

AN
PRODUCTION

**DIGITAL
VIDEO: AN
INTRODUCTION
TO MPEG-2**

Barry G. Haskell,
Atul Puri, and
Arun N. Netravali

Join Us on the Internet

WWW: <http://www.thomson.com>
EMAIL: findit@kiosk.thomson.com

thomson.com is the on-line portal for the products, services and resources available from International Thomson Publishing (ITP). This Internet kiosk gives users immediate access to more than 34 ITP publishers and over 20,000 products. Through *thomson.com* Internet users can search catalogs, examine subject-specific resource centers and subscribe to electronic discussion lists. You can purchase ITP products from your local bookseller, or directly through *thomson.com*.

Visit Chapman & Hall's Internet Resource Center for information on our new publications, links to useful sites on the World Wide Web and an opportunity to join our e-mail mailing list. Point your browser to: <http://www.chaphall.com/chaphall.html> or <http://www.chaphall.com/chaphall/electeng.html> for Electrical Engineering

A service of **ITP**[®]

Digital
Multimedia
Standards
Series



DIGITAL VIDEO: AN INTRODUCTION TO MPEG-2

Barry G. Haskell

Head, Image Processing Research Department, AT&T Labs

Atul Puri

Principal Member of Technical Staff,
Image Processing Research Department, AT&T Labs

Arun N. Netravali

Vice President of Research, Bell Labs,
Lucent Technologies



CHAPMAN & HALL



INTERNATIONAL THOMSON PUBLISHING

New York • Albany • Bonn • Boston • Cincinnati • Detroit • London • Madrid • Melbourne
Mexico City • Pacific Grove • Paris • San Francisco • Singapore • Tokyo • Toronto • Washington

Cover design: Curtis Tow Graphics

Copyright © 1997 by Chapman & Hall

Printed in the United States of America

Chapman & Hall
115 Fifth Avenue
New York, NY 10003

Chapman & Hall
2-6 Boundary Row
London SE1 8HN
England

Thomas Nelson Australia
102 Dodds Street
South Melbourne, 3205
Victoria, Australia

Chapman & Hall GmbH
Postfach 100 263
D-69442 Weinheim
Germany

International Thomson Editores
Campos Eliseos 385, Piso 7
Col. Polanco
11560 Mexico D.F.
Mexico

International Thomson Publishing-Japan
Hirakawacho-cho Kyowa Building, 3F
1-2-1 Hirakawacho-cho
Chiyoda-ku, 102 Tokyo
Japan

International Thomson Publishing Asia
221 Henderson Road #05-10
Henderson Building
Singapore 0315

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without the written permission of the publisher.

1 2 3 4 5 6 7 8 9 10 XXX 01 00 99 98 97

Library of Congress Cataloging-in-Publication Data

Haskell, Barry G.

Digital video : an introduction to MPEG-2 / Barry G. Haskell, Atul Puri, and Arun N. Netravali.

p. cm.

Includes bibliographical references and index.

ISBN 0-412-08411-2

1. Digital video. 2. Video compression -- Standards. 3. Coding theory. I. Puri, Atul. II. Netravali, Arun N. III. Title.

~~TK6680.5.T137~~ 1996
621.388'33--dc20

96-14018
CIP

British Library Cataloguing in Publication Data available

"Digital Video: An Introduction to MPEG-2" is intended to present technically accurate and authoritative information from highly regarded sources. The publisher, editors, authors, advisors, and contributors have made every reasonable effort to ensure the accuracy of the information, but cannot assume responsibility for the accuracy of all information, or for the consequences of its use.

To order this or any other Chapman & Hall book, please contact **International Thomson Publishing, 7625 Empire Drive, Florence, KY 41042**. Phone: (606) 525-6600 or 1-800-842-3636. Fax: (606) 525-7778. e-mail: order@chaphall.com.

For a complete listing of Chapman & Hall titles, send your request to **Chapman & Hall, Dept. BC, 115 Fifth Avenue, New York, NY 10003**.

To our families
Ann, Paul, and Andrew
Diane
Chitra, Ilka, and Ravi

TK 6680
5
H37
1997
Copy 2

CONTENTS

Dedication	v
Preface	xiii
Acknowledgments	xiv

1	Introduction to Digital Multimedia, Compression, and MPEG-2	1
1.1	The Digital Revolution	1
1.2	The Need for Compression	3
1.3	Considerations for Compression	4
1.4	Compression Technology	6
1.5	Videophone and Compact Disk Standards—H.320 and MPEG-1	9
1.6	The Digital Entertainment TV Standard—MPEG-2	9
1.7	Complexity/Cost	12
	References	13
2	Anatomy of MPEG-2	14
2.1	MPEG-1	14
2.2	MPEG-2	16
2.2.1	MPEG-2 Systems	17
2.2.2	MPEG-2 Video	18
2.2.3	MPEG-2 Audio	21
2.2.4	MPEG-2 Conformance	22
2.2.5	Software	25
2.2.6	DSM-CC	25
2.2.7	MPEG-2 NBC Audio	26
2.2.8	10-bit video (cancelled!)	27
2.2.9	MPEG-2 RTI	27
2.2.10	MPEG-2 DSM-CC Conformance	28

