# Silicon Solutions for Digital Video

## 1. Introduction and Overview

Digital video and audio have become an increasingly important component of the Internet sensory experience. Today's rapidly proliferating web sites with first generation streaming media features only begin to reveal the central role that Internet-delivered video will play in the coming years. Video conferencing, video on demand (VOD), distance learning, interactive gaming, and the inexhaustible need to convey compelling, informative, persuasive, and entertaining information will become a growing part of our business and personal lives.

More than bandwidth limitations at the last mile, however, has inhibited streaming media's growth. The processing and computational requirements of digital video media, combined with the current, distributed architecture of the Internet, have introduced enormous challenges to delivering an acceptable video experience through IP-based delivery. Advanced ASIC designs with MPEG-2 video compression has solved, for the most part, the processing requirements for digital video in "closed systems", spawning the enormously successful growth in digital video players (DVDs), video recorders, (DVRs), digital camcorders, and computer-based editing and playback (decoding) systems. Advanced ASIC design will also solve the unique requirements of delivering high-quality video experience over distributed architectures, such as WANs, LANs, and next generation wireless networks.

This white paper will provide an introduction to the markets for digital video, a brief overview of the MPEG standard, an overview of current ASIC design challenges for MPEG encoders, and introduce advanced solutions for optimizing MPEG delivery over IP and other streaming environments. In addition to radical improvements in cost and performance that has fueled the rapid commercialization of DVRs and DVDs, additional features are now available that provide motion estimation, rate control, low-delay and error resiliency algorithms necessary to deliver high quality video over IP-based and other networks where latency and packet loss are common.

## 2. The Market For Video Over IP: "Please Sir, I Want More."

The enormous impact that digital video has made upon our lives, and the major contribution of MPEG standards and codec technology, is attested to by the fact that digital video disk players (DVDs) have become the fastest growing consumer electronics product of all time. Since their introduction in 1997, sales of DVD players grew from 1 million units in 1998 to over 8.4 million in 2000 in the US alone (Consumer Electronics Manufacturers Association, 2001).

Spurred by declining prices and improved picture and sound quality, digital camcorders now represent 26 % of the 5.8 million camcorders shipped to US dealers (or 1.5 million units). In personal video recorders, one out of 10 consumers hopes to purchase a digital hard drive recorder by the summer of 2001 (Consumer Electronics Association, Market Research Study, May 2000).

Digital TV is also growing; in 2000, sales in US alone grew over 400%, representing $1.4 billion in factory dollar volume, according to CEA. The dramatic growth in digital video is also indicated by the sales of digital set-top boxes. Researchers from IDC, Gartner Group, and Joe Peddie and Associates indicate that worldwide sales of digital set-top boxes—including interactive, satellite, cable and DSL-enabled systems—will exceed 80 million units in 2004 (Streaming Media, Electronic News, June 11, 2001).

Beyond consumer electronics applications, digital video is also making an enormous impact in networked applications, including Internet video streaming, video conferencing, video-on-demand, and LAN-delivered video applications.

Today's common streaming applications on corporate, news, and entertainment web sites are just the first generation that attempts to exploit the ubiquity of the networked PC with high quality monitors and sound for video decoding. Initiated during the web's "go-go" days of the late 1990s these applications enhanced and complimented the more critical base of text and graphic content. Next generation video over IP will be driven by economic imperative. These video applications will be cost justified by the savings in travel, productivity improvements, and revenue generation through content sales and advertising.

The market for next generation streaming media is only recently emerging and apparently little affected by the demise of "dot.coms." "Big enterprises are lining up to put streaming on the network," according Jim Ricotta, general manager of Cisco's content networking business (Industry Standard, June 21, 2001). InternetNews reports, "streaming media looks to be the hot sector of the moment—and the next several years" (InternetNews, June 18, 2001).

According to Streaming Media, Inc. among just the sampled companies of an industry survey indicated that over $100 million a year will be spent on streaming in the enterprise market (Enterprise Streaming-Return on Investment, June 20, 2001). Only three of the 100+companies surveyed deployed streaming media before the year 2000, but many companies that have implemented streaming technology in the past 18 months have reported immediate and tangible benefits.

