

ITU/SMPTE TUTORIAL DIGITAL TERRESTRIAL TELEVISION BROADCASTING (DTTB)

Edited by Stanley N. Baron

Published by The Society of Motion Picture and Television Engineers

SONY 1008

The Society of Motion Picture and Television Engineers, White
Plains, NY 10607

Copyright © 1994 by the Society of Motion Picture and Television
Engineers, Inc. All rights reserved.

This book was reproduced photographically from original copy submitted by the authors. Statements and opinions contained herein are expressions of the contributors and do not necessarily represent the policies or opinions of the SMPTE.

No part of this book may be reproduced without written permission of the publisher.

Printed in the United States of America

Library of Congress Catalog Card Number: 94-65591

ISBN 0-940690-24-1

Cover Design: Mathew V. Kuriakose

**ITU/SMPTE TUTORIAL
DIGITAL TERRESTRIAL TELEVISION BROADCASTING
(DTTB)**

CONTENTS

- vii *Preface*, Kenneth P. Davies, SMPTE Engineering Vice President (CBC, Canada)
- viii *Introduction*, Richard C. Kirby, Director and Richard L. Nickelson, Senior Counsellor, ITU Radiocommunication Bureau

Source Coding and Service Multiplex

- 3 *An Overview of the DTTB Model*, S. N. Baron (NBC, U.S.A.)
- 8 *Video Compression Techniques and Multilevel Approaches*, Marzio Barbero and Mario Stroppiana (Radiotelevisione Italia)
- 28 *MPEG Overview*, Stan Baron (NBC, U.S.A.) and W. Robin Wilson (Grass Valley Group, U.S.A.)
- 37 *The Service Multiplex*, Gary Tonge (ITC, England)
- 43 *Encipherment and Conditional Access*, Louis Claude Guillou and Jean-Luc Giachetti (CCETT, France)

Planning Factors and Implementation Strategies

- 57 *Planning Factors and Their Influence on System Aspects*, Jorgen Weber (Telecom, Denmark)
- 69 *Digital Television Broadcasting: Issues for Successful Implementation*, M.D. Windram and A.G. Mason (NTL, England)
- 83 *Digital Television Broadcasting — A Swedish Perspective*, Per Appelquist (Sveriges Television AB, Stockholm)
- 85 *Digital Terrestrial Television Broadcasting*, Keith G. Malcolm (Department of Transport and Communications, Australia)
- 87 *Receiver Characteristics*, M. Artigas and D. Westerkamp (Thomson Consumer Electronics, France)

Channel Coding and Modulation Technology

- 97** *Modulation and Channel Coding for ATV Terrestrial Transmission*, Yiyan Wu, (Communications Research Centre, Canada)
- 107** *Error Management in Digital Terrestrial Television Broadcasting*, Yuichi Ninomiya (NHK, Japan)
- 129** *Channel Coding Approaches and Consequences — Single and Dual Carriers*, Glenn Reitmeier, Richard J. Klensch, and Hugh E. White (David Sarnoff Research Center, U.S.A.)
- 145** *Digital Terrestrial Television Using the COFDM Technique*, Bertrand Sueur, Damien Castelain, Gabriel Degoulet, Marc Rivière, and Bernard Le Floch (CCETT, France)

Display Requirements and Interoperability

- 155** *Displays and Colorimetry for Future Television*, L. DeMarsh (U.S.A.)
- 167** *Digital Television System Scalability and Interoperability*, Stanley Baron (NBC, U.S.A.)

MPEG¹ OVERVIEW

Stan Baron (NBC) and W. Robin Wilson (Grass Valley Group)

Introduction

The MPEG standard address the compression and decompression (restoration) of moving pictures (video) and sound (audio) and the formation of a multiplexed common data stream that includes the compressed video and audio data plus any associated ancillary service data. The MPEG standard further addresses the synchronization of the video, audio, and ancillary data during playback of the decompressed signals.

The MPEG standard is intended to serve a wide variety of applications and services such as digital storage media, television broadcast and communications.