2.3	Summary	28
	References	29
3	MPEG-2 Systems	32
3.1	Multiplexing	32
3.2	Synchronization	32
3.3	Elementary Streams	33
3.4	Programs and Nonelementary Streams	37
3.4.1	Program Streams	37
3.4.2	Transport Streams	40
3.4.3	Descriptors	50
3.5	Practical Systems Decoders	52
3.6	Practical Systems Encoders	53
3.7	Conclusion	54
	References	54
4	Audio	55
4.1	Introduction	55
4.2	Audio Compression Fundamentals	56
4.3	MPEG-1	58
4.3.1	Psychoacoustic Model	60
4.3.2	MPEG-1 Syntax and Semantics	62
4.3.3	Layer I Coding	63
4.3.4	Intensity_Stereo Mode	64
4.3.5	Layer II Coding	64
4.3.6	Layer III	67
4.3.7	MS_Stereo Mode	69
4.4	MPEG-2 Audio	71
4.4.1	MPEG-2 Backward Compatible (BC)	71
4.4.2	Subband Groups	72
4.4.3	Dynamic Transmission Channel Switching	72
4.4.4	Dynamic Crosstalk	72
4.4.5	Adaptive Multichannel Prediction	73
4.4.6	Phantom Coding of the Center Channel	73
4.4.7	Multilingual Channels	73
4.4.8	Syntax and Semantics of MPEG-2-BC Audio	73
4.4.9	MPEG-Non Backward Compatible (NBC)	78
	References	79
5	Video Basics	80
5.1	Video Imaging	80
5.2	Visual System and Color TV	86
5.3	From Analog to Digital Video	91
5.4	Spatial Conversions	95
5.5	Temporal Conversions	102
5.6	Mixing and Keying	105
5.7	Summary	107
	References	108

6	Digital Compression: Fundamentals	110
6.1	Limitations of the Television Camera	110
6.2	Numerical Representation	110
6.3	Basic Compression Techniques	116
6.4	Predictive Coding (DPCM)	117
6.5	Block Transform Coding	132
6.6	Hybrid Transform Coding	136
6.7	Vector Quantization	138
6.8	Motion Adaptive Interpolation	139
6.9	Incorporation of Perceptual Factors	140
	References	145
7	Motion Compensation Modes in MPEG	146
7.1	Coding of Progressive Scanned Video	146
7.2	Coding of Interlaced Video	149
7.3	Choosing the Mode of Motion Compensation	154
7.4	Coding the Motion Vectors	154
	References	155
8	MPEG-2 Video Coding and Compression	156
8.1	MPEG-2 Chrominance Sampling	156
8.2	Requirements of the MPEG-2 Video Standard	157
8.3	Main Profile Algorithm Overview	158
8.4	Overview of the MPEG-2 Video Stream Syntax	161
8.5	Typical Encoder Architecture	170
8.6	Typical Decoder Architecture	173
8.7	MPEG-2 Error Concealment	174
8.8	Performance of MPEG-2	175
	References	181
9	MPEG-2 Scalability Techniques	183
9.1	Scalability Basics	184
9.2	Data Partitioning	185
9.3	SNR Scalability	190
9.4	Spatial Scalability	197
9.5	Temporal Scalability	213
9.6	Hybrid Scalability	224
9.7	Scalability Summary	228
	References	229
10	Video Stream Syntax and Semantics	230
10.1	Video Stream Syntax	230
10.2	Semantics for the Syntax Variables	231
	References	257
11	Requirements and Profiles	258
11.1	MPEG-2 Requirements	258
11.2	The Development Process	260

11.3	Basic Profiles Hierarchy	261
11.4	Basic Profiles Interoperability	268
11.5	Basic Profiles Verification Tests	271
11.6	New Profiles	274
11.7	New Profiles Interoperability	276
11.8	Decoders and Bitstream Compatibility	277
11.9	Summary	277
	References	278
12	Digital Networks	280
12.1	Current Methods of Access ¹⁻⁵	282
12.2	New Broadband Wired Access Scenarios ⁶⁻¹²	283
12.3	Wired Access Compared	285
12.4	Terrestrial Broadcasting ¹²	286
12.5	Direct Broadcast Satellites ¹⁴	289
	References	291
13	Interactive Television	292
13.1	Customer Premises Equipment	294
13.2	Interactive Television Servers	296
13.3	User Interfaces ¹⁰⁻¹²	302
13.4	Content Creation for ITV ¹⁴⁻¹⁹	304
	References	306
14	High Definition Television (HDTV)	307
14.1	Characteristics of HDTV	308
14.2	Problems of Current Analog TV	310
14.3	Grand Alliance HDTV System	311
14.4	Performance of MPEG-2 on HDTV Pictures	316
14.5	Evolution of HDTV	320
	References	320
15	Three-Dimensional TV	322
15.1	Basics of 3D	323
15.2	Techniques for 3D Viewing	327
15.3	Stereoscopic Imaging and Display	334
15.4	Stereoscopic Geometry and Human Factors	338
15.5	Video Compression for Stereoscopic 3DTV	344
15.6	Multiviewpoint Video	353
15.7	Summary	358
	References	358
16	Processing Architecture and Implementation	361
	Dr. Horng-Dar Lin	
16.1	MPEG Processing Requirements	361
16.2	Architectural Choices and Examples	362
16.3	Design Issues	366
16.4	Directions in MPEG Chip Development	367

