ETSI TS 103 466 V1.2.1 (2019-09)

TECHNICAL SPECIFICATION

Digital Audio Broadcasting (DAB);
DAB audio coding (MPEG Layer II)

Reference
RTS/JTC-DAB-96

Keywords
audio, broadcasting, coding, DAB, digital

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables 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 (https://ipr.etsi.org/).

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.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

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

NOTE 1: 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

The Eureka Project 147 was established in 1987, with funding from the European Commission, to develop a system for the broadcasting of audio and data to fixed, portable or mobile receivers. Their work resulted in the publication of European Standard, ETSI EN 300 401 [1], for DAB (see note 2) which now has worldwide acceptance.

NOTE 2: DAB is a registered trademark owned by one of the Eureka Project 147 partners.

The DAB family of standards is supported by WorldDAB, an organization with members drawn from broadcasting organizations and telecommunication providers together with companies from the professional and consumer electronics industry.

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

1 Scope

The present document defines the method to code and transmit audio services using the MPEG Layer II audio coder for Digital Audio Broadcasting (DAB) (ETSI EN 300 401 [1]) and details the necessary mandatory requirements for decoders. The permitted audio modes and the data protection and encapsulation are detailed. This audio coding scheme permits the full use of the PAD channel for carrying dynamic labels and user applications.

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at https://docbox.etsi.org/Reference/.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are necessary for the application of the present document.

[1] ETSI EN 300 401 (V2.1.1): "Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers".

[2] ISO/IEC 11172-3 (1993): "Information technology - Coding of moving pictures and associated audio for digital storage media at up to 1,5 Mbit/s - Part 3: Audio".

[3] IEC 60958 (all parts): "Digital audio interface".

[4] ISO/IEC 13818-3: "Information technology - Generic coding of moving pictures and associated audio information - Part 3: Audio".

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

Not applicable.

3 Definition of terms, abbreviations, mathematical symbols and convention

3.1 Terms

For the purposes of the present document, the terms given in ETSI EN 300 401 [1] and the following apply:

alias component: mirrored signal component resulting from sub-Nyquist sampling

audio bit stream: sequence of consecutive audio frames

audio frame: frame of a duration of 24 ms (at 48 kHz sampling frequency) or of 48 ms (at 24 kHz sampling frequency) which contains a Layer II encoded audio signal ISO/IEC 11172-3 [2], ISO/IEC 13818-3 [4], corresponding to 1 152 consecutive audio samples

NOTE: It is the smallest part of the audio bit stream which is decodable on its own.

audio mode: audio coding system provides single channel, stereo and joint stereo audio modes

NOTE: In each mode, the complete audio signal is encoded as one audio bit stream.

bark: unit of the critical band

NOTE: The Bark scale is a non-linear mapping of the frequency scale over the entire audio frequency range.

bit allocation: time-varying assignment of bits to samples in different sub-bands according to a psychoacoustic model

bound: lowest sub-band in which Intensity stereo coding is used, in the case of joint stereo mode

Common Interleaved Frame (CIF): serial digital output from the main service multiplexer which is contained in the Main Service Channel part of the transmission frame

NOTE: It is common to all transmission modes and contains 55 296 bits (i.e. 864 CUs).

convolutional coding: coding procedure which generates redundancy in the transmitted data stream in order to provide ruggedness against transmission distortions

critical band: psychoacoustic measure in the frequency domain which corresponds to the frequency selectivity of the human ear

DAB audio frame: Same as audio frame, but includes all specific DAB audio-related information.

dual channel mode: audio mode, in which two audio channels with independent programme contents are encoded within one audio bit stream

NOTE: This audio mode is not used in DAB.

Equal Error Protection (EEP): error protection procedure which ensures a constant protection of the bit stream

Extended Programme Associated Data (X-PAD): extended part of the PAD carried towards the end of the DAB audio frame, immediately before the Scale Factor Cyclic Redundancy Check (CRC)

NOTE: Its length is variable.

Fixed Programme Associated Data (F-PAD): fixed part of the PAD contained in the last two bytes of the DAB audio frame

intensity stereo coding: method of exploiting stereo irrelevance or redundancy in stereophonic audio programmes

NOTE: It is based on retaining only the energy envelope of the right and left channels at high frequencies. At low frequencies, the fine structure of the left and right channel of a stereophonic signal is retained.

joint stereo mode: audio mode, in which two channels forming a stereo pair (left and right) are encoded within one bit stream and for which stereophonic irrelevance or redundancy is exploited for further bit reduction

NOTE: The method used in the DAB system is Intensity stereo coding.

logical frame: data burst, contributing to the contents of a sub-channel, during a time interval of 24 ms

EXAMPLE: Data bursts at the output of an audio encoder, a Conditional Access scrambler and a convolutional encoder are referred to as logical frames. The number of bits contained in a specific logical frame depends on the stage in the encoding process and the bit rate associated with the sub-channel.

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

masking: property of the human auditory system by which an audio signal cannot be perceived in the presence of another audio signal

masking threshold: function of frequency and time, specifying the sound pressure level below which an audio signal cannot be perceived by the human auditory system

N: length of Fast Fourier Transform (FFT)

polyphase filter bank: set of equal-bandwidth filters with special phase relationship, allowing for efficient implementation of a filter bank

Programme Associated Data (PAD): information which is related to the audio data in terms of contents and synchronization

NOTE: The PAD field is located at the end of the DAB audio frame.

protection level: level specifying the degree of protection, provided by the convolutional coding, against transmission errors

protection profile: scheme of convolutional coding applied

psychoacoustic model: mathematical model of the masking behaviour of the human auditory system

Scale Factor (ScF): factor by which a set of values is scaled before quantization

NOTE: The numerical code for the Scale Factor is called the Scale Factor Index.

Scale Factor Select Information (ScFSI): 2-bit code which indicates for each sub-band how many Scale Factors are coded within the audio frame

service: user-selectable output which can be either a programme service or a data service

service component: part of a service which carries either audio (including PAD) or data

NOTE: The service components of a given service are linked together by the Multiplex Configuration Information. Each service component is carried either in a sub-channel or in the Fast Information Data Channel.

side information: information in the encoded audio bit stream which is necessary for controlling the audio decoder

NOTE: This information includes Bit Allocation, Scale Factor Select Information and Scale Factors.

single channel mode: audio mode, in which a monophonic audio programme is encoded within one bit stream

stereo mode: audio mode, in which two channels forming a stereo pair (left and right) are encoded within one bit stream

stuffing: one or more bits which may be inserted into the audio bit stream

NOTE: Stuffing bits are ignored by the audio decoding process. The purpose is to fill up a data field when required.

sub-band: subdivision of the audio frequency range

NOTE: In the audio coding system, 32 sub-bands of equal bandwidth are used.

sub-band samples: sub-band filter bank in the audio encoder creates a filtered and sub-sampled representation of the input audio signal

NOTE: The filtered samples are called sub-band samples. From 384 consecutive input audio samples, 12 consecutive sub-band samples are generated for each of the 32 sub-bands.

syncword: 12-bit code embedded in the MPEG Audio Layer II bit stream ISO/IEC 11172-3 [2], ISO/IEC 13818-3 [4] that identifies the beginning of an audio frame

Unequal Error Protection (UEP): error protection procedure which allows the bit error characteristics to be matched with the bit error sensitivity of the different parts of the bit stream

X-PAD data group: package of data used for one user application in the Extended Programme Associated Data (X-PAD)

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in ETSI EN 300 401 [1] and the following apply:

AES	Audio Engineering Society
ASCTy	Audio Service Component Type
CIF	Common Interleaved Frame
CRC	Cyclic Redundancy Check
DAB	Digital Audio Broadcasting
DRC	Dynamic Range Control
EBU	European Broadcasting Union
EEP	Equal Error Protection
FFT	Fast Fourier Transform
F-PAD	Fixed Programme Associated Data
ID	Identifier of audio coding algorithm
ISO	International Organization for Standardization
LSb	Least Significant bit
LSF	Lower Sampling Frequency
M/S	Music/Speech
MPEG	Moving Pictures Expert Group
MSb	Most Significant bit
MSB	Most Significant Byte
PAD	Programme Associated Data
PCM	Pulse Coded Modulation
ScF	Scale Factor
ScF-CRC	audio Scale Factor - Cyclic Redundancy Check (error check)
ScFSI	Scale Factor Select Information
SMR	Signal-to-Mask Ratio
SPL	Sound Pressure Level
UEP	Unequal Error Protection
X-PAD	eXtended Programme Associated Data

3.3 Mathematical symbols

3.3.1 Arithmetic operators

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

^	Power
/	Integer division with truncation of the result toward zero; for example, 7/4 and -7/-4 are truncated to 1 and -7/4 and 7/-4 are truncated to -1
Q(a/b)	Q(a/b) is the quotient part of the division of a by b (a and b positive integers)
R(a/b)	R(a/b) is the remainder of the division of a by b
mod(a,b) (b positive integer)	$\mathrm{mod}(a,b)=\begin{cases}\mathrm{R}(a/b) \text{ if } a \text{ is a positive integer}\\ \mathrm{R}((b - \mathrm{R}(-a/b))/b) \text{ if } a \text{ is a negative integer}\end{cases}$
(mod p)	Modulo p operation

3.3.2 Logical and set operators

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

max [,...,]	The maximum value in the argument list
min [,...,]	The minimum value in the argument list
$\oplus$	Exclusive or
$\cap$	Set intersection
$\cup$	Set union
\	Set exclusion: {-3, -2, ..., 3} \ {0} is the set of integers {-3, -2, -1, 1, 2, 3}

3.3.3 Functions

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

sin	Sine
cos	Cosine
exp	Exponential
$e^{(.)}$	Exponential function
$\sqrt{\ }$	Square root
log10	Logarithm to base 10
j	Imaginary unit, $j^2 = -1$
Rect	$\mathrm{Rect}(x) = \begin{cases}1 \text{ if } 0 \le x < 1\\ 0 \text{ elsewhere}\end{cases}$
δ	Kronecker symbol $\delta(\mathrm{i},\ \mathrm{j}) = \begin{cases}1 \text{ if } i = j\\ 0 \text{ if } i \ne j\end{cases}$

3.3.4 Constants

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

π	3,14159265359...
e	2,71828182846...

3.4 C-language mathematical symbols

3.4.1 Arithmetic operators

For the purposes of the present document, the following C-language mathematical symbols apply:

\+ Addition
\- Subtraction (as a binary operator) or negation (as a unary operator)
++ Increment
\- - Decrement
* Multiplication
DIV Integer division with truncation of the result toward $-\infty$
% Modulo operator. Defined only for positive numbers
$\log_{10}$ Logarithm to base 10

3.4.2 Logical operators

For the purposes of the present document, the following C-language mathematical symbols apply:

|| Logical OR

3.4.3 Relational operators

For the purposes of the present document, the following C-language mathematical symbols apply:

> Greater than
≥ Greater than or equal to
< Less than
≤ Less than or equal to
== Equal to
!= Not equal to

3.4.4 Assignment

For the purposes of the present document, the following C-language mathematical symbols apply:

= Assignment operator

3.4.5 Mnemonics

The following mnemonics are defined to describe the different data types used in the coded DAB audio bit-stream:

bslbf Bit string, left bit first, where "left" is the order in which bit strings are written in the present document. Bit strings are written as a string of 1 s and 0 s within single quote marks, e.g. "1000 0001". Blanks within a bit string are for ease of reading and have no significance
bound Number of first sub-band in joint stereo mode
ch Channel. If ch has the value 0 the left channel of a stereo signal or the first of two independent audio signals is indicated
chlimit Number of channels
dscf Difference between two Scale Factors
gr Granule of three sub-band samples per sub-band
nbal Number of allocated bits per sub-band sample
nch Number of channels; equal to 1 for single channel mode, 2 in other modes
rpchof Remainder polynomial coefficients, highest order first
sb Sub-band
sblimit The number of the lowest sub-band for which no bits are allocated
scfsi Scale Factor selection information
uimsbf Unsigned integer, most significant bit first

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

3.4.6 Method of describing bit stream syntax

The bit stream described in clause 5 is the bit stream that exists in the DAB-receiver at the interface between channel decoder and audio decoder. The bit stream is described using the "C" software language which is used to program the processor which assembles the programme audio and associated data for channel coding. Each data item in the bit stream 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 annex B. The following constructs are used to express the conditions when data elements are present, and are in normal type.

