GUIDE
TO THE USE OF THE
ATSC DIGITAL TELEVISION STANDARD

Blank Page

GUIDE TO THE USE OF THE ATSC DIGITAL TELEVISION STANDARD

Table of Contents

List of Figures

List of Tables

GUIDE TO THE USE OF THE ATSC DIGITAL TELEVISION STANDARD

FOREWORD

The *Guide to the Use of the ATSC Digital Television Standard* was prepared by the Advanced Television Systems Committee (ATSC) Technology Group on Distribution (T3). The document was approved by the members of T3 on February 23, 1995 for submission by letter ballot to the membership of the full ATSC to accompany the Digital Television Standard. The document was approved by the Members of the ATSC on April 12, 1995.

This guide provides tutorial information and an overview of the advanced television (ATV) system defined by ATSC Standard A/53, *ATSC Digital Television Standard.*

Annex A, *Informative References* lists informative references which, through reference in this text, provide information useful to the understanding of the text.

1. SCOPE

This document provides a guide to the understanding of the system characteristics of the U. S. advanced television system as documented in ATSC Standard A/53, *ATSC Digital Television Standard.*

2. NORMATIVE REFERENCES

There are no normative references.

3. DEFINITIONS

3.1 Definitions

With respect to definition of terms, abbreviations and units, the practice of the Institute of Electrical and Electronics Engineers (IEEE) as outlined in the Institute's published standards shall be used. Where an abbreviation is not covered by IEEE practice, or industry practice differs from IEEE practice, then the abbreviation in question will be described in Section 3.3 of this document. Many of the definitions included therein are derived from definitions adopted by MPEG.

3.2 Treatment of syntactic elements

This document contains symbolic references to syntactic elements used in the audio, video, and transport coding subsystems. These references are typographically distinguished by the use of a different font (e.g., restricted), may contain the underscore character (e.g., sequence_end_code) and may consist of character strings that are not English words (e.g., dynrng).

3.3 Terms employed

For the purposes of the Digital Television Standard, the following definition of terms apply:

ACATS: Advisory Committee on Advanced Television Service.

access unit: A coded representation of a presentation unit. In the case of audio, an access unit is the coded representation of an audio frame. In the case of video, an access unit includes all the coded data for a picture, and any stuffing that follows it, up to but not including the start of the next access unit. If a picture is not preceded by a group_start_code or a sequence_header_code, the access unit begins with a the picture start code. If a picture is preceded by a group_start_code and/or a sequence_header_code, the access unit begins with the first byte of the first of these start codes. If it is the last picture preceding a sequence_end_code in the bit stream all bytes between the last byte of the coded picture and the sequence_end_code (including the sequence_end_code) belong to the access unit.

A/D: Analog to digital converter.

AES: Audio Engineering Society.

anchor frame: A video frame that is used for prediction. I-frames and P-frames are generally used as anchor frames, but B-frames are never anchor frames.

ANSI: American National Standards Institute.

Asynchronous Transfer Mode (ATM): A digital signal protocol for efficient transport of both constant-rate and bursty information in broadband digital networks. The ATM digital stream consists of fixed-length packets called "cells," each containing 53 8-bit bytes—a 5-byte header and a 48-byte information payload.

ATEL: Advanced Television Evaluation Laboratory.

ATM: See asynchronous transfer mode.

ATTC: Advanced Television Test Center.

ATV: The U. S. advanced television system.

bidirectional pictures or **B-pictures** or **B-frames:** Pictures that use both future and past pictures as a reference. This technique is termed *bidirectional prediction*. B-pictures provide the most compression. B-pictures do not propagate coding errors as they are never used as a reference.

bit rate: The rate at which the compressed bit stream is delivered from the channel to the input of a decoder.

block: A block is an 8-by-8 array of pel values or DCT coefficients representing luminance or chrominance information.

bps: Bits per second.

byte-aligned: A bit in a coded bit stream is byte-aligned if its position is a multiple of 8-bits from the first bit in the stream.

CDTV: See conventional definition television.

channel: A digital medium that stores or transports a digital television stream.

coded representation: A data element as represented in its encoded form.

compression: Reduction in the number of bits used to represent an item of data.

constant bit rate: Operation where the bit rate is constant from start to finish of the compressed bit stream.

conventional definition television (CDTV): This term is used to signify the *analog* NTSC television system as defined in ITU-R Recommendation 470. See also *standard definition television* and ITU-R Recommendation 1125.

CRC: The cyclic redundancy check to verify the correctness of the data.

D-frame: Frame coded according to an MPEG-1 mode which uses DC coefficients only.

data element: An item of data as represented before encoding and after decoding.

DCT: See discrete cosine transform.

decoded stream: The decoded reconstruction of a compressed bit stream.

decoder: An embodiment of a decoding process.

decoding (process): The process defined in the Digital Television Standard that reads an input coded bit stream and outputs decoded pictures or audio samples.

decoding time-stamp (DTS): A field that may be present in a PES packet header that indicates the time that an access unit is decoded in the system target decoder.

digital storage media (DSM): A digital storage or transmission device or system.

discrete cosine transform: A mathematical transform that can be perfectly undone and which is useful in image compression.

DSM-CC: Digital storage media command and control.

DSM: Digital storage media.

DTS: See decoding time-stamp.

DVCR: Digital video cassette recorder

ECM: See entitlement control message.

editing: A process by which one or more compressed bit streams are manipulated to produce a new compressed bit stream. Conforming edited bit streams are understood to meet the requirements defined in the Digital Television Standard.

elementary stream (ES): A generic term for one of the coded video, coded audio or other coded bit streams. One elementary stream is carried in a sequence of PES packets with one and only one stream_id.

elementary stream clock reference (ESCR): A time stamp in the PES Stream from which decoders of PES streams may derive timing.

EMM: See entitlement management message.

encoder: An embodiment of an encoding process.

encoding (process): A process that reads a stream of input pictures or audio samples and produces a valid coded bit stream as defined in the Digital Television Standard.

entitlement control message (ECM): Entitlement control messages are private conditional access information which specify control words and possibly other stream-specific, scrambling, and/or control parameters.

entitlement management message (EMM): Entitlement management messages are private conditional access information which specify the authorization level or the services of specific decoders. They may be addressed to single decoders or groups of decoders.

entropy coding: Variable length lossless coding of the digital representation of a signal to reduce redundancy.

entry point: Refers to a point in a coded bit stream after which a decoder can become properly initialized and commence syntactically correct decoding. The first transmitted picture after an entry point is either an I-picture or a P-picture. If the first transmitted picture is not an I-picture, the decoder may produce one or more pictures during acquisition.

ES: See elementary stream.

ESCR: See elementary stream clock reference.

event: An event is defined as a collection of elementary streams with a common time base, an associated start time, and an associated end time.

field: For an interlaced video signal, a "field" is the assembly of alternate lines of a frame. Therefore, an interlaced frame is composed of two fields, a top field and a bottom field.

forbidden: This term, when used in clauses defining the coded bit stream, indicates that the value shall never be used. This is usually to avoid emulation of start codes.

FPLL: Frequency and phase locked loop.

frame: A frame contains lines of spatial information of a video signal. For progressive video, these lines contain samples starting from one time instant and continuing through successive lines to the bottom of the frame. For interlaced video a frame consists of two fields, a top field and a bottom field. One of these fields will commence one field later than the other.

GOP: See group of pictures.

Group of pictures (GOP): A group of pictures consists of one or more pictures in sequence.

HDTV: See high definition television.

high definition television (HDTV): High definition television has a resolution of approximately twice that of conventional television in both the horizontal (H) and vertical (V) dimensions and a picture aspect ratio (HxV) of 16:9. ITU-R Recommendation 1125

further defines "HDTV quality" as the delivery of a television picture which is subjectively identical with the interlaced HDTV studio standard.

high level: A range of allowed picture parameters defined by the MPEG-2 video coding specification which corresponds to high definition television.

Huffman coding: A type of source coding that uses codes of different lengths to represent symbols which have unequal likelihood of occurrence.

IEC: International Electrotechnical Commission.

intra coded pictures or **I-pictures** or **I-frames:** Pictures that are coded using information present only in the picture itself and not depending 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 pel blocks and provide only moderate compression.

ISO: International Organization for Standardization.

ITU: International Telecommunication Union.

JEC: Joint Engineering Committee of EIA and NCTA.

layer: One of the levels in the data hierarchy of the video and system specification.

level: A range of allowed picture parameters and combinations of picture parameters.

macroblock: In the ATV system a macroblock consists of four blocks of luminance and one each Cr and Cb block.

main level: A range of allowed picture parameters defined by the MPEG-2 video coding specification with maximum resolution equivalent to ITU-R Recommendation 601.

main profile: A subset of the syntax of the MPEG-2 video coding specification that is expected to be supported over a large range of applications.

Mbps: 1,000,000 bits per second.

motion vector: A pair of numbers which represent the vertical and horizontal displacement of a region of a reference picture for prediction.

MP@HL: Main profile at high level.

MP@ML: Main profile at main level.

MPEG: Refers to standards developed by the ISO/IEC JTC1/SC29 WG11, *Moving Picture Experts Group.* MPEG may also refer to the Group.

MPEG-1: Refers to ISO/IEC standards 11172-1 (Systems), 11172-2 (Video), 11172-3 (Audio), 11172-4 (Compliance Testing), and 11172-5 (Technical Report).

MPEG-2: Refers to ISO/IEC standards 13818-1 (Systems), 13818-2 (Video), 13818-3 (Audio), 13818-4 (Compliance).

pack: A pack consists of a pack header followed by zero or more packets. It is a layer in the system coding syntax.

packet data: Contiguous bytes of data from an elementary data stream present in the packet.

packet identifier (PID): A unique integer value used to associate elementary streams of a program in a single or multi-program transport stream.

packet: A packet consists of a header followed by a number of contiguous bytes from an elementary data stream. It is a layer in the system coding syntax.

padding: A method to adjust the average length of an audio frame in time to the duration of the corresponding PCM samples, by continuously adding a slot to the audio frame.

payload: Payload refers to the bytes which follow the header byte in a packet. For example, the payload of a transport stream packet includes the PES_packet_header and its PES_packet_data_bytes or pointer_field and PSI sections, or private data. A PES_packet_payload, however, consists only of PES_packet_data_bytes. The transport stream packet header and adaptation fields are not payload.

PCR: See program clock reference.

pel: See pixel.

PES packet header: The leading fields in a PES packet up to but not including the PES_packet_data_byte fields where the stream is not a padding stream. In the case of a padding stream, the PES packet header is defined as the leading fields in a PES packet up to but not including the padding_byte fields.

PES packet: The data structure used to carry elementary stream data. It consists of a packet header followed by PES packet payload.

PES Stream: A PES stream consists of PES packets, all of whose payloads consist of data from a single elementary stream, and all of which have the same stream_id.

PES: An abbreviation for packetized elementary stream.

picture: Source, coded or reconstructed image data. A source or reconstructed picture consists of three rectangular matrices representing the luminance and two chrominance signals.

PID: See packet identifier.

pixel: "Picture element" or "pel." A pixel is a digital sample of the color intensity values of a picture at a single point.

predicted pictures or **P-pictures** or **P-frames:** Pictures that are coded with respect to the nearest *previous* I or P-picture. This technique is termed *forward prediction*. 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.

presentation time-stamp (PTS): A field that may be present in a PES packet header that indicates the time that a presentation unit is presented in the system target decoder.

presentation unit (PU): A decoded audio access unit or a decoded picture.

profile: A defined subset of the syntax specified in the MPEG-2 video coding specification

program clock reference (PCR): A time stamp in the transport stream from which decoder timing is derived.

program element: A generic term for one of the elementary streams or other data streams that may be included in the program.

program specific information (PSI): PSI consists of normative data which is necessary for the demultiplexing of transport streams and the successful regeneration of programs.

program: A program is a collection of program elements. Program elements may be elementary streams. Program elements need not have any defined time base; those that do have a common time base and are intended for synchronized presentation.

PSI: See program specific information.

PTS: See presentation time-stamp.

PU: See presentation unit.

quantizer: A processing step which intentionally reduces the precision of DCT coefficients

random access: The process of beginning to read and decode the coded bit stream at an arbitrary point.

reserved: This term, when used in clauses defining the coded bit stream, indicates that the value may be used in the future for Digital Television Standard extensions. Unless otherwise specified within this Standard, all reserved bits shall be set to "1".

SCR: See system clock reference.

scrambling: The alteration of the characteristics of a video, audio or coded data stream in order to prevent unauthorized reception of the information in a clear form. This alteration is a specified process under the control of a conditional access system.

SDTV: See standard definition television.

slice: A series of consecutive macroblocks.

SMPTE: Society of Motion Picture and Television Engineers.

source stream: A single, non-multiplexed stream of samples before compression coding.

splicing: The concatenation performed on the system level or two different elementary streams. It is understood that the resulting stream must conform totally to the Digital Television Standard.

standard definition television (SDTV): This term is used to signify a *digital* television system in which the quality is approximately equivalent to that of NTSC. This equivalent quality may be achieved from pictures sourced at the 4:2:2 level of ITU-R Recommendation 601 and subjected to processing as part of the bit rate compression. The results should be such that when judged across a representative sample of program

material, subjective equivalence with NTSC is achieved. Also called standard digital television. See also *conventional definition television* and ITU-R Recommendation 1125.

start codes: 32-bit codes embedded in the coded bit stream that are unique. They are used for several purposes including identifying some of the layers in the coding syntax. Start codes consist of a 24 bit prefix (0x000001) and an 8 bit stream_id.

STD input buffer: A first-in, first-out buffer at the input of a system target decoder for storage of compressed data from elementary streams before decoding.

STD: See system target decoder.

still picture: A coded still picture consists of a video sequence containing exactly one coded picture which is intra-coded. This picture has an associated PTS and the presentation time of succeeding pictures, if any, is later than that of the still picture by at least two picture periods.

system clock reference (SCR): A time stamp in the program stream from which decoder timing is derived.

system header: The system header is a data structure that carries information summarizing the system characteristics of the Digital Television Standard multiplexed bit stream.

system target decoder (STD): A hypothetical reference model of a decoding process used to describe the semantics of the Digital Television Standard multiplexed bit stream.

time-stamp: A term that indicates the time of a specific action such as the arrival of a byte or the presentation of a presentation unit.

TOV: Threshold of visibility.

transport stream packet header: The leading fields in a transport stream packet up to and including the continuity_counter field.

variable bit rate: Operation where the bit rate varies with time during the decoding of a compressed bit stream.

VBV: See video buffering verifier.

Video buffering verifier (VBV): A hypothetical decoder that is conceptually connected to the output of an encoder. Its purpose is to provide a constraint on the variability of the data rate that an encoder can produce.

video sequence: 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.

8 VSB: Vestigial sideband modulation with 8 discrete amplitude levels.

16 VSB: Vestigial sideband modulation with 16 discrete amplitude levels.

3.4 Symbols, abbreviations, and mathematical operators

3.4.1 Introduction

The symbols, abbreviations, and mathematical operators used to describe the Digital Television Standard are those adopted for use in describing MPEG-2 and are similar to those used in the "C" programming language. However, integer division with truncation and rounding are specifically defined. The bitwise operators are defined assuming two's-complement representation of integers. Numbering and counting loops generally begin from 0.

3.4.2 Arithmetic operators

+	Addition.
-	Subtraction (as a binary operator) or negation (as a unary operator).
++	Increment.
- -	Decrement.
* or ×	Multiplication.
^	Power.
/	Integer division with truncation of the result toward 0. For example, 7/4 and -7/-4 are truncated to 1 and -7/4 and 7/-4 are truncated to -1.
//	Integer division with rounding to the nearest integer. Half-integer values are rounded away from 0 unless otherwise specified. For example 3//2 is rounded to 2, and -3//2 is rounded to -2.
DIV	Integer division with truncation of the result towards $-\infty$.
%	Modulus operator. Defined only for positive numbers.

Sign( ) $\text{Sign}(x) = \begin{cases} 1 & x > 0 \\ 0 & x == 0 \\ -1 & x < 0 \end{cases}$

NINT ( )   Nearest integer operator. Returns the nearest integer value to the real-valued argument. Half-integer values are rounded away from 0.

sin	Sine.
cos	Cosine.
exp	Exponential.
√	Square root.
$\log_{10}$	Logarithm to base ten.
$\log_e$	Logarithm to base e.

3.4.3 Logical operators

|| Logical OR.

&&   Logical AND.

!   Logical NOT.

3.4.4 Relational operators

\>   Greater than.

≥   Greater than or equal to.

\<   Less than.

≤   Less than or equal to.

==   Equal to.

!=   Not equal to.

max [,...,] The maximum value in the argument list.

min [,...,] The minimum value in the argument list.

3.4.5 Bitwise operators

&   AND.

| OR.

\>>   Shift right with sign extension.

<<   Shift left with 0 fill.

3.4.6 Assignment

=   Assignment operator.

3.4.7 Mnemonics

The following mnemonics are defined to describe the different data types used in the coded bit stream.

bslbf	Bit string, left bit first, where "left" is the order in which bit strings are written in the Standard. Bit strings are written as a string of 1s and 0s within single quote marks, e.g. '1000 0001'. Blanks within a bit string are for ease of reading and have no significance.
uimsbf	Unsigned integer, most significant bit first.

The byte order of multi-byte words is most significant byte first.

3.4.8 Constants

$$\pi \qquad 3.14159265359...$$

$$e \qquad 2.71828182845...$$

3.4.9 Method of describing bit stream syntax

Each data item in the coded bit stream described below is in bold type. It is described by its name, its length in bits, and a mnemonic for its type and order of transmission.

The action caused by a decoded data element in a bit stream depends on the value of that data element and on data elements previously decoded. The decoding of the data elements and definition of the state variables used in their decoding are described in the clauses containing the semantic description of the syntax. The following constructs are used to express the conditions when data elements are present, and are in normal type.

Note this syntax uses the "C" code convention that a variable or expression evaluating to a non-zero value is equivalent to a condition that is true.

while (condition) {
    data_element
    . . .
}
	If the condition is true, then the group of data elements occurs next in the data stream. This repeats until the condition is not true.

do {
    data_element
    . . . }
 while (condition)
	The data element always occurs at least once. The data element is repeated until the condition is not true.

if (condition) {
    data_element
    . . .
}
	If the condition is true, then the first group of data elements occurs next in the data stream.

else {
    data_element
    . . .
}
	If the condition is not true, then the second group of data elements occurs next in the data stream.

for (i = 0;i<n;i++) {
    data_element
    . . .
}
	The group of data elements occurs n times. Conditional constructs within the group of data elements may depend on the value of the loop control variable i, which is set to zero for the first occurrence, incremented to 1 for the second occurrence, and so forth.

As noted, the group of data elements may contain nested conditional constructs. For compactness, the {} are omitted when only one data element follows.

data_element [ ]	data_element [ ] is an array of data. The number of data elements is indicated by the context.
data_element [n]	data_element [n] is the n+1th element of an array of data.
data_element [m][n]	data_element [m][n] is the m+1,n+1 th element of a two-dimensional array of data.

data_element [l][m][n]	data_element [l][m][n] is the l+1,m+1,n+1 th element of a three-dimensional array of data.
data_element [m..n]	data_element [m..n] is the inclusive range of bits between bit m and bit n in the data_element.

Decoders must include a means to look for start codes and sync bytes (transport stream) in order to begin decoding correctly, and to identify errors, erasures or insertions while decoding. The methods to identify these situations, and the actions to be taken, are not standardized.

3.4.9.1 Definition of bytealigned function

The function bytealigned( ) returns 1 if the current position is on a byte boundary; that is, the next bit in the bit stream is the first bit in a byte. Otherwise it returns 0.

3.4.9.2 Definition of nextbits function

The function nextbits( ) permits comparison of a bit string with the next bits to be decoded in the bit stream.

3.4.9.3 Definition of next_start_code function

The next_start_code( ) function removes any zero bit and zero byte stuffing and locates the next start code.

This function checks whether the current position is byte-aligned. If it is not, 0 stuffing bits are present. After that any number of 0 bytes may be present before the start-code. Therefore start-codes are always byte-aligned and may be preceded by any number of 0 stuffing bits.

Table 3.1 Next Start Code

Syntax	No. of bits	Mnemonic

next_start_code( ) {
    while (!bytealigned( ))
        zero_bit
    while (nextbits( )!='0000 0000 0000 0000 0000 0001')
        zero_byte
}
	
	1
	
	8
		
		'0'
		
		'00000000'

4. SYSTEM OVERVIEW

4.1 Objectives

The Digital Television Standard describes a system designed to transmit high quality video and audio and ancillary data over a single 6 MHz channel. The system can deliver reliably about 19 Mbps of throughput in a 6 MHz terrestrial broadcasting channel and about 38 Mbps of throughput in a 6 MHz cable television channel. This means that encoding a video source whose resolution can be as high as five times that of conventional television (NTSC) resolution requires a bit rate reduction by a factor of 50 or higher. To achieve this bit rate reduction, the system is designed to be efficient in utilizing available channel capacity by exploiting complex video and audio compression technology.

The objective is to maximize the information passed through the data channel by minimizing the amount of data required to represent the video image sequence and its associated audio. The objective is to represent the video, audio, and data sources with as few bits as possible while preserving the level of quality required for the given application.

Although the RF/Transmission subsystems described in the Digital Television Standard are designed specifically for terrestrial and cable applications, the objective is that the video, audio, and service multiplex/transport subsystems be useful in other applications.

4.2 System block diagram

A basic block diagram representation of the system is shown in Figure 4.1. This representation is based on one adopted by the International Telecommunication Union, Radiocommunication Sector (ITU-R), Task Group 11/3 (Digital Terrestrial Television Broadcasting). According to this model, the digital television system can be seen to consist of three subsystems.[1]

1. Source coding and compression,

2. Service multiplex and transport, and

3. RF/Transmission.

"Source coding and compression" refers to the bit rate reduction methods, also known as data compression, appropriate for application to the video, audio, and ancillary digital data streams. The term "ancillary data" includes control data, conditional access control data, and data associated with the program audio and video services, such as closed captioning. "Ancillary data" can also refer to independent program services. The purpose of the coder is to minimize the number of bits needed to represent the audio and video information. The digital television system employs the MPEG-2 video stream syntax for the coding of video and the Digital Audio Compression (AC-3) Standard for the coding of audio.

---

[1] ITU-R Document TG11/3-2, "Outline of Work for Task Group 11/3, Digital Terrestrial Television Broadcasting," June 30, 1992.
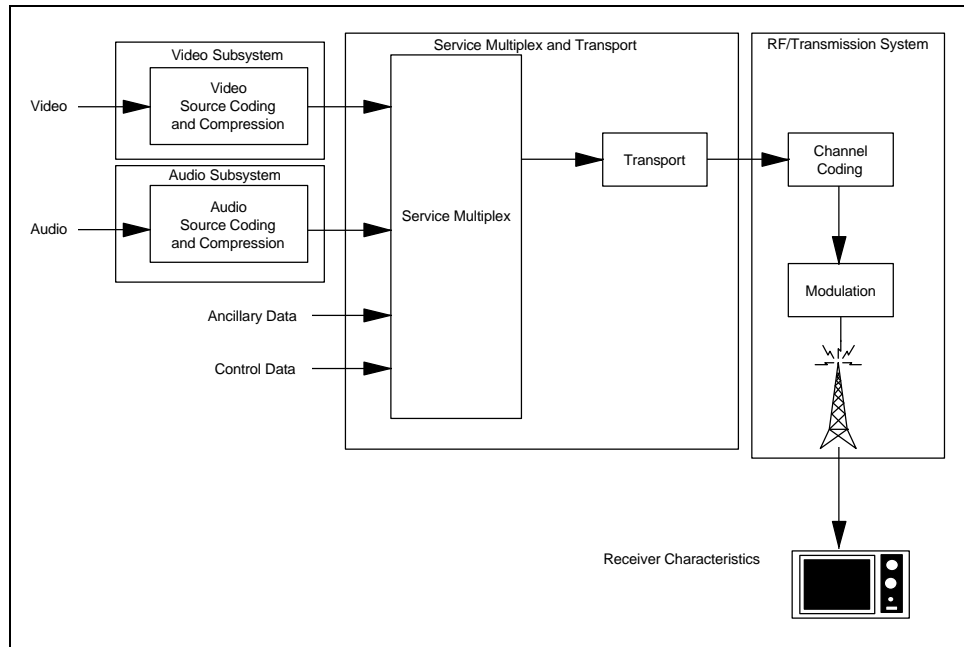
Figure 4.1. ITU-R digital terrestrial television broadcasting model.

"Service multiplex and transport" refers to the means of dividing the digital data stream into "packets" of information, the means of uniquely identifying each packet or packet type, and the appropriate methods of multiplexing video data stream packets, audio data stream packets, and ancillary data stream packets into a single data stream. In developing the transport mechanism, interoperability among digital media, such as terrestrial broadcasting, cable distribution, satellite distribution, recording media, and computer interfaces, was a prime consideration. The digital television system employs the MPEG-2 transport stream syntax for the packetization and multiplexing of video, audio, and data signals for digital broadcasting systems.[2] The MPEG-2 transport stream syntax was developed for applications where channel bandwidth or recording media capacity is limited and the requirement for an efficient transport mechanism is paramount. It was designed also to facilitate interoperability with the ATM transport mechanism.

"RF/Transmission" refers to channel coding and modulation. The channel coder takes the data bit stream and adds additional information that can be used by the receiver to reconstruct the data from the received signal which, due to transmission impairments, may not accurately represent the transmitted signal. The modulation (or physical layer) uses the digital data stream information to modulate the transmitted signal. The modulation subsystem offers two modes: a terrestrial broadcast mode (8 VSB), and a high data rate mode (16 VSB).

The Chapters that follow consider the characteristics of the subsystems necessary to accommodate the services envisioned.

---

[2] Chairman, ITU-R Task Group 11/3, "Report of the Second Meeting of ITU-R Task Group 11/3, Geneva, October 13-19, 1993," January 5, 1994, p. 40.

5. VIDEO SYSTEMS

5.1 Overview of video compression and decompression

The need for compression in a digital HDTV system is apparent from the fact that the bit rate required to represent an HDTV signal in uncompressed digital form is about 1 Gbps, and the bit rate that can reliably be transmitted within a standard 6 MHz television channel is about 20 Mbps. This implies a need for about a 50:1 or greater compression ratio.

The Digital Television Standard specifies video compression using a combination of compression techniques, and for reasons of compatibility these compression algorithms have been selected to conform to the specifications of MPEG-2, which is a flexible internationally accepted collection of compression algorithms.

The purpose of this tutorial exposition is to identify the significant processing stages in video compression and decompression, giving a clear explanation of what each processing step accomplishes, but without including all the details that would be needed to actually implement a real system. Those necessary details in every case are specified in the normative part of the standards documentation, which shall in all cases represent the most complete and accurate description of the video compression. Because the video coding system includes a specific subset of the MPEG-2 toolkit of algorithmic elements, another purpose of this tutorial is to clarify the relationship between this system and the more general MPEG-2 collection of algorithms.

5.1.1 MPEG-2 levels and profiles

The MPEG-2 specification is organized into a system of profiles and levels, so that applications can ensure interoperability by using equipment and processing that adhere to a common set of coding tools and parameters.[3] The Digital Television Standard is based on the MPEG-2 Main Profile. The Main Profile includes three types of frames for prediction (I-frames, P-frames, and B-frames), and an organization of luminance and chrominance samples (designated 4:2:0) within the frame. The Main Profile does not include a scalable algorithm, where scalability implies that a subset of the compressed data can be decoded without decoding the entire data stream. The High Level includes formats with up to 1152 active lines and up to 1920 samples per active line, and for the Main Profile is limited to a compressed data rate of no more than 80 Mbps. The parameters specified by the Digital Television Standard represent specific choices within these constraints.

5.1.2 Compatibility with MPEG-2

The video compression system does not include algorithmic elements that fall outside the specifications for MPEG-2 Main Profile. Thus video decoders which conform

---

[3] For more information about profiles and levels see ISO/IEC 13818-2, Section 8.

to the MPEG-2 MP@HL can be expected to decode bit streams produced in accordance with the Digital Television Standard. Note that it is not necessarily the case that all video decoders which are based on the Digital Television Standard will be able to properly decode all video bit streams which comply to MPEG-2 MP@HL.

5.1.3 Overview of video compression

The video compression system takes in an analog video source signal and outputs a compressed digital signal that contains information that can be decoded to produce an approximate version of the original image sequence. The goal is for the reconstructed approximation to be imperceptibly different from the original for most viewers, for most images, for most of the time. In order to approach such fidelity, the algorithms are flexible, allowing for frequent adaptive changes in the algorithm depending on scene content, history of the processing, estimates of image complexity and perceptibility of distortions introduced by the compression.

Figure 5.1 shows the overall flow of signals in the ATV system. Note that analog signals presented to the system are digitized and sent to the encoder for compression, and the compressed data then are transmitted over a communications channel. On being received, the possibly error-corrupted compressed signal is decompressed in the decoder, and reconstructed for display.
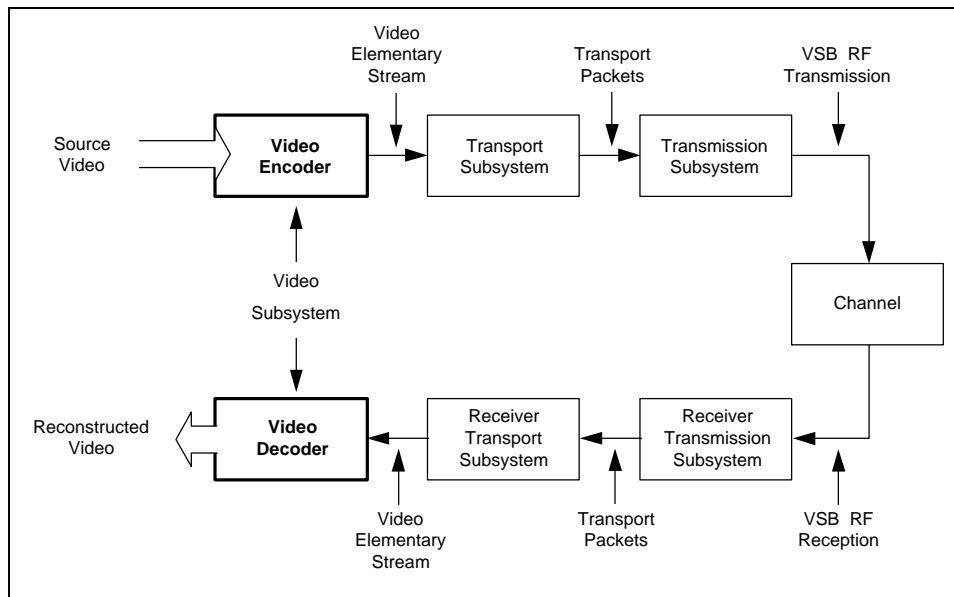
Figure 5.1. Video coding in relation to the ATV system.

5.2 Video preprocessing

Video preprocessing converts the analog input signals to digital samples in the form needed for the subsequent compression. The analog input signals are red (R), green (G), and blue (B) signals.

5.2.1 Video compression formats

Table 5.1 lists the compression formats allowed in the Digital Television Standard.

Table 5.1 Compression Formats

Vertical lines	Pixels	Aspect ratio	Picture rate
1080	1920	16:9	60I, 30P, 24P
720	1280	16:9	60P, 30P, 24P
480	704	16:9 and 4:3	60P, 60I, 30P, 24P
480	640	4:3	60P, 60I, 30P, 24P

In Table 5.1, "vertical lines" refers to the number of active lines in the picture. "Pixels" refers to the number of pixels during the active line. "Aspect ratio" refers to the picture aspect ratio. "Picture rate" refers to the number of frames or fields per second. In the values for picture rate, "P" refers to progressive scanning, "I" refers to interlaced scanning. Note that both 60.00 Hz and 59.94 (60x1000/1001) Hz picture rates are allowed. Dual rates are allowed also at the picture rates of 30 Hz and 24 Hz.

5.2.1.1 Possible video inputs

While not required by the Digital Television Standard, there are certain television production standards, shown in Table 5.2, that define video formats that relate to compression formats specified by the Standard.

Table 5.2 Standardized Video Input Formats

Video standard	Active lines	Active samples/ line
SMPTE 274M	1080	1920
SMPTE S17.392	720	1280
ITU-R BT.601-4	483	720

The compression formats may be derived from one or more appropriate video input formats. It may be anticipated that additional video production standards will be developed in the future that extend the number of possible input formats.

5.2.1.2 Sampling rates

For the 1080-line format, with 1125 total lines per frame and 2200 total samples per line, the sampling frequency will be 74.25 MHz for the 30.00 frames per second (fps) frame rate. For the 720-line format, with 750 total lines per frame and 1650 total samples per line, the sampling frequency will be 74.25 MHz for the 60.00 fps frame rate. For the 480-line format using 704 pixels, with 525 total lines per frame and 858 total samples per line, the sampling frequency will be 13.5 MHz for the 59.94 Hz field rate. Note that both 59.94 fps and 60.00 fps are acceptable as frame or field rates for the system.

For the 480-line format, there may be 704 or 640 pixels in the active line. If the input is based on ITU-R BT.601-4, it will have 483 active lines with 720 pixels in the active line. Only 480 of the 483 active lines are used for encoding. Only 704 of the 720

pixels are used for encoding; the first eight and the last eight are dropped. The 480-line, 640 pixel picture format is not related to any current video production format. It does correspond to the IBM VGA graphics format and may be used with ITU-R BT.601-4 sources by using appropriate resampling techniques.

5.2.1.3 Colorimetry

For the purposes of the Digital Television Standard, "colorimetry" means the combination of color primaries, transfer characteristics, and matrix coefficients. Video inputs conforming to SMPTE 274M and S17.392 have the same colorimetry; in this document, this will be referred to as SMPTE 274M colorimetry. Note that SMPTE 274M colorimetry is the same as ITU-R BT.709 (1990) colorimetry. The Digital Television Standard specifies SMPTE 274M colorimetry as the default, and preferred, colorimetry. Video inputs corresponding to ITU-R BT.601-4 may have SMPTE 274M colorimetry or SMPTE 170M colorimetry.

ISO/IEC 13818-2 allows the encoder to signal the input colorimetry parameter values to the decoder. If sequence_display_extension() is not present in the bit stream, or if color_description is zero, the color primaries, transfer characteristics, and matrix coefficients are assumed to have the value 0x01, which is the correct set of values for SMPTE 274M. As a result, if the input colorimetry is SMPTE 274M, it is not necessary to signal the colorimetry parameter values in the bit stream. (In a proposed amendment to ISO/IEC 13818-2, if sequence_display_extension() is not present in the bit stream, or if color_description is zero, the color primaries, transfer characteristics, and matrix coefficients are assumed to be implicitly defined by the application. Because the Digital Television Standard defines the colorimetry as SMPTE 274M, if the input colorimetry is SMPTE 274M, it is not necessary to signal the colorimetry parameter values in the bit stream.)

In generating bit streams, broadcasters should understand that many receivers will likely display all inputs, regardless of colorimetry, according to the default SMPTE 274M colorimetry. Some receivers may include circuitry to properly display also SMPTE 170M colorimetry (color primaries, transfer characteristics, and matrix coefficients have the value 0x06). It is believed that few receivers will display properly the other colorimetry combinations allowed by ISO/IEC 13818-2.

5.2.2 Precision of samples

Samples are typically obtained using analog-to-digital converter circuits with 8-bit precision. After preprocessing, the various luminance and chrominance samples will typically be represented using 8 bits per sample of each component.

5.2.3 Source-adaptive processing

The image sequences that constitute the source signal can vary in spatial resolution (480 lines, 720 lines, or 1080 lines), in temporal resolution (60 fps, 30 fps, or 24 fps), and in scanning format (2:1 interlaced or progressive scan). The video compression system accommodates the differences in source material to maximize the efficiency of compression.

5.2.4 Film mode

When a large fraction of pixels do not change from one frame in the image sequence to the next, a video encoder may automatically recognize that the input was film with an underlying frame rate less than 60 fps.

In the case of 24 fps film material that is sent at 60 Hz using a 3:2 pull-down operation, the processor may detect the sequences of three nearly identical pictures followed by two nearly identical pictures, and only encode the 24 unique pictures per second that existed in the original film sequence. When 24 fps film is detected by observation of the 3:2 pull-down pattern, the input signal is converted back to a progressively scanned sequence of 24 frames per second prior to compression. This avoids sending redundant information, and allows the encoder to provide an improved quality of compression. The encoder indicates to the decoder that the film mode is active.

In the case of 30 fps film material that is sent at 60 Hz, the processor may detect the sequences of two nearly identical pictures followed by two nearly identical pictures. In that case, the input signal is converted back to a progressively scanned sequence of 30 frames per second.

5.2.5 Color component separation and processing

The input video source to the ATV video compression system is in the form of RGB components matrixed into luminance (Y) and chrominance (Cb and Cr) components using a linear transformation (3-by-3 matrix, specified in the standard). The luminance component represents the intensity, or black-and-white picture, while the chrominance components contain color information. The original RGB components are highly correlated with each other; the resulting Y, Cb, and Cr signals have less correlation and are thus easier to code efficiently; The luminance and chrominance components correspond to functioning of the biological vision system; that is, the human visual system responds differently to the luminance and chrominance components.

The coding process may take advantage also of the differences in the ways that humans perceive luminance and chrominance. In the Y, Cb, Cr color space, most of the high frequencies are concentrated in the Y component; the human visual system is less sensitive to high frequencies in the chrominance components than to high frequencies in the luminance component. To exploit these characteristics the chrominance components are low-passed filtered in the ATV video compression system and sub-sampled by a factor of two along both the horizontal and vertical dimensions, producing chrominance components that are one-fourth the spatial resolution of the luminance component.

5.2.6 Anti-alias filtering

The Y, Cb, and Cr components are applied to appropriate low-pass filters that shape the frequency response of each of the three components. Prior to horizontal and

vertical sub-sampling of the two chrominance components, they may be processed by half-band filters in order to prevent aliasing.[4]

5.2.7 Number of lines encoded

The video coding system requires that the coded picture area has a number of lines that is a multiple of 32 for an interlaced format, and a multiple of 16 for a non-interlaced format. This means that for encoding the 1080-line format, a coder must actually deal with 1088 lines (1088 = 32 x 34). The extra eight lines are in effect "dummy" lines having no content, and the coder designers will choose dummy data that simplifies the implementation. The extra eight lines are always the last eight lines of the encoded image. These dummy lines do not carry useful information, but add little to the data required for transmission.

5.3 Representation of picture data

Digital television uses a digital representation of the image data, which allows the data to be processed using computer-like digital processing. The process of digitization involves sampling of the analog television signals and their components, and representing each sample with a digital code.

5.3.1 Pixels

The analog video signals are sampled in a sequence that corresponds to the scanning raster of the television format; *i.e.*, from left to right within a line, and in lines from top to bottom. The collection of samples in a single frame, or in a single field for interlaced images, is treated together, as if they all corresponded to a single point in time (in the case of film modes, they do in fact correspond to a single time or exposure interval). The individual samples of image data are referred to as picture elements, or "pixels," or "pels." A single frame or field can then be thought of as a rectangular array of pixels.

5.3.1.1 Square pixels

When the ratio of active pixels per line to active lines per frame is the same as the display aspect ratio, which is 16:9, the format is said to have "square" pixels. The term refers to the spacing of samples and does not refer to the shape of the pixel, which might ideally be a point with zero area from a mathematical sampling point of view.

5.3.1.2 Spatial relationship between luminance and chrominance samples

As described in Section 5.2.5, the chrominance component samples are sub-sampled by a factor of two in both horizontal and vertical directions. This means the chrominance samples are spaced twice as far apart as the luminance samples, and it is

---

[4] For more information about aliasing and sampling theory, see James A. Cadzow, *Discrete Time Systems*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1973.

necessary to specify the location of chrominance samples relative to the luminance samples.

