VDM product line takes advantage of the CCSDS packetized data format that is common to all Apogee Labs MITC and AMITC products to pass all metadata from the encoder to the decoder. This structure assures that the decompression and video reproduction process is totally synchronized. Typically, other systems occasionally overflow or underflow the data stream. When this occurs, frames of video information are either lost or repeated. This is not an acceptable situation where precision is a key parameter of an application.

*Apogee Labs understands video,  
making realization of the toughest applications possible.*