NOTE 1: 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) {	If the condition is true, then the group of data elements occurs next
data_element	in the data stream. This repeats until the condition is not true.
. . .	
}	
do {	
data_element	The data element always occurs at least once.
. . .	
} while (condition)	The data element is repeated until the condition is not true.
if (condition) {	If the condition is true, then the first group of data elements occurs next
data_element	in the data stream.
. . .	
}	
else {	If the condition is not true, then the second group of data elements
data_element	occurs next in the data stream.
. . .	
}	
for (expr1; expr2; expr3) {	expr1 is an expression specifying the initialization of the loop.
data_element	Normally it specifies the initial state of the counter. expr2 is a condition specifying a test made before each iteration of the loop.
. . .	The loop terminates when the condition is not true.
}	expr3 is an expression that is performed at the end of each iteration of the loop, normally it increments a counter.

NOTE 2: The most common usage of this construct is as follows.

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

As noted, the group of data elements may contain nested conditional constructs. For compactness, the {} may be 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+1)th 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]	is the inclusive range of bits between bit m and bit n in the data_element.

3.5 Convention

Unless otherwise stated, the following notation, regarding the order of bits within each step of processing is used:

- in figures, the bit shown in the left hand position is considered to be first;
- in tables, the bit shown in the left hand position is considered to be first;
- in byte fields, the Most Significant bit (MSb) is considered to be first and denoted by the higher number. For example, the MSb of a single byte is denoted "b_7" and the Least Significant bit (LSb) is denoted "b_0";
- in vectors (mathematical expressions), the bit with the lowest index is considered to be first.

NOTE: Due to time-interleaving, this order of bits is not the true transmission order.

4 Introduction

The DAB system standard [1] allows audio (programme) services to be carried using either DAB audio or DAB+ audio. The present document defines the way that audio (programme) services are carried when using DAB audio (MPEG layer II audio coding).

Two sampling rates are permitted for DAB audio coding, 48 kHz and 24 kHz. Each audio frame contains samples for 24 ms or 48 ms, respectively, and each frame has a constant size. The audio frames are carried in one or two DAB logical frames, respectively. A range of bit rates and audio modes are available and the addition of Programme Associated Data (PAD) allows supplementary content to be provided.

5 DAB audio coding

5.1 Introduction

An overview of the principal functions of the audio coding scheme is shown in the simplified block diagram of the DAB audio encoder (see figure 1). The main characteristics of the audio coding system, like audio modes, bit rates and audio frame length are given in clause A.2, whereas the characteristics of the input audio signal are given in clause A.1.

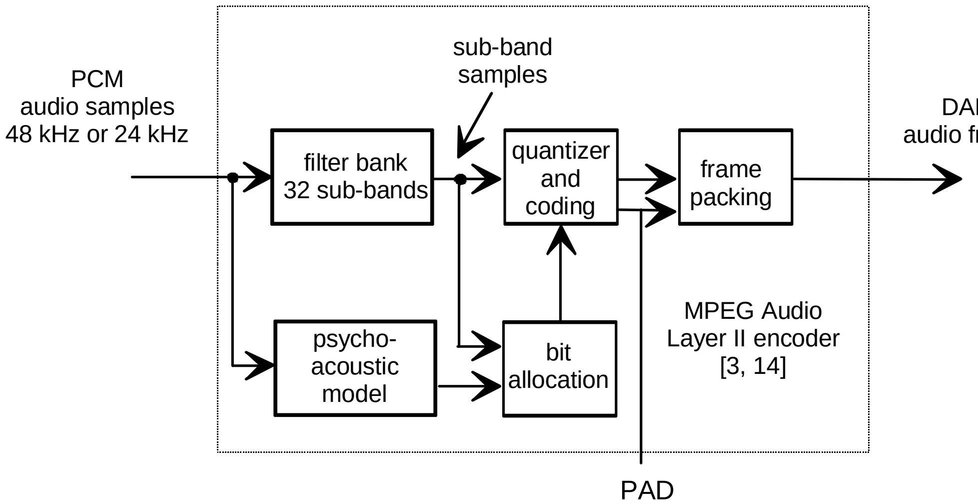

Figure 1: Simplified block diagram of the DAB audio encoder

The input PCM audio samples are fed into the audio encoder. A filter bank creates a filtered and sub-sampled representation of the input audio signal. The filtered samples are called sub-band samples. A psychoacoustic model of the human ear should create a set of data to control the quantizer and coding. These data can be different depending on the actual implementation of the encoder. An estimation of the masking threshold can be used to obtain these quantizer control data. The quantizer and coding block shall create a set of coding symbols from the sub-band samples. The frame packing block shall assemble the actual audio bit stream from the output data of the previous block, and shall add other information, such as header information, CRC words for error detection and Programme Associated Data (PAD), which are intimately related with the coded audio signal. For a sampling frequency of 48 kHz, the resulting audio frame corresponds to 24 ms duration of audio and shall comply with the Layer II format, ISO/IEC 11172-3 [2]. The audio frame shall map on to the logical frame structure in such a way that the first bit of the DAB audio frame corresponds to the first bit of a logical frame. For a sampling frequency of 24 kHz, the resulting audio frame corresponds to 48 ms duration of audio and shall comply with the Layer II LSF format, ISO/IEC 13818-3 [4]. The audio frame shall map on to the logical frame structure in such a way that the first bit of the DAB audio frame corresponds to the first bit of a logical frame (this may be associated with either an "even" or an "odd" logical frame count). The formatting of the DAB audio frame shall be done in such a way that the structure of the DAB audio frame conforms to the audio bit stream syntax described in clause 5.4.

The simplified block diagram of the audio decoder in the receiver, shown in figure 2, accepts the DAB audio frame in the syntax defined in clause 5.4.2 which is a conformant subset of the MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) bit stream syntax defined in clause 5.4.1. This allows the use of an MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) decoder. The DAB audio frame shall be fed into the audio decoder, which unpacks the data of the frame to recover the various elements of information. The reconstruction block shall reconstruct the quantized sub-band samples. An inverse filter bank shall transform the sub-band samples back to produce digital PCM audio signals in the case of ISO/IEC 11172-3 [2] at 48 kHz sampling frequency and in the case of ISO/IEC 13818-3 [4] at 24 kHz according to annex B.

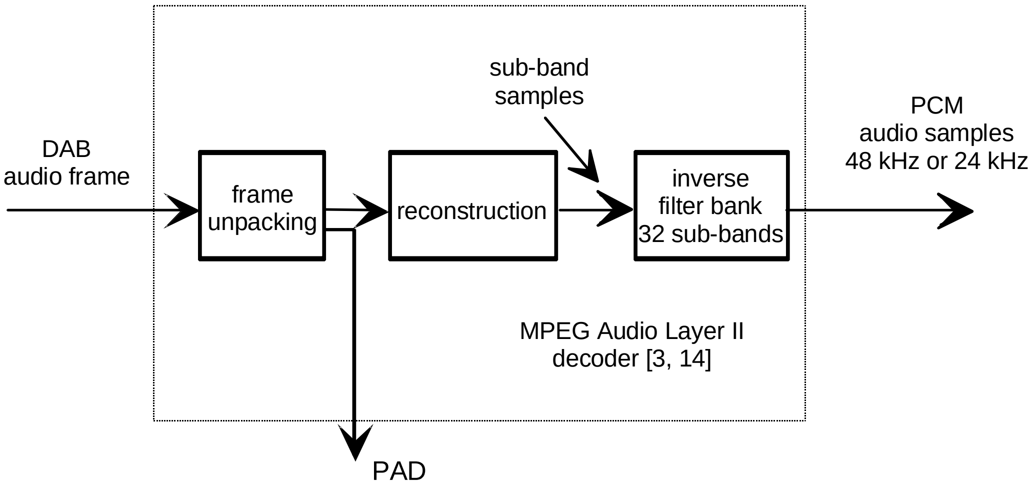

Figure 2: Simplified block diagram of the DAB audio decoder

5.2 Audio encoding

5.2.0 General

The source encoder for the DAB system is the MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) encoder with restrictions on some parameters and some additional protection against transmission errors. In the ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4] International Standards only the encoded audio bit stream, rather than the encoder, and the decoder are specified. In subsequent clauses, both normative and informative parts of the encoding technique are described. An example of one complete suitable encoder with the corresponding flow diagram (figure 3) is given in the following clauses.

The DAB source coding algorithm is based on a perceptual coding technique. The six primary parts of such an audio encoding technique are:

1) analysis sub-band filter (clauses 5.2.1 and C.1);

2) Scale Factor calculation (clauses 5.2.2 to 5.2.4);

3) psychoacoustic model (clauses 5.2.5 and C.2);

4) bit allocation procedure (clauses 5.2.6 and C.3);

5) quantizing and coding (clauses 5.2.7 and 5.2.8);

6) bit stream formatter (clause 5.2.9).

5.2.1 Analysis sub-band filter

An analysis sub-band filter should be used to split the broadband audio signal with sampling frequency f_s into 32 equally spaced sub-bands, each with a sampling frequency of $f_s/32$. This filter, called a poly-phase analysis filter bank, is critically sampled (i.e. there are as many samples in the sub-band domain as there are in the time domain). A detailed description of a suitable analysis sub-band filter bank with the appropriate formulae, coefficients and flow charts is provided in clause C.1.

The encoding algorithm provides a frequency response down to 0 Hz. However, in applications where this is not desirable, a high-pass filter should be included at the audio input of the encoder. The application of such a high-pass filter avoids an unnecessarily high bit rate requirement for the lowest sub-band and may increase the overall audio quality. The cut-off frequency should be in the range of 2 Hz to 10 Hz.

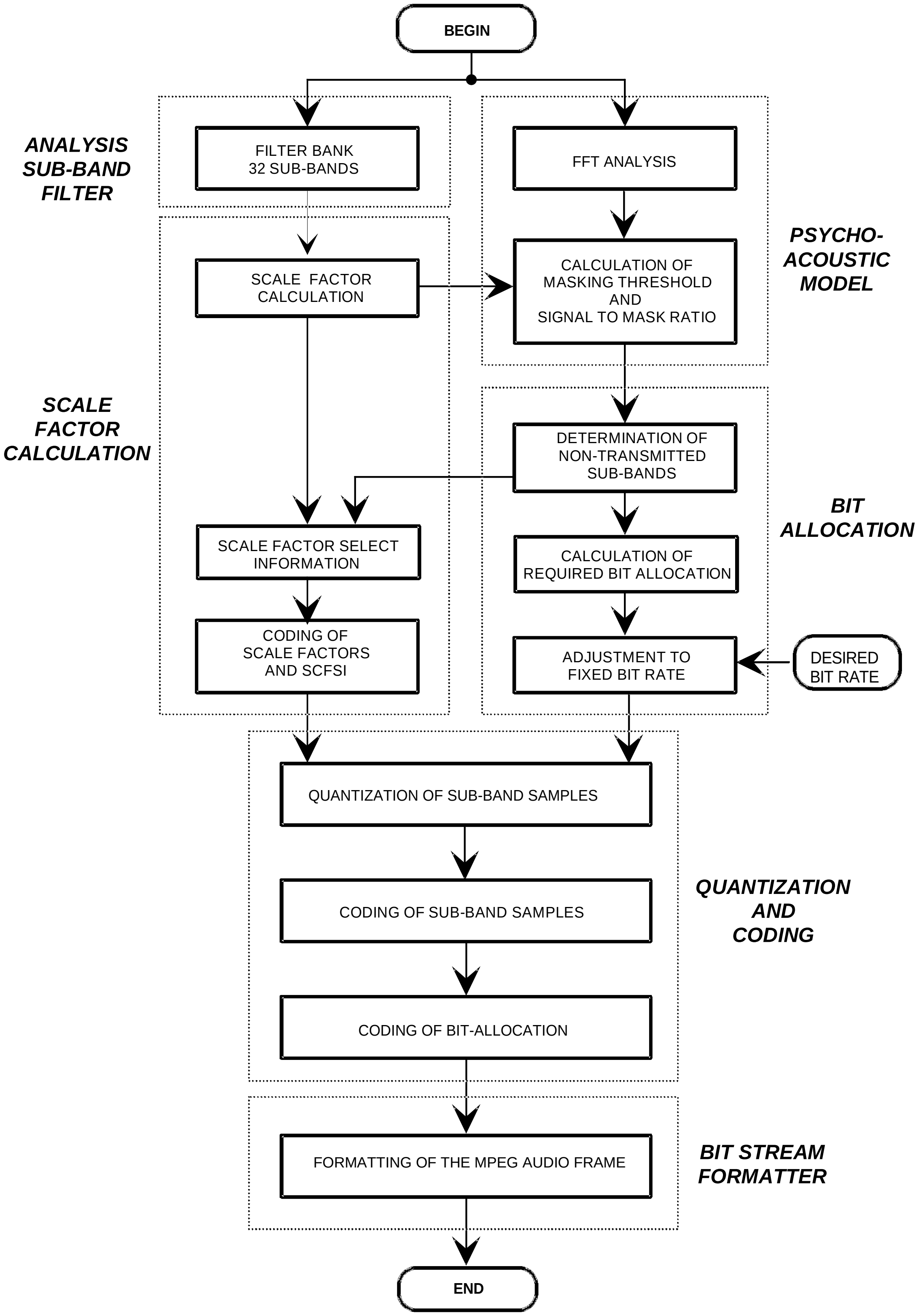

Figure 3: Flow diagram of the MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) encoder