Figure 5.2 illustrates the spatial relationship between chrominance and luminance samples. For every four luminance samples, there are one each of the Cb and Cr chroma samples. The Cb and Cr chroma samples are located in the same place.
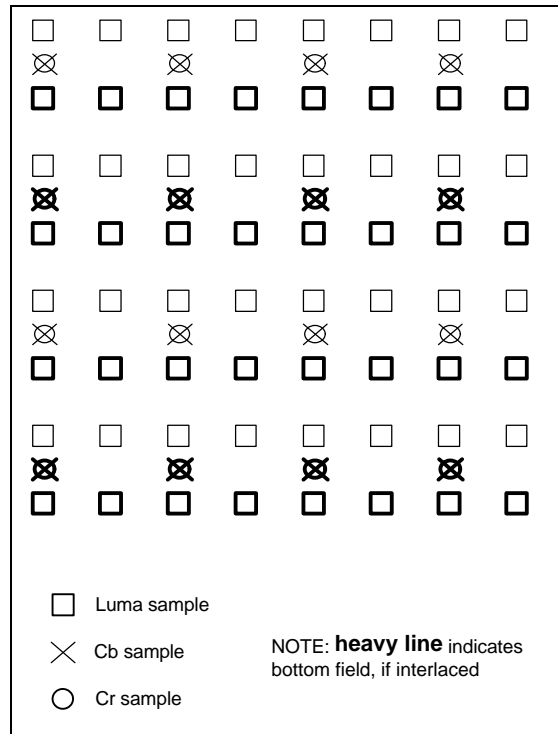
Figure 5.2. Placement of luma/chroma samples for 4:2:0.

Note that the vertical spatial location of chrominance samples does not correspond to an original sample point, but lies halfway between samples on two successive lines. The 4:2:0 sampling structure thus requires the Cb and Cr samples to be interpolated. For progressively scanned source pictures the processor may simply average the two adjacent (upper and lower) values to compute the sub-sampled values.

In the case of interlaced pictures, it can be seen in Figure 5.2 that the vertical positions of the chrominance samples in a field are not halfway between the luminance samples of the same field. This is done so that the spatial locations of the chrominance samples in the frame are the same for both interlaced and progressive sources.

5.3.2 Blocks of pixels

The pixels are organized into blocks for the purpose of further processing. A block consists of an array of pixel values or an array that is some transform of pixel values. A block for the ATV system is defined as an array of 8-by-8 values representing either luminance or chrominance information (see Figure 5.3a).
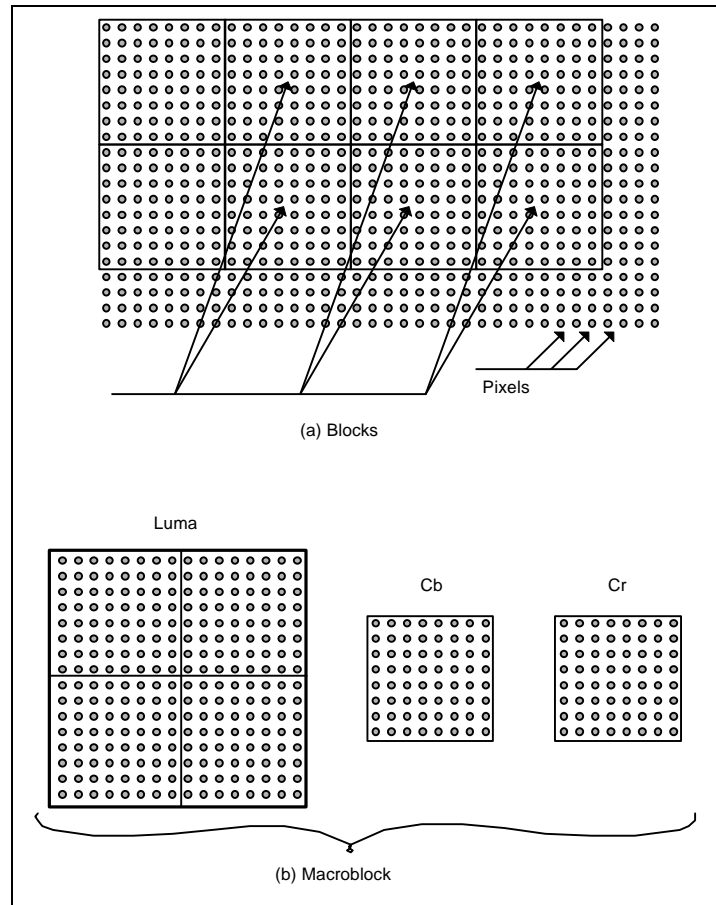
Figure 5.3. Blocks and macroblocks.

5.3.3 Macroblocks

Blocks of information are organized into macroblocks. A macroblock consists of four blocks of luminance (or a 16 pixel by 16 line region of values) and two chroma (Cb and Cr) blocks. The term *macroblock* may be used to refer directly to pel data or to the transformed and coded representation of pel data. As shown in Figure 5.3b, this yields 256 luminance samples and 64 Cb samples and 64 Cr samples (total of 384) per macroblock.

For the 1080-line format (with 1920 samples per line), there are 68 rows of macroblocks (including the last row that adds eight dummy lines to create the 1088 lines for coding), with 120 macroblocks per row. For the 720 line format (with 1280 samples per line), there are therefore 45 rows of macroblocks, with 80 macroblocks per row. For the 480-line format with 704 samples per line, there are 30 rows of macroblocks, with 44 macroblocks per row. For the 480-line format with 640 samples per line, there are 30 rows of macroblocks, with 40 macroblocks per row.

5.3.4 Slices

One or more contiguous macroblocks within the same row 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.

Slices provide a convenient mechanism for limiting the propagation of errors. Since the coded bit stream consists mostly of variable-length codewords, any uncorrected transmission errors will cause a decoder to lose its sense of codeword alignment. Each slice begins with a slice start code. Since the MPEG codeword assignment guarantees that no legal combination of codewords can emulate a start code, the slice start code can be used to regain the sense of codeword alignment after an error. When an error occurs in the data stream, the decoder can thus skip to the start of the next slice and resume correct decoding.

The number of slices affects the compression efficiency; partitioning the data stream to have more slices provides for better error recovery but uses bits that could otherwise be used to improve picture quality. The slice is the minimum unit for resynchronization after an error.

In the ATV system, the initial macroblock of every horizontal row of macroblocks is also the beginning of a slice, with possibly several slices across the row.

5.3.5 Pictures, groups of pictures, and sequences

The primary coding unit of a video sequence is the individual video frame or picture. A video picture consists of the collection of slices which constitute the active picture area.

A video sequence consists of a collection of one or more consecutive pictures. A video sequence commences with a sequence header and is terminated by an end-of-sequence code in the data stream. A video sequence can contain additional sequence headers. Any video sequence header can serve as an entry point. An entry point is a point in the coded video bit stream after which a decoder can become properly initialized and correctly parse the bit stream syntax.

One or more pictures (frames) in sequence may be combined into a Group of Pictures (GOP) to provide boundaries for inter-picture coding and registration of time code. GOPs are optional within both MPEG-2 and the ATV system.

Figure 5.4 illustrates a time sequence of video frames consisting of intra-coded pictures (I-frames); predictive coded pictures (P-frames), and bidirectionally predictive coded pictures (B-frames).

5.3.5.1 I-frames

Some elements of the compression process exploit only the spatial redundancy within a single picture (frame or field). These processes are called intraframe coding, and do not take advantage of the temporal correlation addressed by temporal prediction, which is referred to as interframe coding. Frames that do not use any interframe coding are

referred to as I-frames (where "I" denotes *intraframe* coded). The ATV video compression system utilizes both intraframe coding and interframe coding.
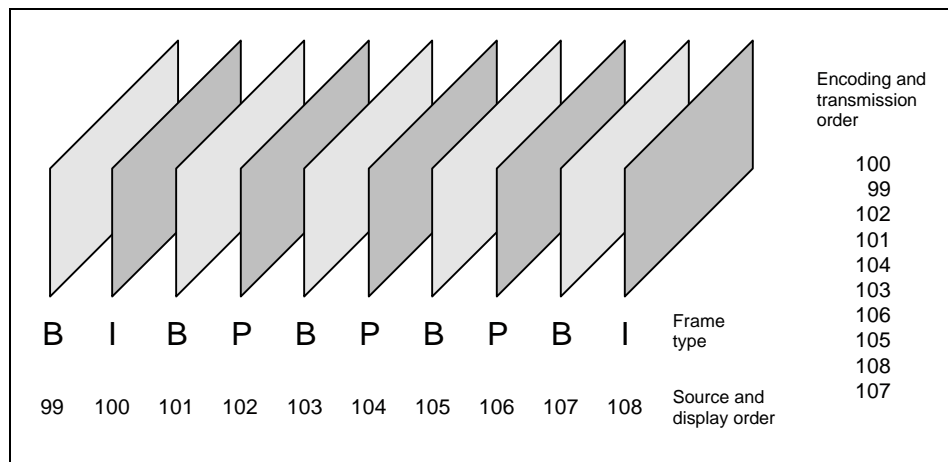
Figure 5.4. Illustration of I-frames, P-frames, and B-frames.

The use of periodic I-frames facilitates receiver initializations and channel acquisition (when the receiver is turned on or the channel is changed). The decoder can take advantage of the intraframe coding mode when non-correctable channel errors occur. With motion-compensated prediction, an initial frame must be available at the decoder to start the prediction loop. Therefore, a mechanism must be built into the system so that if the decoder loses synchronization for any reason, it can rapidly reacquire tracking.

The frequency of occurrence of I-pictures may vary and is selected at the encoder. This allows consideration to be given to the need for random access and the location of scene cuts in the video sequence.

5.3.5.2 P-frames

P-frames (where "P" denotes *predicted*) are frames where the temporal prediction is in the forward direction only (i.e., predictions for the P-frame are formed only from pixels in the most recently decoded I or P-frame). These forward-predicted frames allow the exploitation of interframe coding techniques to improve the overall compression efficiency and picture quality. P-frames may include portions that are only intraframe coded. Each macroblock within a P-frame can be either forward-predicted or intraframe coded.

5.3.5.3 B-frames

The B-frame (where "B" denotes *bidirectionally* predicted) is a picture type within the coded video sequence that includes prediction from a future frame as well as from a previous frame. The referenced future or previous frames, sometimes called "anchor" frames, are in all cases either I or P-frames.

The basis of the B-frame prediction is that a video frame is correlated both with frames which occur in the past and frames which occur in the future. Consequently, if a

future frame is available to the decoder, a superior prediction can be formed, thus saving bits and improving performance. Some of the consequences of using future frames in the prediction are: the B-frame cannot be used for predicting future frames, the transmission order of frames is different from the displayed order of frames, and the encoder and decoder must reorder the video frames thereby increasing the total latency. In the example illustrated in Figure 5.4, there is one B-frame between each pair of I/P-frames. Each frame is labeled with both its display order and transmission order. The I and P frames are transmitted out of sequence so the video decoder has both anchor frames decoded and available for prediction.

B-frames are used for increasing the compression efficiency and perceived picture quality when encoding latency is not an important factor. The use of B-frames increases coding efficiency for both interlaced and progressively scanned material. B-frames are included in the ATV system because the increase in compression efficiency is noticeable especially with progressive scanning where techniques such as dual prime (see Section 5.5.2) are not available. The choice of number of bidirectional pictures between any pair of reference (I or P) frames can be determined at the encoder.

5.4 Motion estimation

As explained in Section 5.5, "Encoder prediction loop," the compression algorithm depends on creating an estimate of the image being compressed, and subtracting from the image to be compressed the pixel values of the estimate or prediction. If the estimate is good, the subtraction will leave a very small residue to be transmitted; in fact, if the estimate or prediction were perfect, the difference would be zero for all the pixels in the frame of differences, and no new information would need to be sent (that condition can be approached for still images).

If the estimate is not close to zero for some pixels or many pixels, those differences represent information that needs to be transmitted so the decoder can reconstruct a correct image. The kinds of image sequences that cause large prediction differences include severe motion and/or sharp details.

5.4.1 Vector search algorithm

The video coding system uses motion compensated prediction as part of the data compression process. Thus macroblocks in the current frame of interest are predicted by macroblock-sized regions in previously transmitted frames. Motion compensation refers to the fact that the locations of the macroblock-sized regions in the reference frame can be offset to account for local motions. The macroblock offsets are known as *motion vectors*.

This standard does not specify how encoders should determine motion vectors. One possible approach might be to perform an exhaustive search to determine the vertical and horizontal offsets that minimize the total difference between the offset region in the reference frame and the macroblock in the frame to be coded.

5.4.2 Motion vector precision

The estimation of interframe displacement is calculated with half-pel precision, in both vertical and horizontal dimensions. That means the displaced macroblock from the previous frame can be displaced by non-integer displacements, and will require interpolation to compute the values of displaced picture elements at locations not in the original array of samples. Estimates for half-pel locations are computed by averages of adjacent sample values.

5.4.3 Motion vector coding

Motion vectors within a slice are differenced, so that the first value for a motion vector is transmitted directly, and the following sequence of motion vectors differences is sent using variable-length codes (VLC).

5.4.4 Estimation at frame boundaries

Motion vectors are constrained so that all pixels from the motion compensated prediction region in the reference picture fall within the picture boundaries.

5.5 Encoder prediction loop

The best way to understand the way the different algorithmic elements combine to achieve video compression is to examine the encoder prediction loop. This closed feedback loop, shown in the simplified block diagram of Figure 5.5, is the heart of the video compression for the ATV system.

5.5.1 Prediction loop block diagram

The prediction loop contains a prediction function that estimates, or predicts, the picture values of the next picture to be encoded in the sequence of successive pictures that constitute the TV program. This prediction is based on previous information that is available within the loop, derived from earlier pictures. The transmission of the predicted compressed information works because the very same information used to make the prediction is available also at the receiving decoder (barring transmission errors, which are expected to be infrequent within the coverage area).

The subtraction of the predicted picture values from the new picture to be coded is at the core of predictive coding. The goal is to do such a good job of predicting the new values that the result of the subtraction function at the beginning of the prediction loop is zero or close to zero for most of the time.

The prediction differences are computed separately for the luminance and two chrominance components before further processing.

As discussed under I-frames, there are times when prediction is not used, for part of a frame or for an entire frame. Those portions are said to be "intraframe" coded, while the portions that use the prediction from previous or future pictures are said to be "interframe" coded.
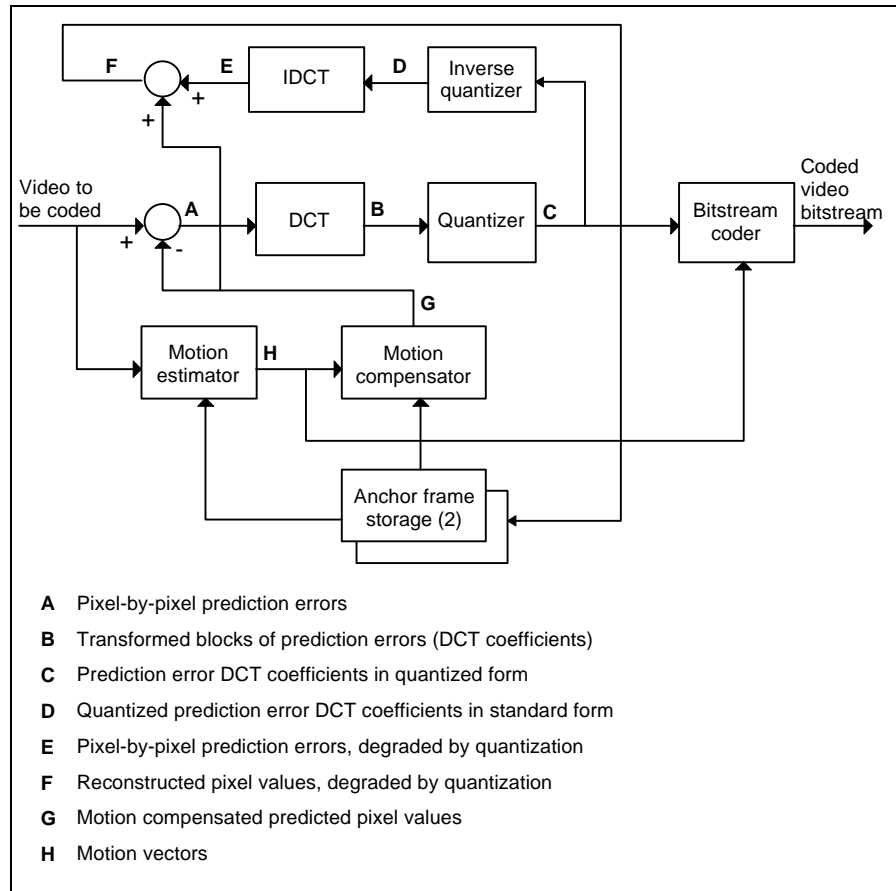
Figure 5.5. Encoder prediction loop.

5.5.1.1 Spatial transform block — DCT

The prediction differences (sometimes referred to as prediction errors) are grouped into 8-by-8 blocks and a spatial transform is applied to the blocks of difference values. In the intraframe case, the spatial transform is applied to the raw, undifferenced picture data. The luminance and two chrominance components are separately transformed. Since the chrominance data is sub-sampled vertically and horizontally, each 8-by-8 block of chrominance (Cb or Cb) data corresponds to a 16-by-16 macroblock of luminance data, which is not sub-sampled.

The spatial transform used is the discrete cosine transform, or DCT. The formula for transforming the data is given by:

$$F(u,v) = \frac{1}{4} C(u)C(v) \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y) \cos\left[\frac{(2x+1)u\boldsymbol{p}}{16}\right] \cos\left[\frac{(2y+1)v\boldsymbol{p}}{16}\right]$$

where x and y are pixel indices within an 8-by-8 block, u and v are DCT coefficient indices within an 8-by-8 block, and:

$$C(w) = \frac{1}{\sqrt{2}} \quad for \ w = 0$$

$$C(w) = 1 \qquad for \ w = 1, 2, \dots, 7$$

Thus an 8-by-8 array of numbers $f(x, y)$ is the input to a mathematical formula, and the output is an 8-by-8 array of different numbers, $F(u, v)$. The inverse transform[5] is given by:

$$f(x, y) = \frac{1}{4} \sum_{u=0}^{7} \sum_{v=0}^{7} C(u)C(v)F(u,v)\cos\left[\frac{(2x+1)u\boldsymbol{p}}{16}\right]\cos\left[\frac{(2y+1)v\boldsymbol{p}}{16}\right]$$

In principle, applying the inverse DCT transform to the transformed array would yield exactly the same array as the original. In that sense, transforming the data doesn't modify the data but merely represents the data in a different form.

The decoder uses the inverse transformation to approximately reconstruct the arrays that were transformed at the encoder, as part of the process of decoding the received compressed data. The approximation in that reconstruction is controlled in advance during the encoding process so as to minimize the visual effects of coefficient inaccuracies, while reducing the quantity of data that needs to be transmitted.

DCT transforms are discussed further in Section 5.7.

5.5.1.2 Quantizer

The process of transforming the original data organizes the information in a way that exposes the spatial frequency components of the images or image differences. Using knowledge about the response of the human visual system to different spatial frequencies, the encoder can selectively adjust the precision of transform coefficient representation. The goal is to include as much information about a particular spatial frequency as needed (and possible, given constraints on data transmission), but not to use more precision than is needed, based on visual perception criteria.

For example, in a portion of a picture that is very "busy" with much detail, imprecision in reconstructing spatial high frequency components in a small region might be masked by the picture's local "busy-ness." On the other hand, very precise representation and reconstruction of the average value or DC term of the DCT block (the F(0,0) term of the transformed coefficients represents the average of the original 64 coefficients and is referred to as the DC term) would be important in a smooth area of sky.

Recall that the DCT of each 8-by-8 block of pixel values produces an 8-by-8 array of DCT coefficients. The relative precision accorded to each of the 64 DCT coefficients can be selected according to its relative importance in human visual perception. The relative coefficient precision information is represented by a *quantizer matrix*, which is an

---

[5] The IDCT is required to conform to *IEEE Standard Specifications for the Implementation of 8x8 Inverse Discrete Cosine Transform*, Std 1180-1990, December 6, 1990.

8-by-8 array of values. Each value in the quantizer matrix represents the coarseness of quantization of the related DCT coefficient.

Two types of quantizer matrices are supported — one which is used for macroblocks which are intraframe coded, and the other which is used for non-intraframe coded macroblocks. The video coding system defines default values for both the intra-quantizer and the non-intra-quantizer matrices. Either or both of the quantizer matrices can be overridden at the picture level by transmitting the appropriate arrays of 64 values. Any quantizer matrix overrides stay in effect until the following sequence start code.

The transform coefficients, which represent the bulk of the actual coded video information, are quantized to various degrees of coarseness. As indicated above, the appearance of some portions of the picture will be more affected than others to the loss of precision through coefficient quantization. This phenomenon is exploited by the availability of the quantizer scale factor, which allows the overall level of quantization to vary for each macroblock. Thus entire macroblocks which are deemed to be visually less important can be quantized more coarsely, which results in decreasing the number of bits needed to represent the picture.

For each coefficient other than the DC coefficient of intraframe coded blocks, the quantizer scale factor is multiplied by the corresponding value in the appropriate quantizer matrix to form the quantizer step size. Quantization of the DC coefficients of intra-coded blocks is unaffected by the quantizer scale factor, and is only governed by the $(0, 0)$ element of the intra-quantizer matrix, which is always set to be 8 by ISO/IEC 13818-2.

Quantization is discussed further in Section 5.8.

5.5.1.3 Entropy coder

An important effect of the quantization of transform coefficients is that many coefficients will be rounded to zero after quantization. In fact, a primary method of controlling the encoded data rate is the control of quantization coarseness, since a coarser quantization leads to an increase in the number of zero-value quantized coefficients. Entropy coding is discussed in greater detail in Section 5.9.

5.5.1.4 Inverse quantizer

At the decoder the coded coefficients are decoded and an 8-by-8 block of quantized coefficients is reconstructed. Each of these 64 coefficients is *inverse quantized* according to the prevailing quantizer matrix, quantizer scale, and frame type. The result of inverse quantization is a block of 64 DCT coefficients.

5.5.1.5 Inverse spatial transform block — IDCT

The decoded and inverse quantized coefficients are organized as 8-by-8 blocks of DCT coefficients and the inverse discrete cosine transform (IDCT) is applied to each block. This results in a new array of pixel values, or pixel difference values that correspond to the output of the subtraction at the beginning of the prediction loop. If the prediction loop was in the interframe mode, the values will be pixel differences. If the

mode was in the intraframe mode, then the inverse transform will produce pixel values directly.

5.5.1.6 Motion compensator

If a portion of the image has not moved, then it is easy to see that a subtraction of the old portion from the new portion of the image will produce zero or nearly zero pixel differences, which is the goal of the prediction.

If there has been movement in the portion of the image under consideration, the direct pixel-by-pixel differences will in general not be zero, and might be statistically very large. However, the motion in most natural scenes is organized, and can in most cases be approximately represented locally as a translation. For this reason the video coding system allows for *motion compensated* prediction, whereby macroblock sized regions in the reference frame may be translated vertically and horizontally with respect to the macroblock being predicted, to compensate for local motion.

The pixel-by-pixel differences between the current macroblock and the motion compensated prediction are transformed by the DCT and quantized using the composition of the non-intra-quantizer matrix and the quantizer scale factor. The quantized coefficients are then coded.

5.5.1.7 Anchor frames

In the case of I-frames, the entire frame is encoded without reference to any other coded frames. P-frames are referenced to the most recently decoded I or P-frame). B-frames, however, permit the use of two frames as prediction references. One of the reference frames occurs earlier than the coded frame in display order (which can be used for forward prediction), and the other occurs later in display order (which can be used for backward prediction).

For a given macroblock within the B-frame, the encoder has four options. They are: forward prediction, backward prediction, bidirectional prediction, and intraframe coding. When bidirectional prediction is used, the forward and backward predictors are averaged and then subtracted from the target macroblock to form the prediction error. The prediction error is then transformed, quantized and transmitted in the usual manner. Note that both of the frames used as references in coding a B-frame are coded and transmitted prior to the coding of the actual B-frame. This results in the need for frame reordering within the decoder to produce the proper display order.

5.5.2 Dual prime prediction mode

The dual prime prediction mode is an alternative "special" prediction mode which is based on field-based motion prediction but requires fewer transmitted motion vectors than conventional field-based prediction. This mode of prediction is available only for interlaced material and only when the encoder configuration does not use B-frames. This mode of prediction may be particularly useful for improving encoder efficiency for low delay applications.

The basis of dual prime prediction is that field-based predictions of both fields in a macroblock are obtained by averaging two separate predictions which are predicted from the two nearest decoded fields in time. Each of the macroblock fields is predicted separately, although the four vectors (one pair per field) used for prediction are all derived from a single transmitted field-based motion vector. In addition to the single field-based motion vector, a small "differential" vector (limited to vertical and horizontal component values of +1, 0 and -1) is also transmitted for each macroblock. Together, these vectors are used to calculate the pairs of motion vectors for each macroblock. The first prediction in the pair is simply the transmitted field-based motion vector. The second prediction vector is obtained by combining the differential vector with a scaled version of the first vector. Once both predictions are obtained, a single prediction for each macroblock field is calculated simply by averaging each pel in the two original predictions. The final averaged prediction is then subtracted from the macroblock field being encoded.

5.5.3 Adaptive field/frame prediction mode

Interlaced pictures may be coded in either of two ways — either as two separate fields or as a single frame. When coded as separate fields, all of the codes for the first field are transmitted as a unit before the codes for the second field. When coded as a frame, information for both fields is coded for each macroblock.

When frame-based coding is used with interlaced pictures, each macroblock may be selectively coded using either field prediction or frame prediction. When frame prediction is used, a motion vector is applied to a picture region which is made up of both parity fields interleaved together. When field prediction is used, a motion vector is applied to a region made up of scan lines from a single field. Field prediction allows the selection of either parity field to be used as a reference for the field being predicted.

5.6 Image refresh

As described in the preceding sections, a given picture may be sent by describing the differences between it and one or two previously transmitted pictures. In order for this scheme to work, there must be some way for decoders to become initialized with a valid picture upon tuning into a new channel, or to become re-initialized with a valid picture after experiencing transmission errors. Additionally, it is necessary to limit the number of consecutive predictions which can be performed in a decoder to control the buildup of errors due to *IDCT mismatch*.

IDCT mismatch arises from the fact that the video coding system, by intention, does not completely specify the results of the IDCT operation.[6] It is thus possible for the reconstructed pictures in a decoder to "drift" away from those in the encoder if many successive predictions are used, even in the absence of transmission errors. The amount of drift is controlled by requiring that each macroblock be coded without prediction (intra-coded) at least once in any 132 consecutive frames.

---

[6] MPEG did not fully specify the results of the IDCT to allow for evolutionary improvements in implementations of this computationally intensive operation.

The process whereby a decoder becomes initialized or re-initialized with valid picture data without reference to previously transmitted picture information is termed *image refresh*. Image refresh is accomplished by the use of intraframe coded macroblocks. There are two general classes of image refresh which can be used either independently or jointly, periodic transmission of I-frames and progressive refresh.

5.6.1 Periodic transmission of I-frames

One approach to image refresh is to periodically code an entire frame using only intraframe coding. In this case the intra-coded frame is typically an I-frame.[7] The period between successive intra-coded frames may be constant or it may vary.

When a receiver tunes into a new channel where I-frame coding is used for image refresh it may perform the following steps:

- Ignore all data until receipt of the first sequence header

- Decode the sequence header and configure circuits based on sequence parameters

- Ignore all data until the next received I-frame

- Commence picture decoding and presentation

When a receiver processes data which contains uncorrectable errors in an I or P-frame there will typically be a propagation of picture errors due to the use of predictive coding. Pictures received after the error may be decoded incorrectly until an error-free I-frame is received.

5.6.2 Progressive refresh

An alternative method for accomplishing image refresh is to encode only a portion of each picture using the intraframe mode. In this case the intraframe coded regions of each picture should be chosen such that, over the course of a reasonable number of frames, all macroblocks are coded intraframe at least once. In addition constraints might be placed on motion vector values in order to avoid possible contamination of refreshed regions through predictions using unrefreshed regions in an uninitialized decoder.

5.7 Discrete cosine transform (DCT)

Predictive coding in the ATV compression algorithm exploits the temporal correlation in the sequence of image frames. Motion compensation is a refinement of that temporal prediction that allows the coder to account for apparent motions in the image that can be estimated. Aside from temporal prediction, another source of correlation that represents redundancy in the image data is the spatial correlation within an image frame or

---

[7] Note that MPEG allows a field-structured I-frame to consist of a first field which is coded with type I, and a second field which is coded with type P, using the first field as its prediction reference. Although prediction is used within the frame, no reference is made to previously transmitted frames.

field. This spatial correlation of images, including parts of images that contain apparent motion, can be accounted for by a spatial transform of the prediction differences. In the intraframe coding case (I-frames), where there is by definition no attempt at prediction, the spatial transform applies to the actual picture data. The effect of the spatial transform is to concentrate a large fraction of the signal energy in a few transform coefficients.

In order to exploit spatial correlation in intraframe and predicted portions of the image, the image prediction residual pixels are represented by their DCT coefficients. For typical images, a large fraction of the energy is concentrated in a few of these coefficients. This makes it possible to code only a few coefficients without seriously affecting the picture quality. The DCT is chosen because it has good energy compaction properties, and in addition, results in real coefficients, and there exist numerous fast computational algorithms for its implementation.

5.7.1 Blocks of 8-by-8 pixels

Theoretically, a DCT of larger size will outperform a DCT of smaller size in terms of coefficient decorrelation and block energy compaction. Better overall performance can be achieved, however, by subdividing the frame into many smaller regions each of which is individually processed. The motivation for this can be understood by the following. If we compute the DCT of the entire frame, we treat the whole frame equally. For a typical image, some regions contain a large amount of detail and other regions contain very little detail. By exploiting the changing characteristics of different images and of different portions of the same image, significant improvements in performance can be realized. In order to take advantage of the varying characteristics of the frame over its spatial extent, the frame is partitioned into blocks of 8-by-8 pixels. The blocks are then independently transformed and adaptively processed based on their local characteristics. The partitioning of the frame into small blocks before taking the transform not only allows spatially adaptive processing, but also reduces the computational and memory requirements. Partitioning the signal into small blocks before computing the DCT is referred to as the Block DCT.

An additional advantage of using the DCT domain representation is that the DCT coefficients contain information about the spatial frequency content of the block. By utilizing the spatial frequency characteristics of the human visual system, the precision with which the DCT coefficients are transmitted can be in accordance with their perceptual importance. This is achieved through the quantization of these coefficients, as explained in the following section.

5.7.2 Adaptive field/frame DCT

As noted above the DCT allows taking advantage of the typically high degree of spatial correlation in natural scenes. When coding interlaced pictures on a frame basis, however, it is possible that significant amounts of motion result in relatively low spatial correlation in some regions. This situation is accommodated by allowing the DCTs to be computed either on a field basis or on a frame basis. The decision to use field or frame-based DCT is made individually for each macroblock.

5.8 Adaptive quantization

The goal of video compression is to maximize the video quality at a given bit rate. This requires a wise distribution of the limited number of available bits. By exploiting the perceptual irrelevancy (as explained in Section 5.7.1) and statistical redundancy (as explained in Section 5.9) within the DCT domain representation, an appropriate bit allocation can yield significant improvements in performance. Quantization is performed to reduce the precision of the DCT coefficient values, and through quantization and codeword assignment, the actual bit rate compression is achieved. The quantization process is the source of virtually all of the loss of information in the compression algorithm. This is important, as it simplifies the design process and facilitates fine tuning of the system.

5.8.1 Adaptive step sizes

The degree of subjective picture degradation caused by coefficient quantization tends to depend on the nature of the scenery being coded. Within a given picture distortions of some regions may be less apparent than in others. The video coding system allows for the level of quantization to be adjusted for each macroblock in order to save bits, where possible, through coarse quantization.

5.8.2 Perceptual weighting

The human visual system is not uniformly sensitive to coefficient quantization error. Perceptual weighting of each source of coefficient quantization error is used to increase quantization coarseness in order to lower the bit rate. The amount of visible distortion resulting from quantization error for a given coefficient depends on the coefficient number, or frequency, the local brightness in the original image, and the duration or the temporal characteristic of the error. DC coefficient error results in mean value distortion for the corresponding block of pels which can expose block boundaries. This is more visible than higher frequency coefficient error which appears as noise or texture.

Displays and human visual systems exhibit non-uniform sensitivity to detail as a function of local average brightness. Loss of detail in dark areas of the picture is not as visible as it is in brighter areas. Another opportunity for bit savings is presented in textured areas of the picture where high frequency coefficient error is much less visible than in relatively flat areas. Brightness and texture weighting require analysis of the original image since these areas may be well predicted. Finally, distortion is easily masked by limiting its duration to one or two frames. This effect is most profitably used after scene changes where the first frame or two can be greatly distorted without perceptible distortion at normal speed.

When quantizing transform coefficients, the differing perceptual importance of the various coefficients can be exploited by "allocating the bits" to shape the quantization noise into the perceptually less important areas. This can be accomplished by varying the relative step-sizes of the quantizers for the different coefficients. The perceptually important coefficients may be quantized with a finer step size than the others. For

example, low spatial frequency coefficients may be quantized finely, while the less important high frequency coefficients may be quantized more coarsely. A simple method to achieve different step-sizes is to normalize or weight each coefficient based on its visual importance. All of the normalized coefficients may then be quantized in the same manner, such as rounding to the nearest integer (uniform quantization). Normalization or weighting effectively scales the quantizer from one coefficient to another. The video compression system utilizes perceptual weighting, where the different DCT coefficients are weighted according to a perceptual criterion prior to uniform quantization. The perceptual weighting is determined by quantizer matrices. The ATV video compression system allows for modifying the quantizer matrices before each picture.

5.8.3 Adaptive intra-quantizer and non-intra-quantizer matrices

The video coding syntax allows the quantizer matrices to be specified for every picture for improved coding efficiency. A certain probability distribution is associated with the variable-length codes (VLC) for quantized coefficients. Although one cannot change the VLC distribution to match the actual distribution of the data, the quantizer matrices can be adjusted to help match the distribution of the data to the distribution of the VLC. Over the course of encoding the frame data, the variance of each spatial frequency band may be calculated for both intra-data and non-intra-data. One method for choosing quantizer matrices involves applying upper and lower bounds per band to ensure reasonable operation in all cases.

Transmitting the quantizer matrices costs bits in the compressed data stream. If sent with every picture in the 60 fps progressive mode, the matrices consume 0.32% of the channel bandwidth. This modest amount of overhead can be reduced by updating the quantizer matrix less frequently, or only when the difference between the desired quantizer matrix and the prevailing quantizer matrix becomes significant.

Sufficient compression cannot be achieved unless a large fraction of the DCT coefficients are dropped and therefore not selected for transmission. The coefficients which are not selected are assumed to have zero value in the decoder.

The DC coefficients are coded differently to take advantage of high spatial correlation. For example, when intra-coded, the first DC coefficient in a slice is sent absolutely; the following DC coefficients are sent as differences.

5.9 *Entropy coding of video data*

Quantization creates an efficient discrete representation for the data to be transmitted. Codeword assignment takes the quantized values and produces a digital bit stream for transmission. Hypothetically, the quantized values could be simply represented using uniform or fixed-length codewords. Under this approach, every quantized value would be represented with the same number of bits. Greater efficiency, in terms of bit rate, can be achieved by employing entropy coding. Entropy coding attempts to exploit the statistical properties of the signal to be encoded. A signal, whether it is a pixel value or a transform coefficient, has a certain amount of information, or entropy, based on the probability of the different possible values or events occurring. For example, an event that

occurs infrequently conveys much more new information than one that occurs often. By realizing that some events occur more frequently than others, the average bit rate may be reduced.

5.9.1 Huffman coding

Huffman coding, which is utilized in the video compression system, is one of the most common entropy coding schemes. In Huffman coding, a code book is generated which can approach the minimum average description length (in bits) of events, given the probability distribution of all the events. Events which are more likely to occur will be assigned shorter length codewords while those which are less likely to occur will be assigned longer length codewords.

5.9.2 Run-length coding

In video compression, most of the transform coefficients are frequently quantized to zero. There may be a few non-zero low-frequency coefficients and a sparse scattering of non-zero high-frequency coefficients, but the great majority of coefficients may have been quantized to zero. To exploit this phenomenon the two-dimensional array of transform coefficients is reformatted and prioritized into a one-dimensional sequence through either a zigzag or alternate scanning process. This results in most of the important non-zero coefficients (in terms of energy and visual perception) being grouped together early in the sequence. They will be followed by long runs of coefficients that are quantized to zero. These zero-valued coefficients can be efficiently represented through run-length encoding. In run-length encoding, the number (run) of consecutive zero coefficients before a non-zero coefficient is encoded, followed by the non-zero coefficient value. The run-length and the coefficient value can be entropy coded, either separately or jointly. The scanning separates most of the zero and the non-zero coefficients into groups, thereby enhancing the efficiency of the run-length encoding process. Also, a special end-of-block (EOB) marker is used to signify when all of the remaining coefficients in the sequence are equal to zero. This approach can be extremely efficient, yielding a significant degree of compression.

5.9.3 Zigzag scan and alternate scan

As indicated above, the array of 64 DCT coefficients is arranged in a one-dimensional vector before run-length/amplitude code-word assignment. Two different one-dimensional arrangements, or *scan types*, are allowed, which are generally referred to as zigzag scan (shown in Figure 5.6a) and alternate scan (shown in Figure 5.6b). The scan type is specified before coding each picture, and is permitted to vary from picture-to-picture.

5.10 Channel buffer

Whenever entropy coding is employed, the bit rate produced by the encoder is variable and is a function of the video statistics. Since the bit rate permitted by the transmission system is less than the peak bit rate produced by the variable-length coder, a

channel buffer is necessary at the decoder. The buffering must be carefully designed. What is needed is some form of buffer control that would allow efficient allocation of bits to encode the video while ensuring that no overflow or underflow occurs.

(a) Zigzag scan ordering of coefficients

Blocks of DCT coefficients

(b) Alternate scan ordering of coefficients
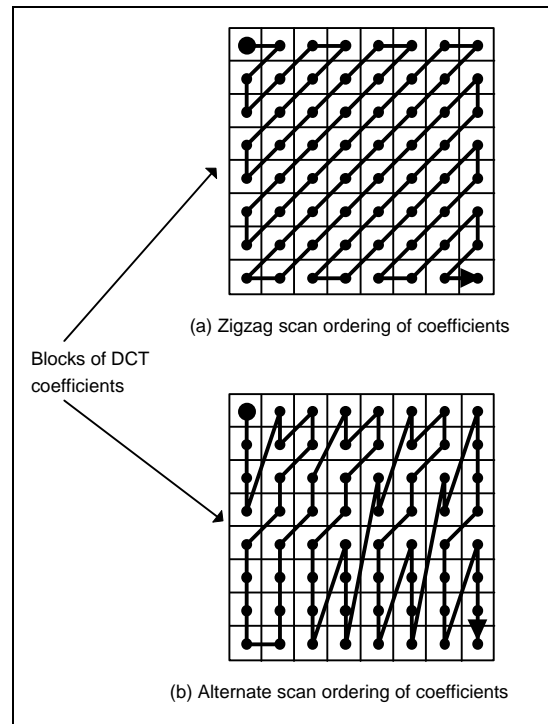
Figure 5.6. Scanning of coefficient blocks.

The buffer control typically involves a feedback mechanism to the compression algorithm whereby the amplitude resolution (quantization) and/or spatial, temporal and color resolution may be varied in accordance with the instantaneous bit rate requirements. If the bit rate decreases significantly, a finer quantization can be performed to increase it.

As indicated above the peak bit rate produced by variable-length coding will typically fluctuate dramatically, and will frequently exceed the desired peak channel bandwidth allocation.[8] This situation is handled by the presence of a channel buffer at each decoder for temporary storage of the coded bit stream. The Digital Television Standard specifies a channel buffer size of 8 Mbits.

A model buffer is defined in the video coding system as a reference for manufacturers of both encoders and decoders to ensure interoperability. An encoder will control its production of bits so that the model buffer does not overflow or underflow.[9]

---

[8] Note that the average bit rate, by definition, cannot be permitted to exceed the peak allocated channel bandwidth.

[9] Buffer underflow is actually permitted in the case of low-delay bit streams which do not contain B-frames. In such cases underflow might occur due to an unusually difficult picture which requires a particularly large number of bits. The result of underflow is the repeat of one or more pictures, which are presented in lieu of pictures that were skipped in the encoder.

In order to avoid overflow or underflow of the model buffer, an encoder may maintain measures of buffer occupancy and scene complexity. When the encoder needs to reduce the number of bits produced, it may do so by increasing the general value of the quantizer scale, which will increase picture degradation. When it is able to produce more bits, it may decrease the quantizer scale thus decreasing picture degradation.

5.11 Interface to system multiplexer

The bit stream produced by the video encoder is passed to the transport encoding system for multiplexing with audio and ancillary data, "lip-synch", and scheduling for delivery.

5.12 Decoder block diagram

As shown in Figure 5.7 the video decoder contains elements which invert, or undo, the processing performed in the encoder. The incoming coded video bit stream is placed in the channel buffer. Bits are removed from the channel buffer by a variable-length decoder (VLD).

