Doc. A/54A
4 December 2003

Recommended Practice: Guide to the Use of the ATSC Digital Television Standard

Advanced Television Systems Committee
1750 K Street, N.W.
Suite 1200
Washington, D.C. 20006
www.atsc.org

The Advanced Television Systems Committee, Inc., is an international, non-profit organization developing voluntary standards for digital television. The ATSC member organizations represent the broadcast, broadcast equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries. Specifically, ATSC is working to coordinate television standards among different communications media focusing on digital television, interactive systems, and broadband multimedia communications. ATSC is also developing digital television implementation strategies and presenting educational seminars on the ATSC standards.

ATSC was formed in 1982 by the member organizations of the Joint Committee on InterSociety Coordination (JCIC): the Electronic Industries Association (EIA), the Institute of Electrical and Electronic Engineers (IEEE), the National Association of Broadcasters (NAB), the National Cable and Telecommunications Association (NCTA), and the Society of Motion Picture and Television Engineers (SMPTE). Currently, there are approximately 160 members representing the broadcast, broadcast equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries.

ATSC Digital TV Standards include digital high definition television (HDTV), standard definition television (SDTV), data broadcasting, multichannel surround-sound audio, and satellite direct-to-home broadcasting.

Table of Contents

List of Figures and Tables

Recommended Practice A/54A: Guide to the Use of the ATSC Digital Television Standard

1. SCOPE

This guide provides tutorial information and an overview of the digital television system defined by ATSC Standard A/53, *ATSC Digital Television Standard*. In addition, recommendations are given for operating parameters for certain aspects of the DTV system.

2. REFERENCES

2.1 Normative References

There are no normative references.

2.2 Informative References

1. 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," Audio Engineering Society, New York, N.Y.
2. ANSI S1.4-1983: "Specification for Sound Level Meters."
3. ATSC IS-191 (2003): "DTV Lip Sync at Emission Encoder Input: ATSC IS Requirements for a Recommended Practice," Advanced Television Systems Committee, Washington, D.C.
4. ATSC Standard A/52A (2001): "Digital Audio Compression (AC-3)," Advanced Television Systems Committee, Washington, D.C., August 20, 2001.
5. ATSC Standard A/53B (2001) with Amendment 1 (2002) and Amendment 2 (2003): "ATSC Digital Television Standard," Advanced Television Systems Committee, Washington, D.C., carrying the cover date of August 7, 2001.
6. ATSC Standard A/65B (2003): "Program and System Information Protocol," Advanced Television Systems Committee, Washington, D.C., March 18, 2003.
7. ATSC Standard A/70 (2000): "Conditional Access System for Terrestrial Broadcast with Amendment," Advanced Television Systems Committee, Washington, D.C., May 31, 2000.
8. IEC 651 (1979): "Sound Level Meters."
9. IEC 804 (1985), Amendment 1 (1989): "Integrating/Averaging Sound level Meters."
10. IEEE Standard 100-1992: *The New IEEE Standard Dictionary of Electrical and Electronic Terms*, Institute of Electrical and Electronics Engineers, New York, N.Y.
11. ISO/IEC 11172-1, "Information Technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 1: Systems."
12. ISO/IEC 11172-2, "Information Technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 2: Video."
13. ISO/IEC IS 13818-1:2000 (E), International Standard, Information technology – Generic coding of moving pictures and associated audio information: Systems.
14. ISO/IEC IS 13818-2, International Standard (1996), MPEG-2 Video.
15. ISO/IEC IS 13818-1:2000 (E), International Standard, Information technology – Generic coding of moving pictures and associated audio information: Systems.
16. ISO/IEC CD 13818-4, MPEG Committee Draft (1994): "MPEG-2 Compliance."
17. ITU-R BT. 601-4 (1994): "Encoding parameters of digital television for studios."
18. ITU-R BT.601-5 (1995): Encoding Parameters of Digital Television for Studios.

19. SMPTE 125M (1995): "Standard for Television—Component Video Signal 4:2:2, Bit-Parallel Digital Interface," Society of Motion Picture and Television Engineers, White Plains, N.Y.
20. SMPTE 170M (1999): "Standard for Television—Composite Analog Video Signal, NTSC for Studio Applications," Society of Motion Picture and Television Engineers, White Plains, N.Y.
21. SMPTE 267M (1995): "Standard for Television—Bit-Parallel Digital Interface, Component Video Signal 4:2:2 16 9 Aspect Ratio," Society of Motion Picture and Television Engineers, White Plains, N.Y.
22. SMPTE 274M (1998): "Standard for Television—1920 1080 Scanning and Analog and Parallel Digital Interfaces for Multiple Picture Rates," Society of Motion Picture and Television Engineers, White Plains, N.Y.
23. SMPTE 293M (2003): "Standard for Television—720 483 Active Line at 59.94-Hz Progressive Scan Production, Digital Representation," Society of Motion Picture and Television Engineers, White Plains, N.Y.
24. SMPTE 296M (2001): :Standard for Television—1280 720 Progressive Image Sample Structure, Analog and Digital Representation and Analog Interface, Society of Motion Picture and Television Engineers, White Plains, N.Y.
25. SMPTE/EBU: "Task Force for Harmonized Standards for the Exchange of Program Material as Bitstreams - Final Report: Analyses and Results," Society of Motion Picture and Television Engineers, White Plains, N.Y., July 1998.
26. SMPTE Recommended Practice 202 (2002): "Video Alignment for MPEG Coding," Society of Motion Picture and Television Engineers, White Plains, N.Y., 2002.
27. Digital TV Group: "Digital Receiver Implementation Guidelines and Recommended Receiver Reaction to Aspect Ratio Signaling in Digital Video Broadcasting," Issue 1.2, August 2000.

3. DEFINITIONS

The following definitions are included here for reference but the precise meaning of each may vary slightly from standard to standard. 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.

3.1 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).

3.2 Terms Employed

For the purposes of the Digital Television Standard, the following definitions 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 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.

AFT Active format description.

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.

ATM See *asynchronous transfer mode*.

ATTC Advanced Technology Test Center.

AWGN Additive white Gaussian noise.

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.

channel A digital medium that 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 used to verify the correctness of the data.

D-frame A frame coded according to an MPEG-1 mode that 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 which indicates the time that an access unit is decoded in the system target decoder.

DFS Data field synchronization.

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.

DSS Data segment synchronization.

DTV Digital television, the system described in the ATSC Digital Television Standard.

DTS See *decoding time-stamp*.

D/U Desired (signal) to undesired (signal) ratio.

DVCR Digital video cassette recorder

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 that 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 that 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*.

essence In its simplest form, *essence* = content – metadata. In this context, (video) essence is the image itself without any of the transport padding (H and V intervals, ancillary data, etc).

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.

FIR Finite-impulse-response.

forbidden This term, when used in clauses defining the coded bit stream, indicates that the value must 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 provides significantly improved picture quality relative to conventional (analog NTSC) television and a wide screen format (16:9 aspect ratio). The ATSC Standard enables transmission of HDTV pictures at several frame rates and one of two picture formats; these are listed in the top two lines of Table 5.1. The ATSC Standard also enables the delivery digital sound in various formats.

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

Huffman coding A type of source coding that uses codes of different lengths to represent symbols that 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.

ISI Intersymbol interference.

ISO International Organization for Standardization.

ITU International Telecommunication Union.

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.

LMS Least mean squares.

macroblock In the DTV 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.

Mbps 1,000,000 bits per second.

motion vector A pair of numbers that 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 itself.

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 that 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 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 that 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*.

PSIP Program and System Information Protocol, as defined in ATSC A/65.

PTS See *presentation time-stamp*.

quantizer A processing step that 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, all reserved bits are set to "1".

ROM Read-only memory.

SAW filter Surface-acoustic-wave filter.

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 of 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 bit rate compression. The results should be such that when judged across a representative sample of program material, subjective equivalence with NTSC is achieved. 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.

STC System time clock.

STD See *system target decoder*.

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.

still picture A coded still picture consists of a video sequence containing exactly one coded picture that 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, defined as 2.5 data segment errors per second.

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.

3.3 Symbols, Abbreviations, and Mathematical Operators

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.

3.3.1 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 – .
%	Modulus operator. Defined only for positive numbers.
Sign( )	Sign(x) = 1 $x > 0$ = 0 $x == 0$ = –1 $x < 0$

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.

3.3.2 Logical Operators

|| Logical OR.
&& Logical AND.
! Logical NOT.

3.3.3 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.

3.3.4 Bitwise Operators

& AND.
| OR.
>> Shift right with sign extension.
<< Shift left with 0 fill.

3.3.5 Assignment

= Assignment operator.

3.3.6 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.

3.3.7 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.

3.3.7.1 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.

3.3.7.2 Definition of nextbits Function

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

3.3.7.3 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 (Table 3.1). 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	**1**	**'0'**
while (nextbits()!='0000 0000 0000 0000 0000 0001')		
zero_byte	**8**	**'00000000'**
}		

4. OVERVIEW OF THE ATSC DIGITAL TELEVISION SYSTEM

The digital television (DTV) standard has ushered in a new era in television broadcasting. The impact of DTV is more significant than simply moving from an analog system to a digital system. Rather, DTV permits a level of flexibility wholly unattainable with analog broadcasting. An important element of this flexibility is the ability to expand system functions by building upon the technical foundations specified in ATSC standards such as the ATSC Digital Television Standard (A/53) and the Digital Audio Compression (AC-3) Standard (A/52).

With NTSC, and its PAL and SECAM counterparts, the video, audio, and some limited data information are conveyed by modulating an RF carrier in such a way that a receiver of relatively simple design can decode and reassemble the various elements of the signal to produce a program consisting of video and audio, and perhaps related data (e.g., closed captioning). As such, a complete program is transmitted by the broadcaster that is essentially in finished form. In the DTV system, however, additional levels of processing are required after the receiver demodulates the RF signal. The receiver processes the digital bit stream extracted from the received signal to yield a collection of program elements (video, audio, and/or data) that match the service(s) that the consumer selected. This selection is made using system and service information that is also transmitted. Audio and video are delivered in digitally compressed form and must be decoded for presentation. Audio may be monophonic, stereo, or multi-channel. Data may supplement the main video/audio program (e.g., closed captioning, descriptive text, or commentary) or it may be a stand-alone service (e.g., a stock or news ticker).

The nature of the DTV system is such that it is possible to provide new features that build upon the infrastructure within the broadcast plant and the receiver. One of the major enabling developments of digital television, in fact, is the integration of significant processing power in the receiving device itself. Historically, in the design of any broadcast system—be it radio or television—the goal has always been to concentrate technical sophistication (when needed) at the transmission end and thereby facilitate simpler receivers. Because there are far more receivers than transmitters, this approach has obvious business advantages. While this trend continues to

be true, the complexity of the transmitted bit stream and compression of the audio and video components require a significant amount of processing power in the receiver, which is practical because of the enormous advancements made in computing technology. Once a receiver reaches a certain level of sophistication (and market success) additional processing power is essentially "free."

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 about 19 Mbps in a 6 MHz terrestrial broadcasting channel and about 38 Mbps in a 6 MHz cable television channel. This means that encoding HD video essence at 1.106 Gbps[1] (highest rate progressive input) or 1.244 Gbps[2] (highest rate interlaced picture input) requires a bit rate reduction by about a factor of 50 (when the overhead numbers are added, the rates become closer). 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 compression scheme optimizes the throughput of the transmission channel by representing the video, audio, and data sources with as few bits as possible while preserving the level of quality required for the given application.

The RF/transmission subsystems described in the Digital Television Standard are designed specifically for terrestrial and cable applications. The structure is such that the video, audio, and service multiplex/transport subsystems are useful in other applications.

4.1 System Block Diagram

A basic block diagram representation of the system is shown in Figure 4.1. According to this model, the digital television system consists of four major elements, three within the broadcast plant plus the receiver.

[1] 720 × 1280 × 60 × 2 × 10 = 1.105920 Gbps (the 2 represents the factor needed for 4:2:2 color subsampling, and the 10 is for 10-bit systems)

[2] 1080 × 1920 × 30 × 2 × 10 = 1.244160 Gbps (the 2 represents the factor needed for 4:2:2 color subsampling, and the 10 is for 10-bit systems)

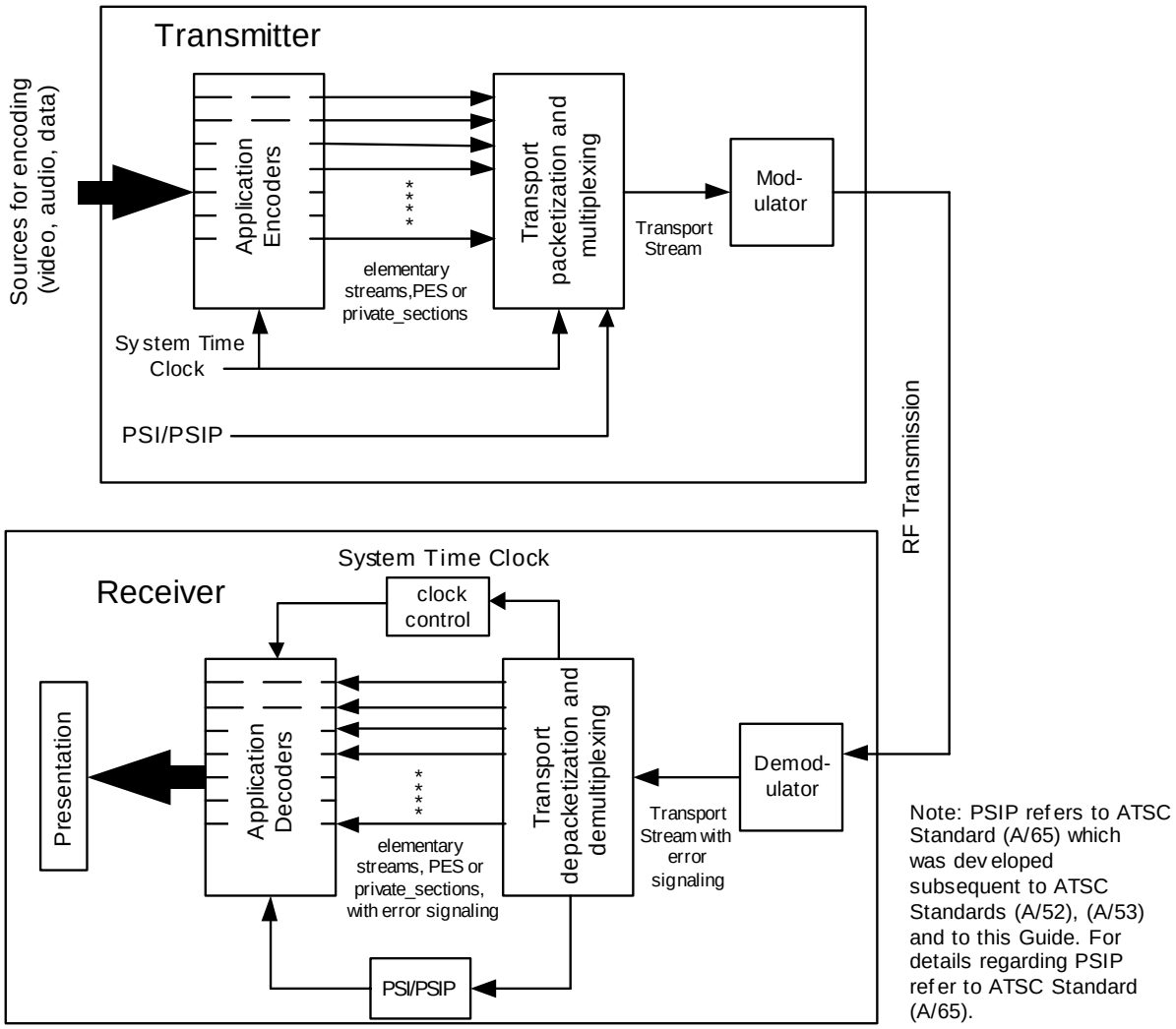

Figure 4.1 Block diagram of functionality in a transmitter/receiver pair.

4.1.1 Application Encoders/Decoders

The "application encoders/decoders," as used in Figure 4.1, refer to the bit rate reduction methods, also known as data compression, appropriate for application to the video, audio, and ancillary digital data streams. The purpose of compression is to minimize the number of bits needed to represent the audio and video information. The DTV system employs the MPEG-2 video stream syntax for the coding of video and the ATSC Standard "Digital Audio Compression (AC-3)" for the coding of audio.

The term "ancillary data" dates from the original drafting of A/53 and is a broad term that includes control data, and data associated with the program audio and video services. As standards were developed to define how to transport and process data, it became clear that different forms of data served very different purposes and different standards were needed for metadata and essence [22]. The discussion of data is out of scope for this document (contact ATSC for more information). Including system information as ancillary data is not strictly proper, as some such data is needed to re-assemble the audio, video, and data services. Data delivered as a separate payload can provide independent services as well as data elements related

to an audio- or video-based service. Accordingly, the term ancillary data is not used hereinafter in this document, to avoid confusion.

4.1.2 Transport (de)Packetization and (de)Multiplexing

"Transport (de)Packetization and (de)Multiplexing" refers to the means of dividing each bit stream into "packets" of information, the means of uniquely identifying each packet or packet type, and the appropriate methods of interleaving or multiplexing video bit stream packets, audio bit stream packets, and data bit stream packets into a single transport mechanism. The structure and relationships of these essence bit streams is carried in service information bit streams, also multiplexed in the single transport mechanism. 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 DTV system employs the MPEG-2 Transport Stream syntax for the packetization and multiplexing of video, audio, and data signals for digital broadcasting systems. 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.

4.1.3 RF Transmission

"RF Transmission" refers to channel coding and modulation. The channel coder takes the digital 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 bit stream information to modulate a carrier for the transmitted signal. The modulation subsystem offers two modes: an 8-VSB mode and a 16-VSB mode.

4.1.4 Receiver

The ATSC receiver recovers the bits representing the original video, audio, and other data from the modulated signal. In particular, the receiver performs the following functions:

- Tune the selected 6 MHz channel
- Reject adjacent channels and other sources of interference
- Demodulate (equalize as necessary) the received signal, applying error correction to produce a transport bit stream
- Identify the elements of the bit stream using a transport layer processor
- Select each desired element and send it to its appropriate processor
- Decode and synchronize each element
- Present the programming

Issues affecting receiver design are discussed thoroughly in Section 9 (Receiver Subsystem) of this Guide. In general, most attention in Section 9 is paid to recovery and demodulation of the terrestrial-broadcast RF signal, because it is the most challenging of receiver processes. Noise, interference, and multipath are elements of the terrestrial transmission path, and the receiver circuits are expected to deal with these impairments. Innovations in equalization, automatic gain control, interference cancellation, and carrier and timing recovery create product performance differentiation and improve signal coverage.

The decoding of transport elements that make up the programming is usually considered to be a more straightforward implementation of specifications, although opportunities for innovation in circuit efficiency or power usage exist. In particular, innovations in video decoding offer opportunities for savings in memory and circuit speed and complexity. The user interface and new data-based services are important areas of product differentiation.

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

5. VIDEO SYSTEMS

5.1 Overview of Video Compression and Decompression

The need for compression in a digital television 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 that required to represent a standard-definition television signal is about 200 Mbps, while the bit rate that can reliably be transmitted within a standard 6 MHz television channel is about 19 Mbps. This implies a need for about a 50:1 or greater compression ratio for HDTV and 10:1 or greater for standard definition.

The Digital Television Standard specifies video compression using a combination of compression techniques. 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 in all cases, represents 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 that conform to the MPEG-2 MP@HL can be expected to decode bit streams produced in accordance with the Digital Television Standard. Note 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 that comply to MPEG-2 MP@HL.

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

5.1.3 Overview of Video Compression

The video compression system takes in an analog or uncompressed digital 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 temporal and spatial complexity and perceptibility of distortions introduced by the compression.

Figure 5.1 shows the overall flow of signals in the ATSC DTV system. Video signals presented to the system are first digitized (if not already in digital signal form) and sent to the encoder for compression; 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.

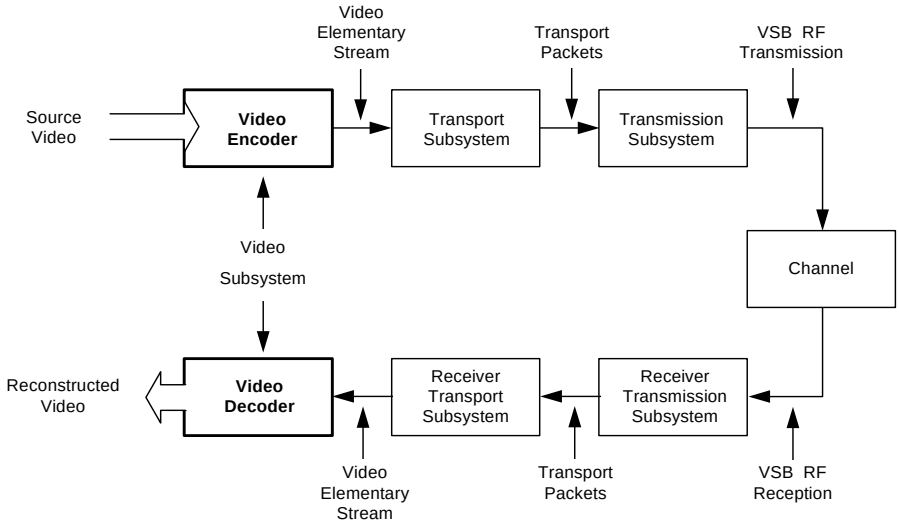

Figure 5.1 Video coding in relation to the DTV system.

5.2 Video Preprocessing

Video preprocessing converts the input signals to digital samples in the form needed for subsequent compression. Analog input signals are typically composite for standard definition signals or components consisting of luminance (Y) and chrominance (Pb and Pr) for high definition signals, are first decoded (for composite signals) then digitized as component luminance (Y) and chrominance (Cb and Cr) signals. Digital input signals, both standard definition and high definition, are typically serial digital signals carrying Y, Cb, Cr components. The input signals may undergo pre-processing for noise reduction and/or other processing algorithms that improve the efficiency of the compression encoding. Further processing is then carried out for chrominance and luminance filtering and sub-sampling (see Section 5.2.5 for more information).

5.2.1 Video Compression Formats

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

Table 5.1 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

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.

Receiver designers should be aware that a larger range of video formats is allowed under SCTE 43, and that consumers may expect receivers to decode and display these as well. One format likely to be frequently encountered is 720 pixels by 480 lines (encoded from ITU-R BT. 601 input signals with 720 pixels by 483 lines). See SCTE 43, Table 3.

5.2.1.1 Possible Video Inputs

While not required by the Digital Television Standard, there are certain digital 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	Picture Rate
SMPTE 274M-1998	1080	1920	24P, 30P, 60I
SMPTE 296M-2001	720	1280	24P, 30P, 60P
SMPTE 293M-2003	483	720	60P
ITU-R BT. 601-5	483	720	60I

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 both the 1080- and 720-line formats, other frame rates, specifically 23.976, 24.00, 29.97, and 30.00 fps rates are acceptable as input to the system. The sample frequency will be either 74.25 MHz (for 24.00 and 30.00 fps) or 74.25/1.001 MHz for the other rates. The number of

total samples per line is the same for either of the paired picture rates. See SMPTE 274M and SMPTE 296M.

The six frame rates noted are the only allowed frame rates for the Digital Television Standard. In this document, references to 24 fps include both 23.976 and 24.00 fps, references to 30 fps include both 29.97 and 30.00 fps, and references to 60 fps include both 59.94 and 60.00 fps.

For the 480-line format, there may be 704 or 640 pixels in the active line. The interlaced formats are based on ITU-R BT. 601-5; the progressive formats are based on SMPTE 294M. If the input is based on ITU-R BT. 601-5 or SMPTE 294M, it will have 483 or more active lines with 720 pixels in the active line. Only 480 of these active lines are encoded. The lines to be encoded should be lines 23–262 and 286–525 for 480I and lines 45–524 for 480P, as specified in SMPTE Recommended Practice RP-202, "Video Alignment for MPEG Coding." 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. 601-5 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 SMPTE 296M 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 Part 2 colorimetry. Video inputs corresponding to ITU-R BT. 601-5 should have 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 be implicitly defined by the application. Therefore, the colorimetry should always be explicitly signaled using sequence_display_extension(). If this information is not transmitted, receiver behavior cannot be predicted.

In generating bit streams, broadcasters should understand that some receivers will display 480-line formats according to SMPTE 170M colorimetry (value 0x06) and 720- and 1080-line formats according to SMPTE 274M colorimetry (value 0x01). It is believed that few receivers will display properly the other colorimetry combinations allowed by ISO/IEC 13818-2. Legacy material using SMPTE 240M colorimetry should be treated as if it used ITU-R BT. 709 Part 2 colorimetry.

5.2.2 Precision of Samples

Samples are typically obtained using analog-to-digital converter circuits with 10-bit precision. After studio processing, the various luminance and chrominance samples will typically be represented using 8 or 10 bits per sample for luminance and 8 bits per sample for each chrominance component. The limit of precision of the MPEG-2 Main Profile is 8 bits per sample for each of the luminance and chrominance components.