16.5	Summary	367
	References	368
17	MPEG-4 and the Future	369
17.1	H.263: The ITU-T Low-Bitrate Standard	370
17.2	Coding Advances	382
17.3	Introduction to MPEG-4	387
17.4	MPEG-4 Requirements	388
17.5	MPEG-4 Video	391
17.6	MPEG-4 Audio	399
17.7	MPEG-4 Syntax Description Language (MSDL)	399
17.8	Synthetic and Natural Hybrid Coding (SNHC)	405
17.9	The Future	408
17.10	Summary	410
	References	410
	Appendix A: Resource Guide	413
	Appendix B: MPEG Committee	422
	Index	423

Anatomy of MPEG-2

After many fits and starts, the era of Digital Television²⁷ is finally here. Although a variety of audio and video coding standards predate MPEG standardization activity, the MPEG-1 and the MPEG-2 standards are truly unique integrated audio-visual standards.^{25,28}

Neither the MPEG-1 nor the MPEG-2 standards prescribe which encoding methods to use, the encoding process, or details of encoders. These standards only specify formats for representing data input to the decoder, and a set of rules for interpreting these data. These formats for representing data are referred to as *syntax* and can be used to construct various kinds of valid data streams referred to as *bitstreams*. The rules for interpreting the data are called *decoding semantics*; An ordered set of decoding semantics is referred to as a *decoding process*. Thus, we can say that MPEG standards specify a decoding process; However, this is still different than specifying a decoder implementation.

Given audio or video data to be compressed, an encoder must follow an ordered set of steps called the *encoding process*. This encoding process, however, is not standardized and typically varies, since encoders of different complexities may be used in different applications. Also, since the encoder is not standardized, continuing improvements in quality are still possible because of encoder opti-

mizations even after the standard is complete. The only constraint is that the output of the encoding process result in a syntactically correct bitstream that can be interpreted by following the decoding semantics by a standards compliant decoder.

2.1 MPEG-1

Although we plan to discuss the anatomy of MPEG-2, we first need to understand the anatomy of MPEG-1. This will allow us to make comparisons between the standards to clearly show similarities and differences, and wherever appropriate, compatibilities between the two standards.

The MPEG-1 standard is formally referred to as ISO 11172 and consists of the following parts:

- 11172-1: Systems
- 11172-2: Video
- 11172-3: Audio
- 11172-4: Conformance
- 11172-5: Software

Initially, the MPEG-1 standard was intended for video coding at bitrates of about 1.2 Mbit/s with stereo audio coding at bitrates of around 250 kbits/s.

2.1.1 MPEG-1 Systems

The MPEG-1 Systems¹ part specifies a system coding layer for combining coded audio and video data and to provide the capability for combining with it user-defined private data streams as well as streams that may be specified at a later time. To be more specific,¹² the MPEG-1 Systems standard defines a packet structure for multiplexing coded audio and video data into one stream and keeping it synchronized. It thus supports multiplexing of multiple coded audio and video streams where each stream is referred to as an *elementary stream*. The systems syntax includes data fields to assist in parsing the multiplexed stream after random access and to allow synchronization of elementary streams, management of decoder buffers containing coded data, and identification of timing information of coded program. The MPEG-1 Systems thus specifies syntax to allow generation of systems bitstreams and semantics for decoding these bitstreams.

Since the basic concepts of timing employed in MPEG-1 Systems are also common to MPEG-2 Systems, we briefly introduce the terminology employed. The Systems Time Clock (STC) is the reference time base, which operates at 90 kHz, and may or may not be phase locked to individual audio or video sample clocks. It produces 33-bit time representations that are incremented at 90 kHz. In MPEG Systems, the mechanism for generating the timing information from decoded data is provided by the *Systems Clock Reference* (SCR) fields that indicate the current STC time and appear intermittently in the bitstream spaced no further than 700 ms apart. The presentation playback or display synchronization information is provided by *Presentation Time Stamps* (PTSs), that represent the intended time of presentation of decoded video pictures or audio frames. The audio or video PTS are sampled to an accuracy of 33 bits.

To ensure guaranteed decoder buffer behavior, MPEG Systems employs a *Systems Target Decoder* (STD) and *Decoding Time Stamp* (DTS). The DTS differs from PTS only in the case of video pictures that require additional reordering delay during the decoding process.