C	Prediction error DCT coefficients in quantized form
D	Quantized prediction error DCT coefficients in standard form
E	Pixel-by-pixel prediction errors, degraded by quantization
F	Reconstructed pixel values, degraded by quantization
G	Motion compensated predicted pixel values
H	Motion vectors
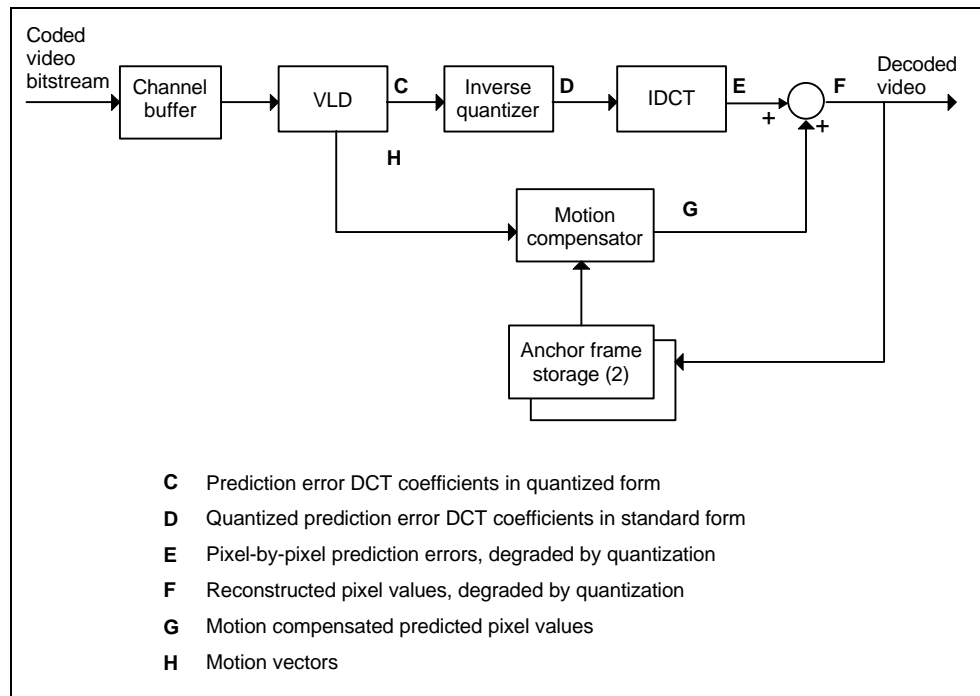
Figure 5.7. Decoder block diagram.

The VLD reconstructs 8-by-8 arrays of quantized DCT coefficients by decoding run-length/amplitude codes and appropriately distributing the coefficients according to the scan type used. These coefficients are de-quantized and transformed by the inverse discrete cosine transform (IDCT) to obtain pixel values or prediction errors.

In the case of interframe prediction the decoder uses the received motion vectors to perform the same prediction operation as was done in the encoder. The prediction

errors are summed with the results of motion compensated prediction to produce pixel values.

5.12.1 Error concealment capability

When transmission errors occur, a decoder may act to minimize the perceived picture degradation. This process is discussed in more detail in Section 10.2.6.

5.12.2 Frame store for decoded pictures

As described above, pixel values are decoded from the incoming bit stream. In the case of decoded anchor frames (I or P-frames) these values must be stored in a frame buffer for subsequent use as prediction references. When B-frames are used the anchor frame storage also allows for the necessary frame re-ordering for display.

5.13 Concatenated sequences

The MPEG-2 standard which underlies the Digital Television Standard clearly specifies the behavior of a compliant video decoder when processing a single video sequence. A coded video sequence commences with a sequence header, may contain some repeated sequence headers and one or more coded pictures, and is terminated by an end-of-sequence code. A number of parameters are specified in the sequence header that are required to remain constant throughout the duration of the sequence. The sequence level parameters include, but are not limited to:

- Horizontal and vertical resolution

- Frame rate

- Aspect ratio

- Chroma format

- Profile and level

- All-progressive indicator

- Video buffering verifier (VBV) size

- Maximum bit rate

It is envisioned that it will be common for coded bit streams to be spliced for editing, insertion of commercial advertisements, and other purposes in the video production and distribution chain. If one or more of the sequence level parameters differ between the two bit streams to be spliced, then an end-of-sequence code must be inserted to terminate the first bit stream and a new sequence header must exist at the start of the second bit stream. Thus the situation of concatenated video sequences arises.

While the MPEG-2 standard specifies the behavior of video decoders when processing a single sequence, it does not place any requirements on the handling of concatenated sequences. Specification of the decoding behavior in the former case is feasible because the MPEG-2 standard places constraints on the construction and coding

of individual sequences. These constraints prohibit channel buffer overflow and coding the same field parity for two consecutive fields. The MPEG-2 standard does not prohibit these situations at the junction between two coded sequences and it likewise does not specify the behavior of decoders in this case.

While it is recommended, the Digital Television Standard does not require the production of well-constrained concatenated sequences. Well-constrained concatenated sequences are defined as having the following characteristics:

- The extended decoder buffer never overflows, and may only underflow in the case of low-delay bit streams. Here "extended decoder buffer" refers to the natural extension of the MPEG-2 decoder buffer model to the case of continuous decoding of concatenated sequences.

- When field parity is specified in two coded sequences which are concatenated, the parity of the first field in the second sequence is opposite that of the last field in the first sequence.

- Whenever a progressive sequence is inserted between two interlaced sequences, the exact number of progressive frames shall be such that the parity of the interlaced sequences is preserved as if no concatenation had occurred.

5.14 Guidelines for refreshing

While the Digital Television Standard does not require refreshing at less than the intra-macroblock refresh rate as defined in IEC/ISO 13818-2, the following is recommended:

- In a system which uses periodic transmission of I-frames for refreshing, the frequency of occurrence of I-frames will determine the channel-change time performance of the system. In this case, it is recommended that I-frames be sent at least once every 0.5 second in order to have acceptable channel-change performance. It is recommended also that sequence layer information be sent before every I-frame.

- In order to spatially localize errors due to transmission, intra-coded slices should contain fewer macroblocks than the maximum number allowed by the Standard. It is recommended that there be four to eight slices in a horizontal row of intra-coded macroblocks for the intra-coded slices in the I-frame refresh case as well as for the intraframe coded regions in the progressive refresh case. The size of non-intra-coded slices can be larger than that of intra-coded slices.

6. AUDIO SYSTEMS

This section describes the audio coding technology and gives guidelines as to its use. Information of interest to both broadcasters (and other program providers) and receiver manufacturers is included. The audio system is fully specified in Annex B of the Digital Television Standard and is based on the Digital Audio Compression (AC-3) Standard, with some limitations on bit rate, sampling rate, and audio coding mode.

6.1 Audio system overview

As illustrated in Figure 6.1, the audio subsystem comprises the audio encoding/decoding function and resides between the audio inputs/outputs and the transport subsystem. The audio encoder(s) is (are) responsible for generating the audio elementary stream(s) which are encoded representations of the baseband audio input signals. The flexibility of the transport system allows multiple audio elementary streams to be delivered to the receiver. At the receiver, the transport subsystem is responsible for selecting which audio streams(s) to deliver to the audio subsystem. The audio subsystem is responsible for decoding the audio elementary stream(s) back into baseband audio.
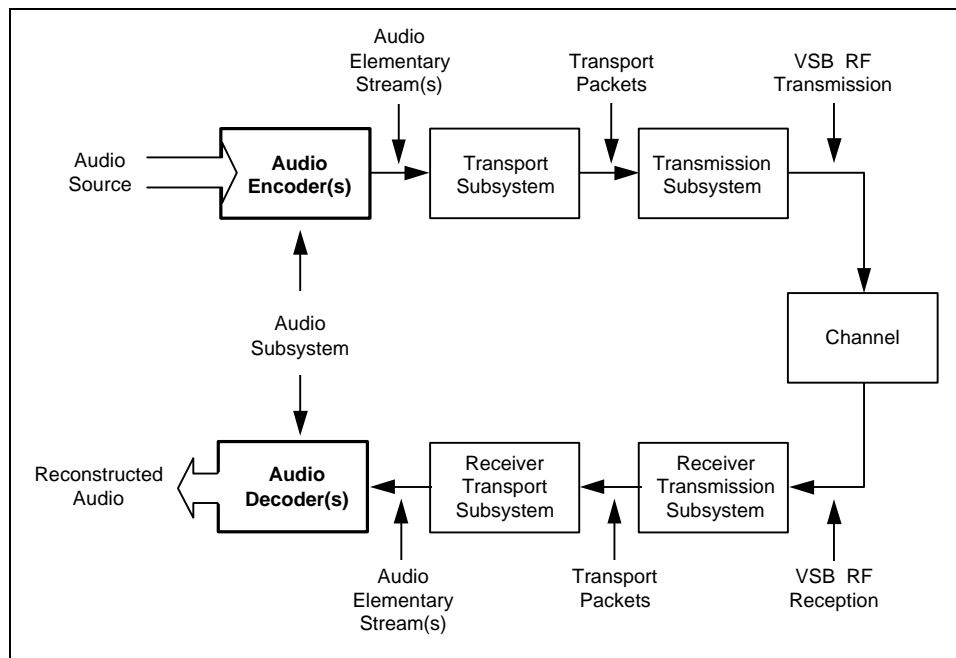
Figure 6.1. Audio subsystem within the digital television system.

An audio program source is encoded by a digital television audio encoder. The output of the audio encoder is a string of bits that represent the audio source, and is referred to as an *audio elementary stream*. The transport subsystem packetizes the audio data into PES packets which are then further packetized into transport packets. The transmission subsystem converts the transport packets into a modulated RF signal for transmission to the receiver. At the receiver, the received signal is demodulated by the receiver transmission subsystem. The receiver transport subsystem converts the received

audio packets back into an audio elementary stream which is decoded by the digital television audio decoder. The partitioning shown is conceptual, and practical implementations may differ. For example, the transport processing may be broken into two blocks; one to perform PES packetization, and the second to perform transport packetization. Or, some of the transport functionality may be included in either the audio coder or the transmission subsystem.

6.2 Audio encoder interface

The audio system accepts baseband audio inputs with up to six audio channels per audio program bit stream. The channelization is consistent with ITU-R Recommendation BS-775, "*Multi-channel stereophonic sound system with and without accompanying picture*". The six audio channels are: Left, Center, Right, Left Surround, Right Surround, and Low Frequency Enhancement (LFE). Multiple audio elementary bit streams may be conveyed by the transport system.

The bandwidth of the LFE channel is limited to 120 Hz. The bandwidth of the other (main) channels is limited to 20 kHz. Low frequency response may extend to DC, but is more typically limited to approximately 3 Hz (-3 dB) by a DC blocking high-pass filter. Audio coding efficiency (and thus audio quality) is improved by removing DC offset from audio signals before they are encoded.

6.2.1 Input source signal specification

Audio signals which are input to the audio system may be in analog or digital form.

6.2.1.1 High-pass filtering

Audio signals should have any DC offset removed before being encoded. If the audio encoder does not include a DC blocking high-pass filter, the audio signals should be high-pass filtered before being applied to the audio encoder.

6.2.1.2 Analog input

For analog input signals, the input connector and signal level are not specified. Conventional broadcast practice may be followed. One commonly used input connector is the 3-pin XLR female (the incoming audio cable uses the male connector) with pin 1 ground, pin 2 hot or positive, and pin 3 neutral or negative.

6.2.1.3 Digital input

For digital input signals, the input connector and signal format are not specified. Commonly used formats such as the AES 3-1992 two-channel interface may be used. When multiple two-channel inputs are used, the preferred channel assignment is:

> Pair 1:   Left, Right
>
> Pair 2:   Center, LFE
>
> Pair 3:   Left Surround, Right Surround

6.2.1.4 Sampling frequency

The system conveys digital audio sampled at a frequency of 48 kHz, locked to the 27 MHz system clock. If analog signal inputs are employed, the A/D converters should sample at 48 kHz. If digital inputs are employed, the input sampling rate shall be 48 kHz, or the audio encoder shall contain sampling rate converters which convert the sampling rate to 48 kHz. The sampling rate at the input to the audio encoder must be locked to the video clock for proper operation of the audio subsystem.

6.2.1.5 Resolution

In general, input signals should be quantized to at least 16-bit resolution. The audio compression system can convey audio signals with up to 24-bit resolution.

6.2.2 Output signal specification

Conceptually, the output of the audio encoder is an elementary stream which is formed into PES packets within the transport subsystem. It is possible that digital television systems will be implemented wherein the formation of audio PES packets takes place within the audio encoder. In this case, the output(s) of the audio encoder(s) would be PES packets. Physical interfaces for these outputs (elementary streams and/or PES packets) may be defined as voluntary industry standards by SMPTE or other standards organizations.

6.3 AC-3 digital audio compression

6.3.1 Overview and basics of audio compression

The audio compression system conforms with the Digital Audio Compression (AC-3) Standard specified in ATSC Doc. A/52. The audio compression system is considered a constrained subset of that Standard. The constraints are specified in Annex B of the Digital Television Standard. By conforming with the standardized syntax in ATSC Doc. A/52, the system employs an audio compression system which is interoperable across many different media, and is appropriate for use in a multitude of applications.

A major objective of audio compression is to represent an audio source with as few bits as possible, while preserving the level of quality required for the given application. Audio compression has two major applications. One is efficient utilization of channel bandwidth for video transmission systems. The other is reduction of storage requirements. Both of these applications apply to the digital television system.

The audio compression system consists of three basic operations, as shown in Figure 6.2. In the first stage, the representation of the audio signal is changed from the time domain to the frequency domain, which is a more efficient domain in which to perform psychoacousticly based audio compression. The resulting frequency domain coefficients are what are then encoded. The frequency domain coefficients may be coarsely quantized because the resulting quantizing noise will be at the same frequency as the audio signal, and relatively low signal to noise ratios are acceptable due to the phenomena of

psychoacoustic masking. The bit allocation operation determines, based on a psychoacoustic model of human hearing, what actual SNR is acceptable for each individual frequency coefficient. Finally, the frequency coefficients are coarsely quantized to the necessary precision and formatted into the audio elementary stream. The basic unit of encoded audio is the AC-3 sync frame, which represents 1536 audio samples. Each sync frame of audio is a completely independent encoded entity. The elementary bit stream contains the information necessary to allow the audio decoder to perform the identical (to the encoder) bit allocation. This allows the decoder to unpack and de-quantize the elementary bit stream frequency coefficients, resulting in the reconstructed frequency coefficients. The synthesis filterbank is the inverse of the analysis filterbank, and converts the reconstructed frequency coefficients back into a time domain signal.
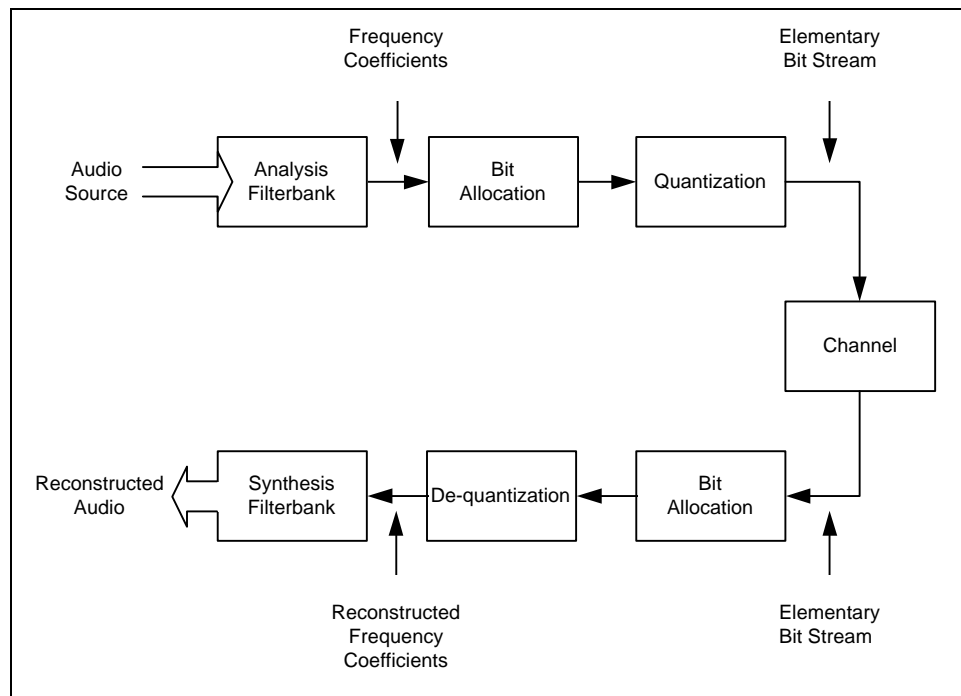
Figure 6.2. Overview of audio compression system.

6.3.2 Transform filter bank

The process of converting the audio from the time domain to the frequency domain requires that the audio be blocked into overlapping blocks of 512 samples. For every 256 new audio samples, a 512 sample block is formed from the 256 new samples, and the 256 previous samples. Each audio sample is represented in two audio blocks, and thus the number of samples to be processed initially is doubled. The overlapping of blocks is necessary in order to prevent audible blocking artifacts. New audio blocks are formed every 5.33 ms. A group of 6 blocks are coded into one AC-3 sync frame.

6.3.2.1 Window function

Prior to being transformed into the frequency domain, the block of 512 time samples is windowed. The windowing operation involves a vector multiplication of the 512 point block with a 512 point window function. The window function has a value of 1.0 in its center, and tapers down to almost zero at its ends. The shape of the window function is such that the overlap/add processing at the decoder will result in a reconstruction free of blocking artifacts. The window function shape also determines the shape of each individual filterbank filter.

6.3.2.2 Time division aliasing cancellation transform

The analysis filterbank is based on the fast Fourier transform. The particular transformation employed is the oddly stacked time domain aliasing cancellation (TDAC) transform. This particular transformation is advantageous because it allows the 100% redundancy which was introduced in the blocking process to be removed. The input to the TDAC transform is 512 windowed time domain points, and the output is 256 frequency domain coefficients.

6.3.2.3 Transient handling

When extreme time domain transients exist (such as an impulse or a castanet click), there is a possibility that quantization error, incurred in coarsely quantizing the frequency coefficients of the transient, will become audible due to time smearing. The quantization error within a coded audio block is reproduced throughout the block. It is possible for the portion of the quantization error which is reproduced prior to the impulse to be audible. Time smearing of quantization noise may be reduced by altering the length of the transform which is performed. Instead of a single 512 point transform, a pair of 256 point transforms may be performed, one on the first 256 windowed samples, and one on the last 256 windowed samples. A transient detector in the encoder determines when to alter the transform length. The reduction in transform length prevents quantization error from spreading more than a few milliseconds in time, which is adequate to prevent its audibility.

6.3.3 Coded audio representation

The frequency coefficients which result from the transformation are converted to a binary floating point notation. The scaling of the transform is such that all values are smaller than 1.0. An example value in binary notation (base 2) with 16-bit precision would be:

$$0.0000\ 0000\ 1010\ 1100_2$$

The number of leading zeroes in the coefficient, 8 in this example, becomes the raw exponent. The value is left shifted by the exponent, and the value to the right of the decimal point (1010 1100) becomes the normalized mantissa to be coarsely quantized. The exponents and the coarsely quantized mantissas are encoded into the bit stream.

6.3.3.1 Exponent coding

Some processing is applied to the raw exponents in order to reduce the amount of data required to encode them. First, the raw exponents of the 6 blocks to be included in a single AC-3 sync frame are examined for block-to-block differences. If the differences are small, a single exponent set is generated which is useable by all 6 blocks, thus reducing the amount of data to be encoded by a factor of 6. If the exponents undergo significant changes within the frame, then exponent sets are formed over blocks where the changes are not significant. Due to the frequency response of the individual filters in the analysis filter bank, exponents for adjacent frequencies rarely differ by more than ±2. To take advantage of this fact, exponents are encoded differentially in frequency. The first exponent is encoded as an absolute, and the difference between the current exponent and the following exponent is then encoded. This reduces the exponent data rate by a factor of 2. Finally, where the spectrum is relatively flat, or an exponent set only covers 1-2 blocks, differential exponents may be shared across 2 or 4 frequency coefficients, for an additional savings of a factor of 2 or 4.

The final coding efficiency for exponents is typically 0.39 bits/exponent (or 0.39 bits/sample since there is an exponent for each audio sample). Exponents are only coded up to the frequency needed for the perception of full frequency response. Typically, the highest audio frequency component in the signal which is audible is at a frequency lower than 20 kHz. In the case that signal components above 15 kHz are inaudible, only the first 75% of the exponent values are encoded, reducing the exponent data rate to <0.3 bits/sample.

The exponent processing changes the exponent values from their original values. The encoder generates a local representation of the exponents which is identical to the decoded representation which will be used by the decoder. The decoded representation is then used to shift the original frequency coefficients to generate the normalized mantissas which are quantized.

6.3.3.2 Mantissas

The frequency coefficients produced by the analysis filterbank have useful precision dependent on the wordlength of the input PCM audio samples, and the precision of the transform computation. Typically this precision is on the order of 16-18 bits, but may be as high as 24 bits. Each normalized mantissa is quantized to a precision between 0 and 16 bits. The goal of audio compression is to maximize the audio quality at a given bit rate. This requires an optimum (or near optimum) allocation of the available bits to the individual mantissas.

6.3.4 Bit allocation

The number of bits allocated to each individual mantissa value is determined by the bit allocation routine. The identical core routine is run in both the encoder and the decoder, so that each generates the identical bit allocation.

6.3.4.1 Backward adaptive

The core bit allocation algorithm is considered backwards adaptive, in that some of the encoded audio information within the bit stream (fed back into the encoder) is used to compute the final bit allocation. The primary input to the core allocation routine is the decoded exponent values, which give a general picture of the signal spectrum. From this version of the signal spectrum, a masking curve is calculated. The calculation of the masking model is based on a model of the human auditory system. The masking curve indicates, as a function of frequency, the level of quantizing error which may be tolerated. Subtraction (in the log power domain) of the masking curve from the signal spectrum yields the required SNR as a function of frequency. The required SNR values are mapped into a set of bit allocation pointers (baps) which indicate which quantizer to apply to each mantissa.

6.3.4.2 Forward adaptive

The AC-3 encoder may employ a more sophisticated psychoacoustic model than that used by the decoder. The core allocation routine used by both the encoder and the decoder makes use of a number of adjustable parameters. If the encoder employs a more sophisticated psychoacoustic model than that of the core routine, the encoder may adjust these parameters so that the core routine produces a better result. The parameters are inserted into the bit stream by the encoder and fed forward to the decoder.

In the event that the available bit allocation parameters do not allow the ideal allocation to be generated, the encoder can insert explicit codes into the bit stream to alter the computed masking curve, and thus the final bit allocation. The inserted codes indicate changes to the base allocation, and are referred to as delta bit allocation codes.

6.3.5 Rematrixing

When the AC-3 coder is operating in a two-channel stereo mode, an additional processing step is inserted in order to enhance interoperability with Dolby Surround 4-2-4 matrix encoded programs. The extra step is referred to as *rematrixing*.

The signal spectrum is broken into four distinct rematrixing frequency bands. Within each band, the energy of the Left, Right, Sum, and Difference signals are determined. If the largest signal energy is in the Left or Right channels, the band is encoded normally. If the dominant signal energy is in the Sum or Difference channel, then those channels are encoded instead of the Left and Right channels. The decision as to whether to encode Left and Right, or Sum and Difference is made on a band-by-band basis and is signaled to the decoder in the encoded bit stream.

6.3.6 Coupling

In the event that the number of bits required to encode the audio signals transparently exceeds the number of bits which are available, the encoder may invoke coupling. Coupling involves combining the high frequency content of individual channels and sending the individual channel signal envelopes along with the combined coupling channel. The psychoacoustic basis for coupling is that within narrow frequency bands the

human ear detects high frequency localization based on the signal envelope rather than the detailed signal waveform.

The frequency above which coupling is invoked, and the channels which participate in the process, are determined by the AC-3 encoder. The encoder also determines the frequency banding structure used by the coupling process. For each coupled channel and each coupling band, the encoder creates a sequence of coupling coordinates. The coupling coordinates for a particular channel indicate what fraction of the common coupling channel should be reproduced out of that particular channel output. The coupling coordinates represent the individual signal envelopes for the channels. The encoder determines the frequency with which coupling coordinates are transmitted. When coupling is in use, coupling coordinates are always sent in block 0 of a frame. If the signal envelope is steady, the coupling coordinates do not need to be sent every block, but can be reused by the decoder until new coordinates are sent. The encoder determines how often to send new coordinates, and can send them as often as every block (every 5.3 ms).

6.4 Bit stream syntax

6.4.1 Sync frame

The audio bit stream consists of a repetition of audio frames which are referred to as AC-3 sync frames. Shown in Figure 6.3, each AC-3 sync frame is a self contained entity consisting of synchronization information (SI), bit stream information (BSI), 32 ms of encoded audio, and a CRC error check code. Every sync frame is the same size (number of bits) and contains six encoded audio blocks. The sync frame may be considered an audio access unit. Within SI is a 16-bit sync word, an indication of audio sample rate (48 kHz for the digital television system), and an indication of the size of the audio frame (which indicates bit rate).
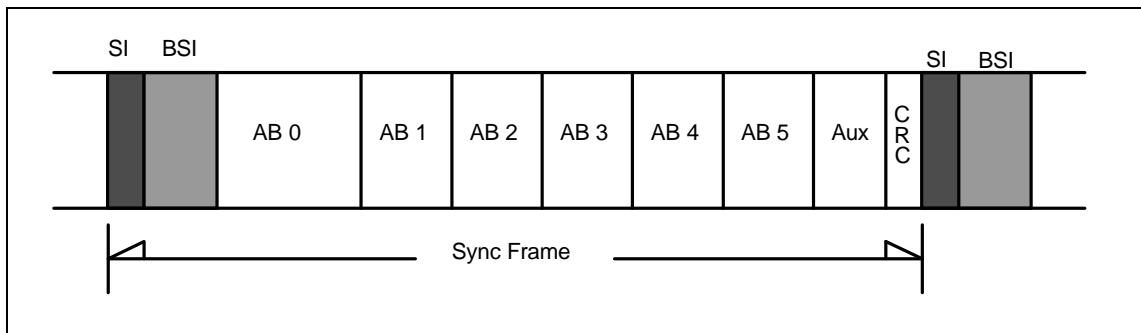
Figure 6.3. AC-3 synchronization frame.

6.4.2 Splicing, insertion

The ideal place to splice encoded audio bit streams is at the boundary of a sync frame. If a bit stream splice is performed at the sync frame boundary, the audio decoding will proceed without interruption. If a bit stream splice is performed randomly, there will be an audio interruption. The frame which is incomplete will not pass the decoder's error

detection test and this will cause the decoder to mute. The decoder will not find sync in its proper place in the next frame, and will enter a sync search mode. Once the sync code of the new bit stream is found, synchronization will be achieved, and audio reproduction may begin once again. The outage will be on the order of two frames, or about 64 ms. Due to the windowing process of the filterbank, when the audio goes to mute there will be a gentle fade down over a period of 2.6 ms. When the audio is recovered, it will fade up over a period of 2.6 ms. Except for the approximately 64 ms of time during which the audio is muted, the effect of a random splice of an AC-3 elementary stream is relatively benign.

6.4.3 Error detection codes

Each AC-3 sync frame ends with a 16-bit CRC error check code. The decoder may use this code to determine whether a frame of audio has been damaged or is incomplete. Additionally, the decoder may make use of error flags provided by the transport system. In the case of detected errors, the decoder may try to perform error concealment, or may simply mute.

6.5 Loudness and dynamic range

6.5.1 Loudness normalization

It is important for the digital television system to provide uniform subjective loudness for all audio programs. Consumers find it very annoying when audio levels fluctuate between broadcast channels (observed when channel hopping), or between program segments on a particular channel (commercials much louder than the entertainment). One element which is found in most audio programming is the human voice. Achieving an approximate level match for dialogue (spoken in a normal voice, not shouting or whispering) amongst all audio programming is a desirable goal. The AC-3 audio system provides syntactical elements which make this goal achievable.

There is (currently) no regulatory limit as to how loud dialogue may be in an encoded bit stream. Since the digital audio coding system can provide more than 100 dB of dynamic range, there is no technical reason for dialogue to be encoded anywhere near 100% as is commonly done in NTSC television. However, there is no assurance that all program channels, or all programs or program segments on a given channel, will have dialogue encoded at the same (or even similar) level. Lacking a uniform coding level for dialogue (which would imply a uniform headroom available for all programs) there would be inevitable audio level fluctuations between program channels or even between program segments.

Encoded AC-3 elementary bit streams are tagged with an indication (dialnorm) of the subjective level at which dialogue has been encoded. Different audio programs may be encoded with differing amounts of headroom above the level of dialogue in order to allow for dynamic music and sound effects. The digital television receiver (and all AC-3 decoders) are able to use the value of dialnorm to adjust the reproduced level of audio programs so that different received programs have their spoken dialogue reproduced at a

uniform level. Some receiver designs may even offer the listener an audio volume control calibrated in absolute sound pressure level. The listener could dial up the desired SPL for dialogue, and the receiver would scale the level of every decoded audio program so that the dialogue is always reproduced at the desired level.

The BSI portion of the sync frame contains the 5-bit dialnorm field which indicates the level of average spoken dialogue within the encoded audio program. The indication is relative to the level of a full scale 1 kHz sinewave. The measurement of dialogue level is done by a method which gives a subjectively accurate value. The measurement of subjective loudness is not an exact science, and new measurement techniques will be developed in the future. A measurement method which is currently available and quite useful is the "A" weighted integrated measurement (L_{Aeq}). This measurement method should be used until a more accurate method is standardized and available in practical equipment. Any new measurement methodology which is developed should be normalized (scaled) so that its results generally match those of the L_{Aeq} method.

It is important for broadcasters and others who deliver encoded audio bit streams to ensure that the value of dialnorm is correct. Incorrect values will lead to unwelcome level fluctuations in consumer homes. The worst case example of incorrect (or abusive) setting of dialnorm would be to broadcast a commercial message which indicates dialogue at a low level, but which is actually encoded with dialogue at full level. This would result in the commercial message being reproduced at the same level as a full scale explosion in a feature film (>100 dB SPL in some home theatre setups!). If such abuses occur, there may be a demand for regulatory enforcement of audio levels. Fortunately, bit streams which contain an incorrect value of dialnorm are easily corrected by simply changing the value of the 5-bit dialnorm field in the BSI header.

There are two primary methods which broadcast organizations may employ to ensure that the value of dialnorm is set correctly. The first method is to select a suitable dialogue level for use with all programming and conform all baseband audio programs to this level prior to AC-3 encoding. Then the value of dialnorm can be set to one common value for all programs which are encoded. Conforming all programs to a common dialogue level may mean that for some programs the audio level never approaches 100% digital level (since they have to be reduced in gain), while for other programs non-reversible (by the receiver) limiting must be engaged in order to prevent them from going over digital 100% (since they had to be increased in gain). Pre-encoded programs can be included in broadcasts if they have had the value of dialnorm correctly set, and the receiver will then conform the level.

The second (and generally preferred) method is to let all programming enter the encoder at full level, and correct for differing levels by adjusting the encoded value of dialnorm to be correct for each program. In this case, the conforming to a common level is done at the receiver. This method will become more practical as computer remote control of the encoding equipment becomes commonplace. The data base for each audio program to be encoded would include (along with items such as number of channels, language, etc.) the dialogue level. The master control computer would then communicate the value of dialogue level to the audio encoder which would then place the appropriate value in the bit stream.

In the case where a complete audio program is formed from the combination of a main and an associated service, each of the two services being combined will have a value of dialnorm, and the values may not be identical. In this case, the value of dialnorm in each bit stream should be used to alter the level of the audio decoded from that bit stream, prior to the mixing process which combines the audio from the two bit streams to form the complete audio program.

6.5.2 Dynamic range compression

It is common practice for high quality programming to be produced with wide dynamic range audio, suitable for the highest quality audio reproduction environment. Broadcasters, serving a wide audience, typically process audio in order to reduce its dynamic range. The processed audio is more suitable for the majority of the audience which does not have an audio reproduction environment which matches that of the original audio production studio. In the case of NTSC, all viewers receive the same audio with the same dynamic range, and it is impossible for any viewer to enjoy the original wide dynamic range audio production.

The audio coding system provides an embedded dynamic range control system which allows a common encoded bit stream to deliver programming with a dynamic range appropriate for each individual listener. A dynamic range control value (dynrng) is provided in each audio block (every 5 ms). These values are used by the audio decoder in order to alter the level of the reproduced audio for each audio block. Level variations of up to ± 24 dB may be indicated. The values of dynrng are generated in order to provide a subjectively pleasing but restricted dynamic range. The unaffected level is dialogue level. For sounds louder than dialogue, values of dynrng will indicated gain reduction. For sounds quieter than dialogue, values of dynrng will indicate a gain increase. The broadcaster is in control of the values of dynrng, and can supply values which generated the amount of compression which the broadcaster finds appropriate. The use of dialogue level as the unaffected level further improves loudness uniformity.

By default, the values of dynrng will be used by the audio decoder. The receiver will thus reproduce audio with a reduced dynamic range, as intended by the broadcaster. The receiver may also offer the viewer the option to scale the value of dynrng in order to reduce the effect of the dynamic range compression which was introduced by the broadcaster. In the limiting case, if the value of dynrng is scaled to zero, then the audio will be reproduced with its full original dynamic range. The optional scaling of dynrng can be done differently for values indicating gain reduction (which reduces the levels of loud sounds) and for values indicating gain increases (which makes quiet sounds louder). Thus the viewer may be given independent control of the amount of compression applied to loud and quiet sounds. Therefore, while the broadcaster may introduce dynamic range compression to suit the needs of most of the audience, individual listeners may have the option to choose to enjoy the audio program with more or all of its original dynamic range intact.

The dynamic range control words may be generated by the AC-3 encoder. They may also be generated by a processor located before or after the encoder. If the dynamic range processor is located prior to the encoder, there is a path to convey the dynamic

range control words from the processor to the encoder, or to a bit stream processor, so that the control words may be inserted into the bit stream. If the dynamic range processor is located after the encoder, it can act upon an encoded stream and directly insert the control words without altering the encoded audio. In general, encoded bit streams may have dynamic range control words inserted or modified without affecting the encoded audio.

When it is necessary to alter subjectively the dynamic range of audio programs, the method built into the audio coding subsystem should be used. The system should provide a transparent pathway, from the audio program produced in the audio post production studio, into the home. Signal processing devices such as compressors or limiters which alter the audio signal should not be inserted into the audio signal chain. Use of the dynamic range control system embedded within the audio coding system allows the broadcaster or program provider to appropriately limit the delivered audio dynamic range without actually affecting the audio signal itself. The original audio is delivered intact and is accessible to those listeners who wish to enjoy it.

In the case where a complete audio program is formed from the combination of a main and an associated service, each of the two services being combined may have a dynamic range control signal. In most cases, the dynamic range control signal contained in a particular bit stream applies to the audio channels coded in that bit stream. There are three exceptions: a single-channel visually impaired (VI) associated service containing only a narrative describing the picture content, a single-channel commentary (C) service containing only the commentary channel, and a voice-over (VO) associated service. In these cases, the dynamic range control signal in the associated service elementary stream is used by the decoder to control the audio level of the main audio service. This allows the provider of the VI, C, or VO service the ability to alter the level of the main audio service in order to make the VI, C, or VO services intelligible. In these cases the main audio service level is controlled by both the control signal in the main service and the control signal in the associated service.

6.6 Main, associated, and multi-lingual services

6.6.1 Overview

An AC-3 elementary stream contains the encoded representation of a single audio service. Multiple audio services are provided by multiple elementary streams. Each elementary stream is conveyed by the transport multiplex with a unique PID. There are a number of audio service types which may (individually) be coded into each elementary stream. Each elementary stream is tagged as to its service type using the bsmod bit field. There are two types of *main service* and six types of *associated service*. Each associated service may be tagged (in the AC-3 audio descriptor in the transport PSI data) as being associated with one or more main audio services. Each AC-3 elementary stream may also be tagged with a language code.

Associated services may contain complete program mixes, or may contain only a single program element. Associated services which are complete mixes may be decoded

and used as is. They are identified by the full_svc bit in the AC-3 descriptor (see A/52, Annex A). Associated services which contain only a single program element are intended to be combined with the program elements from a main audio service.

This Section describes each type of service and gives usage guidelines. In general, a complete audio program (what is presented to the listener over the set of loudspeakers) may consist of a main audio service, an associated audio service which is a complete mix, or a main audio service combined with one associated audio service. The capability to simultaneously decode one main service and one associated service is required in order to form a complete audio program in certain service combinations described in this Section. This capability may not exist in some receivers.

6.6.2 Summary of service types

The service types which correspond to each value of bsmod are defined in the Digital Audio Compression (AC-3) Standard and in Annex B of the Digital Television Standard. The information is reproduced in Table 6.1 and the following paragraphs briefly describe the meaning of these service types.

Table 6.1 Table of Service Types

bsmod	Type of service
000 (0)	Main audio service: complete main (CM)
001 (1)	Main audio service: music and effects (ME)
010 (2)	Associated service: visually impaired (VI)
011 (3)	Associated service: hearing impaired (HI)
100 (4)	Associated service: dialogue (D)
101 (5)	Associated service: commentary (C)
110 (6)	Associated service: emergency (E)
111 (7)	Associated service: voice-over (VO)

6.6.2.1 Complete main audio service (CM)

This is the normal mode of operation. All elements of a complete audio program are present. The audio program may be any number of channels from 1 to 5.1.

6.6.2.2 Main audio service, music and effects (ME)

All elements of an audio program are present except for dialogue. This audio program may contain from 1 to 5.1 channels. Dialogue may be provided by a D associated service (that may be simultaneously decoded and added to form a complete program).

6.6.2.3 Associated service: visually impaired (VI)

This is typically a single-channel service, intended to convey a narrative description of the picture content for use by the visually impaired, and intended to be decoded along with the main audio service. The VI service also may be provided as a complete mix of all program elements, in which case it may use any number of channels (up to 5.1).

6.6.2.4 Associated service: hearing impaired (HI)

This is typically a single-channel service, intended to convey dialogue which has been processed for increased intelligibility for the hearing impaired, and intended to be decoded along with the main audio service. The HI service also may be provided as a complete mix of all program elements, in which case it may use any number of channels (up to 5.1).

6.6.2.5 Associated service: dialogue (D)

This service conveys dialogue intended to be mixed into a main audio service (ME) which does not contain dialogue.

6.6.2.6 Associated service: commentary (C)

This service typically conveys a single-channel of commentary intended to be optionally decoded along with the main audio service. This commentary channel differs from a dialogue service, in that it contains optional instead of necessary program content. The C service also may be provided as a complete mix of all program elements, in which case it may use any number of channels (up to 5.1).

6.6.2.7 Associated service: emergency message (E)

This is a single-channel service, which is given priority in reproduction. If this service type appears in the transport multiplex, it is routed to the audio decoder. If the audio decoder receives this service type, it will decode and reproduce the E channel while muting the main service.

6.6.2.8 Associated service: voice-over (VO)

This is a single-channel service intended to be decoded and added into the center loudspeaker channel.

6.6.3 Multi-lingual services

Each audio bit stream may be in any language. In order to provide audio services in multiple languages a number of main audio services may be provided, each in a different language. This is the (artistically) preferred method, because it allows unrestricted placement of dialogue along with the dialogue reverberation. The disadvantage of this method is that as much as 384 kbps is needed to provide a full 5.1-channel service for each language. One way to reduce the required bit-rate is to reduce the number of audio channels provided for languages with a limited audience. For instance, alternate language versions could be provided in 2-channel stereo with a bit-rate of 128 kbps. Or, a mono version can be supplied at a bit-rate of approximately 64-96 kbps.

Another way to offer service in multiple languages is to provide a main multi-channel audio service (ME) which does not contain dialogue. Multiple single-channel dialogue associated services (D) can then be provided, each at a bit-rate of approximately 64-96 kbps. Formation of a complete audio program requires that the appropriate

language D service be simultaneously decoded and mixed into the ME service. This method allows a large number of languages to be efficiently provided, but at the expense of artistic limitations. The single-channel of dialogue would be mixed into the center reproduction channel, and could not be panned. Also, reverberation would be confined to the center channel, which is not optimum. Nevertheless, for some types of programming (sports, etc.) this method is very attractive due to the savings in bit rate it offers. Some receivers may not have the capability to simultaneously decode an ME and a D service.