Availability of low-cost bandwidth, combined with declining infrastructure costs and penetration of set-top boxes, is expected to fuel the growth of video-on-demand. Yankee Group estimates the VOD revenues will exceed $65 million in 2001, growing to nearly $2 billion in 2005 (Yankee Group, June 25, 2000). Forecasted shipments in the digital media server market also point to the recent and rapid escalation of streaming applications. According to researchers at Cahners In-Stat Group, the worldwide dollar value of digital media servers will top $1 billion in 2001 and surpass $4.4 billion in four years (Cahners In-Stat Group, May 21, 2001). Currently, Internet streaming solutions represent the fastest growing segment, but the report predicts the VOD server segment will also be on a strong growth trend, surpassing streaming applications after 2005. During the five year forecast period, over 500,000 media servers will be sold.

Much of this streaming media growth is being fueled by the expectation for incremental advertising revenues. In a report by DFC Intelligence, rich media advertising is expected to surpass other forms of online advertising within five years (DFC Intelligence, June 6, 2001). According to the report, over 20% of the video-centric destinations have deployed in-stream advertising with the percentages growing rapidly. According to the Yankee Group, streaming media advertising revenue is estimated to grow to $3.1 billion in 2005 (June 18, 2001).

Beyond traditional corporate and Internet applications, a new arena of streaming video applications is expected to emerge with next generation wireless networks. According to a recent study by IGI Consulting Services, operators of third generation wireless networks must deliver a vast choice of streaming media at affordable prices or face financial disaster (IGI Consulting, May 4, 2001). As Dr. Paul Polishuk, the author of the report states, "Wireless streaming will make or break operators…they need scalable content servers, portable media players, and industry-wide support for micropayments."

## 3. Fundamentals of Data Compression: Why Do You Want to Play?

The goal of data compression is to represent a message (audio, video or both) using as few bits as possible. Although the cost of compressing and decompressing the message can be significant in terms of computational

complexity, the cost of transmitting, storing and caching the uncompressed message is even higher.
An uncompressed television signal requires a bit rate near 216Mbps (or 27MBps). In other words, 3 minutes can be stored on the new 4.7GB DVD. This enormous bit-rate must be reduced by a factor of 100 before it can be economically stored. Furthermore, the need for compression is equally compelling when the video is transmitted over satellite, cable or Internet applications. How many people have a 216Mbps Internet connection?

This paper briefly describes the MPEG-1, MPEG-2, MPEG-4 and H.263 international standards as well as motion JPEG and motion wavelet even though these are not standards.

## *3.1 International Standard: MPEG-1, MPEG-2, MPEG-3, MPEG-4 and H.263*

All three MPEG algorithms use the following techniques:

1. Frame-to-frame (interframe) prediction to remove temporal redundancy
2. The Discrete Cosine Transform (DCT) is used to convert the image from the time domain to the frequency domain, which is more efficient. The result of the process is a set of weighted coefficients corresponding to the DCT waveform.
3. Quantization is the process of deleting some of the weighted coefficients. This process is irreversible and is the only source of loss in the coding scheme.

To perform interframe prediction, the image is divided into 16x16 pixel arrays called macroblocks. The encoder searches the previously compressed picture looking for a macroblock that closely resembles the current macroblock. The macroblock is coded as a displacement from the best-matched macroblock (commonly referred to as a motion vector). The process of finding the motion vectors is known as motion estimation and the resulting prediction is known as motion compensation prediction.

The MPEG algorithms code the pictures as any of the following:

1. Intra (I)-frames, which do not contain any interframe predictions. I-frames can be decoded independently and can be used as random access points into the stream.
2. Predicted (P)-frames, which are predictions from the previous frame. When no prediction can be made the P-frame’s macroblock is coded as an Intra-frame macroblock.
3. Bi-directional predicted (B)-frames, which are predictions from either a previous or future frame.

### 3.1.1 MPEG-1

In 1988, in response to a growing need for a common format for coding and storing digital video, ISO established the Moving Pictures Expert Group (MPEG). MPEG-1 was born to compress data and store it onto a digital storage device such as a CD-ROM. MPEG-1 was optimized for video rates around 1.5Mbps, resolutions around 352x240 (or 288) and progressively scanned images (not TV signals).

MPEG-1 has been a very successful standard. It is, currently, used for video CDs, Internet traffic and surveillance.

