

**Digital Radio Mondiale (DRM) –
System specification
for digital transmissions
in the broadcasting bands
below 30 MHz**

Withdrawn

PUBLICLY AVAILABLE SPECIFICATION



INTERNATIONAL
ELECTROTECHNICAL
COMMISSION



Reference number
IEC/PAS 62272-1

Withdrawn

ETSI TS 101 980 V1.1.1 (2001-09)

Technical Specification

Digital Radio Mondiale (DRM); System Specification

Withdrawn



Documents

Document Sharing Hub

Withdrawn

INTERNATIONAL ELECTROTECHNICAL COMMISSION

DIGITAL RADIO MONDIALE (DRM) –

System specification for digital transmissions
in the broadcasting bands below 30 MHz

FOREWORD

A PAS is a technical specification not fulfilling the requirements for a standard, but made available to the public and established in an organization operating under given procedures.

IEC-PAS 62272-1 is identical to ETSI specification TS 101980. The ETSI document is the result of a co-operation between Digital Radio Mondiale Consortium (DRM) and the European Telecommunications Standards Institute (ETSI)

Both ETSI and DRM expressed their willingness to have this ETSI document published as an IEC PAS.

IEC-PAS 62272-1 has been processed by IEC technical committee 103: Transmitting equipment for radio-communication.

The text of this PAS is based on the following document:

This PAS was approved for publication by the P-members of the committee concerned as indicated in the following document:

Draft PAS	Report on voting
103/24/PAS	103/25/RVD

Following publication of this PAS, the technical committee or subcommittee concerned will investigate the possibility of transforming the PAS into an International Standard.

- 1) The IEC (International Electrotechnical Commission) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of the IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, the IEC publishes International Standards. Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. The IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
- 2) The formal decisions or agreements of the IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested National Committees.
- 3) The documents produced have the form of recommendations for international use and are published in the form of standards, technical specifications, technical reports or guides and they are accepted by the National Committees in that sense.
- 4) In order to promote international unification, IEC National Committees undertake to apply IEC International Standards transparently to the maximum extent possible in their national and regional standards. Any divergence between the IEC Standard and the corresponding national or regional standard shall be clearly indicated in the latter.
- 5) The IEC provides no marking procedure to indicate its approval and cannot be rendered responsible for any equipment declared to be in conformity with one of its standards.
- 6) Attention is drawn to the possibility that some of the elements of this PAS may be the subject of patent rights. The IEC shall not be held responsible for identifying any or all such patent rights.

Contents

Intellectual Property Rights	6
Foreword	6
Introduction	6
1 Scope	8
2 References	8
3 Definitions, symbols and abbreviations	9
3.1 Definitions	9
3.2 Symbols	9
3.3 Abbreviations	10
4 General characteristics	11
4.1 System overview	11
4.2 System architecture	11
4.3 Source coding	13
4.4 Transmission modes	14
4.4.1 Signal bandwidth related parameters	14
4.4.2 Transmission efficiency related parameters	14
4.4.2.1 Coding rates and constellations	14
4.4.2.2 OFDM parameter set	14
5 Source coding modes	16
5.1 Overview	16
5.1.1 AAC Audio Coding	17
5.1.2 MPEG CELP coding	17
5.1.3 MPEG HVXC coding	18
5.1.4 SBR coding	18
5.2 UEP and audio super framing	18
5.3 AAC coding	19
5.3.1 AAC	19
5.3.1.1 AAC audio super frame	20
5.3.2 AAC + SBR	22
5.4 MPEG CELP coding	23
5.4.1 MPEG CELP	23
5.4.1.1 CELP audio super frame	25
5.5 HVXC	27
5.5.1 Definitions	28
5.5.1.1 HVXC source coder parameters	28
5.5.1.2 CRC bits for fixed bit rate modes	28
5.5.2 HVXC decoder	28
5.5.3 HVXC encoder	29
5.5.3.1 LPC analysis and LSP quantization	30
5.5.3.2 Open loop pitch search	30
5.5.3.3 Harmonic magnitude and fine pitch estimation	30
5.5.3.4 Vector quantization of harmonic magnitudes	30
5.5.3.5 Voiced/Unvoiced decision	30
5.5.3.6 VXC coding of unvoiced signals	30
5.5.4 HVXC channel coding	31
5.5.4.1 Protected Bit Selection	31
5.5.4.2 Syntax of DRM HVXC error resilience (ErHVXCfixframe_CRC)	33
5.5.5 Category Interleaving	37
5.5.6 HVXC error detection and concealment	38
5.5.6.1 Cyclic Redundancy Code (CRC)	38
5.5.6.2 Error concealment	38
5.5.6.2.1 Parameter replacement	39
5.6 SBR	41

5.6.1	Conceptual overview	41
5.6.2	AAC + SBR Decoding Process	42
5.6.2.1	Analysis Filterbank	42
5.6.2.2	Synthesis Filterbank	42
5.6.2.3	Gain Control	43
5.6.2.4	T/F Grid Control	43
5.6.2.5	Huffman Decoder	43
5.6.3	AAC + SBR Protocol	46
5.6.3.1	AAC + SBR syntax	46
5.6.3.2	SBR bit stream element definitions	52
6	Multiplex definition	56
6.1	Introduction	56
6.2	Main Service Channel	56
6.2.1	Introduction	56
6.2.2	Structure	56
6.2.3	Building the MSC	57
6.2.3.1	Multiplex frames	57
6.2.3.2	Hierarchical frames	57
6.2.4	Reconfiguration	57
6.3	Fast Access Channel	57
6.3.1	Introduction	57
6.3.2	Structure	57
6.3.3	Channel parameters	58
6.3.4	Service parameters	59
6.3.5	CRC	61
6.3.6	FAC repetition	61
6.4	Service Description Channel	61
6.4.1	Introduction	61
6.4.2	Structure	61
6.4.3	Data Entities	62
6.4.3.1	Multiplex description data entity - type 0	63
6.4.3.2	Label data entity - type 1	64
6.4.3.3	Conditional access parameters data entity - type 2	64
6.4.3.4	Frequency information data entity - type 3	64
6.4.3.5	Frequency schedule data entity - type 4	67
6.4.3.6	Application information data entity - type 5	68
6.4.3.7	Announcement support and switching entity - type 6	69
6.4.3.8	Region definition data entity - type 7	70
6.4.3.9	Time and date information data entity - type 8	71
6.4.3.10	Audio information data entity - type 9	71
6.4.3.11	FAC channel parameters data entity - type 10	73
6.4.3.12	Linkage data entity - type 11	75
6.4.3.13	Language and country data entity - type 12	75
6.4.3.14	Other data entities	76
6.4.4	Summary of data entity characteristics	76
6.4.5	Changing the content of the SDC	76
6.4.6	Signalling of reconfigurations	77
6.4.6.1	Service reconfigurations	77
6.4.6.2	Channel reconfigurations	77
6.5	Text message application	78
6.5.1	Structure	78
6.6	Packet mode	79
6.6.1	Packet structure	80
6.6.1.1	Header	80
6.6.1.2	Data field	81
6.6.2	Asynchronous streams	81
6.6.3	Files	81
6.6.4	Choosing the packet length	81
7	Channel coding and modulation	82
7.1	Introduction	82

7.2	Transport multiplex adaptation and Energy dispersal	82
7.2.1	Transport multiplex adaptation	82
7.2.1.1	MSC	82
7.2.1.2	FAC	85
7.2.1.3	SDC	85
7.2.2	Energy dispersal	86
7.3	Coding	87
7.3.1	Multilevel coding	87
7.3.1.1	Partitioning of bitstream in SM and HMsym	89
7.3.1.2	Partitioning of bitstream in HMmix	90
7.3.2	Component Code	91
7.3.3	Bit Interleaving	97
7.3.3.1	FAC	97
7.3.3.2	SDC	98
7.3.3.3	MSC	98
7.4	Signal constellations and mapping	100
7.5	Application of coding to the channels	104
7.5.1	Coding the MSC	104
7.5.1.1	SM	104
7.5.1.2	HMsym	105
7.5.1.3	HMmix	105
7.5.2	Coding the SDC	106
7.5.3	Coding the FAC	106
7.6	MSC Cell Interleaving	106
7.7	Mapping of MSC cells on the transmission super frame structure	108
8	Transmission structure	109
8.1	Transmission frame structure and modes	109
8.2	Propagation-related OFDM parameters	111
8.3	Signal bandwidth related parameters	111
8.3.1	Parameter definition	111
8.3.2	Simulcast transmission	113
8.4	Pilot cells	113
8.4.1	Functions and derivation	113
8.4.2	Frequency references	113
8.4.2.1	Cell positions	113
8.4.2.2	Cell gains and phases	114
8.4.3	Time references	114
8.4.3.1	Cell positions and phases	115
8.4.3.2	Cell gains	116
8.4.4	Gain references	117
8.4.4.1	Cell positions	117
8.4.4.2	Cell gains	117
8.4.4.3	Cell phases	118
8.4.4.3.1	Procedure for calculation of cell phases	118
8.4.4.3.2	Robustness mode A	118
8.4.4.3.3	Robustness mode B	119
8.4.4.3.4	Robustness mode C	119
8.4.4.3.5	Robustness mode D	119
8.5	Control cells	120
8.5.1	General	120
8.5.2	FAC cells	121
8.5.2.1	Cell positions	121
8.5.2.2	Cell gains and phases	122
8.5.3	SDC cells	123
8.5.3.1	Cell positions	123
8.5.3.2	Cell gains and phases	123
8.6	Data cells	123
8.6.1	Cell positions	123
8.6.2	Cell gains and phases	123

Annex A (informative):	Simulated system performance	124
Annex B (informative):	Definition of channel profiles.....	125
Annex C (informative):	Example of mapping of logical frames to multiplex frames	128
Annex D (normative):	Calculation of the CRC word	129
Annex E (informative):	Indicative RF Protection ratios	131
Annex F (informative):	Guidelines for emitter implementation.....	133
Annex G (informative):	Guidelines for receiver implementation	134
G.1	Alternative frequency checking and switching (AFS)	134
G.2	Character sets	136
Annex H (informative):	Service capacity and bit rates	137
Annex I (normative):	Audio tables.....	138
Annex J (informative):	Numbers of input bits.....	141
Annex K (informative):	Simulcast transmission.....	146
Annex L (informative):	Pilot reference illustrations.....	148
Annex M (informative):	MSC configuration examples.....	152
Annex N (informative):	HVXC parameters.....	155
Annex O (informative):	Bibliography.....	157
History		158

Withdrawing

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (<http://www.etsi.org/legal/home.htm>).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Technical Specification (TS) has been produced by Joint Technical Committee (JTC) of the European Broadcasting Union (EBU), Comité Européen de Normalization Electrotechnique (CENELEC) and the European Telecommunications Standards Institute (ETSI).

NOTE: The EBU/ETSI JTC Broadcast was established in 1990 to co-ordinate the drafting of standards in the specific field of broadcasting and related fields. Since 1995 the JTC Broadcast became a tripartite body by including in the Memorandum of Understanding also CENELEC, which is responsible for the standardization of radio and television receivers. The EBU is a professional association of broadcasting organizations whose work includes the co-ordination of its members' activities in the technical, legal, programme-making and programme-exchange domains. The EBU has active members in about 60 countries in the European broadcasting area; its headquarters is in Geneva.

European Broadcasting Union
CH-1218 GRAND SAconnex (Geneva)
Switzerland
Tel: +41 22 717 21 11
Fax: +41 22 717 24 81

Introduction

The frequency bands used for broadcasting below 30 MHz are:

- Low frequency (LF) band - from 148,5 kHz to 283,5 kHz, in ITU Region 1 [1] only;
- Medium frequency (MF) band - from 526,5 kHz to 1 606,5 kHz, in ITU Regions 1 [1] and 3 [1] and from 525 kHz to 1 705 kHz in ITU Region 2 [1];
- High frequency (HF) bands - a set of individual broadcasting bands in the frequency range 2,3 MHz to 27 MHz, generally available on a Worldwide basis.

These bands offer unique propagation capabilities that permit the achievement of:

- Large coverage areas, whose size and location may be dependent upon the time of day, season of the year or period in the (approximately) 11 year sunspot cycle;
- Portable and mobile reception with relatively little impairment caused by the environment surrounding the receiver.

There is thus a desire to continue broadcasting in these bands, perhaps especially in the case of international broadcasting where the HF bands offer the only reception possibilities which do not also involve the use of local repeater stations.

However, broadcasting services in these bands:

- use analogue techniques;
- are subject to limited quality;
- are subject to considerable interference as a result of the long-distance propagation mechanisms which prevail in this part of the frequency spectrum and the large number of users.

As a direct result of the above considerations, there is a desire to effect a transfer to digital transmission and reception techniques in order to provide the increase in quality which is needed to retain listeners who, increasingly, have a wide variety of other programme reception media possibilities, usually already offering higher quality and reliability.

In order to meet the need for a digital transmission system suitable for use in all of the bands below 30 MHz, the Digital Radio Mondiale (DRM) consortium was formed in early 1998. The DRM consortium is a non-profit making body which seeks to develop and promote the use of the DRM system worldwide. Its members include broadcasters, network providers, receiver and transmitter manufacturers and research institutes. More information is available from their website (<http://www.drm.org/>).

Withdrawal

DIGITAL RADIO MONDIALE (DRM) –

System specification for digital transmissions in the broadcasting bands below 30 MHz

1 Scope

The present document gives the specification for the Digital Radio Mondiale (DRM) system for digital transmissions in the broadcasting bands below 30 MHz.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication and/or edition number or version number) or non-specific.
 - For a specific reference, subsequent revisions do not apply.
 - For a non-specific reference, the latest version applies.
- [1] ITU-R Radio Regulations.
 - [2] ISO/IEC 14496-3: "Information technology - Coding of audio-visual objects - Part 3: Audio".
 - [3] ISO/IEC 14496-3/Am1: "Audio extensions".
 - [4] ETSI EN 300 401: "Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers".
 - [5] EN 50067: "Specification of the radio data system (RDS) for VHF/FM sound broadcasting in the frequency range from 87,5 to 108,0 MHz".
 - [6] ISO/IEC 10646-1: "Information technology - Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane".
 - [7] ISO 639-2: "Codes for the representation of names of languages - Part 2: Alpha-3 code".
 - [8] ISO 3166 (all parts): "Codes for the representation of names of countries and their subdivisions".
 - [9] ISO 8859-1: "Information technology - 8-bit single-byte coded graphic character sets - Part 1: Latin alphabet No. 1".
 - [10] ETSI TS 101 756: "Digital Audio Broadcasting (DAB); Registered Tables".
 - [11] ITU-R Recommendation BS.559-2: "Objective measurement of radio-frequency protection ratios in LF, MF and HF broadcasting".
 - [12] ITU-R Recommendation SM.328-10: "Spectra and bandwidth of emissions".

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

cell: sine wave portion of duration T_s , transmitted with a given amplitude and phase and corresponding to a carrier position

NOTE 1: Each OFDM symbol is the sum of K such sine wave portions equally spaced in frequency.

energy dispersal: operation involving deterministic selective complementing of bits in the logical frame, intended to reduce the possibility that systematic patterns result in unwanted regularity in the transmitted signal

Fast Access Channel (FAC): channel of the multiplex data stream which contains the information that is necessary to find services and begin to decode the multiplex

Main Service Channel (MSC): channel of the multiplex data stream which occupies the major part of the transmission frame and which carries all the digital audio services, together with possible supporting and additional data services

mod: the modulo operator

NOTE 2: $(x \bmod y) = z$, where $y > 0$, such that $x = qy + z$, q is an integer, and $0 \leq z < y$.