Stereo (two-channel) service without artistic limitation can be provided in multiple languages with added efficiency by transmitting a stereo ME main service along with stereo D services. The D and appropriate language ME services are simply combined in the receiver into a complete stereo program. Dialogue may be panned, and reverberation may be placed included in both channels. A stereo ME service can be sent with high quality at 192 kbps, while the stereo D services (voice only) can make use of lower bit-rates, such as 128 or 96 kbps per language. Some receivers may not have the capability to simultaneously decode an ME and a D service.

Note that during those times when dialogue is not present, the D services can be momentarily removed, and their data capacity used for other purposes.

6.6.4 Detailed description of service types

6.6.4.1 CM — complete main audio service

The CM type of main audio service contains a complete audio program (complete with dialogue, music, and effects). This is the type of audio service normally provided. The CM service may contain from 1 to 5.1 audio channels. The CM service may be further enhanced by means of the VI, HI, C, E, or VO associated services described below. Audio in multiple languages may be provided by supplying multiple CM services, each in a different language.

6.6.4.2 ME — main audio service, music and effects

The ME type of main audio service contains the music and effects of an audio program, but not the dialogue for the program. The ME service may contain from 1 to 5.1 audio channels. The primary program dialogue is missing and (if any exists) is supplied by providing a D associated service. Multiple D services in different languages may be associated with a single ME service.

6.6.4.3 VI — visually impaired

The VI associated service typically contains a narrative description of the visual program content. In this case, the VI service is a single audio channel. Simultaneous reproduction of the VI service and the main audio service allows the visually impaired user to enjoy the main multi-channel audio program, as well as to follow the on-screen activity. This allows the VI service to be mixed into one of the main reproduction channels (the choice of channel may be left to the listener) or to be provided as a separate output (which, for instance, might be delivered to the VI user via open-air headphones).

The dynamic range control signal in this type of VI service is intended to be used by the audio decoder to modify the level of the main audio program. Thus the level of the main audio service will be under the control of the VI service provider, and the provider may signal the decoder (by altering the dynamic range control words embedded in the VI audio elementary stream) to reduce the level of the main audio service by up to 24 dB in order to assure that the narrative description is intelligible.

Besides providing the VI service as a single narrative channel, the VI service may be provided as a complete program mix containing music, effects, dialogue, and the narration. In this case, the service may be coded using any number of channels (up to 5.1), and the dynamic range control signal applies only to this service. The fact that the service is a complete mix is indicated in the AC-3 descriptor (see A/52, Annex A).

6.6.4.4 HI — hearing impaired

The HI associated service typically contains only a single-channel of dialogue and is intended for use by those whose hearing impairments make it difficult to understand the dialogue in the presence of music and sound effects. The dialogue may be processed for increased intelligibility by the hearing impaired. The hearing impaired listener may wish to listen to a mixture of the single-channel HI dialogue track and the main program audio. Simultaneous reproduction of the HI service along with the CM service allows the HI listener to adjust the mixture to control the emphasis on dialogue over music and effects. The HI channel would typically be mixed into the center channel. An alternative would be to deliver the HI signal to a discrete output (which, for instance, might feed a set of open-air headphones worn only by the HI listener.)

Besides providing the HI service as a single narrative channel, the HI service may be provided as a complete program mix containing music, effects, and dialogue with enhanced intelligibility. In this case, the service may be coded using any number of channels (up to 5.1). The fact that the service is a complete mix shall be indicated in the AC-3 descriptor (see A/52, Annex A).

6.6.4.5 D — dialogue

The dialogue associated service is employed when it is desired to most efficiently offer multi-channel audio in several languages simultaneously, and the program material is such that the restrictions (no panning, no multi-channel reverberation) of a single dialogue channel may be tolerated. When the D service is used, the main service is of type ME (music and effects). In the case that the D service contains a single-channel, simultaneously decoding the ME service along with the selected D service allows a complete audio program to be formed by mixing the D channel into the center channel of the ME service. Typically, when the main audio service is of type ME, there will be several different language D services available. The transport demultiplexer may be designed to select the appropriate D service to deliver to the audio decoder based on the listener's language preference (which would typically be stored in memory in the receiver). Or, the listener may explicitly instruct the receiver to select a particular language track, overriding the default selection.

If the ME main audio service contains more than two audio channels, the D service will be monophonic (1/0 mode). If the main audio service contains two channels, the D service may contain two channels (2/0 mode). In this case, a complete audio program is formed by simultaneously decoding the D service and the ME service, mixing the left channel of the ME service with the left channel of the D service, and mixing the right channel of the ME service with the right channel of the D service. The result will be a two-channel stereo signal containing music, effects, and dialogue.

6.6.4.6 C — commentary

The commentary associated service is similar to the D service, except that instead of conveying primary program dialogue, the C service conveys optional program commentary. When C service(s) are provided, the receiver may notify the listener of their presence. The listener should be able to call up information (probably on-screen) about the various available C services, and optionally request one of them to be selected for decoding along with the main service. The C service may be added to any loudspeaker channel (the listener may be given this control). Typical uses for the C service might be optional added commentary during a sporting event, or different levels (novice, intermediate, advanced) of commentary available to accompany documentary or educational programming.

The C service may be a single audio channel containing only the commentary content. In this case, simultaneous reproduction of a C service and a CM service will allow the listener to hear the added program commentary.

The dynamic range control signal in the single-channel C service is intended to be used by the audio decoder to modify the level of the main audio program. Thus the level of the main audio service will be under the control of the C service provider, and the provider may signal the decoder (by altering the dynamic range control words embedded in the C audio elementary stream) to reduce the level of the main audio service by up to 24 dB in order to assure that the commentary is intelligible.

Besides providing the C service as a single commentary channel, the C service may be provided as a complete program mix containing music, effects, dialogue, and the commentary. In this case the service may be provided using any number of channels (up to 5.1). The fact that the service is a complete mix shall be indicated in the AC-3 descriptor (see A/52, Annex A).

6.6.4.7 E — emergency

The E associated service is intended to allow the insertion of emergency announcements. The normal audio services do not necessarily have to be replaced in order for the emergency message to get through. The transport demultiplexer shall give first priority to this type of audio service. Whenever an E service is present, it is delivered to the audio decoder by the transport subsystem. When the audio decoder receives an E type associated service, it stops reproducing any main service being received and only reproduces the E service. The E service may also be used for non-emergency applications.

It may be used whenever the broadcaster wishes to force all decoders to quit reproducing the main audio program and substitute a higher priority single-channel.

6.6.4.8 VO — voice-over

It is possible to use the E service for announcements, but the use of the E service leads to a complete substitution of the voice-over for the main program audio. The voice-over associated service is similar to the E service, except that it is intended to be reproduced along with the main service. The systems demultiplexer shall give second priority to this type of associated service (second only to an E service). The VO service is intended to be simultaneously decoded and mixed into the center channel of the main audio service which is being decoded. The dynamic range control signal in the VO service is intended to be used by the audio decoder to modify the level of the main audio program. Thus the level of the main audio service will be under the control of the broadcaster, and the broadcaster may signal the decoder (by altering the dynamic range control words embedded in the VO audio bit stream) to reduce the level of the main audio service by up to 24 dB during the voice-over. The VO service allows typical voice-overs to be added to an already encoded audio bit stream, without requiring the audio to be decoded back to baseband and then re-encoded. However, space must be available within the transport multiplex to make room for the insertion of the VO service.

6.7 Audio bit rates

6.7.1 Typical audio bit rates

The information in Table 6.2 provides a general guideline as to the audio bit rates which are expected to be most useful. For main services, the use of the LFE channel is optional and will not affect the indicated data rates.

6.7.2 Audio bit rate limitations

The audio decoder input buffer size (and thus part of the decoder cost) is determined by the maximum bit rate which must be decoded. The syntax of the AC-3 standard supports bit rates ranging from a minimum of 32 kbps up to a maximum of 640 kbps per individual elementary bit stream. The bit rate utilized in the digital television system is restricted in order to reduce the size of the input buffer in the audio decoder, and thus the receiver cost. Receivers can be expected to support the decoding of a main audio service, or an associated audio service which is a complete service (containing all necessary program elements), at a bit rate up to and including 384 kbps. Transmissions may contain main audio services, or associated audio services which are complete services (containing all necessary program elements), encoded at a bit rate up to and including 384 kbps. Transmissions may contain single-channel associated audio services intended to be simultaneously decoded along with a main service encoded at a bit rate up to and including 128 kbps. Transmissions may contain dual-channel dialogue associated services intended to be simultaneously decoded along with a main service encoded at a bit rate up to and including 192 kbps. Transmissions have a further limitation that the combined bit

rate of a main and an associated service which are intended to be simultaneously reproduced is less than or equal to 512 kbps.

Table 6.2 Typical Audio Bit Rate

Type of service	Number of channels	Typical bit rates
CM, ME, or associated audio service containing all necessary program elements	5	320-384 kbps
CM, ME, or associated audio service containing all necessary program elements	4	256-384 kbps
CM, ME, or associated audio service containing all necessary program elements	3	192-320 kbps
CM, ME, or associated audio service containing all necessary program elements	2	128-256 kbps
VI, narrative only	1	48-128 kbps
HI, narrative only	1	48-96 kbps
D	1	64-128 kbps
D	2	96-192 kbps
C, commentary only	1	32-128 kbps
E	1	32-128 kbps
VO	1	64-128 kbps

7. ANCILLARY DATA SERVICES

The digital television system affords the opportunity to augment the basic television video and audio service with ancillary digital data services. The flexibility of the MPEG-2 transport layer employed in the system allows new digital services to be easily introduced at any time in a completely backward-compatible manner. Sections 7.1 and 7.2 describe two ancillary services that are specified as a part of the Digital Television Standard from the outset. These basic services provide for the transmission of Program Subtitles (PS), Emergency Messages (EM), and Program Guide information. Section 7.3 describes how certain system information may be transmitted in the transport stream. Section 7.4 describes the required elements for new services that may be added in the future and the method of specifying such future services.

7.1 Text-based ancillary services

The Digital Television Standard provides for two text-based services which augment the basic television service. These are Program Subtitles, which are completely analogous to the Closed Caption service transmitted on line 21 of the analog NTSC signal, and Emergency Messages which are mixed into the baseband video signal in the NTSC signal. The ancillary services envisioned are a mix of associated audio services and textual services. The associated audio services are simply transported as separate audio streams, and are not described here because they follow the specifications of the audio service. Similarly, the text-based services are not described here because they are carried as video user bits and follow the specifications of the video service.

7.2 Program guide

At the option of broadcasters, an interactive program guide database may be transmitted in the transport stream. The program guide database contains information relating to the programs currently being transmitted and information for programs that will be transmitted at future times. It also contains time and control information to facilitate navigation and as such it allows suitably-equipped receivers to build an interactive on-screen grid of program information. The program guide shall be formatted according to the structure and syntax described in the standard developed from ATSC Document T3/S8-050 "Program Guide for Digital Television". The program guide is defined in four major tables, a master program guide, a special program guide and two information parcels. In order to facilitate simple acquisition of the master program guide by a receiver, a specific transport PID is reserved for its transmission. The PIDs for the other tables are specified in the master program guide.

7.2.1 Master program guide PID

If present, the master program guide data stream shall be transported in PID 0x1FFD. This PID shall be reserved exclusively for the program guide.

7.2.2 Program guide STD model

Each program guide bit stream shall adhere to an STD model that can be described by an MPEG smoothing buffer descriptor (Section 2.6.30 in ISO/IEC 13818-1) with the following constraints:

- sb_leak_rate shall be 250 (indicating a leak rate of 100,000 bps)

- sb_size shall be 1024 (indicating a smoothing buffer size of 1024 bytes)

Note that the smoothing buffer descriptor is referred to here to describe the STD model for the program guide, and does not imply that a smoothing buffer descriptor for the program guide is to be included in the PMT.

7.3 System information

At the option of broadcasters, certain system information may be transmitted in the transport stream. In general, system information contains information on the locations of available transport streams and on how to acquire these streams and their constituent services. System information is especially useful in networks where the operator has control over more than one transport multiplex. To allow for interoperability between broadcast and other transmission environments, the system information mechanisms identified for broadcast applications are a compatible subset of a protocol which provides also for cable, satellite, MMDS, and SMATV. The system information shall be formatted according to the structure and syntax described in the standard developed from ATSC Document T3/S8-079 "System Information for Digital Television". Constraints applying to specific transmission media are given in that standard. In order to facilitate simple acquisition of the system information by a receiver, a specific transport PID is reserved for its transmission.

7.3.1 System information PID

If present, the system information data stream shall be transported in PID 0x1FFC. This PID shall be reserved exclusively for the system information.

7.3.2 System information STD model

The system information bit stream shall adhere to an STD model that can be described by an MPEG smoothing buffer descriptor (Section 2.6.30 in ISO/IEC 13818-1) with the following constraints:

- sb_leak-rate shall be 50 (indicating a lead rate of 20,000 bps)

- sb_size shall be 1024 (indicating a smoothing buffer size of 1024 bytes)

Note that the smoothing buffer descriptor is referred to here to describe the STD model for the system information, and does not imply that a smoothing buffer descriptor for the system information is to be included in the PMT.

7.4 Specification of additional ancillary services

The MPEG-2 transport layer adopted for use in the Digital Television Standard allows new ancillary services to be added to the basic service at any time in the future. To ensure interoperability of service encoders, broadcast multiplexers and switchers, and consumer receivers, it is important that the signal format and delivery characteristics of any new ancillary services that are added in the future be specified completely and unambiguously.

One way to do this would be to specify precisely the encoder and decoder characteristics. A better way, as it provides more latitude in the design of encoders and decoders while still guaranteeing interoperability, is to specify the syntax and semantics of the transmitted bit stream and to specify an idealized decoder. This is the method that was used by MPEG and it is the method used in the Digital Television Standard.

Section 5.6.2 of Annex C of the Digital Television Standard requires that the standards for any additional ancillary services be specified in terms of a verification model by defining the characteristics of the transmitted syntax and an idealized decoder.

In MPEG-2 parlance, these ancillary services are referred to as private-data. Private data provides a means to add new ancillary services to the basic digital television service specified in this Standard. Private data is supported in two bit stream locations:

1. Private data can be transmitted within the adaptation header of transport packets (Sections 2.4.3.4 and 2.4.3.5 of ISO/IEC 13818-1).

2. Private data can be transmitted as a separate transport stream with its own PID. The contents can be identified as being ATSC private by using the private_data_indicator_descriptor (Section 2.6.29 of ISO/IEC 13818-1) within the PMT.

In either case, it is necessary that the standards which specify the characteristics of such private_streams be consistent with the Digital Television Standard. Standards for private_streams shall specify precisely the semantics of the transmitted syntax as described in Sections 7.4.1 and 7.4.2.

7.4.1 Verification model

The standard shall be specified in terms of a verification model by defining the characteristics of the transmitted syntax and an idealized decoder. In ISO/IEC 13818-1 and 13818-2, this is accomplished by using the T-STD and VBV models, respectively. The elements required for specification by the Digital Television Standard are described in the following Sections.

7.4.1.1 Syntax and semantics

The syntax and semantics of the transmitted bit stream that implements the ancillary service must be completely and unambiguously specified. The decoding process must also be completely and unambiguously specified.

7.4.1.2 Ancillary service target decoder (ASTD)

An idealized decoder model must be precisely defined for the service. Figure 7.1 introduces a concrete model for pedagogic purposes It is modeled after the T-STD.

The salient features of the model are the size of the transport demultiplexing buffer (TB), the minimum transfer rate out of the transport demultiplex buffer (R_{leak}), the required system buffering (BS_{sys}), and optionally the partitioning of BS_{sys} between the smoothing portion and the decoder portion. The decoding process, represented as the decoding times T_decoder(i), must be completely specified. The behavior of the BS_{sys} buffer must be completely modeled with respect to its input process and its output process. Certain parameters such as bit rate, etc., should also be specified.
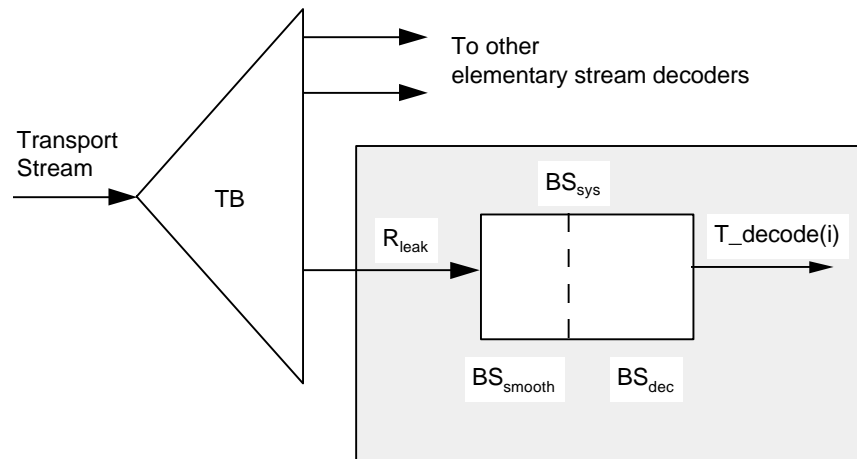
Figure 7.1. Ancillary service target decoder.

7.4.2 Stream type and PMT descriptors

A new ancillary service is described as a program or elementary stream through documented Program Specific Information.

7.4.2.1 Stream type

Several identifiers that are part of the transport section of the Digital Television Standard may be used to identify either the signal or constituent parts thereof; however, the fundamental identifier is the User Private stream type. The stream_type codes are to be unambiguously assigned within the range 0x80 to 0xAF. The AC-3 audio stream is identified by the code 0x81.

7.4.2.2 PMT descriptors

The Ancillary Service specification must include all pertinent descriptors that are found within the Program Map Table. Specifically, it is recommended that either the private_stream_identifier or the registration_descriptor, or both, be included. The value of the 32-bit private_stream_identifier is assigned by the entity specifying the syntax of the Ancillary

Service bit stream. Consequently, its value is not guaranteed to be unique. The value of the 32-bit format_identifier in the registration_descriptor is assigned by the ISO-designated registration authority and is guaranteed to be unique. Although inclusion of the private_stream_identifier and/or registration_descriptor is not required for a stream with a unique stream_type code within the Digital Television Standard, it will enhance interoperability in the case where the stream is stored outside this Standard, or transmitted in some other network that has its own set of stream_type codes.

8. SERVICE MULTIPLEX AND TRANSPORT SYSTEMS

8.1 Introduction

This Chapter provides a tutorial description of the functionality and format of the transport system employed in the digital television system. It is intended to aid the reader in understanding and applying the precise specification of the transport layer given in the underlying normative standards documents. The precise transport system standard is based on MPEG-2 and is given in ISO/IEC 13818-1 as constrained and extended for the Digital Television Standard. The constraints and extensions to the MPEG-2 standard that apply are given in Annex C of the Digital Television Standard.

The transport system employs the fixed-length transport stream packetization approach defined by MPEG. This approach to the transport layer is well suited to the needs of terrestrial broadcast and cable television transmission of digital television. The use of moderately long, fixed-length packets matches well with the needs and techniques for error protection in both terrestrial broadcast and cable television distribution environments. At the same time it provides great flexibility to accommodate the initial needs of the service to multiplex video, audio, and data while providing a well-defined path to add additional services in the future in a fully backward-compatible manner. By basing the transport layer on MPEG-2, maximum interoperability with other media and standards is maintained.

8.1.1 Advantages of the fixed-length packetization approach

A transport layer based on a fixed-length packetization approach offers a great deal of flexibility and some significant advantages when attempting to multiplex data related to several applications on a single bit stream. These are described in some detail in this Section.

8.1.1.1 Dynamic capacity allocation

While digital systems are generally described as flexible, the use of fixed-length packets offers complete flexibility to allocate channel capacity among video, audio and auxiliary data services. The use of a packet identifier (or PID) in the packet header as a means of bit stream identification makes it possible to have a mix of video, audio and auxiliary data which is flexible and which need not be specified in advance. The entire channel capacity can be reallocated in bursts for data delivery. This capability could be used to distribute decryption keys to a large audience of receivers during the seconds preceding a popular pay-per-view program, or download program-related, computer software to a "smart receiver."

8.1.1.2 Scalability

The transport format is scalable in the sense that the availability of a larger bandwidth channel may also be exploited by adding more elementary bit streams at the input of the multiplexer, or even multiplexing these elementary bit streams at the second

multiplexing stage with the original bit stream. This is a valuable feature for network distribution, and also serves interoperability with a cable plant's capability to deliver a higher data rate within a 6 MHz channel.

8.1.1.3 Extensibility

Because there will be possibilities for future services that we cannot anticipate today, it is extremely important that the transport architecture provide open-ended extensibility of services. New elementary bit streams could be handled at the transport layer without hardware modification by assigning new packet IDs at the transmitter and filtering on these new PIDs in the bit stream at the receiver. Backward compatibility is assured when new bit streams are introduced into the transport system as existing decoders will automatically ignore new PIDs. This capability could possibly be used to compatibly introduce "1000-line progressive formats" or "3D-HDTV" by sending augmentation data along with the basic signal.

8.1.1.4 Robustness

Another fundamental advantage of the fixed-length packetization approach is that the fixed-length packet can form the basis for handling errors that occur during transmission. Error correction and detection processing (which precedes packet demultiplexing in the receiver subsystem) may be synchronized to the packet structure so that one deals at the decoder with units of packets when handling data loss due to transmission impairments. Essentially, after detecting errors during transmission, one recovers the data bit stream from the first good packet. Recovery of synchronization within each application is also aided by the transport packet header information. Without this approach, recovery of synchronization in the bit streams would have been completely dependent on the properties of each elementary bit stream.

8.1.1.5 Cost effective receiver implementations

A transport system based on fixed-length packets enables simple decoder bit stream demultiplex architectures, suitable for high speed implementations. The decoder does not need detailed knowledge of the multiplexing strategy or the source bit-rate characteristics to extract individual elementary bit streams at the demultiplexer. All the receiver needs to know is the identity of the packet, which is transmitted in each packet header at fixed and known locations in the bit stream. The only important timing information is for bit-level and packet-level synchronization.

8.1.1.6 MPEG-2 compatibility

The transport system is based on the MPEG-2 System specification. While the MPEG-2 System layer has been designed to support many different transmission and storage scenarios, care has been taken to limit the burden of protocol inefficiencies caused by this generality in definition.

An additional advantage of MPEG-2 compatibility is interoperability with other MPEG-2 applications. The MPEG-2 format is likely to be used for a number of other

applications, including storage of compressed bit streams, computer networking, and non-HDTV delivery systems. MPEG-2 transport system compatibility implies that digital television transport bit streams may be handled directly in these other applications (ignoring for the moment the issue of bandwidth and processing speed).

While the transport format conforms to the MPEG-2 Systems standard, it will not exercise all the capabilities defined in the MPEG-2 standard. Therefore, a digital television decoder need not be fully MPEG-2 Systems compliant, in that it will not need to decode any arbitrary MPEG-2 Systems bit stream. However, all MPEG-2 decoders should be able to decode the digital television bit stream syntax at the transport system level. Documents defining the extent to which the MPEG capabilities are supported in this digital television transport specification have been submitted to the MPEG committee and have contributed to the standard. MPEG-2 features not supported in ATSC Standard A/53 are constrained[10] when deemed not applicable to broadcast/cable delivery of digital television.

8.1.2 Overview of the transport subsystem

Figure 8.1 illustrates the organization of a digital television transmitter-receiver pair and the location of the transport subsystem in the overall system. The transport resides between the application (e.g., audio or video) encoding/decoding function and the transmission subsystem. At its lowest layer, the encoder transport subsystem is responsible for formatting the encoded bits and multiplexing the different components of the program for transmission. At the receiver, it is responsible for recovering the bit streams for the individual application decoders and for the corresponding error signaling. (At a higher layer, multiplexing and demultiplexing of multiple programs within a single bit stream can be achieved with an additional system level multiplexing or demultiplexing stage before the modem in the transmitter and after the modem in the receiver.) The transport subsystem also incorporates other higher level functionality related to identification of applications and, as illustrated, synchronization of the receiver. This document will describe these functions in greater detail.

As described earlier, the data transport mechanism is based on the use of fixed-length packets that are identified by headers. Each header identifies a particular application bit stream (also called an *elementary bit stream*) which forms the payload of the packet. Applications supported include video, audio, data, program and system control information, etc. The elementary bit streams for video and audio are themselves wrapped in a variable-length packet structure called the packetized elementary stream (PES) before transport processing. The PES layer provides functionality for identification, and synchronization of decoding and presentation of the individual application. The format and functionality of a PES packet is described in Section 8.5.[11]

---

[10] The constraint takes the form of a limitation of functionality. In this instance, certain flags will be permanently configured, and some fields will not appear in the bit stream. This allows a simpler decoder, as it will not be necessary to handle elements that are not used.

[11] Note that the PES layer is not required for all applications. Its use is required for both the video and audio in the Digital Television Standard.
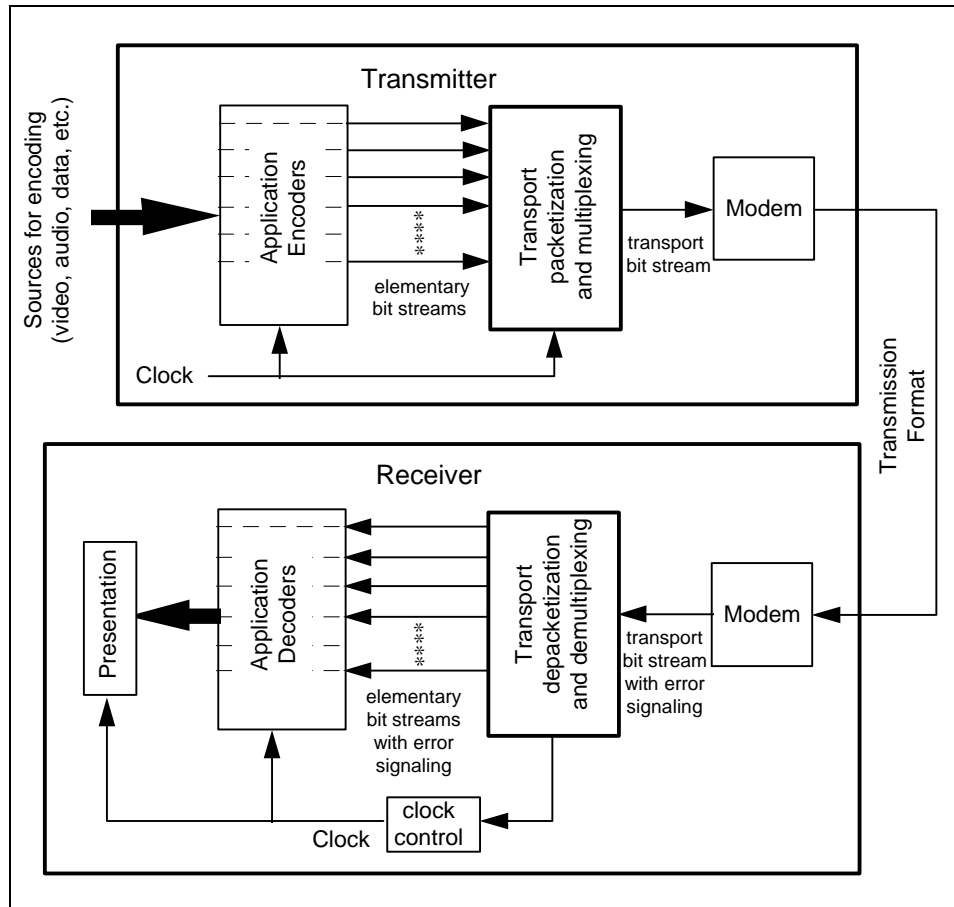
Figure 8.1. Sample organization of functionality in a transmitter-receiver pair for a single digital television program.

Moving up one level in the description of the general organization of the bit streams, elementary bit streams sharing a common time base are multiplexed, along with a control data stream, into *programs*. Note that a program in the digital television system is analogous to a channel in the NTSC system in that it contains all of the video, audio, and other information required to make up a complete television program. These programs and an overall system control data stream are then asynchronously multiplexed to form a multiplexed *system*. The organization is described in detail in Section 8.3.

At this level, the transport is also quite flexible in two aspects:

1. It permits programs to be defined as any combination of elementary bit streams; specifically, the same elementary bit stream can be present in more than one program (e.g., two different video bit streams with the same audio bit stream); a program can be formed by combining a basic elementary bit stream and a supplementary elementary bit stream (e.g., bit streams for scaleable decoders); programs can be tailored for specific needs (e.g., regional selection of language for broadcast of secondary audio); etc.

2. Flexibility at the systems layer allows different programs to be multiplexed into the system as desired, and allows the system to be reconfigured easily when

required. The procedure for extraction of separate programs from within a system is also simple and well defined.

The transport system provides other features that are useful for both normal decoder operation and for the special features required in broadcast and cable applications. These include:

1. Decoder synchronization

2. Conditional access

3. Local program insertion

The elements of these features that are relevant to the Digital Television Standard will be discussed in detail.

The transport bit stream definition directly addresses issues relating to the storage and playback of programs. Although this is not directly related to the transmission of digital television programs, it is a fundamental requirement for creating programs in advance, storing them, and playing them back at the desired time. The programs are stored in the same format in which they are transmitted, i.e., as transport bit streams. The bit stream format also contains the hooks needed to support the design of consumer digital products based on recording and playback of these bit streams, including the use of the "trick modes" that one is familiar with for current analog VCRs. It should be noted that the issues related to storage and playback of digitally compressed video bit streams are quite different from those that apply to analog systems such as NTSC.

8.1.3 General bit stream interoperability issues

Bit stream interoperability at the transport level is an important feature of the digital television system. There are two aspects of interoperability to consider. One is whether the transport bit stream can be carried on other communication systems, and the other is the ability of the system to carry bit streams generated by other communication systems.

In general, there is nothing that prevents the transmission of a bit stream as the payload of a different transmission system. It may be simpler to achieve this functionality in certain systems, e.g., cable television, DBS, ATM, etc., than in others, e.g., computer networks based on protocols such as FDDI, IEEE 802.6, etc., but it is always possible. Because ATM is expected to form the basis of future broadband communications networks, the issue of bit stream interoperability with ATM networks is especially important. ATM interoperability has been specifically addressed in the design of the protocol. This is discussed in more detail in Section 8.5.4.2.

The other aspect of interoperability is the transmission of other, non-television, bit streams within the digital television system. This makes more sense for bit streams linked to television broadcast applications, e.g., cable television, DBS, etc., but is also possible for other "private" bit streams. This function is achieved by transmitting these other bit streams as the payload of identifiable transport packets. The only requirement is to have the general nature of these bit streams recognized within the system context. Note that

there is also a certain minimum system-level processing requirement defined by the Digital Television Standard that needs to be implemented to extract all (even private) bit streams. The details are made clearer in the Sections that follow. It is also important to remember that the system is essentially a broadcast system and hence any private transmissions that are based on a two-way communications protocol will not be directly supported, unless this functionality is provided external to the system definition.

8.2 The packetization approach and functionality

The transport bit stream consists of fixed-length packets with a fixed and a variable component to the header field as illustrated in Figure 8.2.
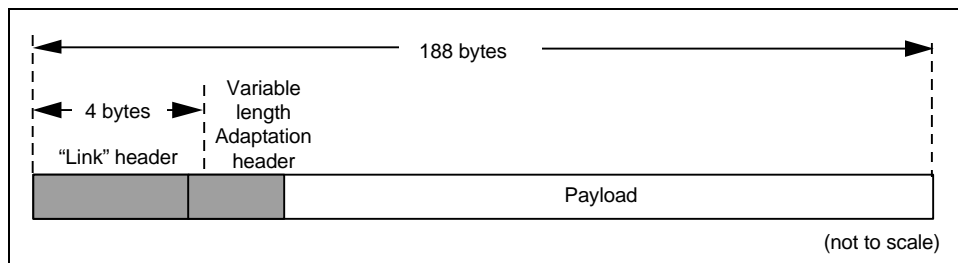
Figure 8.2. Transport packet format.

Each packet consists of 188 bytes and is constructed in accordance with the MPEG-2 transport syntax and semantics. The choice of this packet size was motivated by a few factors. The packets need to be large enough so that the overhead due to the transport headers do not become a significant portion of the total data carried. They should not be so large that the probability of packet error becomes significant under standard operating conditions (due to inefficient error correction). It is also desirable to have packet lengths consistent with the block sizes of typical, block oriented, error correction methods, so that packets may be synchronized to error correction blocks, and the physical layer of the system can aid the packet level synchronization process in the decoder. Another reason for the particular packet length selection is interoperability with the ATM format. The general philosophy is to transmit a single MPEG-2 transport packet in four ATM cells. There are, in general, several approaches to achieve this functionality. Section 8.5.4.2 includes a discussion of one possible approach.

The contents of each packet are identified by the packet headers. The packet header structure is layered and may be described as a combination of a fixed-length "link" layer and a variable-length adaptation layer. Each layer serves a different function similar to the link and transport layer functions in the OSI layered model of a communications system. In the digital television system, this link and adaptation level functionality are used directly for the terrestrial broadcast link on which the MPEG-2 transport bit stream is transmitted. However, in a different communications system, e.g., ATM, the MPEG-2 headers would not play a role in implementing a protocol layer in the overall transmission system. The MPEG-2 headers would be carried as part of the payload in such a case and would continue to serve as identifiers for the contents of the data stream.

8.2.1 The "link" layer

The link layer is implemented using a four byte header field. The format of the header field is described in greater detail in Section 8.4.1. Some of the important functions that are enabled by the header elements are described here.

8.2.1.1 Packet synchronization

Packet synchronization is enabled by the sync_byte, which is the first byte in a packet. The sync_byte has the same fixed, pre-assigned, value for all MPEG-2 bit streams. In some implementations of decoders the packet synchronization function is done at the physical layer of the communication link (which precedes the packet demultiplexing stage), in which case this sync_byte field may be used for verification of the packet synchronization function. In other decoder implementations this byte may be used as the primary source of information for establishing packet synchronization. The standard does not specify the details of the approach to be used to implement this function in a decoder but only provides the hooks in the bit stream to facilitate the function.

8.2.1.2 Packet identification

As discussed earlier, an important element in the link header is a 13 bit field called the PID or Packet ID field. This provides the mechanism for multiplexing and demultiplexing bit streams, by enabling identification of packets belonging to a particular elementary or control bit stream. As the location of the PID field in the header is always fixed, extraction of the packets corresponding to a particular elementary bit stream is very simple to achieve once packet synchronization is established by filtering packets based on PIDs. The fixed packet length makes for simple filter and demultiplexing implementations suitable for high speed transmission systems.

8.2.1.3 Error handling

Error detection is enabled at the packet layer in the decoder through the use of the continuity_counter field. At the transmitter end, the value in this field cycles from 0 through 15 for all packets with the same PID that carry a data payload (as will be seen later, the transport protocol allows the definition of packets that have no data payload). At the receiver end, under normal conditions, the reception of packets in a PID stream with a discontinuity in the continuity_counter value indicates that data has been lost in transmission. The transport processor at the decoder then signals the decoder for the particular elementary stream about the loss of data. This signaling approach is not included in the standard.

Because certain information (such as adaptation headers, time stamps, and program maps) is very important to the smooth and continuous operation of a system, the transport system has a means to increase the robustness of this information to channel errors by providing a mechanism for the encoder to duplicate packets. Those packets that contain important information can be duplicated at the encoder. At the decoder, the duplicate packets are either used if the original packet was in error or are dropped.

Semantics for identifying duplicate packets are described in the description of the continuity_counter.

8.2.1.4 Conditional access

The transport format allows for scrambling of data in the packets. Each elementary bit stream in the system can be scrambled independently. The Digital Television Standard specifies the descrambling approach to be used but does not specify the descrambling key and how it is obtained at the decoder. The key must be delivered to the decoder within a time interval of its usefulness. There is "private" data capacity at several locations within the transport stream where this key data might be carried. Two likely locations are 1) a separate private stream with it's own PID, or 2) a private field within an adaptation header carried by the PID of the signal being scrambled. The security of the conditional access system is ensured by encrypting the descrambling key when sending it to the receiver, and by updating the key frequently. There are no constraints imposed by this Standard on the number of keys that can be used and the rate at which they may be changed. As mentioned before, the key encryption, transmission, and decryption approaches are not a part of the Standard and could differ in different applications of the Digital Television Standard.

The EIA and NCTA have a Joint Engineering Committee which is working to standardize receiver interfaces to implement the key decryption and bit stream descrambling functions. This is described in more detail in Section 8.5.3.

Information in the link header of a transport packet indicates whether the payload in the packet is scrambled and if so, flags the key to be used for descrambling. The header information in a packet is always transmitted in the clear, i.e., unscrambled. The amount of data to be scrambled in a packet is variable depending on the length of the adaptation header. It should be noted that some padding of the adaptation field may be necessary for certain block mode algorithms. Conditional access is discussed in greater detail in Section 8.5.3.

Note that the general MPEG-2 transport definition provides the mechanism to scramble at two levels, within the PES packet structure and at the transport layer. Scrambling at the PES packet layer is primarily useful in the program stream (which is not supported in the Digital Television Standard), where there is no protocol layer similar to the transport layer to enable this function. In this system, scrambling will be implemented only at the transport layer.

8.2.2 The adaptation layer

The adaptation header in the MPEG-2 packet is a variable-length field. Its presence is flagged in the link level section of the header. The functionality of these headers is basically related to the decoding of the elementary bit stream that is extracted using the link level functions. Some of the functions of this layer that are important to the functioning of the digital television system are described here.

8.2.2.1 Synchronization and timing

Synchronization of the decoding and presentation process for the applications running at a receiver is a particularly important aspect of real-time digital data delivery systems such as the digital television system. Because received data is expected to be processed at a particular rate (to match the rate at which it is generated and transmitted), loss of synchronization leads to either buffer overflow or underflow at the decoder, and as a consequence, loss of presentation and/or display synchronization. The problems in dealing with this issue for a digital compressed bit stream are different from those for analog NTSC. In NTSC, information is transmitted for the pictures in a synchronous manner, so that one can derive a clock directly from the picture synchronizing signal. In a digital compressed system the amount of data generated for each picture is variable (based on the picture coding approach and complexity), and timing cannot be derived directly from the start of picture data. Indeed, there is really no natural concept of synchronism between display and transmission (that one is familiar with in NTSC) in a compressed digital television bit stream.

The solution to this issue is to transmit timing information in the adaptation headers of selected packets, to serve as a reference for timing comparison at the decoder. This is done by transmitting a sample of a 27 MHz clock in the program_clock_reference (PCR) field, which indicates the expected time at the completion of the reading of that field from the bit stream at the transport decoder. The phase of the local clock running at the decoder is compared to the PCR value in the bit stream at the instant at which it is obtained, to determine whether the decoding process is synchronized. In general, the PCR from the bit stream does not directly change the phase of the local clock but only serves as an input to adjust the clock rate. Exceptions are during channel change and insertion of local programming. As mentioned earlier, the nominal clock rate in the decoder system is 27 MHz. A point to note here is that the Standard only specifies the means of transmitting synchronization information to a receiver but does not specify the implementation of the sync recovery process. Note also that the audio and video sample clocks in the decoder system are locked to the system clock derived from the PCR values. This simplifies the receiver implementation in terms of the number of local oscillators required to drive the complete decoding process, and has other advantages such as rapid sync acquisition.

Details of the format for the PCR are given in Section 8.4.2.1. Note that in this implementation the encoder and decoder system clocks are set completely independently of the modem clock. This makes for a clean separation of functionality when implementing the two subsystems, and leads to simpler interfaces. This also makes it simpler for the transport system to interface to modems that are used for transmission on other media such as cable television, DBS, computer networks, etc.

8.2.2.2 Random entry into the compressed bit stream

Random entry into the application bit streams such as video and audio is necessary to support functions such as program acquisition and program switching. Random entry into an application is possible only if the coding for the elementary bit stream for the application supports this functionality directly. For example, the video bit stream supports random entry through the concept of intraframes (or I-frames) that are coded without any

prediction, and can therefore be decoded without any prior information. The beginning of the video sequence header information preceding data for an I-frame could serve as a random entry point into a video elementary bit stream. In general, random entry points should also coincide with the start of PES packets where they are used, e.g., for video and audio. The support for random entry at the transport layer comes from a flag in the adaptation header of the packet that indicates whether the packet contains a random access point for the elementary bit stream. In addition, the data payload of packets that are random access points starts with the data that forms the random access point of entry into the elementary bit stream itself. This approach allows packets to be discarded directly at the transport layer when switching channels and searching for a resynchronization point in the transport bit stream, and also simplifies the search for the random access point in the elementary bit stream once transport level resynchronization is achieved.