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

Material originated at 24 frames per second, such as that shot on film, is typically converted to 30 or 60 frame-per-second video for broadcast. In the case of 30 fps interlaced television, this means that each four frames of film are converted to ten fields, or five frames of video. In the case of 60 fps progressive-scan television, each four frames of film are converted into ten frames of video. This conversion is done using the so-called 3:2 pulldown sequence; prior to the introduction of 24P video equipment it was an inherent part of the telecine process.

In the 3:2 pulldown, the first frame of film is converted to two pictures (frames or fields, depending on whether the output format is 60P or 30I respectively). The second frame is converted to three pictures, the third to two pictures and the fourth to three pictures.

When describing the sequence, the film frames are conventionally labelled A, B, C and D; the video fields or frames 1-5 (interlaced) or 1-10 (progressive). In the interlaced case, the third field generated from film frame A is field 1 of Frame 3; the third field generated from film frame C is field 2 of Frame 5. Note that in the interlaced case, Frame 3 will contain video from film Frames B and C and Frame 4 will contain video from film Frames C and D.

It is inefficient to code these sequences directly; not only is there a great deal of repeated information, but in interlace, Frames 3 and 4 each contain fields from two different film frames, so there may be motion differences between the two fields. MPEG therefore provides tools specifically for coding these sequences; these are top_field_first and repeat_first_field (see 13818-2, clauses 6.2.3.1 and 6.3.10)

It is relatively straightforward for the encoder to detect the repeated frames in progressive-scan video derived from 24 fps material. It is less straightforward to detect the repeated fields in interlaced video. Particularly with interlaced material, it is important that the 3:2 pulldown sequence be maintained; if it is not, encoder efficiency and picture quality may suffer. For this reason, it is becoming more common for material to be edited at 24 Hz before frame-rate conversion to 30I or 60P.

5.2.5 Color Component Separation and Processing

The input video source to the ATSC DTV 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. While 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 ATSC DTV 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.

It must be noted that the luminance component Y is not true luminance as this term is used in color science; this is because the RGB-to-YC matrixing operation is performed after the optoelectronic transfer characteristic (gamma) is applied. For this reason, some experts refer to the Y component as luma rather than luminance. While the preponderance of the luminance information is present in the luma, some of it ends up in the chroma, and it can be lost when the chroma components are sub-sampled.

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 Concatenated Sequences

The MPEG-2 video standard that 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, typically contains 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 (unless the insertion equipment is capable of scaling those parameters in real time). Thus the situation of concatenated video sequences arises.

---

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

Regarding concatenated sequences, the MPEG-2 Video Standard (13818-2) states:

> The behavior of the decoding process and display process for concatenated sequences is not within the scope of this Recommendation | International Standard. An application that needs to use concatenated sequences must ensure by private arrangement that the decoder will be able to decode and play concatenated sequences.

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 that 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 should be such that the parity of the interlaced sequences is preserved as if no concatenation had occurred.

5.4 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:

- Sequence layer information is very helpful and it is important that it be sent before every I-frame, independent of the interval between I-frames. Use of intra-macroblock refresh in the decoder can improve receiver channel-change performance.
- Some receivers rely on periodic transmission of I-frames for refreshing. The frequency of occurrence of I-frames may determine the channel-change time performance of the receiver. It is recommended that I-frames be sent at least once every 0.5 second in order to have acceptable channel-change performance in such receivers.
- 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.

5.5 Active Format Description (AFD)

With the approval of Amendment 1 to A/53B, active format description data has been added to the ATSC Digital Television Standard. The term "active format" in this context refers to that portion of the coded video frame containing "useful information." For example, when 16:9 aspect ratio material is coded in a 4:3 format (such as 480i), letterboxing may be used to avoid cropping the left and right edges of the widescreen image. The black horizontal bars at the top and bottom of the screen contain no useful information, and in this case the AFD data would indicate 16:9 video carried inside the 4:3 rectangle. The AFD solves a troublesome problem in the transition from conventional 4:3 display devices to widescreen 16:9 displays, and also

addresses the variety of aspect ratios that have been used over the years by the motion picture industry to produce feature films.

There are, of course, a number of different types of video displays in common usage—ranging from 4:3 CRTs to widescreen projection devices and flat-panel displays of various design. Each of these devices may have varying abilities to process incoming video. In terms of input interfaces, these displays may likewise support a range of input signal formats—from composite analog video to IEEE 1394.

Possible video source devices include cable, satellite, or terrestrial broadcast set-top (or integrated receiver-decoder) boxes, media players (such as DVDs), analog or digital VHS tape players, and personal video recorders.

Although choice is good, this wide range of consumer options presented two problems to be solved:

- First, no standard method had been agreed upon to communicate to the display device the "active area" of the video signal. Such a method would be able, for example to signal that the 4:3 signal contains within it a letterboxed 16:9 video image.
- Second, no standard method had been agreed upon to communicate to the display device, for all interface types, that a given image is intended for 16:9 display.

The AFD solves these problems and, in the process, provides the following benefits:

- Active area signaling allows the display device to process the incoming signal to make the highest-resolution and most accurate picture possible. Furthermore, the display can take advantage of the knowledge that certain areas of video are currently unused and can implement algorithms that reduce the effects of uneven screen aging.
- Aspect ratio signaling allows the display device to produce the best image possible. In some scenarios, lack of a signaling method translates to restrictions in the ability of the source device to deliver certain otherwise desirable output formats.

5.5.1 Active Area Signaling

A consumer device such as a cable or satellite set-top box cannot reliably determine the active area of video on its own. Even though certain lines at the top and bottom of the screen may be black for periods of time, the situation could change without warning. The only sure way to know active area is for the service provider to include this data at the time of video compression and to embed it into the video stream.

Figure 5.8 shows 4:3- and 16:9-coded images with various possible active areas. The group on the left is either coded explicitly in the MPEG-2 video syntax as 4:3, or the uncompressed signal provided in NTSC timing and aspect ratio information (if present) indicates 4:3. The group on the right are coded explicitly in the MPEG-2 video syntax as 16:9, provided with NTSC timing and an aspect ratio signal indicating 16:9, or provided uncompressed with 16:9 timing across the interface.

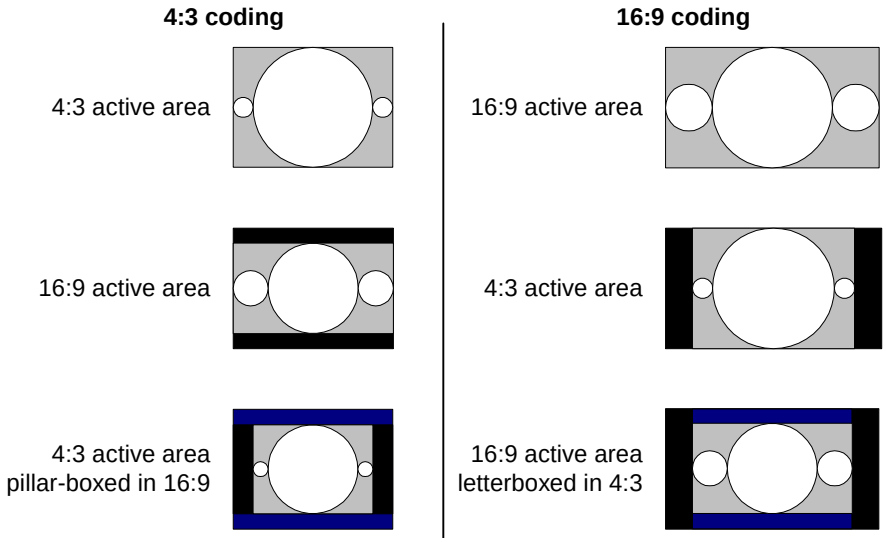

Figure 5.8 Coding and active area.

As can be seen in the figure, a pillar-boxed display results when a 4:3 active area is displayed within a 16:9 area, and a letterboxed display results when a 16:9 active area is displayed within a 4:3 area. It is also apparent that double-boxing can also occur, for example when 4:3 material is delivered within a 16:9 letterbox to a 4:3 display. Or, when 16:9 material is delivered within a 4:3 pillar-box to a 16:9 display.

For the straight letter- or pillar-box cases, if the display is aware of the active area it may take steps to mitigate the effects of uneven screen aging. Such steps could, for example, involve using gray instead of black bars. Some amount of linear or nonlinear stretching and/or zooming may be done as well using the knowledge that video outside the active area can safely be discarded.

The two double-boxed cases can occur as a result of poor or uninformed production choices made by the service provider, in some cases in concert with the content provider. Whenever 4:3 material is coded as 16:9, double boxing occurs when the 4:3 display places the 16:9 coded frame on screen. Whenever 16:9 material is coded as 4:3, double boxing occurs when the 16:9 display pillar-boxes the 4:3 coded frame.

A common situation that will cause double-boxing on a 16:9 digital TV display occurs when a 4:3 NTSC signal is encoded as 480i MPEG video, but the NTSC image is a letterboxed widescreen movie. Regardless of the cause, two aspects of the problem are of prime importance:

- The display device should not be expected to process the double-boxed image to fill the screen to make up for incorrectly coded content.
- Content and service providers should be expected to deliver properly coded content. Native 4:3 content must be delivered coded as 4:3. Native 16:9 content must be delivered coded as 16:9. Letterboxed widescreen video in NTSC should not be coded as 4:3, but should be coded into a 16:9 coded frame.[5]

5.5.2 Existing Standards

Several industry standards include some form of active area information, including EIA-608-B and DVB Active Format Description (AFD). The EIA-608-B standard is applicable only to NTSC analog video. The DVB description data applies only to compressed MPEG-2 video. Also

[5] Letterboxed content inside a 4:3 results in vertical resolution less than standard-definition television.

working in this area, MPEG has adopted an amendment to the MPEG-2 Video standard to include active area data.

Letterboxed movies can be seen on cable, satellite, and terrestrial channels today. If one observes closely, considerable variability in the size of the black bar areas can be seen. In fact, variations can be seen even over the course of one movie.

As mentioned previously, a display device may wish to mitigate the effects of uneven screen aging by substituting gray video for the black areas. It is problematic for the display to be required to actively track a varying letterbox area, and real-time tracking of variations from frame to frame would be difficult (if not impossible).

Clearly, two approaches are possible. First, include—on a frame-by-frame basis—a video parameter identifying the number of black lines (for letterbox) or number of black pixels (for pillar-box). Second, standardize on just two standard aspect ratios: 16:9 and 4:3.

5.5.3 Treatment of Active Areas Greater than 16:9

Any wide-screen source material can be coded into a 16:9-coded frame. No aspect ratio for coded frames exceeding 16:9 is standardized for cable, terrestrial broadcast, or satellite transmission in the U.S. If the aspect ratio of given content exceeds 16:9, the coded image will be letterboxed inside the 16:9 frame, as shown in Figure 5.9, where 2.35:1 material is letterboxed inside the 16:9 frame.

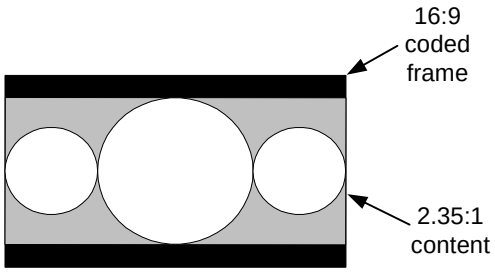

Figure 5.9 Example of active video area greater than 16:9 aspect ratio.

It can be helpful for a display to know the actual aspect ratio of the active portion of the 16:9 coded frame for a variety of reasons, including:

- Reduction in the effects of uneven screen aging. The display device controller may wish to use gray instead of black for the bars.
- The display may offer the user a "zoom" option to make better use of available display area, and knowledge of the aspect ratio can automate the selection of this display option. The zoom feature can be vertical scaling only, or a combination of horizontal and vertical where the leftmost and rightmost portions of the image are sacrificed to fill the screen area vertically.

Several standards include aspect ratio data. The MPEG-2 video syntax includes horizontal and vertical size data and aspect ratio indication of the coded image. An NTSC signal is normally thought to be intended for 4:3 display, but this is not always the case. EIA-608-B includes a "squeezed" bit, and IEC 61880 defines a method for NTSC VBI line 20. The line-20 method is currently used for playback of anamorphically coded DVDs, when the DVD player supports it and is properly set up by the user.

5.5.4 Active Format Description (AFD) and Bar Data

In recognition of these issues, the ATSC undertook a study of available options and—in Amendment 1 to A/53B—decided to endorse the basic signaling structure developed by the DVB consortium. The benefits of common active format description signaling across many different markets is easily understood.

Some common active video formats represented by the 4-bit AFD field include:

- The aspect ratio of the active video area is 16:9; when associated with a 4:3 coded frame, the active video is top-justified
- Active video area is 16:9; when associated with a 4:3 coded frame, the active video is centered vertically
- Active video area is 4:3; when associated with a 16:9 coded frame, the active video is centered horizontally
- Active video area exceeds 16:9 aspect ratio; active video is centered vertically (in whatever coded frame is used)
- Active video area is 14:9; when associated with a 4:3 coded frame the active video is centered vertically; when associated with a 16:9 coded frame, it is centered horizontally

It should be noted that certain active formats signal to the receiver that active video may be safely cropped in the receiver display (see [27], "Digital Receiver Implementation Guidelines and Recommended Receiver Reaction to Aspect Ratio Signaling in Digital Video Broadcasting."

In addition to AFD, Amendment 1 defined another data structure, bar_data(), also for use in the video Elementary Stream. The bar_data() structure, like AFD, appears in the picture user_data() area of the video syntax. While the AFD gives a general view of the relationship between the coded frame and the geometry of the active video within it, bar_data() is able to indicate precisely the number of lines of black video at the top and bottom of a letterboxed image, or the number of black pixels at the left and right side of a pillar-boxed image.

For the ATSC system, AFD and/or bar_data() are included in video user data whenever the rectangular picture area containing useful information does not extend to the full height or width of the coded frame. Such data may optionally also be included in user data when the rectangular picture area containing useful information extends to the full height and width of the coded frame.

The AFD and bar_data() are carried in the user data of the video Elementary Stream. After each sequence start (and repeat sequence start) the default aspect ratio of the area of interest is signalled by the sequence header and sequence display extension parameters. After introduction, each type of active format data remain in effect until the next sequence start or until another instance is introduced. Receivers are expected to interpret the absence of AFD and bar_data() in a sequence start to mean the active format is the same as the coded frame. Since it is not able to represent non-standard video aspect ratios, AFD may be only an approximation of the actual active video area. However when bar_data() is present, it should be assumed to be exact. If the bar_data() and the AFD are in conflict, the bar_data() should take precedence.

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) that 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.

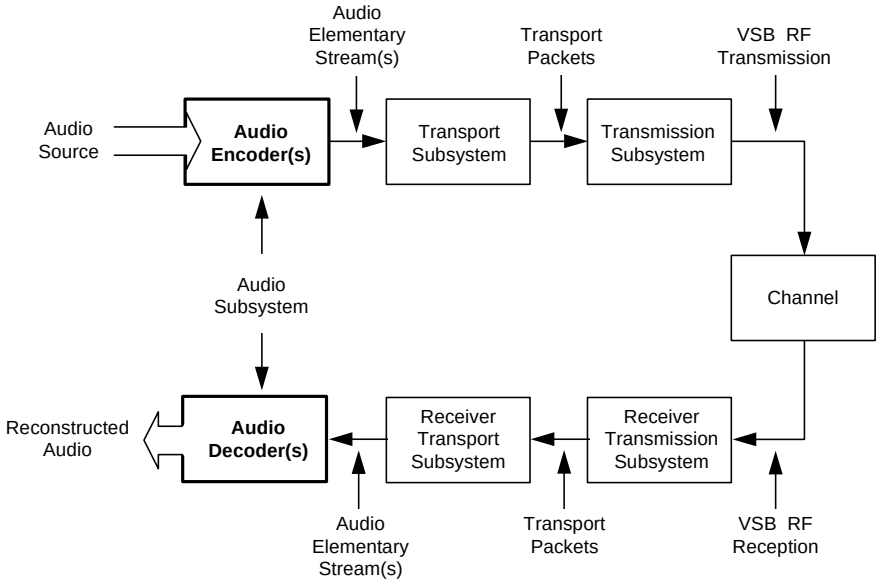

Figure 6.1 Audio subsystem within the digital television system.

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 that 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 should be 48 kHz, or the audio encoder should contain sampling rate converters that 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 that 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 that 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.

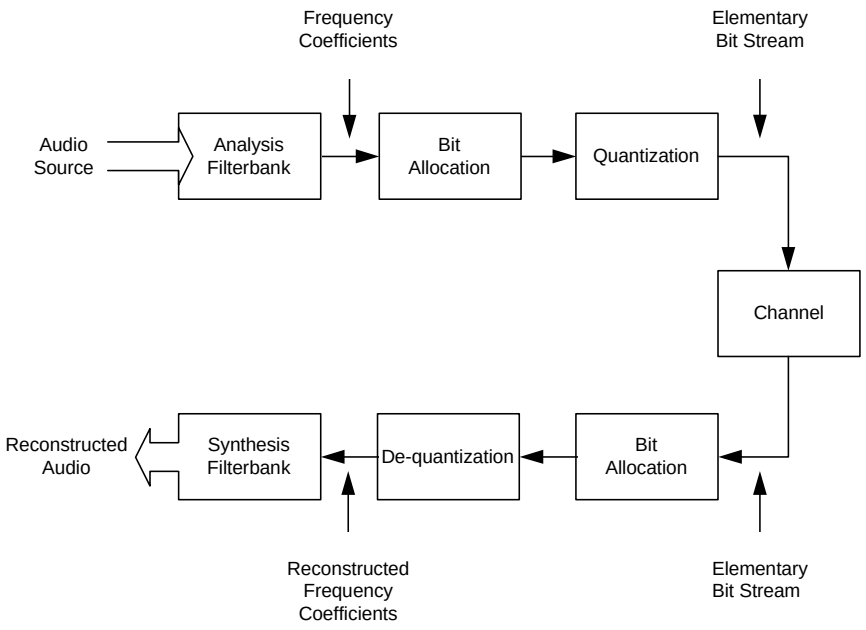

Figure 6.2 Overview of audio compression system.

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 percent redundancy that 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 that is reproduced prior to the impulse to be audible. Time smearing of quantization noise may be reduced by altering the length of the transform that 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 that 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 that 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 that 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 percent 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 that is identical to the decoded representation that will be used by the decoder. The decoded representation is then used to shift the original frequency coefficients to generate the normalized mantissas that 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 that 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 that 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 that 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 that 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).

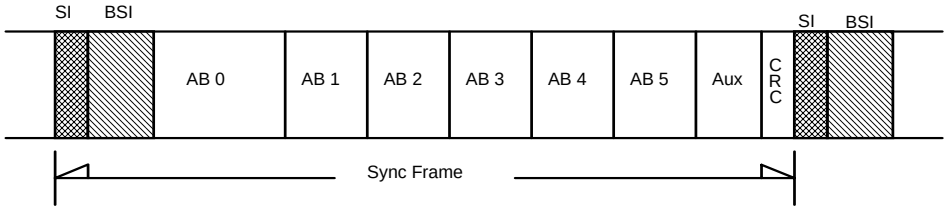

Figure 6.3 AC-3 synchronization frame.

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 that 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 that 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 that 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 that 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 that 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 that 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 that 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 that 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 that 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 that 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 that are encoded. Conforming all programs to a common dialogue level may mean that for some programs the audio level never approaches 100 percent 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 percent (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 that 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 that 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 that 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 that 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 that 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 that 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
- 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 that 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 that are complete mixes may be decoded and used as is. They are identified by the full_svc bit in the AC-3 descriptor (see [4], Annex A). Associated services that 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 that 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 that 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] 5.1 channel sound refers to a system reproducing the following signals: right, center, left, right surround, left surround, and low-frequency enhancement (LFE).

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 that 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) that 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) that 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 is indicated in the AC-3 descriptor (see [4], 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 is indicated in the AC-3 descriptor (see [4], 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 gives 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 gives 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 that 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 that 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 that 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 to 448 kbps 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 that is a complete service (containing all necessary program elements), at a bit rate up to and including 448 kbps. Transmissions may contain main audio services, or associated audio services that are complete services (containing all necessary program elements), encoded at a bit rate up to and including 448 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 that are intended to be simultaneously reproduced is less than or equal to 576 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	384–448 kbps
CM, ME, or associated audio service containing all necessary program elements	4	320-384 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	64-128 kbps
HI, narrative only	1	64-96 kbps
D	1	64-128 kbps
D	2	96-192 kbps
C, commentary only	1	64-128 kbps
E	1	64-128 kbps
VO	1	64-128 kbps

7. DTV TRANSPORT

7.1 Introduction

The ATSC DTV system described in core documents A/52 and A/53 provides the framework for conveying information to consumers. Built into this framework is a toolkit of features that can be used to extend the capabilities of the DTV system far beyond what the initial designers might have envisioned. This extensibility is, perhaps, the greatest benefit of digital technology, as well as the source of some confusion about what is required by the Digital Television Standard.

This section provides a tutorial description of the functionality and format of the transport subsystem employed in the ATSC DTV system. It is intended to aid the reader in understanding and applying the precise specification of the transport subsystem given in the underlying normative standards documents. The ATSC transport subsystem standard is based on the MPEG-2 Systems standard (ISO/IEC 13818-1) [13] and is further constrained and extended by Annex C of the Digital Television Standard (A/53) [5]. The MPEG-2 Standard was developed by the Moving Picture Experts Group, part of the International Standards Organization.

The transport subsystem employs the fixed-length transport stream packetization approach defined in ISO/IEC13818-1, which is usually referred to as the MPEG-2 Systems Standard. This approach is well-suited to the needs of terrestrial broadcast and cable television transmission of digital television. The use of relatively short, fixed-length packets matches well with the needs and techniques for error protection in both terrestrial broadcast and cable television distribution environments.

The ATSC DTV transport may carry a number of television programs. The MPEG-2 term "program" corresponds to an individual digital TV channel or data service, where each program is composed of a number of MPEG-2 program elements (i.e., related video, audio, and data streams). The MPEG-2 Systems Standard support for multiple channels or services within a single, multiplexed bit stream enables the deployment of practical, bandwidth efficient digital broadcasting systems. It also 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 subsystem on MPEG-2, maximum interoperability with other media and standards is maintained.

Figure 4.1 illustrates the organization of a digital television transmitter-receiver pair and the location of the transport subsystem in the overall system. The transport subsystem resides between the application (e.g., audio or video) encoding and decoding functions 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. The transport subsystem also incorporates other higher-level functionality related to identification of applications and synchronization of the receiver.

7.2 MPEG-2 Basics

The MPEG-2 standards are built upon the foundations of the MPEG-1 standards [11]. While the MPEG-1 standards were developed primarily to address the then upcoming video CD marketplace's need for an interoperable solution for compressed digital video storage and real-time playback at rates of about 1.5 Mbps, MPEG-2 was developed to primarily address the broadcast digital television and DVD markets and includes new features such as:

- Improved video and audio compression technologies
- Encoding support for both 4:2:0 and 4:2:2 video

- Support for the transmission of the coded bit streams in error-prone environments
- Support for multiple programs ("channels") in a single, multiplexed stream. This includes improved synchronization with the capability for each program to have a unique time-base, and the ability to describe and identify a network consisting of multiple multiplexed streams, each containing multiple programs
- Conditional access support
- Stream buffer management including buffer initialization
- Private data transport support

In contrast to previously developed standards, the MPEG-2 standards were designed to support full ITU-R 601 standard-definition resolutions, high-definition resolutions, and interlaced sequences. The MPEG-2 standards were also designed to support multi-channel networks carried in error-prone environments (such as terrestrial broadcasting), and the basic constructs used to encapsulate private data and a multitude of data essence formats. MPEG-2 standards are the foundation of several digital television technologies including digital set top boxes (STB), high-definition television (HDTV), and data broadcasting.

The MPEG-2 Systems Standard [13] defines the bit stream syntax and the methods necessary for (de)multiplexing, transporting, and synchronizing coded video, coded audio, and other data (including data essence not defined by the MPEG standards, referred to as "private data"). The standard includes the definition of packet formats, the synchronization and timing model, the mechanism for identifying content carried in the bit stream, and the buffer models used to enable a receiving device to properly decode and reconstruct the video, audio, and/or data presentation. The MPEG-2 Systems Standard as constrained and extended by the ATSC is the basis for the remainder of this section.

7.2.1 Standards Layering

The MPEG-2 Systems Standard (ISO/IEC 13818-1) [13] provides a toolkit that can be used to create the DTV transport bit stream. This toolkit can be thought of as providing general purpose functionality. Users of the MPEG-2 standards (such as the ATSC) choose tools from the toolkit and specify how they may be used (i.e., specify constraints on the syntax and semantics of the MPEG-2 standards). A/53 describes which portions of the MPEG-2 Systems Standard are to be used in creating the ATSC bit stream and also describes the constraints imposed.