OFDM symbol: transmitted signal for that portion of time when the modulating amplitude and phase state is held constant on each of the equally-spaced carriers in the signal

reserved for future addition (rfa): bits with this designation shall be set to zero

NOTE 3: Receivers shall ignore these bits.

reserved for future use (rfu): bits with this designation shall be set to zero

NOTE 4: Receivers shall check that these bits are zero in order to determine the valid status of the other fields in the same scope.

Service Description Channel (SDC): channel of the multiplex data stream which gives information to decode the services included in the multiplex

NOTE 5: The SDC also provides additional information to enable a receiver to find alternative sources of the same data.

Single Frequency Network (SFN): network of transmitters sharing the same radio frequency to achieve a large area coverage

transmission frame: a number of consecutive OFDM symbols (duration of 400 ms), whereby the first OFDM symbol contains the time reference cells

transmission super frame: three consecutive transmission frames (duration of 1 200 ms), whereby the first OFDM symbols contain the SDC block

logical frame: contains data of one stream during 400 ms

multiplex frame: logical frames from all streams form a multiplex frame (duration of 400 ms)

NOTE 6: It is the relevant basis for coding and interleaving.

3.2 Symbols

For the purposes of the present document, the following symbols apply:

$E[\]$ expectation value of the expression in brackets

f_c	reference frequency of the emitted signal
K	number of active carriers in the OFDM symbol
$K_{\max}$	carrier index of the upper active carrier in the OFDM signal
$K_{\min}$	carrier index of the lower active carrier in the OFDM signal
L_{MUX}	number of input bits per multiplex frame for the multilevel encoding
N_{MUX}	number of MSC cells (QAM symbols) per multiplex frame
T	elementary time period, equal to $83^{1/3} \mu\text{s}$ (1/12 kHz)
T_f	duration of the transmission frame, equal to 400 ms
T_g	duration of the guard interval
T_s	duration of an OFDM symbol
T_{sf}	duration of the transmission super-frame built from three transmission frames
T_u	duration of the useful (orthogonal) part of an OFDM symbol, excluding the guard interval
X^*	complex conjugate of value X
$\lceil \]$	round towards plus infinity
$\lfloor \]$	round towards minus infinity

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

AAC	Advanced Audio Coding
AFS	Alternative Frequency Switching
BER	Bit Error Ratio
CELP	Code Excited Linear Prediction
CRC	Cyclic Redundancy Check
DFT	Discrete Fourier Transform
EEP	Equal Error Protection
FAC	Fast Access Channel
HF	High Frequency
HVXC	Harmonic Vector eXcitation Coding
IFFT	Inverse Fast Fourier Transform
ISO	International Organization for Standardization
LF	Low Frequency
LPC	Linear Predictive Coding
LSb	Least Significant bit
LSP	Line Spectral Pairs
MF	Medium Frequency
MPEG	Moving Picture Experts Group
MSb	Most Significant bit
MSC	Main Service Channel
OFDM	Orthogonal Frequency Division Multiplexing
PRBS	Pseudo-Random Binary Sequence
QAM	Quadrature Amplitude Modulation
RF	Radio Frequency
rfa	reserved for future addition
rfu	reserved for future use
SBR	Spectral Band Replication
SDC	Service Description Channel
SFN	Single Frequency Network
SM	Simple Modulation
SPP	Standard Protected Part
UEP	Unequal Error Protection
uimsbf	unsigned integer most significant bit first
VSPP	Very Strongly Protected Part

4 General characteristics

4.1 System overview

The DRM system is designed to be used at any frequency below 30 MHz, i.e. within the long wave, medium wave and short wave broadcasting bands, with variable channelization constraints and propagation conditions throughout these bands. In order to satisfy these operating constraints, different transmission modes are available. A transmission mode is defined by transmission parameters classified in two types:

- signal bandwidth related parameters;
- transmission efficiency related parameters.

The first type of parameters defines the total amount of frequency bandwidth for one transmission. Efficiency related parameters allow a trade-off between capacity (useful bit rate) and ruggedness to noise, multipath and Doppler.

4.2 System architecture

This clause gives a general presentation of the system architecture, based on the synoptic diagram of figure 1, which gives reference to the clauses defining the individual parts of the system.

Figure 1 describes the general flow of different classes of information (audio, data, ...) and does not differentiate between different services that may be conveyed within one or more classes of information. A detailed description on the distribution of services onto those classes can be found in clause 6.

Withhold

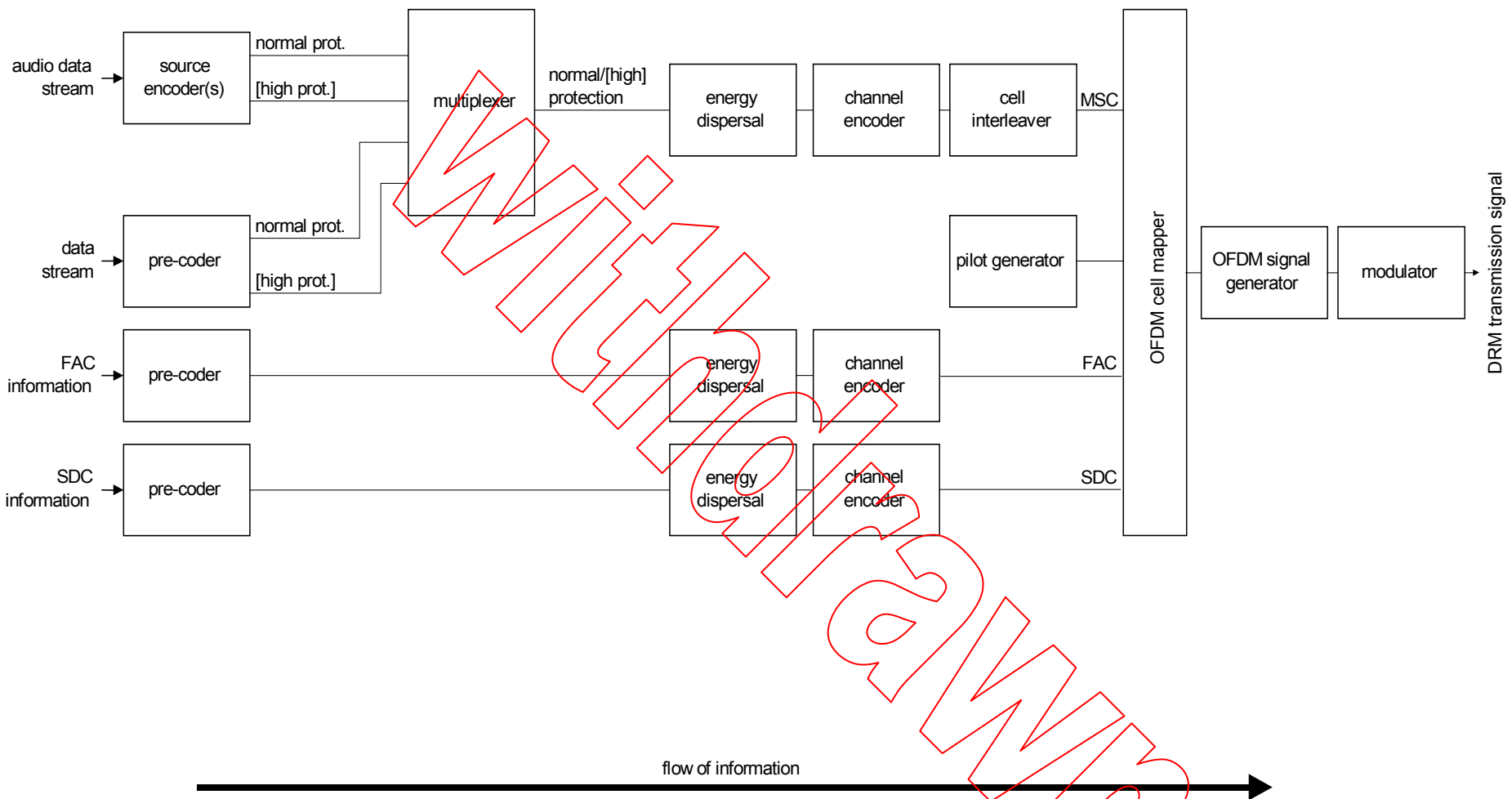


Figure 1: Conceptual DRM transmission block diagram

The source encoder and pre-coders ensure the adaptation of the input streams onto an appropriate digital transmission format. For the case of audio source encoding, this functionality includes audio compression techniques as described in clauses 4.3 and 5. The output of the source encoder(s) and the data stream pre-coder may comprise two parts requiring different levels of protection within the subsequent channel encoder. All services have to use the same two levels of protection.

The multiplexer combines the protection levels of all data and audio services as described in clause 6.

The energy dispersal provides a deterministic selective complementing of bits in order to reduce the possibility that systematic patterns result in unwanted regularity in the transmitted signal.

The channel encoder adds redundant information as a means for quasi error-free transmission and defines the mapping of the digital encoded information onto QAM cells as described in clause 7.

Cell interleaving spreads consecutive QAM cells onto a sequence of cells quasi-randomly separated in time and frequency, in order to provide robust transmission in time-frequency dispersive channels. The pilot generator provides means to derive channel state information in the receiver, allowing for a coherent demodulation of the signal.

The OFDM cell mapper collects the different classes of cells and places them on the time-frequency grid as specified in clause 7.

The OFDM signal generator transforms each ensemble of cells with same time index to a time domain representation of the signal. Consecutively, the OFDM symbol is obtained from this time domain representation by inserting a guard interval as a cyclic repetition of a portion of the signal, as specified in clause 7.

The modulator converts the digital representation of the OFDM signal into the analogue signal in the air. This operation involves digital-to-analogue conversion and filtering that have to comply with spectrum requirements as described in annex E.

4.3 Source coding

Within the constraints of broadcasting regulations in broadcasting channels below 30 MHz and the parameters of the coding and modulation scheme applied, the bit rate available for source coding is in the range from 8 kbit/s (half channels) to ≈ 20 kbit/s (standard channels) to up to ≈ 72 kbit/s (double channels).

In order to offer optimum quality at a given bit rate, the system offers different source coding schemes:

- a subset of MPEG-4 AAC (Advanced Audio Coding) including error robustness tools for generic mono and stereo audio broadcasting;
- a subset of MPEG-4 CELP speech coder for error robust speech broadcasting in mono, for cases when only a low bit rate is available or especially high error robustness is required;
- a subset of MPEG-4 HVXC speech coding for very low bit rate and error robust speech broadcasting in mono, especially well suited also for speech data base applications;
- Spectral Band Replication (SBR), an audio coding enhancement tool that allows to achieve full audio bandwidth at low bit rates. It can be applied to AAC and CELP.

The bit-stream transport format of the source coding schemes has been modified to meet the requirements of the DRM system (audio superframing). Unequal error protection (UEP) can be applied to improve the system behaviour in error prone channels.

Provision is made for further enhancement of the audio system by linking two DRM signals together.

4.4 Transmission modes

4.4.1 Signal bandwidth related parameters

The current channel widths for radio broadcasting below 30 MHz are 9 kHz and 10 kHz. The DRM system is designed to be used:

- within these nominal bandwidths, in order to satisfy the current planning situation;
- within half these bandwidths (4,5 kHz or 5 kHz) in order to allow for simulcast with analogue AM signals;
- within twice these bandwidths (18 kHz or 20 kHz) to provide for larger transmission capacity where and when the planning constraints allow for such facility.

These signal bandwidth related parameters are specified in clause 8.

4.4.2 Transmission efficiency related parameters

For any value of the signal bandwidth parameter, transmission efficiency related parameters are defined to allow a trade off between capacity (useful bit rate) and ruggedness to noise, multipath and Doppler. These parameters are of two types:

- coding rate and constellation parameters, defining which code rate and constellations are used to convey data;
- OFDM symbol parameters, defining the structure of the OFDM symbols to be used as a function of the propagation conditions.

4.4.2.1 Coding rates and constellations

As a function of the desired protection associated within each service or part of a service, the system provides a range of options to achieve one or two levels of protection at a time. Depending on service requirements, these levels of protection may be determined by either the code rate of the channel encoder (e.g. 0,6 ...), by the constellation order (e.g. 4-QAM, 16-QAM, 64-QAM), or by hierarchical modulation. Detailed definition of these options is given in clause 7.

4.4.2.2 OFDM parameter set

The OFDM parameter set is presented in this clause. The specification of the signal waveform is given in clause 8. These values are defined for different propagation-related transmission conditions to provide various robustness modes for the signal. In a given bandwidth, the different robustness modes provide different available data rates. Table 1 illustrates typical uses of the robustness modes.

Table 1: Robustness Mode uses

Robustness Mode	Typical Propagation Conditions
A	Gaussian channels, with minor fading
B	Time and frequency selective channels, with longer delay spread
C	As robustness mode B, but with higher Doppler spread
D	As robustness mode B, but with severe delay and Doppler spread

The transmitted signal comprises a succession of OFDM symbols, each symbol being made of a guard interval followed by the so-called useful part of the symbol. Each symbol is the sum of K sine wave portions equally spaced in frequency. Each sine wave portion, called a "cell", is transmitted with a given amplitude and phase and corresponds to a carrier position. Each carrier is referenced by the index k , k belonging to the interval $[k_{\min}, k_{\max}]$ ($k = 0$ corresponds to the reference frequency of the transmitted signal).

The time-related OFDM symbol parameters are expressed in multiples of the elementary time period T , which is equal to $83^{1/3} \mu\text{s}$. These parameters are:

- T_g : duration of the guard interval;
- T_s : duration of an OFDM symbol;
- T_u : duration of the useful (orthogonal) part of an OFDM symbol (i.e. excluding the guard interval).

The OFDM symbols are grouped to form transmission frames of duration T_f .

As specified in clause 8, a certain number of cells in each OFDM symbol are transmitted with a predetermined amplitude and phase, in order to be used as references in the demodulation process. They shall be called "reference pilots" and represent a certain proportion of the total number of cells.

Table 2: OFDM symbol parameters

Parameters list	Robustness Mode			
	A	B	C	D
T (μs)	$83^{1/3}$	$83^{1/3}$	$83^{1/3}$	$83^{1/3}$
T_u (ms)	24 ($288 \times T$)	$21^{1/3}$ ($256 \times T$)	$14^{2/3}$ ($176 \times T$)	$9^{1/3}$ ($112 \times T$)
T_g (ms)	$2^{2/3}$ ($32 \times T$)	$5^{1/3}$ ($64 \times T$)	$5^{1/3}$ ($64 \times T$)	$7^{1/3}$ ($88 \times T$)
T_g / T_u	1/9	1/4	4/11	11/14
$T_s = T_u + T_g$ (ms)	$26^{2/3}$	$26^{2/3}$	20	$16^{2/3}$
T_f (ms)	400	400	400	400

5 Source coding modes

5.1 Overview

The source coding options in the DRM system are shown in figure 2.