### 3.1.2 MPEG-2

In 1990, MPEG started the second phase of its work to develop extensions to MPEG-1 that would allow for greater input-format flexibility, higher data rates and better error resiliency. This standard became known as MPEG-2. MPEG-2 added support for interlaced pictures, video bit rates from 1.1Mbps to 20Mbps, enhanced Chroma fidelity and scalability. MPEG-2, also added a number of improvements to the compression efficiency including: more flexible DCT transforms, more sophisticated quantization and more complex inter-frame predictions.

MPEG-2 is arguably the most commercially important video compression algorithm on the market. It is used in digital cable, digital satellite, DVD, HDTV and Super VCD players. An MPEG-2 decoder can play an MPEG-1 or MPEG-2 compressed video stream.

#### 3.1.3 MPEG-3

MPEG-3 was formed to enhance MPEG-2 to high-definition (HD) resolutions. It was found that MPEG-2 supported HD so MPEG-3 was eliminated.

#### 3.1.4 MPEG-4

MPEG-4 was started as an attempt to improve the quality of low bit-rate encoded video; however, as it progressed, it evolved into much more. It has been refocused to meet the anticipated needs of the entertainment TV, computing and telecommunication convergence.

MPEG-4 adds many refinements to the basic MPEG motion compensated DCT approach, including changes that increased MPEG-4's robustness to transmission errors.

MPEG-4 brings higher levels of interaction with content, increased networking interaction and far more reusability versus today's video. The following features have been added:

1. Scenes are decomposed into media objects. These media objects can be encoded separately, can be natural or synthetic in origin and can be packetized together or separately.
2. The system level, responsible for multiplexing the encoded objects together, contains features for interactivity and network processing, with limited QoS.

MPEG-4 may ultimately supplant MPEG-2 for all applications, however, initially it is targeted at low bit-rate applications from 64Kbps to 2Mbps.

#### 3.1.5 H.263

H.263 was built as an enhancement of the H.261 video coding standard. H.261 is similar to MPEG-1 in that it uses a combination of DCT and differential coding. The primary differences are that only I and P-pictures are used and at low-bit rates some macroblocks (or entire frames) can be skipped. H.263 added some of the same features as MPEG-4. In fact, the MPEG-4 short-header format is compatible with the baseline for H.263. The H.263 standard is targeted at low-bit rate applications, primarily, video conferencing.

H.263 standard, just like MPEG, has evolved into many different versions over time. H.263 version 1 is commonly known as H.263, H.263 Version 2 is commonly known H.263+, while H.263 version 3 is the latest standard in the H.263 family, known as H.263++. The enhanced versions achieve higher compression efficiency, more robust error handling and higher resolution support. Many features in MPEG-4 and H.263 are very similar, and some people believe that the two standards will unite in the near future.

### *3.2 Popular Alternatives, But Not International Standards: Motion JPEG and Motion Wavelets*

#### 3.2.1 Motion JPEG

JPEG is an international standard that is very similar to MPEG's intra-frame coding. JPEG was invented before MPEG and is common throughout the Internet. Each frame is totally, independent in the JPEG standard; thus, there is no such thing as motion JPEG.

However, JPEG is used in two different ways to provide proprietary motion algorithms:

1. Each frame is independently compressed and transmitted in their totality.
2. Each frame is independently compressed and the differences between adjacent frames are transmitted (this is a degenerated form of MPEG).

JPEG's primary advantages are in algorithm complexity. The algorithms can be easily performed on today's computers without hardware assistance.

### 3.2.2 Motion Wavelet

Wavelet compression was developed in the 1980s and can be thought of as a standard digital filtering technique. Instead of breaking the image into blocks, the entire image is quantized. At high compression ratios, JPEG images tend to look blocky. Wavelet, on the other hand, does not divide the image into blocks so the entire image gets fuzzy as the compression is increased.

To generate motion wavelet each frame is independently compressed. Differential coding is not performed frame-to-frame, thus, the overall bit rate is higher for motion wavelet compared to its MPEG alternatives.

Wavelet is used primarily in surveillance applications and in Internet still images. The underlying video techniques behind wavelet have been incorporated into MPEG-4.

