HEWLETT® PACKARD

# An Overview of MPEG-2

**The 1997 Digital Video Test Symposium**

Author: Dragos Ruiu

**Discover the latest techniques and products for real-world video testing**

MPEG stands for Moving Picture Experts Group. The name was given to the International Standards Organization (ISO) committee that specified a standard compression, transmission, and decompression scheme for video. Today MPEG refers to a series of standards documents officially known as ISO/IEC 11172 (MPEG-1) and ISO/IEC 13818 (MPEG-2). MPEG actually specifies the syntax of the compressed video and audio bitstreams, which enables a great deal of flexibility at the encoder.

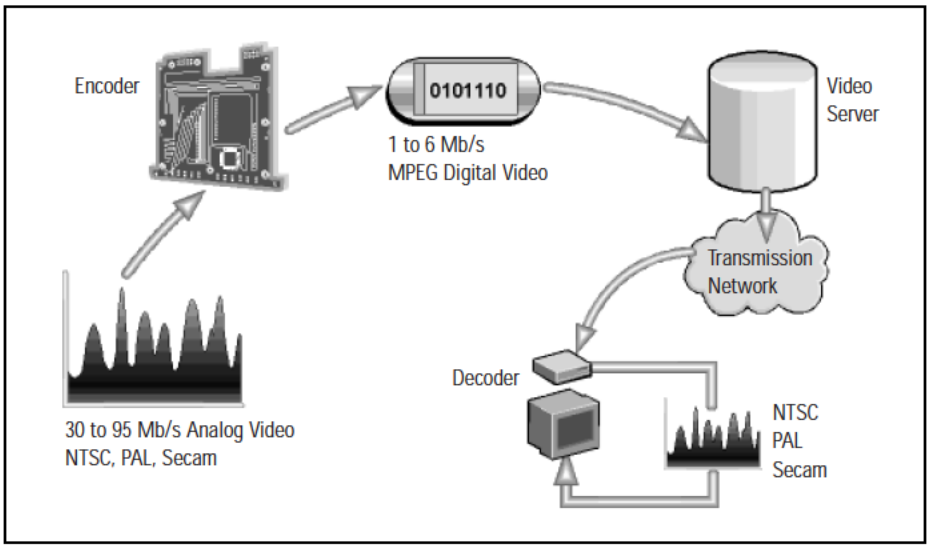


Diagram 1: Video Compression Process

The basic job of MPEG is to take analog or digital video signals and convert them into packets of digital information that are more efficiently transported on modern networks. MPEG compresses the video into much less information, consuming less transmission bandwidth-only one-sixth to one-thirtieth of the capacity is needed. A digital transmission, MPEG maintains the transmission quality end to end. As progressively longer transmission networks are used, the signal does not degrade and the picture does not get fuzzy.

There are many important ramifications of the technologies incorporated into the MPEG specifications, but what seems to get the most press is the video compression system. MPEG shares much of its compression technique with a related standard called JPEG. JPEG stands for Joint Pictures Experts Group,



Peloton - EX1025, Page 2
IPR2023-01239 (Peloton v. NEC)

which is another work-group under the same sub-committee (29) of the Joint Technical Committee 1 of the International Electrical Commission of the International Standards Organization. That name is quite a mouth-full so it is commonly referred to as JPEG.

JPEG, a standard for compression of still images, takes advantage of the optical characteristics of the human eye and removes picture information that is not very visible to viewers. The image is stored in a format that removes information, such as high frequency color transitions, which is not readily discerned by the human eye. The eye is much more sensitive to high frequency luminance (brightness) changes than to color changes, due to its rod (luminance) and cone (color) structure. So when the information is converted back into a picture, it is hoped that the human eye does not notice any change in the picture caused by the loss of information in this compression system. The information is lost in the Discrete Cosine Transform, quantization, and entropy coding processes used for the compression.

MPEG takes advantage of spatial and temporal redundancies in video material to compress the information. The following explains the video compression process.

## Sampling the Chrominance Information

The first of several steps in the compression is to translate the information in the picture into the frequency domain. The red, green and blue intensity information in each pixel is translated into luminance/brightness values (Y), as well color vectors (U, V).

The chrominance information can then be subsampled. There are three formats:

- 4:4:4
  The chrominance and luminance planes are sampled at the same resolution.
- 4:2:2
  The chrominance planes are subsampled at half resolution in the horizontal direction.