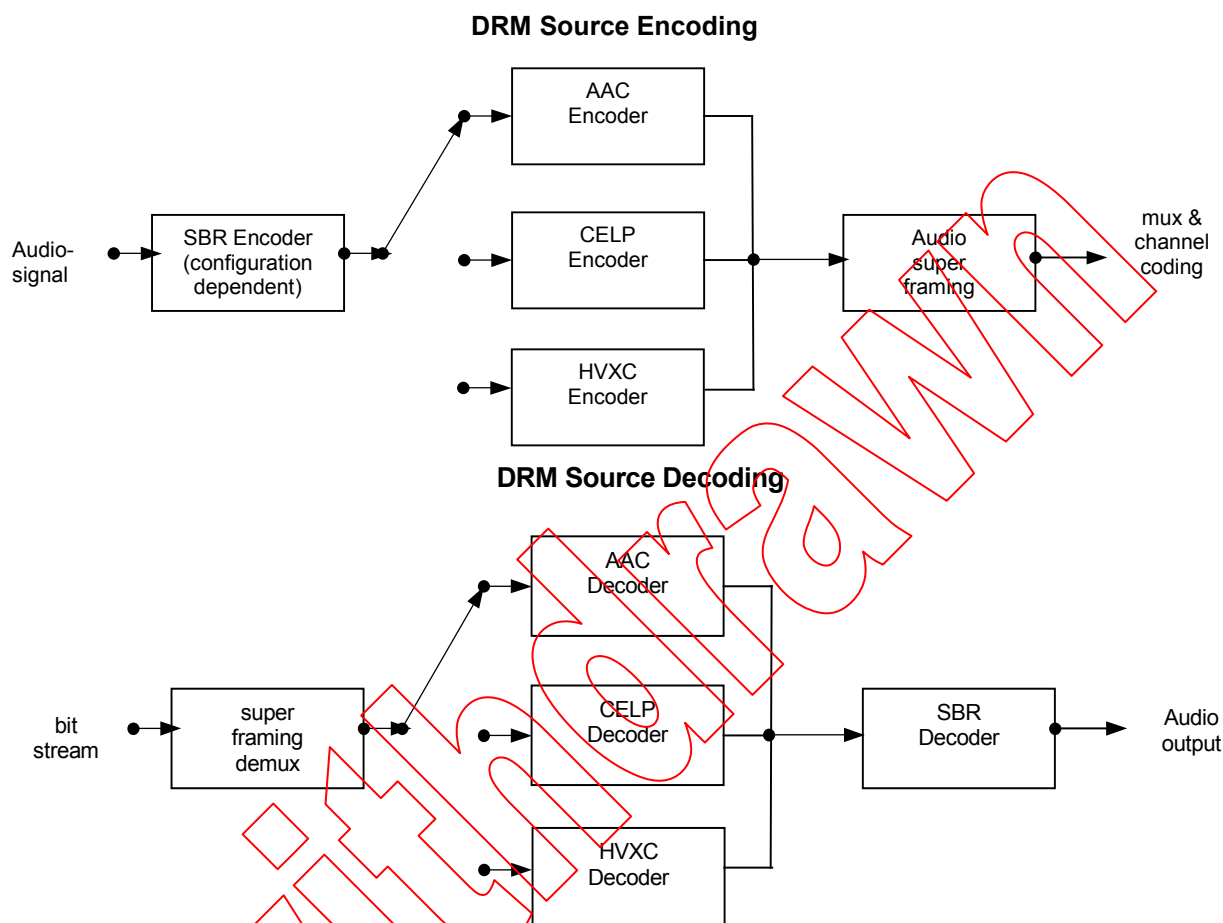


Figure 2: Source coding overview

As described in clause 4.3, the DRM system offers audio coding (AAC) and speech coding (CELP and HVXC). In general, a high frequency reconstruction method (SBR) can be used to enhance the perceptual audio quality. However, at present the use of SBR is only defined for use with AAC. Special care is taken so that the encoded audio can be composed into audio super frames of constant length. Multiplexing and UEP of audio/speech services is done by means of the multiplex and channel coding units. Audio specific configuration information is transmitted in the SDC (see clause 6.4.3.10).

5.1.1 AAC Audio Coding

For generic audio coding, a subset of the MPEG-4 Advanced Audio Coding (AAC) toolbox chosen to best suit the DRM system environment is used. For example a standard configuration for use in one short wave channel could be 20 kbit/s mono AAC.

Specific features of the AAC stream within the DRM system are:

- Bit rate: AAC can be used at any bit rate. Byte-alignment of the 400 ms audio super frame leads to a granularity of 20 bit/s for the AAC bit rate;
- Sampling rates: permitted sampling rates are 12 kHz and 24 kHz;
- Transform length: the transform length is 960 to ensure that one audio frame corresponds to 80 ms or 40 ms in time. This is required to harmonize CELP and AAC frame lengths and thus to allow the combination of an integer number of audio frames to build an audio super frame of 400 ms duration;
- Error resilience: a subset of MPEG-4 tools is used to improve the AAC bit stream error resilience in error prone channels;
- Audio super framing: 5 (12 kHz) or 10 (24 kHz) audio frames are composed into one audio super frame, which always corresponds to 400 ms in time. The audio frames in the audio super frames are encoded together such that each audio super frame is of constant length, i.e. that bit exchange between audio frames is only possible within an audio super frame. One audio super frame is always placed in one logical frame (see clause 6). In this way no additional synchronization is needed for the audio coding. Retrieval of frame boundaries and provisions for UEP are also taken care of within the audio super frame;
- UEP: better graceful degradation and better operation at higher BERs is achieved by applying UEP to the AAC bit stream. Unequal error protection is realized via the multiplex/coding units.

5.1.2 MPEG CELP coding

MPEG CELP speech coding is offered in the DRM system to allow for reasonable speech quality at bit rates significantly below the standard rate (for example "half rate" operation at 8 kbit/s). Possible scenarios for the use of the speech coder are:

- Dual/triple speech applications: instead of one audio program at 20 kbit/s to 24 kbit/s, the channel contains two or three speech signals of 8 kbit/s to 10 kbit/s each, allowing simultaneous speech transmissions;
- Speech services in addition to an audio service;
- Simulcast transmissions: in case of analogue/digital simulcast only bit rates as low as 8 kbit/s may be available;
- Very robust speech applications: due to its nature a speech coder can be expected to offer higher robustness against channel errors. Therefore 8 kbit/s speech coding can be used to do ultra robust speech coding in one channel.

Basic features of MPEG CELP coding are:

- 8 kHz or 16 kHz sampling rate;
- Bit rates between 4 kbit/s and 20 kbit/s;
- Error robustness;
- Composition of an integer number of CELP frames to build one audio super frame.

5.1.3 MPEG HVXC coding

MPEG-4 HVXC (Harmonic Vector eXcitation Coding) speech coding is offered in the DRM system to allow for reasonable speech quality at very low bit rates such as 2,0 kbit/s. The operating bit rates of HVXC open up new applications for DRM such as:

- Speech services in addition to an audio service;
- Multi-language application;
- Solid-state storage of multiple programs such as news, data base in a card radio (e.g. total of about 4,5 hours of radio programs can be stored in 4 MByte Flash memory);
- Time scale modification for fast playback/browsing of stored program;
- Highly error robust transmission with or without hierarchical modulation scheme.

Basic features of HVXC coding are:

- 8 kHz sampling rate;
- Bit rates of 2,0 kbit/s and 4,0 kbit/s for fixed rate coding;
- Time scale and pitch modification of arbitrary amounts;
- Error robust syntax is supported, and a CRC tool can be used to improve the error resilience of the HVXC bitstream in error prone channels;
- Composition of a constant integer number of HVXC frames (20) to build one audio super frame.

5.1.4 SBR coding

To maintain a reasonable perceived audio quality at low bit rates, classical audio or speech source coding algorithms need to limit the audio bandwidth and to operate at low sampling rates. It is desirable to be able to offer high audio bandwidth also in very low bit rate environments. This can be realized by the use of SBR (Spectral Band Replication).

The purpose of SBR is to recreate the missing high frequency band of the audio signal that could not be coded by the encoder. In order to do this in the best possible way, some side information needs to be transmitted in the audio bitstream, removing a small percentage of the available data rate from the audio coder. This side information is computed on the full bandwidth signal, prior to encoding and aids the reconstruction of the high frequencies after audio/speech decoding.

SBR exists in two versions. SBR-LC, a tool of low complexity, that offers intermediate sound quality, and SBR-HQ a tool that offers higher sound quality compared to SBR-LC, albeit at a somewhat higher complexity. Both versions are encoder and bitstream compatible and thus offer a "future-proof" upgrade concept. The version difference is only reflected in the decoder design.

The SBR technique is described in detail in clause 5.6.

5.2 UEP and audio super framing

Today's coding schemes are highly optimized in terms of coding efficiency and according to information theory this should lead to the fact, that the entropy of the bits is nearly equal. If this assumption is true, then the channel coding must be optimized, such that the total amount of residual errors usually referred to as bit error rate (BER) is minimized. This criteria can be fulfilled by a channel coding method called equal error protection (EEP), were all information bits are protected with the same amount of redundancy.

However, the audible effects of errors are not independent of the part of the bitstream that was hit by the error. This behaviour of unequal error sensitivity is well known for source coding schemes that are used in broadcast and communication systems, like Eureka DAB or GSM. The optimized solution to cope with this unequal error sensitivity is called unequal error protection (UEP). In such a system, higher protection is assigned to the more sensitive information, whereas lower protection is assigned to the less sensitive part of the bitstream.

To accommodate for UEP channel coding, it is necessary to have frames with a constant length and a UEP profile that is constant as well for a given bit rate. Since AAC is a coding scheme with a variable length, several coded frames are grouped together to build one audio super frame. The bit rate of the audio super frame is constant. Since the channel coding is based on audio super frames, the audio super frames itself consist of two parts: a higher protected part and a lower protected part. Therefore, the coded audio frames itself have to be split into these two part. Further details on the audio super frame structure of AAC, CELP and HVXC is provided in the subsequent clauses. Note that HVXC is intended for use with the EEP scheme only.

Table 3: Syntax of audio_super_frame()

Syntax	No. of bits	Note
<pre> Audio_super_frame(audio_info) // audio info from the SDC { switch (audio_info.audio_coding) { case AAC: aac_super_frame(audio_info) break; case CELP: celp_super_frame(audio_info) break; case HVXC: hvxc_super_frame(audio_info) break; } } </pre>		

NOTE: The SDC describes the audio coder used, and the parameters associated with that coder. It also provides information about the sampling rate and bit rate used (see clause 6).

5.3 AAC coding

The following two clauses explain how the AAC and AAC + SBR frames fit into the audio super frame.

5.3.1 AAC

ISO/IEC 14496-3 [2] together with ISO/IEC 14496-3/Amd 1 [3] define the MPEG-4 Audio standard. The audio coding standard MPEG-4 AAC is part of the MPEG-4 Audio standard. Two versions are defined, but only version 2 is intended for the use in error prone channels. The AAC bitstreams in the DRM system are therefore MPEG-4 version 2 bitstreams. From the possible audio object types, only the Error Resilient (ER) AAC Low Complexity object type (Object Type ID = 17), which is part of the High Quality Audio Profile, will be used in the DRM system.

DRM specific usage of MPEG-4 AAC: Three error resilience tools may be used within an MPEG-4 version 2 AAC bitstream: HCR (Huffman Codeword Reordering), VCB11 (Virtual Codebooks for Codebook 11) and RVLC (Reversible Variable Length Coding). In the DRM system, all AAC bitstreams shall use the HCR tool, since this tool reduces the error sensitivity of the bitstream significantly with a minimum of overhead. The VCB11 tool shall be used, since for low bit rates, the VCB11 overhead is less than 1 %. The RVLC tool shall not be used, since it introduces a significant bit rate overhead that would be a major drawback for the low bit rates used by DRM.

The MPEG-4 AAC tool PNS (Perceptual Noise Substitution) shall not be used in DRM since SBR provides this functionality more appropriately.

For DRM the 960 transform shall be used.

When 12 kHz sampling is used, 5 AAC frames shall be combined into one audio super frame.

When 24 kHz sampling is used, 10 AAC frames shall be combined into one audio super frame.

The AAC sampling frequency shall be 24 kHz when the stereo mode is used.

No standard extension_payload() shall be used and the only allowed extension is SBR (signalled via SDC).

The left and the right channel in one stereo audio frame are transmitted in an interleaved way to achieve a decreasing error sensitivity within the stereo frame.

The element_instance_tag for a single_channel_element(), respectively a channel_pair_element(), shall not be used in order to save 4 bits.

Any DRM AAC bitstream can easily be translated into an MPEG-4 V2 compliant bitstream by applying the above rules.

The MPEG-4 standard defines how the bits for one raw error resilient AAC audio frame are stored. Each element of the error resilient AAC bitstream is assigned an error sensitivity category. In the DRM system there are two possible error resilient AAC audio frames:

mono audio frame

One mono audio frame consists of three consecutive parts, hereinafter called mono1, mono2 and mono3. Mono1 contains the side information (SI) bits, mono2 contains the TNS bits and mono3 contains the spectral data bits. The error sensitivity decreases from mono1 to mono3.

stereo audio frame

One stereo audio frame consists of seven consecutive parts, hereinafter called stereo1 (common side info), stereo2 (side info left channel), stereo3 (side info right channel), stereo4 (TNS left channel), stereo5 (TNS right channel), stereo6 (spectral data left channel), stereo7 (spectral data right channel). With this interleaving of left and right channel, the error sensitivity is decreasing from stereo1 to stereo7.

5.3.1.1 AAC audio super frame

Table 4: Syntax of aac_super_frame()

Syntax	No. of bits	Note
<pre> aac_super_frame(audio_info) // audio info from the SDC { switch (audio_info.audio_sampling_rate) { // only 12 000 and 24 000 is allowed case 12000: num_frames = 5; break; case 24000: num_frames = 10; break; } aac_super_frame_header(num_frames - 1) for (f = 0; f < num_frames; f++) { // higher_protected_block for (b = 0; b < num_higher_protected_bytes; b++) audio_frame[f][b] aac_crc_bits[f] } // lower_protected_part for (f = 0; f < num_frames; f++) { num_lower_protected_bytes = frame_length[f] - num_higher_protected_bytes; for (b = 0; b < num_lower_protected_bytes; b++) audio_frame[f][num_higher_protected_bytes + b] } } } </pre>	8 8 8	see annex D
NOTE 1: num_higher_protected_bytes is derived from the UEP profile used (see clause 6). NOTE 2: audio_frame is either an AAC or an AAC + SBR frame.		

Table 5: Syntax of aac_super_frame_header()

Syntax	No. of bits	Note
<pre> aac_super_frame_header(num_borders) { previous_border = 0; for (n = 0; n < num_borders; n++) { frame_length[n] = frame_border - previous_border; // frame border in bytes previous_border = frame_border; } frame_length[num_borders] = audio_payload_length - previous_border; if (num_borders == 9) reserved // byte-alignment } </pre>	12	
	4	
NOTE: The audio_payload_length is derived from the length of the audio super frame (data_length_of_part_A + data_length_of_part_B) subtracting the audio super frame overhead (bytes used for the audio super frame header() and for the aac_crc_bits).		

higher protected part

The higher protected part contains one header followed by num_frames higher protected blocks. num_frames is the number of audio frames in the audio super frame.

header

The header contains information to recover the frame lengths of the num_frames AAC frames stored in the audio super frame.

