# Video compression makes big gains

*New standards for video compression plus new IC chips will change the world of computing, broadcasting, and communication*

Peng H. Ang, Peter A. Ruetz, and David Auld
LSI Logic Corp.

Digitally encoded video is upon us. Just as audio compact discs have revolutionized the recording industry, so this new technology promises to unleash a whole slew of video applications—among them, the digital laserdisc, electronic camera, videophone and -conferencing system, image and interactive video tools on personal computers and workstations, program delivery on cable and satellite, and high-definition television (HDTV). Unlike the digital audio technology of the '80s, however, many of the applications of digital video hinge on the use of data compression. The audio bandwidth, after all, is about 20 kHz, which translates into a digital data rate of about 1.4 megabits per second for high-quality stereo sound. Sampled video source signals, on the other hand, require much higher bit rates, ranging from 10 Mb/s for broadcast-quality video to more than 100 Mb/s for HDTV signals.

Even when still pictures are involved, as in image archival systems, a mountain of data is needed to represent them. For example, a color image with resolution of 1000 by 1000 picture elements (pixels) at 24 bits each will occupy 3 megabytes of storage in an uncompressed form. This will not fit onto a high-density floppy diskette, which can hold just 1.2 Mbytes.

Meanwhile, to facilitate industry growth, three standards are being developed for still and moving pictures and for videoconferencing. Sets of chips already exist for all three purposes, some of them proprietary designs, and others in agreement with the standards closest to completion.

**HOW IT WORKS.** Compression methods build on both redundancies in the data and the nonlinearities of human vision. They exploit correlation in space for still images and in both space and time for video signals. Compression in space is known as intra-frame compression, while compression in time is called inter-frame compression. Generally, methods that achieve high compression ratios (10:1 to 50:1 for images and 50:1 to 200:1 for video) are lossy in that the reconstructed data are not identical to the original.

Lossless methods do exist, but their compression ratios are far lower, perhaps no better than 3:1. Such techniques are used only in sensitive applications such as medical images. For example, artifacts introduced by a lossy algorithm into an X-ray radiograph may suggest an incorrect interpretation and alter the diagnosis of a medical condition. Conversely, for commercial, industrial and consumer applications, lossy algorithms are preferred because they save on storage and communication bandwidth.

These lossy algorithms also generally exploit aspects of the human visual system. For instance, the eye is much more receptive to fine detail in the luminance (or brightness) signal than in the chrominance (or color) signals. Consequently, the luminance signal is usually sampled at a higher spatial resolution. (For example, in broadcast quality television, the digital resolution of the sampled luminance signal is 720 by 480 pixels, while for the color signals it may be only 360 by 240 pixels.) Second, the encoded (or compressed) representation of the luminance signal is assigned more bits (a higher dynamic range) than are the chrominance signals.

> Video data, even after 100:1 compression, can be decompressed with close to analog videotape quality

Also, the eye is less sensitive to energy with high spatial frequency than with low spatial frequency. Indeed, if an image on a 13-inch personal computer monitor were formed by an alternating spatial signal of black and white, the human viewer would see a uniform gray instead of the alternating checkerboard pattern. This deficiency is exploited by coding the high-frequency coefficients with fewer bits and the low ones with more bits.

All these techniques add up to powerful lossy compression algorithms. In many subjective tests, reconstructed images that were encoded with a 20:1 compression ratio are hard to distinguish from the original. Video data, even after compression at ratios at 100:1, can be decompressed with close to analog videotape quality.

**STANDARDS.** Lack of open standards could slow the growth of this technology and its applications. Three digital video standards that have been proposed are the Joint Photographic Experts Group (JPEG) standard for still picture compression; the Consultative Committee on International Telephony and Telegraphy (CCITT) Recommendation H.261 for video teleconferencing; and the Moving Pictures Experts Group (MPEG) for full-motion compression on digital storage media (DSM).

JPEG's proposed standard is a still-picture-coding algorithm developed by a research team under the auspices of the International Standards Organization (ISO). The team convened in 1987 and the algorithm is currently an ISO committee draft 10918 recommendation. The scope of the algorithm is broad: it comprises a baseline lossy approach and an extended lossless approach, as well as independent functions using coding techniques different from the baseline one. Only the first approach will be discussed here.