2.1.2 MPEG-1 Video

The MPEG-1 Video² standard was originally aimed at coding of video of SIF resolution (352×240 at 30 noninterlaced frames/s or 352×288 at 25 noninterlaced frames/s) at bitrates of about 1.2 Mbit/s. In reality it also allows much larger picture sizes and correspondingly higher bitrates. Besides the issue of bitrates, the other significant issue is that MPEG includes functions to support interactivity such as fast forward (FF), fast reverse (FR), and random access into the stored bitstream. These functions exist since MPEG-1 was originally aimed at digital storage media such as video compact discs (CDs).

The MPEG-1 Video standard basically specifies the video bitstream syntax and the corresponding video decoding process. The MPEG-1 syntax supports encoding methods^{13,14,15} that exploit both the spatial redundancies and temporal redundancies. Spatial redundancies are exploited by using block-based Discrete Cosine Transform (DCT) coding of 8×8 pel blocks followed by quantization, zigzag scan, and variable length coding of runs of zero quantized indices and amplitudes of these indices. Moreover, quantization matrix allowing perceptually weighted quantization of DCT coefficients can be used to discard perceptually irrelevant information, thus increasing the coding efficiency further. Temporal redundancies are exploited by using motion compensated prediction, which results in a significant reduction of interframe prediction error.

The MPEG-1 Video syntax supports three types of coded pictures, Intra (I) pictures, coded separately by themselves, Predictive (P) pictures, coded with respect to immediately previous I- or P-pictures, and Bidirectionally Predictive (B) pictures coded with respect to the immediate previous I- or P-picture, as well as the immediate next I- or P-picture. In terms of coding order, P-pictures are causal, whereas B-pictures are noncausal and use two surrounding causally coded pictures for prediction. In terms of compression efficiency, I-pictures are least efficient, P-pictures are somewhat better, and B-pictures are the most efficient. In typical MPEG-1 encoding^{13,15} an input video sequence is divided

into units of group-of-pictures (GOPs), where each GOP consists of an arrangement of one I-picture, P-pictures, and B-pictures. A GOP serves as a basic access unit, with the I-picture serving as the entry point to facilitate random access. Each picture is divided further into one or more *slices* that offer a mechanism for resynchronization and thus limit the propagation of errors. Each slice is composed of a number of macroblocks; each macroblock is basically a 16×16 block of luminance pels (or alternatively, four 8×8 blocks) with corresponding chrominance blocks. MPEG-1 Video encoding is performed on a macroblock basis. In P-pictures each macroblock can have one motion vector, whereas in B-pictures each macroblock can have as many as two motion vectors.

2.1.3 MPEG-1 Audio

The MPEG-1 Audio³ standard basically specifies the audio bitstream syntax and the corresponding audio decoding process. MPEG-1 Audio is a generic standard and does not make any assumptions about the nature of the audio source, unlike some vocal-tract-model coders that work well for speech only. The MPEG syntax permits exploitation of perceptual limitations of the human auditory system while encoding, and thus much of the compression comes from removal of perceptually irrelevant parts of the audio signal. MPEG-1 Audio coding^{15,23} allows a diverse assortment of compression modes and in addition supports features such as random access, audio fast forwarding, and audio fast reverse.

The MPEG-1 Audio standard consists of three Layers—I, II, and III. These Layers also represent increasing complexity, delay, and coding efficiency. In terms of coding methods, these Layers are somewhat related, since a higher Layer includes the building blocks used for a lower Layer. The sampling rates supported by MPEG-1 Audio are 32, 44.2, and 48 kHz. Several fixed bitrates in the range of 32 to 224 kbits/s per channel can be used. In addition, Layer III also supports variable bitrate coding. Good quality is possible with Layer I above 128 kbits/s, with Layer II around 128 kbit/s, and with Layer III at 64 kbit/s. MPEG-1 Audio has four modes: mono, stereo, dual with two separate

channels, and joint stereo. The optional joint stereo mode exploits interchannel redundancies.

A polyphase filter bank is common to all Layers of MPEG-1 Audio coding. This filter bank subdivides the audio signal into 32 equal-width frequency subbands. The filters are relatively simple and provide good time-resolution with a reasonable frequency-resolution. To achieve these characteristics, some exceptions had to be made. First, the equal widths of subbands do not precisely reflect the human auditory system's frequency-dependent behavior. Second, the filter bank and its inverse are not completely lossless transformations. Third, adjacent filter bands have significant frequency overlap. However, these exceptions do not impose any noticeable limitations, and quite good audio quality is possible. The Layer I algorithm codes audio in frames of 384 samples by grouping 12 samples from each of the 32 subbands. The Layer II algorithm is a straightforward enhancement of Layer I; it codes audio data in larger groups (1 152 samples per audio channel) and imposes some restrictions on possible bit allocations for values from the middle and higher subbands. The Layer II coder gets better quality by redistributing bits to better represent quantized subband values. The Layer III algorithm is more sophisticated and uses audio spectral perceptual entropy coding and optimal coding in the frequency domain. Although based on the same filter bank as used in Layer I and Layer II it compensates for some deficiencies by processing the filter outputs with a modified DCT.