All the frame lengths are derived from the absolute positions of the frame borders. These frame borders are stored consecutively in the header. Each frame border occupies 12 bits (unsigned integer, most significant bit first). The frame border is measured in bytes from the start of the AAC bitstream sequence. 4 padding bits are added in case num_frames==10. num_frames-1 frame borders are stored in the header.

higher protected block

One higher protected block contains a certain amount of bytes from the start of each AAC frame, dependent upon the UEP profile. One 8-bit CRC check derived from the CRC-bits of the corresponding AAC frame follows (see annex D for CRC calculation). For a mono signal, the CRC-bits cover (mono1, mono2). For a stereo signal, the CRC-bits cover (stereo1, stereo2, stereo3, stereo4, stereo5).

lower protected part

The lower protected bytes (the remaining bytes not stored in the higher protected part) of the AAC frames are stored consecutively in the lower protected part.

Figure 3 illustrates an example audio super frame for a 24 kHz sampled signal.

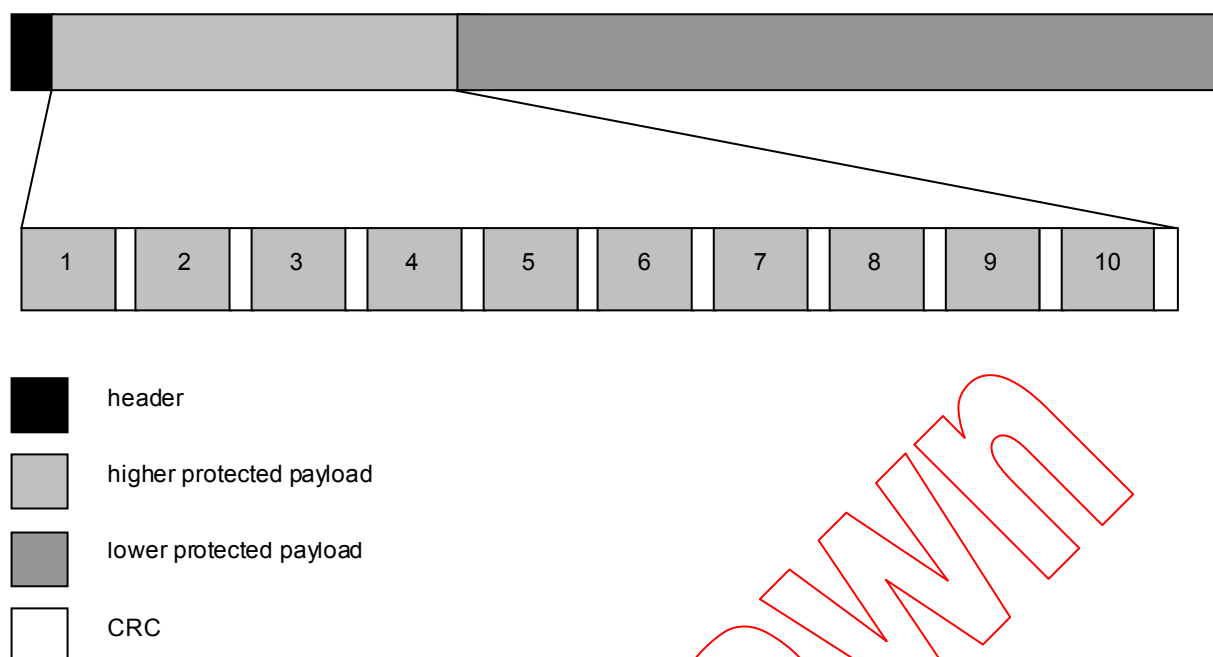


Figure 3: Example AAC audio super frame (24 kHz)

5.3.2 AAC + SBR

The SBR sampling frequency shall be 48 kHz and the AAC sampling frequency shall be 24 kHz. One raw AAC + SBR frame contains an AAC part and a SBR part. The SBR part of the data is located at the end of the frame. The first bit in the SBR-bitstream is the last bit in the frame, and the SBR bits are thus written/read in reversed order. In this way, the starting points of respective part of the frame data are always easily found.

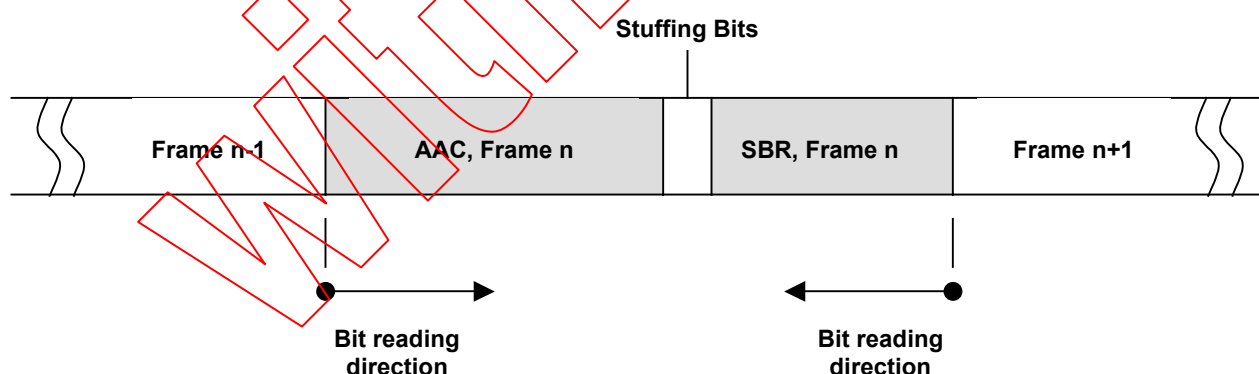


Figure 4: AAC + SBR frame

Both AAC and SBR data-sizes vary from frame to frame. The total size of the individual frames, now including the SBR data, can be derived from the `aac_super_frame_header()` as described in clause 5.3.1. Thus no extra signalling due to the varying SBR bit rate is needed.

The AAC + SBR frames are inserted into the audio super frame in the same manner as when SBR is not used.

For source coding bit rates at 20 kbit/s or greater, SBR shall be used. For bit rates below 20 kbit/s, SBR may be used.

The details of the SBR-bitstream are described in clause 5.5.3.

5.4 MPEG CELP coding

5.4.1 MPEG CELP

ISO/IEC 14496-3 [2] together with ISO/IEC 14496-3/Amd 1[3] define the MPEG-4 Audio standard. The speech coding standard MPEG-4 CELP (Code Excited Linear Prediction) is part of the MPEG-4 Audio standard. Two versions are defined, but only version 2 is intended for the use in error prone channels. The CELP bitstreams in the DRM system are therefore MPEG-4 version 2 bitstreams. From the possible audio object types, only the Error Resilient (ER) CELP object type (Object Type ID = 24), which is part of the High Quality Audio Profile, will be used in the DRM system.

The MPEG-4 CELP covers the compression and decoding of natural speech sound at bit rates ranging between 4 kbit/s and 24 kbit/s. MPEG-4 CELP is a well-known coding algorithm with new functionality. Conventional CELP coders offer compression at a single bit rate and are optimized for specific applications. Compression is one of the functionalities provided by MPEG-4 CELP, but MPEG-4 also enables the use of one basic coder in multiple applications. It provides scalability in bit rate and bandwidth, as well as the ability to generate bitstreams at arbitrary bit rates. The MPEG-4 CELP coder supports two sampling rates, namely, 8 kHz and 16 kHz. The associated bandwidths are 100 Hz to 3 800 Hz for 8 kHz sampling and 50 Hz to 7 000 Hz for 16 kHz sampling.

A basic block diagram of the CELP decoder is given in figure 5.

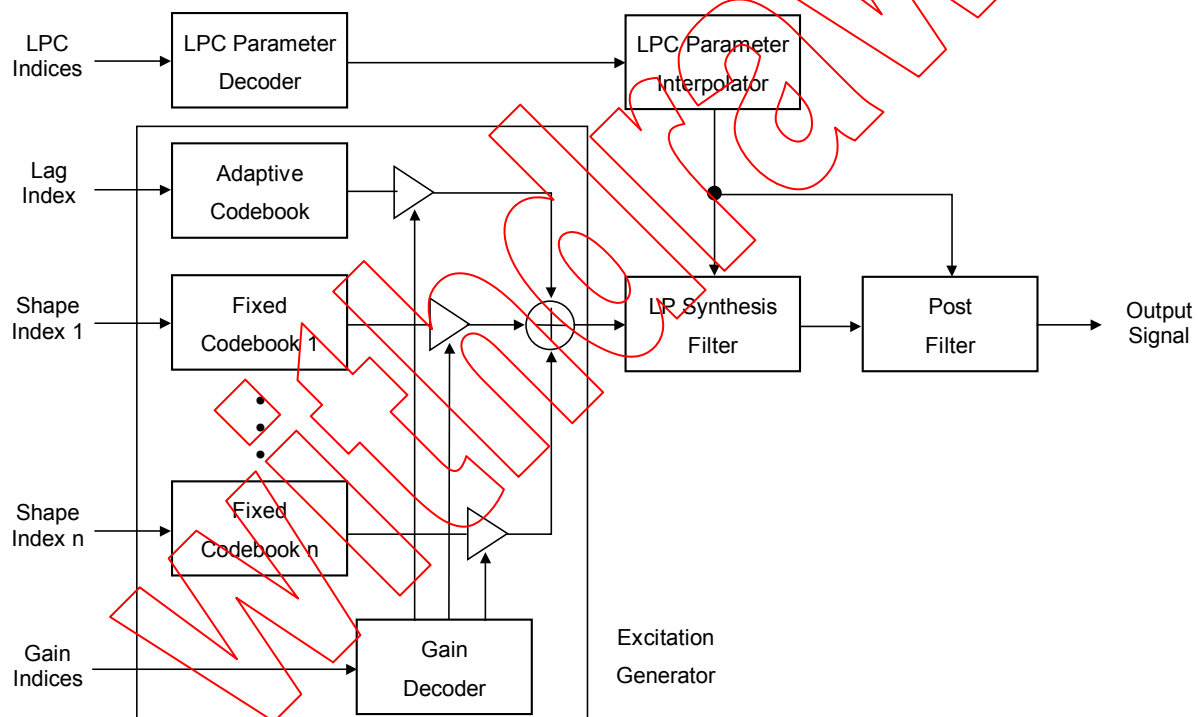


Figure 5: Block diagram of a CELP decoder

The CELP decoder primarily consists of an excitation generator and a synthesis filter. Additionally, CELP decoders often include a post-filter. The excitation generator has an adaptive codebook to model periodic components, fixed codebooks to model random components and a gain decoder to represent a speech signal level. Indices for the codebooks and gains are provided by the encoder. The codebook indices (pitch-lag index for the adaptive codebook and shape index for the fixed codebook) and gain indices (adaptive and fixed codebook gains) are used to generate the excitation signal. It is then filtered by the linear predictive synthesis filter (LP synthesis filter). Filter coefficients are reconstructed using the LPC indices, then are interpolated with the filter coefficients of successive analysis frames. Finally, a post-filter can optionally be applied in order to enhance the speech quality.

The MPEG-4 CELP coder offers the following functionalities: Multiple bit rates, Bit rate Scalability, Bandwidth Scalability, and FineRate Control. DRM shall only use the multiple bit rates functionality.

Multiple bit rates: the available bit rates depend on the sampling rate. The following fixed bit rates can be used:

Table 6: Fixed bit rates for the CELP coder

Bit rates for the 8 kHz sampling rate (bit/s)	Bit rates for the 16 kHz sampling rate (bit/s)
3 850, 4 250, 4 650, 5 700, 6 000, 6 300, 6 600, 6 900, 7 100, 7 300, 7 700, 8 300, 8 700, 9 100, 9 500, 9 900, 10 300, 10 500, 10 700, 11 000, 11 400, 11 800, 12 000, 12 200	10 900, 11 500, 12 100, 12 700, 13 300, 13 900, 14 300, 14 700, 15 900, 17 100, 17 900, 18 700, 19 500, 20 300, 21 100, 13 600, 14 200, 14 800, 15 400, 16 000, 16 600, 17 000, 17 400, 18 600, 19 800, 20 600, 21 400, 22 200, 23 000, 23 800

The algorithmic delay of the CELP coder comes from the frame length and an additional look ahead length. The frame length depends on the coding mode and the bit rate. The look ahead length, which is an informative parameter, also depends on the coding mode. The delays presented below are applicable to the modes used in DRM.

Table 7: Delay and frame length for the CELP coder at 8 kHz sampling rate

Bit rate (bit/s)	Delay (ms)	Frame Length (ms)
3 850, 4 250, 4 650	45	40
5 700, 6 000, 6 300, 6 600, 6 900, 7 100, 7 300, 7 700, 8 300, 8 700, 9 100, 9 500, 9 900, 10 300, 10 500, 10 700	25	20
11 000, 11 400, 11 800, 12 000, 12 200	15	10

Table 8: Delay and frame length for the CELP coder at 16 kHz sampling rate

Bit rate (bit/s)	Delay (ms)	Frame Length (ms)
10 900, 11 500, 12 100, 12 700, 13 300, 13 900, 14 300, 14 700, 15 900, 17 100, 17 900, 18 700, 19 500, 20 300, 21 100	25	20
13 600, 14 200, 14 800, 15 400, 16 000, 16 600, 17 000, 17 400, 18 600, 19 800, 20 600, 21 400, 22 200, 23 000, 23 800	15	10

Withheld

5.4.1.1 CELP audio super frame

CELP frames have a fixed frame length. The CELP audio frames are grouped together to form audio super frames of 400 ms duration. UEP is applicable. The start of each audio frame is mapped into the higher protected part, the remaining bits are allocated to the lower protected part. The partitioning of the CELP frames is given in tables 10 and 11. The CELP bit rate index is signalled in the SDC.

Table 9: Syntax of celp_super_frame()

Syntax	No. of bits	Note
<pre> celp_super_frame(celp_table_ind) // CELP table index from the SDC { switch (audio_info.audio_sampling_rate) { // only 8 000 and 16 000 is allowed case 8000: (num_frames, num_higher_protected_bits, num_lower_protected_bits) = read_table_10 (CELP_index) break; case 16000: (num_frames, num_higher_protected_bits, num_lower_protected_bits) = read_table_11 (CELP_index) break; } for (f = 0; f < num_frames; f++) { // higher_protected_block for (b = 0; b < num_higher_protected_bits; b++) celp_frame[f][b] if (audio_info.CELP_CRC == 1) celp_crc_bits[f] } // lower_protected_part for (f = 0; f < num_frames; f++) { for (b = 0; b < num_lower_protected_bits; b++) celp_frame[f][num_higher_protected_bits + b] } } </pre>	1 8 1	see annex D

Table 10: UEP parameters for 8 kHz sampling CELP

CELP bit rate index	Bit rate (bits/s)	Audio frame length (ms)	Higher Protected Part (bits/audio frame)	Lower Protected Part (bits/audio frame)	Higher Protected Part (bytes/audio super frame)	Lower Protected Part (bytes/audio super frame)	Audio super frame length (bytes)
0	3 850	40	36	118	45	148	193 (see note)
1	4 250	40	36	134	45	168	213 (see note)
2	4 650	40	36	150	45	188	233 (see note)
6	5 700	20	24	90	60	225	285
7	6 000	20	24	96	60	240	300
8	6 300	20	24	102	60	255	315
9	6 600	20	24	108	60	270	330
10	6 900	20	24	114	60	285	345
11	7 100	20	24	118	60	295	355
12	7 300	20	24	122	60	305	365
13	7 700	20	36	118	90	295	385
14	8 300	20	36	130	90	325	415
15	8 700	20	36	138	90	345	435
16	9 100	20	36	146	90	365	455
17	9 500	20	36	154	90	385	475
18	9 900	20	36	162	90	405	495
19	10 300	20	36	170	90	425	515
20	10 500	20	36	174	90	435	525
21	10 700	20	36	178	90	445	535
22	11 000	10	24	86	120	430	550
23	11 400	10	24	90	120	450	570
24	11 800	10	24	94	120	470	590
25	12 000	10	24	96	120	480	600
26	12 200	10	24	98	120	490	610

NOTE: For these bit rates, the last four bits of the audio super frame are padded with 0s.