In addition to constraining the MPEG-2 Systems Standard, the ATSC has also created compatible extensions to the standard. Some syntactical fields in the MPEG-2 Systems Standard are user defined—other fields have user private ranges. The ATSC is considered a "user" of the MPEG-2 standards and has used the user private areas to create ATSC standardized extensions to the MPEG-2 standards.

7.3 MPEG-2 Transport Stream Packet

An MPEG-2 Transport Stream is a continuous series of MPEG-2 Transport Stream packets. An MPEG-2 Transport Stream packet is 188 bytes in length and always begins with the synchronization byte 0x47.

7.3.1 MPEG-2 TS Packet Structure

The first four bytes of the MPEG-2 Transport Stream packet are the Transport Stream packet header. The remaining 184 bytes of an MPEG-2 Transport Stream packet may contain an optional adaptation field and up to 184 bytes of Transport Stream packet payload. If the adaptation field is present, it immediately follows the last byte of the Transport Stream packet header. The adaptation field is not part of the Transport Stream packet header nor the Transport

Stream packet payload. When the adaptation field is present, the MPEG-2 Transport Stream packet payload's size is 184 bytes minus the length of the adaptation field.

The definition of the contents of an MPEG-2 Transport Stream packet payload may differ depending upon the MPEG-2 stream_type and the encapsulation method.

7.3.2 MPEG-2 Transport Stream Packet Syntax

In the packet header, the Packet Identifier (PID) is a 13-bit value used to identify multiplexed packets within the MPEG-2 Transport Stream. Assigning a unique PID value to each bit stream allows Transport Stream packets from up to 8192 (2^{13}) separate bit streams to be simultaneously carried within the MPEG-2 Transport Stream. Note that not all bit streams are MPEG-2 Program Elements (e.g., PSI), but all Program Elements are bit streams. The PID provides a unique bit stream (and, therefore, Program Element) association for each Transport Stream packet.

The payload_unit_start_indicator is used to signal to the decoder (by being set to '1') that the first byte of something "interesting" can be found within the payload of the current MPEG-2 Transport Stream Packet (an MPEG-2 PES packet (see Section 7.4.4) or MPEG-2 section (see Section 7.4.1). This form of signaling, combined with hardware filtering in the decoder, allows for considerable efficiencies in decoding the contents of the stream. A PES packet must always commence as the first byte of the Transport Stream packet payload and only a single PES packet may begin in a Transport Stream packet. Thus, two PES packets (or portions thereof) are not permissible in a single Transport Stream packet.

For MPEG-2 sections (PSI and private sections) carried as payload, when the payload_unit_start_indicator field is set to '1', then the first byte of the MPEG-2 Transport Stream packet payload carries the pointer_field, which indicates the byte offset from the start of the Transport Stream packet payload to the beginning of the next PSI or private section. If the payload_unit_start_indicator field is set to '0', then the first byte of the Transport Stream packet payload is not a pointer_field. Instead, the Transport Stream packet payload contains the continuation of a previously started PSI or private section along with any necessary stuffing bytes.

The transport_scrambling_control field indicates if the MPEG-2 Transport Stream packet payload has been scrambled. The MPEG-2 Transport Stream packet header, the optional adaptation field, and the payload of a Null MPEG-2 Transport Stream packet (see Section 7.3.2.1) are never scrambled.

The adaptation_field_control field signals the inclusion of the optional adaptation field. The most significant bit of the two-bit field always indicates the presence of the adaptation field. The least significant bit indicates the presence of payload.

The continuity_counter field is a 4-bit rolling counter associated with MPEG-2 Transport Stream packets carrying the same PID. The counter is incremented by one for each consecutive Transport Stream packet for a given PID except when the adaptation_field_control field is set to indicate that the Transport Stream packet contains an adaptation field only (no payload) or if it is set to the 'reserved' value, or if the Transport Stream packet is a duplicate[7] (these exception cases are known as "non-incrementing conditions"). The continuity_counter is considered "continuous" if it has incremented by one from the continuity_counter value in the previous Transport Stream packet of the same PID or when any of the non-incrementing conditions have been met. The continuity

[7] The MPEG-2 Systems Standard defines a duplicate Transport Stream packet to be the second of two—and only two—consecutive Transport Stream packets having the same PID that are carrying payload and contain identical byte-by-byte contents (except for the program clock reference, if present). Duplicate Transport Stream packets may be used for additional error resilience purposes.

counter is considered "discontinuous" if it has not incremented by one from the continuity counter value in the previous Transport Stream packet having the same PID and a non-incrementing condition has not been met. Except in the case when the discontinuity_indicator flag[8] has been set to '1' to signal a discontinuous continuity_counter, if a receiver encounters a situation where the continuity_counter is discontinuous, then it should assume that some number of MPEG-2 Transport Stream packets have been lost.

Two other fields, the transport_error_indicator and the transport_priority, which are not typically used in ATSC transport Streams, are also carried in the packet header. The transport_error_indicator may be used to indicate that at least one uncorrectable bit error exists in the Transport Stream packet. The transport_priority field may be used to indicate that a Transport Stream packet with the field set to '1' is of higher priority than other Transport Stream packets having the same PID which do not have the field set to '1'.

The payload field carries the data content. The data content can be one of many types; for example, an MPEG-2 PES packet (which itself may contain an elementary stream) or one or more PSI or private sections.

7.3.2.1 The MPEG-2 Transport Stream Null Packet

The MPEG-2 Transport Stream Null packet is a special Transport Stream packet designed to pad an MPEG-2 Transport Stream. While individual MPEG-2 Programs (services) within a multiplexed bit stream may have variable bit-rate characteristics, the overall MPEG-2 Transport Stream must have a constant bit rate. MPEG-2 Transport Stream Null packets are transmitted when there are no other packets ready to be transmitted. This is necessary, since the MPEG-2 equipment creating the Transport Stream must maintain a constant bit rate output. Note that null packets may be added and/or removed by any re-multiplexing process within the data path.

MPEG-2 Transport Stream Null packets are always identified by a PID with value 0x1FFF. The Transport Stream Null packet payload may contain any data values. The continuity_counter of a Null Transport Stream packet is undefined, carries no information, and should be ignored.

7.4 MPEG-2 Transport Stream Data Structures

MPEG-2 Systems defines two fundamental bit stream data structures. The first, generically called a "section," is used to encapsulate either descriptive information about the data essence streams (coded video, coded audio, or data) within the Transport Stream service multiplex (e.g., stream type, information needed to extract the streams, program guide information) or a "private data" essence stream itself. The second, called a "Packetized Elementary Stream (PES) packet" is used to encapsulate elementary stream data essence (e.g., coded video, coded audio, or data).

7.4.1 Tables and Sections

The MPEG-2 Systems Standard defines tables that provide information necessary to act on or to further describe the data essence streams within the Transport Stream service multiplex. The logical tables are constructed by using one or more Sections. For example, the Program Map Table (PMT) contains information about what elementary streams are parts of which MPEG-2 programs. The PMT is composed of one or more TS_program_map_section sections. A Table is the aggregation of the Sections that comprise it. A Section is divided as necessary to be packetized

[8] The MPEG-2 Systems Standard defines the discontinuity_indicator as a flag in the adaptation field syntax. Among other uses, it may be set to indicate a discontinuous continuity_counter value. See 13818-1 subclause 2.4.3.5 for details.

into the payload of one or more MPEG-2 Transport Stream packets so that it may be incorporated into the Transport stream service multiplex along with other bit streams.

The MPEG-2 Systems Standard defines several different tables, collectively called Program Specific Information (PSI). Using the private_section, which is the MPEG-2 Systems-defined generic section data structure, the ATSC standards define many other tables.

7.4.2 MPEG-2 Private Section

The term "section" is a generic term referring to any data structure that is based on the MPEG-2 private_section syntax. The MPEG-2 private_section defines a data encapsulation method used to place private data (that is, data that the MPEG-2 standards do not define, including ATSC-defined sections) into an MPEG-2 Transport Stream packet with a minimum amount of structure [13].

A section, or more specifically the MPEG-2 private_section, always begins with an 8-bit table_id, which uniquely identifies the table of which the section is part. Another field, the section_syntax_indicator, determines whether the "short" or "long" form of the private_section syntax is used. The short form section includes a minimal amount of header information and is limited to carrying a payload of at most 4093 bytes. The long form section incorporates additional header fields, which allow the segmentation of large data structures into multiple parts. A collection of long form sections may accommodate 256 * 4084 bytes of payload (maximum size of 1,045,504 bytes).

In practice, most receivers' incorporate hardware section filtering allowing the receiver to specify filtering criteria for the first eight bytes of a section. This length equates to the byte count necessary to filter the long form private_section header. Hardware assisted filtering offloads the processing burden from the host processor and enables the receiver to specify exact section identification syntax for the section it is interested in acquiring.

The long form section header contains a version_number field, which identifies the revision of the contents of the section. Any time the section's payload bytes are modified, the version_number must be incremented so that a receiver will be able to determine that the section's contents have changed.

The long form section contains a CRC_32 field as the first byte following the last payload byte, which is used for error detection purposes. The receiver's 32-bit CRC decoder (the CRC decoder model is described in MPEG-2 Systems, Annex A) calculates the CRC result over all the bytes that comprise a section beginning with the table_id through the last byte of the CRC_32 field itself. A CRC accumulator result of zero indicates that the section was received without error.

One or more sections may be placed into an MPEG-2 Transport Stream packet depending on the section's size. If the section length is smaller than a Transport Stream packet's payload, then there may be multiple sections contained within the single MPEG-2 Transport Stream packet. Sections that are larger than a single MPEG-2 Transport Stream packet are segmented across multiple MPEG-2 Transport Stream packets. Once the process of packetizing a section commences, a new section will not be packetized into Transport Stream packets having the same PID until the previous section's packetization has completed. When a section does not completely fill an MPEG-2 Transport Stream packet's payload area and there is no new section ready to begin filling the remainder of the payload area, the remaining bytes of the MPEG-2 Transport Stream packet are stuffed, or filled, with the value 0xFF. To prevent stuffing byte emulation, the MPEG-2 Systems Standard forbids the use of 0xFF as a table_id value.

7.4.3 MPEG-2 PSI

MPEG-2 Program Specific Information (PSI) provides data necessary to identify an MPEG-2 Program (i.e., the desired service) and to demultiplex (i.e., separate and extract) the Program and its Program Elements from the MPEG-2 single or multi-program Transport Stream service multiplex. The MPEG-2 Systems Standard currently defines five PSI tables: the Program Association Table (PAT), the Program Map Table (PMT), the Conditional Access Table (CAT), the Network Information Table (NIT), and the Transport Stream Description Table (TSDT).

The Program Association Table provides a complete list of all the MPEG-2 Programs (services) within the Transport Stream. The PAT establishes a relationship between each MPEG-2 Program, via the program_number, and its corresponding program map section (properly defined as TS_program_map_section), via the PID value assigned to the corresponding program map section. Transport Stream packets that contain the PAT are assigned to PID 0x0000.

Each program map section contains the mapping between an MPEG-2 Program and the Program Elements. that define the Program (this mapping is called a *program definition*). Specifically, a program definition establishes a mapping (establishing the relationship) between an MPEG-2 Program Number and the list of the PIDs that identify the individual Program Elements comprising the MPEG-2 Program. The PMT is defined as the complete collection of individual Program Definitions within the Transport Stream, with one TS_program_map_section per MPEG-2 Program. The PMT is unique among the PSI tables in that its contents may be carried as part of different bit streams (i.e., within Transport Stream packets that have different PIDs). This simplifies the addition, deletion, or modification of the PSI for individual MPEG-2 programs, as each can be altered independently. This also simplifies the demultiplexing process as only relevant portions of the Transport Stream need to be parsed by the receiver. In comparison, the other PSI tables are each required to be in its own unique bit stream (within Transport Stream packets of a single, unique PID).

However, even though an MPEG-2 Program is announced in a TS_program_map_section, there is no requirement in MPEG-2 that the individual Program Elements are currently present in the Transport Stream. Furthermore, there is no MPEG-2 requirement that all PIDs currently in use are described by any PSI table.

Whenever an MPEG-2 Program's bit stream is scrambled (i.e., the contents are only decodable with the use of a conditional access system process), a CAT must be present in the Transport Stream. The CAT associates aspects of the conditional access system (CA system or CAS), such as access rights sent in entitlement management messages (EMMs), with the scrambled streams. Transport Stream packets which contain the CAT are assigned to PID 0x0001. CA systems provide scrambling of MPEG-2 Programs or individual Program Elements along with end user authorization. While MPEG-2 Programs or individual Program Elements may be scrambled, all of the tables that comprise the PSI are never scrambled. The MPEG standards do not define the contents of the CAT payload. For details of how the ATSC defines CA, see ATSC standard A/70 [7].

The function of the Network Information Table (NIT) is to carry information that applies network-wide (i.e., to all Transport Stream service multiplexes in the delivery/emission network). ATSC standards do not specify the use of the NIT.

The function of the Transport Stream Description Table (TSDT) is to carry descriptors that apply to an entire MPEG-2 Transport Stream service multiplex. A/53 neither constrains nor specifies the use of the TSDT.

7.4.4 MPEG-2 Packetized Elementary Stream (PES) Packet

The MPEG-2 Systems Standard includes a mechanism for efficiently and reliably conveying continuous streams of data (bit streams of compressed audio, compressed video, and/or data) in real-time over a variety of network environments, including terrestrial broadcasting. Each bit stream (Program Element) is segmented into variable-length packets, called Packetized Elementary Stream (PES) packets, which are conveyed in the MPEG-2 Transport Stream and then reassembled at the receiver. MPEG-2 PES packets are used to segment and encapsulate elementary streams such as coded video, coded audio, and private data streams, along with stream synchronization information. Elementary streams are each independently carried in separate PES packets; thus, a PES packet contains data from one and only one elementary stream. A PES packet is further segmented into fixed-length packets, called MPEG-2 Transport Stream packets (see Section 7.3.1). The set of TS packets so created all share a single, common packet identifier (PID).

The MPEG-2 PES packet consists of a PES packet header followed by the PES packet payload. Each PES packet may have a variable length. A length field allows explicitly signaling the size of the PES packet (up to 65,536 Bytes) or, in the case of video elementary streams, the size may be indicated as unbounded by setting the packet length field to zero. When encapsulating data into a PES packet, the elementary stream is first segmented into variable byte-sized segments and these segments are encapsulated using the MPEG-2 PES packet syntax. ATSC Standard A/53 has placed constraints on PES packets that encapsulate video elementary streams: an MPEG-2 PES packet may only contain one coded video frame and must be signaled as being unbounded in size by defining the length field as 0x0000.

MPEG-2 PES packets carry stream synchronization information in the PES packet header using Presentation Time Stamps (PTS) and Decoding Time Stamps (DTS) fields. The timestamps enable decoding the access units and presenting the access units respectively. The PTS and the DTS are each 33-bits long with units in 90 kHz clock periods.

7.4.4.1 MPEG-2 PES Packet Segmentation, Encapsulation, and Packetization

In order to transport an MPEG-2 PES packet, it is first segmented into the payload of one or more MPEG-2 Transport Stream packets (see Section 7.3.1). The first byte of a PES packet must always be the first byte of a Transport Stream packet payload field. When the first byte of a PES packet appears in an MPEG-2 Transport Stream packet, the MPEG-2 Transport Stream packet header's payload_unit_start_indicator flag must be set to '1'. The payload_unit_start_indicator is set to '0' in all subsequent MPEG-2 Transport Stream packets carrying the remaining portion of the PES packet. PES packets are typically much larger than an MPEG-2 Transport Stream packet; however, they can be smaller than an MPEG-2 Transport Stream packet. Only a single PES packet may be packetized into an MPEG-2 Transport Stream packet.

7.4.4.2 Stuffing and the MPEG-2 PES Packet

Since the MPEG-2 Transport Stream is composed of autonomous units of Transport Stream packets, "stuffing" is needed when there is insufficient PES packet data to completely fill a Transport Stream packet payload. "Stuffing" is the process of filling out the remainder of a Transport Stream packet with data bytes that carry no useful information, but only take up the remaining available Transport Stream packet payload bytes. For Transport Stream packets carrying PES packets, stuffing is accomplished by defining an adaptation field longer than the sum of the lengths of the data elements in the adaptation field, so that the payload bytes remaining after the adaptation field exactly accommodate the available PES packet data. This extra space in the adaptation field is filled with stuffing bytes.

7.5 Multiplex Concepts

The MPEG-2 term "program" corresponds to an individual digital TV channel or data service. The MPEG-2 Systems Standard's support for multiple channels or services within a single, multiplexed bit stream (known as a multi-program Transport Stream or service multiplex) enables the deployment of practical, bandwidth-efficient digital broadcasting systems. This approach enables the delivery of services at various bit rates in one defined construct.

The packet identifier (PID), contained in each Transport Stream packet, is the key to sorting out the components or elements in the Transport Stream. The PID is used to reassemble higher level constructs that make up different bit stream elements within the multiplex and can change from Transport Stream packet to packet. This identification mechanism enables the time-based interleaving or multiplexing of services at differing bit rates. For example, video essence typically requires a much higher bit rate than audio essence. A series of Transport Stream packets identified by the same PID contain either a Program Element or descriptive information about one or more Program Elements (a series of Transport Stream packets having the same PID is often referred to as a bit stream).

The MPEG-2 Systems standard has set aside a few special PIDs to directly identify Transport Stream packets that contain constructs that assist in locating the individual MPEG-2 Programs and their associated Program Elements. These constructs are collectively called Program Specific Information (PSI).

A related set of one or more Program Elements is called an MPEG-2 Program. Figure 7.1 illustrates how two MPEG-2 Programs each consisting of a video and audio Program Element (in these cases each Program Element is also an Elementary Stream) might be multiplexed into an MPEG-2 Transport Stream. The Transport Stream packet payload contents are reassembled into a higher level construct (with different packet sizes and structure). For coded audio and video, this higher layer of packetization is called a Packetized Elementary Stream (PES) packet (see Section 7.4.4).

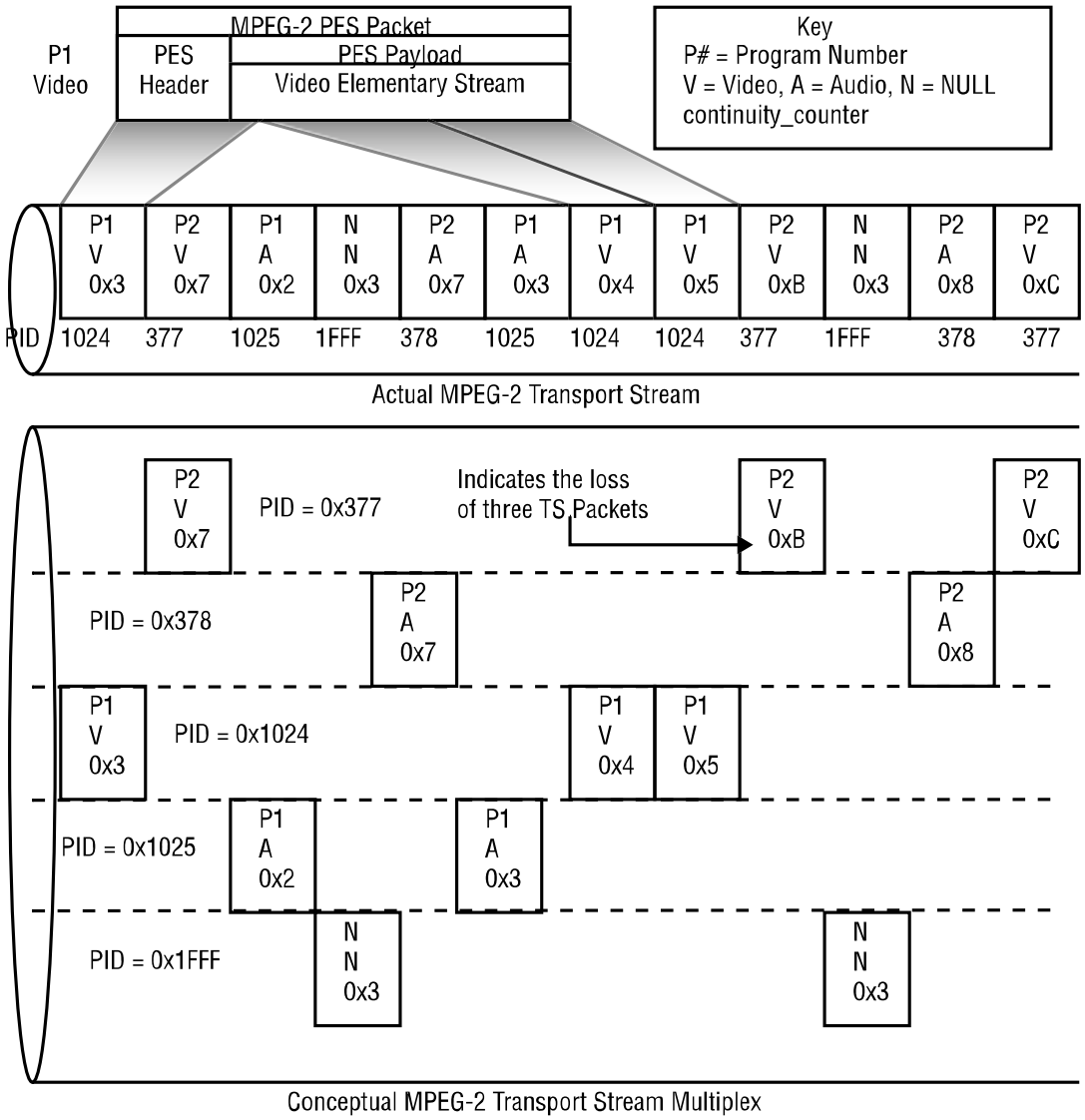

Figure 7.1 MPEG-2 transport stream program multiplex.

In Figure 7.1, Program P1's video stream is illustrated to consist of three MPEG-2 Transport Stream packets identified by PID 0x1024. Each MPEG-2 Transport Stream packet has a continuity_counter associated with the specific PID that enables a receiver to determine if a loss has occurred. In this example, the continuity_counter values begin at 0x3 and end with 0x5 for Program P1's video stream. The individual MPEG-2 Transport Stream packets that contain this PID are extracted from the multiplexed bit stream and reassembled, in this case making up part of an MPEG-2 PES packet carrying a video elementary stream.

Program P1 also has an associated audio stream of packets identified by PID 0x1025. Two MPEG-2 Transport Stream packets from Program P1's audio stream are shown with the continuity_counter values of 0x2 and 0x3 respectively. Similarly, in Figure 7.1, Program P2's packet composition is illustrated. In Program P2's video stream identified by PID 0x0377, the second to last MPEG-2 Transport Stream packet's continuity_counter is 0xB rather than the expected value of 0x9. This condition may indicate an error and the loss of possibly three MPEG-2 Transport Stream packets having this PID. The next expected and received continuity_counter value is 0xC as illustrated in the diagram.

As discussed later, the mechanism for recreating the original System Time Clock (STC) in the decoder uses the actual arrival time of the packets carrying the individual Program Clock References (PCR) as compared to the value carried in the PCR field.

Because of this, MPEG-2 Transport Stream packets with a given PID value cannot casually be rearranged in the MPEG-2 Transport Stream. This limitation exists because shifting the relative location of a Transport Stream packet carrying the PCR introduces jitter into the data stream, which may cause the decoder's System Time Clock (STC) to vary. The temporal location of the individual MPEG-2 Transport Stream packet payload delivery conforms to the buffer model associated with the encapsulation type. Shifting or rearranging the MPEG-2 Transport Stream packets potentially causes buffer model violations by either overflowing or underflowing the buffer, unless such is done without violation of these constraints.

Also notice the Null MPEG-2 Transport Stream packets that were interleaved. These MPEG-2 Transport Stream packets (identified by PID 0x1FFF) may appear anywhere in the stream and are often used to set the Transport Stream service multiplex at a known, fixed overall bit rate, regardless of the total bit rate of all the MPEG-2 programs it contains. For illustrative purposes, a value of 0x03 is shown in the figure for the continuity_counter for the Null packets. In practice any value may be used, as the continuity_counter for Null packets is ignored.

7.6 MPEG-2 Timing and Buffer Model

Key elements of the MPEG-2 Systems Standard include a model for system timing and another for buffering. The timing model allows the synchronization of the components making up MPEG-2 programs. The buffer model ensures interoperability between encoders and decoders for information delivery (i.e., ensuring that the necessary information is always available when needed for decoding).

7.6.1 MPEG-2 System Timing

One of the basic concepts of the MPEG-2 standards revolves around the system timing model. The timing model was developed to enable the synchronization of video and audio Program Elements that are delivered as separate streams, with differing delivery rates and different sized Presentation Units. As will be discussed below, elements that enable the synchronization are clock references, which allow the decoder to recreate a clock that very closely tracks that used in the encoder, and time stamps, which are used to temporally coordinate the presentation of video and audio Presentation Units. This basic timing model is applicable to other forms of Program Elements, including data.