The JPEG baseline algorithm falls under the heading of transform-based image coding. A color image can be represented in different color systems. Those in wide use today include R-G-B (the three primary colors red, green, and blue) in the computer industry; Y-U-V (Y for luminance or brightness, U and V for color difference signals Y-R and Y-B, respectively) in the television industry; and C-M-Y-K (cyan, magenta, yellow, and black) in the printing industry. Within each color system, the constituent parts are called components. Thus, there are three color components in the R-G-B system, four in the C-M-Y-K system.

Each component of the source image in the JPEG encoder and decoder is divided into non-overlapping blocks of 8 by 8 pixels [Fig. 1]. Each block is then transformed using the two-dimensional discrete cosine transform (DCT) with an 8-by-8 kernel.

The resulting 64 coefficients, computed as a 2-D array of 8-by-8 numbers, represent the frequency contents of the given block. The DCT coefficient value in the upper left-

0018-9235/91 $1.00©1991 IEEE

hand corner of the 2-D array measures the energy of the zero-frequency or direct current (dc) term. (For example, if the original 8-by-8 image has a constant value, then only the dc term in the transformed space is nonzero.) The other 63 entries are alternating current (ac) coefficients; they give the relative strengths of signal terms with increasing horizontal frequency from left to right and for terms with increasing vertical frequency from top to bottom.

Next, the DCT coefficients are quantized. The quantization step size varies with frequency and component. The dependence on frequency reflects the fact that the high-frequency coefficients subjectively matter less than the low-frequency ones and may therefore be quantized with a larger step size (that is, more coarsely). In addition, an individual component may have its own quantization table. In the committee draft of the JPEG algorithm, up to four quantization tables are allowed.

Following quantization, the coefficients are re-ordered into a one-dimensional array by reading out the entries of the two-dimensional array along a zigzag route. In this way, the quantized coefficients are "approximately" arranged in order of ascending frequency.

Next, the dc and ac coefficients are loss-

## Defining terms

**Discrete cosine transform (DCT):** conversion of time-domain into frequency-domain data.

**Entropy coding:** an efficient coding method that encodes frequent events with fewer bits than it does infrequent events.

**Huffman coding:** a type of entropy coding that uses predetermined variable-length code words.

**Inverse discrete cosine transform ($DCT^{-1}$):** data conversion from the frequency into the spatial domain.

**Inverse quantization:** scaling up of previously quantized data into larger-range numbers.

**Joint Photographic Experts Group (JPEG):** a group of engineers and scientists from many companies worldwide seeking to formulate a standard method for the compression and decompression of still-frame, continuous-tone, photographic (gray-scale or color), digitized images.

**Moving Pictures Experts Group (MPEG):** a group of industry experts seeking a standard for coded representation of video and associated audio for digital storage media.

**Predictive coding:** the coding of each pixel by quantizing the difference between its current value and a predicted value, computed from past values.