5.2.2 Scale Factor calculation

In each sub-band, 36 samples shall be grouped for processing. Before quantization, the output samples of the filter bank should be normalized. The calculation of the Scale Factor (ScF) for each sub-band shall be performed every 12 sub-band samples. The maximum of the absolute value of these 12 samples shall be determined. The lowest value, given by the column "Scale Factor" in table 1, which is larger than this maximum shall be used as the ScF of the 12 sub-band samples.

Table 1: Scale Factors

Index iscf	Scale Factor ScF	Index iscf	Scale Factor ScF
0	2,00000000000000	32	0,00123039165029
1	1,58740105196820	33	0,00097656250000
2	1,25992104989487	34	0,00077509816991
3	1,00000000000000	35	0,00061519582514
4	0,79370052598410	36	0,00048828125000
5	0,62996052494744	37	0,00038754908495
6	0,50000000000000	38	0,00030759791257
7	0,39685026299205	39	0,00024414062500
8	0,31498026247372	40	0,00019377454248
9	0,25000000000000	41	0,00015379895629
10	0,19842513149602	42	0,00012207031250
11	0,15749013123686	43	0,00009688727124
12	0,12500000000000	44	0,00007689947814
13	0,09921256574801	45	0,00006103515625
14	0,07874506561843	46	0,00004844363562
15	0,06250000000000	47	0,00003844973907
16	0,04960628287401	48	0,00003051757813
17	0,03937253280921	49	0,00002422181781
18	0,03125000000000	50	0,00001922486954
19	0,02480314143700	51	0,00001525878906
20	0,01968626640461	52	0,00001211090890
21	0,01562500000000	53	0,00000961243477
22	0,01240157071850	54	0,00000762939453
23	0,00984313320230	55	0,00000605545445
24	0,00781250000000	56	0,00000480621738
25	0,00620078535925	57	0,00000381469727
26	0,00492156660115	58	0,00000302772723
27	0,00390625000000	59	0,00000240310869
28	0,00310039267963	60	0,00000190734863
29	0,00246078330058	61	0,00000151386361
30	0,00195312500000	62	0,00000120155435
31	0,00155019633981		

5.2.3 Coding of Scale Factors

This clause is partly of informative, and partly of normative nature. The index "iscf" in table 1 is represented by 6 bits, MSb first. The ScF of a certain sub-band shall be transmitted only if a non-zero number of bits has been allocated to this sub-band.

A DAB audio frame corresponds to 36 sub-band samples and therefore contains three ScFs per sub-band. Some may not be transmitted. This clause gives information about which ScFs should be transmitted, and how they shall be encoded.

The two differences $dscf_1$ and $dscf_2$ of the successive ScF indices $iscf_1$, $iscf_2$ and $iscf_3$ shall be calculated as follows:

- $dscf_1 = iscf_1 - iscf_2$;
- $dscf_2 = iscf_2 - iscf_3$.

Five classes of ScF difference shall be defined. The class of each of the differences should be determined by the following table 2.

Table 2: ScF difference classes

class	dscf
1	dscf ≤ -3
2	-3 < dscf < 0
3	dscf = 0
4	0 < dscf < 3
5	dscf ≥ 3

Table 3: ScF transmission patterns

$Class_1$	$Class_2$	Scale Factors used in Encoder	Transmission Pattern	Scale Factor Select. Information (ScFSI)	Code
1	1	1 2 3	1 2 3	0	00
1	2	1 2 2	1 2	3	11
1	3	1 2 2	1 2	3	11
1	4	1 3 3	1 3	3	11
1	5	1 2 3	1 2 3	0	00
2	1	1 1 3	1 3	1	01
2	2	1 1 1	1	2	10
2	3	1 1 1	1	2	10
2	4	4 4 4	4	2	10
2	5	1 1 3	1 3	1	01
3	1	1 1 1	1	2	10
3	2	1 1 1	1	2	10
3	3	1 1 1	1	2	10
3	4	3 3 3	3	2	10
3	5	1 1 3	1 3	1	01
4	1	2 2 2	2	2	10
4	2	2 2 2	2	2	10
4	3	2 2 2	2	2	10
4	4	3 3 3	3	2	10
4	5	1 2 3	1 2 3	0	00
5	1	1 2 3	1 2 3	0	00
5	2	1 2 2	1 2	3	11
5	3	1 2 2	1 2	3	11
5	4	1 3 3	1 3	3	11
5	5	1 2 3	1 2 3	0	00

The pair of difference classes shall indicate the entry point in table 3. For each pair of difference classes the actual transmission pattern of Scale Factors and the actual Scale Factor Selection Information (ScFSI) shall be determined from table 3.

Only the Scale Factors indicated in the "transmission pattern" shall be transmitted. A "1", "2" or "3" means that the first, second or third Scale Factor, respectively, is transmitted within an audio frame. A "4" means that the maximum of the three Scale Factors is transmitted. If two or three of the Scale Factors are the same, not all Scale Factors should be transmitted for a certain sub-band within one audio frame. The information describing the number and the position of the Scale Factors in each sub-band is called "Scale Factor Select. Information" (ScFSI).

5.2.4 Coding of Scale Factor Selection Information

The ScFSI shall be coded by an unsigned two bit binary word, MSb first, which is also to be found in table 3, showing the Scale Factor transmission patterns. Only the ScFSI for the sub-bands which will have a non-zero bit allocation shall be transmitted.

5.2.5 Psychoacoustic model

A psychoacoustic model should calculate a just-noticeable noise-level for each sub-band in the filter bank. This noise level should be used in the bit allocation procedure to determine the actual quantizer for each sub-band. The final output of the model is a Signal-to-Mask Ratio (SMR) for each sub-band. For a high coding efficiency, it is recommended to use a psychoacoustic model with an appropriate frequency analysis. An example of a reference psychoacoustic model is presented in clause C.2.

5.2.6 Bit allocation

A bit allocation procedure shall be applied. Different strategies for allocating the bits to the sub-band samples of the individual sub-bands are possible. A reference model of the bit allocation procedure is described in clause C.3. The principle used in this allocation procedure is minimization of the total noise-to-mask ratio over the audio frame with the constraint that the number of bits used does not exceed the number of bits available for that DAB audio frame. The allocation procedure should consider both the output samples from the filter bank and the Signal-to-Mask-Ratios (SMRs) from the psychoacoustic model. The procedure should assign a number of bits to each sample (or group of samples) in each sub-band, in order to simultaneously meet both the bit rate and masking requirements. At low bit rates, when the demand derived from the masking threshold cannot be met, the allocation procedure should attempt to spread bits in a psychoacoustically inoffensive manner among the sub-bands.

After determining, how many bits should be distributed to each sub-band signal, the resulting number shall be used to code the sub-band samples, the ScFSI and the ScFs. Only a limited number of quantizations is allowed for each sub-band.

In the case of 48 kHz sampling frequency tables 5 and 6 indicate for every sub-band the number of quantization steps which shall be used to quantize the sub-band samples. Table 4 shall be used for bit rates of 56 kbit/s to 192 kbit/s in single channel mode as well as for 112 kbit/s to 384 kbit/s in all other audio modes. The number of the lowest sub-band for which no bits are allocated, called "sblimit", equals 27, and the total number of bits used for the bit allocation per audio frame is defined by the sum of "nbal". If "sblimit" is equal to 27, the sum of "nbal" is equal to 88 for single channel mode, whereas the sum of "nbal" is equal to 176 for stereo mode. This number is smaller, if the joint stereo mode is used. Table 5 shall be used for bit rates of 32 kbit/s and 48 kbit/s in single channel mode, as well as for 64 kbit/s and 96 kbit/s in all other audio modes. In this case "sblimit" is equal to 8, and the total number of bits used for the bit allocation per audio frame, i.e. sum of "nbal" is equal to 26 for single channel mode, whereas the sum of "nbal" is equal to 52 for stereo mode. This number is 40, if joint stereo mode with mode_extension "00" is used.

In the case of 24 kHz sampling frequency, table 6 indicates for every sub-band the number of quantization steps which shall be used to quantize the sub-band samples. Other than in the case of 48 kHz sampling frequency, table 6 shall be used for all bit rates which are specified for MPEG-2 Audio Layer II ISO/IEC 13818-3 [4] low sampling frequency coding, in the range of 8 kbit/s to 160 kbit/s, independent of the audio mode.

The number of the lowest sub-band for which no bits are allocated, called "sblimit", equals 30, and the total number of bits used for the bit allocation per audio frame is defined by the sum of "nbal". The sum of "nbal" is equal to 75 for single channel mode, whereas the sum of "nbal" is equal to 150 for stereo mode. This number is smaller, if the joint stereo mode is used.

The number of bits required to represent these quantized sub-band samples shall be derived from the last two columns of table 8.

5.2.7 Bit allocation coding

In order to increase the coding efficiency, only a limited number of possible quantizations are permitted. Both the number and the quantizations may be different from one sub-band (denoted as "sb" in tables 5 to 7) to another. Only the index with word length "nbal" given in tables 5 to 7, which depends on the bit rate and audio mode, shall be transmitted, MSb first.