7.6.1.1 Timing Model

The MPEG-2 timing model requires that the clock used to encode the content be regenerated (within specified tolerances) at the receiver and used to decode the content. Video and audio consist of discrete Presentation Units, which must be delivered from the decoder at the same rate as they entered the encoder in order to achieve correct reproduction. For video, the Presentation Unit is a picture (a frame or field of video). For audio, the Presentation Unit is a block of audio samples (also known as an audio frame). The Presentation Unit for data is dependant upon the form of the data, but the basic concept is similar. The output rate at the decoder must match the input rate at the encoder.

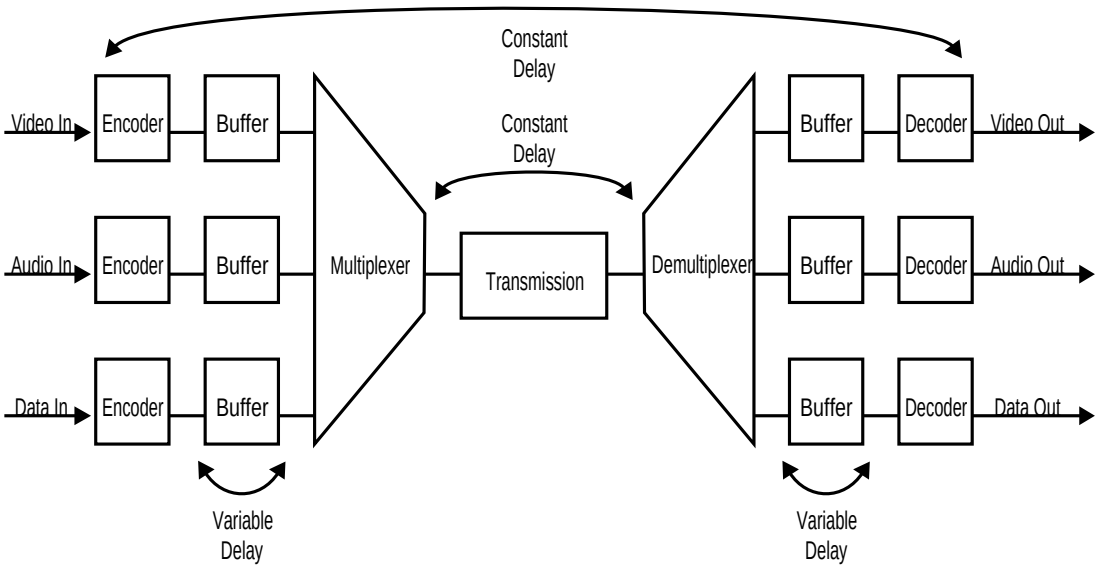

Figure 7.2 MPEG-2 constant delay buffer model.

In developing the timing model, the MPEG-2 Systems Standard adopted two basic concepts: a constant end-to-end delay and an instantaneous decoding process (see Figure 7.2). The MPEG-2 systems standard does not specify how the encoders or decoders operate; rather, it specifies the format of the bit stream (the syntax and semantics) and a theoretical decoding buffer model. With these concepts applied to the bit stream, it is possible to develop implementations of both encoders and decoders that consider real-world constraints and will interoperate. In real systems, the delay through the encoding and decoding buffers is variable [13] and the decoding process takes a finite, non-zero and possibly variable, amount of time.

The MPEG-2 Systems Standard's timing and buffer models solve the issues of synchronization of individual elements by use of a common time reference shared by all individual Program Elements of an MPEG-2 Program. This common time clock is referred to as the System Time Clock (STC).

7.6.1.2 System Time Clock (STC)

The System Time Clock (STC) is the master clock reference for all encoding and decoding processes. Each encoder samples the STC as needed to create timestamps associated with the data's Presentation Units. A timestamp associated with a Presentation Unit is referred to as the Presentation Time Stamp (PTS). A timestamp associated with the decoding start time, known as Decoding Time Stamp (DTS), may also appear in the bit stream.

The STC is not a normative element in the MPEG-2 Systems standard; however, it is required for synchronized services (including video and audio), meaning that all practical implementations require its use. The STC is represented by a 42-bit counter in units of 27 MHz (27 MHz equals approximately 37 ns per clock period).

The STC must be recreated in the decoder in such a way that it very closely matches (within specified tolerances) the STC at the encoder for both buffer management and synchronization reasons. In order for a decoder to reconstruct this clock, the STC is periodically sampled and transmitted in the MPEG-2 Transport Stream packet's adaptation_field, as clock references known as Program Clock References (PCRs). Figure 7.3 illustrates a general decoder circuit used to recreate the STC.

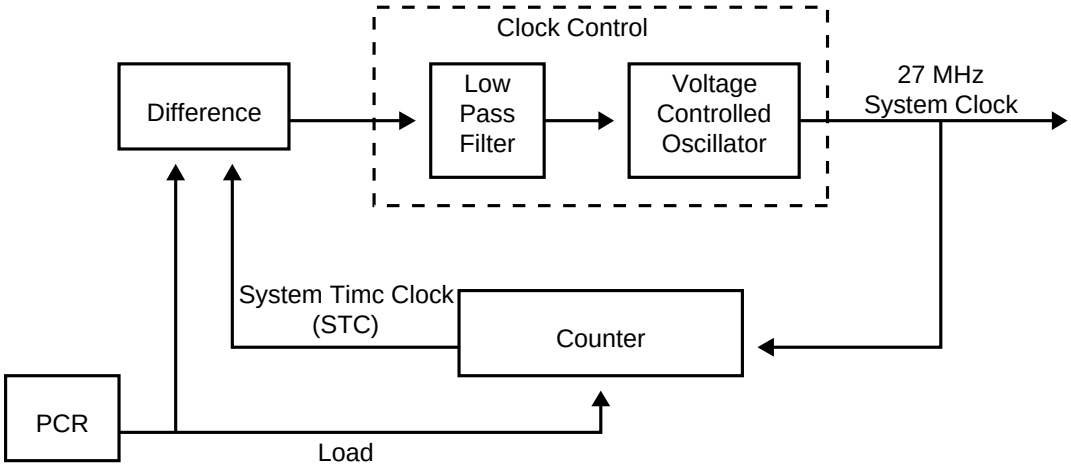

Figure 7.3 MPEG-2 system time clock.

Each MPEG-2 Program may have its own STC or multiple MPEG-2 Programs may share a common STC (by referring to the same Program Element that carries the PCR values). There may be situations where an MPEG-2 Program does not require any form of synchronization and will not need an STC. Also, Program Elements may or may not reference a Program's STC.

The STC increases linearly in time and monotonically. The exception is when there are discontinuities, which are discussed below. Since the STC value is contained within a finite size field, it wraps back to zero when the maximum bit count is achieved, approximately every 26.5 hours.

7.6.1.3 System Clock Frequency

The System Time Clock is derived from the system_clock_frequency specified as 27,000,000 Hz 810 Hz. The STC period is 1/27 MHz or approximately 37 ns per clock period.

7.6.1.4 Program Clock Reference

The Program Clock Reference (PCR) is a 42-bit value used to lock the decoder's 27 MHz clock to the encoder's 27 MHz clock, thereby matching the decoder's STC to the encoder's STC. The PCR is carried in the MPEG-2 Transport Stream packet's adaptation_field using the program_clock_reference_base and the program_clock_reference_extension fields. The MPEG-2 Systems standard mandates that the PCR be sent at least every 100 ms or 10 times a second. The PCR may be sent more frequently if desired. In addition, the standard limits the amount of PCR jitter for a compliant stream to no more than 500 ns.

The decoder uses the arrival time of the MPEG-2 Transport Stream packet carrying a PCR value, and the PCR value itself, in comparison to the current value of the STC to adjust the clock control component. Figure 7.3 illustrates an example of how the PCR is used to exactly recreate the STC.

The program_clock_reference_base is constructed by dividing the value of the 27 MHz clock reference count by 300. This operation creates a 33-bit value in units of 90 kHz clock periods. The program_clock_reference_extension contains the remainder of the previous division (i.e., the 27 MHz clock modulo 300).

The location of the Program Element carrying the PCR for an MPEG-2 Program is signaled in the TS_program_map_section PCR_PID field. The PCR may be carried on the same PID as a video, audio, or data Program Element as the PCR field is independent of the encapsulated data

payload. Different MPEG-2 Programs may share the same STC, by referring to the same PCR_PID.

MPEG-2 Programs not requiring synchronized decoding and presentation to an STC set the PCR_PID field to the value 0x1FFF indicating that there is not a Program Element carrying a PCR.

7.6.1.5 Presentation Time Stamp (PTS)

The Presentation Time Stamp (PTS) is a 33-bit quantity measured in units of 90 kHz clock periods (approximately 11.1 microsecond ticks) carried in the MPEG-2 PES packet header's PTS or DTS fields. The PTS, when compared against the System Time Clock (STC), indicates when the associated Presentation Unit should be "presented" to the viewer. In the case of video, a picture is displayed and in the case of audio the next audio frame is emitted by the receiver. The PTS must be contained in the MPEG-2 Transport Stream at intervals no longer than 700 ms and the ATSC requires that the PTS be inserted at the beginning of every access unit (i.e., coded picture or audio frame).

The PTS, when included, is divided into two fifteen-bit quantities and a 3-bit quantity spread across 36 bits. There are also three "marker bits", always set to '1', interspersed among the three groups. This division into three parts, along with the inclusion of the marker_bits, avoids start_code emulation in the MPEG-2 PES packet header. Avoiding the emulation of the start_code prevents decoders from incorrectly identifying the start of an elementary stream. Figure 7.4 illustrates the PTS and marker_bits.

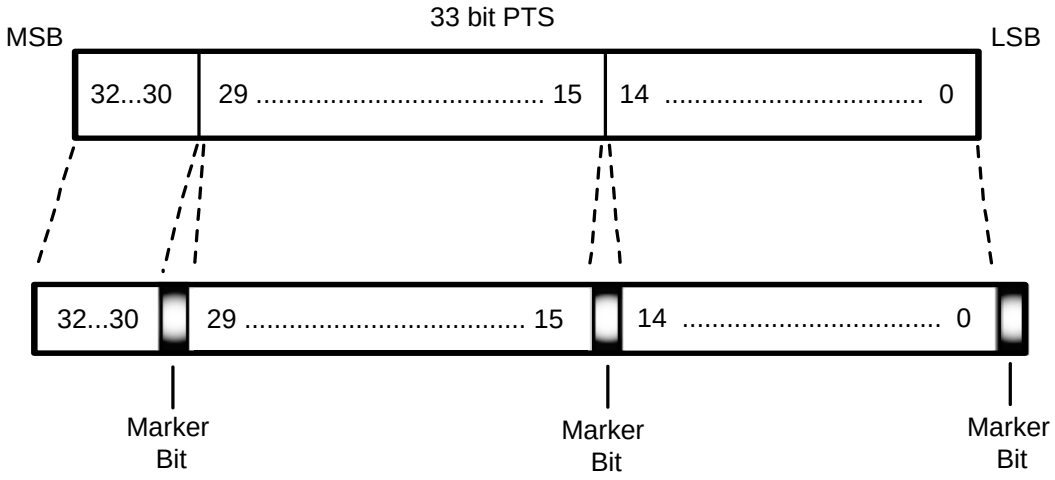

Figure 7.4 The MPEG-2 PTS and marker_bits.

7.6.1.6 Decoding Time Stamp (DTS)

The Decoding Time Stamp (DTS) is a 33-bit quantity, measured in units of 90 kHz clock periods (approximately 11.1 microseconds) that may be carried in the MPEG-2 PES packet header's DTS field. The MPEG-2 Systems Standard only defines a normative meaning for the DTS field for video. Generally speaking, a video stream is the only stream type that may need the DTS due to picture re-ordering (bi-directionally interpolated pictures are decoded after the "future" frame it references has been decoded. The DTS value, compared to the System Time Clock (STC), indicates when the access unit should be removed from the buffer and decoded.

The DTS must always be accompanied by a PTS. If the DTS contains the same value as the PTS, then the DTS is omitted and the decoder assumes that the DTS is equal to the PTS. The DTS, if present, must be contained in the MPEG-2 Transport Stream at intervals no longer than

700 ms. The ATSC mandates that the DTS must be inserted at the beginning of every access unit (i.e., coded picture or audio frame), except when the DTS value matches the PTS value.

The DTS is encoded in the same manner as the PTS—splitting the 33-bit quantity into three portions and incorporating the marker bits.

7.6.1.7 Discontinuities

MPEG-2 Program and MPEG-2 Transport Stream discontinuities are a reality in digital television. Planned discontinuities, where the interruption is not the result of an error, can occur in any number of situations. As an example, the splicing of a commercial into the video and audio streams is a typical planned discontinuity scenario. Other planned discontinuity scenarios include switching between content sources or a new MPEG-2 Program commencing. In each of these cases, the System Time Clock (STC) may be interrupted and set to some new random value from which the count then continues, thus creating a discontinuity in the timeline.

The decoder, in all of the above instances, should be notified of the upcoming interruption by the MPEG-2 PES packet header's discontinuity_indicator. The discontinuity_indicator is used to indicate a discontinuity in the STC or a disruption in the continuity_counter. The signaling of continuity_counter disruptions via the discontinuity_indicator is limited in its practical usefulness. The discontinuity_indicator can be used by the multiplexing process to indicate a known and expected discontinuity in the program time line. MPEG-2 Transport Stream packet loss is not signaled via this indicator bit.

An STC interruption is the result of a receiver receiving a new PCR value associated with the MPEG-2 Program that is out of range of a reasonable variance from the expected value, regardless of whether or not the discontinuity_indicator had been set. Receivers receiving the next PCR after either an explicit discontinuity or a PCR out of range must adjust themselves accordingly. In cases where a discontinuity has been signaled explicitly, a receiver typically will simply use the next PCR value received to reset its internal clock phase circuitry without making any frequency adjustments. In cases where a discontinuity has not be signalled explicitly, a receiver typically will begin a clock-error recovery process. This may include tracking PCR values during a predefined time window to make an "intelligent" determination of what adjustments need to be made to the STC, if any. Besides the STC reference changing, another discontinuity that may be encountered as part of the stream changeover or Program interruption involves the MPEG-2 Transport Stream packet header's continuity_counter value. The continuity_counter may skip to a new value when the newly encoded stream is inserted. Thus, the decoder upon seeing the discontinuity_indicator is made aware of an upcoming continuity_counter change and this change should not be treated as an error or indicative of lost packets.

7.6.2 Buffer Model

The MPEG-2 standards define the bit stream syntax itself and the meaning (or semantics) of the bit stream syntax. In order to ensure interoperability of equipment designed to the specifications, the MPEG-2 Systems Standard also precisely specifies the exact definitions of byte arrival and decoding events—and the times at which these occur--through the use of a hypothetical decoder called the Transport Stream System Target Decoder (T-STD). The T-STD is a conceptual model that is used solely for the purpose of defining terms precisely and to model the decoding process during construction or verification of Transport Streams; neither the architecture of the T-STD nor the timing described is meant to preclude the design of practical solutions that implement the buffer and timing model in other ways. The buffers are therefore "virtual" since they may or may not exist within real physical decoders.

The T-STD model is structured as follows:

- Defines several "virtual buffers"
- Rules for when bytes enter and leave each buffer
- Rules that constrain buffer fullness
- The number of "virtual buffers" in the model varies depending upon the number of streams in the Transport Stream

For video elementary streams, the T-STD consists of three buffers: the transport buffer, the multiplex buffer, and the elementary stream buffer. For audio elementary streams and system-related streams (e.g., PSI tables), the T-STD consists of two buffers, the transport buffer and the main buffer. The rules define when bytes enter and leave the buffers in terms of where they occur within the bit stream and either the transport rate or the specified maximum bit rate for the type of elementary stream, depending on the buffer type and the elementary stream type. The transport rate computation is determined by a mathematical formula based on the program clock reference fields encoded in the bit stream, and the maximum bit rate is determined from the profile and level or similar inherent definition. Buffer sizes are defined by specific mathematical formulas based on the buffer type and the elementary stream type. The decoding time is specified in terms of embedded or inferred decoding or presentation time stamps (DTS or PTS, respectively) and may be delayed due to any re-ordering of pictures that is needed (in the case of video elementary streams only).

Buffer management is needed to ensure that none of the buffers overflow or, in some cases, underflow. Constraints on buffer fullness say whether a particular buffer is allowed to underflow, whether a buffer must empty occasionally, etc. These rules are all clearly defined in the T-STD model.

A suitable analysis tool can verify whether or not a bit stream conforms to the T-STD. It is more difficult to verify that a decoder conforms to the T-STD because true conformance may only be determined by demonstrating that the decoder is capable of decoding all conformant bit streams properly. However, several organizations have prepared special bit streams to stress decoder implementations. These known MPEG-conformant bit streams cause buffers to fill near their capacity, to operate near-empty, and/or to require high-speed transmission between buffers. Receiver implementers are advised to test as many combinations as possible.

7.7 Supplemental Information

The remainder of this section describes additional concepts that are important in the understanding of the ATSC transport subsystem.

7.7.1 MPEG-2 Descriptors

The MPEG-2 descriptor is a generic structure used to carry information within other MPEG-2 data structures, typically sections (PSI or private). The use of standardized descriptors is often optional. Descriptors can be viewed as a mechanism to extend the information conveyed within another MPEG-2 structure.

A descriptor cannot stand alone in the MPEG-2 Transport Stream; rather, it must be contained within a larger syntactic structure, typically within a *descriptor loop* (an area set aside to carry an arbitrary number of descriptors). The basic descriptor format is a tag byte, followed by a length byte followed by data [13]. The tag byte uniquely identifies the descriptor and the length byte specifies the number of data bytes that immediately follow the length field. The form of the data varies for each specific descriptor.

In future versions of a given ATSC standard, additional information may be added to any defined descriptor by simply adding new semantic fields at the end. Receiver designers should always process the length field of all descriptors to ensure that if a receiver finds any information

beyond the known fields, then it discards such information but continues parsing the stream at the first byte beyond that indicated by the length field. Receivers may also encounter descriptors that they do not recognize (such as could be added to a new version of the standard created after the receiver was built). To ensure that newly defined descriptors do not cause operational problems in existing equipment, all descriptors defined will adhere to the existing structure. This provides an inherent escape mechanism to allow receivers that don't understand a particular descriptor to easily skip over it. By jumping the number of bytes listed in the length field, the receiver can proceed to the next item in the loop.

Because a receiver that does not recognize a descriptor of a certain type is expected to simply ignore it, the addition of new features via new descriptor definitions is a powerful way to add new features to the protocol while maintaining backward compatibility.

7.7.1.1 ATSC Descriptors

ATSC-defined descriptors follow the same behavior as described previously for MPEG-2 descriptors and may be used for similar purposes. ATSC standards have also described the usage of some MPEG descriptors. The following descriptors are defined by ATSC Standards A/52 [4] and A/53 [5].

7.7.1.1.1 AC-3 Audio Descriptor

An AC-3 Audio Descriptor [4] describes an audio service present in an ATSC Transport Stream. In addition to describing a possible audio service(s) that a broadcaster might send, this descriptor(s) provides the receiver with audio set up information such as whether the program is in stereo or surround sound. This descriptor is optionally present in the program element loop of the TS_program_map_section that describes the AC-3 audio elementary stream. See A/65B [6] for required placements.

7.7.1.1.2 ATSC Private Information Descriptor

The ATSC Private Information Descriptor [5] provides a method for carrying private information within this descriptor and for unambiguous identification of its registered owner. Since both the identification and the private information are self-contained within a single descriptor, more than one ATSC private information descriptor may appear within a single descriptor loop.

The format identifier field appears in both the ATSC Private Information Descriptor and the MPEG Registration Descriptor. Its purpose is the same in both: it identifies the company or organization that has supplied the associated private data. Only values of the format identifier field registered by the ISO-assigned Registration Authority, the Society of Motion Picture Engineers (SMPTE), may be used.

7.7.1.2 MPEG Descriptors Constrained by ATSC

The following descriptors have been defined in MPEG-2 Systems (13818-1) [13], but their usage has been constrained by A/53 [5].

7.7.1.2.1 Data Stream Alignment Descriptor

The ATSC requires this descriptor to be present in the program element loop of the TS_program_map_section that describes the video elementary stream. In this context, the descriptor specifies the alignment of video stream syntax with respect to the start of the PES packet payload. The ATSC has constrained the alignment to be the first byte of the start code for a video access unit (alignment_type 0x02). Because a video access unit follows immediately after a GOP or Sequence header, this does not preclude alignment from the beginning of a GOP or Sequence. It does, however, prevent alignment from being at the start of a slice.

The use of this descriptor for other stream types is not defined.

7.7.1.2.2 ISO 639 Language Descriptor

In the ATSC digital television system, if the ISO_639_language_descriptor (defined in ISO/IEC 13818-1 Section 2.6.18 [13]) is present then it is used to indicate the language of audio Elementary Stream components.

If present, then the ISO_639_language_descriptor is included in the descriptor loop immediately following the ES_info_length field in the TS_program_map_section for each Elementary Stream of stream_type 0x81 (AC-3 audio). This descriptor will be present when the number of audio Elementary Streams in the TS_program_map_section having the same value of bit stream mode (bsmod in the AC-3 Audio Descriptor) is two or more.

As an example, consider an MPEG-2 program that includes two audio ES components: a Complete Main (CM) audio track (bsmod = 0) and a Visually Impaired (VI) audio track (bsmod = 2). Inclusion of the ISO_639_language_descriptor is optional for this program. If a second CM track were to be added, however, it would then be necessary to include ISO_639_language_descriptors in the TS_program_map_section.

The audio_type field in any ISO_639_language_descriptor used ATSC standards is set to 0x00 (meaning "undefined"). An ISO_639_language_descriptor may be present in the TS_program_map_section in other positions as well, for example to indicate the language or languages of a textual data service program element.

7.7.1.2.3 MPEG-2 Registration Descriptor

Under certain circumstances, the MPEG-2 Registration Descriptor (MRD) is used to provide unambiguous identification of privately defined fields or private data bytes in associated syntactical structures. The detailed rules for the use of the MRD are found in A/53 [5]. Note that no more than one MRD should appear in any given descriptor loop, since the semantics of this situation are unspecified. This usage restriction does not apply to the ATSC private information descriptor discussed previously. The MRD does not contain the private data itself, while the ATSC private information descriptor is designed to carry the actual private data.

7.7.1.2.4 Smoothing Buffer Descriptor

A/53 requires the TS_program_map_section that describes each program to have a Smoothing Buffer Descriptor pertaining to that program [5]. This descriptor signals the required size and leak rate of the smoothing buffer (SBn) to avoid errors in decoding that could be caused by over- or under-flow. During the continuous existence of a program, the value of the elements of the Smoothing Buffer Descriptor are not allowed to change.

7.7.2 Code Point Conflict Avoidance

MPEG standards have numerous syntactical fields set aside for private use. When fields, tags, and table identifier fields are assigned a value by MPEG or by an MPEG user, such as ATSC, the values are then known as "code points."In addition, many other fields have ranges defined as *user private*. The user (not a standards body) may define one or more of these private fields, tags, and table values. Without some type of coordination mechanism, use of ATSC user private fields and ranges may lead to conflicts between privately defined services. Furthermore, without some form of scoping and registration, different organizations may inadvertently choose to use the same values for these fields, but with different meanings for the semantics of the information carried. The ATSC Digital Television Standard A/53 [5] has placed constraints on the use of

private fields and ranges to avoid code point conflicts, through the use of the MPEG-2 Registration Descriptor mechanism.

If an organization uses user private fields and/or ranges, to comply with the ATSC standards, one or more MRDs are used as described in A/53.

7.7.2.1 The MPEG-2 Registration Descriptor

MPEG-2 Systems defines a registration descriptor: "The registration_descriptor provides a method to uniquely and unambiguously identify formats of private data." The Society of Motion Picture and Television Engineers (SMPTE) is the ISO-designated registration authority for the 32 bit format_identifier field carried within this descriptor, guaranteeing that every assigned value will be unique.

The following sections discuss the use of the MRD to avoid collisions. There are some circumstances where an MRD cannot be used for scoping (for example, the MRD has no significance for other descriptors in the same descriptor loop). The ATSC Private Information Descriptor (described previously) has been defined to allow the carriage of private information in a descriptor.

7.7.2.1.1 Private Information in an MPEG-2 Program

The scoping of the use of private structures within an MPEG-2 Program may be done by placing an MRD in the program loop (see Section 7.7.3.3 for an explanation of loops in MPEG-2 syntax) in the PMT (otherwise known as the "outer" loop—the descriptor loop following the program_info_length field). When used in this location, the scope of the MRD is the entire MPEG-2 Program, meaning all of the program elements defined in this instance of the Program Map Table. When the MRD is used to identify the owner of private data, then the identification applies to all program elements comprising the MPEG-2 Program.