Peloton - EX1025, Page 3
IPR2023-01239 (Peloton v. NEC)

- 4:2:0
  The chrominance information is subsampled at half the rate both vertically and horizontally.

After the RGB to YUV conversion, pixels are grouped together into rectangular areas called blocks and groups of blocks called macro-blocks.

## Spatial Compression

These blocks are then translated into frequency information using a Discrete Cosine Transform, similar to the more familiar Fourier transform. This yields a series of coefficients indicating the magnitude of cosine functions at increasing frequencies so that the original signal can be reassembled in the spatial domain.

## Quantization and Entropy Coding

The quantization step truncates some of the least significant bits of information, making some coefficients go to zero. This is then entropy coded, a system that converts the coefficients into variable bit-length codes with the most common coefficients being coded with the fewest number of bits. This coding scheme is sometimes called Huffman encoding.

An example of encoding English text in this fashion would code the most common letter, e, as one bit (1), the next most common letter as three bits (010), and so on (011, 0010, 0011, 00010, 00011, 000010, 000011, etc.).

MPEG and JPEG use a special form of two-dimensional Huffman code that also encodes the number of zero coefficients preceding the encoded value. The DCT component number is also encoded as the difference from the last sample by subtracting it from the last one (DPCM coding). This kind of compression only removes redundant information within one frame, and one of the properties of video signals is that there is much redundant information repeated from frame to frame. More compression can be achieved by not re-transmitting these static portions of the picture.



Peloton - EX1025, Page 4
IPR2023-01239 (Peloton v. NEC)

## Temporal Compression

Some blocks can be predicted from blocks in previous frames. Frames that contain these predicted blocks are called P-frames. But what happens if transmission errors occur in a sequence of P-frames? And what happens if the first frame of a video sequence is lost? Decoder errors do occur. To avoid the propagation of errors, and allow resynchronization, a complete frame that does not rely on information from other frames is periodically transmitted (approximately once every 12 frames, or 2 to 3 times a second). These stand-alone frames are named intra-coded or I-frames. There is also a third kind of frame which borrows information/blocks from frames that occur both before and after it. These bi-directional frames are called B-frames.

This process can be taken one step further by encoding motion vectors so that only portions of a picture that move, or can be borrowed from other locations in previous frames of the video, are encoded using fewer bits. To do this, the 8x8 pixel blocks are grouped into fours to create 16x16 macroblocks. Macroblocks that do not change are not re-encoded in subsequent frames. With P-frames, the encoder searches the previous frame (or frames before and after for B-frames) in half-pixel increments for other macroblock locations that are a close match to the information contained in the current macroblock. If no adequate matching macroblocks are found in the neighboring region, the macroblock is intra-coded and the DCT coefficients are fully encoded. If an adequate matching macroblock is found in the search region (which is dictated both by the search algorithm and the speed of the encoder), the full coefficients are not transmitted, but a motion vector is used instead to point to the similar block(s). The decoder then uses information from the other frame when decoding the current frame. In this fashion, the full DCT coefficients for the block do not need to be transmitted and transmission bits are saved.



Peloton - EX1025, Page 5
IPR2023-01239 (Peloton v. NEC)