Table 4: Bit allocation and possible quantization per sub-band for 48 kHz sampling frequency

Bit rates: 56 kbit/s, 64 kbit/s, 80 kbit/s, 96 kbit/s, 112 kbit/s, 128 kbit/s, 160 kbit/s and 192 kbit/s (single channel mode) Bit rates: 112 kbit/s, 128 kbit/s, 160 kbit/s, 192 kbit/s, 224 kbit/s, 256 kbit/s, 320 kbit/s and 384 kbit/s (all other audio modes) **index --->**																	
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
sb	**nbal**								**nlevels**								
0	**4**	-	3	7	15	31	63	127	255	511	1 023	2 047	4 095	8 191	16 383	32 767	65 535
1	**4**	-	3	7	15	31	63	127	255	511	1 023	2 047	4 095	8 191	16 383	32 767	65 535
2	**4**	-	3	7	15	31	63	127	255	511	1 023	2 047	4 095	8 191	16 383	32 767	65 535
3	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	65 535
4	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	65 535
5	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	65 535
6	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	65 535
7	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	65 535
8	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	65 535
9	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	65 535
10	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	65 535
11	**3**	-	3	5	7	9	15	31	65 535								
12	**3**	-	3	5	7	9	15	31	65 535								
13	**3**	-	3	5	7	9	15	31	65 535								
14	**3**	-	3	5	7	9	15	31	65 535								
15	**3**	-	3	5	7	9	15	31	65 535								
16	**3**	-	3	5	7	9	15	31	65 535								
17	**3**	-	3	5	7	9	15	31	65 535								
18	**3**	-	3	5	7	9	15	31	65 535								
19	**3**	-	3	5	7	9	15	31	65 535								
20	**3**	-	3	5	7	9	15	31	65 535								
21	**3**	-	3	5	7	9	15	31	65 535								
22	**3**	-	3	5	7	9	15	31	65 535								
23	**2**	-	3	5	65 535												
24	**2**	-	3	5	65 535												
25	**2**	-	3	5	65 535												
26	**2**	-	3	5	65 535												
27	**0**	-															
28	**0**	-															
29	**0**	-															
30	**0**	-															
31	**0**	-															

Table 5: Bit allocation and possible quantization per sub-band for 48 kHz sampling frequency

Bit rates: 32 kbit/s and 48 kbit/s (single channel mode) Bit rates: 64 kbit/s and 96 kbit/s (all other audio modes) **index --->**																	
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
sb	**nbal**	**nlevels**															
0	**4**	-	3	5	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	16 383	32 767
1	**4**	-	3	5	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	16 383	32 767
2	**3**	-	3	5	9	15	31	63	127								
3	**3**	-	3	5	9	15	31	63	127								
4	**3**	-	3	5	9	15	31	63	127								
5	**3**	-	3	5	9	15	31	63	127								
6	**3**	-	3	5	9	15	31	63	127								
7	**3**	-	3	5	9	15	31	63	127								
8	**0**	-															
9	**0**	-															
10	**0**	-															
11	**0**	-															
12	**0**	-															
13	**0**	-															
14	**0**	-															
15	**0**	-															
16	**0**	-															
17	**0**	-															
18	**0**	-															
19	**0**	-															
20	**0**	-															
21	**0**	-															
22	**0**	-															
23	**0**	-															
24	**0**	-															
25	**0**	-															
26	**0**	-															
27	**0**	-															
28	**0**	-															
29	**0**	-															
30	**0**	-															
31	**0**	-															

Table 6: Bit allocation and possible quantization per sub-band for 24 kHz sampling frequency

Bit rates: 8 kbit/s, 16 kbit/s, 24 kbit/s, 32 kbit/s, 40 kbit/s, 48 kbit/s, 56 kbit/s, 64 kbit/s, 80 kbit/s, 96 kbit/s, 112 kbit/s, 128 kbit/s, 144 kbit/s and 160 kbit/s (all audio modes) **index --->**																	
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
sb	**nbal**	**nlevels**															
0	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	16 383
1	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	16 383
2	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	16 383
3	**4**	-	3	5	7	9	15	31	63	127	255	511	1 023	2 047	4 095	8 191	16 383
4	**3**	-	3	5	9	15	31	63	127								
5	**3**	-	3	5	9	15	31	63	127								
6	**3**	-	3	5	9	15	31	63	127								
7	**3**	-	3	5	9	15	31	63	127								
8	**3**	-	3	5	9	15	31	63	127								
9	**3**	-	3	5	9	15	31	63	127								
10	**3**	-	3	5	9	15	31	63	127								
11	**2**	-	3	5	9												
12	**2**	-	3	5	9												
13	**2**	-	3	5	9												
14	**2**	-	3	5	9												
15	**2**	-	3	5	9												
16	**2**	-	3	5	9												
17	**2**	-	3	5	9												
18	**2**	-	3	5	9												
19	**2**	-	3	5	9												
20	**2**	-	3	5	9												
21	**2**	-	3	5	9												
22	**2**	-	3	5	9												
23	**2**	-	3	5	9												
24	**2**	-	3	5	9												
25	**2**	-	3	5	9												
26	**2**	-	3	5	9												
27	**2**	-	3	5	9												
28	**2**	-	3	5	9												
29	**2**	-	3	5	9												
30	**0**	-															
31	**0**	-															

5.2.8 Quantization and coding of sub-band samples

A quantization process of the sub-band samples shall be applied. The following description of this process is informative, but the coding of the sub-band samples has to follow normative rules.

Each of the 12 consecutive sub-band samples, which are grouped together for the scaling process, should be normalized by dividing its value by the Scale Factor to obtain a value denoted X and quantized using the following procedure:

- calculate $A \times X + B$;
- take the n most significant bits;
- invert the MSb.

Table 7: Quantization coefficients

No. of steps	A	B
3	0,750000000	-0,250000000
5	0,625000000	-0,375000000
7	0,875000000	-0,125000000
9	0,562500000	-0,437500000
15	0,937500000	-0,062500000
31	0,968750000	-0,031250000
63	0,984375000	-0,015625000
127	0,992187500	-0,007812500
255	0,996093750	-0,003906250
511	0,998046875	-0,001953125
1 023	0,999023438	-0,000976563
2 047	0,999511719	-0,000488281
4 095	0,999755859	-0,000244141
8 191	0,999877930	-0,000122070
16 383	0,999938965	-0,000061035
32 767	0,999969482	-0,000030518
65 535	0,999984741	-0,000015259

The quantization coefficients **A** and **B** can be found in table 7. The number n of bits per codeword, given in table 8, represents the number of bits necessary to encode the number of quantization steps. The inversion of the MSb shall be done in order to avoid the all "1" code that is used for the synchronization word in the MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) header.

Three consecutive sub-band samples, called a granule, shall be considered for coding. Table 8 gives the number of quantization steps that the samples will be quantized to. The same table specifies, whether grouping of a granule shall be used or not. If grouping is not required, the three samples shall be coded with three individual codewords.

If grouping of a granule is required, which depends on the number of quantization steps m (m = 3, 5 or 9), the three consecutive sub-band samples shall be coded with one codeword. Only one value v_m, Most Significant Byte (MSB) first, shall be transmitted for this grouped granule. The relationship between the coded value v_m and the three samples x, y, z of a granule shall be one of the following:

- v3 = 9z + 3y + x (v3 in 0... 26)
- v5 = 25z + 5y + x (v5 in 0...124)
- v9 = 81z + 9y + x (v9 in 0...728)

Table 8: Classes of quantization

No. of steps	Grouping	Samples per codeword	n bits per codeword
3	yes	3	5
5	yes	3	7
7	no	1	3
9	yes	3	10
15	no	1	4
31	no	1	5
63	no	1	6
127	no	1	7
255	no	1	8
511	no	1	9
1 023	no	1	10
2 047	no	1	11
4 095	no	1	12
8 191	no	1	13
16 383	no	1	14
32 767	no	1	15
65 535	no	1	16

5.2.9 Formatting of the audio bit stream

The frame formatter of the audio encoder shall take the bit allocation, ScFSI, ScF and the quantized sub-band samples together with header information and a few code words used for error detection to format the MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) bit stream. It shall further divide this bit stream into audio frames, each corresponding to 1152 PCM audio samples, which is equivalent to a duration of 24 ms in the case of 48 kHz sampling frequency and 48 ms in the case of 24 kHz sampling frequency. The principal structure of such an MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) frame with its correspondence to the DAB audio frame can be seen in figure 4.

Each audio frame starts with a header, consisting of a syncword and audio system related information. A Cyclic Redundancy Check (CRC), following the header protects a part of the header information, the bit allocation, and the ScFSI fields. After the CRC follows bit allocation, ScFSI and Scale Factors. The sub-band samples, which will be used by the decoder to reconstruct the PCM audio signal, are the last audio data part in the MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) frame before the ancillary data field. This ancillary data field, which is of variable length, is located at the end of the MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) frame. The details of the content of the audio frame can be found in clause 5.4.

An adaptation of the MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) frame to the DAB audio frame is performed in order to introduce:

- specific DAB Scale Factor Error Check (ScF-CRC);
- a fixed and a variable field of Programme Associated Data (F-PAD and X-PAD).

The lower part of figure 4 indicates how this additional specific information, necessary for DAB, shall be inserted into the ancillary data field of the MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) frame.

For MPEG-1 Audio (ISO/IEC 11172-3 [2]) the whole DAB audio frame fits exactly into a DAB logical frame. However, for LSF-coding which is standardized in MPEG-2 Audio (ISO/IEC 13818-3 [4]), the DAB LSF audio frame shall be divided into two subframes of equal length and each subframe fits into two consecutive DAB logical frames.

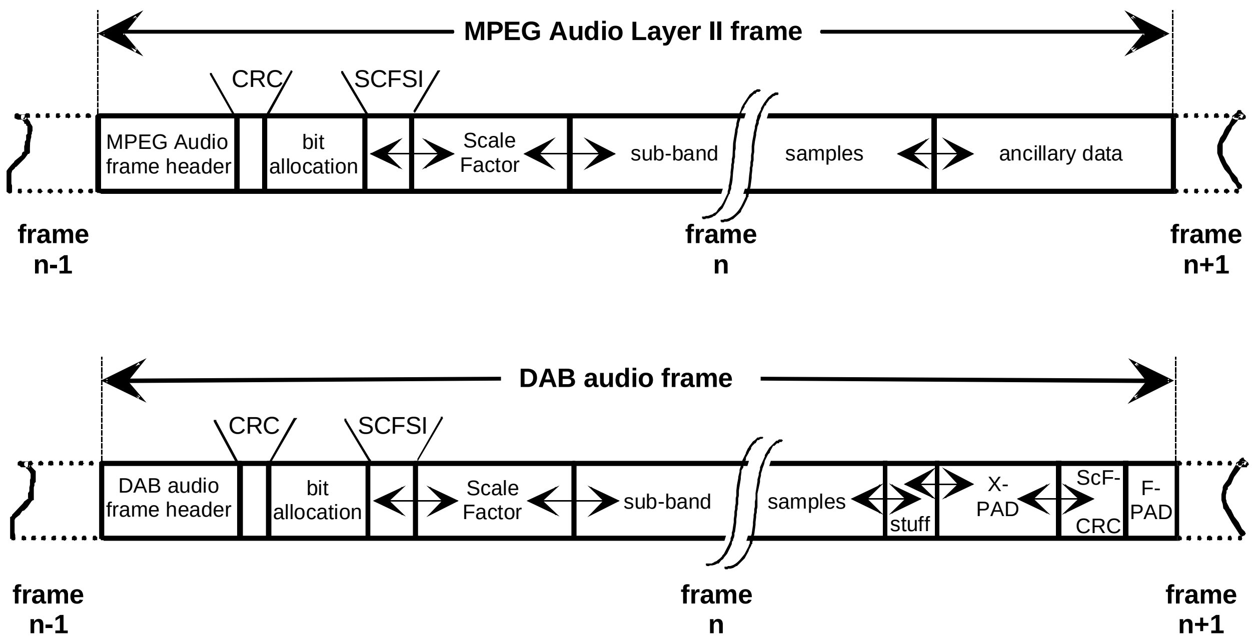

Figure 4: Frame structure of MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) and corresponding DAB audio frame

The first four bytes of the DAB audio frame contain the MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) header. This header carries information for the audio decoder. In the DAB system, some of this information is currently defined as static information. This is:

- **syncword:** set to external synchronization of the audio decoder;
- **layer:** set to Layer II (layer = Layer II);

- **protection_bit:** set to CRC protection on.

5.3 Semantics of the audio bit stream

5.3.1 MPEG Audio Layer II bit stream

5.3.1.1 Audio sequence

The DAB audio coding system uses the MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) format. A graphic representation of an audio frame in MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) format is given in the upper part of figure 4.

Audio_frame: part of the bit stream that is decodable by itself. It contains information for 1 152 samples. It starts with a syncword, and ends just before the next syncword. It consists of an integer number of slots. A slot contains one byte.

5.3.1.2 Audio frame

header: part of the bit stream containing synchronization and state information.

error_check: part of the bit stream containing information for error detection of significant audio side information.

audio_data: part of the bit stream containing information on the audio samples.

ancillary_data: part of the bit stream that may be used for ancillary data.

5.3.1.3 Audio frame header

The first 32 bit (four bytes) are header information.

syncword: the bit string "1111 1111 1111".