7.7.2.1.2 Private Information in an MPEG-2 Program Element

MRDs may be placed in the program element loop in the PMT (otherwise known as the "inner" loop—the descriptor loop following the es_info_length field). When used in this location, the scope of the MRD is the individual program element to which the MRD is bound. When the MRD is used to identify the owner of private data, then the identification applies to the single program element. The scope of the MRD also covers the stream_type used for this program element, in the case that a privately defined stream_type is used.

7.7.2.1.3 Multiple MRDs

At most, one MRD for any entity at any level will appear; in other words, no more than one MRD will appear in the PMT program loop; no more than one MRD will appear in the PMT program element loop for a particular program element. There is no guarantee of how remultiplexing equipment will behave in the presence of multiple MRDs in a single loop, especially in regards to retaining the original ordering of descriptors in the loop. Multiple MRD's at the same level would be ambiguous to a receiver.

MRDs used at different levels are intended to be complimentary, with a deeper level MRD refining the meaning of a higher level MRD. However, certain combinations of MRDs at different levels may result in streams that may cause problems for standard receivers if the combinations are not expected. As an example, a combination of MRDs that identify the program as defined by company X and a particular program element as defined by company Y would lead to contradictions in interpreting semantic elements. The behavior of a receiver upon

receiving a non-conformant stream of this type cannot be specified and construction of streams of this type should be avoided.

7.7.3 Understanding MPEG Syntax Tables

ATSC and MPEG-2 standards use a common convention for specifying how to construct the data structures defined in the standards. This convention consists of a table specifying the syntax (the in-order concatenation of the fields), following by a section specifying the semantics (the detailed definitions of the syntax fields). The syntax is specified using C-language "like" ("C-like") constructs, meaning statements that take the form of the computer language, but would not necessarily be expected to produce reasonable results if run through a compiler.

The tables typically have three columns, as shown in the fragment in Table 7.1:

- Syntax: The name of the field or a "C-like" construct
- No. of Bits: The size of the field in bits.
- Format: Either how to order the bits in the field (an acronym or mnemonic is used, which is defined earlier in the standard—in this example, uimsbf means unsigned integer, most significant bit first)—or, when the field has a pre-defined value, the value itself (typically in either binary or hex notation).

Table 7.1 Table Format

Syntax	No. of Bits	Format
typical_PSI_table() {		
table_id	8	uimsbf
section_syntax_indicator	1	'1'
….	…	…
}		

It should be noted that when the data structures are constructed, the fields are concatenated using big-endian byte-ordering. This means that for a multi-byte field, the most significant byte is encountered first. A common practice for implementation is to step through the syntax structure and copy the values for the fields to a memory buffer. The end result of this type of operation may vary for multi-byte fields, depending upon the computer architecture. Implementors are cautioned when working with little-endian machines (least significant byte encountered first). It is therefore recommended that implementations use byte-oriented instructions to construct the data (i.e., mask and shift operations).

7.7.3.1 Formatting

The curly-bracket characters ('{' and '}') are used to group a series of fields together. In the sample shown in Table 7.1, the curly-bracket characters are used to indicate that all of the fields between the paired curly-brackets belong to the "typical_PSI_table()". For the conditional and loop statements that follow below, curly-bracket pairs are used to indicate the fields affected by either the conditional or loop statements.

The syntax column uses indentation as an aid to the reader (in a similar fashion to a common convention when writing C-code). When a series of fields is grouped, then the convention is to indent them.

7.7.3.2 Conditional Statements

In many of the constructs, a series of fields is included only if certain conditions are met. This situation is indicated using an "if (condition) { }" statement as shown in Table 7.2*a*. When this type of statement is encountered, the fields grouped by brackets are included only if the condition is true.

Table 7.2a IF Statement

Syntax	No. of Bits	Format
typical_PSI_table() {		
field_1	8	uimsbf
if (condition) {		
field_2	8	uimsbf
field_3	8	uimsbf
}		
....	…	…
}		

As with C-code, an alternate path may be indicated by an "else" statement, as illustrated in Table 7.2*b*:

Table 7.2b IF Statement

Syntax	No. of Bits	Format
typical_PSI_table() {		
field_1	8	uimsbf
if (condition) {		
field_2	8	uimsbf
} else {		
field_3	8	uimsbf
}		
....	…	…
}		

If the condition is true, then field_2 is used; otherwise, field_3 is used.

7.7.3.3 Loop Statements

For-loop statements are commonly used in the syntax tables and have the widest variation in style and interpretation[9]. The for-loop takes the following form:

```
for ( i=0; i<N; i++ ) {
…fields
}
```

This type of statement indicates that the fields between curly-brackets should be included a number of times, but can't necessarily be interpreted the way a C compiler would, due to variations in the meaning of the end-point (N in the example above) and nesting of for-loops with

[9] The for-loop statement represents the biggest divergence from actual C-code usage.

re-use of the counter variables. The syntax tables always provide enough information to understand the meaning of the for-loop. Unfortunately, in many cases, common-sense and insight must be used. The following example fragments, taken from the syntax tables in A/65B [6], illustrate how to interpret the for-loop for different types of usage:

Table 7.3 For-Loop Example 1

Syntax	No. of Bits	Format
…		
field_1	8	uimsbf
private_data_length	8	uimsbf
for (I=0; I<private_data_length; I++) {		
private_data_bytes	8	uimsbf
}		
….	…	…

In the example shown in Table 7.3, the interpretation is quite straightforward: the end-point variable, private_data_length, which is given a value in the field immediately above the for-loop. It simply indicates the number of private_data_bytes that follow.

Table 7.4 For-Loop Example 2

Syntax	No. of Bits	Format
…	8	uimsbf
num_channels_in_section	8	uimsbf
for(i=0; i<num_channels_in_section; i++) {		
field_1	8	uimsbf
…		
descriptors_length	10	uimsbf
for (i=0;i<N;i++) {		
descriptor()		
}		
}		
…	6	'111111'

Upon examination of Table 7.4, one quickly notices that there are two nested for-loops, both using the same counter variable (i). As opposed to the interpretation in a real C-program where a change in the value held by the variable in the inner loop will be reflected in the outer, these counter variables do not affect each other. In practice, the actual variable name chosen should be ignored; one should simply view the for-loop construct as a loop that should be traversed some number of times.

Furthermore, the inner and outer loops of this example represent different ways of interpreting how many times to traverse the loop. The outer loop represents a fairly conventional interpretation—the loop is to be traversed "num_channels_in_section" times; each traversal includes all of the fields between this level of paired curly-brackets. As in the example in Table 7.3, the value for "num_channels_in_section" is set by the field preceding the loop.

The inner loop has a different interpretation. In this case, contents of the loop are descriptors. Different descriptors have different lengths, but as shown in Table 7.5, the second byte of each descriptor always specifies the length (in bytes) of the remaining descriptor. The end-point variable "N" is not explicitly set, but may be inferred from a knowledge of how descriptors are constructed. For the inner loop example, the "descriptors_length" field specifies how many bytes make up the fields included in all traversals of this particular loop. In practice, one would follow the steps listed below when parsing this portion of the syntax:

1. Read descriptors_length field
2. Read the descriptor_tag and descriptor_length fields that follow
3. If the descriptor_tag is understood, interpret the following descriptor_length bytes according to the syntax of the particular descriptor, otherwise skip over these bytes
4. Increment the number of bytes traversed by the descriptor_length field + 2 bytes (to include the descriptor_tag and descriptor_length fields)
5. If the number of bytes traversed is less than the value in the descriptors_length field, go to step 2; otherwise, the loop has been fully traversed.

Table 7.5 General Descriptor Format

Syntax	No. of Bits	Format
descriptor () {		
descriptor_tag	8	uimsbf
descriptor_length	8	uimsbf
fields		
…		
}		

Upon examining Table 7.6 one encounters a noteworthy usage of the for-loop. In this case, the end-point variable is listed simply as N, with no indication of what N might be[10]. In some cases, the for-loop is immediately preceded by a length field. For this type of situation, N would take the value of the length field and be interpreted as in one of the cases above.

Table 7.6 For-Loop Example 3

Syntax	No. of Bits	Format
system_time_table_section () {		
table_id	8	0xCD
…		
section_length	12	uimsbf
…		
daylight_savings	16	uimsbf
for (i= 0;i< N;i++) {		
descriptor()		
}		
CRC_32	32	rpchof
}		

[10] Note: This type of usage is more common in the MPEG-2 standards than in the ATSC standards.

The particular case illustrated in Table 7.6 (taken from the System Time Table) requires a little more insight to understand. The field immediately above the for-loop provides no information as to what value to use for N. If one examines the entire table, one finds that the only portion of the syntax with variable length is the for-loop. In addition, one of the earlier fields in the table is the section_length field, which specifies the size of the overall table. These two observations provide enough information to allow the calculation of how many bytes would be included in the for-loop carrying the descriptors.

7.7.3.4 Length Fields

Many of the MPEG-2 syntactical structures include length fields, which indicate the number of bytes remaining in the structure; for example, the general descriptor illustrated in Table 7.5. Some of these structures have an extension mechanism, where non-standard information may be placed at the end of the defined syntax, with the length field increased to account for the extra information. Of course, with this form of extension, there is no expectation that a generic receiver will be able to understand the extra information.

For the above reason, the length field should always be parsed and the information used to determine the offset to the next structure. A common implementation mistake is to assume that the bit stream being parsed contains only standardized usage and only account for the fields defined in the appropriate standard—especially for descriptors. Not correctly accounting for the length information could result in trying to interpret the remainder of the bit stream incorrectly.

8. RF TRANSMISSION

8.1 System Overview

The VSB system offers two modes: a terrestrial broadcast mode (8-VSB) and a high data rate mode (16-VSB) intended for cable applications. Both modes provide a pilot, segment syncs, and a training sequence (as part of data field sync) for acquisition and operation. The two system modes can use the same carrier recovery, demodulation, sync recovery, and clock recovery circuits. Adaptive equalization for the two modes can use the same equalizer structure with some differences in the decision feedback and adaptation of the filter coefficients. Furthermore, both modes use the same Reed-Solomon (RS) code and circuitry for forward error correction (FEC). The terrestrial broadcast mode is optimized for maximum service area and provides a data payload of 19.4 Mbps in a 6 MHz channel. The high data rate mode, which provides twice the data rate at the cost of reduced robustness for channel degradations such as noise and multipath, provides a data payload of 38.8 Mbps in a single 6 MHz channel.

In order to maximize service area, the terrestrial broadcast mode incorporates trellis coding, with added precoding that allows the data to be decoded after passing through a receiver comb filter, used selectively to suppress analog co-channel interference. The high-data-rate mode is designed to work in a cable environment, which is less severe than that of the terrestrial system. It is transmitted in the form of more data levels (bits/symbol). No trellis coding or precoding for an analog broadcast interference rejection (comb) filter is employed in this mode.

VSB transmission with a raised-cosine roll-off at both the lower edge (pilot carrier side) and upper edge (Nyquist slope at 5.38 MHz above carrier) permits equalizing just the in-phase (I) channel signal with a sampling rate as low as the symbol (baud) rate. The raised-cosine shape is obtained from concatenating a root-raised cosine in the transmitter with the same shape in the receiver. Although energy in the vestigial sideband and in the upper band edge extends beyond the Nyquist limit frequencies, the demodulation and sampling process aliases this energy into the baseband to suppress intersymbol interference (ISI) and thereby avoid distortion. With the carrier

frequency located at the –6dB point on the carrier-side raised-cosine roll-off, energy in the vestigial sideband folds around zero frequency during demodulation to make the baseband DTV signal exhibit a flat amplitude response at lower frequencies, thereby suppressing low-frequency ISI. Then, during digitization by synchronous sampling of the demodulated I signal, the Nyquist slope through 5.38 MHz suppresses the remnant higher-frequency ISI. With ISI due to aliasing thus eliminated, equalization of linear distortions can be done using a single A/D converter sampling at the symbol rate of 10.76 Msamples/s and a real-only (not complex) equalizer also operating at the symbol rate. In this simple case, equalization of the signal beyond the –6dB points in the raised-cosine roll-offs at channel edges is dependent on the in-band equalization and cannot be set independently. A complex equalizer does not have this limitation, nor does a fractional equalizer sampling at a rate sufficiently above symbol rate.

The 8-VSB signal is designed to minimize interference and RF channel allocation problems. The VSB signal is designed to minimize peak-energy-to-average-energy ratio, thereby minimizing interference into other signals, especially adjacent and taboo channels. To counter the man-made noise that often accompanies over-the-air broadcast signals, the VSB system includes an interleaver that allows correction of an isolated single burst of noise up to 190 microseconds in length by the (207,187) RS FEC circuitry, which locates as well as corrects up to 10 byte errors per data segment. This was done to allow VHF channels, which are often substantially affected by man-made noise, to be used for DTV broadcasting. If soft-decision techniques are used in the trellis decoder preceding the RS circuitry, the location of errors can be flagged, and twice as many byte errors per data segment can be corrected, allowing correction of an isolated burst of up to 380 microseconds in length.

The parameters for the two VSB transmission modes are shown in Table 8.1.

Table 8.1 Parameters for VSB Transmission Modes

Parameter	Terrestrial Mode	High Data Rate Mode
Channel bandwidth	6 MHz	6 MHz
Guard bandwidth	11.5 percent	11.5 percent
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	19.39 Mbps	38.78 Mbps
Analog co-channel rejection	Analog rejection filter in receiver	N/A
Pilot power contribution	0.3 dB	0.3 dB
C/N threshold	~ 14.9 dB	~ 28.3 dB

8.2 Bit Rate Delivered to a Transport Decoder by the Transmission Subsystem

All data in the ATSC system is transported in MPEG-2 transport packets. The useful data rate is the amount of MPEG-2 transport data carried end-to-end including MPEG-2 packet headers and sync bytes. The exact symbol rate of the transmission subsystem is given by

$$\frac{4.5}{286} \times 684 = 10.7 \text{ ... million symbols/second (megabaud)} \quad (8.1)$$

The symbol rate must be locked in frequency to the transport rate.

The numbers in the formula for the ATSC symbol rate in 6 MHz systems are related to NTSC scanning and color frequencies. Because of this relationship, the symbol clock can be used as a basis for generating an NTSC color subcarrier for analog output from a set top box. The repetition rates of data segments and data frames are deliberately chosen not to have an integer relationship to NTSC or PAL scanning rates, to insure that there will be no discernible pattern in co-channel interference.

The particular numbers used are:

- 4.5 MHz = the center frequency of the audio carrier offset in NTSC. This number was traditionally used in NTSC literature to derive the color subcarrier frequency and scanning rates. In modern equipment, this may start with a precision 10 MHz reference, which is then multiplied by 9/20.
- 4.5 MHz/286 = the horizontal scan rate of NTSC, 15734.2657+...Hz (note that the color subcarrier is 455/2 times this, or 3579545 +5/11 Hz).
- 684: this multiplier gives a symbol rate for an efficient use of bandwidth in 6 MHz. It requires a filter with Nyquist roll-off that is a fairly sharp cutoff (11 percent excess bandwidth), which is still realizable with a reasonable surface acoustic wave (SAW) filter or digital filter.

In the terrestrial broadcast mode, channel symbols carry three bits/symbol of trellis-coded data. The trellis code rate is 2/3, providing 2 bits/symbol of gross payload. Therefore the gross payload is

$$10.76 \times 2 = 21.52 \text{ ... Mbps (megabits/second)} \quad (8.2)$$

To find the net payload delivered to a decoder it is necessary to adjust (8.2) for the overhead of the Data Segment Sync, Data Field Sync, and Reed-Solomon FEC.

To get the net bit rate for an MPEG-2 stream carried by the system (and supplied to an MPEG transport decoder), it is first noted that the MPEG sync bytes are removed from the data stream input to the 8-VSB transmitter and replaced with segment sync, and later reconstituted at the receiver. For throughput of MPEG packets (the only allowed transport mechanism) segment sync is simply equivalent to transmitting the MPEG sync byte, and does not reduce the net data rate. The net bit rate of an MPEG-2 stream carried by the system and delivered to the transport decoder is accordingly reduced by the data field sync (one segment of every 313) and the Reed-Solomon coding (20 bytes of every 208):

$$21.52 \text{ Mbps} \times \frac{312}{313} \times \frac{188}{208} = 19.39 \text{ ... Mbps} \quad (8.3)$$

The net bit rate supplied to the transport decoder for the high data rate mode is

$$19.39 \text{ Mbps} \times 2 = 38.78 \text{ ... Mbps} \quad (8.4)$$

8.3 Performance Characteristics of Terrestrial Broadcast Mode

The terrestrial 8-VSB system can operate in a signal-to-additive-white-Gaussian-noise (S/N) environment of 14.9 dB. The 8-VSB segment error probability curve including 4-state trellis decoding and (207,187) Reed-Solomon decoding in Figure 8.1 shows a segment error probability of 1.93×10^{-4}. This is equivalent to 2.5 segment errors/second, which was established by

measurement as the TOV (threshold of visibility) of errors in the prototype equipment. Particular product designs may achieve somewhat better performance for subjective TOV by means of error masking.

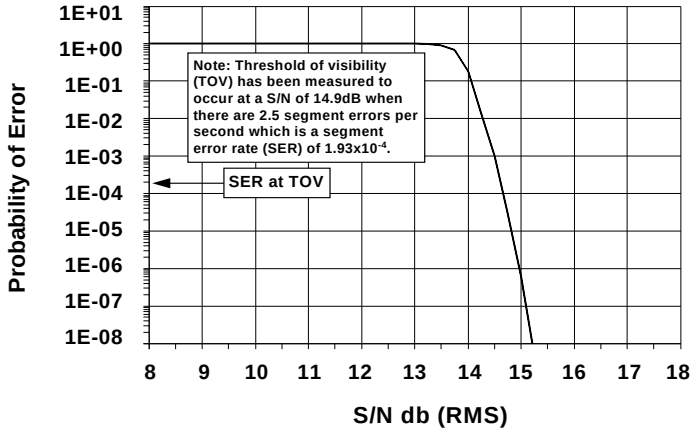

Figure 8.1 Segment error probability, 8-VSB with 4 state trellis decoding, RS (207,187).

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 8.2. The plot shows that 99.9 percent of the time the transient peak power is within 6.3 dB of the average power.

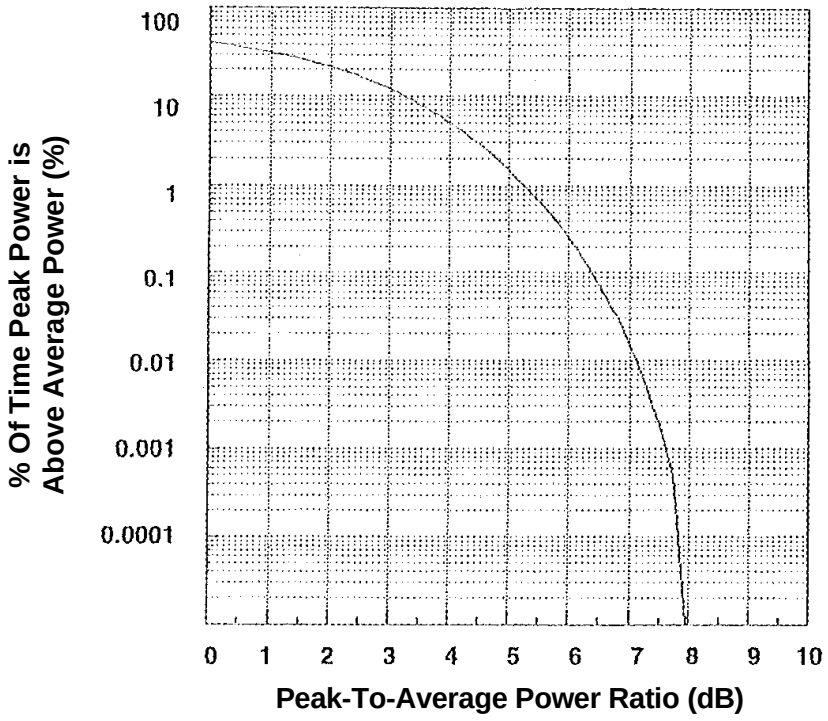

Figure 8.2 Cumulative distribution function of 8-VSB peak-to-average power ratio (in ideal linear system).

8.4 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 significant roll off at band edges. Pre-equalization is typically required in order to compensate the high-order filter used to meet a stringent out-of-band emission mask, such as the U.S. FCC required mask[11]. This linear distortion can be measured by an equalizer in a reference demodulator ("ideal" receiver) employed at the transmitter site. A directional coupler, which is recommended to be located at the sending end of the antenna feed transmission line, supplies the reference demodulator a small sample of the antenna signal feed. The equalizer tap weights of the reference demodulator are transferred to the transmitter pre-equalizer for pre-correction of transmitter linear distortion. This is a one-time procedure of measurement and transmitter pre-equalizer adjustment. Alternatively, the transmitter pre-equalizer can be made continuously adaptive. In this arrangement, the reference demodulator is provided with a fixed-coefficient equalizer compensating for its own deficiencies in ideal response.

A pre-equalizer suitable for many applications 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 microsecond correction range. The pre-equalizer operates on the I channel data signal (there is no Q channel data signal 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

[11] FCC Memorandum Opinion and Order on Reconsideration of the Sixth Report and Order, February 17, 1998.

spectrum of the transmitted signal. If desired, complex equalizers or fractional equalizers (with closer-spaced taps) can provide independent control of the outer portions of the spectrum (beyond the Nyquist slopes).

The transmitter vestigial sideband filtering is sometimes implemented by sideband cancellation, using the phasing method. In this method, the baseband data signal is supplied to digital filtering that generates in-phase and quadrature-phase digital modulation signals for application to respective D/A converters. This filtering process provides the root raised cosine Nyquist filtering and provides compensation for the (sin x)/x frequency responses of the D/A converters, as well. The baseband signals are converted to analog form. The in-phase signal modulates the amplitude of the IF carrier at zero degrees phase, while the quadrature signal modulates a 90-degree shifted version of the carrier. The amplitude-modulated quadrature IF carriers are added to create the vestigial sideband IF signal, canceling the unwanted sideband and increasing the desired sideband by 6 dB. The nominal frequency of the IF carrier (and small in-phase pilot) in the prototype hardware used in ACATS testing was 46.69 MHz, which is equal to the IF center frequency (44.000 MHz) plus the symbol rate divided by 4

$$\frac{10.762}{4} = 2.6905 \text{ MHz}$$

See also the discussion on frequency offsets in Section 8.5.

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 that may include the prefilter of the previous section.

8.5 Upconverter and RF Carrier Frequency Offsets

Modern analog 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 digital 8-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 analog transmitter (at the same frequency), the average power of the DTV signal is on the order of 12 dB less than the analog peak sync power (when operating on the same frequency).

The nominal frequency of the RF upconverter oscillator in DTV terrestrial broadcasts will typically be the same as that used for analog transmitters, (except for offsets required in particular situations).

Note that all examples in this section relate to a 6MHz DTV system. Values may be modified easily for other channel widths.

8.5.1 Nominal DTV Pilot Carrier Frequency

The nominal DTV pilot carrier frequency is determined by fitting the DTV spectrum symmetrically into the RF channel. This is obtained by taking the bandwidth of the DTV signal—5,381.1189 kHz (the Nyquist frequency difference or one-half the symbol clock frequency of 10,762.2378 kHz)—and centering it in the 6 MHz TV channel. Subtracting 5,381.1189 kHz from 6,000 kHz leaves 618.881119 kHz. Half of that is 309.440559 kHz, precisely the standard pilot offset above the lower channel edge. For example, on channel 45 (656–662 MHz), the nominal pilot frequency is 656.309440559 MHz.

8.5.2 Requirements for Offsets

There are two categories of requirements for pilot frequency offsets:

1) Offsets to protect lower adjacent channel analog broadcasts, mandated by FCC rules in the United States, and which override other offset considerations.
2) Recommended offsets for other considerations such as co-channel interference between DTV stations or between DTV and analog stations.

8.5.3 Upper DTV Channel into Lower Analog Channel

This is the overriding case mandated by the FCC rules in the United States—precision offset with a lower adjacent analog station, full service or Low Power Television (LPTV).

The FCC Rules, Section 73.622(g)(1), states that:

> "DTV stations operating on a channel allotment designated with a "c" in paragraph (b) of this section must maintain the pilot carrier frequency of the DTV signal 5.082138 MHz above the visual carrier frequency of any analog TV broadcast station that operates on the lower adjacent channel and is located within 88 kilometers. This frequency difference must be maintained within a tolerance of ±3 Hz."

This precise offset is necessary to reduce the color beat and high-frequency luminance beat created by the DTV pilot carrier in some receivers tuned to the lower adjacent analog channel. The tight tolerance assures that the beat will be visually cancelled, since it will be out of phase on successive video frames.

Note that the frequency is expressed with respect to the lower adjacent analog video carrier, rather than the nominal channel edge. This is because the beat frequency depends on this relationship, and therefore the DTV pilot frequency must track any offsets in the analog video carrier frequency. The offset in the FCC rules is related to the particular horizontal scanning rate of NTSC, and can easily be modified for PAL. The offset O_f was obtained from

$$O_f = 455 \times \left(\frac{F_h}{2}\right) + 191 \times \left(\frac{F_h}{2}\right) - 29.97 = 5,082,138 \text{ Hz} \quad (8.5)$$

Where F_h = NTSC horizontal scanning frequency = 15,734.264 Hz.