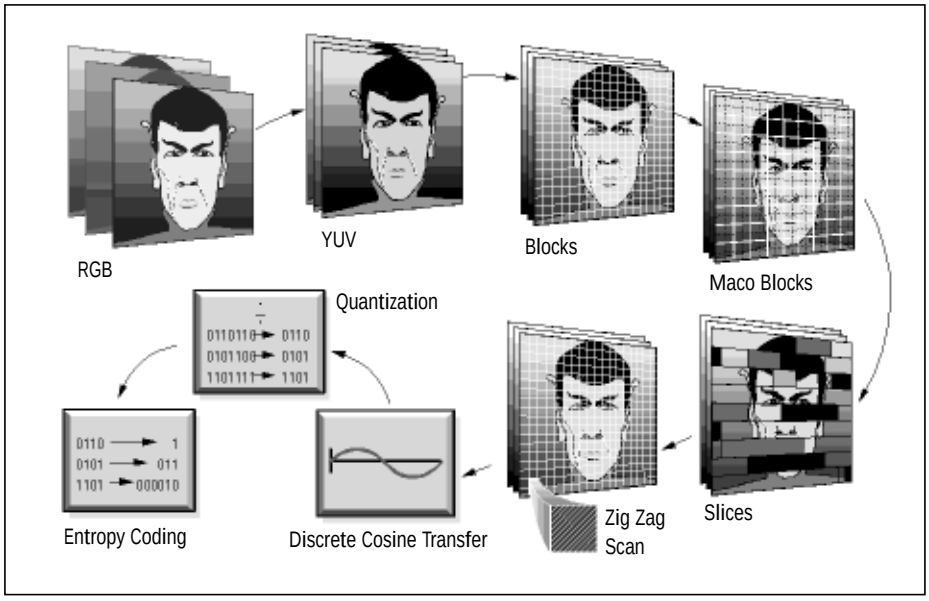


Diagram 2: MPEG Compression Process

This kind of a compression scheme is also used in the H.261 (P*64) videoconferencing standard. The original MPEG standard, MPEG-1 was targeted at producing a trade-off between compression, ease of implementation, and processing power required for decompression. While H.261 was intended for videoconferencing, and therefore optimized for low-resolution images transmitted with low encoding latency (ruling out B-frames that require more buffering and latency in decoding) on ISDN and other low-speed links, MPEG was designed for broadcast and multimedia (i.e. CD-ROM video) applications. Due to the motion vector searching, the encoding effort for MPEG is asymmetrical-it is more work and requires more expensive, powerful hardware for encoding than decoding. This design was justified because a video signal will need to be decoded many times for viewers, while it only needs to be encoded once. Optimizing the cost of the decoder for mass production was a key concern. MPEG-1 was designed for optimum efficiency at the bandwidths available on CD-ROM drives (~1.5 Mb/s) and with a restricted degree of complexity so that it can be implemented on a single chip in the 0.8 micron logic prevalent in the late eighties. As microprocessor CPUs sped up it soon became feasible to decode this information in software.



Peloton - EX1025, Page 6
IPR2023-01239 (Peloton v. NEC)

To achieve this decrease in data volume, MPEG-1 reduces a standard CCIR-601 television signal from NTSC in North America and Japan, and Europe's PAL and SECAM to something called SIF (Standard Image Format). While MPEG-1 can encode images at up to 4096 x 4096 x 60 frames per second, most applications use a format called CPB (Constrained Parameter Bitstream). CPB has a maximum rate of 1.86 Mb/s to closely match the transfer rate of CD-ROM drives and can adequately compress SIF resolution. SIF is half of the horizontal and vertical resolution of CCIR-601, and only half again of the chrominance in the vertical direction (i.e., 4:2:0 format). Some implementations also have the frame rate to ease decoding processing to a level that can be implemented on slower processors using software decoders.

## MPEG-2

MPEG-2 is the successor to MPEG-1. MPEG-2 was optimized for the digital compression of TV broadcast material, and yields little visible degradation in quality from CCIR-601 when transmitted at 1.5 to 6 Mb/s.

## MPEG Audio Compression

The MPEG standards cover compression of audio as well as visual information. Audio compression takes advantage of the psychoacoustic properties of the human hearing systems. Several levels of compression are defined. The basic process is to transform the fixed-rate audio samples (from 11 to 48 thousand 16-bit samples per second) into the frequency domain with a sine transform. The frequency components are then divided into sub-bands, or ranges of frequencies and entropy encoded. Some of the sub-bands are inaudibly removed.

The key principle guiding the removal of sub-band information (or energy as it is sometimes called) is related to the design of the ear's cochlea. The cochlea is the spiraling cavity in human ears and is covered with fine hairs that detect the air vibrations caused by sound frequencies between 100 to 20,000 Hz (we feel-rather than hear-frequencies lower than this, and few people can actually hear all the way to the upper end of the



Peloton - EX1025, Page 7
IPR2023-01239 (Peloton v. NEC)

range). Depending on their position in the spiral, the hairs in the cochlea are sensitive to different frequencies. When we hear an intense sound at a particular frequency the firings of the nerves in adjacent hairs are masked by the neuron response to the intense sounds. So a loud sound at one frequency masks adjacent frequencies. Sending the bits in the audio stream that belong to masked signal is completely unnecessary because we can't hear that information anyway. Empirically, through years of subjective listening tests, masking curves for the human ear have been derived. But our knowledge of the human ear and nervous system is far from complete, so designing efficient audio compression is part art and part science.