ID (Identifier): this 1-bit flag shall identify the audio coding algorithm, as follows:

0: ISO/IEC 13818-3 [4] or MPEG-2 Audio extension to lower sampling frequencies;

1: ISO/IEC 11172-3 [2].

Layer: this 2-bit field shall indicate which layer is used, according to table 9.

Table 9: Indication of MPEG Audio Layer

Code	Layer
"11"	not used in DAB
"10"	Layer II
"01"	not used in DAB
"00"	reserved

protection_bit: this 1-bit flag shall indicate whether redundancy has been added into the audio bit stream in order to facilitate error detection and concealment. The bit shall be set to "0" because redundancy is added for DAB application.

bit_rate_index: indicates the bit rate. The bit_rate_index is an index to a specified bit rate, shown in table 10 for 48 kHz sampling frequency and shown in table 11 for 24 kHz sampling frequency.

Table 10: Specified total bit rates per audio programme for 48 kHz sampling frequency

bit_rate_index	bit rate specified
"0000"	not used in DAB
"0001"	32 kbit/s
"0010"	48 kbit/s
"0011"	56 kbit/s
"0100"	64 kbit/s
"0101"	80 kbit/s
"0110"	96 kbit/s
"0111"	112 kbit/s
"1000"	128 kbit/s
"1001"	160 kbit/s
"1010"	192 kbit/s
"1011"	224 kbit/s
"1100"	256 kbit/s
"1101"	320 kbit/s
"1110"	384 kbit/s
"1111"	forbidden

Table 11: Specified total bit rates per audio programme for 24 kHz sampling frequency

bit_rate_index	bit rate specified
"0000"	not used in DAB
"0001"	8 kbit/s
"0010"	16 kbit/s
"0011"	24 kbit/s
"0100"	32 kbit/s
"0101"	40 kbit/s
"0110"	48 kbit/s
"0111"	56 kbit/s
"1000"	64 kbit/s
"1001"	80 kbit/s
"1010"	96 kbit/s
"1011"	112 kbit/s
"1100"	128 kbit/s
"1101"	144 kbit/s
"1110"	160 kbit/s
"1111"	forbidden

The bit_rate_index indicates the total bit rate irrespective of the mode (stereo, joint_stereo, single_channel). The total bit rate includes all bits in an audio frame, i.e. all bits necessary for header, audio signal, PAD and error detection information.

NOTE: In order to provide the smallest possible delay and complexity, the decoder is not required to support a continuously variable bit rate. However, the bit rate may change from time to time during continuing service. The smallest resolution for changing the bit rate is 6 seconds (see ETSI EN 300 401 [1], clause 6.5 on Multiplex reconfiguration).

The encoder in the DAB transmitter should support at least one of the bit rates given in tables 11 and 12, whereas the audio decoder shall be capable of working at all these bit rates. For 48 kHz sampling frequency, not all combinations of total bit rates and audio modes are allowed. Table 12 shows the audio modes which can be chosen, dependent on the bit rate.

Table 12: Combinations of total bit rates per audio programme and audio modes for 48 kHz sampling frequency

Total bit rate	Audio modes
32 kbit/s	single_channel
48 kbit/s	single_channel
56 kbit/s	single_channel
64 kbit/s	single_channel, stereo, intensity_stereo
80 kbit/s	single_channel
96 kbit/s	single_channel, stereo, intensity_stereo
112 kbit/s	single_channel, stereo, intensity_stereo
128 kbit/s	single_channel, stereo, intensity_stereo
160 kbit/s	single_channel, stereo, intensity_stereo
192 kbit/s	single_channel, stereo, intensity_stereo
224 kbit/s	stereo, intensity_stereo
256 kbit/s	stereo, intensity_stereo
320 kbit/s	stereo, intensity_stereo
384 kbit/s	stereo, intensity_stereo

sampling_frequency: indicates, depending on the value of the ID, the sampling frequency, according to table 13. The DAB system uses the value of "01", indicating for the ID bit equals "1" a sampling frequency of 48 kHz and for the ID bit equals "0" a sampling frequency of 24 kHz.

Table 13: Specified sampling frequencies per PCM audio input/output signal

sampling_frequency	ID = "0" frequency specified	ID = "1" frequency specified
"00"	not used in DAB	not used in DAB
"01"	24 kHz	48 kHz
"10"	not used in DAB	not used in DAB
"11"	reserved for future use	reserved for future use

padding_bit: fixed value of "0". No padding is necessary for 24 kHz and 48 kHz sampling frequency.

private_bit: bit for private use. This bit will not be used in the future by MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]), and is not interpreted by an MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) decoder.

mode: indicates the audio mode according to table 14. In Layer II the joint_stereo mode is intensity_stereo.

Table 14: Audio modes which can be selected in the audio encoder

mode	Audio mode specified
"00"	stereo
"01"	joint_stereo (intensity_stereo)
"10"	not used in DAB
"11"	single_channel

In all audio modes, except joint_stereo, the value of bound equals sblimit. In joint_stereo mode the bound is determined by the mode_extension.

mode_extension: these bits are used in joint_stereo mode, and indicate which sub-bands are in intensity_stereo. All other sub-bands are coded in stereo mode. The figures are given by table 15.

Table 15: Sub-bands in intensity stereo mode, indicated by the mode_extension

mode_extension	Sub-bands in intensity_stereo	Bound
"00"	4 to 31	bound == 4
"01"	8 to 31	bound == 8
"10"	12 to 31	bound == 12
"11"	16 to 31	bound == 16

copyright: if this bit equals "0" there is no copyright on the MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) audio bit stream; "1" means copyright protected.

original/copy: this bit equals "0" if the bit stream is a copy, "1" if it is an original.

emphasis: indicates the type of de-emphasis that shall be used by an MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) decoder. The DAB system shall use a fixed value of "00", indicating "no emphasis", and therefore the use of pre/de-emphasis is excluded (see table 16).

Table 16: Emphasis of the input PCM audio signal

emphasis	emphasis specified
"00"	no emphasis
"01"	not used in DAB
"10"	not used in DAB
"11"	not used in DAB

5.3.1.4 Error check

crc_check: a 16 bit parity check word used for error detection of the most error-sensitive part of the audio information within the encoded audio bit stream. This information includes the third and fourth bytes of the MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) header, bit allocation and Scale Factor Select Information (see for more details clause B.2).

5.3.1.5 Audio data

allocation[ch][sb]: contains information concerning the quantizers used for the samples in sub-band sb in channel ch, whether the information on three consecutive samples of a granule has been grouped to one code, and on the number of bits used to code the samples. The meaning and length of this field depends on the number of the sub-band, the bit rate, and the sampling frequency. The bits in this field form an unsigned integer used as an index to the relevant bit allocation tables 5 to 7, which gives the number of levels "nlevels" used for quantization. For sub-bands in intensity_stereo mode the bit-stream contains only one allocation data element per sub-band.

Table 17: Transmission of Scale Factors dependent on ScFSI

scfsi [ch][sb]	action
"00"	three ScFs transmitted, for parts 0, 1, 2 respectively
"01"	two ScFs transmitted, first one valid for parts 0 and 1, second one for part 2
"10"	one ScF transmitted, valid for all three parts
"11"	two ScFs transmitted, first one valid for part 0, the second one for parts 1 and 2

scfsi[ch][sb]: Scale Factor Selection Information. This gives information on the number of Scale Factors transferred for sub-band sb in channel ch and for which parts of the signal in this frame they are valid (see table 17). The frame is divided into three equal parts of 12 sub-band samples each per sub-band.

scalefactor[ch][sb][p]: indicates the factor by which the re-quantized samples of sub-band sb in channel ch and of part p of the frame should be multiplied. The six bits constitute an unsigned integer, index to table 1, showing the Scale Factors.

grouping[ch][sb]: is a function that determines, whether grouping is applied for coding of samples in sub-band sb of channel ch. Grouping means, that three consecutive samples of the current sub-band sb in channel ch which form the granule gr are coded and transmitted using one common codeword and not using three separate codewords. Grouping[ch][sb] is true, if in the bit allocation table currently in use (see either tables 5 to7) the value found under sb (first row) and allocation[ch][sb] (column) is either 3, 5 or 9. Otherwise it is false. For sub-bands in intensity stereo mode the grouping is valid for both channels.

samplecode[ch][sb][gr]: coded representation of the three consecutive samples in the granule gr in sub-band sb of channel ch. For sub-bands in intensity stereo mode the coded representation of the samplecode is valid for both channels.

sample[ch][sb][s]: coded representation of the s-th sample in sub-band sb of channel ch. For sub-bands in intensity stereo mode the coded representation of the sample is valid for both channels.

5.3.1.6 Ancillary data

ancillary_bit: user definable.

The number of ancillary bits (no_of_ancillary_bits) equals the available number of bits in an audio frame minus the number of bits actually used for header, error check and audio data. The no_of_ancillary_bits corresponds to the distance between the end of the audio data in an MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) frame, and the beginning of the header of the next audio frame.

5.3.2 DAB audio bit stream

5.3.2.0 Introduction

The DAB system uses the MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) format with additional specific information, necessary for the DAB application.

The field for this additional specific information is defined in DAB to contain the DAB fields Extended Programme Associated Data (X-PAD), Audio Scale Factor Error Check (ScF-CRC) and Fixed Programme Associated Data (F-PAD) (see clauses 5.3.2.7 to 5.3.2.9).

5.3.2.1 DAB audio sequence

A detailed graphic representation of the content and the structure of a DAB audio frame is given in figures 5 and 6.

DAB_audio_frame: part of the bit stream that is decodable by itself. Besides the information for 1 152 audio samples, it contains all specific DAB audio information (see also definition given in clause 5.3.1.1).

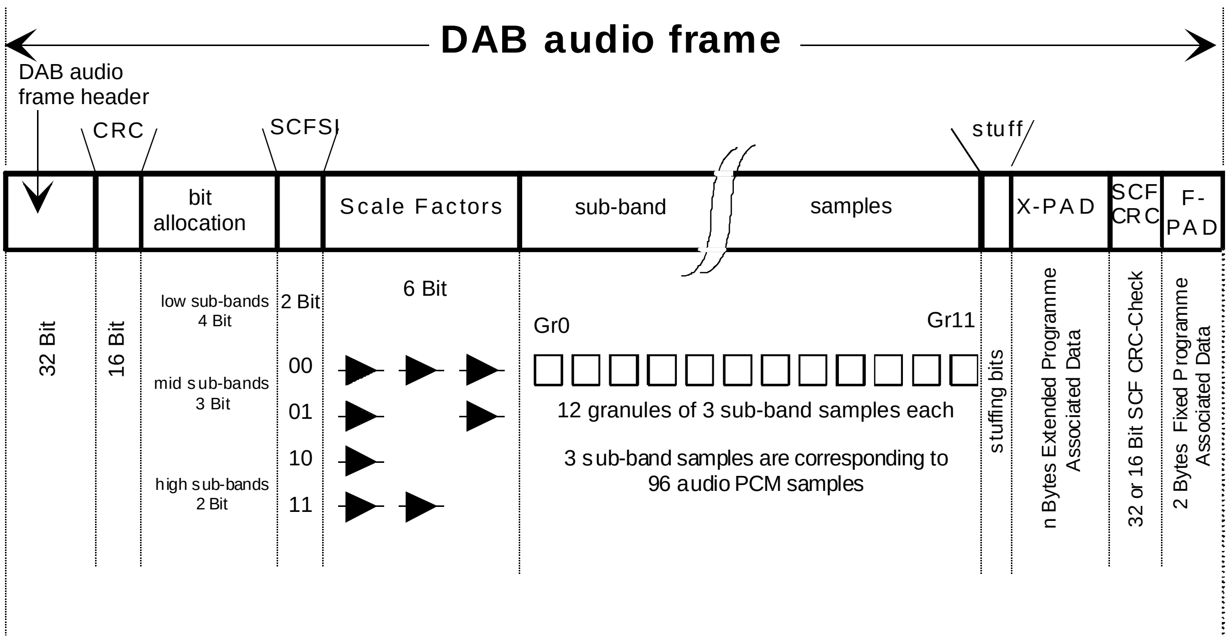

Figure 5: Structure of the DAB audio frame

5.3.2.2 DAB audio frame

DAB_audio_frame_header: part of the bit stream (the first 32 bits) containing relevant state information for the DAB audio decoder.

error_check: see definition given in clause 5.3.1.2.

audio_data: see definition given in clause 5.3.1.2.

audio_stuffing_bits: number of stuffing bits inserted between the end of audio_data and the beginning of x_prog_ass_data.

x_prog_ass_data: part of the DAB audio frame with variable length in multiples of bytes, that may be used for Programme Associated Data.

scf_error_check: part of the DAB audio frame containing information for error detection of ScFs.

f_prog_ass_data: part of the frame with constant length of two bytes, that may be used for Programme Associated Data.