2.2 MPEG-2

After this brief overview of the parts of the MPEG-1 standard, we are now well prepared to delve into a discussion of the various parts of the MPEG-2 standard. This standard is formally referred to as ISO 13818 and consists of the following parts:

- 113818-1: Systems
- 113818-2: Video
- 113818-3: Audio
- 113818-4: Conformance
- 113818-5: Software

- 113818-6: Digital Storage Media—Command and Control (DSM-CC)
- 113818-7: Non Backward Compatible (NBC) Audio
- 113818-8: 10-Bit Video (This work item has been dropped!)
- 113818-9: Real Time Interface
- 113818-10: Digital Storage Media—Command and Control (DSM-CC) Conformance

We now present a brief overview of the aforementioned parts of the MPEG-2 standard.

2.2.1 MPEG-2 Systems

Since the MPEG-1 standard was intended for audio-visual coding for Digital Storage Media (DSM) applications and since DSMs typically have very low or negligible transmission bit error rates, the MPEG-1 Systems part was not designed to be highly robust to bit errors. Also, the MPEG-1 Systems was intended for software oriented processing, and thus large variable length packets were preferred to minimize software overhead.

The MPEG-2 standard on the other hand is more generic and thus intended for a variety of audio-visual coding applications. The MPEG-2 Systems⁴ was mandated to improve error resilience plus the ability to carry multiple programs simultaneously without requiring them to have a common time base. Additionally, it was required that MPEG-2 Systems should support ATM networks. Furthermore, MPEG-2 Systems was required to solve the problems addressed by MPEG-1 Systems and be somewhat compatible with it.

The MPEG-2 Systems^{17,18} defines two types of streams: the Program Stream and the Transport Stream. The Program Stream is similar to the MPEG-1 Systems stream, but uses a modified syntax and new functions to support advanced functionalities. Further, it provides compatibility with MPEG-1 Systems streams. The requirements of MPEG-2 Program Stream decoders are similar to those of MPEG-1 System Stream decoders. Furthermore, Program Stream decoders are expected to be forward compatible with MPEG-1 System Stream decoders, i.e., capable of decoding MPEG-1 System Streams. Like MPEG-1 Systems

decoders, Program streams decoders typically employ long and variable-length packets. Such packets are well suited for software based processing and error-free environments, such as when the compressed data are stored on a disk. The packet sizes are usually in range of 1 to 2 kbytes chosen to match disk sector sizes (typically 2 kbytes); however, packet sizes as large as 64 kbytes are also supported. The Program Stream includes features not supported by MPEG-1 Systems such as scrambling of data; assignment of different priorities to packets; information to assist alignment of elementary stream packets; indication of copyright; indication of fast forward, fast reverse, and other trick modes for storage devices; an optional field for network performance testing; and optional numbering of sequence of packets.

The second type of stream supported by MPEG-2 Systems is the Transport Stream, which differs significantly from MPEG-1 Systems as well as the Program Stream. The Transport Stream offers robustness necessary for noisy channels as well as the ability to include multiple programs in a single stream. The Transport Stream uses fixed length packets of size 188 bytes, with a new header syntax. It is therefore more suited for hardware processing and for error correction schemes. Thus the Transport Stream stream is well suited for delivering compressed video and audio over error-prone channels such as coaxial cable television networks and satellite transponders. Furthermore, multiple programs with independent time bases can be multiplexed in one Transport Stream. In fact the Transport Stream is designed to support many functions such as asynchronous multiplexing of programs, fast access to desired program for channel hopping, multiplex of programs with clocks unrelated to transport clock, and correct synchronization of elementary streams for playback. It also allows control of decoder buffers during startup and playback for both constant bitrates and variable bitrate programs.

A basic data structure that is common to the organization of both the Program Stream and Transport Stream data is called the Packetized Elementary Stream (PES) packet PES packets are generated by packetizing the continuous streams of compressed data generated by video and audio

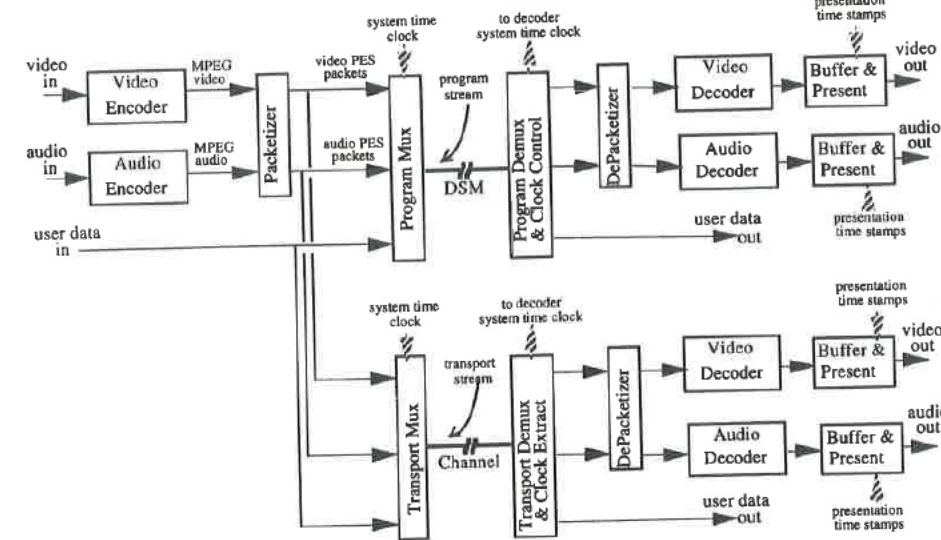


Fig. 2.1 MPEG-2 Systems Multiplex showing Program and Transport Streams.