| Rate | Application |
|---|---|
| 1.5 Mb/s | VHS Quality Video for Film Material |
| 2 to 3 Mb/s | Sports |
| 4 Mb/s | Most Users Detect No Visible Degradation |
| 6 Mb/s | Broadcast Quality |

Diagram 3: MPEG-2 Quality Levels

With sub-band-masking and some other tricks, a 1.5 Mb/s stereo CD audio stream can easily be compressed to 128 kb/s or lower with little, if any, perceived loss in audio quality. MPEG defines several audio compression levels that are optimized for different amounts of processing power and delay. More compression means more delay lag in the algorithm and that more processing power is needed. The lower compression levels can be done in software by a microprocessor while the higher levels require dedicated silicon or fast digital signal processors.



Peloton - EX1025, Page 8
IPR2023-01239 (Peloton v. NEC)

Other tricks used in MPEG audio involve removing the redundant information in the multiple channels of audio. All of the compression levels use some form of master/slave stereo, which converts the two left and right channels into a master channel and a slave channel derived from the information contained in the master channel. The layer two and three algorithms also use intensity stereo, removing stereo imaging at frequencies above 2 kHz because humans can't really discern positional information at high frequencies. (Where is that high-pitched whine coming from?) The layer three algorithm uses Huffman encoding (variable bit length) of the resulting information.

| Layer | Rate | Maximum Delay | Typical Minimum Delay | Original Technology |
|---|---|---|---|---|
| 1 | 192 kb/s | 50 ms | 19 ms | Philips DCC |
| 2 | 64 kb/s | 100 ms | 35 ms | MUSICAM |
| 3 | 64 kb/s | 150 ms | 59 ms | MUSICAM, ASPEC |

Diagram 4: MPEG Audio Compression

The art of audio compression is constantly progressing, and recent advances use different sub-band widths depending on frequency and calculations in the time domain of pre- and post- masking of frequency impulses. MPEG-2 adds more audio channels to MPEG-1's L/R stereo. MPEG-2 has five compressed audio channels (left, center, right, right surround rear, and left surround rear) as well as a special channel for low-frequency effects. MPEG-2 also adds an option for up to seven channels of alternate language audio. The basic compression systems of MPEG-2 are backward compatible with MPEG-1, but recent advances in the quality of compression such as the Dolby AC-3 algorithm have opened the issue again, and a new, non-backward compatible high quality system is being specified.



Peloton - EX1025, Page 9
IPR2023-01239 (Peloton v. NEC)

## Video Implementation Optimization

The MPEG-2 standards are a strict superset of MPEG-1 (MPEG-2 decoders can decode MPEG-1) and the data format is upwardly compatible. MPEG-2 aims for a different set of optimal compression parameters than the earlier standard. While MPEG-1 was optimized for 0.8 micron logic integrated circuit implementation, MPEG-2 increases the compression processing to the level that can be achieved on a single chip with the more dense 0.5 micron logic that can now be fabricated. Today's denser logic allows more functions and processing power to be put on one chip.

## Scalability

Another goal of MPEG-2 was to extend the video formats capable of being carried. This included:

- Spatial scalability

  Carries a video signal in a two-part format that lets inexpensive decoders extract a low-resolution signal, and with additional processing in more capable decoders extract a higher resolution picture using more data (and bandwidth). In this way one broadcast can be used to transmit an HDTV image and still be compatible with NTSC, PAL and SECAM decoders.

- Temporal Scalability

  MPEG-2 allows one signal to be transmitted and displayed at different frame rates.

- Signal-to-Noise-Ratio (SNR) scalability

  Allows one encoded signal to be compatible with different levels of decoding quality.

- Data partitioning

  Allows the MPEG signal to be transmitted over a two priority channel. One channel contains critical information such as the DC values, low-order DCT coefficients, and motion vectors.



Peloton - EX1025, Page 10
IPR2023-01239 (Peloton v. NEC)

The other channel carries less critical information such as higher-order coefficients and information that refines picture detail. This two-part transmission is well-suited to systems like ATM where the Cell Loss Priority bit in the cell header can be used to mark cell discard priority. Portions of the video can be separated so that the most important parts of the video transmission will have a greater chance of arriving uncorrupted at the destination.