5.3.2.3 DAB audio frame header

The DAB audio frame header is identical to the MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) header.

See definitions given in clause 5.3.1.3 for the semantic meaning of the following parameters:

- bitrate_index;
- sampling_frequency;
- padding_bit;
- private_bit;
- mode;
- mode_extension;
- copyright;
- original/copy;
- emphasis.

5.3.2.4 Error check

See definitions given in clause 5.3.1.4.

5.3.2.5 Audio data

See definitions given in clause 5.3.1.5.

5.3.2.6 Audio stuffing bits

The total number of bits available for audio_data per DAB audio frame for a sampling frequency of 48 kHz equals (bit_rate $\times$ 0,024) minus bits used by DAB_audio_frame_header(), error_check(), x_prog_ass_data(), scf_error_check(), and f_prog_ass_data(), and in the case of 24 kHz sampling frequency, this number equals (bit_rate $\times$ 0,048) minus bits used by DAB_audio_frame_header(), error_check(), x_prog_ass_data(), scf_error_check(), and f_prog_ass_data(). The number of bits actually used by audio_data may be less. In this case a number of stuffing bits are inserted between the end of audio_data and the beginning of x_prog_ass_data().

stuff_bit: single bit without useful information. This bit is not defined in DAB. Stuffing bits fill the space from the start of the MPEG Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) ancillary data field up to the beginning of the X-PAD field.

5.3.2.7 Extended Programme Associated Data (X-PAD)

x-pad_byte[i]: i^{th} byte of X-PAD. A variable number of bytes, no_of_x-pad_bytes is used for X-PAD, i.e. the length of this field is variable in multiples of bytes. The actual value is given in ETSI EN 300 401 [1], clause 7.4.

5.3.2.8 Scale Factor Error Check (ScF-CRC)

scf-crc_check[i]: i^{th} word used for Cyclic Redundancy Check (CRC) words, each protecting the Scale Factors of a group of sub-bands of the following DAB audio frame. The actual number of CRC words, no_of_scf_error_checks depends upon the bit-rate and audio mode (see clause B.3). In the case of ISO/IEC 11172-3 [2], either two or four 8-bit cyclic redundancy check words are used for error detection of two or four different spectral groups of Scale Factors within the encoded bit stream. In the case of ISO/IEC 13818-3 [4] always four 8-bit cyclic redundancy check words are used for error detection of four different spectral groups of Scale Factors within the encoded audio bit stream.

5.3.2.9 Fixed Programme Associated Data (F-PAD)

f-pad_byte[i]: i^{th} byte of F-PAD. A fixed number of two bytes for Fixed Programme Associated Data (F-PAD), comprising the last two bytes of the DAB audio frame, is used.

Frame structure of coded bit stream: valid for 1152 PCM audio input samples (stereo mode)

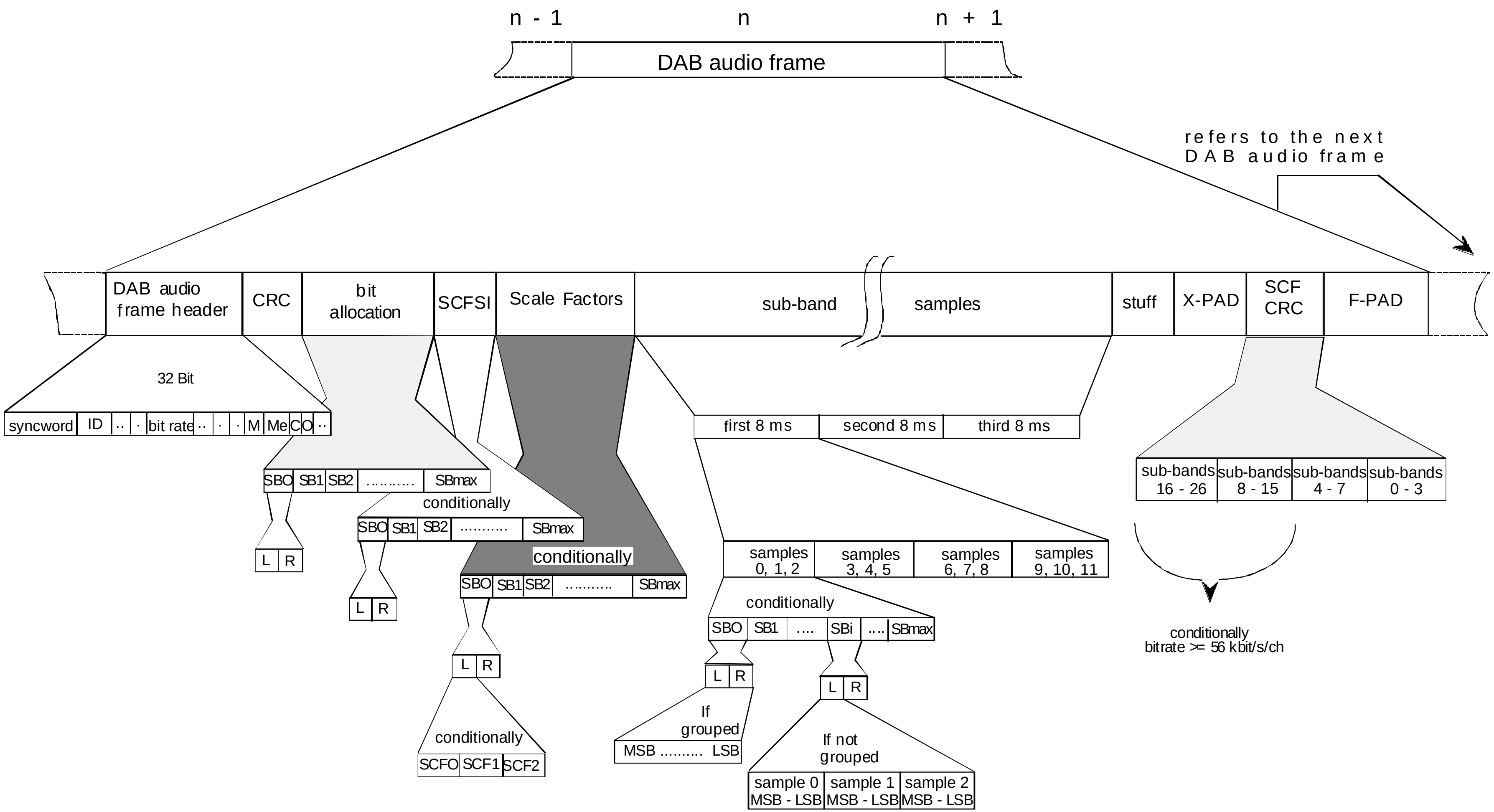

Figure 6: DAB audio frame structure

5.4 Audio bit stream syntax

5.4.0 Introduction

The details of the audio frame can be found in this clause, which describes the MPEG Audio Layer II bit stream syntax (clause 5.4.1), and the DAB audio bit stream syntax (clause 5.4.2). Apart from the audio stuffing bits, there is no difference in the ISO/IEC 11172-3 [2] Layer II bit stream syntax for 48 kHz sampling frequency and the ISO/IEC 13818-3 [4] Layer II bit stream syntax for 24 kHz sampling frequency. A detailed structure of the DAB audio frame is given in figure 6.

5.4.1 ISO/IEC 11172-3 and ISO/IEC 13818-3 Layer II bit stream syntax

5.4.1.0 General

This syntax is valid at the output of an MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) Layer II audio encoder and at the input of an MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) Layer II decoder.

5.4.1.1 Audio sequence

Syntax	No. of bits	Mnemonic
audio sequence()		
{		
while (nextbits()==syncword) {		
frame()		
}		
}		

5.4.1.2 Audio frame

Syntax	No. of bits	Mnemonic
frame()		
{		
header()		
error_check()		
audio_data()		
ancillary_data()		
}		

7.3.1.3 Header

Syntax	No. of bits	Mnemonic
header()		
{		
syncword	12	bslbf
ID	1	bslbf
layer	2	bslbf
protection_bit	1	bslbf
bitrate_index	4	bslbf
sampling_frequency	2	bslbf
padding_bit	1	bslbf
private_bit	1	bslbf
mode	2	bslbf
mode_extension	2	bslbf
copyright	1	bslbf
original/copy	1	bslbf
emphasis	2	bslbf
}		

5.4.1.4 Error check

Syntax	No. of bits	Mnemonic
error_check()		
{		
if (protection_bit==0)		
crc_check	16	rpchof
}		

5.4.1.5 Audio data

Syntax	No. of bits	Mnemonic
audio_data()		
{		
for (sb=0; sb<bound; sb++)		
for (ch=0; ch<nch; ch++)		
allocation[ch][sb]	2...4	uimsbf
for (sb=bound; sb<sblimit; sb++) {		
allocation[0][sb]	2...4	uimsbf
allocation[1][sb]=allocation|0]|sb]		
}		
for (sb=0; sb<sblimit; sb++)		
for (ch=0; ch<nch; ch++)		
if (allocation[ch][sb]!=0)		
scfsi[ch][sb]	2	bslbf
for (sb=0; sb<sblimit; sb++)		
for (ch=0; ch<nch; ch++)		
if (allocation[ch][sb]!=0) {		
if (scfsi[ch][sb]==0) {		
scalefactor[ch][sb][0]	6	uimsbf
scalefactor[ch][sb][1]	6	uimsbf
scalefactor[ch][sb][2]	6	uimsbf
}		
if ((scfsi[ch][sb]==1) || (scfsi[ch][sb]==3)){		
scalefactor[ch][sb][0]	6	uimsbf
scalefactor[ch][sb][2]	6	uimsbf
}		
if (scfsi[ch][sb]==2)		
scalefactor[ch][sb][0]	6	uimsbf
}		
for (gr=0; gr<12; gr++) {		
for (sb=0; sb<bound; sb++)		
for (ch=0; ch<nch; ch++)		
if (allocation[ch][sb]!=0) {		
if (grouping[ch][sb])		
samplecode[ch][sb][gr]	5...10	uimsbf
else		
for (s=0; s<3; s++)		
sample[ch][sb][3*gr+s]	3..16	uimsbf
}		
for (sb=bound; sb<sblimit; sb++)		
if (allocation[0][sb]!=0) {		
if (grouping[0][sb])		
samplecode[0][sb][gr]	5...10	uimsbf
else		
for (s=0; s<3; s++)		
sample[0][sb][3*gr+s] }	3...16	uimsbf
}		
}		
}		

5.4.1.6 Ancillary data

Syntax	No. of bits	Mnemonic
ancillary_data()		
{		
if ((layer == 1) || (layer == 2))		
for (b=0; b<no_of_ancillary_bits; b++)		
ancillary_bit	1	bslbf
}		

5.4.2 DAB audio bit stream syntax

5.4.2.0 General

This syntax is valid at the output of the DAB audio encoder and at the input of the DAB audio decoder.

5.4.2.1 DAB audio sequence

Syntax	No. of bits	Mnemonic
dab_audio sequence()		
{		
while true {		
dab_audio_frame()		
}		
}		

5.4.2.2 DAB audio frame

Syntax	No. of bits	Mnemonic
dab_audio_frame()		
{		
dab_audio_frame_header()		
error_check()		
audio_data()		
audio_stuffing_bits()		
x_prog_ass_data()		
scf_error_check()		
f_prog_ass_data()		
}		

5.4.2.3 DAB audio frame header

The DAB audio frame header is defined identically to the header defined in clause 5.4.1.3.