The MPEG-2 standard is an extremely flexible standard. It is a set of defined compression and systemization algorithms and techniques with well defined rules and guidelines. The rules and guidelines allow variation in the values assigned to many of the parameters providing for a broad range of products and interoperability. This set of algorithms and techniques is has been integrated into a single MPEG "*tool-kit*" or *syntax*.

MPEG-2 Levels and Profiles

The MPEG syntax addresses a variety of cost versus performance grades. These different grades are described in a matrix of **Profiles** and **Levels**. A *profile* is a defined subset of the entire MPEG syntax. A *level* is defined set of constraints imposed on parameters in the bit stream. The constraints may include limits on the values of the parameters such as limits on the structure and output data rate that the encoder can process. A given coder will work at its own (or lower) level and profile. The matrix is described in Table I.

The initial focus of the MPEG-2 standards group was to define a highly efficient audio and video compression scheme mainly for entertainment applications. This initial focus later became defined as the *main profile*, *main level*.

The need for and the relationships between profiles and levels is still a topic of discussion at MPEG meetings. The matrix and many of the relationships may change. It is considered desirable to keep the number and levels of profiles as small as possible and to define Interoperability such that it makes sense in real world applications without burdening an MPEG system

¹ MPEG is understood to reference ISO/IEC Moving Picture Experts Group Draft Standard on "Coded representation of picture, audio, and multimedia/hypermedia information".

with excessive overhead.

Table I

Levels↓/Profiles→	Simple	Main	SNR	Spacial	High
High					
High-1440					
Main					
Low					

	Input	Output
<u>Levels:</u>	<u>Structure</u>	<u>Data Rate</u>
High ⇒	<1920x1152	<80 Mbps
High-1440 ⇒	<1440x1152	<60 Mbps
Main ⇒	<720x576	<15 Mbps
Low ⇒	<352x288	<4 Mbps

<u>Profiles:</u>	<u>Implication:</u>
Simple ⇒	No B-pictures
Main ⇒	No picture, sequence, or temporal scalable extensions
SNR ⇒	Scalable Noise Level (previously denoted as "Main+")
Spatial ⇒	Spatially scalable
High ⇒	4:2:0 or 4:2:2 Chroma

Two (greatly simplified) examples of MPEG profiles would be:

MAIN profile at MAIN level-

- 4:2:0,
- 720 pels/line,
- 480 active lines for 525-line systems,
- 576 active lines for 625-line systems,
- Intra-, predicted-, and bi-directional frames.

SIMPLE profile at MAIN level-

- 4:2:0,
- 720 pels/line,
- 480 active lines for 525-line systems,
- 576 active lines for 625-line systems,

- I-, and P-, but no B-frames.

HDTV Profiles

All of the current MPEG-2 profiles with the exception of the Simple Profile have parameter value ranges appropriate for HDTV operation. The Grand Alliance proposed system for consideration as the HDTV transmission standard for the U.S.A. is most likely to use the MAIN profile at the HI level.

MPEG Data Stream Layers

The MPEG data stream in its most general form is made up of two layers:

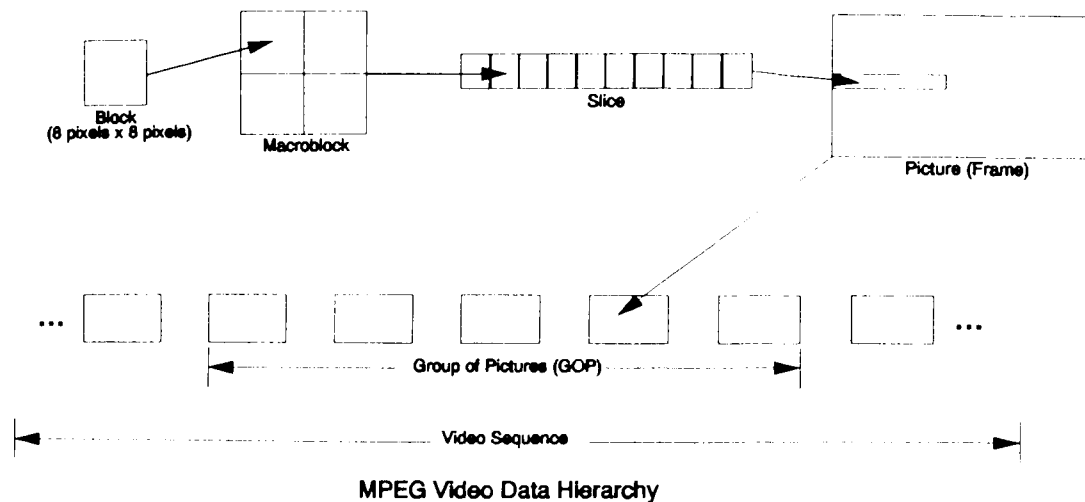


Figure 1

The **system layer** contains timing and other information needed to demultiplex the audio, video, and ancillary data streams and to provide synchronization during playback.

The **compression layer** includes the compressed audio and video data streams.

Video Data Stream

Video Stream Data Hierarchy

Television in its analog representation consists of lines of video and fields and frames of video and is, therefore, sampled in the vertical and temporal dimensions. Transforming the video into its digital representation requires sampling the video in the horizontal dimension. Each picture element or **pixel** is a sample in three dimensions x (horizontal), y (vertical), and t (time). Each pixel consists of

tristimulus color information quantized to an appropriate level of bits, usually of 8 or 10 bit accuracy. The tristimulus representation for television is usually along the R,G,B (Red, Green, Blue) axes or represented by a luminance gray scale value (Y) and two color difference signals $C_R = K_R (R - Y)$ and $C_B = K_B (B - Y)$, where K_B and K_R are scaling coefficients. The color-difference signals may be transported at one-half the horizontal and vertical resolution than the luminance information without noticeable degradation in the reconstructed visual display. MPEG system implementations take advantage of this factor to reduce the amount of data required for processing.

The algorithm uses block-based motion compensation to reduce temporal redundancy. Motion compensation is defined as the use of motion vectors to improve the

efficiency of the prediction of pel values. the prediction uses motion vectors to provide offsets into the past and/or future reference pictures containing previously coded pel values that are used to form the prediction error signal (Reference ISO/IEC 13818-2, §3.59, Nov. 1993). Motion compensation is used for prediction of the current picture from a previous picture and interpolative prediction from past and future pictures. Motion vectors are defined for each 16 pixel by 16 line region of the picture. The prediction error (difference signal) is further compressed using the discrete cosine transform (DCT) to remove spatial correlation before it is quantized in an irreversible process that discards less important information. The motion vectors and residual DCT information are combined and encoded using variable length codes.

The pixels are organized into **Blocks** for the purpose of further processing. A Block normally consists of an array of pixel values or some transform of those picture values which constitute an array of values. A Block represents a set of 8 by 8 values representing the luminance or chrominance information (Reference Figure 1.).

Blocks of information are organized into **Macroblocks**. As defined in ISO/IEC 13818-2, a Macroblock consists of four Blocks of luminance (Y) information (or a 16 pixel by 16 line of values) and a number of chroma (C_R and C_B) blocks. When the number of chroma blocks is two

(one each C_R and C_B blocks) as shown in Figure 2, a 4:2:0 chroma format is assumed. When the number of chroma blocks is four (two each C_R and C_B blocks) a 4:2:2 chroma format is assumed. When the number of chroma blocks is eight (four each C_R and C_B blocks) a 4:4:4 chroma format is assumed. A Macroblock can be used to refer to pel data or to the coded representation of pel data.

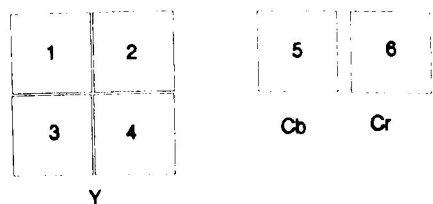
One or more contiguous Macroblocks are grouped together to form **Slices**. The order of the Macroblocks within a slice is the same as the conventional television raster scan being from left to right and from top to bottom.

Slices provide a convenient mechanism for the handling of errors. When an error occurs in the data stream, the decoder can move to the start of the next slice. The number of Slices effects the efficiency of transfer; partitioning the data stream to have more slices provides for better error concealment but uses bits that could otherwise be use to improve picture quality.

The primary coding unit of a video sequence is the individual video **Frame** or **Picture**. A video Picture consists of the group of Slices which constitute the active picture area.