Table 11: UEP parameters for 16 kHz sampling CELP

CELP bit rate index	Bit rate (bits/s)	Audio frame length (ms)	Higher Protected Part (bits/audio frame)	Lower Protected Part (bits/audio frame)	Higher Protected Part (bytes/audio super frame)	Lower Protected Part (bytes/audio super frame)	Audio super frame length (bytes)
0	10 900	20	64	154	160	385	545
1	11 500	20	64	166	160	415	575
2	12 100	20	64	178	160	445	605
3	12 700	20	64	190	160	475	635
4	13 300	20	64	202	160	505	665
5	13 900	20	64	214	160	535	695
6	14 300	20	64	222	160	555	715
8	14 700	20	92	202	230	505	735
9	15 900	20	92	226	230	565	795
10	17 100	20	92	250	230	625	855
11	17 900	20	92	266	230	665	895
12	18 700	20	92	282	230	705	935
13	19 500	20	92	298	230	745	975
14	20 300	20	92	314	230	785	1 015
15	21 100	20	92	330	230	825	1 055
16	13 600	10	50	86	250	430	680
17	14 200	10	50	92	250	460	710
18	14 800	10	50	98	250	490	740
19	15 400	10	50	104	250	520	770
20	16 000	10	50	110	250	550	800
21	16 600	10	50	116	250	580	830
22	17 000	10	50	120	250	600	850
24	17 400	10	64	110	320	550	870
25	18 600	10	64	122	320	610	930
26	19 800	10	64	134	320	670	990
27	20 600	10	64	142	320	710	1 030
28	21 400	10	64	150	320	750	1 070
29	22 200	10	64	158	320	790	1 110
30	23 000	10	64	166	320	830	1 150
31	23 800	10	64	174	320	870	1 190

5.5 HVXC

The MPEG-4 HVXC (Harmonic Vector eXcitation Coding) speech coding toolset as defined in ISO/IEC 14496-3 [2] covers the compression and decoding of natural speech sound at bit rates of 2,0 kbit/s and 4,0 kbit/s. HVXC employs harmonic coding of LPC residual signals for voiced segments and Vector eXcitation Coding (VXC) for unvoiced segments. HVXC provides communications-quality to near-toll-quality speech in the 100 Hz to 3 800 Hz band at 8 kHz sampling rate. In addition, the functionality of pitch and speed change during decoding is supported. This functionality is useful for fast speech database search or browsing. HVXC has a syntax providing error sensitivity categories that can be used with an error protection tool. Additionally the error concealment functionality is supported for the use in error-prone channels.

DRM uses a subset of the HVXC description in ISO/IEC 14496-3 [2], limiting the syntax to the error resilient syntax given in ISO/IEC 14496-3/Amd 1 [3] and the data rates to the two options of 2,0 kbit/s and 4,0 kbit/s. Further, HVXC is used with the non-scalable syntax only. For robust decoding in error-prone channels a low-complexity error concealment tool (CRC and intra-frame interleaving) is defined specifically for DRM.

The syntax of the HVXC audio super frame is identical for all possible HVXC modes, since HVXC does not support UEP functionality and the length of a HVXC audio frame is always 20 ms.

Table 12: Syntax of hvxc_super_frame()

Syntax	No. of bits	Note
<pre> hvxc_super_frame(audio_info) { num_frames = 20; } </pre>	// audio info from the SDC	

The number of bits contained in one audio frame is given by the audio information from the SDC (HVXC_rate, HVXC_CRC).

In case the 4,0 kbit/s fixed rate HVXC coder is used with the CRC tool, the last 4 bits of each audio super frame are padded with zeros and the receiver shall ignore these bits. The resulting bit rate therefore is 4,66 kbit/s.

Only the fixed rate modes of HVXC (2,0 kbit/s or 4,0 kbit/s) shall be used in audio super frames. Variable rate modes shall only be applicable for use with packet mode applications and the variable bit rate syntax will be defined in conjunction with such data applications.

5.5.1 Definitions

5.5.1.1 HVXC source coder parameters

The definition of the basic data entities of a MPEG-4 compliant HVXC speech coding system is given in ISO/IEC 14496-3 [2] and is reproduced as table N.1 in annex N.

5.5.1.2 CRC bits for fixed bit rate modes

Table 13 describes the various CRC bits added for error protection in the DRM system.

Table 13: CRC bits for fixed bit rate modes

Parameter	Description	length (bits)
CRC0_2k	CRC bits for ESC0 at 2 kbit/s	6
CRC1_2k	CRC bits for ESC1 at 2 kbit/s	1
CRC2_2k	CRC bits for ESC2 at 2 kbit/s	1
CRC0_4k	CRC bits for ESC0 at 4 kbit/s	6
CRC1_4k	CRC bits for ESC1 at 4 kbit/s	5
CRC2_4k	CRC bits for ESC2 at 4 kbit/s	1
CRC3_4k	CRC bits for ESC3 at 4 kbit/s	1

5.5.2 HVXC decoder

Figure 6 shows the overall structure of the HVXC decoder. The basic decoding process is composed of four steps; de-quantization of parameters, generation of excitation signals for voiced frames by sinusoidal synthesis (harmonic synthesis) and noise component addition, generation of excitation signals for unvoiced frames by codebook look-up, and LPC synthesis. To enhance the synthesized speech quality spectral post-filtering is used.

For voiced frames, a fixed dimension harmonic spectral vector, obtained by de-quantization of the spectral magnitude, is first converted to the one having the original dimension which varies frame by frame in accordance with the pitch value. This is done by the dimension converter in which a band-limited interpolator generates a set of spectral magnitude values at harmonic frequencies without changing the shape of the spectral envelope. Using the spectral magnitude values, a time domain excitation signal is then generated by the fast harmonic synthesis algorithm using an IFFT. In order to make the synthesized speech sound natural, a noise component is additionally used. A Gaussian noise spectral component, covering 2 kHz to 3,8 kHz, is coloured in accordance with the harmonic spectral magnitudes in the frequency domain, and its IDFT is added to voiced excitation signals in the time domain. The amount and bandwidth of this additive noise is controlled by the transmitted two-bit V/UV value, which is encoded based on the normalized maximum autocorrelation of the LPC residual signal. Noise added harmonic excitation signals for voiced segments are then fed into the LPC synthesis filter followed by the postfilter.

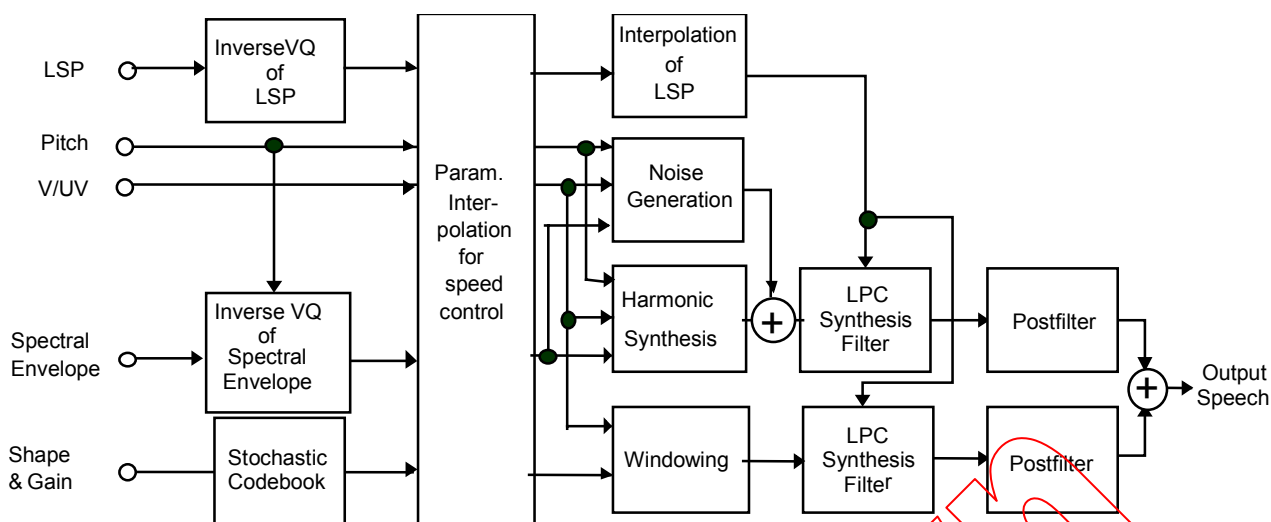


Figure 6: Block diagram of the HVXC decoder

For unvoiced segments, the usual VXC decoding algorithm is used where an excitation signal is generated by multiplying the gain value with the stochastic codevector. The result is then fed into the LPC synthesis filter followed by the postfilter. Finally, the synthesized speech components for voiced and unvoiced segments are added to form the output signal. The description of the time-scale modification standard can be found in the MPEG-4 [2].

5.5.3 HVXC encoder

Figure 7 shows the overall structure of the encoder of the MPEG-4 HVXC. Table N.2 of annex N shows the bit allocations for the 2,0 kbit/s and 4,0 kbit/s coders using fixed rate coding. The parameters followed by (enh) are used only for the 4,0 kbit/s mode. Operation of each part is described below.

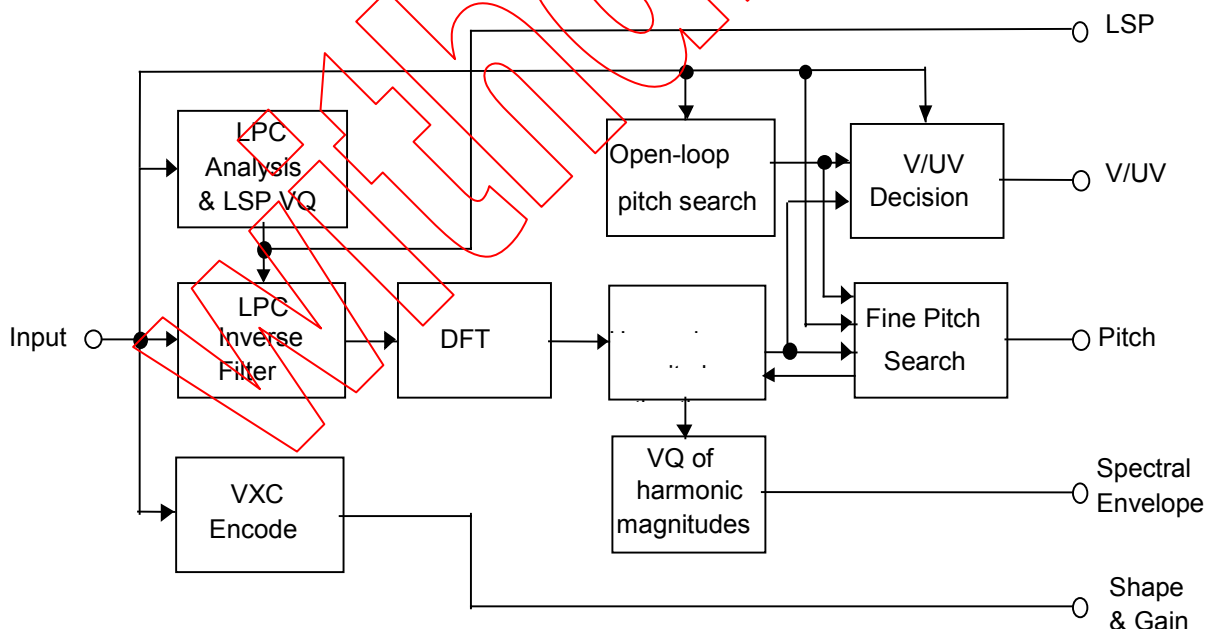


Figure 7: Block diagram of the HVXC encoder

5.5.3.1 LPC analysis and LSP quantization

The speech input at a sampling rate of 8 kHz is formed into frames with a length and interval of 256 samples and 160 samples, respectively. Tenth order LPC analysis is carried out using windowed input data over one frame. LPC parameters are converted to LSP parameters and vector quantized with a partial prediction and multi-stage vector quantization scheme. LPC residual signals are computed by inverse filtering the input data using quantized and interpolated LSP parameters.

5.5.3.2 Open loop pitch search

The open loop pitch value is estimated based on the peak values of the autocorrelation of the LPC residual signals. Using estimated past and current pitch values, pitch tracking is conducted to have a continuous pitch contour and to make the reliability of the pitch estimation higher. The voiced/unvoiced decision of the previous frame is also used to ensure the pitch tracking operation.

5.5.3.3 Harmonic magnitude and fine pitch estimation

The power spectrum of the LPC residual signal is then fed into the fine pitch and harmonic magnitude estimation block, where the harmonic spectral envelope of the LPC residual signal is estimated as follows: A basis spectrum representing one harmonic spectrum is gain scaled and arranged with the spacing of the fundamental frequency obtained by the open loop pitch search. The gain scaling for each harmonic and fundamental frequency is adjusted simultaneously so that the difference between the synthesized power spectrum and actual LPC residual spectrum is minimized. The harmonic spectral envelope for a voiced segment is then vector quantized.

5.5.3.4 Vector quantization of harmonic magnitudes

In order to vector quantize a spectral envelope composed of a variable number of harmonics, the harmonic spectral vector is first converted to a fixed-dimension vector. Band-limited interpolation by a polyphase filter bank is used for the dimensional conversion. A fixed-dimension spectral vector $\mathbf{x}$ is then quantized with weighted distortion measure D ;

$$D = \|\mathbf{WH}(\mathbf{x} - \mathbf{g}(s_0 + s_1))\|^2$$

where s_0 is the output of the first shape codebook, s_1 is the output of the second shape codebook, and $\mathbf{g}$ is the output of the gain codebook. The diagonal components of the matrices $\mathbf{H}$ and $\mathbf{W}$ are the magnitudes of the frequency response of the LPC synthesis filter and the perceptual weighting filter, respectively. In order to reduce the memory requirements and search complexity while maintaining a high performance, two-stage vector quantization scheme is employed for the spectral shape together with a scalar quantizer for the gain under 2,0 kbit/s coding. For the 4,0 kbit/s mode, the quantized harmonic magnitude vector with fixed dimension is first converted to the dimension of original harmonics by the same band-limited interpolation mentioned above. The difference between the original harmonics and de-quantized and dimension converted harmonics are then quantized with additional vector quantizers. This multi-stage structure allows generation of scalable bit-streams.