The equation indicates that the offset with respect to the lower adjacent chroma is an odd multiple (191) of one-half the line rate to eliminate the color beat. However, this choice leaves the possibility of a luma beat. The offset is additionally adjusted by one-half the analog field rate to eliminate the luma beat. While satisfying the exact adjacent channel criteria, this offset is also as close as possible to optimal comb filtering of the analog co-channel in the digital receiver. Note additionally that offsets are to higher frequencies rather than lower, to avoid any possibility of encroaching on the lower adjacent sound. (It also reduces the likelihood of the automatic fine tuning (AFT) in the analog receiver experiencing lock-out because the signal energy including the pilot is moved further from the analog receiver bandpass.)

As an example, if a channel 44 NTSC station is operating with a zero offset, the Channel 45 DTV pilot carrier frequency must be 651.250000 MHz plus 5.082138 MHz or 656.332138 MHz; that is, 22.697 kHz above the nominal frequency. If the lower adjacent NTSC channel is offset ±10 kHz, the DTV frequency will have to be adjusted accordingly.

Note that full power stations are required to cooperate with lower adjacent analog LPTV stations within 32 km of the DTV station to maintain this offset:

The FCC Rules, Section 73.622(g)(2), states that:

> "Unless it conflicts with operation complying with paragraph (g)(1) of this section, where a low power television station or TV translator station is operating on the lower adjacent channel within 32 km of the DTV station and notifies the DTV station that it intends to minimize interference by precisely maintaining its carrier frequencies, the DTV station shall cooperate in locking its carrier frequency to a common reference frequency and shall be responsible for any costs relating to its own transmission system in complying with this provision."

8.5.4 Other Offset Cases

The FCC rules do not consider other interference cases where offsets help. The offset for protecting lower-adjacent analog signals takes precedence. If that offset is not required, other offsets can minimize interference to co-channel analog or DTV signals.

8.5.4.1 Co-Channel DTV into Analog

In co-channel cases, DTV interference into analog TV appears noise-like. The pilot carrier is low on the Nyquist slope of the IF filter in the analog receiver, so no discernable beat is generated. In this case, offsets to protect the analog channel are not required. Offsets are useful, however to reduce co-channel interference from analog TV into DTV. The performance of the analog rejection filter and clock recovery in the DTV receiver will be improved if the DTV carrier is 911.944 kHz below the NTSC visual carrier. In other words, in the case of a 6 MHz NTSC system, if the analog TV station is not offset, the DTV pilot carrier frequency will be 338.0556 kHz above the lower channel edge instead of the nominal 309.44056 kHz. As before, if the NTSC station is operating with a ±10 kHz offset, the DTV frequency will have to be adjusted in the same direction. The formula for calculating this offset is

$$F_{pilot} = F_{vis(n)} - 70.5 \times F_{seg} = 338.0556 \text{ Hz (for no NTSC analog offset)} \quad (8.6)$$

Where:

F_{pilot} = DTV pilot frequency above lower channel edge

$F_{vis(n)}$ = NTSC visual carrier frequency above lower channel edge

= 1,250 kHz for no NTSC offset (as shown)

= 1,240 kHz for minus offset

= 1,260 kHz for plus offset

F_{seg} = ATSC data segment rate; = symbol clock frequency / 832= 12,935.381971 Hz

The factor of 70.5 is chosen to provide the best overall comb filtering of analog color TV co-channel interference. The use of a value equal to an integer +0.5 results in co-channel analog TV interference being out-of-phase on successive data segment syncs.

Note that in this case the frequency tolerance is plus or minus one kHz. More precision is not required. Also note that a different data segment rate would be used for calculating offsets for 7 or 8 MHz systems.

8.5.4.2 Co-channel DTV into DTV

If two DTV stations share the same channel, interference between the two stations can be reduced if the pilot is offset by one and a half times the data segment rate. This ensures that the frame and segment syncs of the interfering signal will each alternate polarity and be averaged out in the receiver tuned to the desired signal.

The formula for this offset is

$$F_{offset} = 1.5 \times F_{seg} = 19.4031 \text{ kHz} \quad (8.7)$$

Where:
F_{offset} = offset to be added to one of the two DTV carriers
F_{seg} = 12,935.381971 Hz (as defined previously)

This results in a pilot carrier 328.84363 kHz above the lower band edge, provided neither DTV station has any other offset.

Use of the factor 1.5 results in the best co-channel rejection, as determined experimentally with the prototype equipment. The use of an integer +0.5 results in co-channel interference alternating phase on successive segment syncs.

8.5.5 Summary: DTV Frequency

8.5.5.1 Table of DTV Pilot Carrier Frequencies

Table 8.2 summarizes the various pilot carrier offsets for different interference situations in a 6 MHz system (NTSC environment). Note that if more than two stations are involved the number of potential frequencies will increase. For example, if one DTV station operates at an offset because of a lower-adjacent-channel NTSC station, a co-channel DTV station may have to adjust its frequency to maintain a 19.403 kHz pilot offset. If the NTSC analog station operates at an offset of plus or minus 10 kHz, both DTV stations should compensate for that. Cooperation between stations will be essential in order to reduce interference.

Table 8.2 DTV Pilot Carrier Frequencies for Two Stations
(Normal offset above lower channel edge: 309.440559 kHz)

Channel Relationship	DTV Pilot Carrier Frequency Above Lower Channel Edge			
	NTSC Station Zero Offset	NTSC Station + 10 kHz Offset	NTSC Station – 10 kHz Offset	DTV Station No Offset
DTV with lower adjacent NTSC	332.138 kHz ± 3 Hz	342.138 kHz ± 3 Hz	322.138 kHz ± 3 Hz	
DTV co-channel with NTSC	338.056 kHz ± 1 kHz	348.056 kHz ± 1 kHz	328.056 kHz ± 1 kHz	
DTV co-channel with DTV	+ 19.403 kHz above DTV	+ 19.403 kHz above DTV	+ 19.403 kHz above DTV	328.8436 kHz ± 10 Hz

8.5.6 Frequency Tolerances

The tightest specification is for a DTV station with a lower adjacent NTSC analog station. If both NTSC and DTV stations are at the same location, they may simply be locked to the same reference. The co-located DTV station carrier should be 5.082138 MHz above the NTSC visual carrier (22.697 kHz above the normal pilot frequency). The co-channel DTV station should set its carrier 19.403 kHz above the co-located DTV carrier.

If there is interference with a co-channel DTV station, the analog station is expected to be stable within 10 Hz of its assigned frequency.

While it is possible to lock the frequency of the DTV station to the relevant NTSC station, this may not be the best option if the two stations are not at the same location. It will likely be easier to maintain the frequency of each station within the necessary tolerances. Where co-channel interference is a problem, that will be the only option.

In cases where no type of interference is expected, a pilot carrier-frequency tolerance of ±1 kHz is acceptable, but in all cases, good practice is to use a tighter tolerance if practicable.

8.5.7 Hardware Options for Tight Frequency Control

Oscillator frequency tolerance is generally expressed as a fraction—in parts per billion, for example. Given a worst case example of a UHF station (NTSC or DTV) on channel 69, the frequency stability required to meet the ±3 Hz tolerance will be 3.7×10^{-9} or 3.7 parts per billion. If two stations are involved and the oscillators are not locked to a common reference, each must maintain its frequency within half that range, or 1.8×10^{-9}.

A high-quality OCXO (oven-controlled crystal oscillator) can achieve stability of 5×10^{-11} per day or 3×10^{-10} per year (30 parts per billion) at oscillator frequencies up to 30 MHz. This is good enough for the short term, but would have to be monitored and adjusted throughout the year. Long term it would not meet the ±10 Hz or 12.5×10^{-9} requirement for co-channel DTV stations.

A rubidium standard is much better and will meet the worst-case requirements. Commercial equipment may have stability of better than 5×10^{-10} (0.5 parts per billion) per year and/or 1×10^{-10} per month with a life of 15 years.

The GPS (Global Positioning System) satellites provide a more stable reference. Commercially available GPS time-reference receivers provide a signal with an accuracy of 1×10^{-12} (0.001 parts per billion). A disciplined clock oscillator with a long time constant is required to eliminate short-term fluctuations in the GPS reference. In the event the GPS signal is lost, they can maintain an accuracy of better than 1×10^{-10} per day (0.1 part per billion) for short periods of time until the signal returns, well within the requirements.

8.5.8 Additional Considerations

Stations may have to modify their NTSC exciters for greater stability. A rubidium or GPS standard combined with a direct digital synthesizer may be the easiest way to do this.

To minimize interference in the case of a DTV transmitter on an upper adjacent channel to an analog transmitter, the chroma subcarrier frequency of the analog signal must be precisely controlled, in addition to the DTV pilot frequency and the analog carrier frequency.

Phase noise specifications should be checked when considering a frequency source. It is recommended that the phase noise of the transmitted signal at 20 kHz be at least –104 dBc/Hz. Some frequency synthesizers may not meet this requirement.

8.6 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 8.3.

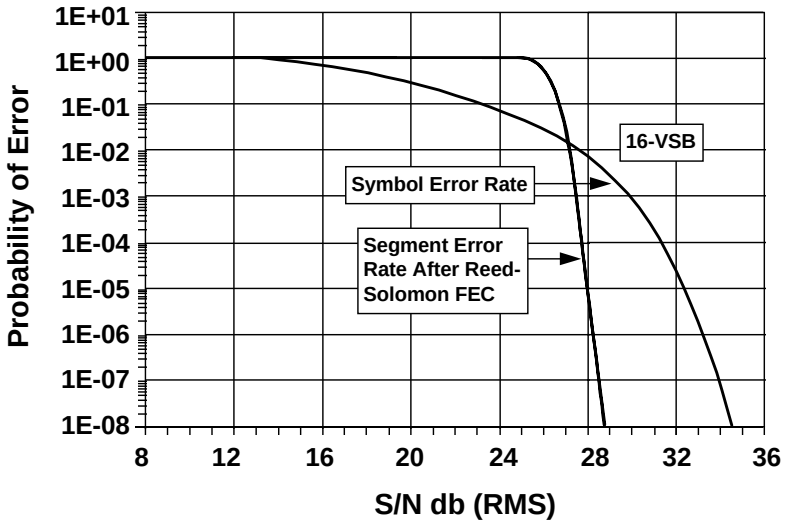

Figure 8.3 16-VSB error probability.

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 8.4 and is slightly higher than that of the terrestrial mode.

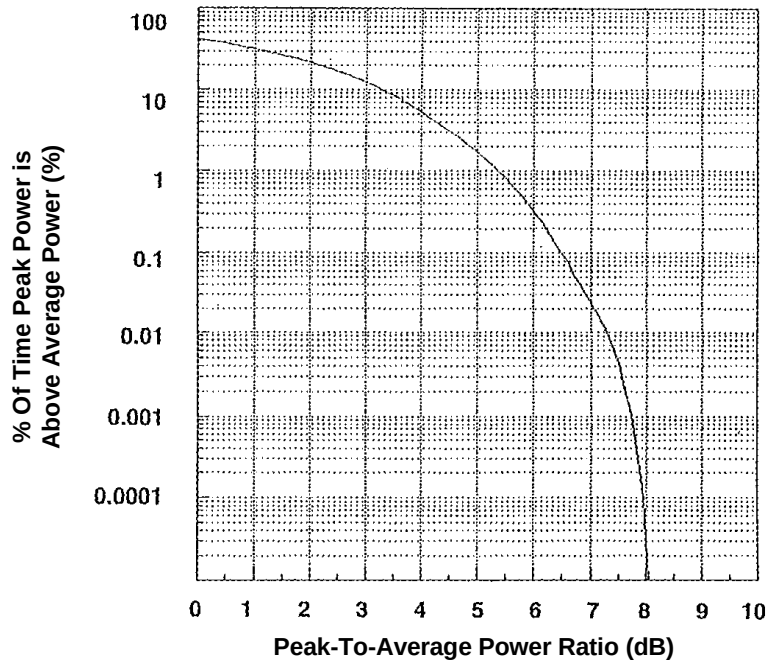

Figure 8.4 Cumulative distribution function of 16-VSB peak-to-average power ratio.

9. RECEIVER SYSTEMS

This section describes various aspects of receiving VSB DTV signals, reproducing transport streams from them, and converting those transport streams to video and audio signals. This section provides references to literature descriptive of VSB DTV receiver design.[12]

9.1 General Issues Concerning DTV Reception

The Working Parties of the Planning and Systems Subcommittees of the Advisory Committee on Advanced Television Service (ACATS) attempted to include all known, theoretically important conditions for signal reception in the original testing of the DTV system. The data gathered by those subcommittees can still help in developing receiver designs.

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. The test results[13] comprise information concerning multipath interference and other impairment conditions, and their effect on the bit error rate of the digital signal. This information is still pertinent.

9.1.1 Planning Factors Used by ACATS PS/WP3

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

The selection of a transmission subsystem began with a series of laboratory tests performed 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. Then, the results of these tests and subsequent VSB modem tests, together with a set of receiver planning factors, were incorporated into the programming of a computer modeling nationwide spectrum utilization. This computer model 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 17 February 1994. That figure shows the performance of 8 VSB and 32 QAM in the presence of thermal noise, co-channel and adjacent-channel interference from DTV and NTSC. Figure 1 also shows the performance of each type of signal as a taboo interferer into NTSC. The change in noise threshold of each type of signal when received subject to specific ensemble multipath characteristics is shown in Figure 1, but did not enter into the computer model.

The receiver planning factors incorporated in the spectrum utilization computer model can be found in the Final Report of PS/WP3, under the heading “Receiver Planning Factors Applicable to All DTV Systems.” Table 9.1 shows these Planning Factors and is excerpted from that final report. The final report indicated that “antenna factors are based on the geometric mean frequencies of the three broadcast bands,” and that, in addition to front-to-back ratio F/B, “a formula is employed for the forward lobe simulating an actual receiving antenna pattern.”

---

[12] Much of this literature is found in U. S. patents. Patent references in Section 9 should be regarded solely as technical literature. ATSC makes no representations concerning the validity or scope of the patents, or efficacy of the technology.

[13] ACATS: “Field Test Results of the Grand Alliance Transmission Subsystem,” Document SS/WP2-1354, September 16, 1994.

Table 9.1 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

The computer analysis, done taking both the ATTC results and receiver planning factors into account, led PS/WP3 to conclude as follows in its Document 296.

> "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 DTV will be broadcast."

The differences between the two systems with regard to performance under multipath reception conditions was later ascribed as being principally due to differences in the way the equalizers were designed in the 8 VSB receiver and in the 32 QAM receiver.[14]

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. The effects of multiple-receiver splitters, VCRs or poor VSWR were not considered. In weak-signal areas a low-noise antenna amplifier (LNA) can be used to compensate for degradation in signal level introduced by such devices.

The planning factors as adopted by PS/WP-3 were intended to provide reception coverage comparable with NTSC coverage and accordingly were based upon similar NTSC planning factors.[15] Improvements in technology, such as reduced feedline losses were included, as well as those allowances known to be unique to the DTV system requirements. Antenna gains were left the same as the NTSC factors. As a practical matter for current consumer outdoor antennas, the UHF gain can average 10 dBd over the UHF band, but may be 8 dBd at a particular channel. (The unit of measurement "dBd" is the gain in dB of the antenna response respective to the response of a half-wave dipole antenna.) Front-to-back (F/B) ratios were reduced somewhat, based on practical experience that reflections from nearby objects often limited the usable directivity.

---

[14] M. Ghosh: "Blind Decision Feedback Equalization for Terrestrial Television Receivers", *Proc. of the IEEE*, vol. 86, no. 10, pp. 2070–2081, October, 1998.

[15] Robert O'Connor: "Understanding Television's Grade A and Grade B Service Contours", *IEEE Transactions on Broadcasting*, vol. BC-14, no. 4, pp. 137–143, December 1968.

The Receiver Planning factors listed in Table 9.1 were useful in comparing DTV and NTSC service areas, but individual consumer installations may experience failure or impairment of reception inside the service area predicted by such factors.[16] [17] Many of these reception problems can be mitigated by use of a mast-mounted low-noise amplifier (LNA), which is available currently from several manufacturers.

Some attempts by ATSC committees and others have been made to analyze DTV reception that uses indoor or set-top antennas ("rabbit ears"). Some estimation of loss of field strength versus antenna height can be made. In going from the planning factor height of 30 ft. to perhaps 6 ft., field strength will be reduced by 12 dB.[18] However, the wide variation of construction materials and techniques of housing cause large variations in indoor signal strength, polarization and multipath.

9.1.2 Noise Figure

A number of factors enter into the ultimate carrier-to-noise ratio within the receiver. Table 9.1 shows the receiver planning factors applicable to UHF that were used by PS/WP3.[19]

In the original PS/WP-3 discussions, dual-conversion tuners were assumed to offer the required protection from interference. Such tuners were also assumed to have a degraded VHF noise figure compared to single-conversion tuners, owing to phase noise in the additional converter.

PS/WP-3 estimated UHF noise figure to be 10 dB, as defined by the present FCC techniques. When the FCC made calculations for channel assignments, this planning factor was changed to 7 dB.

A consumer can improve the noise performance of an installation by using a low-noise amplifier (LNA).

9.1.3 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 made when testing 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. Table 9.2 expresses the results of these tests in equivalent TOV numbers derived from the BER measurements.

---

[16] O. Bendov, J. F. X. Brown, C. W. Rhodes, Y. Wu, and P. Bouchard: "DTV Coverage and Service Prediction, Measurement and Performance Indices", *IEEE Trans. on Broadcasting*, vol. 47, no. 3, pg. 207, September 2001.

[17] "ATSC Transmission System, Equipment and Future Directions: Report of the ATSC Task Force on RF System Performance", revision 1.0 published 12 April 2001, Chapter 6, "Analysis of 8-VSB Performance", pp. 12–21.

[18] J. J. Gibson and R. M. Wilson: The Mini-State — a Small Television Antenna", *RCA Engineer*, vol. 20, pp. 9–19, 1975.

[19] Final Report of the Spectrum Utilization and Alternatives Working Party of the Planning Subcommittee of ACATS.

Table 9.2 DTV Interference Criteria

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

Note that the exact amount of adjacent-channel or co-channel interference entering the receiver terminals depends on the exact overall antenna gain pattern used, rather than just F/B ratio. The exact overall antenna gain pattern of an antenna usually differs from one channel to another, including change in F/B ratio.

9.1.4 Unintentional Radiation

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

9.1.5 Direct Pickup (DPU)

Rules for direct pickup are included in the FCC Rules, Part 15. The image rejection numbers of 50 dB or 60 dB that IS-23 Section 3.28 specifies for acceptable NTSC analog TV reception may be much larger than what is required for acceptable DTV reception. A digital system is much more robust against image interference, and regulation of image rejection depth will not be necessary.

9.2 Grand Alliance Receiver Design

Figure 9.1 shows the receiver block diagram of the VSB terrestrial broadcast transmission system as implemented in the Grand Alliance prototype receiver. Sections 9.2.2 through 9.2.13, following, describe each functional block.

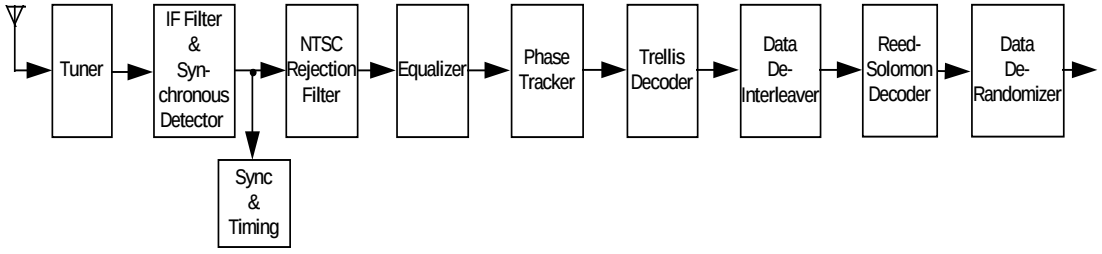

Figure 9.1 Block diagram of Grand Alliance prototype VSB receiver.

Current production designs usually differ somewhat from the prototype receiver with regard to the order in which signal processing is performed. Most recent designs digitize IF signals and perform demodulation (such as synchronous detection) in the digital regime. In some designs the arrangement of the NTSC rejection filtering, equalizer, and phase-tracker (or symbol

synchronizer) portions of the receiver differ from those shown in Figure 9.1. Portions of the sync and timing circuitry are likely to take their input signals after equalization, rather than before, and bright-spectral-line symbol-clock-recovery circuits may respond to envelope detection of IF signals performed separately from synchronous detection. The cascading of trellis decoder, data de-interleaver, Reed-Solomon decoder and data de-randomizer is characteristic of many designs, however.

9.2.1 Tuner

The tuner, illustrated in Figure 9.2, as implemented in the prototype the Grand Alliance submitted for test, receives the 6 MHz signal (UHF or VHF) from the antenna. The tuner is a high-side injection, double-conversion type with a first IF frequency of 920 MHz. This places the image frequencies above 1 GHz, making them easy to reject by a fixed front-end filter. This first IF frequency was chosen high enough that the input band-pass filter selectivity prevents first local oscillations (978–1723 MHz) leaking from the tuner front end and interfering with other UHF channels, but low enough that second harmonics of UHF channels (470–806 MHz) fall above the first IF passband. Although harmonics of cable-channel signals could occur in the first IF passband, they 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 image frequency range of the tuner (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, and introduces very little channel tilt, if any. This contrasts with the tracking filter used in some NTSC single-conversion tuner designs, in which the tracking filter is required to reject image signals only 90 MHz away from the desired channel.

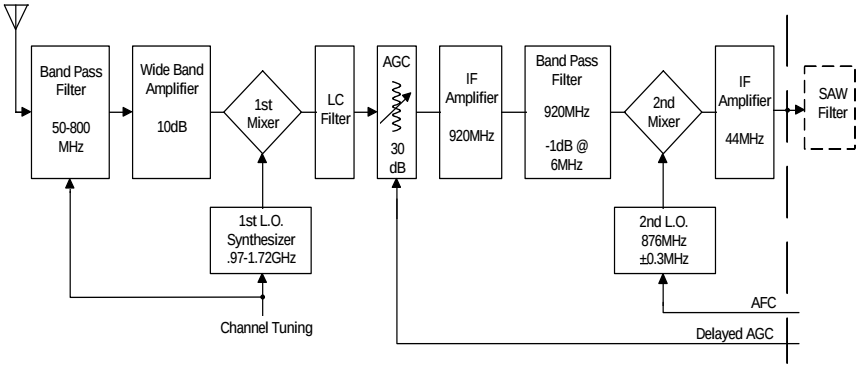

Figure 9.2 Block diagram of the tuner in the prototype VSB receiver.

A 10 dB gain, wideband RF amplifier increases the signal level into the first mixer, and is the predominant factor determining receiver noise figure (7–9 dB over entire VHF, UHF, and cable bands). The first mixer is a highly linear, double-balanced design to minimize even-harmonic generation. The first mixer is of high-side-injection type, being driven by a synthesized low-phase-noise local oscillator (LO) oscillating at a frequency above those of the broadcast signal selected for reception. 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 of low-side-injection type and is driven by the second LO, which is an 876 MHz voltage-controlled SAW oscillator. (Alternatively, the second mixer could be of high-side-injection type.[20]) The second oscillator is controlled by the frequency-and-phase-lock-loop (FPLL) synchronous detector. The second mixer, the output signal of which occupies a 41–47 MHz second IF frequency passband, drives a constant-gain 44 MHz amplifier. The output of the tuner feeds the IF SAW filter and synchronous detection circuitry. The dual-conversion tuner used in the Grand Alliance receiver is made out of standard consumer electronic components, and is housed in a stamped metal enclosure.

Since the testing of the original Grand Alliance systems, alternative tuner designs have been developed. Practical receivers with commercially acceptable performance are now manufactured using both dual-conversion and single-conversion tuners.

Compared to an NTSC analog television receiver, a DTV receiver has two particularly important additional design requirements in its front-end portion up to and including the demodulator(s). One of these requirements is that the phase noise of the local oscillator(s) must be low enough to permit digital demodulation that is reasonably free of symbol jitter. Symbol jitter gives rise to *intersymbol interference* (ISI), which should be kept below levels likely to introduce data-slicing errors. The Grand Alliance receiver could accommodate total phase noise (transmitter and receiver) of –78 dBc/Hz at 20 kHz offset from the carrier. (Note, all figures are measured at 20 kHz offset, and assume a noise-like phase distribution free of significant line frequencies that may be caused by frequency synthesis.) By 2002, fully integrated demodulator designs typically showed an improvement 5 dB or so over the Grand Alliance hardware. It is recommended that the transmitter have a maximum phase noise of –104 dBc/Hz so practically the whole margin is available to the tuner design. For example, a conservative margin of 12 dB based on the Grand Alliance demodulator performance requires a tuner phase noise of –90 dBc/Hz.