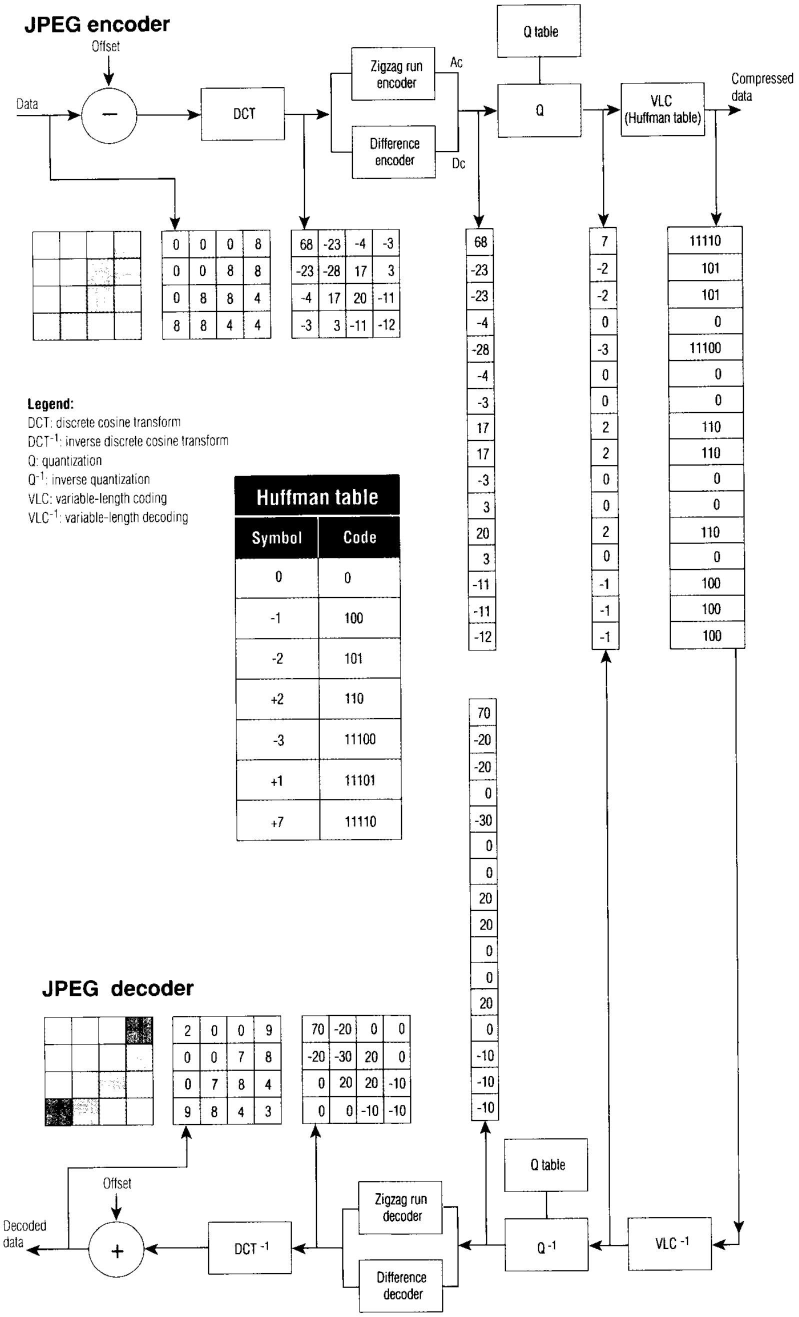
**Quantization:** a process whereby data represented by many bits is scaled down to a lower-precision value requiring fewer bits.

**Transform coding:** mapping (or transforming) pixels of images from one domain (spatial, say) into another (such as frequency).

**Variable-length coding:** encoding so as to create code words of variable number of bits [see Huffman coding].

**Variable-length decoding:** decoding of code words of variable length into raw data.

## Joint Photographic Experts Group encoder and decoder



**Legend:**
DCT: discrete cosine transform
$DCT^{-1}$: inverse discrete cosine transform
Q: quantization
$Q^{-1}$: inverse quantization
VLC: variable-length coding
$VLC^{-1}$: variable-length decoding

**Huffman table**

| Symbol | Code |
|---|---|
| 0 | 0 |
| -1 | 100 |
| -2 | 101 |
| +2 | 110 |
| -3 | 11100 |
| +1 | 11101 |
| +7 | 11110 |

*[1] The baseline algorithm for the compression of still images included in the Joint Photographic Experts Group (JPEG) proposed standard divides the image into 8-by-8 pixel blocks, represented here by a 4-by-4 block for simplicity. In the encoder, the 16-level gray-scale image is first digitized, then undergoes a discrete cosine transform (DCT) that yields 16 frequency coefficients. The two-dimensional array is read in a zigzag fashion to reorder it into a linear array. The coefficients obtained by quantization—here just dividing by 10—are then coded using the Huffman table, shown here in a simplified version.*

*The decoding path takes the variable-length coding (VLC) output as input and recovers the quantized coefficients, multiplies each by 10 (inverse quantization) and turns the linear array into a 2-D one through an inverse zigzag operation. An inverse DCT operation subsequently yields a 40-bit image, representing a compression ratio of 1.6:1 here. Much higher ratios can typically be achieved.*

lessly encoded, both using Huffman-style coding but keyed with different parameters. Huffman coding is a well-known means of reducing the number of bits needed to represent a data set without losing any information. The dc coefficients are differentially encoded so that the dc coefficient of the previous 8-by-8 block of the same component is used to predict the dc coefficient of the current 8-by-8 block and the difference between these two dc terms is encoded. The Huffman code table for the dc term is based on the difference values.

The zigzag-coded ac coefficients are first run-length coded. This process reduces each 8-by-8 block of DCT coefficients to a number of events. Each event represents a nonzero coefficient and the number of preceding zero coefficients. Since the high-frequency coefficients are more likely to be zero, Huffman-coding these events makes it possible to achieve efficient compression.