5.4.2.4 Error check

See clause 5.4.1.4.

5.4.2.5 Audio data

See clause 5.4.1.5.

5.4.2.6 Audio stuffing bits

In the case of 48 kHz sampling frequency, i.e. ISO/IEC 11172-3 [2] Layer II, the following syntax is valid.

Syntax	No. of bits	Mnemonic

audio_stuffing_bits()
{
    while (bitsum < (bit_rate * 0,024 - no_of_x-pad_bytes * 8
        - 2* 8 - no_of_scf_error_checks * 8)) {
            stuff_bit
            bitsum++
    }
}
	1	bslbf

In the case of 24 kHz sampling frequency, i.e. ISO/IEC 13818-3 [4] Layer II, the following syntax is valid.

Syntax	No. of bits	Mnemonic

audio_stuffing_bits()
{
    while (bitsum < (bit_rate * 0,048 - no_of_x-pad_bytes * 8
        - 2* 8 - no_of_scf_error_checks * 8)) {
            stuff_bit
            bitsum++
    }
}
	1	bslbf

5.4.2.7 Extended Programme Associated Data

Syntax	No. of bits	Mnemonic

x_prog_ass_data()
{
    for (i=0; i<no_of_x-pad_bytes; i++)
        x-pad_byte(i)
}
	8	bslbf

5.4.2.8 Scale factor error check

Syntax	No. of bits	Mnemonic

scf_error_check()
{
    for (i=no_of_scf_error_checks-1; i≥0; i--)
        scf-crc_check(i)
}
	8	rpchof

5.4.2.9 Fixed Programme Associated Data

Syntax	No. of bits	Mnemonic

f_prog_ass_data()
{
    for (i=0; i<2; i++)
        f-pad_byte(i)
}
	8	bslbf

5.5 Programme Associated Data (PAD)

5.5.1 Coding

The coding of the F-PAD and X-PAD are described in ETSI EN 300 401 [1], clause 7.4.

5.5.2 Transport

Each DAB audio frame contains a number of bytes which may carry Programme Associated Data (PAD). PAD is information which is synchronous to the audio and its contents may be intimately related to the audio. The PAD bytes in successive audio frames constitute the PAD channel. The functions provided by PAD are given in clause A.4.

The PAD bytes are always located at the end of each DAB audio frame. With a sampling frequency of 48 kHz, the whole DAB audio frame fits into the 24 ms frame structure of the CIF, and a new set of PAD bytes is available at the receiver every 24 ms. However in the case of a 24 kHz sampling frequency, the DAB LSF audio frame is divided into two parts of equal length (i.e. an even and odd partial frame) and spread across two CIFs. In this case, a new set of PAD bytes is available only every 48 ms.

In each DAB audio frame there are two bytes called the fixed PAD (F-PAD) field. Thus, the bit rate of the F-PAD field depends on the sampling frequency used for the audio coding. The bit rate for F-PAD is 0,667 kbit/s for 48 kHz sampling frequency. In the case of 24 kHz sampling frequency, this value is divided by a factor of two. The F-PAD field is intended to carry control information with a strong real-time character and data with a very low bit rate. The PAD channel may be extended using an Extended PAD (X-PAD) field to carry the dynamic label and data to User Applications. The length of the X-PAD field is chosen by the service provider.

The use of PAD is optional. If no information is sent in the F-PAD, all bytes in the F-PAD field shall be set to zero. This also implies that no X-PAD field is present.

The PAD carried in the DAB audio frame n shall be associated with the audio carried in the following frame, n+1.

Figure 7 shows the location of the F-PAD and X-PAD fields within the DAB audio frame.

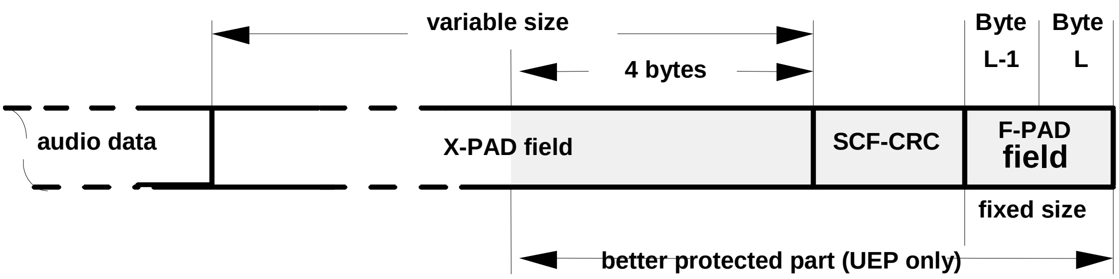

Figure 7: Location of the PAD bytes within the DAB audio frame

The two bytes of the F-PAD field (Byte L-1 and Byte L) are located at the end of the DAB audio frame, following the Scale Factor CRC (ScF-CRC). The X-PAD field is located just before the ScF-CRC. The audio data shall terminate before the beginning of the X-PAD field.

The F-PAD channel carries a two-bit field, "X-PAD Ind", which indicates one of three possibilities for the length of the X-PAD field:

1) **No X-PAD:** only the F-PAD field is available. All bits in the frame up to the ScF-CRC may be filled with audio data.

2) **Short X-PAD:** in this case the length of the X-PAD field is four bytes in every DAB audio frame, and the entire X-PAD field lies in the better protected part of the DAB audio frame when transmitted with UEP (i.e. is as well protected as the ScF-CRC). In total, 6 bytes carry PAD.

3) **Variable size X-PAD:** in this case the length of the X-PAD field may vary from frame to frame. The length of the X-PAD field in the current DAB audio frame can be deduced from the contents information carried within the X-PAD field. When transmitted with UEP, only a part (4 bytes) of the X-PAD field is as well protected as the ScF-CRC. The remainder has a lower protection. Application data carried in the X-PAD channel may require further error protection.

When transmitted with EEP, all bytes have the same level of protection.

5.5.3 Dynamic Range Control data

The DRC data can be used in the receiver to set the gain of a variable gain amplifier. The DRC data shall be coded as follows:

b_7 - b_2 DRC signal;

0 0 0 0 0 0: 0 dB;

0 0 0 0 0 1: +0,25 dB;

0 0 0 0 1 0: +0,50 dB;

↓ – – – – –

– – – – – –: continuous steps of +0,25 dB;

– – – – – –

1 1 1 1 1 1: +15,75 dB.

When DRC is used, the DRC data for each DAB audio frame shall be conveyed in the F-PAD of the preceding DAB audio frame.

Annex A (informative): Main characteristics of the audio coding system

A.1 Audio signal characteristics

The main characteristics of the input audio signal are:

- **audio bandwidth:** For f_s = 48 kHz the input audio signal can cover the whole audio frequency range up to 20,3 kHz, and for f_s = 24 kHz the input audio signal is low-pass filtered before downsampling to 24 kHz to cover a restricted audio frequency range up to about 11,3 kHz;
- **audio Interface:** the digital input signal may conform to the AES/EBU interface specification (see IEC 60958 [3]);
- **input resolution:** the system can support any input resolution up to 22 bits/sample;
- **sampling frequency:** The sampling frequency of the input audio signal prior to encoding at the transmitting end and of a digital output audio signal after decoding at the DAB receiving end is 48 kHz. In the case of MPEG-2 Layer II LSF coding, downsampling from 48 kHz to 24 kHz is applied at the transmitting end, and upsampling from 24 kHz to 48 kHz is applied at the receiving end, in order to avoid any other sampling frequency than 48 kHz for the PCM audio input and output signal.

A.2 Audio coding characteristics

The main characteristics of the audio coding system are:

- **audio modes:** three audio modes are provided:
  - single channel mode (one monophonic audio programme);
  - stereo mode (left and right channels of a stereophonic audio programme);
  - joint stereo mode. In this mode, the encoder exploits redundancy and irrelevancy of stereo signals for further data reduction, using Intensity stereo coding.

NOTE: The dual channel mode is not used in DAB.

- **bit rate:** According to ISO/IEC 11172-3 [2] Layer II, the permitted bit rates of the encoded audio signal for the sampling frequency of 48 kHz in single channel mode are as follows:
  - 32 kbit/s, 48 kbit/s, 56 kbit/s, 64 kbit/s, 80 kbit/s, 96 kbit/s, 112 kbit/s, 128 kbit/s, 160 kbit/s and 192 kbit/s.

  The stereo and joint stereo modes use twice the bit rate of the single channel mode (see also tables 11 and 13 in clause 5.3.1.3).
- According to ISO/IEC 13818-3 [4] Layer II, the permitted bit rates of the encoded audio signal for the sampling frequency of 24 kHz, irrespective of the audio mode, are as follows:
  - 8 kbit/s, 16 kbit/s, 24 kbit/s, 32 kbit/s, 40 kbit/s, 48 kbit/s, 56 kbit/s, 64 kbit/s, 80 kbit/s, 96 kbit/s, 112 kbit/s, 128 kbit/s, 144 kbit/s and 160 kbit/s (see also table 11 in clause 5.3.1.3).
- **DAB audio frame length:** One DAB audio frame covers 1152 PCM audio samples. For 48 kHz sampling frequency the length of a DAB audio frame is 24 ms. For 24 kHz sampling frequency the length of a DAB audio frame is 48 ms. This is twice the length of a DAB audio frame according to MPEG-1 Audio Layer II (ISO/IEC 11172-3 [2]). The DAB LSF audio frame is divided into two parts of equal lengths, an even (subframe "0", containing the first part of the DAB LSF audio frame) and an odd subframe (subframe "1", containing the second part of the DAB LSF audio frame), both fitting perfectly into the frame structure of the DAB logical frames. The X-PAD and F-PAD are available only once per 48 ms period.

A.3 Audio associated data characteristics

Programme Associated Data

Each DAB audio frame contains a number of bytes specifically for carrying Programme Associated Data (PAD). At the end of the DAB audio frame, a capacity of at least two bytes, called Fixed Programme Associated Data (F-PAD), is provided, irrespective of the sampling frequency, bit rate or the audio mode (i.e. single channel, stereo and joint stereo modes will all have the same capacity of F-PAD). But the broadcaster may choose to extend this capacity, called Extended Programme Associated Data (X-PAD) in order to transmit more audio related data.

These PAD comprise mainly information which are intimately associated with the audio signal, and which would become useless if delayed in a queue with other data, or if removed from the channel-coded audio bit stream and sent in a separate data service component. By reserving the limited capacity available for the PAD for information satisfying these criteria, it is possible to make the most effective use of such a data channel, which is strongly linked with the encoded audio signal. Although some capacity of X-PAD can also be provided for programme service information, further capacity can be provided elsewhere in the DAB multiplex (or ensemble) to carry additional information, such as text, relating to the various programmes in the ensemble which may require this, or some similar, facility.

It has to be considered that in the LSF mode the bit rate of F-PAD is reduced by a factor of two compared to the full sampling frequency mode, defined in the present document. Therefore, for a sampling frequency of 48 kHz, a bit rate of 0,667 kbit/s is available for F-PAD, but for a sampling frequency of 24 kHz, only 0,333 kbit/s are available for F-PAD.

For a sampling frequency of 48 kHz F-PAD is transmitted every 24 ms, corresponding to the frame structure of the DAB Main Service Channel. However, for 24 kHz sampling frequency, with the DAB LSF audio frame subdivided into two subframes of equal lengths, PAD is transmitted only once per 48 ms period. Any audio encoding device, which inserts PAD, needs to know whether the present sub-frame is the even or the odd one, whereby PAD is inserted only in the odd subframe.

Error protection of PAD when UEP error protection is used

The F-PAD and some parts of X-PAD are more strongly protected by the convolutional code of the transmission system than most of the other parts of the DAB audio bit stream (see ETSI EN 300 401 [1], clause 11). These fields are protected with different code-rates due to the Unequal Error Protection (UEP). Compared to the audio sub-band samples, a higher protection applies to F-PAD and four bytes of X-PAD. Depending on the requirement of the different types of PAD, this protection may be supplemented by additional protection schemes.