The other particularly important additional design requirement of the front-end portion of the DTV receiver is that the immunity to interference must be better in general than in an analog TV receiver. Table 9.2 shows that such better performance was assumed in making the present channel assignments. During the interim transition period in which both NTSC and DTV signals are broadcast, immunity to NTSC interference from channels other than the one selected for reception will be particularly important. This is because the variations in received levels from stations to adjacent stations can be very large, with variations on the order of 40 dB not uncommon. DTV-to-DTV variations can be the same order of magnitude. DTV receivers can use digital filtering methods for rejecting NTSC co-channel interference, which methods are inapplicable to NTSC receivers.

The principal NTSC interference problem during DTV reception is cross-modulation of a strong out-of-channel NTSC signal with desired-channel DTV signal in the RF amplifier and first mixer stages of the DTV receiver. Actually, the desired-channel DTV signal is less adversely affected by cross-modulation with a strong out-of-channel signal than a desired-

[20] A. L. R. Limberg: "Plural-Conversion TV Receiver Converting 1st IF to 2nd IF Using Oscillations of Fixed Frequency Above 1st IF", U. S. patent No. 6 307 595, 23 October 2001.

channel analog NTSC signal is. However, during the transition period the DTV signal will be subject to a more demanding interference environment. Because each broadcaster was provided with a DTV channel, DTV power had to be reduced to avoid loss of NTSC service area. Furthermore, DTV assignments could not be protected from the UHF taboos nor by adjacent channel restrictions. Thus, the generally accepted conclusion is that DTV receivers should be capable of significantly better interference rejection than NTSC receivers currently produced are.

9.2.2 Channel Filtering and VSB Carrier Recovery

The Grand Alliance prototype receiver generates in-phase and quadrature-phase analog carrier signals at intermediate frequency, which carrier signals are used to implement complex synchronous detection of the VSB IF signal by analog methods. A complex baseband DTV signal is reproduced that has an I baseband DTV signal resulting from the in-phase synchronous detection as a real component and that has a Q baseband DTV signal resulting from the quadrature-phase synchronous detection as an imaginary component. The prototype receiver uses pilot carrier components in both the I and Q baseband DTV signals for controlling carrier recovery.[21] All subsequent signal recovery procedures involve only the I baseband DTV signal.

Figure 9.3 illustrates portions of the prototype receiver, including the frequency- and phase-lock loop (FPLL) circuit used for carrier recovery. The first LO is synthesized by a phase-lock loop (PLL) and controlled by a microprocessor. The third LO is a fixed-frequency reference oscillator. Automatic-phase-and-frequency control (AFPC) signal for the second LO comes from the FPLL synchronous detector, which responds to the small pilot carrier in the received DTV signal. The FPLL provides a pull-in range of ±100 kHz despite the automatic-phase-control low-pass filter having a cut-off frequency less than 2 kHz. When the second LO is in phase lock, the relatively low cut-off frequency of this APC low-pass filter constrains operation of the FPLL to narrow enough bandwidth to be unaffected by most of the modulation of the digital carrier by data and synchronizing signals. However, the FPLL bandwidth remains wide enough to track out any phase noise on the signal (and, hence, on the pilot) of frequencies up to about 2 kHz. Tracking out low-frequency phase noise (as well as low frequency FM components) allows the phase-tracking loop to be more effective.

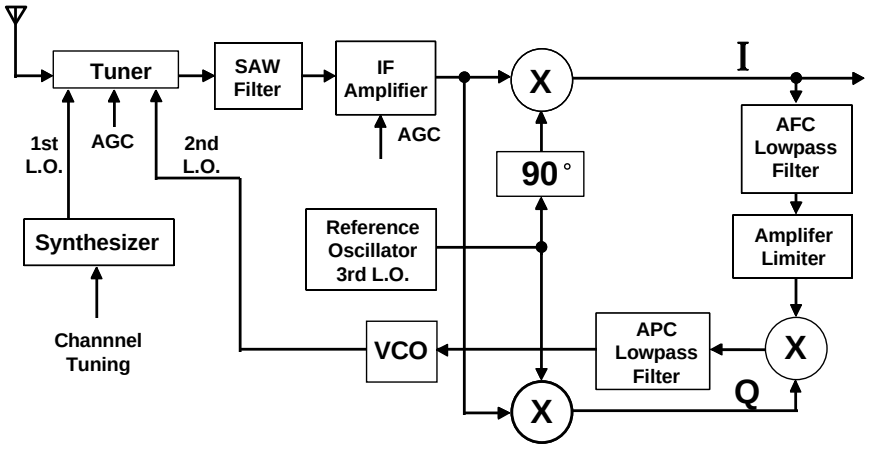

Figure 9.3 Tuner, IF amplifier, and FPLL in the prototype VSB receiver.

[21] R. Citta" "Automatic phase and frequency control system", U. S. patent No. 4 072 909, 7 February 1978.

The pilot beat signal in the Q baseband signal is apt to be at a frequency above the 2 kHz cut-off frequency of the APC low-pass filter, so it cannot appear directly in the AFPC signal applied to the VCO. The automatic-frequency-control (AFC) low-pass filter has a higher cut-off frequency and selectively responds to any pilot beat signal in the I baseband signal that is below 100 kHz in frequency. The response of AFC filter exhibits positive phase shift (lag) of that pilot beat signal that increases with frequency, from a 0 lag at zero-frequency to a 90 lag at a frequency well above the 100 kHz cut-off frequency. The response to pilot beat signal drives the amplifier/limiter well into clipping, to generate a constant amplitude (±A) square wave that is used to multiply the Q baseband DTV signal, to generate a switched-polarity Q baseband DTV signal as product output signal from the multiplier.

Without the AFC filter, the I channel beat note and the Q channel beat note would always be phased at 90 degrees relative to each other, and the direct component of the product output signal from the multiplier would be zero because each half cycle of the A square wave would include equal negative and positive quarter-cycles of the sinusoidal Q beat wave. However, the AFC filter lag delays the square wave polarity transitions, so there is less than a quarter wave of the Q baseband DTV signal before its sinusoidal zero-axis crossing, and more than a quarter wave of the Q baseband signal after its sinusoidal zero-axis crossing. This results in a net dc value for the product output signal from the multiplier. The polarity of the Q baseband DTV signal relative to the square wave depends on the sense of the frequency offset between the IF pilot and the VCO oscillations. Owing to the variable phase shift with frequency of the AFC filter, the amount of phase shift between the square wave and the Q baseband beat note depends on the frequency difference between the VCO oscillation and the incoming pilot. The amount of this phase shift determines the average value of the multiplied Q baseband DTV signal. This zero-frequency component passes through the APC low-pass filter to provide a frequency-control signal that reduces the difference between the VCO frequency and the carrier frequency of the incoming IF signal. (An extended frequency sweep of the response of the FPLL will exhibit the traditional bipolar S-curve AFC characteristic.)

When the frequency difference comes close to zero, the phase shift in the AFC filter approaches zero, and so the AFC control voltage also approaches 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, except for being bi-phase stable. However, the correct phase-lock polarity is determined by forcing the polarity of the pilot to be the same as the known transmitted positive polarity.[22] [23] [24]

The bi-phase stability arises in the following ways. When the VCO is in phase-lock, the detected pilot signal in the real component I of the complex baseband DTV signal is at zero frequency, causing the AFC filter response to be at a constant direct value. This constant direct value conditions the amplifier/limiter output signal either to be continually at +A value or to be continually at –A value. The Q baseband DTV signal resulting from the quadrature-phase synchronous detection is multiplied by this value, and the APC low-pass filter response to the resulting product controls the frequency of the oscillations supplied by the second LO in the tuner. So, there is no longer a zero-frequency AFC component generated by heterodyning higher-frequency demodulation artifacts from the pilot. When the loop settles near zero degrees, the limited I value is +A, and the Q baseband is used as in an ordinary PLL, locking the Q

[22] G. Krishnamurthy, T. G. Laud, and R. B. Lee: "Polarity Selection Circuit for Bi-phase Stable FPLL", U. S. patent No. 5 621 483, 15 April 1997.

[23] G. J. Sgrignoli: "Pilot Recovery and Polarity Detection System", U. S. patent No. 5 675 283, 7 October1997.

[24] V. Mycynek and G. J. Sgrignoli: "FPLL With Third Multiplier in an AC Path in the FPLL", U. S. patent No. 5 745 004, 28 April 1998.

channel at 90 degrees and the I channel at 0 degrees. However, if the loop happens to settle near 180 degrees, the limited I value is –A, causing the multiplied Q channel signal to be reversed in sense of polarity, and therefore driving the loop to equilibrium at 180 degrees.

The prototype receiver can acquire a signal and maintain lock at a signal-to-noise ratio of 0 dB or less, even in the presence of heavy interference. Because its 100 kHz cut-off frequency, the AFC low-pass filter rejects most of the spectral energy in the I baseband DTV signal with 5.38 MHz bandwidth. This includes most of the spectral energy in white noise, in randomized data and in the PN sequences in the DFS signal. The AFC low-pass filter also rejects most of the demodulation artifacts in the I baseband DTV signal that arise from any co-channel NTSC interference, except those arising from the vestigial sideband of the co-channel NTSC interference. So most of the energy from white noise, DTV symbols and any NTSC co-channel interference is removed from the amplifier/limiter input signal. This makes it extremely likely that the beat signal generated by demodulating the pilot is the largest component of the amplifier/limiter input signal so that this beat signal "captures" the limiting and generates square wave essentially unaffected by the other components. Any demodulation artifacts in the Q baseband DTV signal that arise from any co-channel NTSC interference that are down-converted to frequencies below the 2 kHz cut-off frequency of the APC low-pass filter when those artifacts are multiplied by the amplifier/limiter output signal will be within 2 kHz of the digital carrier signal. Only a low-energy portion of the NTSC vestigial sideband falls within this 4-kHz range of signal, so co-channel NTSC interference will have little influence on AFC during pull-in of the second LO towards phase-lock. When the second LO is phase-locked, the narrowband APC low-pass filter rejects most of the energy in white noise, in randomized data, in the DTV sync signals and in any co-channel NTSC interference.

The synchronous detectors and the FPLL loop employed in the prototype VSB receiver were constructed using analog electronic circuitry. Similar processing techniques can be employed in an FPLL constructed in considerable portion using digital electronic circuitry, for use after synchronous detectors constructed in digital circuitry and used for demodulating digitized IF DTV signal.

9.2.3 Segment Sync and Symbol Clock Recovery

The repetitive data segment sync sequences (Figure 9.4) are detected from among the synchronously detected random data by a narrow bandwidth filter. From the data segment sync sequences, 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 9.5.

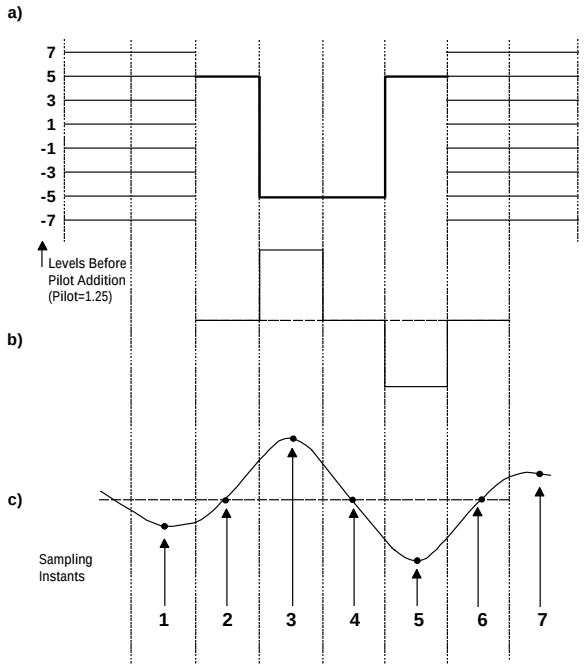

Figure 9.4 Data segment sync.

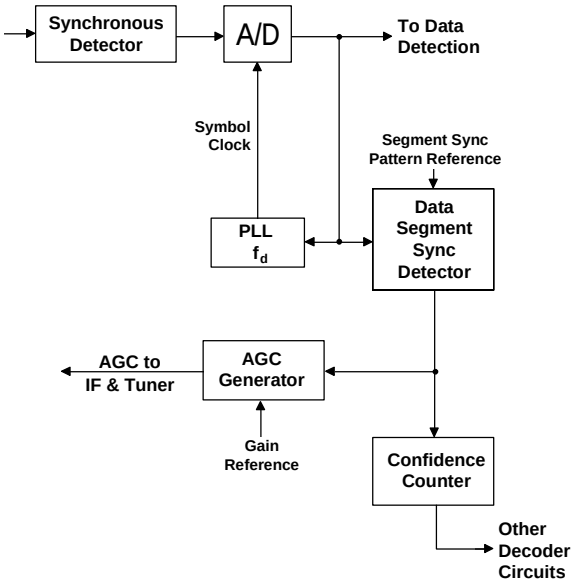

Figure 9.5 Segment sync and symbol clock recovery with AGC.

The synchronous detector supplies 10.76 Msymbols/s I-channel composite baseband data signal comprising data and sync sequences. An oscilloscope can be connected for viewing the traditional analog data eyes in this synchronous detection result. An A/D converter digitizes this synchronous detection result for digital processing. Owing to the sampling process used in the digitization, the data eyes in the digitized synchronous detection result cannot be observed by the conventional oscilloscope procedure.

A phase-lock loop (PLL) derives 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 searches for the two-level data segment sync (DSS) sequences 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. When a confidence counter reaches a predefined level of confidence that the segment sync has been found, subsequent receiver loops are enabled.

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

9.2.4 Non-Coherent and Coherent AGC

Prior to carrier and clock synchronization, non-coherent automatic gain control (AGC) is performed whenever any signal tends to exceed the dynamic range of the A/D converter. This signal may be a locked signal, an unlocked signal, or just noise/interference. The IF and RF gains are reduced to keep the A/D converter within its dynamic range, with the appropriate AGC "delay" being applied.

When data segment syncs are detected, coherent AGC occurs using the measured segment sync amplitudes.[25] The amplitude of the DSS sequences, relative to the discrete levels of the random data, is determined in the transmitter. Once the DSS sequences are detected in the receiver, they are compared to a reference value, with the difference (error) being integrated. The integrator output then controls the IF and "delayed" RF gains, forcing them to whatever values provide the correct DSS amplitudes.

9.2.5 Data Field Synchronization

Data Field Sync detection, shown in Figure 9.6, 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 can predict exactly where a valid field sync correlation should occur within each data segment, if it does occur. The field sync recovery circuit performs a symbol-by-symbol comparison, using a confidence counter to determine whether the valid field sync correlation does or does not occur in each data segment.[26] When a predetermined level of confidence is reached 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.

---

[25] R. W. Citta, D. M. Mutzabaugh, and G. J. Sgrignoli: "Digital Television Synchronization System and Method", U. S. patent No. 5 416 524, 16 May 1995.

[26] R. W. Citta, G. J. Sgrignoli, and R. Turner: "Digital Signal with Multilevel Signals and Sync Recognition", U. S. patent No. 5 598 220, 28 January 1997.

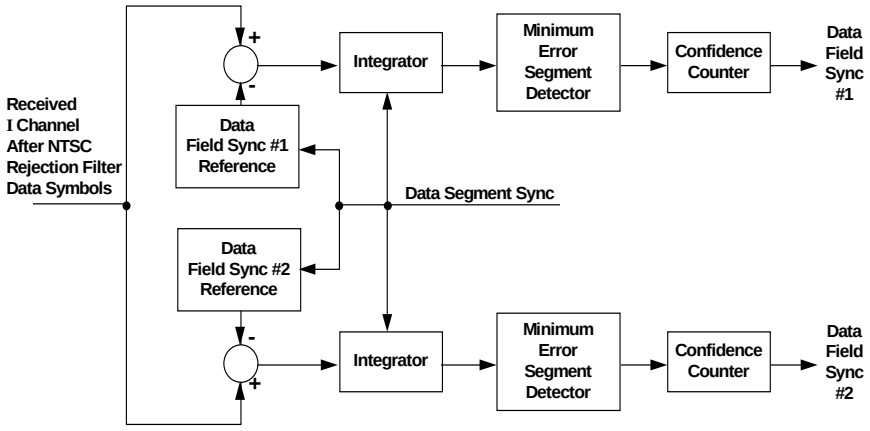

Figure 9.6 Data field sync recovery in the prototype VSB receiver.

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

9.2.6 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 9.7*a* shows the location and approximate magnitude of the three principal NTSC components:

- The video carrier (V) located 1.25 MHz from the lower-frequency edge of the channel allocation
- Chroma subcarrier (C) located 3.58 MHz above the video carrier frequency
- Audio carrier (A) located 4.5 MHz above the video carrier frequency

If there is an expectation that the reception area for a DTV broadcasting station will suffer co-channel interference from an NTSC broadcasting station, the DTV broadcasting station will generally shift its carrier to a frequency 28.615 kHz further from the lower frequency edge of the channel allocation. As Figure 9.7 shows, this places the data carrier (pilot) 338.056 kHz from the lower edge of the channel allocation, rather than the nominal 309.44 kHz.

The NTSC interference rejection filter (comb) is a single-tap linear feedforward filter, as shown in Figure 9.8. Figure 9.7*b* shows the frequency response of the comb filter, which provides periodic spectral nulls spaced 57 * f_H apart. That is, the nulls are 896.853 kHz apart, which is (10.762 MHz / 12). There are 7 nulls within the 6 MHz channel. The demodulation artifact of the video carrier falls 15.091 kHz above the second null in the comb filter response; the demodulation artifact of the chroma subcarrier falls 7.224 kHz above the sixth null in the comb filter response; and the demodulation artifact of the audio carrier falls 30.826 kHz above the seventh null in the comb filter response. Owing to the audio carrier being attenuated by the Nyquist roll-off of the DTV signal response, the demodulation artifact of the audio carrier in the comb filter response is smaller than the demodulation artifact of the video carrier.

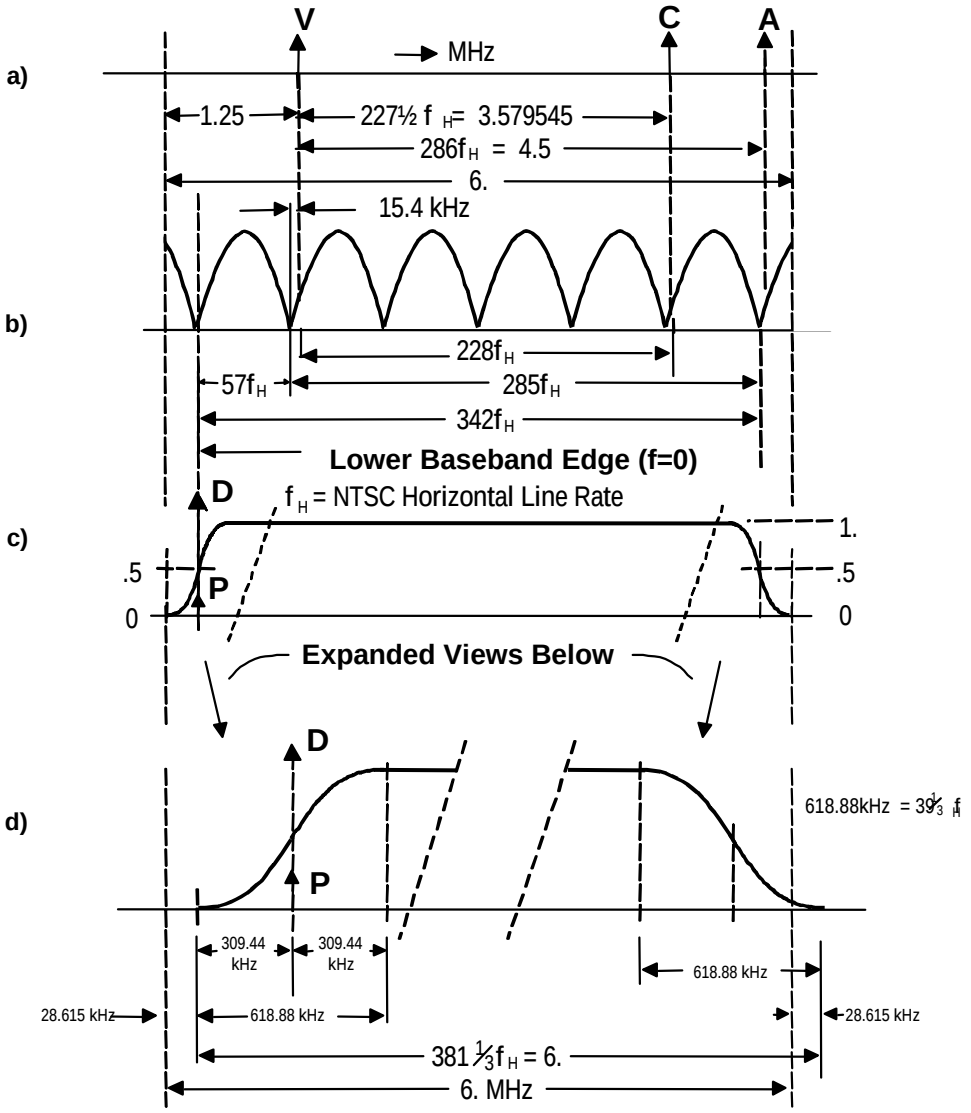

Figure 9.7. Location of NTSC carriers — comb filtering.

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 microseconds). So, 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 DTV 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 disperse the data-slicing errors caused by burst interference and will make it easier for the Reed-Solomon code to correct them. (The Reed-Solomon error-correction circuitry employed in the Grand Alliance receiver locates as well as corrects byte errors and 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). The partial response process causes a special case of intersymbol interference that doubles the number of like-amplitude eyes, but

otherwise leaves them open. The modified data signal can be properly decoded by the trellis decoder. Note that, because of time sampling, only the maximum data eye value is seen after A/D conversion.

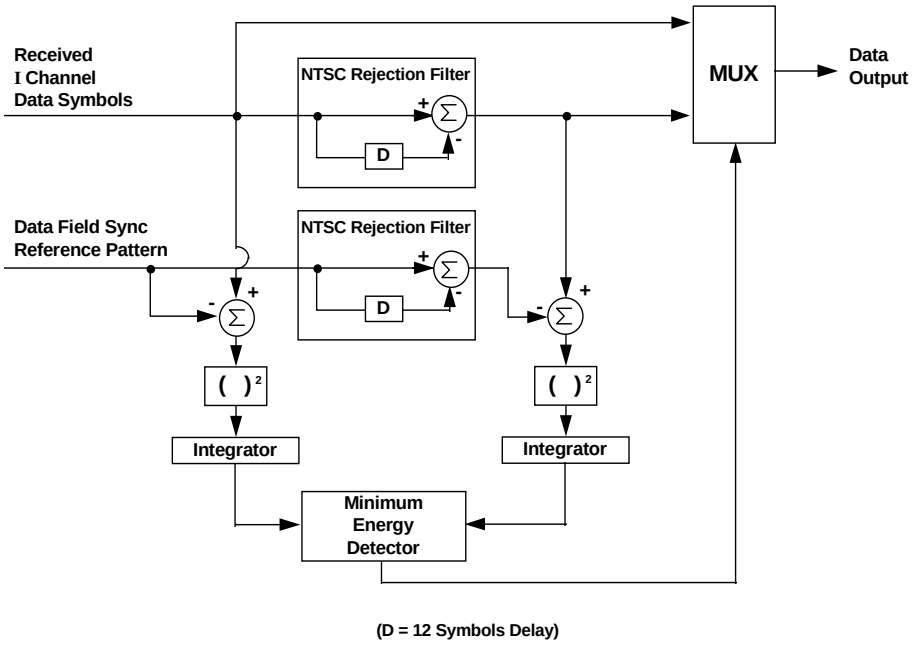

Figure 9.8 NTSC interference rejection filter in prototype VSB receiver.

NTSC interference can be automatically rejected by the circuit shown in Figure 9.8, which determines whether or not the NTSC rejection filter is effective to reduce interference-plus-noise in its response compared to interference-plus-noise in its input signal. The amount of interference-plus-noise accompanying the received signal during the data field sync interval is measured by comparing the received signal with a stored reference of the field sync. The interference-plus-noise accompanying the NTSC rejection filter response during the data field sync interval is measured by comparing that response with a combed version of the internally stored reference field sync. The two interference-plus-noise measurements are each squared and integrated over a few data field sync intervals. After a predetermined level of confidence is achieved, a multiplexer is conditioned to select the signal with the lower interference-plus-noise power for supplying data to the remainder of the receiver.

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 because the filter response is the difference 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. If little or no NTSC interference is present, the comb filter is automatically switched out of the data path. When NTSC broadcasting is discontinued, the comb filter can be omitted from digital television receivers.