A general objective is to have random entry points into the programs as frequently as possible, to enable rapid channel switching.

8.2.2.3 Local program insertion

The transport system supports insertion of local programs and commercials, by use of flags and features dedicated to this purpose in the transport packet Adaptation Header. This important issue has been discussed at great length within MPEG. The syntax allows local program insertion to be supported and its performance to improve as techniques and equipment are developed around these syntax tools. The use of these syntax elements will need to be within some imposed constraints to ensure proper operation of the video decoders. There may also need to be constraints on some current common broadcast practices, imposed not by the transport, but rather by virtue of the compressed digital data format.

The functionality of program segment insertion and switching of channels at a broadcast head-end are quite similar, the difference being in the time constants involved in the splicing process, and also in the fact that in program segment insertion the bit stream is switched back to the original program at the end of the inserted segment, while in the channel switching case one most likely switches over to yet another program at the end of the splice. There are other detailed issues related to the hardware implementation that may differ for these two cases, including input source devices and buffering requirements. For example, if local program insertion is to take place on a bit stream obtained directly from a network feed, and if the network feed does not include place-holders for local program insertion, the input program transport stream will need to be buffered for the duration of the inserted program segment. If the program is obtained from a local source, e.g., a video server or a tape machine, it may be possible to pause the input process for the duration of the inserted program segment. Neither of these is an issue for channel switching.

8.2.2.3.1 Systems level view

There are two layers of processing functionality to address when doing local program insertion. The lower layer functionality is related to splicing transport bit streams for the individual elements of the program. The higher level functionality is related to

coordination of this process between the different elementary bit streams which make up the program transport stream. Figure 8.3 illustrates the correct approach to implement program insertion.
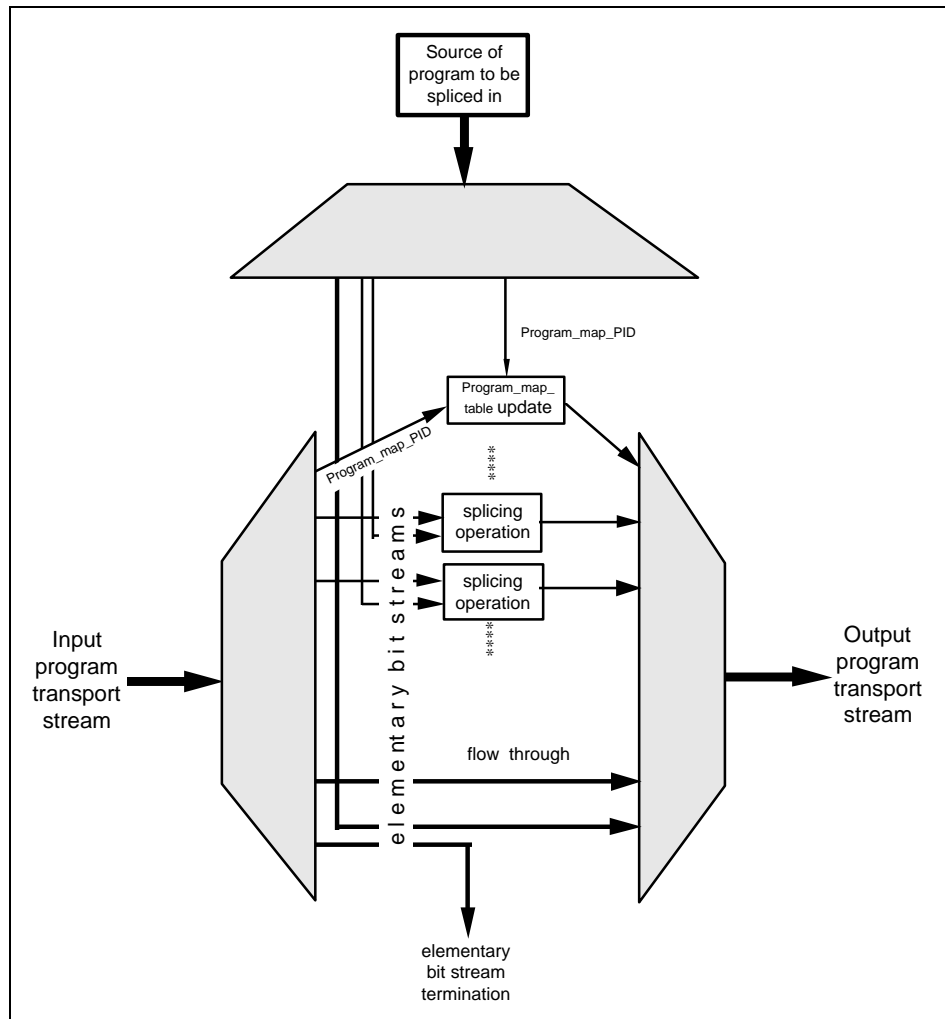
Figure 8.3. Example program insertion architecture.

The first step for program insertion to take place is to extract (by demultiplexing) the packets, identified by the PIDs, of the individual elementary bit streams that make up the program that is to be replaced, including the bit stream carrying the program_map_table. Once these packets have been extracted, as illustrated in Figure 8.3, program insertion can take place on an individual PID basis. If applicable, some packets may be passed through without modification. There is also the flexibility to add and drop elementary bit streams.

8.2.2.3.2 Basics of elementary bit stream insertion

The interface for elementary bit stream insertion is at the transport layer of the protocol. This means that bit stream insertion always takes place in units of transport packets. The primary features enabling local elementary bit stream insertion are the

discontinuity_indicator field and the splice_countdown field in the transport header. The discontinuity_indicator signals the decoder that the PCR is changing to a new time base. This simply informs the decoder that the change in the bit stream is not due to an error in the channel, but rather is intended by the program provider. The implication for the decoder is that it should continue normal decoding, and it is the encoder's responsibility to make sure that the bit stream has been constructed in a compliant manner that will not cause decoders to crash due to overflow or underflow.

The splice_countdown field in the adaptation header is used to signal that a subsequent packet is the point for switching to a new bit stream. The count-down is a positive number which decrements on each subsequent packet of that service.

The Digital Television Standard places some constraints on the transport encoding which are more stringent than the MPEG-2 requirements. One constraint is that a video PES header is followed immediately by a video access unit. This will speed acquisition. The first packet in an insertion will contain the PCR value, with the PCR discontinuity bit set to "1" to inform the decoder that a splice has occurred. The first payload in the stream will begin with a PES Header, which will have a PTS resident, so that the decoder can determine the display time immediately. Because the PES header also has the data_alignment_indicator set, the first data following the header will be the start of the video sequence. Consequently, the decoder has all the information available to begin decoding immediately after receiving the beginning of the spliced program segment. (In general, an MPEG-2 stream does not have these constraints imposed, and hence does not have guaranteed performance at the splice points.)

It is important to note that the process of facilitating frequent opportunities for splicing in a program bit stream is not within the control of the transport layer of the system. The transport only provides the mechanism of implementing the splice itself. Hence the provision of opportunities for inserting program segments requires control over the source coding algorithms used to encode the audio and video.

An important reference for examples of program insertion is contained in Annex L of ISO/IEC 13818-1.

8.3 Higher level multiplexing functionality

As described earlier, the overall multiplexing approach can be described as a combination of multiplexing at two different layers. In the first layer program transport streams are formed by multiplexing one or more elementary bit streams at the transport layer, and in the second layer the program transport streams are combined (using asynchronous packet multiplexing) to form the overall system. The functional layer in the system that contains both this program and system level information is called the PSI or Program Specific Information.

8.3.1 Single program transport multiplex

A program transport stream[12] is formed by multiplexing individual elementary bit streams (with or without PES packetization) that share a common time base. As the elementary streams are multiplexed, they are formed into transport packets and a control bit stream that describes the program (also formed into transport packets) is added. The elementary bit streams, and the control bit stream (also called the elementary stream map in Figure 8.4), are identified by their unique PIDs in the link header field. The organization of this multiplex function is illustrated in Figure 8.4. The control bit stream contains the program_map_table that describes the elementary stream map. The program_map_table includes information about the PIDs of the transport streams that make up the program, the identification of the applications (e.g., audio, video, etc.) that are being transmitted on these bit streams, the relationship between these bit streams, etc. The details of the program_map_table syntax and the functionality of each syntax element are given in Section 8.4.5. The identification of a bit stream carrying a program_map_table is done at the system layers to be described next.
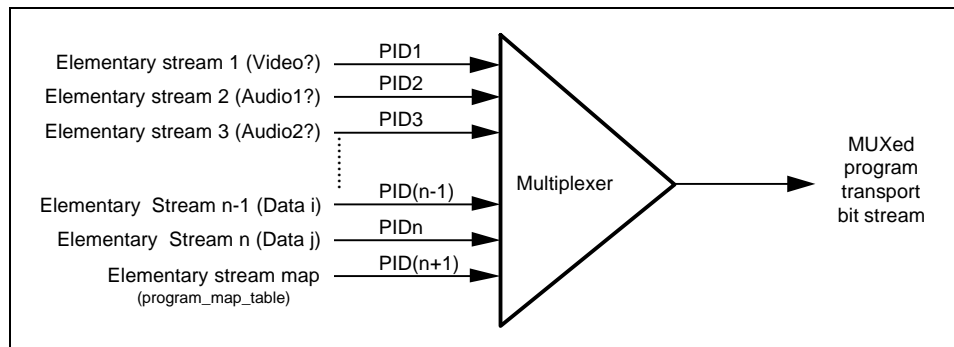
**Figure 8.4. Illustration of the multiplex function to form
a program transport stream.**

The transport syntax allows a program to be comprised of a large number of elementary bit streams, with no restriction on the types of applications required within a program. For example, a program transport stream does not need to contain a single video or audio bit stream; it could be a data "program". On the other hand, a program transport stream could contain multiple related video and audio bit streams, as long as they share a common time base. The data applications that can be carried are flexible; the only constraint is that there should be an appropriate stream_type ID assignment to enable recognition of the application corresponding to the bit stream by a decoder. The list of application types that will be supported in the initial digital television system are given in Chapter 7. Note that the initial selection of applications does not limit the future. (Indeed, it is quite impossible for one to anticipate all possible future applications!)

---

[12] The terminology can be confusing. The term **program** is analogous to a channel in NTSC as discussed in Section 8.1.2. The term **program stream** refers to a particular bit stream format defined by MPEG but not used in the Digital Television Standard. **Program transport stream** is the term used to describe a transport bit stream that has been generated for a program.

Note that, for the different elementary bit streams that make up a program, the link level functions are carried out independently without program level coordination. This includes functions such as PID manipulation, bit stream filtering, scrambling and descrambling, definition of random entry packets, etc. The coordination between the elements of a program is primarily controlled at the presentation (display) stage based on the use of the common time base. This common time base is imposed by having all elementary bit streams in a program derive timing information from a single clock, and then by transmitting this timing information via the PCR on one of the elementary bit streams that constitute the program. The data for timing of presentation is present in the elementary bit stream for each individual application.

8.3.2 System multiplex

The system multiplex is the process of multiplexing different program transport streams. In addition to the transport bit streams (with the corresponding PIDs) that define the individual programs, a system level control bit stream with PID=0 is defined. This bit stream carries the program_association_table that maps program identities to their program transport streams. The program identity is represented by a number in the program_association_table. A program corresponds to what has traditionally been called a channel, e.g., PBS, C-SPAN, etc. The map indicates the PID of the bit stream containing the program_map_table for a program. Thus, the process of identifying a program and its contents takes place in two stages: first one uses the program_association_table in the PID=0 bit stream to identify the PID of the bit stream carrying the program_map_table for the program, and then in the next stage one obtains the PIDs of the elementary bit streams that make up the program from the appropriate program_map_table. Once these steps are completed the filters at a demultiplexer can be set to receive the transport bit streams that correspond to the program of interest. This information can also be used to implement a program guide function in the receiver as discussed in more detail in Section 7.2.
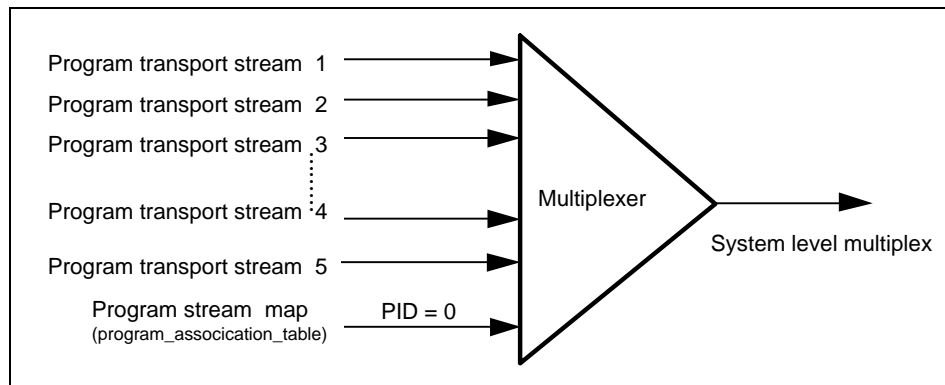
Figure 8.5. Illustration of the multiplex function to form the system level bit stream.

The system layer of multiplexing is illustrated in Figure 8.5. Note that during the process of system level multiplexing, there is the possibility of PIDs on different program streams being identical at the input. This poses a problem because PIDs for different bit streams need to be unique. A solution to this problem lies at the multiplexing stage, where some of the PIDs are modified just before the multiplex operation. The changes have to be

recorded in both the program_association_table and the program_map_table. Hardware implementation of the PID reassignment function in real time is helped by the fact that this process is synchronous at the packet clock rate. The other approach, of course, is to make sure up front that the PIDs being used in the programs that make up the system are unique. This is not always possible with stored bit streams.

Note that the architecture of the bit stream is scalable. Multiple system level bit streams can be multiplexed together on a higher bandwidth channel by extracting the program_association_tables from each system multiplexed bit stream and reconstructing a new PID=0 bit stream. Note again that PIDs may have to be reassigned in this case.

Note also that in all descriptions of the higher level multiplexing functionality, no mention is made of the functioning of the multiplexer and multiplexing policy that should be used. This function is not a part of the Standard and is up to individual designers. Because its basic function is one of filtering, the transport demultiplexer will function on any digital television bit stream regardless of the multiplexing algorithm used.
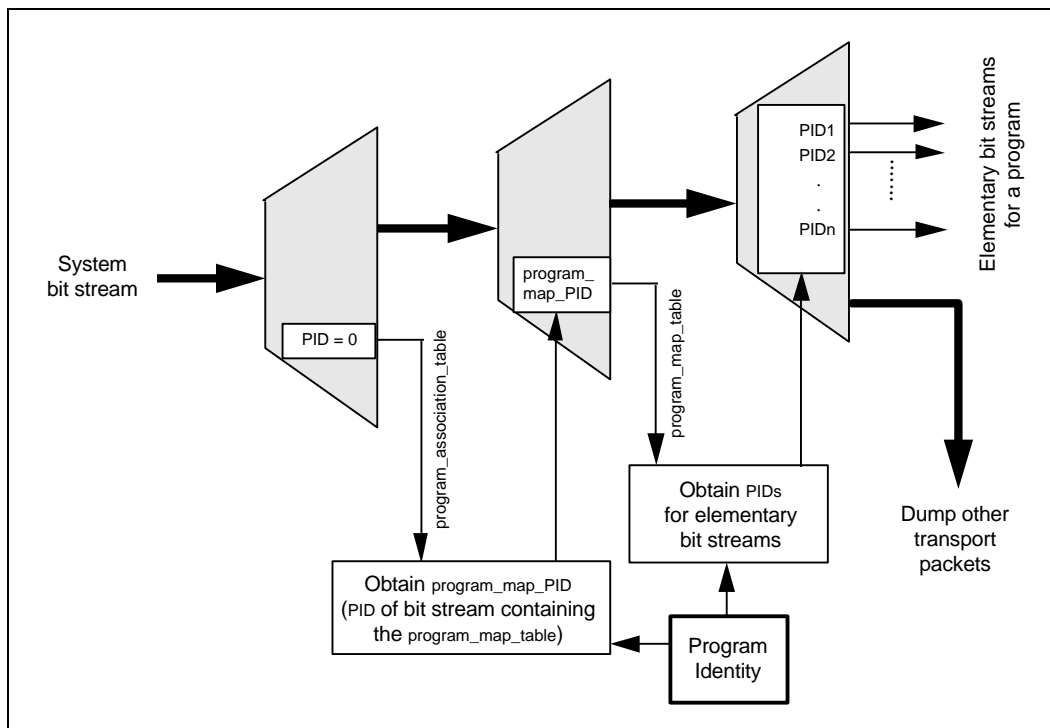
Figure 8.6. Illustration of transport demultiplexing process for a program.

Figure 8.6 illustrates the entire process of extracting elementary bit streams for a program at a receiver. It also serves as one possible implementation approach, although not the most efficient. In practice the same demultiplexer hardware could be used to extract both the program_association_table and the program_map_table control bit streams. This also represents the minimum functionality required at the transport layer to extract any application bit stream (including those that may be private).

Note that once the packets are obtained for each elementary bit stream in the program, further processing to obtain the random entry points for each component bit stream, to achieve decoder system clock synchronization, to obtain presentation (or decoding) synchronization, etc., needs to take place before the receiver decoding process reaches normal operating conditions for receiving a program.

It is important to clarify here that the layered approach to defining the multiplexing function does not necessarily imply that program and system multiplexing should always be implemented in separate stages. A hardware implementation that includes both the program and system level multiplexing within a single multiplexer stage is allowed, as long as the multiplexed output bit stream has the correct properties as described in this document and defined in Annex C of the Digital Television Standard.

It is likely that as the Digital Television Standard is implemented, certain approaches to the multiplexing process will evolve and become standard industry practice. As appropriate, SMPTE or other standardizing organizations may document these approaches and their related interfaces as voluntary standards or recommended practices.

8.4 The transport format and protocol

This Section defines the syntax elements for the transport layer bit stream. All syntax elements need to be recognized at some level in a digital television receiver. Most syntax elements trigger a response in the transport decoder. A few are present for interoperability with MPEG-2.

8.4.1 Link level header

Figure 8.7 shows the link level header with the functionality assigned to each bit. Some of these have been discussed earlier. These general functions may not all be used on the terrestrial broadcast channel, but those that are not used are useful for transmitting the same bit stream over other links, including cable links, computer networks, etc. In short, they provide interoperability features. The remainder of this Section explains the particular usage of these syntax elements.
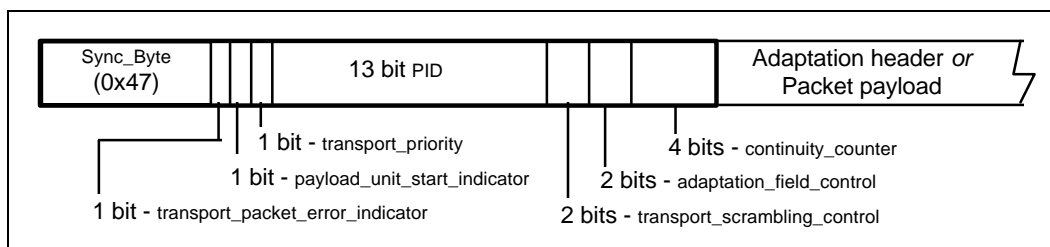
Figure 8.7. Link header format for the transport packet.

The transport_packet_error_indicator is available to signal errors from the modem to transport demultiplexer. If this bit is set the payload of the packet should not be used. As the digital television system does not support prioritized delivery, the transport_priority flag is ignored in the receiver. However this bit could be set during the packetization

process, if the transport layer is also used to simultaneously feed a prioritized delivery system. The payload_unit_start_indicator is used to signal that the packet payload contains the start of a structure which encapsulates the payload data. For audio and video data, this is the PES packet, and for PSI data this is the start of a table.

The adaptation_field control bits signal whether or not an adaptation header field is present and if so, whether or not it is accompanied by a payload in the same transport packet. No adaptation headers are allowed for transport packets containing PSI information other than for signaling with the discontinuity_indicator that the version_number (Section 2.4.4.4 of ISO/IEC 13818-1) may be discontinuous. This restriction allows the PMT bandwidth to be limited to a reasonable level without significantly restricting the opportunities and locations for private data to be transmitted.

The transport_scrambling_control indicators have been defined for the digital television system as shown in Table 8.1.

Table 8.1 Transport Scrambling Control Indicators

transport_ scrambling_ control	Function
00	packet payload not scrambled
01	not scrambled, state may be used as a flag for private use defined by the service provider.
10	packet payload scrambled with "even" key
11	packet payload scrambled with "odd" key

8.4.2 Adaptation level headers

The presence of the adaptation field is signaled by the adaptation_field_control bits of the link level header as described before. The adaptation header itself consists of information useful for higher level decoding functions. The header format is based on the use of flags to indicate the presence of the particular extensions to the field.

The header starts with a fixed-length 2-byte component that is always present (if an adaptation header is transmitted). The format is shown in Figure 8.8.
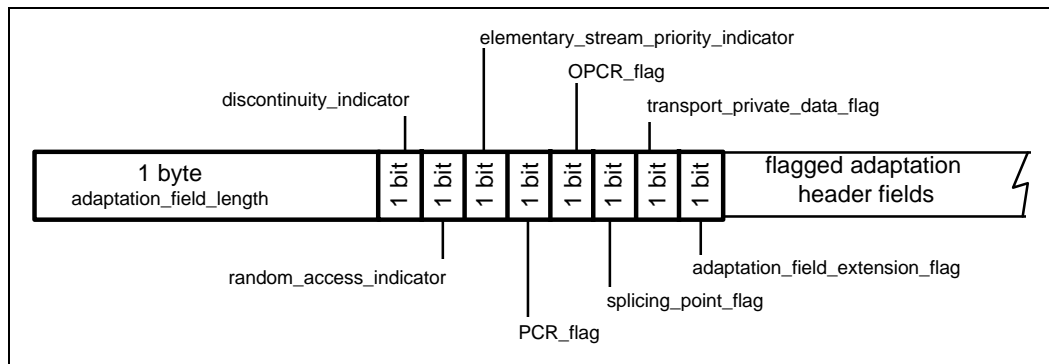
Figure 8.8. Format for the fixed-length component of the adaptation header.

The adaptation_field_length specifies the number of bytes that follow it in the adaptation header. The adaptation header may include stuffing bytes after the last adaptation header component field. (Stuffing bytes have a value of 0xFF and are not

interpreted at the decoder.) If stuffing bytes are included, the adaptation_field_length includes them in the count. The value in the adaptation_field_length field can also be used by the decoder to skip over the entire adaptation header, and to directly advance to the data payload in the packet if desired.

The presence of additional adaptation header fields is indicated by the state of the last five single bit flags shown in Figure 8.8 (with a value of '1' indicating that a particular field is present). The three flags at the beginning do not result in extensions to the adaptation header. The elementary_stream_priority_indicator is not used in the system, and, if set, would be ignored by the decoder.

As mentioned earlier, the other components of the adaptation header appear based on the state of the flags shown in Figure 8.8. The order in which these components appear in the bit stream is the same as the order of the flags. Based on the type of adaptation header information being conveyed, the data in these fields may be either fixed-length or variable-length. These fields are described in detail next.

8.4.2.1 The PCR and OPCR fields

The use of the PCR has been described in detail in Section 8.2.2.1. The PCR is used for synchronization of the system decoding process. Under certain circumstances this field may have to be modified during the transmission process. The PCR will be transmitted at least once every 100 milliseconds.

The original program clock reference (OPCR) indicates the intended time of arrival of the last byte of the original_program_clock_reference_base at the target decoder for a single program. This field is not modified during transmission and may be used for recording and playback of single programs. The OPCR is not used by the receiver in the decoding process.

8.4.2.2 The splice_countdown field

This field is useful for local program insertion as described in Section 8.2.2.3.2. The splice_countdown field indicates the number of packets, with the same PID as the current packet, that remain in the bit stream until a splicing point packet. It thus signals to the receiver the point at which a switch from one program segment to another may occur. Similarly, at the end of an inserted program segment, the count in this field signals the receiver when the inserted segment will end and when the original program will resume. The splicing point packet is defined as the packet containing a point in the elementary bit stream from which point onwards data can be removed and replaced by another bit stream, so that the resulting transport bit stream is valid according to MPEG-2 rules.

8.4.3 PSIs and the pointer_field

The program_association_table and the program_map_tables, as mentioned in Section 8.3, describe the organization of a multiplexed bit stream and are a part of the PSI layer of the digital television system. PSI tables, in general, are transmitted in the appropriate bit stream sequentially, without any gap between the tables. This implies that tables do not

necessarily start at the beginning of a transport packet. This also implies that in order to decode specific tables, there needs to be some indication of where they begin in the bit stream. This functionality is achieved with the pointer_field. The pointer_field, if present, is the first byte of the payload of a packet (after the link and adaptation headers). The pointer_field is present in the packet if a PSI table begins in the packet, an event which is signaled at the link level, by setting the payload_unit_start_indicator to '1'. The pointer_field indicates the number of bytes that follow it before the start of a PSI table. As an example a pointer_field value of 0x00 indicates that a new PSI table begins immediately following it.

8.4.4 The program_association_table

As discussed in Section 8.3.2, the program_association_table is transmitted as the payload of the bit stream with PID=0 and describes how program numbers that are associated with programs, (e.g., PBS, C-SPAN, etc.) map on to bit streams containing the program_map_tables for these programs. The program_association_table may be transmitted as multiple program_association_segments with each segment having a maximum length of 1024 bytes. The transport decoder can extract individual table segments from the bit stream in whatever order it desires.

The variable-length component of the table consists of program_count number of fixed-length entries corresponding to each program, and stuffing_bytes (to make up the program_association_segment_length).

The program identity '0' is reserved for the network_PID, i.e., the PID of the bit stream carrying information about the configuration of the overall system. As discussed in Chapter 7, the Network PID is available for use by broadcasters to provide information relating to the programs currently being transmitted, and also for programs that will be transmitted at future times. For all other program identities, the program_map_PID is the PID of the bit stream containing the program_map_table for the particular program.

8.4.5 The program_map_table

As discussed in the previous Section, the program_map_table is transmitted as the payload of the bit stream with PID=program_map_PID (as indicated in the program_association_table). The program_map_table carries information about the applications (e.g., video, audio, etc.) that make up programs. The format of the first 8 bytes is the same and has similar functionality as that for the program_association_table. The similarity in format is intended to facilitate simple software decoding of the headers.

8.4.6 Descriptors

To describe certain characteristics of the program or the elementary bit stream, descriptors are transmitted in the program_descriptor and the elementary_stream_descriptor fields. In general, each program_descriptor and the elementary_stream_descriptor can consist of a number of individual descriptor field elements transmitted sequentially.

Two factors need to be considered in order to use descriptors. In the first place, there has to be a mechanism for indicating the presence of the descriptors. In the PSI tables

that have been described, this functionality is achieved by the length field that precedes the descriptor, with a value of zero indicating that no descriptor is present. A second function is the identification of the descriptor itself. This is achieved within the descriptor header, which consists of a one byte descriptor_tag field followed by a one byte descriptor_length field that specifies the number of bytes in the descriptor following the descriptor_length field. The set of valid descriptor_tags is the same as that defined for MPEG-2. Additional descriptor_tags specific to the Digital Television Standard are defined in Section 5.7.2 of Annex C to the Standard.

8.4.7 The PSI paradigms and constraints

The time required to acquire programs is of concern in the digital television system design as NTSC users are accustomed to a rapid response time from the receivers when they switch channels. The approach used to ensure rapid access is a combination of defining a repetition rate for transmission of different PSI bit stream elements and assigning PID values to bit streams that are related to the program number under consideration. The program paradigms are such that recognition of the paradigm helps speed up the acquisition of programs in the receiver. A receiver that is unaware of the program paradigms will also be able to acquire the program, except that it could take a longer time to complete the acquisition process.

8.4.7.1 The program paradigms

The program paradigms provide a simple mechanism for quickly recognizing the programs of interest and further identifying the transport bit streams containing elements of the program thus enabling quick acquisition. The basic philosophy is to select PIDs for the transport bit stream that are related to a program number. The program number within a single transport stream serves as the complete identification of the program.

Television programs can only be assigned program numbers 1 to 255. Non-television programs are also not allowed to use these particular program numbers. This allows the receiver to filter the television programs quickly by looking at the MSB of the program number. This process is further aided by specifying television programs to be described only within section 0 of the program_association_table.

We further define:

$$\text{base\_PID} = \text{program number} << 4 \; \textit{(pseudo C code)}$$

where program number refers to each program within one transport multiplex and corresponds to the 16-bit program_number identified in PAT and PMT.

The paradigm to identify the transport bit streams containing certain elements of the program is defined as shown in Table 8.2. Table 8.3 illustrates an example for program 52.

Table 8.2 Transport Bit Streams

Name	PID Definition	Description
PMT_PID	base_PID+0x0000	PID for the bit stream containing the program_map_table for the program.
Video_PID	base_PID+0x0001	PID for the bit stream containing the video for the program.
PCR_PID	base_PID+0x0001	Implies the video bit stream also carries the PCR values for the program
Audio_PID	base_PID+0x0004	PID for the bit stream containing the primary audio for the program. The primary audio shall be a complete main audio service (CM) as defined by ATSC Standard A/52 and shall contain the complete primary audio of the program including all required voice-overs and emergency messages.
Data_PID	base_PID+0x000A	PID for the bit stream containing the data for the program.

Table 8.3 Example for Program 52 (0x0034)

Name	PID value
base_PID	0x0340
PMT_PID	0x0340
Video_PID	0x0341
PCR_PID	0x0341
Audio_PID	0x0344
Data_PID	0x034A

This paradigm enables immediate access to services defined in the paradigm once the program number is known. The program_map_table must be decoded to obtain the PIDs for services not defined by the paradigm but included within the program (such as additional audio service types or a second data channel). The requirement that the audio bit stream identified by the program paradigm be a complete main (CM) channel including all required voice-overs and emergency messages ensures that simple receivers capable of decoding only a single audio bit stream will always receive a complete audio program. Access to all other audio services will require that the program_map_table be decoded to obtain the PIDs for these services. Examples include services that can be decoded by simple receivers, such as a second complete main (CM) channel in another language, as well as services that can be decoded only by receivers that are able to simultaneously decode and add together two or more audio bit streams such as music and effects (ME) plus dialogue (D). According to the program paradigm, every 16th PID is a PMT_PID and may be assigned to a program. If a PMT_PID is assigned to a program by the program paradigm, the next 15 PIDs after that PMT_PID are reserved for elements of that program and shall not be otherwise assigned.

Additional constraints include:

1. Only one program is described in a PSI transport bit stream corresponding to a particular PID value. A transport bit stream containing a program_map_table may not be used to transmit any other kind of PSI table (identified by a different table_id).

2. Adaptation headers shall not occur in transport packets of the PMT_PID for purposes other than for signaling with the discontinuity_indictor that the version_number (Section 2.4.4.5 of ISO/IEC 13818-1) may be discontinuous.

8.4.7.2 Repetition rates

The maximum spacing that is allowed between occurrences of section 0 of the program_association_table is 100 ms.

The maximum spacing allowed between occurrences of a program_map_table containing television program information is 400 ms.

8.5 The PES packet format

The PES Packet may be generated by either the application encoder or the transport encoder; however, for the purposes of explanation the PES encoding is assumed to be a transport function. As described before, some elementary bit streams, including the compressed video and compressed audio, will go through a PES layer packetization process prior to transport layer packetization. The PES header carries various rate, timing, and data descriptive information, as set by the source encoder. The PES packetization interval is application dependent. The resulting PES packets are of variable-length with a maximum size of 2^{16} bytes, when the PES packet length field is set to its maximum value. This value is set to zero for the video stream, indicating that the packet size is unconstrained and that the header information cannot be used to skip over the particular PES packet. Note also that the PES packet format has been defined to also be of use as an input bit stream for Digital Storage Media (DSM) applications. Although the DSM format will not be used for broadcast applications, some of the PES header fields related to the DSM functions are also described in this Section. Note that the ability to handle input bit streams in the DSM format is not essential for a receiver, but may be useful for VCR applications.

Note that the format for carrying the PES packet within the transport layer is a subset of the general definition in MPEG-2. These choices were made to simplify the implementation of the digital television receiver and to also help error recovery.
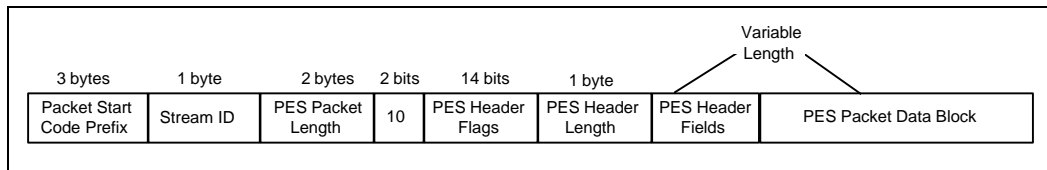

Figure 8.9. Structural overview of a PES packet.

A PES packet consists of a PES_packet_start_code, PES header flags, PES packet header fields, and a payload (or data block), as shown in Figure 8.9. The payload is created by the application encoder. The packet payload is a stream of contiguous bytes of a single elementary stream. For video and audio packets, the payload is a sequence of

access units from the encoder. The access units correspond to the video pictures and audio frames.

Each elementary stream is identified by a unique stream_id. The PES packets formed from elementary streams supplied by each encoder carry the corresponding stream_id. PES packets carrying various types of elementary streams can be multiplexed to form a program or transport stream in accordance with the MPEG-2 Systems standard.

PES packets for video, including new PTS and DTS values, occur once every picture (or video access unit). The PES packets are also aligned to the first occurrence of a sequence, a GOP or a picture start code after the end of a picture, i.e., the first bytes seen within the payload of a video PES packet belong to either a sequence, a GOP or a picture start code. Further, new PES packet data always starts a new transport packet, and stuffing bytes are used in the adaptation header of the transport packets to ensure that PES packets always end on transport packet boundaries.

8.5.1 PES header flags

A breakdown of the PES header flags is shown in Figure 8.10. These flags are a combination of indicators of the properties of the bit stream and indicators of the existence of additional fields in the PES header. The following table describes the flags present in the header. The flags not supported by the system are set to '0' and form the basis of some of the "constraints" discussed earlier. (These entries are shaded in the table.)
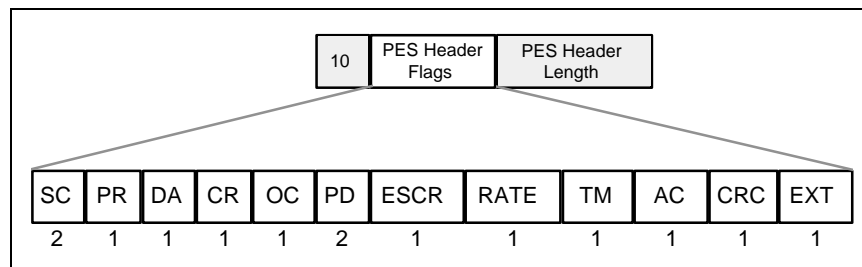
Figure 8.10. PES header flags in their relative positions (all sizes in bits).

8.5.2 The PES header

The PES header immediately follows the PES_header_length field, which indicates the header size in bytes. The size of the header includes all the header fields, any extension fields, and stuffing_bytes. The flags described in the previous Section indicate the organization of the PES header, i.e., which fields it does and does not contain. In essence, all the fields of the PES header are optional. Certain applications require particular fields to be set appropriately. For example, transport of video PES packets in the digital television system requires that the data_alignment_indicator be set. The trick mode flag is not set in this case. For DSM retrieval of video, the opposite is true. It is the application encoder's function to provide the information required to set the appropriate flags, and encode the corresponding fields. The fields are further described in the following Sections. The association between the flags and the corresponding fields is obvious.

The PES header fields are organized according to Figure 8.11 for the PES packets for video elementary streams. Most fields require marker bits to be inserted, as described later, in order to avoid the occurrence of long strings of 0's which could resemble a start code.
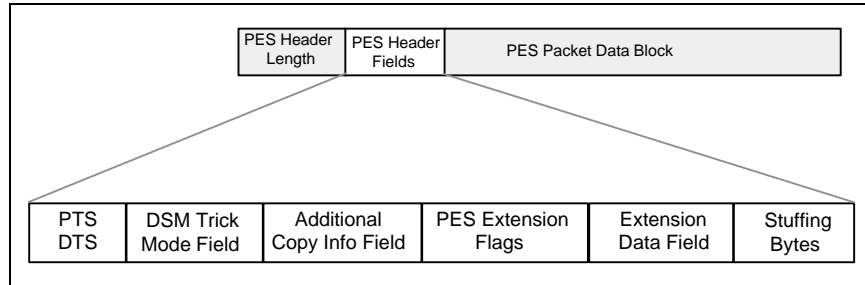
Figure 8.11. Organization of PES header.

8.5.2.1 PTS and DTS

The presentation_time_stamp (PTS) informs the decoder of the intended time of presentation of a presentation unit, and the decoding_time_stamp (DTS) is the intended time of decoding of an access unit. An access unit is an encoded presentation unit. When it is encoded, the PTS refers to the presentation unit corresponding to the first access unit occurring in the packet. If an access unit does not occur in a PES packet, the header shall not contain a PTS. An audio access unit occurs if the first byte of the synchronization word of an audio frame is present. A video access unit may be either a video sequence, a GOP, or a picture header as defined in Section 2.1.1 of ISO/IEC 13818-1. For terrestrial broadcast, each video frame is PES packetized. Under normal conditions, the DTS may be derived from the PTS. The DTS is not encoded when its value is equal to the value of the PTS, such as when the frame is a B-frame or when there are no B-frames in the sequence and hence no frame reordering delay. Under no circumstance does the DTS occur by itself; it must occur along with the PTS, although the converse is not true. The PTS field is organized as shown in Figure 8.12, if it present without the DTS.
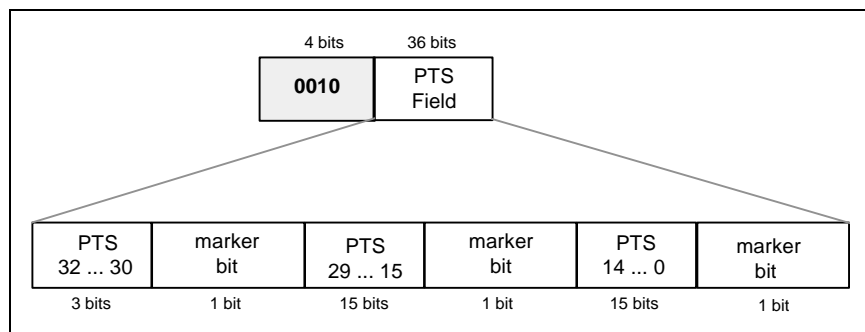
Figure 8.12. Organization of the PTS field when only the PTS is encoded.

If both the PTS and DTS are sent, the organization of Figure 8.13 is required. Here, the DTS field is defined in the same manner as the PTS field.

4 bits	36 bits	4 bits	36 bits
0011	PTS Field	**0001**	DTS Field
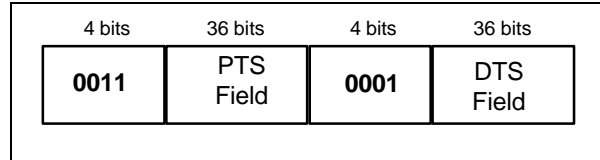
Figure 8.13. Organization of the PTS and DTS field when both PTS and DTS are encoded.

8.5.2.2 PES extension flags

The header may contain additional flags if the EXT flag bit (shown in Figure 8.10) is set. These flags are transmitted in a one byte data field as shown in Figure 8.14.

The flags indicate whether further extensions to the PES header exist. As with the flags defined previously, the flag is set to '1' if the header field is present.

PES private data flag	pack header field flag	program packet sequence counter flag	STD buffer flag	Reserved	PES extension field flag
1 bit	1 bit	1 bit	1 bit	3 bits	1 bit
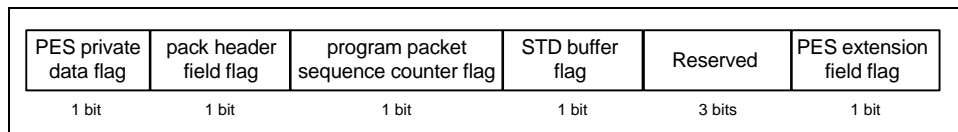
Figure 8.14. Organization of the PES extension flags field.