One or more Pictures (frames) in sequence are combined into a **Group of Pictures (GOP)** to allow random access into the sequence and to provide boundaries for inter-picture coding as explained later.

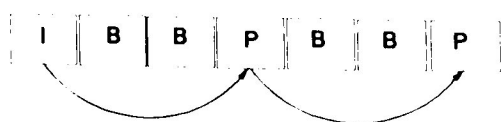
Finally, a **Video Sequence** is represented by a **Sequence Header**, one or more Groups of Pictures, and an **End-of-Sequence** code in the data stream.



Video Macroblock Composition
Figure 2

Inter-picture Coding

The MPEG standard takes advantage



Forward Prediction
Figure 3

of temporal redundancy (the fact that much of the information in a Picture within a video sequence may be similar to the information contained in adjacent Pictures) to represent some Pictures in terms of their differences from a reference Picture.

The MPEG standard defines three types of pictures: *intra*, *predicted*, and *bidirectional*.

Intra coded Pictures or I-pictures are coded using information present only in the picture itself and do not depend on information from other pictures. I-pictures provide a mechanism for random access into the compressed

video data. I-pictures employ transform coding of the Blocks pixels and provide only moderate compression.

Predicted Pictures or P-pictures are coded with respect to the nearest *previous* I- or P-picture. This technique is termed *forward prediction* (see Figure 3). P-pictures provide more compression than I-pictures and serve as a reference for future P-pictures or B-pictures. P-pictures can propagate coding errors when P-pictures (or B-pictures) are predicted from prior P-pictures where the prediction is flawed.

Bidirectional Pictures or B-pictures are pictures that use both future and past pictures as a reference. This



Bidirectional Prediction
Figure 4

technique is termed *bidirectional prediction* (see Figure 4). B-pictures provide the most compression. B-pictures do not propagate coding errors as they are never used as a reference.

Video Data Stream Composition

The MPEG syntax allows the choice of the frequency of occurrence of I-pictures to be selected at the encoder. This allows the system designer to give consideration to the need for random access and the

location of scene cuts in the video sequence. The choice of number of bidirectional pictures between any pair of reference (I- or P-) pictures can also be determined at the encoder. The remaining pictures in the Group of Pictures would be P-pictures. One possible arrangement of I-, P-, and B-pictures is shown in Figure 5.

I B B P B B P B B P B B I B B P B B P B B P B

Two B-pictures between reference pictures

Typical Picture Order
Figure 5

A proper implementation of the MPEG encoder provides for reordering of the Pictures in the Video Data Stream to present the pictures to the decoder in the most efficient sequence. The reference Pictures needed to construct B-Pictures should be sent prior to the associated B-Pictures as shown in Figure 6.

Intra-Picture Transform Coding

The MPEG standard provides for transform coding of I-picture information. The algorithm includes three methods for possible

Display order

I B B P B B P B B P B B

Video stream order

I P B B P B B P B B

Video Data Stream Ordering
Figure 6

compression of the data representing the Picture: discrete cosine transform (DCT), quantization, and run-length coding.

Both image blocks and the prediction-error blocks have high spatial redundancy. To reduce the redundancy, the data in a Block (matrix) representing the pixel information or the error information are converted (transformed) to the frequency domain using the *Discrete Cosine Transform (DCT)*. The result is a Block (matrix) of equal size where the data represents the coefficients of the frequency components. In an information set with high spatial redundancy, many of the frequency coefficient terms may be zero.

Next, the process *quantizes* the frequency coefficients. This process approximates each frequency coefficient as one of a limited number of allowed values. The encoder chooses a quantization matrix that determines how each frequency coefficient in the transformed Block is quantized. Human perceptual

sensitivity of quantization errors are lower for the higher spatial frequencies. Therefore, high frequencies are quantized more coarsely with fewer allowed values than low frequencies. The (0,0) component in the matrix represents the DC value of the Block and is usually carried at full precision.

Some Blocks of pixels need to be coded more accurately than others, such as Blocks with smooth intensity gradients which need accurate coding to avoid visible Block boundaries. The MPEG standard allows the amount of quantization to be modified for each Block.