9.2.7 Channel Equalizer

The equalizer/echo-suppressor compensates for linear channel distortions, such as tilt, and the spectrum variations caused by echoes. These distortions can come from the transmission channel or from imperfect components within the receiver.

The equalizer in the prototype receiver 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 so that future symbol errors in the equalizer response are likely to be reduced.[27] The algorithm does this by generating an estimate of the error currently present in the equalizer response. 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 used in the prototype receiver was designed to achieve equalization by any of three different methods. The equalizer can adapt on the prescribed binary training sequences in the DFS signals; it can adapt on data symbols throughout the frame when the eyes are open; or it can adapt on data symbols throughout the frame when the eyes are closed (blind equalization). The principal differences among these three methods concern how the error estimate is generated.

The prototype receiver stores the prescribed binary training signal in read-only memory, and the field sync recovery circuit determines the times that the prescribed binary training signal is expected to be received. So, when adapting on the prescribed binary training sequences, the receiver generates the exact reception error by subtracting the prescribed training sequence from the equalizer response.

Tracking dynamic echoes requires tap adjustments more often than the training sequence is transmitted. Therefore, the prototype Grand Alliance receiver was designed so that once equalization is achieved, the equalizer switches to a decision-directed equalization mode that bases adaptation on data symbols throughout the frame. In this decision-directed equalization mode, reception errors are estimated by slicing the data with an 8-level slicer and subtracting it from the equalizer response.

For fast dynamic echoes (e.g., airplane flutter) the prototype receiver used 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-correlation between the error signal and the data signal. The size of the correlation corresponds to the amplitude of the residual echo 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 the prototype receiver uses is shown in Figure 9.9. The dc bias of the input signal is removed by subtraction as a preliminary step before equalization. The dc bias is primarily the zero-frequency pilot component of the baseband DTV signal and is subject to change when the receiver selects a different reception channel or when multipath conditions change. The dc offset is tracked by measuring the dc value of the training signal.

[27] B. Widrow and M. E. J. Hoff: "Adaptive Switching Circuits", IRE 1960 Wescon Conv. Record, pp. 563–587.

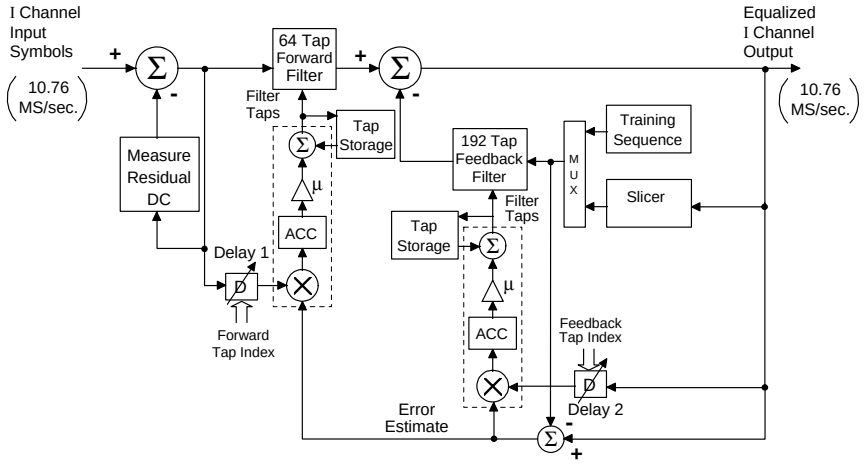

Figure 9.9 Equalizer in the prototype VSB receiver.

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

The output signals of the forward filter and feedback filter are summed to produce the equalizer response. The equalizer response is sliced by either an 8-level slicer (15-level slicer when the comb filter is used) or by a binary slicer, depending upon whether the data eyes are open or closed. (As pointed out in the previous section on interference filtering, the comb filter does not close the data eyes, but generates in its response twice as many eyes of the same magnitude). This sliced signal has the training signal and segment syncs reinserted. 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.

The complete prototype receiver demonstrated in field and laboratory tests showed cancellation of –1 dB amplitude echoes under otherwise low-noise conditions, and –3 dB amplitude echo ensemble with noise. In the latter case, the signal was 2.25 dB above the noise-only reception threshold of the receiver.

9.2.8 Phase Tracker

The term "phase tracker" is a generic term applied to three servo loops—namely, a phase-tracking loop, an amplitude-tracking loop and an offset-tracking loop.[11.5] The phase-tracking loop is an additional decision-feedback loop which degenerates phase noise that is not degenerated by the IF PLL operating on the pilot.[28] [29] [30] Accordingly, two concatenated loops, rather than just one loop, degenerate phase noise in the demodulated DTV signal before its synchronous

[28] T. P. Horwitz, R. B. Lee, and G. Krishnamurthy: "Error Tracking Loop", U. S. patent No. 5 406 587, 11 April 1995.

[29] G. Krishnamurthy and R. B. Lee: "Error Tracking Loop Incorporating Simplified Cosine Look-up Table", U. S. patent No. 5 533 071, 2 July 1996.

[30] J. G. Kim, K. S. Kim, and S. W. Jung: "Phase Error Corrector for HDTV Reception System", U. S. patent No. 5 602 601, 11 February 1997.

equalization. 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. Only first-order loops are used in the phase-tracker of the Grand Alliance receiver. First-order loops were chosen for the phase-tracker because they follow fast changes in phase better than simple higher-order loops do.

The phase-tracking loop suppresses high-frequency phase noise that the IF PLL cannot suppress because of being relatively narrow-band. The IF PLL is relatively narrow-band in order to lock on the pilot carrier with minimum data-induced jitter. The phase tracker takes advantage of decision feedback to determine the proper phase and amplitude of the symbols and corrects a wider bandwidth of phase perturbations. Because of the decision feedback, the phase tracker does not introduce data-induced jitter. However, like any circuit that operates on data, the phase tracker begins to insert excess noise as the S/N ratio drops to approach threshold and bad decisions start to be made. Therefore, it may be desirable to vary phase tracker loop gain or switch it out near threshold signal-to noise-ratio.

A block diagram of the phase-tracking loop is shown in Figure 9.10. The output of the real equalizer operating on the I signal is first gain-controlled by a multiplier and then fed into a filter that recreates an approximation of the Q signal. This is possible because of the VSB transmission method, in which the I and Q components are related by a filter function that closely approximates a Hilbert transform. This filter is of simple construction, being a finite-impulse-response (FIR) digital 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.

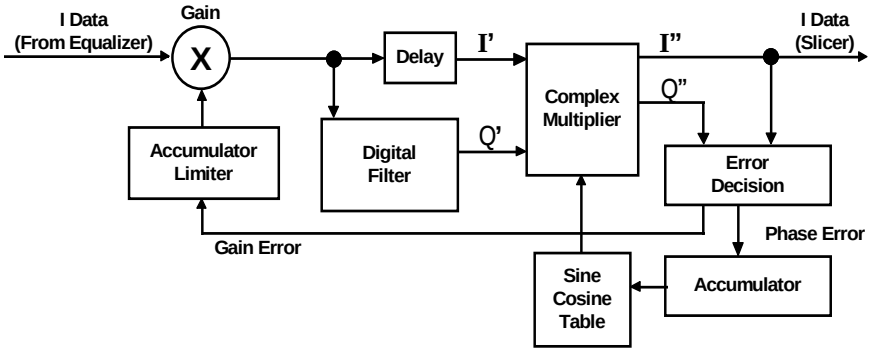

Figure 9.10 Phase-tracking loop portion of the phase-tracker.

These I and Q signals are then supplied to a complex multiplier operated as a de-rotator for suppressing 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. Since the phase tracker adjusts amplitude as well as phase, it can react to small AGC and equalizer errors. The phase tracker time constants are relatively short, as reflected in its 60 kHz bandwidth

The phase tracker provides an amplitude-tracking loop and an offset tracking loop as well as a phase-tracking loop. The phase tracker provides symbol synchronization that facilitates the use of a synchronous equalizer with taps at symbol-epoch intervals. Greater need for the phase tracker is evidenced in the 16-VSB mode than in the 8-VSB mode, being essential for 16-VSB

reception with some tuners. The phase tracker also removes effects of AM and FM hum that may be introduced in cable systems.

9.2.9 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 9.11, the receiver uses 12 trellis decoders in parallel, where each trellis decoder sees every 12th symbol.[31] [32] This code de-interleaving has all the same burst noise benefits of a 12-symbol-code de-interleaver, but also minimizes the resulting code expansion (and hardware) when the NTSC rejection comb filter is active.

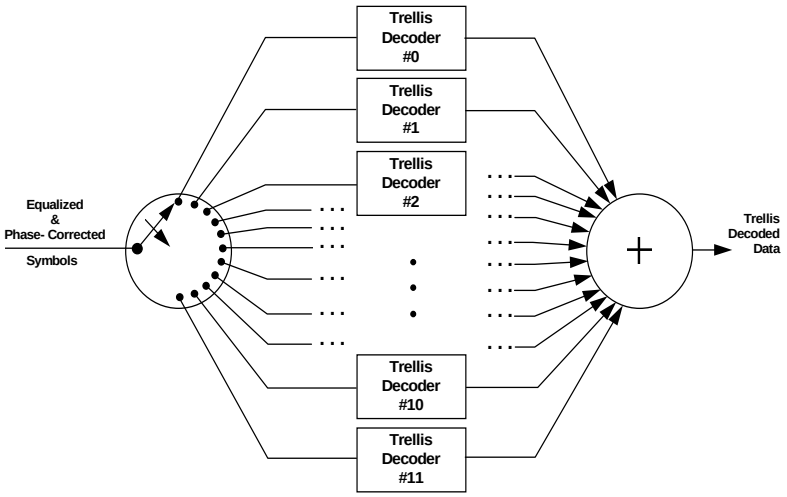

Figure 9.11 Trellis code de-interleaver.

Before the 8 VSB signal can be processed by the trellis decoder, it is necessary to remove the Segment Sync. The Segment Sync is not trellis encoded at the transmitter. The Figure 9.12 circuit block diagram illustrates the Segment Sync removal in the receiver.

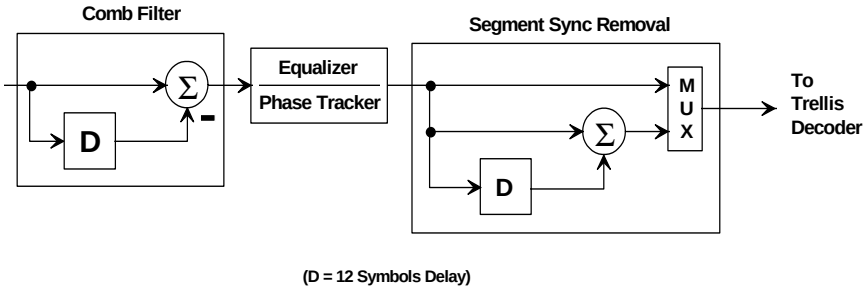

Figure 9.12 Segment sync removal in prototype 8 VSB receiver.

[31] R. W. Citta and D. A. Wilming: "Receiver for a Trellis Coded Digital Television Signal", U. S. patent No. 5 636 251, 3 June 1997.

[32] D. A. Wilming: "Data Frame Structure and Synchronization System for Digital Television Signal", U. S. patent No. 5 629 958, 13 May 1997.

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 interference, and the other when the NTSC rejection filter is not used. This is illustrated in Figure 9.13.

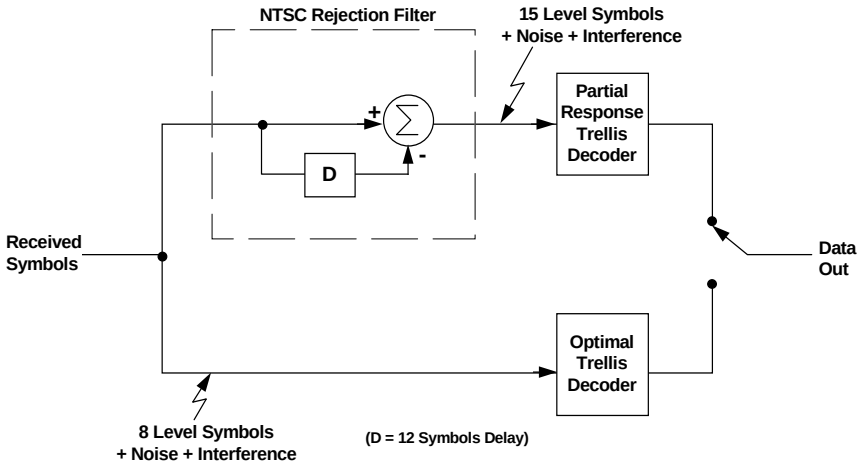

Figure 9.13 Trellis decoding with and without NTSC rejection filter.

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 regenerated in the same order in which they were generated in the encoder.

In the presence of significant NTSC co-channel interference, when the NTSC rejection filter (12 symbol, feedforward 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. 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 noise-threshold degradation beyond the 3 dB attributable to comb filtering is due to the 12-symbol differential coding.

Because Segment Sync is not trellis encoded nor pre-coded, the presence of the Segment Sync sequence in the data stream passed through the comb filter presents a complication that had to be dealt with. Figure 9.12 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 difference 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 depends 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 fewer than 5000 gates.

After the transition period, when NTSC is no longer being transmitted, the NTSC rejection filter and the 8-state trellis decoder can be omitted from digital television receivers.

9.2.10 Data De-Interleaver

The convolutional de-interleaver performs the exact inverse function of the transmitter convolutional interleaver. Its 1/6 data field depth, and inter-segment "dispersion" properties allow noise bursts lasting as long as about 193 microseconds to be corrected by subsequent Reed-Solomon error-correction circuitry of the type that locates byte errors as well as correcting them. Even strong NTSC co-channel signals passing through the NTSC rejection filter, and creating short bursts responsive to NTSC vertical edges, are reliably suppressed by this de-interleaving and RS error-correction process. The de-interleaver uses Data Field Sync for synchronizing to the first data byte of the data field. The functional concept of the convolutional de-interleaver is shown in Figure 9.14.

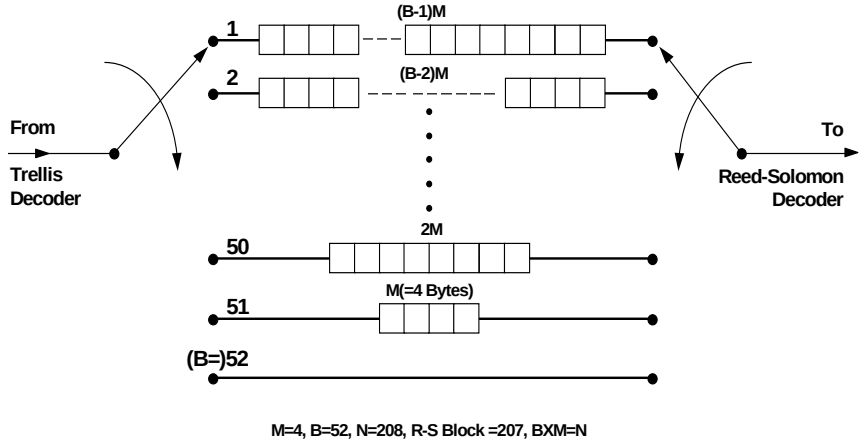

Figure 9.14 Connceptual diagram of convolutional de-interleaver.

In actual practice the convolutional de-interleaver is usually constructed in dual-port random-access memory (RAM).[33]

[33] M. Fimoff, S. F. Halozan, and R. C. Hauge: "Convolutional Interleaver and Deinterleaver", U. S. patent No. 5 572 532, 5 November 1996.

9.2.11 Reed-Solomon Decoder

The trellis-decoded byte data is sent to the (207,187) t =10 RS decoder. The RS decoder in the Grand Alliance receiver uses the 20 parity bytes to perform byte-error location and byte-error correction on a segment-by-segment basis. Up to 10-byte errors per data segment are corrected by the RS decoder. Isolated burst errors created by impulse noise, NTSC co-channel interference, or trellis-decoding errors that are less than 193 microseconds duration are corrected by the combination of the de-interleaving and RS error correction.

9.2.12 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 can be generated the same way 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 to some code word embedded within the potentially noisy data), it is exactly synchronized with the data, and performs reliably.

9.2.13 Receiver Loop Acquisition Sequencing

The Grand Alliance receiver incorporates a "universal reset" that 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 previously have confidence counters associated with them to insure proper operation. However, the build-up or decay 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.

9.2.14 High Data Rate Mode

The VSB digital transmission system provides the basis for a family of DTV 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 was originally designed to offer 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 with or similar to corresponding parts of the terrestrial system. A pilot, data segment sync, and data field sync are used in both systems. The pilot in the high data rate mode also adds 0.3 dB to the data power. The same symbol rate, the same segment rate, and the same field rate are used in the high-data-

rate-mode VSB system, allowing a VSB DTV receiver to lock up on either type of signal transmission. 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 similar to the VSB terrestrial receiver, except for the trellis decoder being 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 the high-data-rate-mode receiver does not employ an NTSC interference rejection filter.

9.3 Receiver Equalization Issues

The multipath distortion of a received DTV signal comprises an ensemble of "echoes" or "ghosts" that accompany the principal component of the received DTV signal because of the signal being received via various paths. Echoes received via paths that are shorter than that over which the principal signal is received are termed "pre-echoes". Echoes received via paths that are longer than that over which the principal signal is received are termed "post-echoes". The DTV receiver contains channel-equalization and echo-suppression filtering for selecting the principal signal and suppressing any accompanying echoes that have sufficient strength to cause errors during data-slicing. This filtering also corrects in some degree for the receiver not having optimal frequency response—e. g., owing to tilt in the IF amplifier response.

The ATSC Digital Television Standard specifies no particular design for the channel-equalization and echo-suppression filtering in the DTV receiver. When designing an adaptive equalizer, usually a designer considers and evaluates various possible ways of implementing this filtering.

Echoes are considered to be of substantial strength if their presence causes a readily observable increase in the number of data-slicing errors. There have been reports of pre-echoes of substantial strength being observed in the field that are advanced as much as 30 microseconds with respect to the principal signal. There have been a significant number of observations of post-echoes of substantial strength that are delayed on the order of 45 microseconds with respect to the principal signal.

Equalizers capable of suppressing pre-echoes advanced no more than 3 microseconds and delayed no more than about 45 microseconds were the general rule in the 1995 to 2000 timeframe. Since 2000 there has been increased awareness that channel equalizers should have capability for suppressing further-advanced pre-echoes and products with over 90 microseconds of total equalization range have been demonstrated.

A detailed discussion of DTV receiver equalization is beyond the scope of this document. Interested readers should consult ATSC Informational Document T3-600, "DTV Signal Reception and Processing Considerations."

9.4 Transport Stream Processing Issues in the Receiver

Receivers will sometimes encounter a transport bitstream that was transmitted in accordance with the MPEG-2 and ATSC standards, but differs somewhat from ordinary practice when it is received. There will often be bit streams within the multiplex that are not announced in the PAT or TS_program map_sections, such as TS packets with PID 0x1FFB (the PSIP base_pid). TS packets containing private data and identified with PID values not indicated in any TS_program_map_section may also be present.

Receiver/decoder designers should be aware that there are times when the TS multiplex contains TS packets with PID values that are not indicated in the current Program Specific Information (PSI) tables. This occurs when TS packets containing private data are transmitted, of

course. There also may be transient conditions where the PID values of received TS packets should be known at the receiver but in fact are not. Updated PSI table entries are sometimes irreparably corrupted during their over-the-air transmission, and accordingly the receiver does not use them to update its PSI table memory. Components described in the discarded update of PSI table entries that are free of uncorrectable error are likely to be received before updated PSI table entries are transmitted again and are successfully received. Since transmission of PSI table entries is as infrequent as every 0.4 second, it is impractical to do anything except immediately discard the data packets with PIDs that are not currently described in the PSI table memory within the receiver.

Receiver/decoder designers are also cautioned that there may be times when the TS multiplex does not contain TS packets with PID values that are referenced in the current PSI tables. For example, a program may comprise a video stream, two audio streams, and a data stream. TS packets of one of the audio streams may be absent from the TS during some time periods between PSI table memory updates, even though the absent audio stream continues to be listed in occurrences of the TS_program_map_section that describes the program. It is strongly advised that PSI tables temporarily stored in the receiver retain PSI entries with long-unused PIDs in their listing until such entries no longer appear in the PSI tables periodically transmitted in the TS.

Receiver/decoder designers should also take into consideration that data packets containing especially important information may each be transmitted more than once. The decoders for such data packets can be designed to discard repeated data packets that have already been successfully received. Alternatively, repeated data packets that have already been successfully received can be discarded during transport stream de-multiplexing.

None of the example Transport Streams described in this section violates any of the requirements concerning signal structure that the ATSC standards specify.

9.5 Receiver Video Issues

Transmissions conforming to the Digital Television Standard are expected to include the video formats as described in Table 9.3. 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 9.3, “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, and “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 9.3 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.

9.5.1 Multiple Video Programs

In the case of multi-video transmissions, receiver designers should consider that the DTV display may have to be capable of also displaying video in the 525-line scanning format. It is necessary to identify and extract the audio stream that 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. A DTV 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.

9.5.2 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 [14]). 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 and scan, and other parameters. Receivers require information concerning all of these parameters in order to conduct their operations.

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 must 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.

Coded bit streams are 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 begin the second bit stream. Accordingly, 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 MPEG-2 Video does not specify the behavior of decoders in this case, unless well-constrained concatenated sequences are produced, channel buffer overflow/underflow will at times occur at the junction between two coded sequences.

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.

9.5.3 D-Frames

The MPEG family of video coding standards (ISO 11172-2 [12] and ISO 13818-2 [14]) 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.

9.5.4 Adaptive Video Error Concealment Strategy

In MPEG video compression, video frames to be coded are formatted into sequences containing intra-coded (I), predictive-coded (P) and bi-directionally 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 frames and B frames, while B-frame errors will be isolated. Therefore, it is desirable to use error concealment techniques to prevent error propagation and, consequently, to improve the quality of reconstructed pictures.

There are two approaches that 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, because motion- processing is sometimes computationally demanding, it is not always available. 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.

9.5.4.1 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 that 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 throughout 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 select between spatial and temporal concealment modes. It is noted that the same idea can be used for intra-coded macroblocks of P frames or 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 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 providing motion vectors for I pictures. Motion information is very useful in concealing losses in P pictures and B pictures, but conventional MPEG-2 does not provide motion information 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. Applying a single estimated motion vector on an entire macroblock to create a temporal replacement can often result in a blocky shearing effect. So, instead, 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.

9.6 Receiver Audio Issues

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

9.6.1 Audio Coding

The audio specification may be found in Annex B of the Digital Television Standard [5]. The audio is encoded with the AC-3 system, which is documented in detail in the Digital Audio Compression (AC-3) Standard, ATSC Document A/52 [4]. Transmissions may contain main audio services, or associated audio services that are complete services (containing all necessary program elements), encoded at a bit rate up to and including 448 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 that are intended to be simultaneously reproduced is less than or equal to 576 kbps.

9.6.2 Audio Channels and Services

In general, a complete audio program 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. Eight types of audio services are 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 that 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 that 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. 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 that is a complete program mix may contain from one to 5.1 audio channels. Except in one case, an associated audio service containing a single program element that is intended to be combined with a main service contains only a single audio channel. In order to decode a main service and an associated service simultaneously, 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). Accordingly, receivers that also support optional decoding of the LFE channel 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 so entails no additional decoder complexity.)

It is not necessary for every receiver to decode completely 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 monophonic decoder are possible. For instance, only a single output buffer is required, so that decoder memory requirements are reduced; and some of the mix-down may occur in the frequency domain, so the complexity of the synthesis filter bank is reduced.

9.6.3 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 percent 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.

9.6.4 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 most members of the audience, who do not have an audio reproduction environment that matches of the one in 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 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). The audio decoder uses these values 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 so as to provide a subjectively pleasing, but restricted dynamic range. The level that is left unaltered 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 that generate the amount of compression 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 dynrng in order to reduce the effect of the dynamic range compression that 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). Accordingly, the viewer can be given independent control of the amount of compression applied to loud and quiet sounds. The details of these control functions can differ in different receiver designs.

9.6.5 Tracking of Audio Data Packets and Video Data Packets

The ATSC Implementation Subcommittee in ATSC Document IS-191 [3] wrote as follows.

> "MPEG-2 models the end-to-end delay from an encoder's signal input to a decoder's signal output as constant. This end-to-end delay is the sum of the delays from encoding, encoder buffering, multiplexing, transmission, demultiplexing, decoder buffering, decoding, and presentation. Presentation time stamps are required in the MPEG bit stream at intervals not exceeding 700 milliseconds. The MPEG System Target Decoder model allows a maximum decoder buffer delay of one second. Audio and video presentation units that represent sound and pictures that are to be presented simultaneously may be separated in time within the transport stream by as much as one second. In order to produce synchronized output, IS finds that the receiver must recover the encoder's System Time Clock (STC) and use the Presentation Time Stamps (PTS) to present the audio-video content to the viewer with a tolerance of +/-15 milliseconds of the time indicated by PTS.

Page 109 of 109