(i.e., elementary stream) encoders. A Program Stream is generated simply by stringing together PES packets from the various encoders with other packets containing necessary data to generate a single bitstream. A Transport Stream consists of packets of fixed length consisting of 4 bytes of header followed by 184 bytes of data, where data is obtained by chopping up data in PES packets.

In Fig. 2.1 we illustrate both types of MPEG-2 Systems, the ones using Program Stream multiplex and the ones using Transport Stream multiplex. An MPEG-2 System is also capable of combining multiple sources of user data along with the MPEG encoded audio and video. The audio and video streams are packetized to form audio and video PES packets that are then sent to either a Program Multiplexer or a Transport Multiplexer, resulting in a Program Stream or Transport Stream, respectively. As mentioned earlier, Program Streams are intended for error-free environments such as DSMs, whereas Transport Streams are intended for noisier environments such as terrestrial broadcast channels.

Transport Streams are decoded by a Transport Demultiplexer (which includes a clock extraction mechanism), unpackitized by a DePacketizer, and

sent for audio and video decoding to Audio and Video Decoders. The decoded signals are sent to respective Buffer and Presentation units that output them to a display device and speaker at the appropriate times. Similarly, if Program Streams are employed, they are decoded by a Program Stream Demultiplexer, DePacketizer, and sent for decoding to Audio and Video Decoders. The decoded signals are sent to respective Buffer and Present and await presentation. Also, as with MPEG-1 Systems, the information about systems timing is carried by Clock Reference fields in the bitstream that are used to synchronize the decoder Systems Time Clock (STC). The presentation of decoded output is controlled by Presentation Time Stamps (PTSs) that are also carried by the bitstream.

2.2.2 MPEG-2 Video

This part of MPEG-2, part 2, addresses coding of video and is referred to as MPEG-2 Video. Originally the MPEG-2 Video⁵ standard was primarily intended for coding of interlaced video of standard TV resolution with good quality in the bitrate range of 4 to 9 Mbit/s. However, the scope

of MPEG-2 Video was considerably revised²⁰ to include video of higher resolutions such as HDTV at higher bitrates as well as hierarchical video coding for a range of applications. Furthermore, it also supports coding of interlaced video of a variety of resolutions.

Actually, MPEG-2 Video does not standardize video encoding; only the video bitstream syntax and decoding semantics are standardized. The standardization process for MPEG-2 Video consisted of a competitive phase followed by a collaborative phase. First an initial set of requirements for MPEG-2 video was defined and used as the basis for a "Call for Proposals" which invited candidate coding schemes for standardization. These schemes underwent formal subjective testing and analysis to determine if they met the listed requirements. During the collaborative phase, the best elements of the top performing schemes in the competitive phase were merged to form an initial *Test Model*. A Test Model contains a description of an encoder, bitstream syntax, and decoding semantics. The encoder description is not to be standardized, but only needed for experimental evaluation of proposed techniques.

Following this, a set of *Core Experiments* were defined that underwent several iterations until convergence was achieved. During the time that various iterations on the Test Model¹⁹ were being performed, the requirements for MPEG-2 Video also underwent further iterations.

In Fig. 2.2 we show a simplified codec^{19,21} for MPEG-2 Nonscalable Video Coding. The MPEG-2 Video Encoder shown for illustration purposes

consists of an Inter frame/field DCT Encoder, a Frame/field Motion Estimator and Compensator, and a Variable Length Encoder (VLE). Earlier we had mentioned that MPEG-2 Video is optimized for coding of interlaced video, which is why both the DCT coding and the motion estimation and compensation employed by the Video Encoder are frame/field adaptive. The Frame/field DCT Encoder exploits spatial redundancies, and the Frame/field Motion Compensator exploits temporal redundancies in interlaced video signal. The coded video bitstream is sent to a systems multiplexer, Sys Mux, which outputs either a Transport or a Program Stream.