## 4. The Guts of Building an MPEG Encoder

To establish standards that enable interoperability between systems, MPEG establishes the composition of the stream structure necessary for widespread decoder adoption. The MPEG standard outlines the general decoding process, the data stream hierarchy and picture-type definitions, and synchronization techniques. The MPEG standard defines the parameters to establish synchronized timing by providing system clock references and presentation timestamp requirements. The MPEG standards do not define an encoding process. They only specify the syntax of the coded bit stream; therefore, the ingenuity behind the MPEG encoding process is left to the developer.

Attesting to the computational requirements of video compression, the advanced knowledge of compression algorithms, and the performance requirements needed for delivering acceptable viewing quality, only a handful of companies have achieved a cost-effective ASIC for consumer electronic applications.

The implementation of the MPEG encoders has resulted in many types of designs, which include many embedded processors, long data paths, millions of gates and significant amounts of external memory. Integrating additional features to make a true "system-on-a-chip" has not been cost effective.

The actual data compression used in MPEG combines three techniques:

1. Quantization
2. Variable length coding
3. Prediction methods

### *4.1 Removing Picture Information / Quantization*

The video data is transformed from the time domain to the frequency domain. This is done using the discrete cosine transform (DCT), which transforms the image in a way that makes it possible to perform an inverse DCT to create an exact copy of the original image.

The resulting frequency coefficients are divided by a certain value, which can cause some of the coefficients to become zero. This step creates picture loss and is termed quantization. As the quantization becomes larger the shades of gray become fewer and fewer. Quantization is a technique that is commonly used to increase the compression ratio at the cost of decreased picture quality.

### *4.2 Variable Length Coding*

After the quantization process, the coefficients are represented by a code. Frequently used coefficients are defined using a code that uses as few bits as possible. Further, it is hoped that many of the coefficients will be zero so the code gives back two values. The first value specifies the number of zeros in front of a non-zero coefficient. This method is termed Variable Length Coding (VLC).

### *4.3 Prediction Methods*

MPEG performs 2 types of predictive modeling: spatial (within a frame) and temporal (across 2 frames). MPEG exploits the fact that the viewer cannot perceive certain characteristics due to our visual systems.

Spatial prediction exploits the fact that the intensity of the current pixel is probably the same as that of the previous pixel. Imagine for the moment, a blue sky or field of grass. The first pixel is coded and each successive pixel is a difference from the first. Furthermore, the pixels can be quantized such that they cause slightly different shades of luminance and chrominance to correlate to the same value. The resultant representation is then runlength encoded to produce a small packet of compressed data.

Temporal prediction applies techniques to find the appropriate macroblock that is close to the previous frames macroblock. Motion estimation is managed by the MPEG standard through identifying the macroblocks, motion vectors, and error correction requirements used to further compress video streams. The motion estimator compares each predicted macroblock with macroblocks in a previously stored reference picture or pictures. It finds the macroblock in the reference picture that most closely matches the new macroblock. The motion estimator then calculates the motion vector to represent the horizontal and vertical displacement from the macroblock being encoded to the matching macroblock-sized area in the reference picture. Error prediction is also generated through motion compensation and sent to the decoder as a residual signal.

## 5. Designing an MPEG Encoder: The Hard Reality + Vweb's Solution

Recent changes in process technology along with ingenuity have led to the development of cost effective solutions. These advances have led to the prediction that the MPEG encoder market will grow from less than 1% of the total MPEG sales to more than 10% by 2005 (according to In-Stat's MPEG Video Chip Market Report).

Vweb Corporation (www.vwebcorp.com) was formed to create MPEG compression and decompression (CODEC) solutions that produce the best image quality while creating a cost-effective solution. In order to achieve these lofty goals, we designed our own proprietary rate control and motion estimation algorithms.

The Vweb patent-pending motion estimation is loosely tied to H.263 and sampling theory. Basically, we perform a multi-path coarse search looking for a correct macroblock. Upon finding a potential candidate, we finely search around the locality of the coarse estimation, which yields a more accurate motion estimation. Using this technique, the mathematical complication has been greatly reduced. This allows us to produce a more compact and faster algorithm for creating the motion vectors. Furthermore, this solution uses less on-board memory.