## Levels and Profiles

This expandability of the MPEG-2 format allows it to serve the needs of many different kinds of applications. This is aided by defining several levels of decoders, and several profiles of video sources. The profiles define limits on the algorithmic complexity that may be used in the video signal (and complexity in the encoder and decoder). The level defines the resolution and the quality of the video. This level flexibility can be used to let HDTV transmissions be encoded using MPEG-2, which brought about the cancellation of the planned MPEG-3 (1920 x 1080 x 30, 20 to 40 Mb/s) specification. The other remaining MPEG specification, MPEG-4, is targeted at very low bit-rate applications (4800 b/s to 64 Kb/s) and videoconferencing quality video (176 x 144 x 10 frames per second) is still in the early stages of development.

Diagram 5: Profiles & Applications

| Profile | Application |
|---|---|
| Simple | Same as main but without B-frames, primarily intended for software decoders |
| Main | Low-cost single chip implementation for cable TV and satellite uplink compression |
| Spatial | Main with spatial scalability, e.g. HDTV |
| SNR | Spatial with SNR scalability |
| High | SNR with 4:4:4 chrominance in the macroblocks |



Peloton - EX1025, Page 11
IPR2023-01239 (Peloton v. NEC)

## Transport and Program Streams

The other major area of extension in MPEG-2 was in the area of transport systems. The MPEG-2 specification shares the same structure as the MPEG-1 specification:

- ISO/IEC 11172-1 MPEG-1 Systems
  ISO/IEC 11172-2 MPEG-1 Video Coding
  ISO/IEC 11172-3 MPEG-1 Audio Coding
- ISO/IEC 13818-1 MPEG-2 Systems
  ISO/IEC 13818-2 MPEG-2 Video Coding
  ISO/IEC 13818-3 MPEG-2 Audio Coding

The systems specification in MPEG-2 was vastly overhauled. The MPEG-1 format was extended with two data stream formats that are very flexible and can be used in many applications. The systems document specifies a system for multiplexing together the video, audio, and data portions of a single program for environments that transmit in a relatively error-free environment such as computer multimedia applications and digital storage recorders called the Program Stream, which is used for applications such as DVD, and another system, the Transport Stream which can be used for broadcast, Video on Demand, and cable TV.

The program stream specifications describe how to encode the data from the multiple sets of video, audio, and data (called Program Element Streams, or PES) within a single set of variable length packets (instead of MPEG-1's fixed 2 k CD sized packets). These packets have headers that specify timing and buffering information so that they maybe delivered to a decode with a limited RAM buffer size. Also included is additional information that describes the program, such as frame and audio sampling rates, that are vital for efficient memory allocation. The program stream has positioning indicators for video streams that are intended to be displayed as a window on a portion of a larger screen.



Peloton - EX1025, Page 12
IPR2023-01239 (Peloton v. NEC)

The program stream updates the pack and packet layer of the MPEG-1 Systems specification with some modifications for currently-desired applications that were not as well understood when MPEG-1 was designed.

Less evolutionary, and more revolutionary, is the transport stream (TS), which is a major extension of the MPEG-1 specifications. The transport stream defines a packetized protocol for multiplexing multiple MPEG compressed programs, and program directory (e.g. Channel Guide) information into a packetized fixed-length format for transmission on digital networks. It is a structure that can be thought of as digital TV. It also includes sophisticated timing distribution, synchronization, and jitter correction mechanisms that are essential for the transmission of video signals over long distances.

The major additions of timing information and small fixed-length packets opens up a whole range of new applications for MPEG-2. The 188-byte transport packets map very well into 48-byte ATM cell payloads, allowing MPEG-2 to be used in switched video architectures that are now possible using the tremendous bandwidth capacity of newly developed ATM switches. With the timing information in the packets, the transport stream is suitable for land-based cable TV distribution, satellite uplink compression, direct satellite digital broadcasting, delivering video streams over long distance cable networks (e.g. SONET/ATM), and video conferencing.

The TS packets contain clock synchronization information in fields called the Program Reference Clock (PCR), Decoding Time Stamp, Presentation Time Stamp (PTS). During the video transmission it is important that the transmitter and receiver maintain synchronization and frame rate. Subtle variation in the clock frequency used at the encoder/transmitter and the decoder/receiver can lead to overflowing or empty buffers-especially because of the high data rates involved with video material. To avoid this, the PCR field is used to synchronize a 27 MHz master clock at both ends.