5.5.3.5 Voiced/Unvoiced decision

The Voiced/Unvoiced decision is made based on the maximum autocorrelation of the LPC residual signals, the number of zero crossing and the harmonic structure of the power spectrum of the LPC residual signals.

5.5.3.6 VXC coding of unvoiced signals

For unvoiced segments, regular VXC coding is carried out, where only stochastic codebooks are used. A 6 bits shape codebook of dimension 80 and 4 bits gain codebook are used for the 2,0 kbit/s mode. For the 4,0 kbit/s mode, the quantization error of the 2,0 kbit/s mode is quantized using a 5 bits shape codebook of dimension 40 and a 3 bits gain codebook at the additional stage.

5.5.4 HVXC channel coding

5.5.4.1 Protected Bit Selection

According to the sensitivity of bits, encoded bits are classified to several Error Sensitivity Categories (ESC). The number of bits for each ESC is shown in table 14 (2,0 kbit/s, voiced sound), table 15 (2,0 kbit/s, unvoiced sound), table 16 (4,0 kbit/s, voiced sound) and table 17 (4,0 kbit/s, unvoiced sound). ESC0 is the group of the most error sensitive bits and ESC4 is the group of the least sensitive bits. Bit rate setting of total of 2,4 kbit/s using 2,0 kbit/s source coder rate, and total of 4,66 kbit/s using 4,0 kbit/s source coder rate are shown.

NOTE: The overall bit rate due to the usage of a CRC is 4,65 kbit/s, but additionally 4 padding bits have to be inserted for each audio super frame, resulting in an overall bit rate of 4,66 kbit/s, see clause 5.5.

Table 14: Number of ESC bits at 2,0 kbit/s fixed rate mode (voiced sound)

Parameters	Voiced frame				
	ESC0 (bits)	ESC1 (bits)	ESC2 (bits)	ESC3 (bits)	total (bits)
LSP1	5	-	-	-	5
LSP2	2	-	-	5	7
LSP3	1	-	-	4	5
LSP4	1	-	-	-	1
VUV	2	-	-	-	2
Pitch	6	-	-	1	7
SE_gain	5	-	-	-	5
SE_shape1	-	4	-	-	4
SE_shape2	-	-	4	-	4
total	22	4	4	10	40
CRC	6	1	1	-	8
total + CRC	28	5	5	10	48

Table 15: Number of ESC bits at 2,0 kbit/s fixed rate mode (unvoiced sound)

Parameters	Unvoiced frame				
	ESC0 (bits)	ESC1 (bits)	ESC2 (bits)	ESC3 (bits)	total (bits)
LSP1	5	-	-	-	5
LSP2	4	3	-	-	7
LSP3	2	1	2	-	5
LSP4	1	-	-	-	1
VUV	2	-	-	-	2
VX_gain1 [0]	4	-	-	-	4
VX_gain1 [1]	4	-	-	-	4
VX_shape1 [0]	-	-	2	4	6
VX_shape1 [1]	-	-	-	6	6
total	22	4	4	10	40
CRC	6	1	1	-	8
total + CRC	28	5	5	10	48

Table 16: Number of ESC bits at 4,0 kbit/s fixed rate mode (voiced sound)

Parameters	Voiced frame					
	ESC0 (bits)	ESC1 (bits)	ESC2 (bits)	ESC3 (bits)	ESC4 (bits)	total (bits)
LSP1	5	-	-	-	-	5
LSP2	4	-	-	-	3	7
LSP3	1	-	-	-	4	5
LSP4	1	-	-	-	-	1
LSP5	1	-	-	-	7	8
VUV	2	-	-	-	-	2
Pitch	6	-	-	-	1	7
SE_gain	5	-	-	-	-	5
SE_shape1	-	-	4	-	-	4
SE_shape2	-	-	-	4	-	4
SE_shape3	5	-	-	-	2	7
SE_shape4	1	9	-	-	-	10
SE_shape5	1	8	-	-	-	9
SE_shape6	1	5	-	-	-	6
Total	33	22	4	4	17	80
CRC	6	5	1	1	-	13
Total + CRC	39	27	5	5	17	93

Table 17: Number of ESC bits at 4,0 kbit/s fixed rate mode (unvoiced sound)

Parameters	Unvoiced frame					
	ESC0 (bits)	ESC1 (bits)	ESC2 (bits)	ESC3 (bits)	ESC4 (bits)	total (bits)
LSP1	5	-	-	-	-	5
LSP2	4	3	-	-	-	7
LSP3	1	4	-	-	-	5
LSP4	1	-	-	-	-	1
LSP5	1	7	-	-	-	8
VUV	2	-	-	-	-	2
VX_gain1 [0]	4	-	-	-	-	4
VX_gain1 [1]	4	-	-	-	-	4
VX_shape1 [0]	-	6	-	-	-	6
VX_shape1 [1]	-	1	4	1	-	6
VX_gain2 [0]	3	-	-	-	-	3
VX_gain2 [1]	3	-	-	-	-	3
VX_gain2 [2]	3	-	-	-	-	3
VX_gain2 [3]	2	1	-	-	-	3
VX_shape2 [0]	-	-	-	3	2	5
VX_shape2 [1]	-	-	-	-	5	5
VX_shape2 [2]	-	-	-	-	5	5
VX_shape2 [3]	-	-	-	-	5	5
total	33	22	4	4	17	80
CRC	6	5	1	1	-	13
total + CRC	39	27	5	5	17	93

5.5.4.2 Syntax of DRM HVXC error resilience (ErHVXCfixframe_CRC)

The bitstream syntax consists of several Error Sensitivity Categories (ESC). Some ESCs include source bits and CRC bits, where CRC bits are computed from source bits within the same ESC.

The HVXC_CRC field in the SDC is used to indicate whether the stream includes CRC parity bits or not (see clause 6.4.3.10). The bitstream syntax of the input of the CRC checker is given in figure 8.

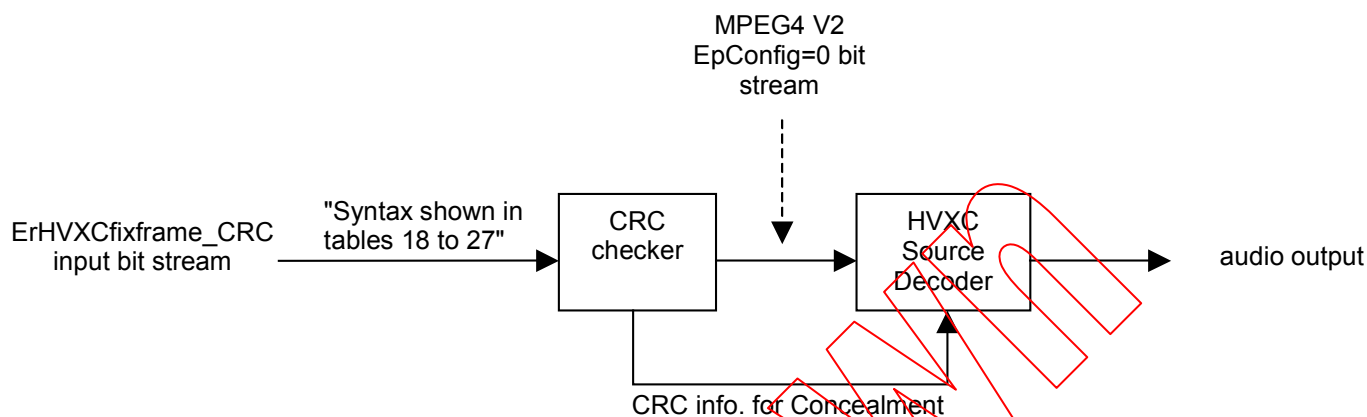


Figure 8: Block Diagram of HVXC Syntax

Table 18: Syntax of ErHVXCfixframe_CRC()

Syntax	No. of bits	Mnemonic
<pre> ErHVXCfixframe_CRC(rate) { if (rate == 2000) { 2k_ESC0_CRC0(); 2k_ESC1_CRC1(); 2k_ESC2_CRC2(); 2k_ESC3_NoCRC(); } else { 4k_ESC0_CRC0(); 4k_ESC1_CRC1(); 4k_ESC2_CRC2(); 4k_ESC3_CRC3(); 4k_ESC4_NoCRC(); } } </pre>		

Table 19: Syntax of 2k_ESC0_CRC0()

Syntax	No. of bits	Mnemonic
2k_ESC0_CRC0() { VUV, 1-0; LSP4, 0; if (VUV!=0) { SE_gain, 4-0; LSP1, 4-0; Pitch, 6-1; LSP2, 6; LSP3, 4; LSP2, 5; } else { VX_gain1[0], 3-0; VX_gain1[1], 3-0; LSP1, 4-0; LSP2, 6-3; LSP3, 4-3; } if (HVXC_CRC==1) { CRC0_2k, 5-0; } }	 2 1 5 5 6 1 1 1 4 4 5 4 2 6	 uimbsf uimbsf uimbsf uimbsf uimbsf uimbsf uimbsf uimbsf uimbsf uimbsf uimbsf uimbsf uimbsf uimbsf

Table 20: Syntax of 2k_ESC1_CRC1()

Syntax	No. of bits	Mnemonic
2k_ESC1_CRC1() { if (VUV!=0) { SE_shape1, 3-0; } else { LSP2, 2-0; LSP3, 2; } if (HVXC_CRC==1) { CRC1_2k, 0; } }	 1	 uimbsf uimbsf uimbsf uimbsf

Table 21: Syntax of 2k_ESC2_CRC2()

Syntax	No. of bits	Mnemonic
2k_ESC2_CRC2() { if (VUV!=0) { SE_shape2, 3-0; } else { LSP3, 1-0; VX_shape1[0], 5-4; } if (HVXC_CRC==1) { CRC2_2k, 0; } }	 1	 uimbsf uimbsf uimbsf uimbsf

Table 22: Syntax of 2k_ESC3_NoCRC()

Syntax	No. of bits	Mnemonic
2k_ESC3_NoCRC() { if (VUV!=0) { LSP2, 4-0; LSP3, 3-0; Pitch, 0; } else { VX_shape1[0], 3-0; VX_shape1[1], 5-0; } }	5 4 1 4 6	uimbsf uimbsf uimbsf uimbsf uimbsf

Table 23: Syntax of 4k_ESC0_CRC0()

Syntax	No. of bits	Mnemonic
4k_ESC0_CRC0() { VUV, 1-0; LSP4, 0; if (VUV!=0) { SE_gain, 4-0; LSP1, 4-0; Pitch, 6-1; LSP2, 6-3; SE_shape3, 6-2; LSP3, 4; LSP5, 7; SE_shape4, 9; SE_shape5, 8; SE_shape6, 5; } else { VX_gain1[0], 3-0; VX_gain1[1], 3-0; LSP1, 4-0; LSP2, 6-3; LSP3, 4; LSP5, 7; VX_gain2[0], 2-0; VX_gain2[1], 2-0; VX_gain2[2], 2-0; VX_gain2[3], 2-1; } if (HVXC_CRC==1) { CRC0_4k, 5-0; } }	2 1 5 5 6 4 5 1 1 1 1 1 4 4 5 4 1 1 3 3 3 3 2 6	uimbsf uimbsf uimbsf

Table 24: Syntax of 4k_ESC1_CRC1()

Syntax	No. of bits	Mnemonic
4k_ESC1_CRC1() { if (VUV!=0) { SE_shape4, 8-0; SE_shape5, 7-0; SE_shape6, 4-0; } else { VX_gain2[3], 0; LSP2, 2-0; LSP3, 3-0; LSP5, 6-0; VX_shape1[0], 5-0; VX_shape1[1], 5; } if (HVXC_CRC==1) { CRC1_4k, 4-0; } }	9 8 5 1 3 4 7 6 1	uimsbf uimsbf uimsbf uimsbf uimsbf uimsbf uimsbf uimsbf uimsbf

Table 25: Syntax of 4k_ESC2_CRC2()

Syntax	No. of bits	Mnemonic
4k_ESC2_CRC2() { if (VUV!=0) { SE_shape1, 3-0; } else { VX_shape1[1], 4-1; } if (HVXC_CRC==1) { CRC2_4k, 0; } }	4 4 1	uimsbf uimsbf uimsbf

Table 26: Syntax of 4k_ESC3_CRC3()

Syntax	No. of bits	Mnemonic
4k_ESC3_CRC3() { if (VUV!=0) { SE_shape2, 3-0; } else { VX_shape1[1], 0; VX_shape2[0], 4-2; } if (HVXC_CRC==1) { CRC3_4k, 0; } }	4 1 3 1	uimsbf uimsbf uimsbf uimsbf

Table 27: Syntax of 4k_ESC4_NoCRC()

Syntax	No. of bits	Mnemonic
4k_ESC4_NoCRC()		
{		
if (VUV!=0) {		
LSP2, 2-0;	3	uimsbf
LSP3, 3-0;	4	uimsbf
LSP5, 6-0;	7	uimsbf
Pitch, 0;	1	uimsbf
SE_shape3, 1-0;	2	uimsbf
} else {		
VX_shape2[0], 1-0;	2	uimsbf
VX_shape2[1], 4-0;	5	uimsbf
VX_shape2[2], 4-0;	5	uimsbf
VX_shape2[3], 4-0;	5	uimsbf
}		
}		

5.5.5 Category Interleaving

In order to improve the robustness of the bitstream to channel errors, bit interleaving is carried out where data bits are reordered inside one frame. Bit interleaving is carried out after the bitstream is ordered and the CRC bits are added as shown in tables 18 to 27. The interleaving is done in two steps. First, the HVXC bitstream is divided into two sequences as:

- $x[0..N_1-1]$: sequence consists of bits of ESC0 where N_1 is the number of bits included in ESC0.
- $y[0..N_2-1]$: sequence consists of bits other than ESC0.

The sequence $y[]$ is composed of ESC1, ESC2 ..., in this order. N_1 is the number of bits of ESC0, and N_2 is the number of bits other than ESC0. The total number of bits is $N = N_1 + N_2$. r is defined as:

~~$$r = \frac{N}{N_1}$$~~

Next, the sequences $x[]$ and $y[]$ are interleaved into one sequence $z[0..N-1]$. Every single bit from $x[]$ is inserted at the beginning of every $\lfloor r \rfloor - 1$ or $\lfloor r \rfloor$ bits from $y[]$, where $\lfloor r \rfloor$ is the largest integer less than or equal to r . The flowchart is shown below in figure 9.