The MPEG-2 Decoder in this codec consists of a Variable Length Decoder (VLD), Inter frame/field DCT Decoder, and the Frame/field Motion Compensator. Sys Demux performs the complementary function of Sys Mux and presents the video bitstream to VLD for decoding of motion vectors and DCT coefficients. The Frame/field Motion Compensator uses a motion vector decoded by VLD to generate motion compensated prediction that is added back to a decoded prediction error signal to generate decoded video out. This type of coding produces video bitstreams called nonscalable, since normally the full spatial and temporal resolution coded is the one that is expected to be decoded. Actually, this is not quite true, since, if B-pictures are used in encoding, because they do not feedback into the interframe coding loop, it is always possible to decode some of them, thus achieving temporal resolution scaling even for non-scalable bitstreams. However, by non-scalable cod-

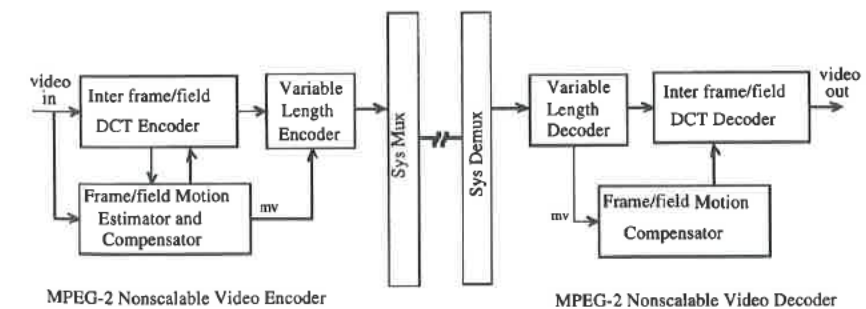


Fig. 2.2 A generalized codec for MPEG-2 Nonscalable Video Coding.

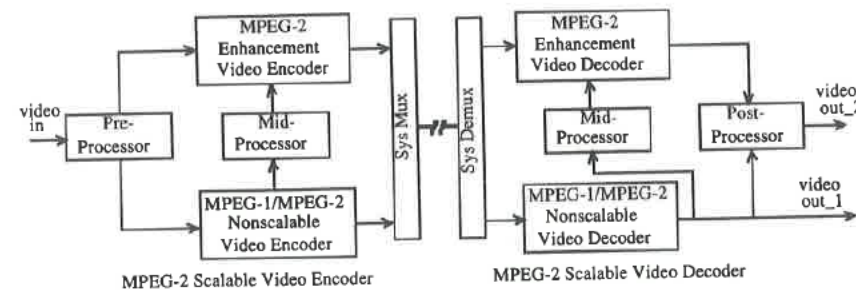


Fig. 2.3 A generalized codec for MPEG-2 Scalable Video Coding.

ing we usually mean that we have not gone out of our way to facilitate scalability.

As you may have guessed by this time, the MPEG-2 Video standard specifies the bitstream syntax and decoding process not only for single layer (MPEG-1 like) video coding but also for scalable video coding. Scalability is the property that allows decoders of various complexities to be able to decode video of resolution/quality commensurate with their complexity from the same bitstream. We will discuss a lot more about MPEG-2 scalable video coding in Chapter 9, but for now we illustrate the general principle by introducing the generalized scalable codec structure shown in Fig. 2.3. In MPEG-2 Video there are many different types of scalability and thus it is difficult to represent all of them with a single generalized codec. Our generalized codec structure basically implies spatial and temporal resolution scalability.

Input video goes through a Pre-Processor that produces two video signals, one of which (the *Base Layer*) is input to an MPEG-1 or MPEG-2 Nonscalable Video Encoder and the other (the *Enhancement Layer*) input to an MPEG-2 Enhancement Video Encoder. Some processing of decoded video from MPEG-1 or MPEG-2 Nonscalable Video Encoder may be needed in the Mid-Processor depending on specific scalability. The two bitstreams, one from each encoder, are multiplexed in Sys Mux (along with coded audio and user data). Thus it becomes possible for two types of decoders to be able to decode a video signal of quality commensurate with their complexity, from the same encoded bitstream. For example, if an MPEG-1 or MPEG-2 Nonscalable Video Decoder is employed, a basic

video signal can be decoded. If in addition, an MPEG-2 Enhancement Video Decoder is employed, an enhanced video signal can be decoded. The two decoded signals may undergo further processing in a Post-Processor.

Since the MPEG-2 Video standard, to be truly generic, had to include a bitstream and semantics description for a variety of coding methods and tools, the implementation of all its features in every decoder was considered too complex and thus the standard was partitioned into Profiles. In MPEG-2 Video, applications with somewhat related requirements are addressed via Profiles; a Profile typically contains a collection of coding techniques (or tools) designed to address a set of such applications. We will discuss a lot more about Requirements and Profiles in Chapter 11.

The MPEG-2 Video standard is a syntactic superset of the MPEG-1 Video standard and is thus able to meet the requirement of forward compatibility, meaning that an MPEG-2 video decoder should be capable of decoding MPEG-1 video bitstreams. The requirement of backward compatibility, meaning that subsets of MPEG-2 bitstreams should be decodable by existing MPEG-1 decoders, is achieved via the use of scalability and supported in specific profiles. To confirm that the MPEG-2 Video standard met its quality objective for various Profiles and for individual applications within a profile, a series of verification tests have been conducted in association with other international standardization bodies and a number of organizations. Thus far, these tests confirm that MPEG-2 Video does indeed meet or exceed the performance bounds of its target applications. MPEG-2 Video

quality has also been judged sufficient for HDTV.³⁴ Recently, newer applications of MPEG-2 Video have also emerged, necessitating combinations of coding tools requiring new Profiles.