8.5.3 Conditional access

The transport protocol implements functions useful for supporting conditional access. The functionality that is available is flexible and complete in the sense of supporting all transmission aspects of applicable key encryption and scrambling approaches that may be used. Conditional access is also flexible in the sense that it can be applied separately to each elementary stream providing the ability to selectively scramble the different elementary streams (e.g., audio, video, etc.) in a program if desired.

A conditional access system operates on the principle of randomizing the transmitted data so that unauthorized decoders cannot decode the signal. Authorized decoders receive a "key" which initializes the circuit which inverts the bit randomization. In subsequent discussion, we use the term scrambling to mean the pseudo-random inversion of data bits based on a "key" which is valid for a short time. We use the term encryption to mean the process of transforming the "key" into an encrypted key by a means which protects the key from unauthorized users. From a cryptographic point of the view, this transformation of the key is the only part of the system which protects the data from a highly motivated pirate. The scrambling portion of the process alone, in the absence of key encryption, can be defeated. Conditional Access (CA) is a blanket term for the system which implements the key encryption and distribution. By virtue of the function of a CA system, the details remain proprietary, however the EIA and NCTA are working to standardize interfaces within consumer equipment to meet the needs of secure applications. The emerging interface standard is referred to as the National Renewable Security System (NRSS).

There are three features of the digital television transport system that support conditional access. The first feature is the two bit transport_scrambling_control field which signals the decoder whether the transport packet is scrambled or not. In the case that it is scrambled, the field identifies which of the scrambling keys was used. The second feature is the ability to insert "private" data at several places in the transport stream. These include entirely private streams and private fields in the adaptation header of the transport bit stream being scrambled. These private fields can be used to transmit the encrypted scrambling key to the decoding device. Thirdly, PID1 is used to identify those private streams which carry program authorizations.

An example of the key distribution process is shown in Figure 8.15. When the bit stream is scrambled, one descrambling key needs to be in use while the other is being received and decrypted. Two keys are transmitted at any time, with the keys being linked to a transport_scrambling_control value as shown in Figure 8.15. The transmission of a key should begin well before it is going to be used, to allow time to decrypt it. Note that this function does not bound the total number of keys that may be used during an entire transmission session. The proper key to be used to descramble is signaled in the transport prefix in the transport_scrambling_control field. The transport_scrambling_control takes on one of the 4 states shown in Table 8.1.
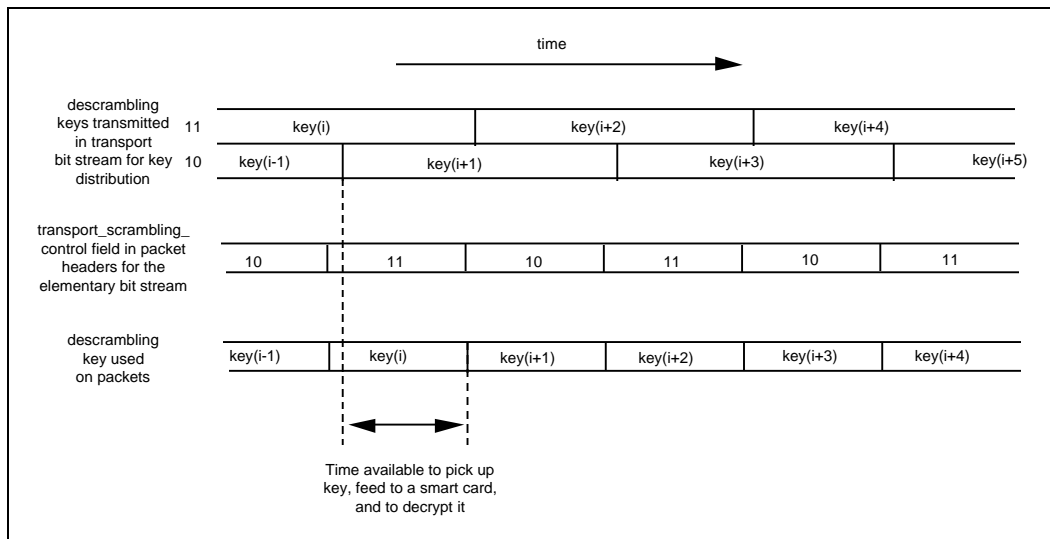
Figure 8.15. Illustration of key distribution and usage process.

8.5.3.1 Conditional access example

Figure 8.16 illustrates an example of a digital television receiver that is operating in an environment which uses an NRSS device to implement the conditional access functions. The NRSS interface provides for 50 Mbps of throughput through a security device. The host microprocessor in the digital television receiver decoder communicates with the security device as to the desired program, etc., over the Serial In/Out control channel. The NRSS device itself filters the transport data stream, presented at the Din input, for all information which pertains to the CA functions. The secure processor decodes the CA

information required to descramble the selected service. The security system algorithms and circuits decode the Entitlement Control Message (ECM) and Entitlement Management Message (EMM) information and supply the scrambling keys to the descrambling block, shown here as a Data Encryption Standard (DES) descrambler. The descrambled data is returned to the digital television receiver on the Dout output connector.
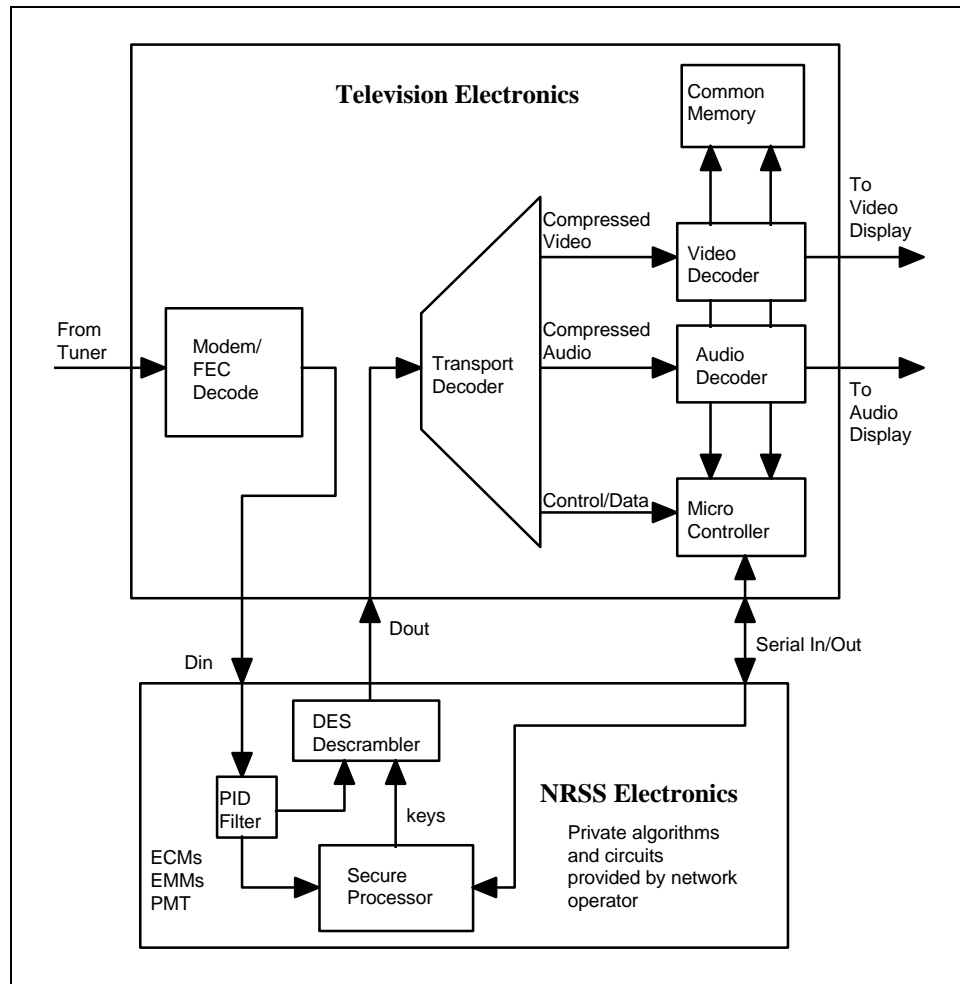
Figure 8.16. Example NRSS based A/V decoder in a secure service environment.

8.5.4 Compatibility with other transport systems

The transport system interoperates with two of the most important alternative transport systems. It is syntactically identical with the MPEG-2 transport stream definition, with the Digital Television Standard being a subset of the MPEG-2 specification. Annex C of the Digital Television Standard completely specifies the special constraints and requirements of the subset. The transport system also has a high degree of interoperability with the ATM definition being finalized for Broadband ISDN. Furthermore, as several of the cable television and Direct Broadcast Satellite (DBS) systems currently in use or being designed employ MPEG-2 transport layer syntax, the

degree of interoperability with such deployed systems should be quite high (possibly requiring a translation if the cable television or DBS system deploys a slightly incompatible MPEG-2 variant).

Interoperability has two aspects. The first is syntactic and refers only to the coded representation of the digital television information. The second relates to the delivery of the bit stream in real time. This aspect of interoperability is beyond the scope of this discussion, but it should be noted that to guarantee interoperability with a digital television receiver conforming to the Standard, the output bit stream of the alternative transport system must have the proper real-time characteristics.

8.5.4.1 Interoperability with MPEG-2

In the development of the digital television transport specification, the intent has never been to limit the design by the scope of the MPEG-2 Systems definition. The system is interoperable with MPEG-2 decoders as the transport is a constrained subset of the MPEG-2 transport syntax. The constraints are imposed for reasons of increased performance with respect to channel acquisition, bandwidth efficiency and decoder complexity.

The system also supports bit streams and services beyond the compressed video and audio services, e.g., text-based services, emergency messages, and other future ancillary services. A means of identifying such bit streams is necessary, but is not part of the MPEG-2 definition. There is a method of encoding such a registration descriptor when an authority to administrate registration and catalog registered codes is identified. The method of encoding the registered value is by means of the registration_descriptor in the PSI stream.

8.5.4.2 Interoperability with ATM

The MPEG-2 transport packet size is such that it can be easily partitioned for transfer in a link layer that supports Asynchronous Transfer Mode (ATM) transmission. The MPEG-2 transport layer and the ATM layer serve different functions in a video delivery application. The MPEG-2 transport layer solves MPEG-2 presentation problems and performs the multi-media multiplexing function. The ATM layer solves switching and network adaptation problems.

There are several possible methods for mapping the MPEG-2 transport packet into the ATM format, and international standards organizations are standardizing the method to be used in different application domains. A popular method for constant bit rate (CBR) sources in the video-on-demand (VOD) application, under standardization by the ITU-T and the ATM Forum, is to perform cell aligned packet mapping with an AAL-5 Packet Data Unit (PDU). This technique is presented here to serve as an example.

8.5.4.2.1 ATM cell and transport packet structures

The ATM cell consists of two parts: a five-byte header and a forty-eight-byte information field. The header, primarily significant for networking purposes, consists of the fields shown in Table 8.4.

Table 8.4 ATM Cell Header Fields

GFC	A four bit Generic Flow Control field used to control the flow of traffic across the User Network Interface (UNI). Exact mechanisms for flow control are under investigation.
VPI	An eight bit network Virtual Path Identifier.
VCI	A sixteen bit network Virtual Circuit Identifier.
PT	A three bit Payload Type (i.e., user information type ID).
CLP	A one bit Cell Loss Priority flag (eligibility of the cell for discard by the network under congested conditions).
HEC	An eight bit Header Error Control field for ATM header error correction
AAL	ATM Adaptation Layer bytes (user specific header).

The ATM User Data Field consists of forty-eight bytes, where up to four of these bytes can be allocated to an Adaptation Layer.

Figure 8.17 illustrates the differences between the format of an ATM cell and the format of the MPEG-2 transport packet.
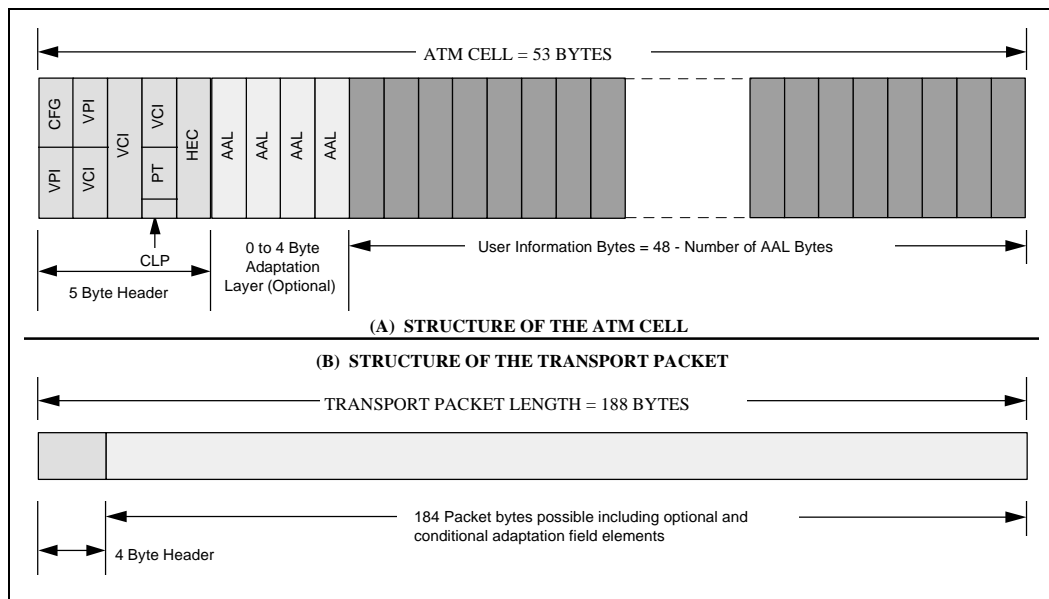

Figure 8.17. Comparison of the ATM cell structure and the MPEG-2 transport packet structure.

8.5.4.2.2 Mapping transport packets to an AAL-5 packet data unit (PDU)

A preferred solution to cell/packet alignment is to encapsulate one or two transport packets into a AAL-5 PDU structure. This is illustrated in Figure 8.18. Two

MPEG-2 transport packets, consisting of 376 bytes, are concatenated and the AAL-5 PDU trailer is computed and appended. The PDU is then segmented and the segments are inserted as the payload of 8 successive ATM cells. The ATM cell which contains the PDU header is identified by a special value in the PT field of the ATM cell header. Two processes can contribute to the time jitter of each MPEG-2 packet as it traverses the ATM network. The first is that ATM cells may be delayed at network nodes as part of the ATM cell scheduling and routing algorithms. This component is not controllable by the encoder-side equipment. The second component is the packetization delay in forming the 2-packet (8-cell) PDU. As the PCR value represents a real-time sample of the Program Clock Reference, jitter in packets containing the PCR is undesirable. The value of jitter is manifest in the decoder as a noise source at the input to the PLL which is used to reconstruct the audio and video sample clocks. This jitter can result in audible and visible artifacts if the effects are not properly mitigated. To ease the problem of jitter mitigation, it is common practice to add the following restriction to PDU creation: transport packets which carry a PCR sample shall be the last packet in a PDU. Consequently, if a new PDU is being formed and a PCR is in the first transport packet, then the encoder will issue a short (4 cell) PDU in order to limit unnecessary contributions to the PCR jitter.
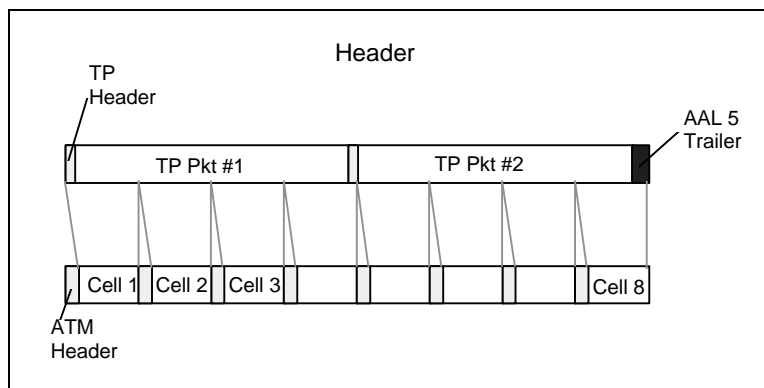
**Figure 8.18. Mapping of two MPEG-2 transport packets
into 8 ATM cells using the AAL-5 PDU.**

Table 8.5 Elements of the AAL-5 PDU Trailer

PDU Field	Size(bytes)	Function
UU	1	User to User information to allow the encoder application to signal the decoder application with application dependent information.
CPI	1	Reserved for future use.
Length	2	16-bit binary number which represents the length of the PDU payload. It allows padding to follow the PDU payload in applications which require it.
CRC-32	4	Cyclic Redundancy Check calculated over the payload of the PDU.

The AAL 5 PDU trailer performs error detection and segmentation and reassembly (SAR) functions. The PDU trailer is an 8-byte structure, with the elements summarized in Table 8.5. The UU field may be defined by the ITU-T and/or the ATM Forum with specific information for the VOD application, as could the CPI field. This may be a long

term enhancement, but early adopters of MPEG/ATM systems will probably restrict consideration to systems which use generic AAL-5 functions. The length field is important to locate any padding which may be required to pad the PDU to be an incremental number of ATM cells. Padding will be required when it is necessary to issue a short PDU because a PCR packet has arrived at the PDU assembly processor. The CRC-32 provides an end-to-end error detection function.

9. RF/TRANSMISSION SYSTEMS

9.1 Introduction and system overview

The VSB system offers two modes: a simulcast terrestrial broadcast mode, and a high data rate mode. The two modes share the same pilot, symbol rate, data frame structure, interleaving, Reed-Solomon coding, and synchronization pulses. The terrestrial broadcast mode is optimized for maximum service area, and supports one ATV signal in a 6 MHz channel. The high data rate mode, which trades off some robustness for twice the data rate, supports two ATV signals in one 6 MHz channel.

Both modes of the VSB transmission subsystem take advantage of a pilot, a segment sync, and a training sequence for robust acquisition and operation. The two system modes also share identical carrier, sync, and clock recovery circuits, as well as phase correctors and equalizers. Additionally, both modes use the same Reed-Solomon (RS) code for forward error correction (FEC).

In order to maximize service area, the terrestrial broadcast mode incorporates both an NTSC rejection filter (in the receiver) and trellis coding. Pre-coding at the transmitter is incorporated in the trellis code. When the NTSC rejection filter is activated in the receiver, the trellis decoder is switched to a trellis code corresponding to the encoder trellis code concatenated with the filter.

The high data rate mode, on the other hand, does not have as severe an environment to work in as that of the terrestrial system. Therefore, a higher data rate is transmitted in the form of more data levels (bits/symbol). No trellis coding or NTSC interference rejection filters are employed.

VSB transmission inherently requires only processing the in-phase (I) channel signal, sampled at the symbol rate, thus optimizing the receiver for low cost implementation. The decoder only requires one A/D converter and a real (not complex) equalizer operating at the symbol rate of 10.76 Msamples/s.

The parameters for the two VSB transmission modes are shown in Table 9.1.

9.2 Bit rate delivered to a transport decoder by the transmission subsystem

The exact symbol rate of the transmission subsystem is given by:

$$(1) \quad 4.5/286 \times 684 = 10.76... \text{ MHz}$$

The symbol rate must be locked in frequency to the transport rate. The transmission subsystem carries 2 information bits per trellis-coded symbol, so the gross payload is:

$$(2) \quad 10.76... \times 2 = 21.52... \text{ Mbps}$$

To find the net payload delivered to a decoder it is necessary to adjust (2) for the overhead of the Data Segment Sync, Data Field Sync, and Reed-Solomon FEC.

Upon doing this the net payload bit rate of the 8 VSB terrestrial transmission subsystem becomes:

(3) 21.52... Mbps x 312/313 x 828/832 x 187/207 = 19.28... Mbps

The factor of 312/313 accounts for the Data Field Sync overhead of one Data Segment per field. The factor of 828/832 accounts for the Data Segment Sync overhead of four symbol intervals per Data Segment, and the factor of 187/207 accounts for the Reed-Solomon FEC overhead of 20 bytes per Data Segment.

The calculation of the net payload bit rate of the high data rate mode is identical except that 16 VSB carries 4 information bits per symbol. Therefore, the net bit rate is twice that of the 8 VSB terrestrial mode:

(4) 19.28... Mbps x 2 = 38.57... Mbps

To get the net bit rate seen by a transport decoder, however, it is necessary to account for the fact that the MPEG sync bytes are removed from the data stream input to the 8 VSB transmitter. This amounts to the removal of one byte per data segment. These MPEG sync bytes are then reconstituted at the output of the 8 VSB receiver. The net bit rate seen by the transport decoder is:

(5) 19.28... Mbps x 188/187 = 19.39... Mbps

The net bit rate seen by the transport decoder for the high data rate mode is:

(6) 19.39... Mbps x 2 = 38.78... Mbps

Table 9.1 Parameters for VSB Transmission Modes

Parameter	Terrestrial mode	High data rate mode
Channel bandwidth	6 MHz	6 MHz
Excess bandwidth	11.5%	11.5%
Symbol rate	10.76 Msymbols/s	10.76 Msymbols/s
Bits per symbol	3	4
Trellis FEC	2/3 rate	None
Reed-Solomon FEC	T=10 (207,187)	T=10 (207,187)
Segment length	832 symbols	832 symbols
Segment sync	4 symbols per segment	4 symbols per segment
Frame sync	1 per 313 segments	1 per 313 segments
Payload data rate	19.28 Mbps	38.57 Mbps
NTSC co-channel rejection	NTSC rejection filter in receiver	N/A
Pilot power contribution	0.3 dB	0.3 dB
C/N threshold	14.9 dB	28.3 dB

9.3 *Performance characteristics of terrestrial broadcast mode*

The terrestrial VSB system can operate in a signal-to-additive-white-Gaussian-noise (S/N) environment of 14.9 dB. The 8 VSB, 4-state segment error probability curve in Figure 9.1 shows a segment error probability of 1.93 x 10^{-4}. This is equivalent to 2.5

segment errors/second which has been established by measurement as the threshold of visibility of errors.

Note: Threshold of visibility (TOV) has been measured to occur at a S/N of 14.9dB when there are 2.5 segment errors per second which is a segment error rate (SER) of 1.93×10^{-4}.

SER at TOV
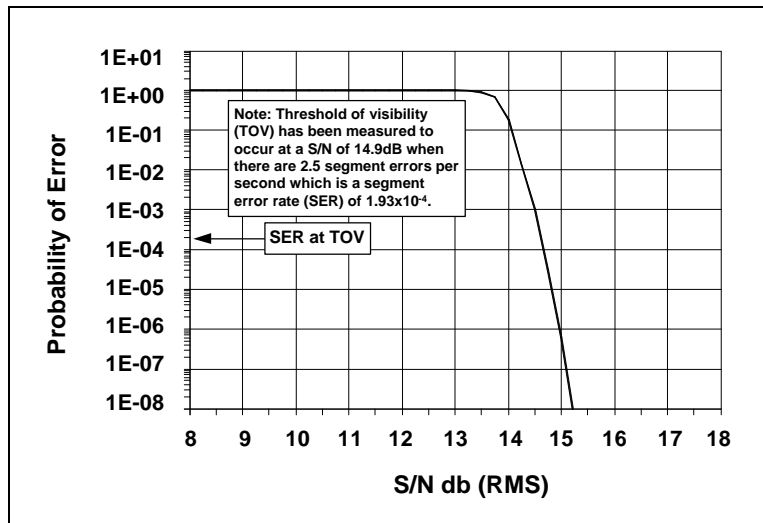
Figure 9.1. Segment error probability, 8 VSB with 4 state trellis, RS (207,187).

The cumulative distribution function (CDF) of the peak-to-average power ratio, as measured on a low power transmitted signal with no non-linearities, is plotted in Figure 9.2. The plot shows that 99.9% of the time the transient peak power is within 6.3 dB of the average power.
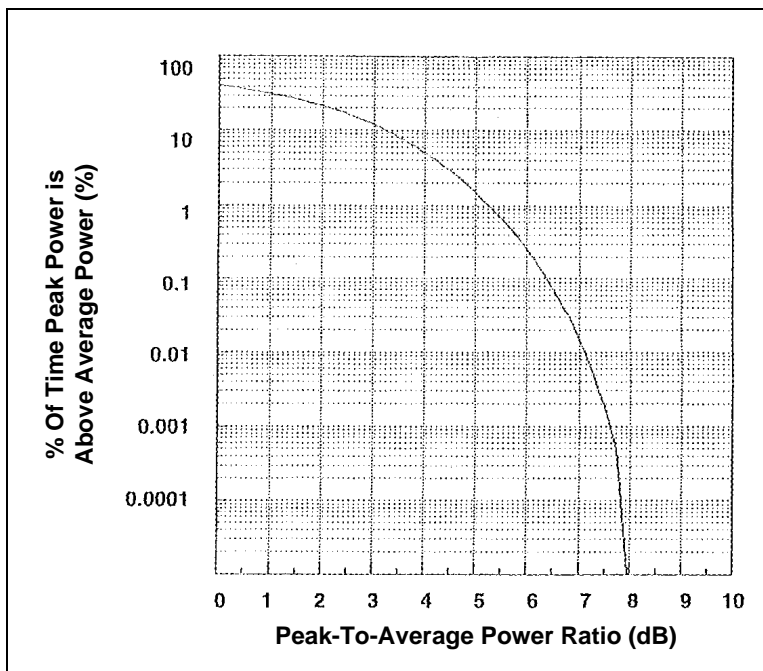

**Figure 9.2. Cumulative distribution function of 8 VSB
peak-to-average power ratio.**

9.4 Transmitter signal processing

A pre-equalizer filter is recommended for use in over-the-air broadcasts where the high power transmitter may have significant in-band ripple or roll off at band edges. This linear distortion can be detected by an equalizer in a reference demodulator ("ideal" receiver) located at the transmitter site that is receiving a small sample of the antenna signal feed provided by a directional coupler which is recommended to be located at the sending end of the antenna feed transmission line. The reference demodulator equalizer tap weights can be transferred into the transmitter pre-equalizer for pre-correction of transmitter linear distortion.

A suitable pre-equalizer is an 80 tap, feed-forward transversal filter. The taps are symbol spaced (93 ns) with the main tap being approximately at the center, giving approximately ±3.7 µs correction range. It operates on the I channel data signal (there is no Q channel data in the transmitter), and shapes the frequency spectrum of the IF signal so that there is a flat in-band spectrum at the output of the high power transmitter that feeds the antenna for transmission. There is no effect on the out-of-band spectrum of the transmitted signal.

The transmitter VSB filtering may be implemented by complex-filtering the baseband data signal, creating precision-filtered and stable in-phase and quadrature-phase modulation signals. This filtering process provides the root raised cosine Nyquist filtering as well as the sin x/x compensation for the D/A converters. The orthogonal baseband signals are converted to analog form (D/A converters) and then modulated on quadrature IF carriers to create the vestigial sideband IF signal by sideband cancellation (phasing method). The nominal frequency of the IF carrier (and small in-phase pilot) in the prototype hardware used in ACATS testing is 46.69 MHz, which is equal to the IF center frequency (44.000 MHz) plus the symbol rate divided by 4 (10.762 MHz / 4 = 2.6905 MHz). Additional adjacent-channel suppression (beyond that achieved by sideband cancellation) may be performed by a linear phase, flat amplitude response SAW filter. Other implementations for VSB filtering are possible which may include the prefilter of the previous section.

9.5 Upconverter and RF carrier frequency offsets

Modern NTSC TV transmitters use a two-step modulation process. The first step usually is modulation of the data onto an IF carrier, which is the same frequency for all channels, followed by translation to the desired RF channel. The VSB transmitter applies this same two-step modulation process. The RF upconverter translates the filtered flat IF data signal spectrum to the desired RF channel. For the same approximate coverage as an NTSC transmitter (at the same frequency), the average power of the ATV signal is 12 dB less than the NTSC peak sync power.

The frequency of the RF upconverter oscillator in ATV terrestrial broadcasts will typically be the same as that used for NTSC (except for NTSC offsets). However, in extreme co-channel situations, the ATV system is designed to take advantage of precise RF carrier frequency offsets with respect to the NTSC co-channel carrier. As the VSB data signal sends repetitive synchronizing information (segment syncs), precise offset

causes NTSC co-channel carrier interference into the VSB receiver to phase alternate from sync to sync. The VSB receiver circuits average successive syncs to cancel the interference and make data segment sync detection more reliable.

For ATV co-channel interference into NTSC, the interference is noise-like and does not change with precise offset. Even the ATV pilot interference into NTSC does not benefit from precise frequency offset because it is so small (11.3 dB below the data power) and falls far down the Nyquist slope (20 dB or more) of NTSC receivers.

The ATV co-channel pilot should be offset in the RF upconverter from the dominant NTSC picture carrier by an odd multiple of half the Data Segment rate. A consequential spectrum shift of the VSB signal into the upper adjacent-channel is required. An additional offset of 0, +10 kHz, or -10 kHz is required to track the principal NTSC interferer.

For ATV-into-ATV co-channel interference, precise carrier offset prevents possible misconvergence of the adaptive equalizer. If perchance the two ATV Data Field Sync signals should fall within the same data segment time, the adaptive equalizer could misinterpret the interference as a ghost. To prevent this, a carrier offset of $f_{seg}/2 = 6.47$ kHz is recommended for close ATV-into-ATV co-channel situations. This causes the interference to have no effect in the adaptive equalizer.

9.6 Performance characteristics of high data rate mode

The high data rate mode can operate in a signal-to-white-noise environment of 28.3 dB. The error probability curve is shown in Figure 9.3.
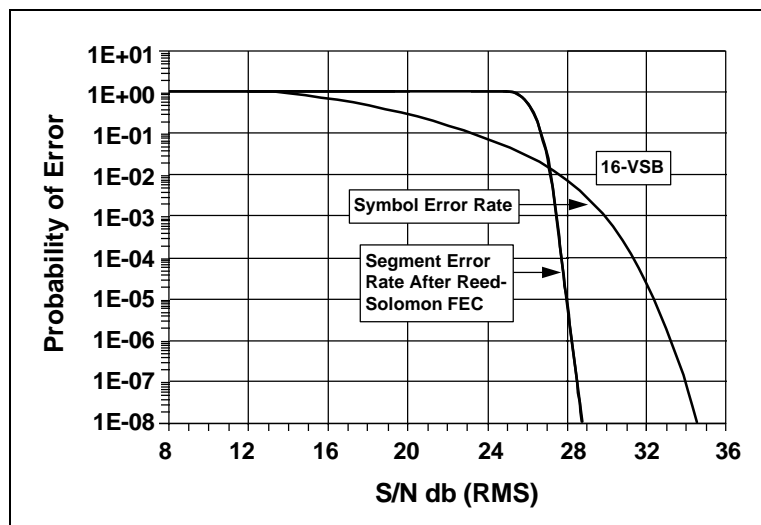

Figure 9.3. 16 VSB error probability.

The cumulative distribution function (CDF) of the peak-to-average power ratio, as measured on a low power transmitted signal with no non-linearities, is plotted in Figure 9.4 and is slightly higher than that of the terrestrial mode.
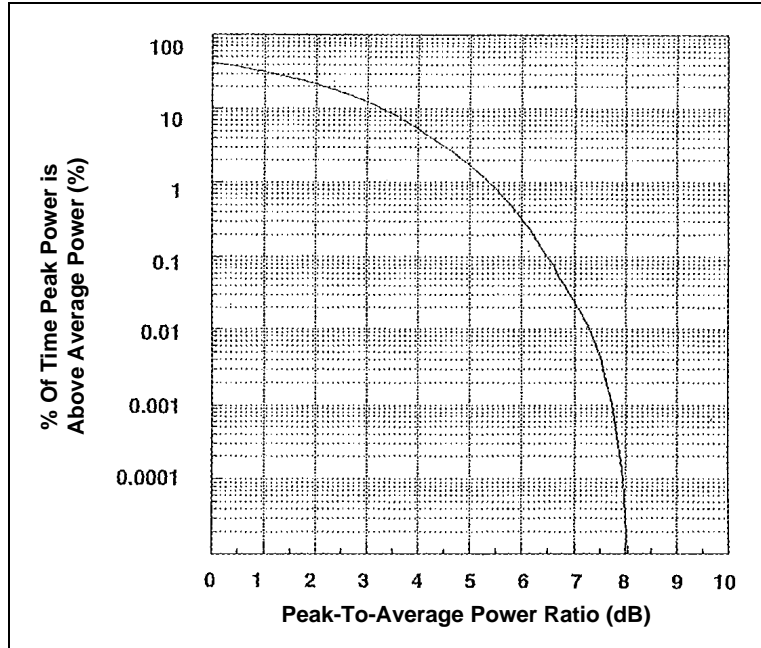
Figure 9.4. Cumulative distribution function of 16 VSB peak-to-average power ratio.

10. RECEIVER CHARACTERISTICS

10.1 Introduction

This Chapter describes the receiver characteristics and design considerations for receivers for the digital television system. It is designed as a tutorial intended for receiver manufacturers to serve as a guideline and literature reference. It illustrates design choices for receivers based on the system implementation example given by the Grand Alliance.

There is no practical implementation of a digital television receiver in existence at the time of release of this Guide. Only a first generation hardware prototype was implemented for laboratory and field tests. The information contained herein will, therefore, not yield a design for a production-ready receiver without substantial development effort by each manufacturer.

In the design of early digital television receivers it will be important to study the test methods, test conditions, and results that were designed and recorded by the Working Parties of the Planning and Systems Subcommittees of ACATS. They attempted to include in the testing all known and theoretically important reception conditions for signals. The data they gathered will allow consumer product designers to prepare a receiver design suitable for most commonly found conditions. Experience gained during the introduction of this new service is likely to teach the industry further important improvements for delivering the service and for receiver design.

10.2 Receiver RF issues

10.2.1 RF characteristics

10.2.1.1 Planning factors used by ACATS PS/WP3

The transmission subsystem is described in the ACATS report of February 24, 1994 under the heading RF/Transmission Characteristics. This summary provides background as to how it was selected and what planning factors were assigned to it.

The selection of a transmission subsystem comprised the following: first, a series of laboratory tests were executed at the Advanced Television Test Center (ATTC) to determine the performance limits of candidate transmission subsystems with respect to channel impairments including noise, interference and multipath. Second, the results of these tests and subsequent VSB modem tests, together with a set of receiver planning factors, were included in a nationwide spectrum utilization computer model which was developed under the direction of the Spectrum Utilization and Alternatives Working Party (PS/WP3) of the Planning Subcommittee of ACATS.

The results of the ATTC tests of transmission subsystems are summarized in Figure 1 of PS/WP3 Document 296 dated February 17, 1994. That figure summarizes the performance of 8 VSB and 32 QAM in the presence of: thermal noise; co-channel and adjacent-channel interference from ATV and NTSC; performance as a taboo interferer

into NTSC; and change in noise threshold due to specific ensemble multipath characteristics (which did not enter into the computer model).

The receiver planning factors that entered into the spectrum utilization computer model can be found in the Final Report of PS/WP3 under the heading "Receiver Planning Factors Applicable to All ATV Systems." Table 10.1 was taken from the latest document of PS/WP3, dated December 1, 1994 and shows these Planning Factors. It is also footnoted that "antenna factors are based on the geometric mean frequencies of the three broadcast bands," and that, in addition to F/B, "a formula is employed for the forward lobe simulating an actual receiving antenna pattern."

Table 10.1 Receiver Planning Factors Used by PS/WP3

Planning factors	Low VHF	High VHF	UHF
Antenna impedance (ohms)	75	75	75
Bandwidth (MHz)	6	6	6
Thermal noise (dBm)	-106.2	-106.2	-106.2
Noise figure: (dB)	10	10	10
Frequency (MHz)	69	194	615
Antenna factor (dBm/dBμ)	-111.7	-120.7	-130.7
Line loss (dB)	1	2	4
Antenna gain (dB)	4	6	10
Antenna F/B ratio (dB)	10	12	14

Taking both the ATTC results and receiver planning factors into account, the computer analysis was done which led PS/WP3 to conclude, in its Document 296, that, "as the natural outcome of employing as planning factors the results of the ATTC interference tests, the 8 VSB transmission subsystem shows an advantage over the 32 QAM subsystem both during the transition period when the spectrum is shared with NTSC and after the transition when only ATV will be broadcast."

The computer model calculates signal conditions at locations throughout the country, based upon transmitter locations, power, propagation models, etc. The performance numbers—such as C/N or D/U ratios—that entered into the computer model as the result of the ATTC tests will, in the real world, be a function of these signal conditions and the entire receiver installation including actual antenna, feedline and receiver performance.

10.2.1.2 Noise figure

A number of factors enter into the ultimate carrier-to-noise ratio within the receiver. For example, the receiver planning factors applicable to UHF used by PS/WP3 (reference Final Report of the Spectrum Utilization and Alternatives Working Party of the Planning Subcommittee of the Advisory Committee on Advanced Television Service) are shown in Table 10.1.

A consumer can effect an improvement in the noise performance of his installation by impacting any contributing factor; examples include installation of a low-noise amplifier (LNA) or a better antenna.

10.2.1.3 Co-channel and adjacent-channel rejection

The assumptions made by PS/WP3 as reported in its Document 296 were based on threshold-of-visibility (TOV) measurements determined during the test of the competing systems. These TOV numbers were correlated to BER test results from the same tests. When testing the Grand Alliance VSB modem hardware at ATTC, and in field tests, only BER measurements were taken. In Table 10.2 the results of these tests are again expressed in equivalent TOV numbers derived from the BER measurements

Table 10.2 ATV Interference Criteria

Co-channel ATV-into-NTSC	33.8 dB
Co-channel NTSC-into-ATV	2.07 dB
Co-channel ATV-into-ATV	15.91 dB
Upper-adjacent ATV-into-NTSC	-16.17 dB
Upper-adjacent NTSC-into-ATV	-47.05 dB
Upper-adjacent ATV-into-ATV	-42.86 dB
Lower-adjacent ATV-into-NTSC	-17.95 dB
Lower-adjacent NTSC-into-ATV	-48.09 dB
Lower-adjacent ATV-into-ATV	-42.16 dB

It should be noted that the exact amount of adjacent-channel or co-channel interference entering the receiver terminals is a function of the exact overall antenna gain pattern used (not just F/B ratio) which is also a function of frequency.

10.2.1.4 Unintentional radiation

This subject is already well covered by the FCC Rules, Part 15.

10.2.1.5 Direct pickup (DPU)

Rules for direct pickup are included in the FCC Rules, Part 15. As advanced television will use digital transmission, rather than analog transmission, the image rejection numbers of IS-23 Section 3.28 of 50 dB or 60 dB may be far too stringent for the digital broadcasting case. A digital system is much more robust against image interference, and regulation of image rejection depth will not be necessary. Again the assumptions of IS-23 based on an IF of 45 MHz must be re-examined.

10.2.2 The modem field test

A preliminary version of the Grand Alliance transmission subsystem was provided for field testing in the summer of 1994. Terrestrial testing was performed using the 8 VSB mode and cable testing was performed using the 16 VSB high data rate mode. Results of

the tests are included in a September 16, 1994 ACATS report "*Field Test Results of the Grand Alliance Transmission Subsystem*," Document SS/WP2-1354.

The test results contain valuable information for receiver manufacturers about multipath interference and other impairment conditions, and their effect on the bit error rate of the digital signal.

10.2.3 Signal conversion and carrier recovery.

The following descriptions were taken from the Grand Alliance HDTV System Specification and are specific to the hardware implementation of the Grand Alliance. Figure 10.1 shows the receiver block diagram of the VSB terrestrial broadcast transmission system. Descriptions of each block follow.
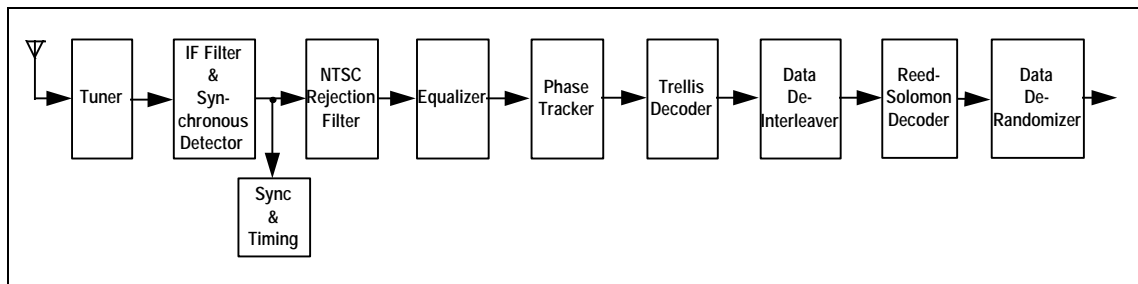
Figure 10.1. VSB receiver.