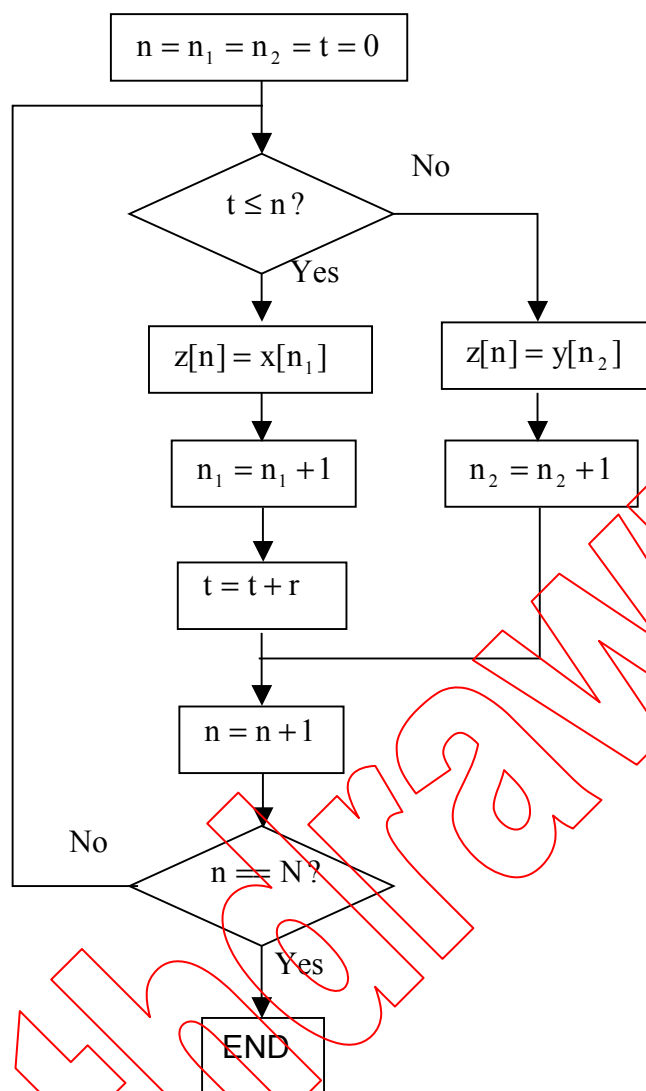


Figure 9: Flowchart of class interleaving method

5.5.6 HVXC error detection and concealment

5.5.6.1 Cyclic Redundancy Code (CRC)

The CRC parity bits are computed from the source bits in the same ESC. A schematic diagram of the CRC checker, the polynomials and initialization procedure is given in annex D.

5.5.6.2 Error concealment

In case a CRC error is detected, error concealment processing (bad frame masking) is carried out in the HVXC decoder. The state transition diagram of the concealment processing is shown in figure 10. A frame masking state of the current frame is updated based on the decoded CRC result of ESC0. If a CRC error is detected in ESC0, the frame is declared to be a "bad" frame. The initial state of the state transition diagram is state = 0. The arrow with a letter "1" denotes the transition for a bad frame, and that with a letter "0" the one for a good frame. At the 2,0 kbit/s rate ESC1 and ESC2 are protected by CRC bits; at the 4,0 kbit/s ESC1 to ESC3 are protected by CRC bits. The results of the CRC checks against these ESCs are used to maintain toll quality as described in detail below.

5.5.6.2.1 Parameter replacement

According to the state value, the following parameter replacement is done. In error free condition, the state value becomes 0, and received source coder bits are used without any concealment processing.

LSP parameters

At state = 1...6, LSP parameters are replaced with those of previous states. At state = 7, if LSP4 = 0 (LSP quantization mode without inter-frame prediction), then LSP parameters are calculated from all LSP indices received in the current frame. If LSP4 = 1 (LSP quantization mode with inter-frame coding), then LSP parameters are calculated with the following method, where LSP parameters belonging to the LSP1 index are interpolated with the previous LSPs.

$$LSP_{base}(n) = p \times LSP_{prev}(n) + (1 - p) \times LSP_{1st}(n) \text{ for } n = 1 \dots 10 \quad (1)$$

$LSP_{base}(n)$ is LSP parameters of the base layer, $LSP_{prev}(n)$ is the decoded LSPs of the previous frame, $LSP_{1st}(n)$ is the decoded LSPs from the current LSP1 index, and p is the interpolation factor. p is changed according to the number of previous bad frames as shown in table 28. LSP indices LSP2, LSP3 and LSP5 are not used, and $LSP_{base}(n)$ computed according to the Equation (1) is used as current LSP parameters.

Table 28: p factor

frame	p
0	0,7
1	0,6
2	0,5
3	0,4
4	0,3
5	0,2
6	0,1
7	0,0

Mute variable

According to the "state" value, a variable "mute" is set to control the output level of the reproduced speech. The "mute" values in table 29 are used. At state = 7, the average of 1,0 and "mute" value of the previous frame (= 0,5 (1,0 + previous "mute value")) is used. However, when this value is more than 0,8, "mute" value is replaced with 0,8.

Table 29: Mute value

state	mute value
0	1,0
1	0,8
2	0,7
3	0,5
4	0,25
5	0,125
6	0,0
7	average/0,8

Replacement and gain control of "voiced" parameters

At state = 1...6, spectrum parameter SE_shape1, SE_shape2, spectrum gain parameter SE_gain, spectrum parameter for 4,0 kbit/s mode SE_shape3 ... SE_shape6 are replaced with the corresponding parameters of the previous frame. Also, to control the volume of the output speech and the harmonic magnitude parameters of LPC residual signal, " $Am[0 \dots 127]$ " is gain controlled as shown in equation (2). In the equation, $Am_{(org)}[i]$ is computed from the received spectrum parameters of the latest error free frame.

$$Am[i] = mute \times Am_{(org)}[i] \text{ for } i = 0 \dots 127 \quad (2)$$

If the previous frame is unvoiced and the current state is state = 7, equation (2) is replaced with equation (3);

$$Am[i] = 0,6 \times mute \times Am_{(org)}[i] \text{ for } i = 0 \dots 127 \quad (3)$$

As described before, SE_shape1 and SE_shape2 are individually protected by 1 bit CRCs. ESC1 applies to SE_shape1 and ESC2 applies to SE_shape2 at 2,0 kbit/s. In the same way, ESC2 applies to SE_shape1 and ESC3 applies to SE_shape2 at 4,0 kbit/s. At state = 0 or 7, if both of the CRCs of SE_shape1 and SE_shape2 are in error at the same time, the quantized harmonic magnitudes with fixed dimension $Am_{qnt}[1..44]$ are gain suppressed as:

$$Am_{qnt}[i] = s[i] \times Am_{qnt(org)}[i] \text{ for } i = 1 \dots 44 \quad (4)$$

$s[i]$ is the factor for the gain suppression. $Am_{qnt(org)}[i]$ is the fixed dimension harmonic magnitudes generated using the SE_shape1 and SE_shape2 which contains bit errors. $Am_{qnt}[i]$ is then dimension converted to obtain $Am[i]$.

Table 30: Factor for gain suppression, $s[0..44]$

i	1	2	3	4	5	6	7...44
$s[i]$	0,10	0,25	0,40	0,55	0,70	0,85	1,00

At 4,0 kbit/s, SE_shape4, SE_shape5, and SE_shape6 are subject to CRC checking as ESC1. When a CRC error is detected, the spectrum parameter of the enhancement layer is not used.

Replacement and gain control of "unvoiced" parameters

At state = 1...6, the stochastic codebook gain parameters $VX_gain1[0]$ and $VX_gain1[1]$ are replaced with the previous frame's $VX_gain1[1]$. Also the stochastic codebook gain parameters for the 4,0 kbit/s mode $VX_gain2[0] \dots VX_gain2[3]$ are replaced with the previous frame's $VX_gain2[3]$.

The stochastic codebook shape parameters $VX_shape1[0]$, $VX_shape1[1]$ and the stochastic codebook shape parameters for the 4,0 kbit/s mode $VX_shape2[0]$, $VX_shape2[1]$, $VX_shape2[2]$ and $VX_shape2[3]$ are generated from randomly generated index values.

Finally, in order to control the volume of the output speech, the LPC residual signal $res[0..159]$ is gain controlled as shown in equation (5). $res_{(org)}[i]$ is the excitation signal obtained by using the stochastic codebook gain and the shape parameters generated as explained above.

$$res[i] = mute \times res_{(org)}[i] \quad (0 \leq i \leq 159) \quad (5)$$

Frame Masking State Transitions

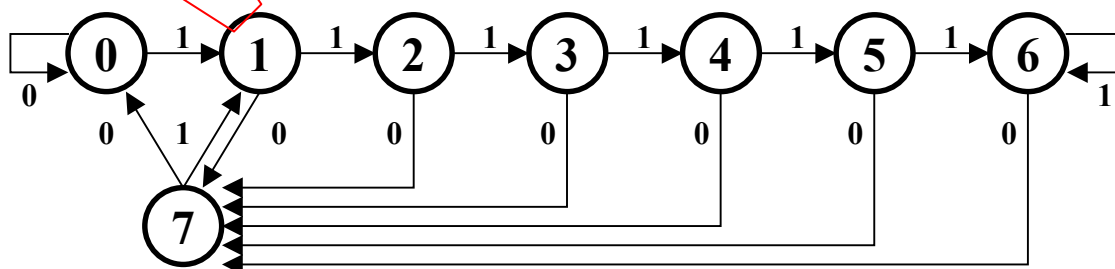


Figure 10: Frame Masking State Transitions

5.6 SBR

5.6.1 Conceptual overview

Spectral Band Replication (SBR) is an audio coding enhancement tool. It improves the performance of low bit rate audio and speech codecs. SBR can increase the bandwidth of a conventional low bit rate codec to equal or exceed the analogue FM audio bandwidth (15 kHz), see figure 11. This can be achieved within the low bit rates available from DRM. SBR can also improve the performance of narrow-band speech codecs, offering the broadcaster 12 kHz commentary audio bandwidth, used for example for multilingual broadcasting. As most speech codecs are narrowband, SBR is important not only for improving speech quality, but also for improving speech intelligibility and speech comprehension. SBR is mainly a post-process, although some pre-processing is performed in the encoder in order to guide the decoding process.

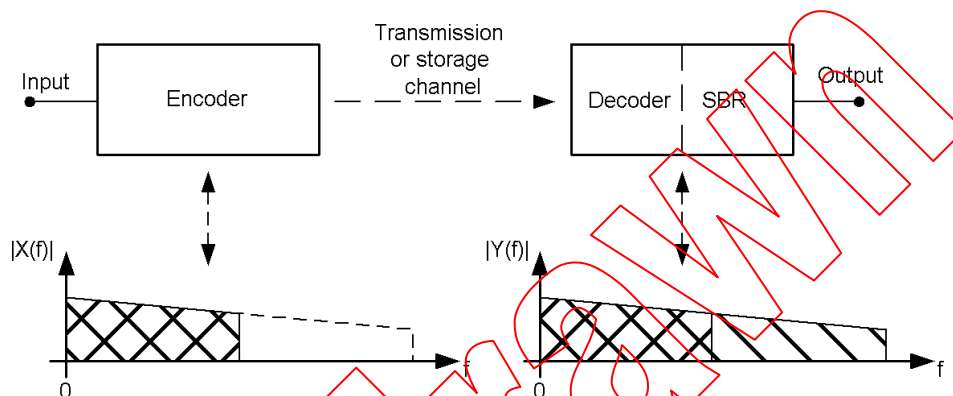


Figure 11: The SBR source coding system

The human voice and most musical instruments generate quasi-stationary excitation signals that emerge from oscillating systems. The wide-band excitation spectrum is created by for example vocal cords, strings or reeds etc and its frequency components form a harmonic series. The harmonic series is filtered by resonators such as the vocal tract, violin body etc., giving the voice or musical instrument its characteristic tone colour or timbre. A bandwidth limitation of such a signal is equivalent to a truncation of the harmonic series, see figure 12. Such a truncation alters the perceived timbre and the audio signal sounds "muffled" or "dull", and the intelligibility may be reduced.

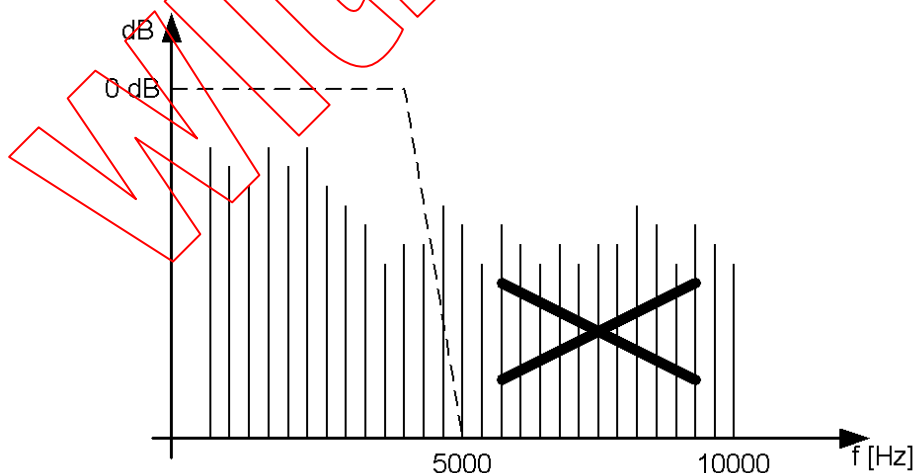


Figure 12: Example of band-limiting of a typical signal

The SBR concept postulates that a truncated harmonic series can be extended based on the relation between lowband and highband spectral components. The spectral envelope, i.e. the coarse spectral distribution, of the replicated highband, must resemble that of the original signal. This is assured by transmitting spectral envelope guidance information from the encoder to decoder at a very low bit rate (approximately 2 kbit/s/channel).

It is important to maintain the ratio between harmonic and noise-like components in the replicated highband. This makes it necessary to selectively add noise components to the SBR replicated high-band.

Two different bit stream protocols are used, one for use with AAC, and one for use with CELP. They are described in the clauses below.

5.6.2 AAC + SBR Decoding Process

The input to the SBR decoder is the SBR bitstream data and the decoded time domain signal from the AAC decoder, which is sampled at half the rate of the output signal. The decoded low-band time domain signal is fed to an analysis QMF bank, as shown in figure 13. The subband signals obtained from the filter bank are re-routed by an LC-repating algorithm to reconstruct the high-band. Using the envelope information derived from the transmitted SBR bitstream a gain adjustment within the filter bank scalefactor bands is performed on the high-band subband signals. The adjusted high-band and the delay compensated low-band are then fed to the common synthesis QMF bank which generates the wide-band time domain output signal.