The JPEG baseline algorithm provides for up to two tables each for dc and ac coefficient coding. In JPEG decoding, the encoding algorithm is simply run in reverse, and it is therefore commonly described as a symmetric algorithm.

As for videotelephony, the CCITT Recommendation H.261 specifies a method of communication for visual telephony. It is often called the p*64 standard because the data rate on the communication channel is p times 64 kilobits per second, where p is a positive integer less than or equal to 32. For p=1, then, a low-quality video signal for use in picture phones can be transmitted over a 64-kb/s line. If p=32, a high-quality video signal for teleconferencing can be transmitted over a 2-Mb/s line.

The standard specifies the organization and interpretation of the transmitted bits so that two coder-decoders (codecs) from different manufacturers may successfully establish a videoconferencing session. A CCITT encoder is more complicated than the JPEG encoder, although one can distinguish familiar functional blocks such as the DCT and the quantizer [Fig. 2, top]. The decoder, however, is less complex [Fig. 2, bottom].

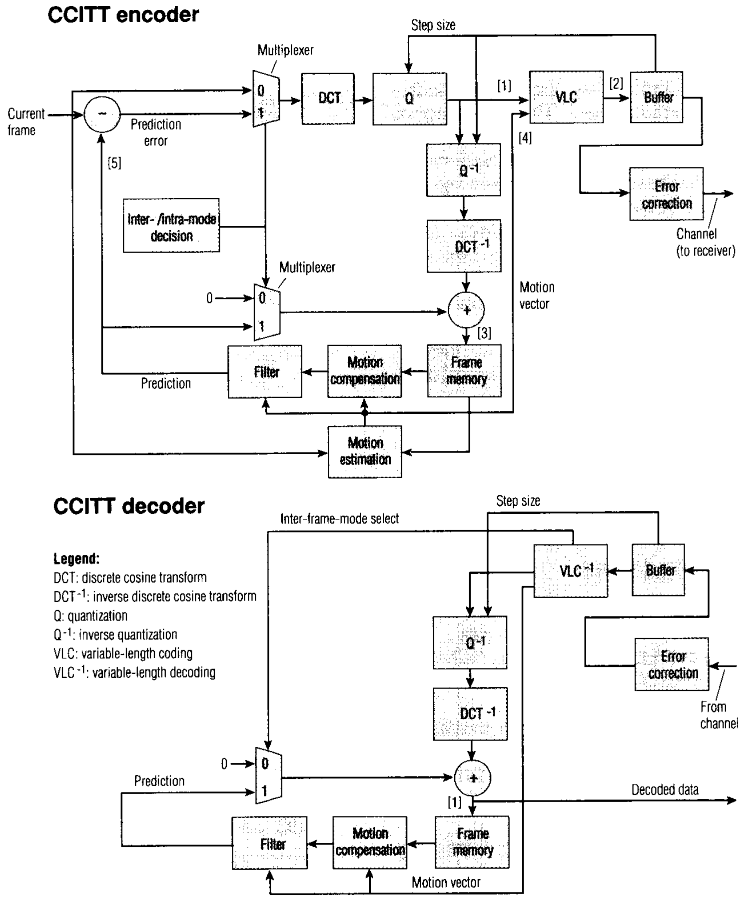
A detailed description of the baseline algorithm is beyond the scope of this article. Instead, an overview of the techniques used will be given.

The CCITT coder is called hybrid because it combines transform coding—it is also DCT-based—with predictive coding, in which a block in the current frame is predicted from a block in the previous frame using a feedback loop. (In contrast, the JPEG algorithm basically operates in an open loop and is reset at the end of each image.) It is this inter-frame prediction that results in a higher compression ratio.

Also, instead of basing predictive coding purely on the difference between the current frame and the reconstructed image in the frame memory, the CCITT H.261 standard has an optional specification for motion compensation. This increases the predictive coder's efficiency at tracking the inter-frame motion and hence its success at bit rate reduction.

**PROPOSED FULL-MOTION STANDARD.** Like H.261, the MPEG proposed standard is a full-motion compression algorithm with both intra- and inter-frame modes. Unlike H.261, the data rate may not exceed the 1.5 Mb/s of today's digital storage media. (There are ongoing development efforts to increase the data transfer rate for next-generation media.) Present media include compact disc drives, digital audio tape (DAT) drives, and magnetic hard disks.