10.2.3.1 Tuner

The tuner, illustrated in Figure 10.2, as implemented in the prototype submitted for test, receives the 6 MHz signal (UHF or VHF) from the antenna. It is a high-side injection double-conversion type with a first IF frequency of 920 MHz. This puts the image frequencies above 1 GHz, making them easy to reject by a fixed front end filter. This selection of first IF frequency is high enough so that the input band-pass filter selectivity prevents the local oscillator (978-1723 MHz) from leaking out the tuner front end and interfering with other UHF channels, yet it is low enough for second harmonics of UHF channels (470-806 MHz) to fall above the first IF band-pass. Harmonics of cable channels could possibly occur in the first IF pass-band but are not a real problem because of the relatively flat spectrum (within 10 dB) and small signal levels (-28 dBm or less) used in cable systems.

The tuner input has a band-pass filter that limits the frequency range to 50-810 MHz, rejecting all other non-television signals that may fall within the tuner's image frequency range (beyond 920 MHz). In addition, a broadband tracking filter rejects other television signals, especially those much larger in signal power than the desired signal power. This tracking filter is not narrow, nor is it critically tuned, as is the case of present day NTSC tuners that must reject image signals only 90 MHz away from the desired channel. Minimal channel tilt, if any, exists due to this tracking filter.

Figure 10.2. Tuner block diagram.

A 10 dB gain, wideband RF amplifier increases the signal level into the first mixer, and is the dominant determining factor of receiver noise figure (7-9 dB over entire VHF, UHF, and cable bands). The first mixer, a highly linear double-balanced design to minimize even harmonic gen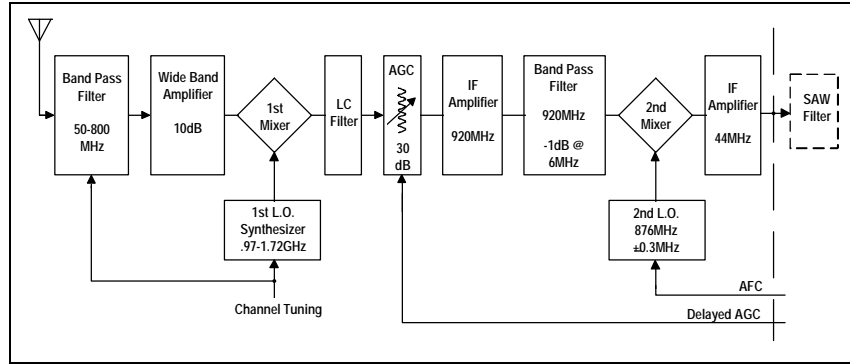eration, is driven by a synthesized low phase noise local oscillator (LO) above the first IF frequency (high-side injection). Both the channel tuning (first LO) and broadband tracking filtering (input band-pass filter) are controlled by microprocessor. The tuner is capable of tuning the entire VHF and UHF broadcast bands as well as all standard, IRC, and HRC cable bands.

The mixer is followed by an LC filter in tandem with a narrow 920 MHz band-pass ceramic resonator filter. The LC filter provides selectivity against the harmonic and sub-harmonic spurious responses of the ceramic resonators. The 920 MHz ceramic resonator band-pass filter has a -1 dB bandwidth of about 6 MHz. A 920 MHz IF amplifier is placed between the two filters. Delayed AGC of the first IF signal is applied immediately following the first LC filter. The 30 dB range AGC circuit protects the remaining active stages from large signal overload.

The second mixer is driven by the second LO, which is an 876 MHz voltage-controlled SAW oscillator. It is controlled by the frequency and phase-locked loop (FPLL) synchronous detector. The second mixer, whose output is the desired 44 MHz second IF frequency, drives a constant gain 44 MHz amplifier. The output of the tuner feeds the IF SAW filter and synchronous detection circuitry.

The tuner is made out of standard consumer electronic components, and is housed in a stamped metal enclosure.

10.2.3.2 Channel filtering and VSB carrier recovery

Carrier recovery is performed on the small pilot carrier by an FPLL circuit, illustrated in Figure 10.3. The first LO is synthesized by a PLL and controlled by a microprocessor. The third LO is a fixed reference oscillator. Any frequency drift or deviation from nominal has to be compensated in the second LO. Control for the second LO comes from the FPLL synchronous detector, which integrally contains both a frequency loop and a phase-locked loop in one circuit. The frequency loop provides a wide frequency pull-in range of ±100 kHz while the phase-locked loop has a narrow bandwidth (less than 2 kHz).
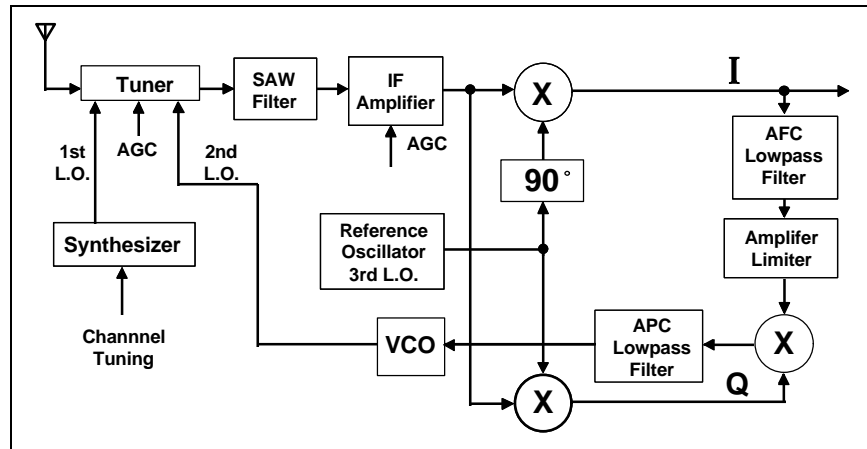
Figure 10.3. Tuner — IF — FPLL.

During frequency acquisition, the frequency loop uses both the in-phase (I) and quadrature-phase (Q) pilot signals. All other data processing circuits in the receiver use only the I channel signal. Prior to phase-lock, as is the condition after a channel change, the automatic frequency control (AFC) low-pass filter acts on the beat signal created by the frequency difference between the VCO and the incoming pilot. The high frequency data (as well as noise and interference) is mostly rejected by the AFC filter, leaving only the pilot beat frequency. After limiting this pilot beat signal to a constant amplitude (±1) square wave, and using it to multiply the quadrature signal, a traditional bipolar S-curve AFC characteristic is obtained. The polarity of the S-curve error signal depends upon whether the VCO frequency is above or below the incoming IF signal. Filtered and integrated by the automatic phase control (APC) low-pass filter, this DC signal adjusts the tuner's second LO to reduce the frequency difference.

When the frequency difference comes close to zero, the APC loop takes over and phase-locks the incoming IF signal to the third LO. This is a normal phase-locked loop circuit, with the exception that it is bi-phase stable. However, the correct phase-lock polarity is determined by forcing the polarity of the pilot to be equal to the known transmitted positive polarity. Once locked, the detected pilot signal is constant, the limiter output feeding the third multiplier is at a constant +1, and only the phase-locked loop is active (frequency loop automatically disabled). The APC low-pass filter is wide enough to reliably allow ±100 kHz frequency pull-in, yet narrow enough to consistently reject all strong white noise (including data) and NTSC co-channel interference signals. The PLL has a bandwidth that is narrow enough to reject most of the AM and PM generated by the data, yet is wide enough to track out any phase noise on the signal (and, hence, on the pilot) out to about 2 kHz. Tracking out low frequency phase noise (as well as low frequency FM components) allows the phase tracking loop, discussed in Section 10.2.3.8, to be more effective.

The prototype receiver can acquire a signal and maintain lock at a signal-to-noise ratio of 0 dB or less, and in the presence of heavy interference.

10.2.3.3 Segment sync and symbol clock recovery

The repetitive data segment syncs (Figure 10.4) are detected from among the synchronously detected random data by a narrow bandwidth filter. From the data segment syncs, a properly phased 10.76 MHz symbol clock is created along with a coherent AGC control signal. A block diagram of this circuit is shown in Figure 10.5.
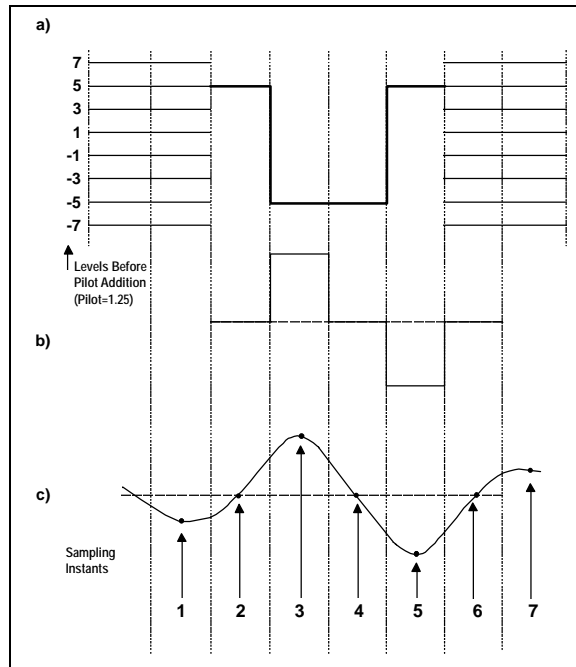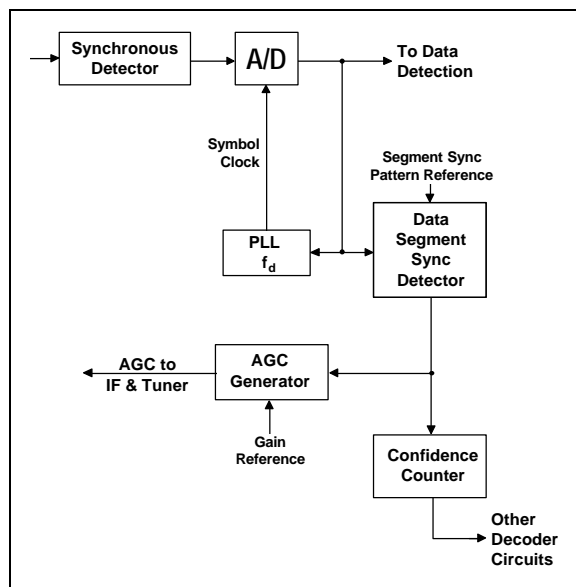

Figure 10.4. Data segment sync.

Figure 10.5. Segment sync & symbol clock recovery with AGC.

The 10.76 Msymbols/s (684 ÷ 286 x 4,500,000 Hz) I channel composite baseband data signal (syncs and data) from the synchronous detector is converted by an A/D converter for digital processing. Traditional analog data eyes can be viewed after synchronous detection. However, after conversion to a digital signal, the data eyes cannot be seen due to the sampling process. A PLL is used to derive a clean 10.76 MHz symbol clock for the receiver.

With the PLL free-running, the data segment sync detector containing a 4-symbol sync correlator looks for the two level syncs occurring at the specified repetition rate. The repetitive segment sync is detected while the random data is not, enabling the PLL to lock on the sampled sync from the A/D converter, and achieve data symbol clock synchronization. Upon reaching a predefined level of confidence (using a confidence counter) that the segment sync has been found, subsequent receiver loops are enabled.

Data segment sync detection and clock recovery both work reliably at signal-to-noise ratios of 0 dB or less, and in the presence of heavy interference.

10.2.3.4 Non-coherent and coherent AGC

Prior to carrier and clock synchronization, non-coherent automatic gain control (AGC) is performed whenever any signal (locked or unlocked signal, or noise/interference) overruns the A/D converter. The IF and RF gains are reduced accordingly, with the appropriate AGC "delay" applied.

When data segment syncs are detected, coherent AGC occurs using the measured segment sync amplitudes. The amplitude of the bipolar syncs, relative to the discrete levels of the random data, is determined in the transmitter. Once the syncs are detected in the receiver, they are compared to a reference value, with the difference (error) integrated. The integrator output then controls the IF and "delayed" RF gains, forcing them to whatever values provide the correct sync amplitudes.

10.2.3.5 Data field synchronization

Data Field Sync detection, shown in Figure 10.6, is achieved by comparing each received data segment from the A/D converter (after interference rejection filtering to minimize co-channel interference) with ideal field #1 and field #2 reference signals in the receiver. Over-sampling of the field sync is NOT necessary as a precision data segment and symbol clock has already been reliably created by the clock recovery circuit. Therefore, the field sync recovery circuit knows exactly where a valid field sync correlation should occur within each data segment, and only needs to perform a symbol by symbol difference. Upon reaching a predetermined level of confidence (using a confidence counter) that field syncs have been detected on given data segments, the Data Field Sync signal becomes available for use by subsequent circuits. The polarity of the middle of the three alternating 63 bit pseudo random (PN) sequences determines whether field 1 or field 2 is detected.
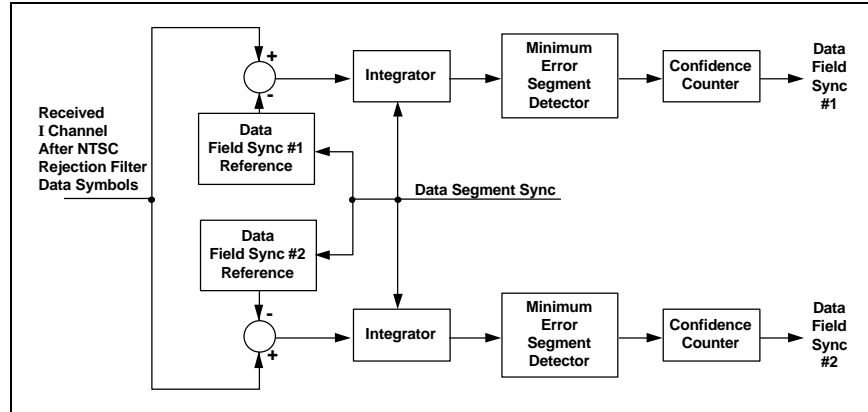
Figure 10.6. Data field sync recovery.

This procedure makes field sync detection robust, even in heavy noise, interference, or ghost conditions. Field sync recovery can reliably occur at signal-to-noise ratios of 0 dB or less, and in the presence of heavy interference.

10.2.3.6 Interference rejection filter

The interference rejection properties of the VSB transmission system are based on the frequency location of the principal components of the NTSC co-channel interfering signal within the 6 MHz television channel and the periodic nulls of a VSB receiver baseband comb filter.

Figure 10.7a shows the location and approximate magnitude of the three principal NTSC components: (1) the visual carrier (V) located 1.25 MHz from the lower band edge, (2) the chrominance subcarrier (C) located 3.58 MHz higher than the visual carrier frequency, and (3) the aural carrier (A) located 4.5 MHz higher than the visual carrier frequency.

The NTSC interference rejection filter (comb) is a one tap linear feed-forward filter, as shown in Figure 10.8. Figure 10.7b shows the frequency response of the comb filter, which provides periodic spectral nulls spaced $57 * f_H$ (10.762 MHz / 12, or 896.85 kHz) apart. There are 7 nulls within the 6 MHz channel. The NTSC visual carrier frequency falls close to the second null from the lower band edge. The 6th null from the lower band edge is correctly placed for the NTSC chrominance subcarrier, and the 7th null from the lower band edge is near the NTSC aural carrier.

Comparing Figure 10.7a and Figure 10.7b shows that the visual carrier falls 2.1 kHz below the second comb filter null, the chroma subcarrier falls near the 6th null, and the aural carrier falls 13.6 kHz above the 7th null. The NTSC aural carrier is at least 7 dB below its visual carrier.[13]

---

[13] Figure 10.7 shows a particular example of ATV offset frequency. The value is under study and may be adjusted based on adjacent-channel and/or co-channel effects.
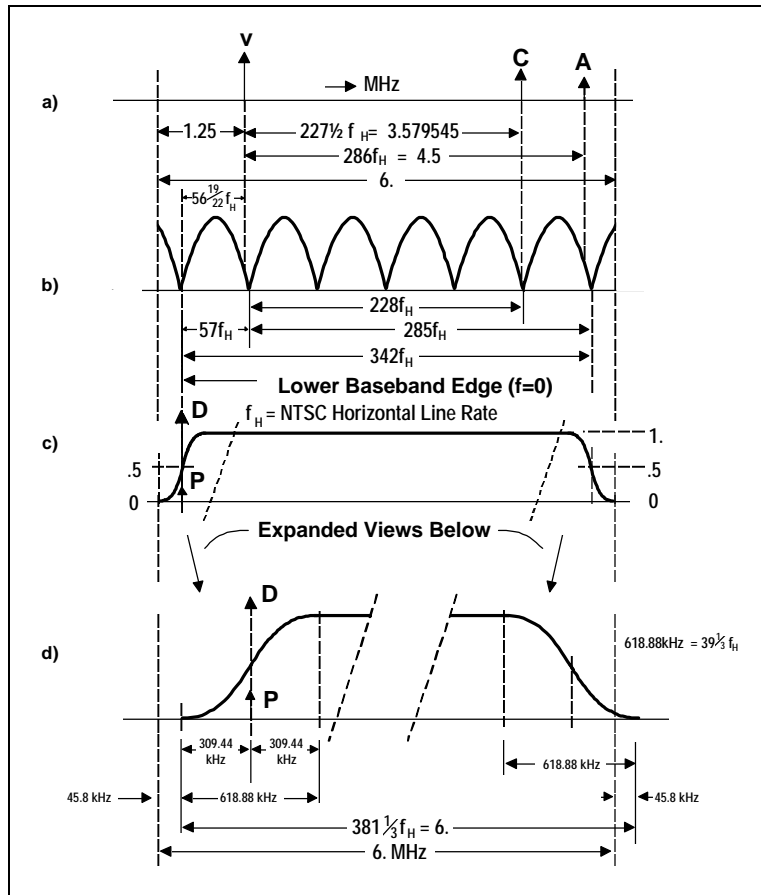
Figure 10.7. Location of NTSC carriers — comb filtering.

The comb filter, while providing rejection of steady-state signals located at the null frequencies, has a finite response time of 12 symbols (1.115 μs). Thus, if the NTSC interfering signal has a sudden step in carrier level (low to high or high to low), one cycle of the zero-beat frequency (offset) between the ATV and NTSC carrier frequencies will pass through the comb filter at an amplitude proportional to the NTSC step size as instantaneous interference. Examples of such steps of NTSC carrier are the leading and trailing edge of sync (40 IRE units). If the desired to undesired (D/U) signal power ratio is large enough, data slicing errors will occur. However, interleaving will spread the interference and will make it easier for the Reed-Solomon code to correct them (RS can correct up to 10 byte errors/segment).

Although the comb filter reduces the NTSC interference, the data is also modified. The 7 data eyes (8 levels) are converted to 14 data eyes (15 levels). This conversion is caused by the partial response process which is a special case of intersymbol interference that does not close the data eye but creates double the number of eyes of the same magnitude. The modified data signal can be properly decoded by the trellis decoder, and will be described in Section 10.2.3.9. Note that, because of time sampling, only the maximum data eye value is seen after A/D conversion.

The detail at the band edges for the overall channel is shown in Figure 10.7c and Figure 10.7d. Figure 10.7d shows that the frequency relationship of $56 \frac{19}{22} * f_H$ between

the NTSC visual carrier and the ATV carrier requires a shift in the ATV spectrum with respect to the nominal channel. The shift equals +45.8 kHz, or about +0.76%. This is slightly higher than currently applied channel offsets and reaches into the upper adjacent-channel at a level of about -40 dB. If that is another ATV channel, its spectrum is also shifted upward, therefore no spectral overlapping occurs. If it is an NTSC channel, the shift is below the (RF equivalent of the) Nyquist slope of an NTSC receiver where there is high attenuation, and it is slightly above its customary lower adjacent-channel sound trap. No adverse effects of the shift have been found, nor are they foreseen. An additional shift of the ATV spectrum is used in order to track the dominant NTSC interferer which may be assigned an offset of -10 kHz, 0 kHz or + 10 kHz.
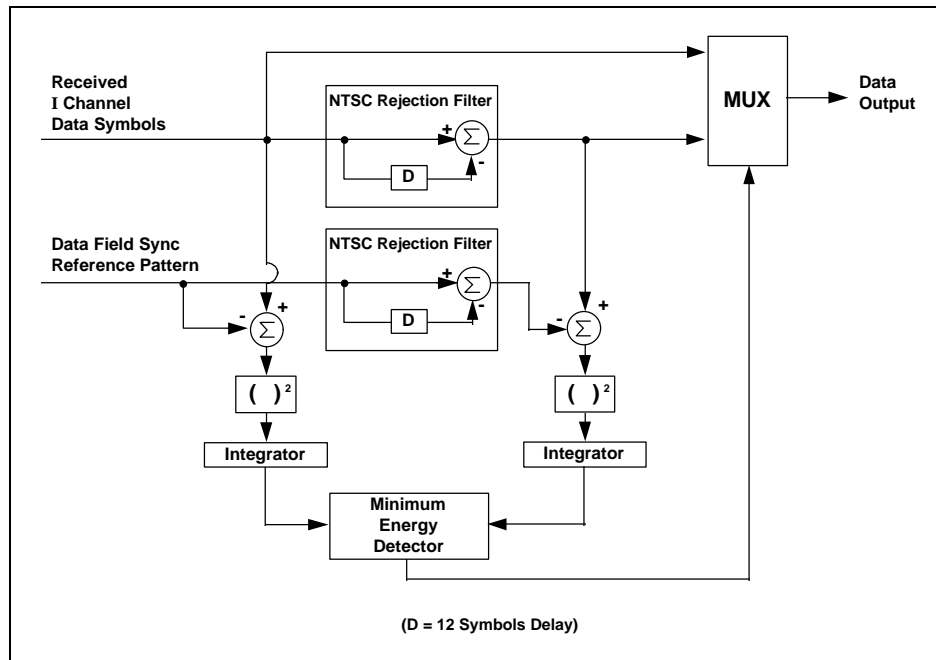
Figure 10.8. NTSC interference rejection filter.

NTSC interference can be detected by the circuit shown in Figure 10.8, where the signal-to-interference plus noise ratio of the binary Data Field Sync is measured at the input and output of the comb filter, and compared to each other. This is accomplished by creating two error signals. The first is created by comparing the received signal with a stored reference of the field sync. The second is created by comparing the rejection filter output with a combed version of the internally stored reference field sync. The errors are squared and integrated. After a predetermined level of confidence is achieved, the path with the largest signal-to-noise ratio (lowest interference energy) is switched in and out of the system automatically.

There is a reason to not leave the rejection comb filter switched in all the time. The comb filter, while providing needed co-channel interference benefits, degrades white noise performance by 3 dB. This is due to the fact that the filter output is the subtraction of two full gain paths, and as white noise is un-correlated from symbol to symbol, the noise power doubles. There is an additional 0.3 dB degradation due to the 12 symbol differential

coding. (See Section 10.2.3.9, Trellis decoder.) If little or no NTSC interference is present, the comb filter is automatically switched out of the data path. When the NTSC service is phased out, the comb filter can be omitted from digital television receivers.

10.2.3.7 Channel equalizer

The equalizer/ghost canceller compensates for linear channel distortions, such as tilt and ghosts. These distortions can come from the transmission channel or from imperfect components within the receiver.

The equalizer uses a Least-Mean-Square (LMS) algorithm and can adapt on the transmitted binary Training Sequence as well as on the random data. The LMS algorithm computes how to adjust the filter taps in order to reduce the error present at the output of the equalizer. It does this by generating an estimate of the error present in the output signal. This error signal is used to compute a cross-correlation with various delayed data signals. These correlations correspond to the adjustment that needs to be made for each tap to reduce the error at the output.

The equalizer algorithm can achieve equalization through three means: it can adapt on the binary training sequence; it can adapt on data symbols throughout the frame when the eyes are open; or, it can adapt on data when the eyes are closed (blind equalization). The principal difference among these three methods is how the error estimate is generated. (See Section 10.3, Receiver equalization issues.)

For adapting on the training sequence, the training signal presents a fixed data pattern in the data stream. Because the data pattern is known, the exact error is generated by subtracting the training sequence from the output.

The training sequence alone, however, may not be enough to track dynamic ghosts as these require tap adjustments more often than the training sequence is transmitted. Therefore, once equalization is achieved, the equalizer can switch to adapting on data symbols throughout the frame, and produce an accurate error estimate by slicing the data with an 8-level slicer and subtracting it from the output signal.

For fast dynamic ghosts (e.g., airplane flutter) it is necessary to use a blind equalization mode to aid in acquisition of the signal. Blind equalization models the multi-level signal as binary data signal plus noise, and the equalizer produces the error estimate by detecting the sign of the output signal and subtracting a (scaled) binary signal from the output to generate the error estimate.

To perform the LMS algorithm, the error estimate (produced using the training sequence, 8-level slicer, or the binary slicer) is multiplied by delayed copies of the signal. The delay depends upon which tap of the filter is being updated. This multiplication produces a cross-correction between the error signal and the data signal. The size of the correlation corresponds to the amplitude of the residual ghost present at the output of the equalizer and indicates how to adjust the tap to reduce the error at the output.

A block diagram of the equalizer is shown in Figure 10.9. The DC bias of the input signal is first removed by subtraction. The DC may be caused by circuit offsets, non-

linearities, or shifts in the pilot caused by ghosts. The DC offset is tracked by measuring the DC value of the training signal.
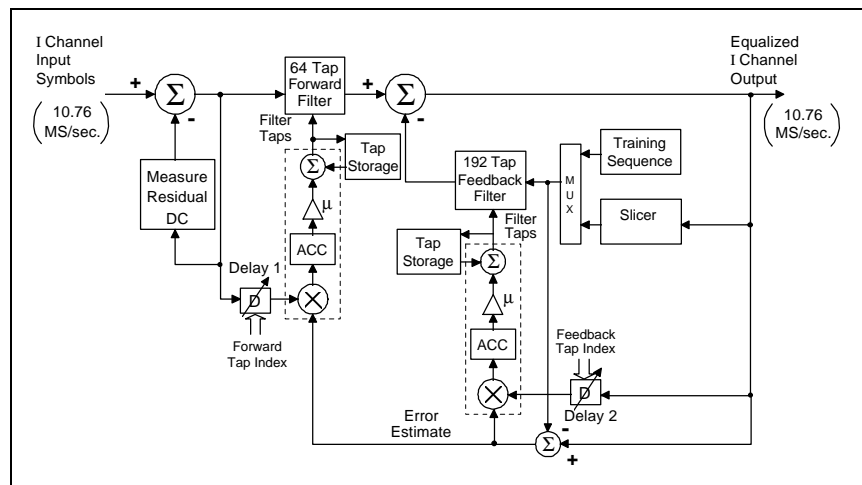
Figure 10.9. VSB receiver equalizer.

The equalizer filter consists of two parts, a 64 tap feed-forward transversal filter followed by a 192 tap decision feedback filter. The equalizer operates at the 10.762 MHz symbol rate (T-sampled equalizer).

The output of the forward filter and feedback filter are summed to produce the output. This output is sliced by either an 8-level slicer (15-level slicer when the comb filter is used) or a binary slicer depending upon whether the data eyes are open or not. (As pointed out in the previous Section on interference filtering, the comb filter does not close the data eyes but creates twice as many of the same magnitude). This sliced signal has the training signal and segment syncs reinserted as these are fixed patterns of the signal. The resultant signal is fed into the feedback filter, and subtracted from the output signal to produce the error estimate. The error estimate is correlated with the input signal (for the forward filter), or by the output signal (for the feedback filter). This correlation is scaled by a step size parameter, and used to adjust the value of the tap. The delay setting of the adjustable delays is controlled according to the index of the filter tap that is being adjusted.

10.2.3.8 Phase tracking loop

The phase tracking loop is an additional decision feedback loop which further tracks out phase noise which has not been removed by the IF PLL operating on the pilot. Thus, phase noise is tracked out by not just one loop, but two concatenated loops. Because the system is already frequency-locked to the pilot by the IF PLL (independent of the data), the phase tracking loop bandwidth is maximized for phase tracking by using a first order loop. Higher order loops, which are needed for frequency tracking, do not perform phase tracking as well as first order loops. Therefore, they are not used in the VSB system.

A block diagram of the phase tracking loop is shown in Figure 10.10. The output of the real equalizer operating on the I signal is first gain controlled by a multiplier and then fed into a filter which recreates an approximation of the Q signal. This is possible because of the VSB transmission method, where the I and Q components are related by a filter function which is almost a Hilbert transform. The complexity of this filter is minor because it is a finite impulse response (FIR) filter with fixed anti-symmetric coefficients and with every other coefficient equal to zero. In addition, many filter coefficients are related by powers of two, thus simplifying the hardware design.
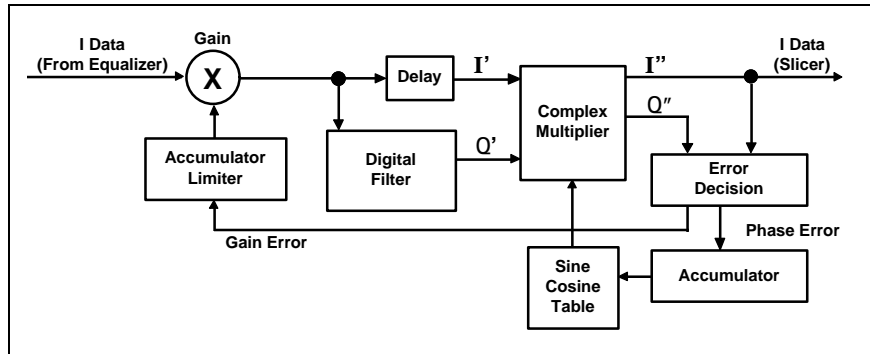
Figure 10.10. Phase tracking loop.

These I and Q signals are then fed into a de-rotator (complex multiplier), which is used to remove the phase noise. The amount of de-rotation is controlled by decision feedback of the data taken from the output of the de-rotator. As the phase tracker is operating on the 10.76 Msymbol/s data, the bandwidth of the phase tracking loop is fairly large, approximately 60 kHz. The gain multiplier is also controlled with decision feedback.

10.2.3.9 Trellis decoder

To help protect the trellis decoder against short burst interference, such as impulse noise or NTSC co-channel interference, 12 symbol code intrasegment interleaving is employed in the transmitter. As shown in Figure 10.11, the receiver uses 12 trellis decoders in parallel, where each trellis decoder sees every 12th symbol. This code interleaving has all the same burst noise benefits of a 12 symbol interleaver, but also minimizes the resulting code expansion (and hardware) when the NTSC rejection comb filter is active.

Before the 8 VSB signal can be processed by the trellis decoder it is necessary to suspend the Segment Sync. The Segment Sync is not trellis encoded at the transmitter. The circuit block diagram which illustrates the Segment Sync suspension is shown in Figure 10.12.

The trellis decoder performs the task of slicing and convolutional decoding. It has two modes; one when the NTSC rejection filter is used to minimize NTSC co-channel, and the other when it is not used. This is illustrated in Figure 10.13. The insertion of the NTSC rejection filter is determined automatically (before the equalizer), with this information passed to the trellis decoder. When there is little or no NTSC co-channel

interference, the NTSC rejection filter is not used, and an optimal trellis decoder is used to decode the 4-state trellis-encoded data. Serial bits are re-created in the same order in which they were created in the encoder.
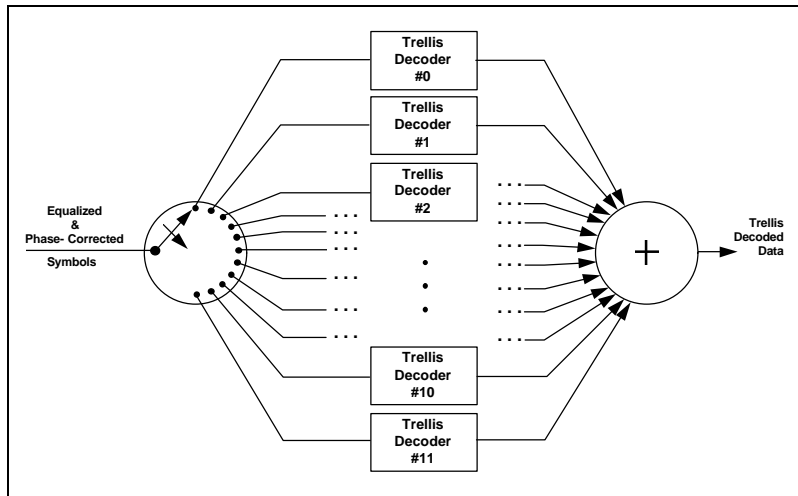
Figure 10.11. Trellis code de-interleaver.

In the presence of significant NTSC co-channel interference, when the NTSC rejection filter (12 symbol, feed-forward subtractive comb) is employed, a trellis decoder optimized for this partial response channel is used. This optimal code requires 8 states. This is necessary because the NTSC rejection filter, which has memory, represents another state machine seen at the input of the trellis decoder. In order to minimize the expansion of trellis states, two measures are taken: (1) special design of the trellis code, and (2) twelve-to-one interleaving of the trellis encoding. The interleaving, which corresponds exactly to the 12 symbol delay in the NTSC rejection filter, makes it so that each trellis decoder only sees a one-symbol delay NTSC rejection filter. By minimizing the delay stages seen by each trellis decoder, the expansion of states is also minimized. Only a 3.5 dB penalty in white noise performance is paid as the price for having good NTSC co-channel performance. The additional 0.5 dB beyond the 3 dB comb filter noise threshold degradation is due to the 12 symbol differential coding.
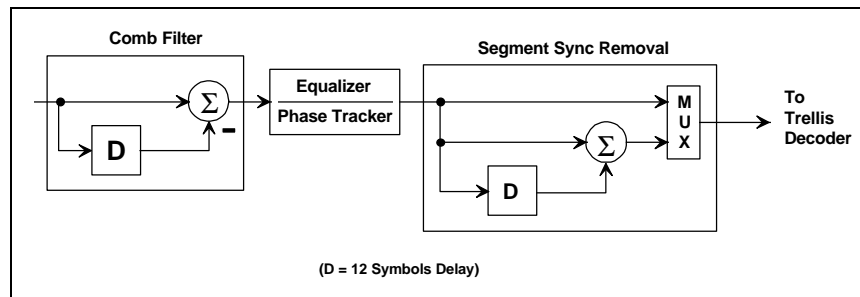
Figure 10.12. 8 VSB receiver segment sync suspension.
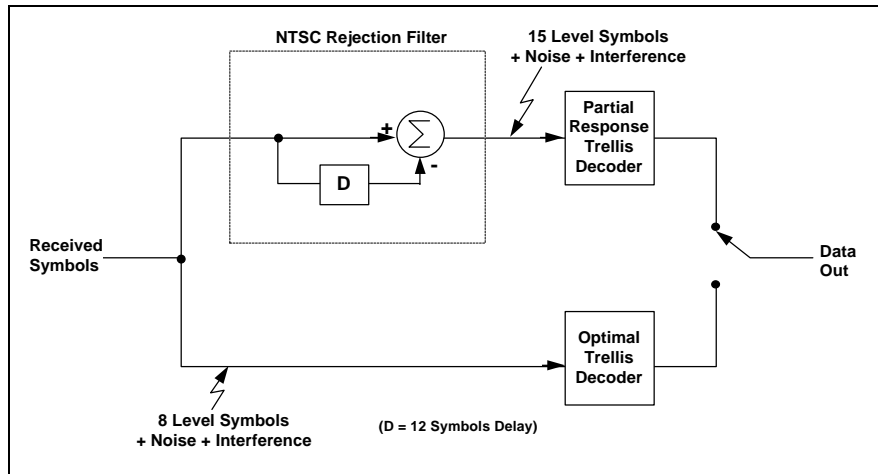
Figure 10.13. Trellis decoding with and without NTSC rejection filter.

The presence of the Segment Sync character in the data stream passed through the comb filter presents a complication which must be dealt with because Segment Sync is not trellis encoded or precoded. Figure 10.12 shows the technique that has been used. It shows the receiver processing that is performed when the comb filter is present in the receiver. The multiplexer in the Segment Sync removal block is normally in the upper position. This presents data that has been filtered by the comb to the trellis decoder. However, because of the presence of the sync character in the data stream, the multiplexer selects its lower input during the four symbols that occur twelve symbols after the segment sync. The effect of this sync removal is to present to the trellis decoder a signal that consists of only the subtraction of two adjacent data symbols that come from the same trellis encoder, one transmitted before, and one after the segment sync. The interference introduced by the segment sync symbol is removed in this process, and the overall channel response seen by the trellis decoder is the single-delay partial response filter.

The complexity of the trellis decoder is dependent upon the number of states in the decoder trellis. Since the trellis decoder operates on an 8-state decoder trellis when the comb filter is active, this defines the amount of processing that is required of the trellis decoder. The decoder must perform an Add-Compare-Select (ACS) operation for each state of the decoder. This means that the decoder is performing 8 ACS operations per symbol time. When the comb filter is not activated, the decoder operates on a 4-state trellis. The decoder hardware can be constructed such that the same hardware that is decoding the 8-state comb filter trellis can also decode the 4-state trellis when the comb filter is disengaged, thus there is no need for separate decoders for the two modes. The 8-state trellis decoder requires less than 5000 gates.

It should be noted that after the transition period when NTSC is no longer being transmitted, the NTSC rejection filter and the 8-state trellis decoder can be eliminated from digital television receivers.

10.2.3.10 Data de-interleaver

The convolutional de-interleaver performs the exact inverse function of the transmitter convolutional interleaver. Its 1/6 data field depth, and intersegment "dispersion" properties allow noise bursts lasting about 193 µs to be handled. Even strong NTSC co-channel signals passing through the NTSC rejection filter, and creating short bursts due to NTSC vertical edges, are reliably handled due to the interleaving and RS coding process. The de-interleaver uses Data Field Sync for synchronizing to the first data byte of the data field. The convolutional de-interleaver is shown in Figure 10.14.
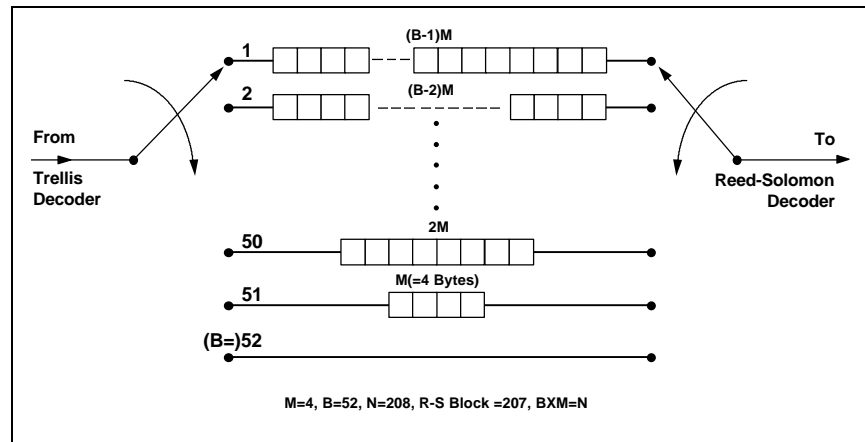
Figure 10.14. Convolutional de-interleaver.

10.2.3.11 Reed-Solomon decoder

The trellis-decoded byte data is sent to the (207,187) t=10 RS decoder, where it uses the 20 parity bytes to perform the byte-error correction on a segment-by-segment basis. Up to 10-byte errors per data segment are corrected by the RS decoder. Any burst errors created by impulse noise, NTSC co-channel interference, or trellis-decoding errors, are greatly reduced by the combination of the interleaving and RS error correction.

10.2.3.12 Data de-randomizer

The data is randomized at the transmitter by a Pseudo Random Sequence (PRS). The de-randomizer accepts the error-corrected data bytes from the RS decoder, and applies the same PRS randomizing code to the data. The PRS code is generated identically as in the transmitter, using the same PRS generator feedback and output taps. Since the PRS is locked to the reliably recovered Data Field Sync (and not some code word embedded within the potentially noisy data), it is exactly synchronized with the data, and performs reliably.

10.2.3.13 Receiver loop acquisition sequencing

The receiver incorporates a "universal reset" which initiates a number of "confidence counters" and "confidence flags" involved in the lock-up process. A universal reset occurs, for example, when tuning to another station or turning on the receiver. The

various loops within the VSB receiver acquire and lock-up sequentially, with "earlier" loops being independent from "later" loops. The order of loop acquisition is as follows:

- Tuner 1st LO synthesizer acquisition

- Non-coherent AGC reduces unlocked signal to within A/D range

- Carrier acquisition (FPLL)

- Data segment sync and clock acquisition

- Coherent AGC of signal (IF and RF gains properly set)

- Data field sync acquisition

- NTSC rejection filter insertion decision made

- Equalizer completes tap adjustment algorithm

- Trellis and RS data decoding begin

Most of the loops mentioned above have confidence counters associated with them to insure proper operation. However, the build-up or let-down of confidence is not designed to be equal. The confidence counters build confidence quickly for quick acquisition times, but lose confidence slowly to maintain operation in noisy environments. The VSB receiver carrier, sync and clock circuits will work in SNR conditions of 0 dB or less as well as in severe interference situations.