At the time of writing of this chapter in early 1996, two new amendments in MPEG-2 video, each adding a new profile, are in progress. One of the two amendments is nearly complete, and the second one is expected to be completed by the end of 1996. Although MPEG-2 video already includes various coding techniques (decoding techniques, really), in reality, these techniques must be included in one of the agreed Profiles also (these profiles can even be defined after completion of the standard). Next, we briefly introduce what application areas are being targeted by these two amendments.

The first amendment to MPEG-2 Video involves a Profile that includes tools for coding of a higher resolution chrominance format called the 4:2:2 format. We will discuss more about this format in Chapter 5; for now, we need to understand why it was relevant for MPEG-2 video to address applications related to this format. When work on MPEG-2 Video was close to completion, it was felt that MPEG-2 video was capable of delivering fairly high quality, and thus it could be used in professional video applications. In sorting out needs of such applications, besides higher bitrates, it was found that some professional applications require higher than normal chrominance spatial (or spatiotemporal) resolution, such as that provided by the 4:2:2 format while another set of applications required higher amplitude resolution for luminance and chrominance signal. Although coding of the 4:2:2 format video is supported by existing tools in the MPEG-2 Video standard, it was necessary to verify that its performance was equal to or better than that of other coding schemes while fulfilling additional requirements such as coding quality after multiple codings and decodings.

The second amendment³² to MPEG-2 Video involves a Profile that includes tools for coding of a number of simultaneous video signals generated in imaging a scene, each representing a slightly different viewpoint; imaging of such scenes is said to constitute multiviewpoint video. An example of multiviewpoint video is stereoscopic video which

uses two slightly different views of a scene. Another example is generalized 3D video where a large number of views of a scene may be used. We will discuss more about this profile in Chapter 5 and a lot more about 3D/stereoscopic video in Chapter 15; for now, as in the case of 4:2:2 coding we need to understand why it was relevant to address multiviewpoint coding applications. As mentioned earlier, MPEG-2 Video already includes several Scalability techniques that allow layered coding which exploits correlations between layers. As the result of a recent rise in applications in video games, movies, education, and entertainment, efficient coding of multiviewpoint video such as stereoscopic video is becoming increasingly important. Not surprisingly, this involves exploiting correlations between different views, and since scalability techniques of MPEG-2 Video already included such techniques, it was considered relatively straightforward to enable use of such techniques via definition of a Profile (which is currently in progress).

2.2.3 MPEG-2 Audio

Digital multichannel audio systems employ a combination of p front and q back channels; for example, three front channels—left, right, center—and two back channels—surround left and surround right—to create surreal experiences. In addition, multichannel systems can also be used to provide multilingual programs, audio augmentation for visually impaired individuals, enhanced audio for hearing impaired individuals, and so forth.

The MPEG-2 Audio⁶ standard consists of two parts, one that allows coding of multichannel audio signals in a forward and backward compatible manner with MPEG-1 and one that does not. Part 3 of the MPEG-2 standard is forward and backward compatible with MPEG-1. Here forward compatibility means that the MPEG-2 multichannel audio decoder can decode MPEG-1 mono or stereo audio signals. Backward compatibility (BC) means that an MPEG-1 stereo decoder can reproduce a meaningful downmix of the original five channels from the MPEG-2 audio bitstream.

While forward compatibility is not as hard to achieve, backward compatibility is a much more

DIGITAL VIDEO: AN INTRODUCTION TO MPEG-2

Digital Video offers comprehensive coverage of the MPEG-2 audio/visual digital compression standard. The treatment includes the specifics needed to implement an MPEG-2 *Decoder*, including the syntax and semantics of the coded bitstreams. Since the MPEG-2 *Encoders* are not specified by the standard, and are actually closely held secrets of many vendors, the book only outlines the fundamentals of encoder design and algorithm optimization.

Building on its coverage of the fundamentals in the opening chapters, **Digital Video** proceeds to explore specific applications, including digital video delivery media, interactive television, high-definition television, three-dimensional stereoscopic television, and VLSI implementation architectures. The book concludes with a progress report on MPEG-4, keeping readers abreast of this fast-changing field.

This outstanding volume will be essential to all those involved in audio/visual digital compression. Engineers in communications, computers, and electrical industries will all find an abundance of practical information unavailable from other sources.

ABOUT THE AUTHORS

Barry G. Haskell, Atul Puri, and Arun N. Netravalli are all at AT&T Bell Labs in Holmdel, New Jersey.



Chapman & Hall
115 Fifth Avenue, New York, NY 10003
2-6 Boundary Row, London SE1 8HN
Printed in U.S.A.

JOIN US ON THE INTERNET <http://www.thomson.com>

Visit Chapman & Hall's Internet Resource Center for information on our new publications, links to useful sites on the World Wide Web and an opportunity to join our e-mail mailing list.
Point your browser to: <http://www.chaphall.com/chaphall.html> or
<http://www.chaphall.com/chaphall/electeng.html> for Electrical Engineering

A service of **ITP**



9 780412 084119