Peloton - EX1025, Page 13
IPR2023-01239 (Peloton v. NEC)

The clock consists of a 42-bit counter that increments at 27 MHz (with the upper 33 bits incrementing at a 90 kHz rate). The PCR field is used to transmit periodic samples of this counter. The receiver can compare its counter with the received values and a simple phase locked loop (PLL) circuit can be used to adjust the local rate to accurately match the transmitter. This clock can be used with the frame time stamps that specify when a frame needs to be displayed, to buffer and smooth jitter that may have occurred when the MPEG stream was transported through variable delay (e.g. ATM switch buffering). These timestamps can also be used for effects such as slow motion and pausing the video (e.g VCR-like control).

There are several buffering and traffic rate policing mechanisms built into the MPEG systems protocol definition. One is specifically intended to reduce delay (buffer latency) for video conferencing applications. Another is used to minimize the amount of RAM the decoder uses. Because of the large size of the bandwidth streams involved in video transmissions, it can take a lot of RAM to buffer a video stream. Optimizing this RAM usage is an important goal because the memory chips are a major cost driver for mass-produced set-top boxes. The bandwidth is also inversely correlated to the memory usage; i.e., a more compressed stream requires more RAM to be decompressed. To ensure the possibility of compatible low-cost implementations, the transport stream defines bounds on the filling of buffers (buffer occupancy). Two buffer algorithms define these limits, one for transport streams and another for the elementary stream decoding.



Peloton - EX1025, Page 14
IPR2023-01239 (Peloton v. NEC)

The transport stream specification also makes it feasible to create a gradual transition from an analog TV broadcast architecture to a digital one. Switching from traditional analog to digital transmission systems offers an immediate advantage in that by using digital modulation schemes (e.g., Quadrature Amplitude Modulation, QAM-64, QAM-256, VSB, COFDM, etc.) a traditional 6 to 8 MHz analog television channel frequency allocation can carry 25 to 40 Mb/s of data allowing the transport of more than 5 times as many channels! Since terrestrial frequency allocations are scarce, better satellites and upgrading land-lines expensive, installing compression is a clear benefit that allows many broadcasters to immediately justify the costs associated with installing digital video equipment in transmission equipment such as CATV and satellite uplinks. Other transport stream features allow the creation of hybrid analog/digital networks to allow the deployment of new user terminals while maintaining compatibility with the huge existing installed base of analog televisions. The digital channels can co-exist with traditional analog basic services, and a user could still access basic services with a TV set. By adding a digital set-top box the user gets access to data, video-on-demand (VOD), and near-video-on-demand architectures (NVOD).

The transport protocol also provides features for broadcast scrambling and transmission of encryption keys. It has a special message format, called Entitlement Management Message (EMM), for individually addressing set-top box decoders in a broadcast. These Entitlement Management Messages (EMM) carry information such as encryption keys and the addresses of the devices that have paid for this program.

The transport stream also specifies data structures that allow the decoder to separate digitally multiplexed channels on one transmission.



Peloton - EX1025, Page 15
IPR2023-01239 (Peloton v. NEC)

## For more information

For more information on Hewlett-Packard Test & Measurement products, publications or services, please call your local HP sales office. A current listing is available at: http://www.hp.com

**United States:**

Hewlett-Packard Company
Test and Measurement Organization
5301 Stevens Creek Blvd.
Building 51L-SC
Santa Clara, CA 95052-8059
1-800-452-4844

**Canada:**

Hewlett-Packard Canada Ltd.
5150 Spectrum Way
Mississauga, Ontario L4W 5G1
905-206-4725

**Europe:**

Hewlett-Packard
International Sales Europe
Geneva, Switzerland
+41-22-780-4111

**Japan:**

Hewlett-Packard Japan Ltd.
Measurement Assistance Center
9-1, Takakura-Cho, Hachioji-Shi
Tokyo 192, Japan
(81) 426-48-3860

**Latin America:**

Hewlett-Packard
Latin America Region Headquarters
5200 Blue Lagoon Drive, 9th Floor
Miami, Florida 33126 U.S.A.305-267-4245, 305-267-4220

**Australia/New Zealand:**

Hewlett-Packard Australia Ltd.
31-41 Joseph Street
Blackburn, Victoria 3130
Australia
131-347 Ext. 2902

**Asia Pacific:**

Hewlett-Packard Asia Pacific Ltd.
17-21/F Shell Tower, Time Square
1 Matheson Street, Causeway Bay
Hong Kong
(852) 2599-7070

5966-1031E 06/1997 Rev A

**Specifications subject to change**