The standards committee for MPEG has



*[2] The first of a sequence of frames is intracoded in the full motion compression scheme proposed by the Consultative Committee for International Telegraphy and Telephony (CCITT). Like JPEG, each 8-by-8-pixel block of the frame is encoded with the discrete cosine transform (DCT) and then quantized (top). At point [1], there are two signal paths: one leads toward a receiver through a lossless coder and optional error-correction circuitry; the other—a feedback—is inverse quantized and undergoes inverse DCT to yield a reconstructed block for storage in frame memory. Reconstruction is needed because inter-frame compression uses predictive coding, which requires the encoder to track the behavior of the decoder to prevent the decoder's reconstructed image from diverging from the original input. The coder can be bypassed (instead of the signal being taken at [2]) because it is lossless and without influence on the reconstruction process. When the entire frame has been processed, a reconstructed image as seen by the decoder is stored in the frame memory at [3]. Next, inter-frame coding is applied. To compensate for motion, each 8-by-8 block in the current frame is matched with a search window in the frame memory. Then the motion vector that represents the offset between the current block and a block in the prior reconstructed image that form the best match [4] is coded and sent to the receiver. The predictor at [5] gives the motion-compensated 8-by-8 block from the reconstructed frame. The difference between this and the original block is transform coded, quantized and coded before being sent to the receiver. The CCITT decoder (bottom) first corrects incoming bit stream errors, and then decodes the data in the variable-length decoder. Passage through the inverse quantizer and the inverse DCT yields the DCT coefficients. At [1], a block like one in the encoder's feedback loop has been reconstructed and stored in the frame memory. In inter-frame mode, motion vectors extracted from the variable-length decoder are used to provide the location of the predicted blocks.*

not completed its work and the system and audio specifications remain in flux. To a large degree, the functionality of the H.261 encoder block diagram applies. However, the specifics of quantization and motion estimation/compensation and coding are different.

Recently a second phase of standardization dubbed MPEG-2 has begun. Its aim is efficient coding for up to 10 Mb/s with the further goal of higher-quality results.

**SILICON SOLUTIONS.** Many digital video applications will require low-cost silicon implementations to become broad-based. Toward the end of 1990, there was a flurry of announcements of chip set solutions for image and video compression. Some are nonstandard approaches; others conform to such draft standards as JPEG and H.261.

Solutions for JPEG were offered first by C-Cube Microsystems Inc., San Jose, Calif., then by LSI Logic Corp., Milpitas, Calif. The former is a single-chip CL550 announced in late 1989 and conforms to an early revision of the JPEG draft. The company is currently reworking the design to meet the latest JPEG specifications. LSI Logic's semicustom two-chip solution combines the L64735 DCT processor with the L64745 quantizer plus JPEG variable-length coder. The partitioning into two chips kept costs down because smaller die have higher yields and can reside in cheaper plastic packages.

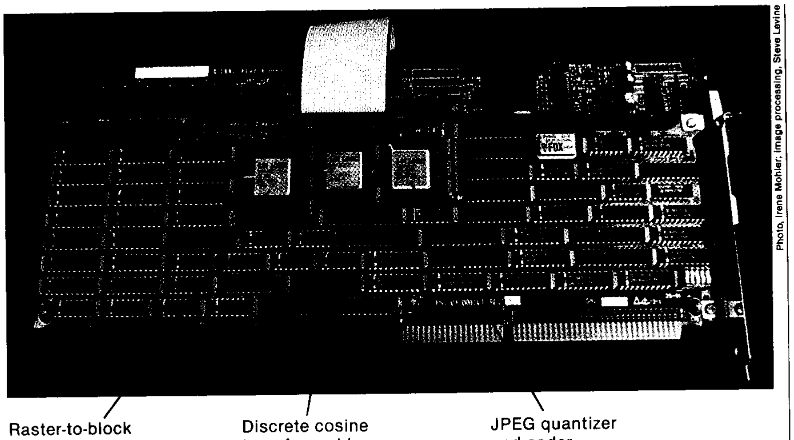
The architectural approaches of C-Cube and LSI Logic are similar. Both the CL550 and the LSI parts implement the JPEG algorithm, but the user may change such JPEG parameters as image resolution, number of components, and the quantization and Huffman code tables used.