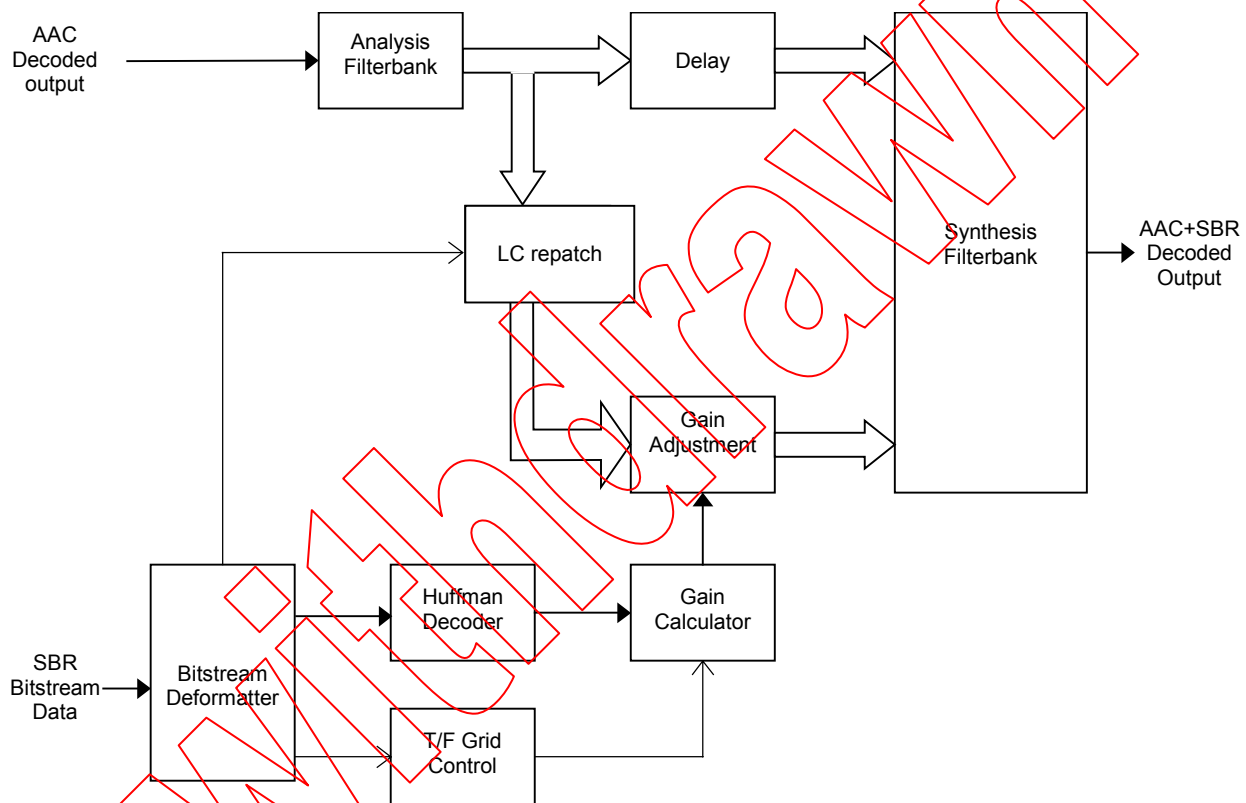


Figure 13: Flow Diagram of the AAC-SBR Decoding Process

5.6.2.1 Analysis Filterbank

Subband filtering of the AAC decoded time domain signal is achieved using a 32-channel QMF bank. The flowchart of figure 14(a) shows the operation. The window coefficients are given in annex I, table I.1. For every loop in the flowchart, the output is 32 subband samples, each representing the output from one filterbank channel. For every frame the filterbank will produce 30 subband samples for every channel, corresponding to a time domain signal of length 960 samples.

5.6.2.2 Synthesis Filterbank

Synthesis filtering of the delayed subband samples and the gain-adjusted subband samples is accomplished using a 64-channel complex QMF bank. The operation is shown in the flowchart of figure 14(b), and the window coefficients are the same as for the analysis bank and tabulated in annex I, table I.1. The delayed subband samples are fed to the lowest 32 channels. The real-valued gain-adjusted subband samples are fed to channels corresponding to higher frequencies. The output for every frame is 1 920 time domain samples.

5.6.2.3 Gain Control

The steps involved for one cycle of the gain control unit, corresponding to one frame, is shown in the flowchart of figure 15. First, an energy calculation of the subband samples is performed, i.e. the squared absolute values of the complex-valued subband samples is computed. The energy samples are subsequently combined within segments being the same as the segments used in the encoder for the current part of the signal. This combination produces the envelope samples $\text{nrg_est}[i]$, where $i = 0, 1, \dots, \text{nNrg}-1$. The envelope data transmitted from the encoder is decoded and dequantized, thus restoring $\text{nrg_ref}[i]$, where $i = 0, 1, \dots, \text{nNrg}-1$. The calculation of gainfactors is then performed by taking the square-root of the ratios of the energy values $\text{nrg_ref}[i]$ and $\text{nrg_est}[i]$ with compensation for the noise floor level; $\text{gain}[i] = \sqrt{\text{nrg_ref}[i]/(\text{nrg_est}[i] \times (1+\text{ns_level}[i]))}$, for $i = 0, 1, \dots, \text{nNrg}-1$. The real part of the subband samples in the segments is multiplied with the corresponding gainfactor, thus adjusting several frequency- and time-adjacent subband samples with one gainfactor. The noise floor level data transmitted from the encoder is decoded and dequantized, which restores and maps the nNoiseNrg values to nNrg noise floor level values, $\text{ns_level}[i]$, where $i = 0, 1, \dots, \text{nNrg}-1$. Finally, random white noise is added to the subband samples using the noise level data, i.e. the noise energy will be $\text{nrg_ref}[i] \times \text{ns_level}[i]/(1+\text{ns_level}[i])$.

5.6.2.4 T/F Grid Control

As stated above, the envelope estimates are obtained by adaptive grouping of subband-samples of a fixed-size filterbank into time- and frequency-regions, for each of which one envelope sample is calculated. The description of those frequency bands and time segments is referred to as a T/F-grid. Clearly, in order to perform the above described gain adjustment, this grid must be transmitted to the decoder. The frequency bands are defined by table-pairs that are stored in both the encoder and the decoder. Such a pair corresponds to high and low frequency resolution respectively, in the sense that more or fewer frequency bands are used. The frequency resolution is selectable on a per-segment basis, and signalled correspondingly to the decoder. The time segment boundaries, referred to as borders, are selected according to the input signal properties, whereby longer segments are used during quasi-stationary passages, and smaller steps in time are used in the vicinity of transients. Again, those borders must be signalled to the decoder. Four frame classes are defined, and the bitstream signals are tailored towards the specific needs of each class. Class "FixFix" corresponds to conventional constant length granules. Class "FixVar" has a movable stop boundary, which allows the frame length to vary. Class "VarFix" has a variable start boundary, whereas the stop border is fixed. The last class, "VarVar", has variable boundaries at both ends. The class is signalled, and, except for class FixFix, followed by the position of the variable boundary(/-ies). Thus, those signals define the frame start and stop boundaries. The borders within a frame are signalled by means of relative distances from the variable boundaries. The noise samples are defined and signalled in a similar manner. The task of the T/F-Grid Control module is to reassemble the grid from the above described bitstream elements.

5.6.2.5 Huffman Decoder

The envelope and noise samples are quantized and delta coded, either in the frequency or time direction, and the direction chosen is signalled to the decoder. The delta values are Huffman coded, and the codewords are transmitted to the decoder. Obviously, the task of the Huffman Decoder module is to reconvert the codewords into the quantized envelope and noise samples generated at the encoder.

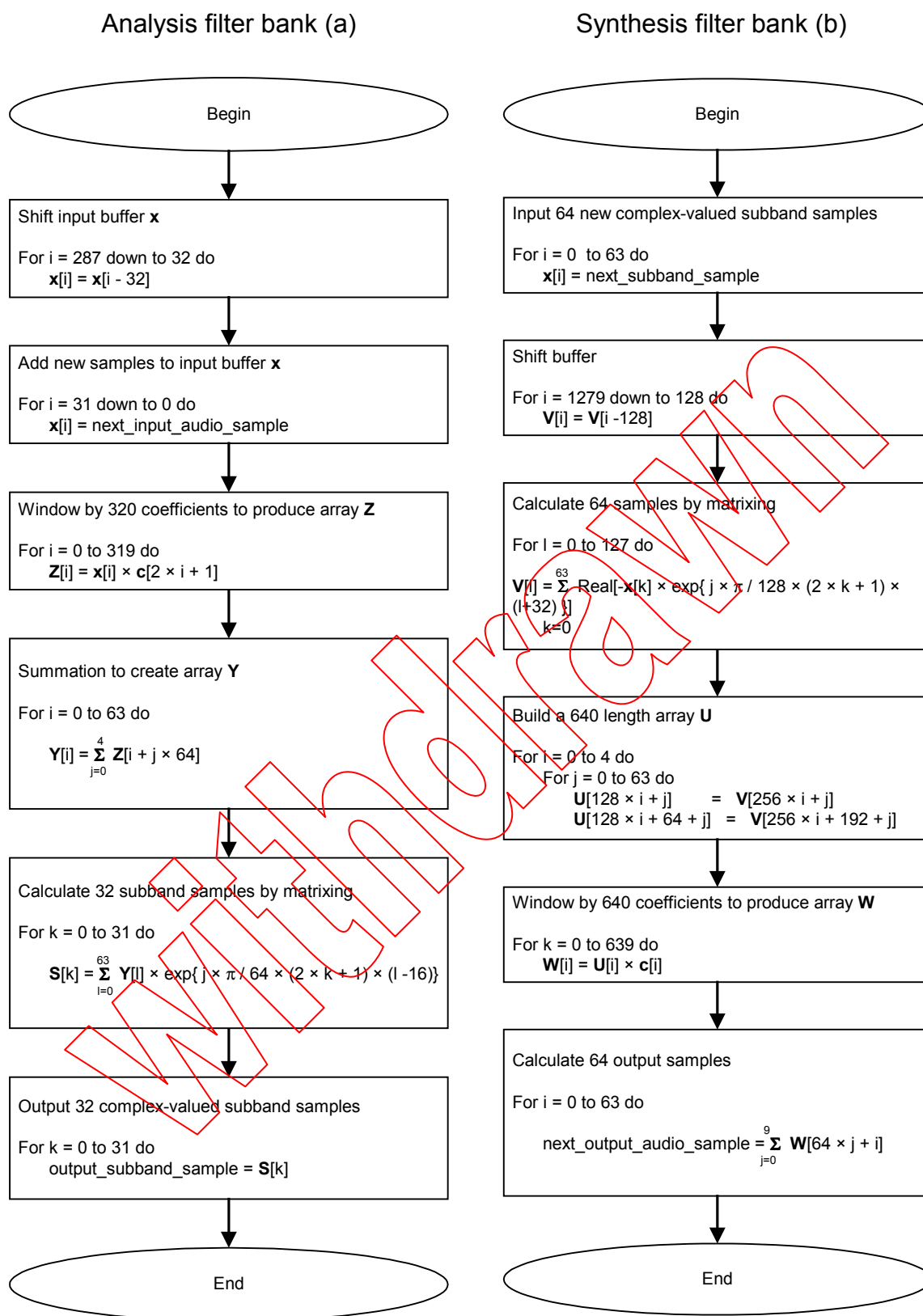


Figure 14: Flow Chart of the Analysis (a) and Synthesis (b) filter bank

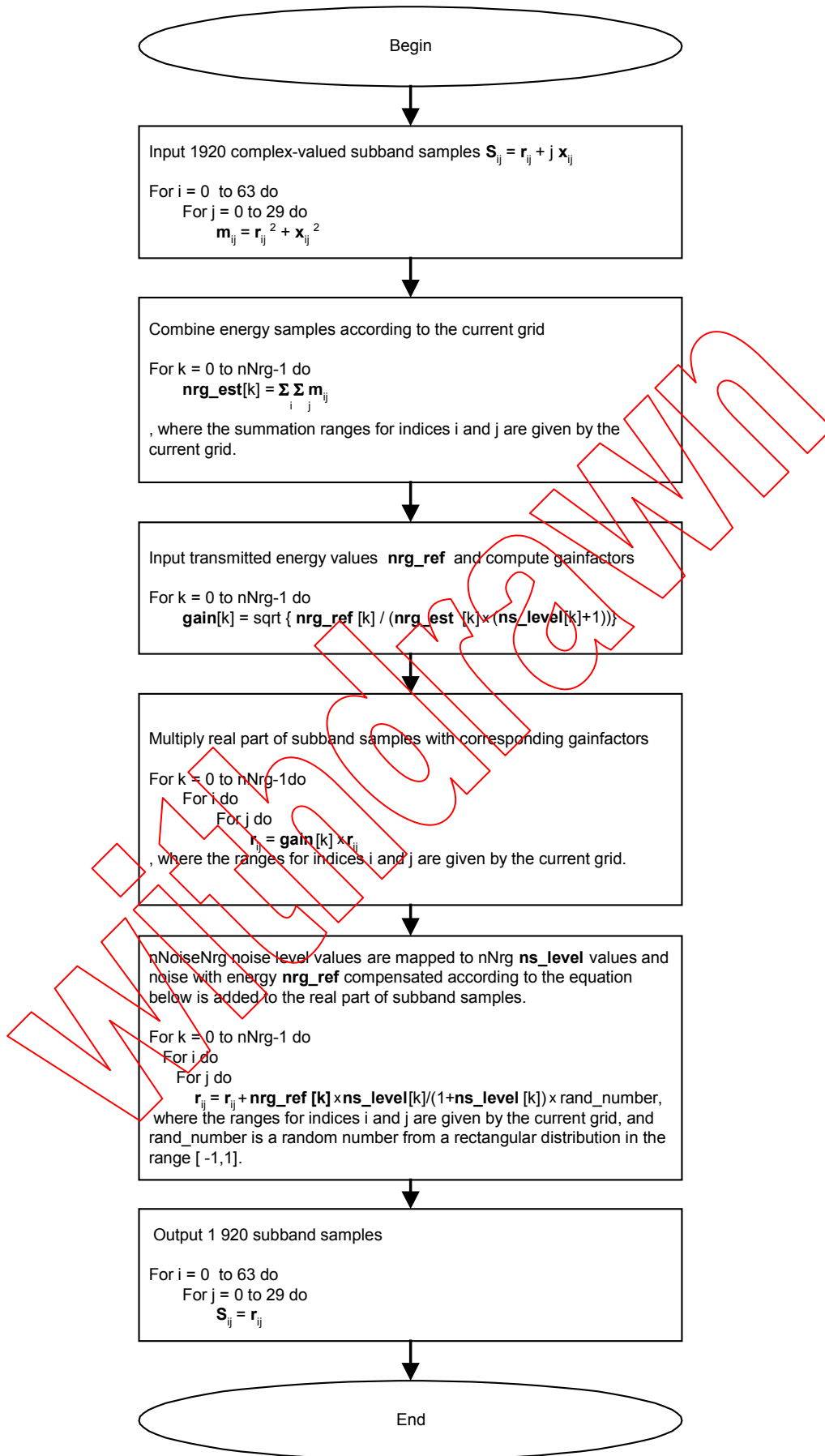


Figure 15: Flow Chart of the Gain Control Module

5.6.3 AAC + SBR Protocol

The SBR bitstream is described by means of pseudo code, using a syntax similar to the one used in the MPEG standard documents. The location of the SBR bitstream within the AAC audio super frames was described in clause 5.3.2. This clause only describes how to extract the elements from the bitstream. The application thereof was briefly described in clause 4.

5.6.3.1 AAC + SBR syntax

Table 31: Syntax of sbr_aac_frame()

Syntax	No. of bits	Note
sbr_aac_frame(audio_mode) { sbr_crc_bits if (header_flag) sbr_header (audio_mode) sbr_data (audio_mode, data extra) // data extra is located in sbr_header }	 8 1 	 see annex D

Table 32: Syntax of sbr_header()

Syntax	No. of bits	Note
sbr_header (audio_mode) { protocol_version amp_res start_frequency stop_frequency xover_band reserved data_extra // enable optional data part header_extra_1 // enable optional header part 1 header_extra_2 // enable optional header part 2 if (audio_mode == LC_STEREO) lc_stereo_mode // optional parts if (header_extra_1) { freq_scale alter_scale noise_bands } if (header_extra_2) { limiter_bands limiter_gains interpol_freq smoothing_mode reserved } }	 2 1 4 4 3 2 1 1 1 2 2 1 2 2 2 1 1 1	