10.2.3.14 High data rate mode

The VSB digital transmission system provides the basis for a family of ATV receivers suitable for receiving data transmissions from a variety of media. This family shares the same pilot, symbol rate, data frame structure, interleaving, Reed-Solomon coding, and synchronization pulses. The VSB system offers two modes: a simulcast terrestrial broadcast mode, and a high data rate mode.

Most parts of the high data rate mode VSB system are identical or similar to the terrestrial system. A pilot, data segment sync, and data field sync are all used to provide robust operation. The pilot in the high data rate mode also adds 0.3 dB to the data power. The symbol, segment, and field signals and rates are all the same, allowing either receiver to lock up on the other's transmitted signal. Also, the data frame definitions are identical. The primary difference is the number of transmitted levels (8 versus 16) and the use of trellis coding and NTSC interference rejection filtering in the terrestrial system.

The high data rate mode receiver is identical to the VSB terrestrial receiver, except that the trellis decoder is replaced by a slicer, which translates the multi-level symbols into data. Instead of an 8-level slicer, a 16-level slicer is used. Also note that no NTSC interference rejection filter is required.

10.3 Receiver equalization issues

10.3.1 Introduction

The VSB signal contains features which allow design of receivers that reliably perform the functions of acquiring and locking to the transmitted signal. The equalization of the signal for channel frequency response and ghosts is facilitated by the inclusion of specific features in the Data Field Sync (see Figure 10.15). Utilization of these features is made more reliable by the provision of means to first acquire and synchronize to the VSB signal, particularly by the Segment Sync. The Data Field Sync then can be used both to identify itself and to further perform equalization of linear transmission distortions.

The VSB signal may also be equalized by data-based or blind equalization methods which do not use the Data Field Sync.
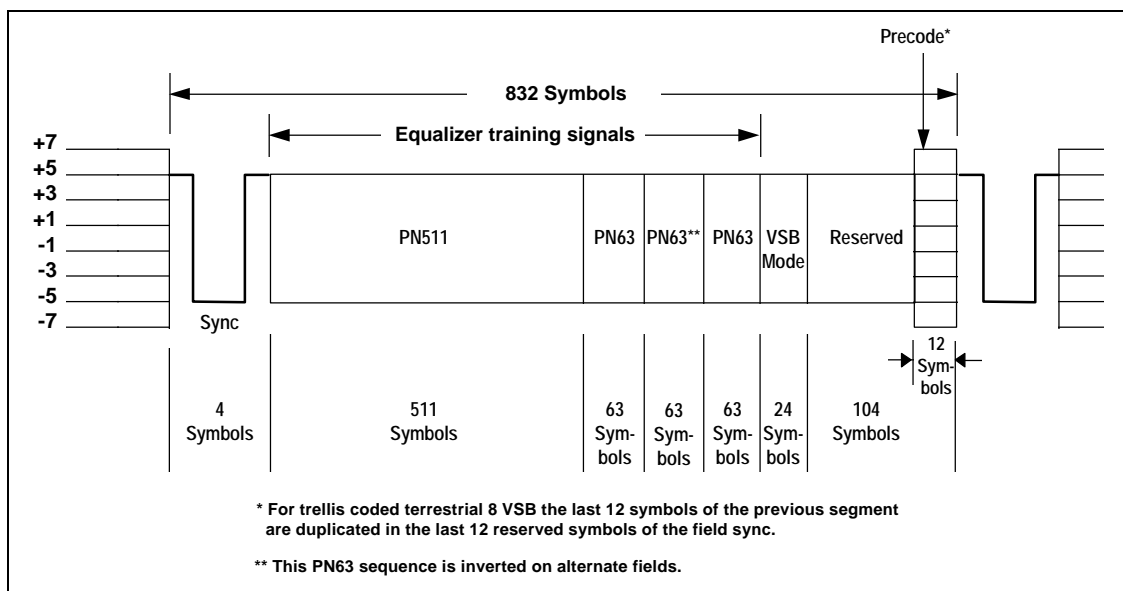
Figure 10.15. Data field sync.

10.3.2 The equalizer training signal

10.3.2.1 Specifications for the data field sync

10.3.2.2 General

Data Field Sync is a unique type of Data Segment in the VSB signal. All payload data in the VSB signal is contained in Data Segments, which are processed with data interleaving, Reed-Solomon error coding, and trellis coding. The Data Field Sync (and the Segment Sync portion of every Data Segment), however, is not processed this way, as its purpose is to provide direct measurement and compensation for channel linear distortion. Equalizer training signals consisting of pseudo-noise (PN) sequences are major parts of the Data Field Sync.

10.3.3 Receiver equalization using the training signal

10.3.3.1 PN511 equalizer training sequence

This is a sequence of 511 two-level symbols (symbol rate is approximately 10.76 MHz, symbol period approximately 93 ns). The PN 511 auto-correlation function consists of one central peak surrounded by small values 30 dB or more lower than the central peak. This residual correlation noise limits the accuracy with which cancellation can be achieved to approximately 27 dB below the main signal. This is more than sufficient for the 8 VSB signal. Because the PN511 signal is not surrounded by a blank area, it will also exhibit random correlation to adjacent parts of the transmitted signal, also at a level of approximately -27 dB compared to the peak correlation. The random correlation noise can be further reduced by various means which effectively average the results of several training sequences.

10.3.3.2 PN63 equalizer training sequences

This portion of the signal consists of three identical sequential sequences of 63 two-level symbols. Because of the repetition, the correlation of one PN63 sequence with the three in the signal shows three equal peaks at 63-symbol spacing. The region of the correlation between the central peak and the outside peaks is a constant equal to -1/63 of the peak value. This allows ghosts within a delay range of 63 symbols total to be compensated exactly with one training signal measurement, as they are not affected by inherent correlation noise or intrusion of echoes from other parts of the Data Field Sync. The receiver designer may divide the total 63-symbol capability arbitrarily between pre-ghost and post-ghost correction regions. The polarity of the middle PN63 sequence is inverted on alternate fields. This removes any correlation to the PN511 sequence, if the result is averaged over an even number of fields.

10.3.3.3 Segment sync

This is the same two-level Segment Sync used on all data segments, either payload or Data Field Sync. It is useful in equalization because it independently establishes the timing of the PN sequences and the symbol clock phase. Because of the resulting known segment timing, it is only necessary to identify which segment contains Data Field Sync. A full correlator is not required because the position of the sequence within the segment has already been established. A simple circuit which compares each segment's data to the known PN511 sequence can accomplish this identification.

10.3.3.4 VSB mode

24 two-level symbols identify the type of data in the following data segments.

10.3.3.5 Precode

This section of twelve symbols repeats the last twelve symbols of payload data from the preceding Data Segment to allow the 12-symbol NTSC-rejection comb filter to

operate properly on payload data in conjunction with trellis coding, while allowing suspension of trellis coding during the other parts of Data Field Sync.

10.3.4 Theoretical equalizer performance using training signals

10.3.4.1 Amplitude and delay of ghosts

Theoretically, in a noise-free signal, ghosts of amplitude up to 0 dB with respect to the largest signal, and within a total window of 63 symbols, can be exactly canceled in one pass using the information in the sequence. Ghosts of any delay can be canceled using the information in the sequence, with a single-pass accuracy of approximately -27 dB, which improves by averaging over multiple passes. Operation of a complete receiver system with 0 dB ghosts may not be achievable due to failure of carrier acquisition, although operation of the equalizer itself at 0 dB may be demonstrable under test by supplying an external carrier.

Practical implementations of receivers will differ in performance depending on the particular techniques used (see discussion on receiver implementation in Sections 10.3.5 and 10.3.6). The complete prototype receiver demonstrated in field and laboratory tests showed cancellation of -1 dB amplitude ghosts under otherwise low-noise conditions, and -3 dB amplitude ghost ensemble with noise. In the latter case, the signal was 2.25 dB above the receivers noise-only reception threshold.

10.3.4.2 Multiple ghosts

The number of ghosts to be canceled has in itself little effect on the theoretical performance of the system. Theoretical limits to cancellation depend on the amount of noise gain that occurs in compensating the frequency response due to the particular ghosted signal. This is affected non-linearly by the coring effect of signal quantization and decision feedback (see discussion of decision feedback in Section 10.3.6.3).

10.3.4.3 Speed of operation

Because the equalizer training signals recur with a period of approximately 24 ms, the receiver cannot perform equalization updates at a faster rate. The signal thus provides information such that the equalization system can theoretically have a bandwidth of 20 Hz. Additional constraints are implied by the desire to average out the correlation between the PN511 sequence and the alternating-polarity middle PN63 sequence.

Speed of convergence is not the only important criterion of performance, however. Ultimate accuracy and response to noise are also of importance. Equalization techniques generally proceed by successive approximation to the desired state, and therefore exhibit a convergence time measured as a number of frame periods. By using the information in the signal in different ways, speed of operation can be traded off for accuracy and noise immunity. The success of such a trade-off may depend on non-linear techniques; for example, switching between a quick acquisition mode and slower refinement mode, or using varying step sizes in a steepest-descent technique. Techniques for affecting the rate of convergence and its final accuracy are mentioned in the literature.

10.3.5 Receiver implementation using blind equalization

This section describes blind equalization techniques, which are not based on a training signal reference. They may be appropriate for use with the VSB transmission system. Blind equalization techniques are particularly useful when the channel impairments vary more rapidly than the transmission of the training wave form.

As in many modern digital data communication systems, an adaptive equalizer is used in the Grand Alliance system to compensate for changing conditions on the broadcasting channel. In communication systems which use an adaptive equalizer, it is necessary to have a method of adapting the equalizer's filter response to adequately compensate for channel distortions. There are several algorithms available for adapting the filter coefficients. The most widely used is the Least Mean Squares (LMS) method.[14]

When the equalizer is first started, the tap weights are usually not set to adequately compensate for the channel distortions. In order to force initial convergence of the equalizer coefficients, a known training signal (i.e., both the transmitter and receiver know the signal) is used as the reference signal. The error signal is formed by subtracting a locally generated copy of the training signal from the output of the adaptive equalizer. When using the training signal, the eye diagram is typically closed. The training signal serves to open the eye. After adaptation with the training signal, the eye has opened, and the equalizer may be switched to a decision directed mode of operation. The decision directed mode uses the symbol values at the output of the decision device instead of the training signal.

A problem arises in the above scenario when a training signal is not available. In this case, a method, typically called blind equalization, of acquiring initial convergence of the equalizer taps and forcing the eye open is necessary without a training signal. Blind equalization has been extensively studied for quadrature amplitude modulated (QAM) systems. For QAM systems, there are several methods typically employed: the constant modulus algorithm (CMA), and the reduced constellation algorithm (RCA) are among the most popular.[15] For VSB systems both of these methods are not directly applicable. CMA relies on the fact that, at the decision instants, the modulus of the detected data symbols should lie on one of several circles of varying diameters. Thus, it inherently relies on the underlying signal to be two-dimensional. Since VSB is essentially a one-dimensional signal (at least for the data carrying portion), CMA is not directly applicable. RCA relies on forming "super constellations" within the main constellation. The data signal is first forced to fit into a "super constellation," and then the "super constellations" are subdivided to include the entire constellation. Again, as typically used, RCA implies a two-dimensional constellation.

However, blind equalization can be performed on the VSB constellation using a modified reduced constellation algorithm (MRCA). The key part of this modification is to realize that there exists a one-dimensional version of the RCA algorithm which is

---

[14] See [1] in the bibliography of Section 10.3.6.6.

[15] See [2] and [3] in the bibliography of Section 10.3.6.6.

appropriate for VSB signals. The MRCA consists of an algorithm to determine appropriate decision regions for a VSB decision device, so as to generate decisions which allow an adaptive equalizer to converge without the use of a training signal.

In VSB systems, the decision regions typically span one data symbol of the full constellation and the upper and lower bound of each decision region are set midway between the constellation points. If these decision regions are used for initial convergence of the equalizer, the equalizer will not converge, since, due to the presence of intersymbol interference, a significant amount of the decisions from the decision device will be incorrect .

In order to force more correct decisions to be made, an algorithm for determining new upper and lower decision region boundaries has been determined. The algorithm clusters the full VSB constellation into several sets, determines upper and lower bounds for decision regions, and appropriate decision device output "symbol" values. These first sets are further divided into smaller sets until each set of symbols contains exactly one symbol, and the decision regions correspond to the standard decision regions for VSB described above. The function of each stage is to allow for more decisions to be correct, and thereby drive the equalizer towards convergence. In this way, each stage in the blind equalization process serves to further open the eye.

In general, the MRCA algorithm consists of clustering the decision regions of the decision device into finer portions of the VSB constellation. The method starts with a binary (two-level) slicer, then switches to a four-level slicer, then eight-level slicer, etc. It should be noted that the MRCA algorithm is applicable to both linear equalization and decision feedback equalization.

10.3.6 Equalizer hardware implementation

10.3.6.1 General

Various system configurations need to be considered and evaluated when designing an adaptive equalizer. Some pertinent items are touched on here; references are given to further information which has appeared in the literature. A particular practical implementation will combine choices from various techniques available for achieving the desired performance, and will be based on current technology and economics as well as theoretical considerations. Choices between the alternatives presented here are not required by the Digital Television Standard, although the particular combination of techniques used in the Grand Alliance prototype system is described for reference.

10.3.6.2 Data and filter coefficient resolution

A consideration which needs brief mention is the resolution of the quantized data and the resolution of the filter coefficients. Obviously, more resolution means better performance, but at increased cost. Generally speaking, it has been found desirable to have data resolution in the range of 8 bits or more, and coefficient resolution in the range of 9 bits or more.

10.3.6.3 Equalizer filter implementation

In the design of equalizers a trade-off is usually made between Finite Impulse Response (FIR) filters vs. Infinite Impulse Response (IIR) filters.

In equalizers, FIR filters have the attractive properties of always being stable and being able to cancel both pre and post ghosts. Unfortunately, in the process of canceling a ghost they generate a secondary ghost, with twice the delay of the original, but with an amplitude that is the square of the original. This secondary ghost is then canceled in a similar manner, with a tertiary ghost being generated. This process continues until the nth ghost generated is so small as to be lost in the noise of the signal, or is lost in the quantization process of the digital filtering, provided the FIR filter is long enough. To cancel long delay ghosts, FIR filters therefore need a great number of coefficients. Also, since the cancellation process involves delaying, scaling, and adding copies of the original signal back to itself, noise enhancement generally results.

IIR filters cancel ghosts without other ghosts being generated, and so they cancel long ghosts with fewer coefficients. With a slicer placed in the feedback path (decision feedback), white noise is removed from the feedback signal (assuming the slicer doesn't make any mistakes), and so the feedback signal does not increase the overall noise level of the output.

IIR filters cannot cancel pre ghosts, and, in actual applications, cannot cancel short post delay ghosts due to latencies inherent in the filtering hardware. Also, care must be taken to insure stability. A useful approach is to use an equalizer that has both an FIR and an IIR section. Here, small clumps of coefficients are positioned, with the help of programmable bulk delays, where they will be most effective in canceling the ghosts. Very long delay ghosts can then be canceled with relatively few coefficients.

10.3.6.4 Real-only vs. complex equalizer structures

In QAM systems, 2 carriers (I and Q), offset 90 degrees from one another, are modulated with independent data streams. Two equalization paths need to be employed in order to correctly equalize this signal. In the 8/16 VSB system, only one carrier (I) is modulated with unique data. The resulting Q carrier is dependent upon the I carrier, and is related to it by a mathematical function that is very similar to a Hilbert transform. (Had the 8/16 VSB system been single sideband (SSB), the Q channel would be identically equal to the Hilbert transform of the I channel. The 8/16 VSB system is not SSB, but it is very close.) Therefore, only one equalization path is needed in the 8/16 VSB system.

10.3.6.5 Symbol-rate vs. fractional-symbol sampling; use of segment sync

Many digital transmission systems do not use any kind of explicit synchronization information for clock recovery, and so must operate on the transmitted data itself in order to regenerate clocks. Any phase error in the sampling clock will result in degraded performance at the equalizer output. One means to overcome this is to over-sample the data (fractionally spaced sampling). A common over-sampling rate is 2 times (designated T/2), so that 2 samples are taken for every transmitted symbol. T/2 sampling reduces the effect of sampling phase error in the recovered data, but it doubles the number of filter

coefficients required for a given ghost coverage range. The 8/16 VSB system periodically transmits a fixed pattern of 4 symbols (Segment Sync), and these can be used to set the phase of the sampling clock. The phase error of this clock is then extremely low, making a symbol spaced (T) equalizer very effective for equalization.

10.3.6.6 Bibliography on equalization

[1] Qureshi, Shahid U.H., "Adaptive Equalization," *Proceedings of the IEEE*, Vol. 73, No. 9, September 1985, pp. 1349-1387.

[2] Ciciora, Walter, et. al., "A Tutorial on Ghost Canceling in Television Systems," *IEEE Transactions on Consumer Electronics*, Vol. CE-25, No. 1, February 1979, pp. 9-44.

[3] Ungerboeck, Gottfried, "Fractional Tap-Spacing Equalizer and Consequences for Clock recovery in Data Modems," *IEEE Transactions on Communications*, Vol. COM-24, No. 8, August 1976, pp. 856-864.

[4] Keeler, R. J., "Construction and Evaluation of a Decision Feedback Equalizer", *Rec. IEEE Int. Conference on. Commun.*, Philadelphia, PA, June 19-21, 1972.

[5] Widrow, B., and Stearns, S., *Adaptive Signal Processing*, Prentice Hall, 1985.

[6] Haykin, S., *Adaptive Filter Theory*, Prentice Hall, 1991.

[7] Proakis, John G., *Digital Communications*, second edition, McGraw-Hill, New York, 1989.

[8] Godard, D. N. "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Transactions on Communications*, Vol. COM-28, No. 11, pp. 1867-1875, November 1980.

[9] Godard, D.N., and Thirion, P.E. "Method and device for training an adaptive equalizer by means of an unknown data signal in a QAM transmission system," U.S. Patent 4 227 152, October 7, 1980.

10.4 Receiver video issues

10.4.1 Video formats

Transmissions conforming to the Digital Television Standard are expected to include the video formats as described in Table 10.3. Receivers will have to extract the picture rates and video format information, and will have to perform the necessary interpolation and format conversion so that these video formats can be displayed in the 'native' display format of the receiver.

In Table 10.3, "vertical lines" refers to the number of active lines in the picture. "Pixels" refers to the number of pixels during the active line. "Aspect ratio" refers to the picture aspect ratio. "Picture rate" refers to the number of frames or fields per second. In the values for picture rate, "P" refers to progressive scanning, "I" refers to interlaced scanning. Note that both 60.00 Hz and 59.94 (60x1000/1001) Hz picture rates are allowed. Dual rates are allowed also at the picture rates of 30 Hz and 24 Hz.

Table 10.3 Digital Television Standard Video Formats

Vertical lines	Pixels	Aspect ratio	Picture rate
1080	1920	16:9	60I, 30P, 24P
720	1280	16:9	60P, 30P, 24P
480	704	16:9 and 4:3	60P, 60I, 30P, 24P
480	640	4:3	60P, 60I, 30P, 24P

Two native display formats were implemented in the Grand Alliance prototype. Receivers were implemented using 787.5 scan lines per vertical scan in progressive mode and using 562.5 lines per vertical scan in interlaced mode (1125 lines per frame). The 480-line formats were not implemented in the prototype.

10.4.2 Multiple video programs

In the case of multi-video transmissions, the receiver designers should consider that the ATV display may have to be capable of also displaying video in the 525-line scanning format. It will be necessary to identify and extract the audio stream which corresponds to the user selected video stream. The assignment of packet identification (PID) values to individual audio, video, and data streams allows flexibility in the creation of a transport multiplex containing one or more audio-visual programs. An ATV receiver can make use of the program_map_table information carried in the transport multiplex to identify the PID values of the audio, video, and auxiliary data elementary streams for a desired program. Alternatively, the program paradigm, as described in ATSC Standard A/53, Annex C, Section 5.3, can be used to identify the audio, video, and data PIDs for programs numbered 1 through 255.

10.4.3 Concatenation of video sequences

The video coding specified in the Digital Television Standard is based on the ISO/IEC Standard 13818-2 (MPEG-2 Video). MPEG-2 Video specifies a number of video related parameters in the sequence header such as profile and level, VBV size, maximum bit rate, field/frame rate information, all progressive scan indicator, horizontal and vertical resolution, picture structure, picture aspect ratio, color field identification, chroma format, colorimetry, pan & scan etc., all of which will be necessary for receivers.

MPEG-2 Video specifies the behavior of a compliant video decoder when processing a single video sequence. A coded video sequence commences with a sequence_start_code symbol, contains one or more coded pictures and is terminated by a sequence_end_code symbol. Parameters specified in the sequence header are required to remain constant throughout the duration of the sequence. Specification of the decoding behavior in this case is feasible because the MPEG-2 Video standard places constraints on the construction and coding of individual sequences. These constraints prohibit channel buffer overflow/underflow as well as coding the same field parity for two consecutive fields.

It is envisioned that it will be common for coded bit streams to be spliced for editing, insertion of commercial advertisements, and other purposes during the video

production and distribution chain. If one or more of the sequence level parameters differ between the two bit streams to be spliced, then a `sequence_end_code` symbol must be inserted to terminate the first bit stream, and a new sequence header must exist at the start of the second bit stream. Thus the situation of concatenated video sequences arises.

While the MPEG-2 Video standard specifies the behavior of video decoders when processing a single sequence, it does not place any requirements on the handling of concatenated sequences. Since the MPEG-2 Video standard does not specify the behavior of decoders in this case, channel buffer overflow/underflow could occur at the junction between two coded sequences unless well-constrained concatenated sequences are produced.

While it is recommended, the Digital Television Standard does not require the production of well-constrained concatenated sequences as described in Section 5.13. If well-constrained concatenated sequences are produced according to these recommendations, then it is recommended that receivers provide a seamless presentation across such concatenated sequences. Seamless presentation occurs when each coded picture is correctly decoded and displayed for the proper duration.

10.4.4 D-frames

The MPEG family of video coding standards (ISO 11172-2 and ISO 13818-2) includes a provision for efficiently coding reduced resolution pictures in "D-frames" by using intraframe DCT DC coefficients. The use of D-frames was envisioned as a means of storing highly compressed intraframe coded pictures for allowing crude fast scan of compressed video stored on digital storage media. The Digital Television Standard does not include syntax for the transmission of D-frame coded pictures; however, receivers may support the decoding of D-frames for all picture formats to allow for the use of this efficient coding mode by VCRs, digital videodisc players, or other digital storage media.

10.4.5 Adaptive video error concealment strategy

10.4.5.1 Error concealment requirements

In MPEG video compression, video frames to be coded are formatted into sequences containing intra-coded (I), predictive-coded (P) and bidirectionally predictive-coded (B) frames. This structure of MPEG implies that if an error occurs within I-frame data, it will propagate for a number of frames. Similarly, an error in a P-frame will affect the related P and B-frames, while B-frame errors will be isolated. Therefore, it is desirable to develop error concealment techniques to prevent error propagation and, consequently, to improve the quality of reconstructed pictures.

There are two approaches which have been used for I-frame error concealment, temporal replacement and spatial interpolation. Temporal replacement can provide high resolution image data as the substitute to the lost data; but in motion areas a significant difference might exist between the current intra-coded frame and the previously decoded frame. In this case, temporal replacement will produce large distortion unless some motion-based processing can be applied at the decoder. However, this type of processing

is not always available since it is a computationally demanding task. In contrast, a spatial interpolation approach synthesizes the lost data from the adjacent blocks in the same frame. In spatial interpolation the intra-frame redundancy between blocks is exploited, while a potential problem of blurring remains due to insufficient high order DCT coefficients for active areas.

10.4.5.2 Error concealment implementation

To address this problem, an adaptive error concealment technique has been developed. In this scheme, temporal replacement or spatial interpolation should be used based on easily obtained measures of image activity from the neighboring macroblocks, i.e., the local motion and the local spatial detail. If local motion is smaller than spatial detail, the corrupted blocks belong to the class on which temporal replacement is applied; when local motion is greater than local spatial detail, the corrupted blocks belong to the class which will be concealed by spatial interpolation. The overall concealment procedure consists of two stages. First, temporal replacement is applied to all corrupted blocks of that class through the whole frame. After the temporal replacement stage, the remaining unconcealed damaged blocks are more likely to be surrounded by valid image blocks. A stage of spatial interpolation is then performed on them. This will now result in less blurring, or the blurring will be limited to smaller areas. Therefore, a good compromise between distortion and blurring can be obtained. This algorithm uses some simple measures, obtainable at the decoder, to adapt between spatial and temporal concealment modes. It is noted that the same idea can be used for intra-coded macroblocks of P and B-frames. The only modification is that the motion-compensation should be applied to the temporal replacement and the motion vectors (if lost) are assumed from ones in the top and bottom macroblocks.

Several new methods have been developed to further improve the accuracy of concealment. The first is a spatial concealment algorithm using directional interpolation. This algorithm utilizes spatially correlated edge information from a large local neighborhood of surrounding pixels and performs directional or multi-directional interpolation to restore the missing block.

The second method is I-picture vectors. Motion information is very useful in concealing losses in P and B pictures, but is not available for I-pictures. If motion vectors are made available for all MPEG pictures (including I-pictures) as an aid for error concealment, good error concealment performance can be obtained without the complexity of adaptive spatial processing. Therefore, a syntax extension has been adopted where motion vectors can be transmitted in an I-picture as the redundancy for error concealment purposes.

The third algorithm is the enhancement version of the adaptive spatio-temporal algorithm. The basic idea of this algorithm is to use a weighted average of spatial and temporal information rather than exclusively using either spatial or temporal information alone to conceal missing blocks. The temporal replacement estimate is further enhanced by the use of sub-macroblock refined motion vectors. Rather than applying a single estimated motion vector on an entire macroblock to create a temporal replacement (which can often

result in a blocky shearing effect), every small sub-macroblock pixel region (e.g. 2x2 or 4x4 pixel regions) that composes the entire macroblock undergoes temporal replacement with its own estimated motion vector. The motion vectors associated with each sub-macroblock region are obtained from a smooth interpolation of the motion vector field, resulting in a temporal replacement estimate that is continuous at macroblock boundaries and fits well with its neighboring macroblocks.

10.4.5.3 Bibliography on error concealment

[1] H. Sun and J. Zdepski, "Adaptive Error Concealment Algorithm for MPEG Compressed Video," *Proceedings of SPIE, Visual Communication and Image Signal Processing*, Vol. 1818, November 1992, pp. 814-824.

[2] H. Sun, J. Zdepski, W. Kwok and D. Raychaudhuri, "Error Concealment Algorithms for Robust Decoding of MPEG Compressed Video," Submitted to *IEEE Trans. on Circuits and Systems for Video Technology*.

[3] W. Kwok and H. Sun, "Multi-directional Interpolation for Spatial Error Concealment," *IEEE Trans. on Consumer Electronics*, Vol. 39, Number 3, August 1993, pp. 455-460.

[4] H. Sun and W. Kwok, "Concealment of Damaged Block Transform Coded Images Using Projections onto Convex Sets," will appear in the April 1995 issue of *IEEE Trans. on Image Processing*.

10.5 Receiver audio issues

This Section summarizes receiver implementation issues related to audio. Further information on the audio system may be found in Chapter 6 of this Guide which contains information of interest to both broadcasters and receiver manufacturers.

10.5.1 Audio coding

The audio specification may be found in Annex B of the Digital Television Standard. The audio is encoded with the AC-3 system which is documented in detail in the Digital Audio Compression (AC-3) Standard, ATSC Doc. A/52. The parameters of AC-3 used in the digital television system are constrained by limiting the audio sample rate to 48 kHz, prohibiting the use of the 1+1 audio coding mode, and restricting the maximum bit rates used for audio services. A main audio service is encoded at a bit rate less than or equal to 384 kbps. A single-channel associated service containing a single program element is encoded at a bit rate less than or equal to 128 kbps. A two-channel associated service containing only dialogue is encoded at a bit rate less than or equal to 192 kbps. The combined bit rate of a main service and an associated service, which are intended to be decoded simultaneously, is less than or equal to 512 kbps.

10.5.2 Audio channels and services

In general, a complete audio program which is presented to a viewer may consist of audio program elements from more than one audio elementary stream. Program

elements are delivered in elementary streams tagged as to audio service type. There are eight types of audio services defined.

Two service types are defined as main audio services: complete main program (CM); and music and effects (ME). Six service types are defined as associated audio services: visually impaired (VI); hearing impaired (HI); dialogue (D); commentary (C); emergency announcement (E); and voice-over (VO). The VI, HI, and C associated service types may be either complete program mixes, or may contain only a single program element. In general, a complete audio program is constructed by decoding one complete main audio service (CM), an associated service which is a complete program mix, or by decoding and combining one main audio service (CM or ME) and one associated audio service (VI, HI, D, C, E, or VO).

The AC-3 audio descriptor in the PSI data provides the receiver information about the audio service types which are present in a broadcast. The transport decoder is responsible for selecting which audio service(s) elementary bit stream(s) to deliver to the audio decoder.

A main audio service may contain from one to 5.1 audio channels. The 5.1 channels are left (L), center (C), right (R), left surround (LS), right surround (RS), and low frequency enhancement (LFE). Decoding of the LFE, channel is receiver optional. The LFE channel provides non-essential low frequency effects enhancement, but at levels up to 10 dB higher than the other audio channels. Reproduction of this channel is not essential to enjoyment of the program, and can be perilous if the reproduction equipment can not handle high levels of low frequency sound energy. Typical receivers may thus only decode and provide five audio channels from the selected main audio service, not six (counting the 0.1 as one).

An associated audio service which is a complete program mix may contain from one to 5.1 audio channels. An associated audio service containing a single program element intended to be combined with a main service, with one exception, contains only a single audio channel. In order to simultaneously decode a main service and an associated service, it is necessary for the audio decoder to be able to decode six audio channels (five from a main service plus one from an associated service). Receivers which also support optional decoding of the LFE channel thus need to support the decoding of seven audio channels. In the case that an ME main audio service is limited to two audio channels (2/0 mode), the D service may also contain two audio channels (2/0 mode). (This exception only requires the decoding of four audio channels and thus entails no additional decoder complexity.)

It is not necessary for every receiver to completely decode all of the encoded audio channels into separate audio signals. For instance, a monophonic receiver only needs to provide a single output channel. While the single monophonic output channel must represent a mix down of all of the audio channels contained in the audio program being decoded, simplifications of the mono decoder are possible. For instance, only a single output buffer is required so that decoder memory requirements are reduced; and, some of the mixdown may occur in the frequency domain thus reducing the complexity of the synthesis filter bank.

10.5.3 Loudness normalization

There is no regulatory limit as to how loud dialogue may be in an encoded bit stream. Since the digital audio coding system can provide more than 100 dB of dynamic range, there is no reason for dialogue to be encoded anywhere near 100% as is commonly done in NTSC television. However, there is no assurance that all program channels, or all programs or program segments on a given channel, will have dialogue encoded at the same (or even similar) level. Encoded AC-3 elementary bit streams are tagged with an indication of the subjective level at which dialogue has been encoded. The receiver should be capable of using this value to adjust the reproduced level of audio programs so that different received programs have their spoken dialogue reproduced at a uniform level. The receiver may then offer the viewer an audio volume control calibrated in absolute sound pressure level. The viewer could dial up the desired SPL for dialogue, and the receiver would scale the level of every decoded audio program so that the dialogue is always reproduced at the desired level.

10.5.4 Dynamic range control

It is common practice for high quality programming to be produced with wide dynamic range audio, suitable for the highest quality audio reproduction environment. Broadcasters, serving a wide audience, typically process audio in order to reduce its dynamic range. The processed audio is more suitable for the majority of the audience which does not have an audio reproduction environment which matches that of the original audio production studio. In the case of NTSC, all viewers receive the same audio with the same dynamic range, and it is impossible for any viewers to enjoy the original wide dynamic range production.

The AC-3 audio coding system provides a solution to this problem. A dynamic range control value (dynrng) is provided in each audio block (every 5 ms). These values are used by the audio decoder in order to alter the level of the reproduced audio for each audio block. Level variations of up to +24 dB may be indicated. The values of dynrng are generated in order to provide a subjectively pleasing but restricted dynamic range. The unaffected level is dialogue level. For sounds louder than dialogue, values of dynrng will indicate gain reduction. For sounds quieter than dialogue, values of dynrng will indicate a gain increase. The broadcaster is in control of the values of dynrng, and can supply values which generated the amount of compression which the broadcaster finds appropriate.

By default, the values of dynrng will be used by the audio decoder. The receiver will thus reproduce audio with dynamic range as intended by the broadcaster. The receiver may also offer the viewer the option to scale the value of in order to reduce the effect of the dynamic range compression which was introduced by the broadcaster. In the limiting case, if the value of dynrng is scaled to zero, then the audio will be reproduced with its full original dynamic range. The optional scaling of dynrng can be done differently for values indicating gain reduction (which makes quiet sounds louder). Thus the viewer may be given independent control of the amount of compression applied to loud and quiet sounds. The details of these control functions are up to each receiver implementation.

10.6 Guide to existing TV receiver requirements and standards

10.6.1 Introduction

The following is a listing of mandatory and voluntary standards, recommended practices, and other reference information for conventional television receivers. A study of these materials will be helpful in understanding what changes will be necessary to accommodate ATV receivers and what may have to be added to cover ATV receivers, if necessary.

10.6.2 Mandatory requirements

47, CFR, (FCC) Part 2, *Marketing and Importation*.

47, CFR, (FCC) Part 15, *Radio Frequency Devices*.

FCC/OET MP-2*, Measurement of UHF Noise Figure of TV Receivers*.

21CFR (FDA), Subchapter J, *Radiological Health*.

   Part 1000, *General*.

   Part 1002, *Records and Reports*.

   Part 1020, *Performance Standard for Ionizing Radiation Emitting Products*.

   Part 1020.10, *Television Receivers*.

ANSI C63.4, *Standard Methods Of Measurement Of Radio-Noise Emissions From Low-Voltage Electrical And Electronic Equipment In The Range Of 9 kHz To 40 GHz*.

FTC-16CFR, Part 410, *Deceptive Advertising As To Sizes Of Viewable Pictures Shown By Television Receiving Sets*.

10.6.3 Mandatory for "cable ready" receivers

EIA IS-6, *Recommended Cable Television Identification Plan*.

EIA IS-23, *RF Interface Specification For Television Receiving Devices And Cable Television Systems*.

EIA-Draft IS-132, NEW, *Channelization Plan for Cable Television Tuners*.

10.6.4 Mandatory in some states of the US and/or in Canada

UL 1410, *Television Receivers And High Voltage Video Product*, (can submit receivers using this standard until 2002).

UL-1492, *Audio-Video Products And Accessories*, (can submit receivers using this standard now and beyond 2002).

UL-1413, *High Voltage Components For Television Type Appliances*.

UL-1418, *Implosion Protected Cathode Ray Tubes For Television Type Appliances*.

NFPA-70, *National Electrical Code*.

CSA-C22.2 No. 1-M90, *Radio, Television And Electronic Apparatus*.

10.6.5 Voluntary standards

EIA IS-16A, *Immunity Of Television Receivers And Video Cassette Recorders (VCRs) To Direct Radiation From Radio Transmissions, O.5 To 30 MHz.*

EIA IS-31, *Recommended Design Guideline - Rejection Of Educational Interference To Ch 6 Television Reception.*

EIA-544, *Immunity Of TV And VCR Tuners To Internally Generated Harmonic Interference From Signals.*

10.6.6 Other related standards and references

10.6.6.1 FCC

47, CFR, (FCC) Part 73, *Radio Broadcast Services*.

47, CFR, (FCC) Part 76, *Cable Television Service*.

10.6.6.2 Safety & x-ray

NOM-001, *Electronic Apparatus-Household Electronic-Apparatus By Different Sources Of Electrical Power-Safety Requirements And Testing-Methods For Type Approval.*

CPEB1, *Standard Method Of Measurement Of Ionizing Radiation From Television Receivers For Factory Quality Assurance.*

CPEB2, *Definition Of Normal Operating Conditions For Television Receivers.*

CPEB3, *Measurement Instrumentation For X-Radiation From Television Receivers.*

EIA-500-A, *Recommended Practice For Measurement Of X-Radiation From Projection Cathode Ray Tubes.*

EIA-503-A, *Recommended Practice For The Measurement Of X-Radiation From Direct View Television Picture Tubes.*

10.6.6.3 Interference & immunity

EIA-378, *Measurement Of Spurious Radiation From FM And TV Broadcast Receivers In The Frequency Range Of 100 To 1000 MHz - Using The EIA Laurel Broadband Antenna.*

CISPR 13, *Limits And Methods Of Measurement Of Radio Interference Characteristics Of Sound And Television Broadcast Receivers And Associated Equipment.*

CISPR 20, *Limits And Methods Of Measurement Of Immunity Characteristics Of Sound And Television Broadcast Receivers And Associated Equipment.*

IEEE-187-90, *IEEE Standard On Radio Receivers: Open Field Method Of Measurement Of Spurious Radiation From FM And Television Broadcast Receivers.*

10.6.6.4 Cathode ray TV display tubes

EIA -256-A, *Deflecting Yokes For Cathode Ray Tubes*.

EIA-266-A, *Registered Screen Dimensions For Monochrome Picture Tubes*.

EIA-324-A, *Registered Screen Dimensions For Color Picture Tubes*.

EIA-493, *Recommended Practice For Conversion Of U.S. To Metric Dimensions For Color And Monochrome Cathode Ray Tubes And Their Component Parts*.

EIA-527, *Screen Definition For Color Picture Tubes*.

10.6.6.5 Voluntary TV receiver recommended practices

EIA-462, *Electrical Performance Standards For Television Broadcast Demodulators*.

EIA-563, *Standard Baseband (Audio/Video) Interface Between NTSC Television Receiving Devices And Peripheral Devices*.

REC-109-CH, *Intermediate Frequencies For Entertainment Receivers*.

TVSB1, *EIA Recommended Practice For Use Of A Vertical Interval Reference (VIR) Signal*.

TVSB3, *A History Of The Vertical Interval Color Reference Signal (VIR)*.

TVSB5, *Multichannel TV Sound System - BTSC System Recommended Practices*.

10.6.6.6 International standards, IEC, etc.

IEC 65, *Safety Requirements For Mains Operated Electronic And Related Apparatus For Household And Other Similar General Use*.

IEC 107-1 to -6, *Methods Of Measurement On Receivers For Television Broadcast Transmissions*, (6 Parts).

IEC 569, *Informative Guide For Subjective Tests On Television Receivers*.

IEC 68, *Environmental Testing - Part 1 : General and Guidance*.

ANNEX A

INFORMATIVE REFERENCES

ATSC Standard A/52 (1995), *Digital Audio Compression (AC-3).*

ATSC Standard A/53 (1995), *ATSC Digital Television Standard.*

AES 3-1992 (ANSI S4.40-1992), *AES Recommended Practice for digital audio engineering — Serial transmission format for two-channel linearly represented digital audio data.*

ANSI S1.4-1983, *Specification for Sound Level Meters.*

IEEE Standard 100-1992, *The New IEEE Standard Dictionary of Electrical and Electronic Terms.*

IEC 651 (1979), *Sound Level Meters.*

IEC 804 (1985), Amendment 1 (1989), *Integrating/Averaging Sound level Meters.*

ISO/IEC IS 13818-1, International Standard (1994), *MPEG-2 Systems.*

ISO/IEC IS 13818-2, International Standard (1994), *MPEG-2 Video.*

ISO/IEC CD 13818-4, MPEG Committee Draft (1994), *MPEG-2 Compliance.*

ITU-R BT.601-4 (1994), *Encoding parameters of digital television for studios.*

SMPTE 274M (1995), *Standard for television, 1920 x 1080 Scanning and Interface.*

SMPTE S17.392 (1995), *Proposed Standard for television, 1280 x 720 Scanning and Interface.*

SMPTE 259 (1993), *Standard for television, 10-Bit 4:2:2 Component and 4fsc NTSC Composite Digital Signals — Serial Digital Interface.*