H.261 products are available from LSI Logic and Graphics Communications America Ltd. (GCA), Glen Ellyn, Ill. LSI Logic announced a seven-chip building block family for general full-motion video encoding and decoding (including H.261) in September 1990, and GCA followed with a 12-chip building block family about a month later. Both are best classified as function-specific video codec building blocks.

The LSI Logic line consists of four processors and three coding and decoding chips. The processors are the L64720 for motion estimation, the DCT L64730, the L64740 for quantization and L64760 for intra/inter-frame decision. The other three chips are the L64715 Bose-Chaudhuri-Hocquenghen (BCH) error-correcting codec, the L64750 H.261 variable-length encoder and L64751 H.261 variable-length decoder, all specific to H.261.

The four processors may be used in any transform-based compression system. For instance, the MPEG video coding loop (minus the variable-length coder) can be built with the LSI chips.

**DESIGN FLEXIBILITY.** Function-specific building blocks give video system designers the freedom to configure their own compression architecture at a system level. The parts do not require complicated microcode and are generally easy to use.



Photo, Irene Mohler; image processing, Steve Levine

*Image compression boards that comply with the Joint Photographic Experts Group (JPEG) proposed standard are becoming available. The one shown here—Visionary JPEG Compression Engine from Rapid Technology Corp., Amherst, N.Y.—has three LSI Logic Corp. ICs performing raster-to-block conversion, the discrete cosine transform, and quantization and coding. Its connection to Rapid Technology's Visionary Video Board (in the back) allows real-time video data or stills to be directly transferred between the two boards.*

In the category of proprietary solutions, there is a two-chip set from Intel Corp., Santa Clara, Calif., and a single-chip from UVC, Irvine, Calif. The two-chip Intel i750 comprises a pixel processor and a display processor, both components of Intel's Digital Video Interactive (DVI) architecture. The pixel processor has a 48-bit-wide microcode instruction and supports the compression and decompression of images and motion video. It also synthesizes graphics primitives. The display processor takes bitmaps from memory and displays them on the CRT. Since May 1990, Intel has been shipping a software product using the chips to perform compression and decompression based on a vector-quantization algorithm achieving real-time performance for 256-by-240-pixel images. A feature of Intel's approach is its programmability.

In its product announcement, UVC claims a chip capable of real-time 500:1 compression, using purely intra-frame techniques.

One other company, SGS-Thomson Microelectronics, Phoenix, Ariz., has announced parts suitable for use in a compression system. It has several DCT processors (STV3200, STV3208, and A121) and a motion estimation processor (STV3220).

**A DIGITAL TREND.** Many new applications in video technology will be digital, as is underscored by the Federal Communications Commission's recent statement that it favors an all-digital HDTV approach. In the computing arena, interest is already growing in the integration of video with graphics and audio. Often referred to as "multimedia" applications, these will be used in interactive education, next-generation graphics systems, network videoconferencing, and other user-friendly systems.

In short, the technology encompassing digital video and audio integrate the worlds of broadcasting and communication with the world of computing. Ten years from now, these three industries will not be distinguishable.

**TO PROBE FURTHER.** Two textbooks with good discussions of general image encoding techniques are *Digital Image Processing* by William K. Pratt (John Wiley and Sons, New York, 1991) and *Digital Image Processing* by Rafael C. Gonzalez and Paul Wintz (Addison-Wesley, Reading, Mass., 1986). Arch C. Luther wrote an *IEEE Spectrum* article on interactive digital video (DVI) technology in September 1988, and subsequently authored a definitive book on DVI technology entitled *Digital Video in the PC Environment*, 2nd edition (McGraw-Hill, New York, 1990). Another recent book on HDTV is entitled *HDTV: Advanced Television for the 1990s* by K.Blair Benson and Donald G. Fink (McGraw-Hill, New York, 1991).

The April 1991 issue of the *Communications of the Association for Computing Machinery* (CACM) was devoted to emerging compression standards with relevance to digital multimedia.

**ABOUT THE AUTHORS.** Peng H. Ang (M) joined LSI Logic Corp., Milpitas, Calif., in July 1984. He is vice president and general manager of the digital signal processing (DSP) business unit, which introduced a line of image and video compression ICs in September 1990.

Peter A. Ruetz (M) joined LSI Logic in June 1986. He is currently a senior member of the DSP group's technical staff.

David Auld joined LSI Logic in 1990. He is currently the applications manager of the DSP group. Prior to joining LSI, he was with Inmos Corp., which later became part of SGS-Thomson Microelectronics.

◆