This combination of DCT processing and quantization results in many of the frequency coefficients being zero. This is particularly so of the higher order spatial frequencies. The coefficients are then organized in a zigzag order to take advantage of this. The result shown in Figure 7 tends to produce long runs of zeroes. The coefficients are then converted to a series of run-amplitude pairs, each pair indicating a number of zero coefficients and the amplitude of a non-zero coefficient. These run-amplitude pairs are then coded with a *variable-length* code which uses shorter codes for commonly occurring pairs and longer codes for less common pairs.

Motion Compensation

Motion compensation is a technique applied to the compression of P- and B-pictures that improves the

compression factor by a factor of three compared to Intra-picture coding by eliminating temporal redundancy. Motion compensation algorithms are applied at the Macroblock level.

Motion compensation applied to P-picture Macroblocks produces two sets of information:

Motion vectors representing the difference between the reference and the Macroblock being coded, and

Error terms representing the differences between the predicted values and the actual results.

When a Macroblock in a P-picture cannot be portrayed using Motion compensation, such as might occur if some unknown object became revealed (uncovered) in a scene, then the Macroblock is coded in same manner as a Macroblock in an I-picture using spatial transform coding techniques.

Motion compensation applied to N-picture Macroblocks can be coded using either a previous or future reference picture. One of four possible codings can result:

Intra-coding, no motion compensation data;

Forward prediction, in which the closest prior I- or P-picture serves as the reference;

Backward prediction, in which the closest future I- or P-picture serves

as the reference; or

Bidirectional prediction, in which two pictures serve as the reference, one being the closest prior I- or P-picture and the other being the closest future I- or P-picture.

Backward prediction can be used to predict uncovered areas that do not appear in the prior pictures.

Synchronization

The MPEG standard provides a timing mechanism that ensures synchronization of audio, video, and associated ancillary data such as captioning. The two parameters are the *system clock reference (SCR)* and the *presentation time stamp (PTS)*.

System Clock Reference

The system clock reference is a snapshot of the encoder system clock. The SCR's used by the audio, video, and associated ancillary data in the decoder must have approximately the same value. To keep the values in agreement, SCRs are inserted into the MPEG stream on a cyclic basis by the encoder and are extracted by the system decoder. The audio and video decoders and the associated ancillary data processors update their internal clocks using the SCR value sent by the system decoder.

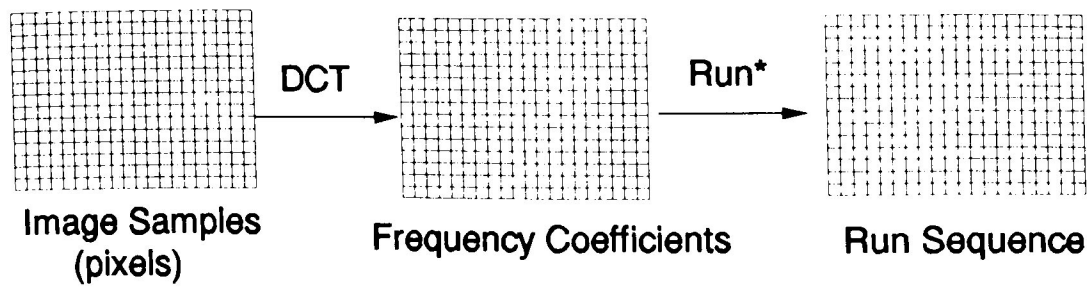
Presentation Time Stamps

Presentation time stamps are

samples of the encoder system clock that are associated with some audio, video, or associated ancillary data *presentation units*. A presentation unit is a decoded video picture, ancillary data or audio time sequence. PTSs are inserted into the MPEG stream on a cyclic basis by the encoder and are extracted by the system decoder. The PTS represents the time at which the video picture is to be displayed, the starting playback time for the audio time sequence, or the time that the associated ancillary data must be activated.

Conclusion

The MPEG standard is a compression standard that uses motion estimation and DCT technology to achieve high efficiency. The MPEG compression standard is flexible enough to be adapted to many different video applications.



*Run = Quantization, Zigzag Scan, Run-length coding
Transform Coding Operations
Figure 7