In addition, we created a flexible rate control unit. The unit is based on a proprietary RISC engine. The rate control unit controls the number of bits that are allocated for each frame using techniques that adjust the quantization appropriately. The rate control monitors several parameters within the engine producing six different possible implementations:

1. Fixed rate "strictly enforced" constant bit rate (CBR). This mode does not allow VBR overflows or underflows. This mode is appropriate for fixed bandwidth applications where the video quality can be sacrificed to ensure bandwidth restrictions.
2. "Section-based" CBR. This mode is much like the fixed rate CBR mode discussed above; however, based on feedback from an external source (usually network notification messages), the rate can be adapted to fit within the existing

conditions of the pipe. This mode is appropriate for network applications where one video stream is sent to one receiver.

3. “Loosely-based” CBR. This mode does not prohibit VBR overflows and underflows. It assumes that the data will be written to a file or that error handling is being performed at the receiver. This mode is appropriate for general applications.
4. “Low-delay mode” CBR. This mode allows frames to be skipped. This is used for video conferencing applications just in case the network feed is heavily congested.
5. “Storage-based” variable bit rate (VBR). This mode does not prohibit VBR overflows and underflows. It uses a “look-behind” model to reduce the latency. The current macroblock is compared to the previous frame’s macroblock to calculate how many bits to apply to this frame. This is appropriate for PVR and high-quality video applications where disk storage is fixed over a period of time, but bursty within that period of time.
6. “DVD-based” VBR. This mode protects against VBR underflows and timing characteristics caused by the DVD write mechanisms. This mode is used for DVD, VCD or SVCD applications.

## 6. Vweb’s Products

Our current offering, the VW2000 chip is roughly half the size of alternative products. It uses 4MB of SDRAM, has a 32-bit SDRAM interface and has less than 500K gates. The MPEG-2 chip delivers outstanding video quality over a broad range of networks, including the Internet as well as in standalone consumer devices. The single-chip solution is ideal for various applications, notably: streaming video with multicasting technology, video on demand, video conferencing, digital camcorders, digital servers, personal video recorders, DVD play and recorders, Internet Appliance, Internet video creators, set-top boxes and video capture cards. The VW2000 supports NTSC (29.97) and PAL (25) frame rates, programmed GOP structures (IP, IBP, and IBBP) and resolutions up to 720 x 480 (or 576 in PAL).

Our next product, the VW2005, is an MPEG-1, MPEG-2 and MPEG-4 audio & video encoder, which addresses the broadcast and distribution of digital video, particularly at low bandwidths. We expect this to fuel the widespread deployment of video over IP, moving the market from its current nascent stage to one of material commercial significance. The VW2005 supports NTSC and PAL, dynamically adjustable frame rates and resolutions, and programmable GOP structures.

Subsequently, we will produce an MPEG-1, MPEG-2 and MPEG-4 audio & video CODEC, named the VW2010. The VW2010 addresses the DVD recordable, PVR, transcoding, transrating and capture card markets. The VW2010 supports NTSC and PAL, dynamically adjustable frame rates and resolutions, programmable GOP structures, five graphic planes and multiple post-processing filters.

## 7. Future Expectations: The MPEG-4 Era

The massive MPEG-4 effort (three versions are completed and additional extensions are in development) introduces a number of new features that address interactivity, object-based bitstreams, enhanced compression, and other enhancements.

Through both the introduction of new compression techniques and more effective transport mechanisms, MPEG-4 will yield up to 25% better compression than MPEG-2. MPEG-4 is poised for streaming video running on a variety of interactive broadband networks. (Broadband Week, February 5, 2001).

Major consumer electronics players also see the impact of MPEG-4. "We see MPEG-4 playing similar role (in IP networks) as MPEG-2 played in digital broadcasting," said Ahmad Ouri, vice president and general manager of

MP4Net, a unit of Philips. MPEG-4 "widens the scope of content delivery beyond the PC and set-top box" (Multichannel News, April 4, 2001).

Furthering the popularity of the MPEG-4 is the Internet Streaming Media Alliance. This industry group was formed late last year by Apple, Cisco, Kasenna, Philips and Sun, with a goal to accelerate adoption of MPEG-4 to stream rich media over the Net.

"We definitely see MPEG-4 as a way to reduce costs for video on the Internet and broadband service providers want that," says Sean Badding, senior analyst for media research firm, The Carmel Group. (Broadband Week, February 5, 2001). "We see it really taking off in the 2001-2002 time frame."

With our core technology, we are poised to assume a leading role in this take-off.