Annex B (normative): Audio decoding

B.1 General

The first action is synchronization of the decoder to the incoming audio bit stream, just after start-up. This may be done by using an external hardware synchronization signal, which is provided by the COFDM channel-demodulator every 24 ms, and thus enables the synchronization in the case of 48 kHz sampling frequency. However, in the case of 24 kHz sampling frequency, the DAB audio frame length is 48 ms, subdivided into two subframes of equal lengths, and the frame start is valid only every second time this external signal is provided. In this case, after start-up, the synchronization of the audio decoder is done by searching in the encoded audio bit stream for the MPEG Audio 12 bit syncword which is conveyed in each even subframe (subframe "0") and which can be compared with the external hardware synchronization signal, thereby allowing an extremely reliable synchronization.

In the DAB application, some parts of the ISO/IEC 11172-3 [2] header information, which are still kept in the DAB audio frame header, are already known to the decoder and need not to be decoded. These are layer, protection_bit, sampling_frequency, padding_bit, private bit and emphasis.

In addition to the ID bit, bitrate_index bits, copyright bit and original/copy bit, the decoder shall read the mode bits, and if these equal "01" also the mode_extension bits. The mode_extension bits set the "bound' as shown in clause 5.3 and thus indicate which sub-bands are coded in the Intensity stereo mode.

B.2 CRC check for audio side information

A CRC-check word for detecting errors within the significant side information of a DAB audio frame has been inserted in the bit stream just after the DAB audio frame header. The error detection method used is "CRC-16' whose generator polynomial is:

$$G_1(x) = x^{16} + x^{15} + x^2 + 1$$

The bits included into the CRC-check are:

- 16 bits of DAB_audio_frame_header( ), starting with bit_rate_index and ending with emphasis;
- a number of bits of audio_data( ), starting with the first bit. These bits include bit allocation information and ScFSI.

The method for the calculation of the CRC word in the decoder is described in annex E of ETSI EN 300 401 [1]. The initial state of the shift register is "1111 1111 1111 1111". If the final output of the shift register and the CRC-check word in the DAB audio frame are not identical, a transmission error has occurred in the protected field of the audio bit stream.

B.3 CRC check for Scale Factors

For detection of errors within the three MSbs of the Scale Factors, CRC-check words shall be inserted in the DAB audio bit stream just in front of the F-PAD field of the preceding DAB audio frame. For 48 kHz sampling frequency coding according to ISO/IEC 11172-3 [2] Layer II, either two or four CRC-check words shall be used, dependent on the bit rate. The CRC-check words are covering the Scale Factors of the following sub-bands:

If the bit rate per channel is greater than or equal to 56 kbit/s (i.e. bit rate ≥ 56 kbit/s for single channel mode, bit rate ≥ 112 kbit/s for all other modes):

- ScF-CRC0: Sub-bands 0 to 3 (sub-band group 0);
- ScF-CRC1: Sub-bands 4 to 7 (sub-band group 1);

- ScF-CRC2: Sub-bands 8 to 15 (sub-band group 2);
- ScF-CRC3: Sub-bands 16 to 26 (sub-band group 3).

If the bit rate per channel is less than 56 kbit/s (i.e. bit rate < 56 kbit/s for single channel mode, bit rate < 112 kbit/s for other audio modes):

- ScF-CRC0: Sub-bands 0 to 3 (sub-band group 0);
- ScF-CRC1: Sub-bands 4 to 7 (sub-band group 1).

To keep the position of the ScF-CRC-check words 1 and 2 independent of the bit rate, the ScF-CRC-check words are put in reverse order in the bit stream:

- ScF-CRC3, ScF- CRC2, ScF-CRC1, ScF-CRC0 bitrate ≥ 56 kbit/s/ch;
- ScF-CRC1, ScF-CRC0 bitrate < 56 kbit/s/ch.

For 24 kHz sampling frequency coding according to ISO/IEC 13818-3 [4] Layer II, always four CRC-check words shall be used. The CRC-check words are covering the Scale Factors of the following sub-bands:

- ScF-CRC0: Sub-bands 0 to 3 (sub-band group 0);
- ScF-CRC1: Sub-bands 4 to 7 (sub-band group 1);
- ScF-CRC2: Sub-bands 8 to 15 (sub-band group 2);
- ScF-CRC3: Sub-bands 16 to 29 (Sub-band group 3).

As in the case of 48 kHz sampling frequency coding, the ScF-CRC-check words are put in reverse order in the bit stream:

- ScF-CRC3, ScF- CRC2, ScF-CRC1, ScF-CRC0.

The error detection method used is "CRC-8", whose generator polynomial is:

$$G_2(X) = x^8 + x^4 + x^3 + x^2 + 1$$

For both, 48 kHz and 24 kHz sampling frequency coding, the bits included in the CRC-check are the 3 MSbs of all Scale Factors of the sub-band group, according to their order in the bit stream.

The method for the calculation of the ScF-CRC word is the same as for the CRC word in clause B.2, and is described in annex E of ETSI EN 300 401 [1]. The initial state of this shift register however is "0000 0000". If the output of the shift register and the transmitted ScF-CRC-check words are not identical, a transmission error has occurred in the three MSbs of one of the Scale Factors of this special sub-band group.

B.4 Decoding of the MPEG Audio Layer II bit stream

For 48 kHz and 24 kHz sampling frequency coding, the principles of the decoding process are given in ISO/IEC 11172-3 [2], chapter 2.4.3.3. In the case of 48 kHz sampling frequency, either tables 5 or 6 (bit allocation and possible quantization per sub-band) shall be used, dependent on the bit rate and audio mode, whereas in the case of 24 kHz sampling frequency, always table 6 (bit allocation and possible quantization per sub-band) shall be used.

Annex C (informative): Audio encoding

C.1 Analysis sub-band filter

The first step in the encoding process of a broadband PCM audio signal should be the filtering into 32 equally spaced sub-bands, each of which is down-sampled by a factor of $f_s/32$. The flow chart of this iterative process with the appropriate formulas is given in figure C.1. The analysis sub-band filtering includes the following steps:

- input 32 PCM audio samples;
- build an input sample vector *X* of 512 elements, so that the 512 most recent PCM audio samples are stored in the vector *X*. In each iteration 32 PCM audio samples are shifted in at positions 0 to 31, the most recent on at position 0, and the 32 oldest samples are shifted out. Position 0 of the vector *X* always contains the most recent sample, and position 511 the oldest one;
- vector *X* is windowed by vector *C*. The coefficients C_i are to be found in the table C.1;
- calculate the 64 intermediate values Y_i according to the formula given in the analysis filter flow chart;
- the 32 sub-band samples S_i are calculated by matrixing. The coefficients for the matrix *M* can be calculated by the following formula:

$$M_{ik} = \cos\left[(2i + 1)(k - 16)\pi/64\right] \qquad 0 \le i \le 31,\ 0 \le k \le 63$$

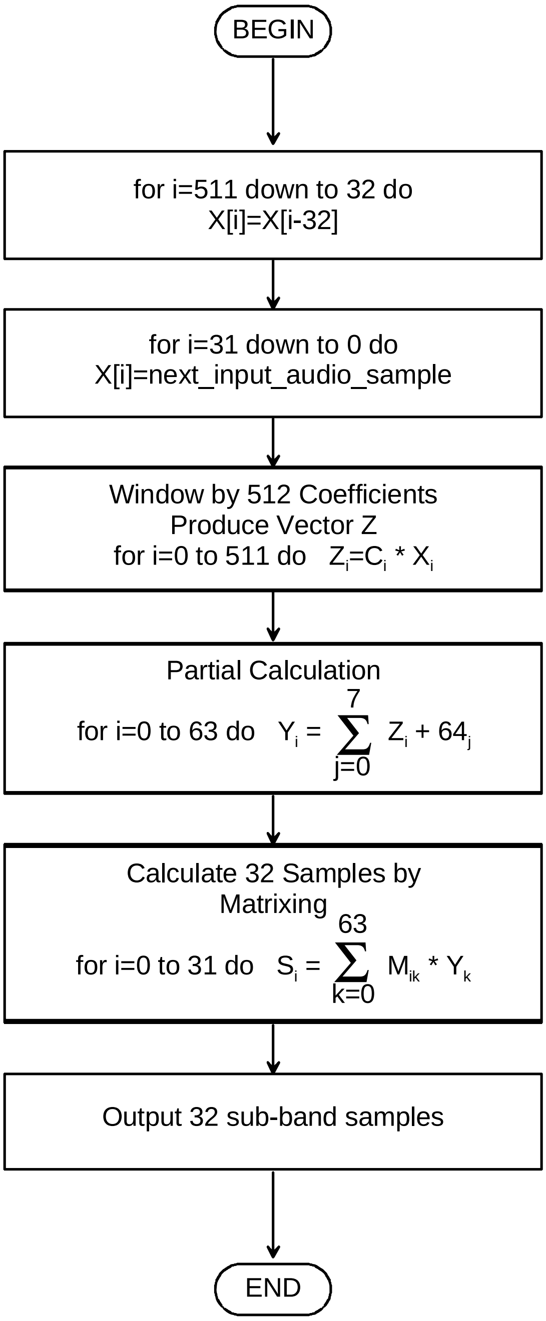

Figure C.1: Analysis sub-band filter flow chart

Table C.1: Coefficients C_i of the analysis window

C[0]= 0,000000000	C[1]=-0,000000477	C[2]=-0,000000477	C[3]=-0,000000477
C[4]=-0,000000477	C[5]=-0,000000477	C[6]=-0,000000477	C[7]=-0,000000954
C[8]=-0,000000954	C[9]=-0,000000954	C[10]=-0,000000954	C[11]=-0,000001431
C[12]=-0,000001431	C[13]=-0,000001907	C[14]=-0,000001907	C[15]=-0,000002384
C[16]=-0,000002384	C[17]=-0,000002861	C[18]=-0,000003338	C[19]=-0,000003338
C[20]=-0,000003815	C[21]=-0,000004292	C[22]=-0,000004768	C[23]=-0,000005245
C[24]=-0,000006199	C[25]=-0,000006676	C[26]=-0,000007629	C[27]=-0,000008106
C[28]=-0,000009060	C[29]=-0,000010014	C[30]=-0,000011444	C[31]=-0,000012398
C[32]=-0,000013828	C[33]=-0,000014782	C[34]=-0,000016689	C[35]=-0,000018120
C[36]=-0,000019550	C[37]=-0,000021458	C[38]=-0,000023365	C[39]=-0,000025272
C[40]=-0,000027657	C[41]=-0,000030041	C[42]=-0,000032425	C[43]=-0,000034809
C[44]=-0,000037670	C[45]=-0,000040531	C[46]=-0,000043392	C[47]=-0,000046253
C[48]=-0,000049591	C[49]=-0,000052929	C[50]=-0,000055790	C[51]=-0,000059605

C[52]=-0,000062943	C[53]=-0,000066280	C[54]=-0,000070095	C[55]=-0,000073433
C[56]=-0,000076771	C[57]=-0,000080585	C[58]=-0,000083923	C[59]=-0,000087261
C[60]=-0,000090599	C[61]=-0,000093460	C[62]=-0,000096321	C[63]=-0,000099182
C[64]= 0,000101566	C[65]= 0,000103951	C[66]= 0,000105858	C[67]= 0,000107288
C[68]= 0,000108242	C[69]= 0,000108719	C[70]= 0,000108719	C[71]= 0,000108242
C[72]= 0,000106812	C[73]= 0,000105381	C[74]= 0,000102520	C[75]= 0,000099182
C[76]= 0,000095367	C[77]= 0,000090122	C[78]= 0,000084400	C[79]= 0,000077724
C[80]= 0,000069618	C[81]= 0,000060558	C[82]= 0,000050545	C[83]= 0,000039577
C[84]= 0,000027180	C[85]= 0,000013828	C[86]=-0,000000954	C[87]=-0,000017166
C[88]=-0,000034332	C[89]=-0,000052929	C[90]=-0,000072956	C[91]=-0,000093937
C[92]=-0,000116348	C[93]=-0,000140190	C[94]=-0,000165462	C[95]=-0,000191212
C[96]=-0,000218868	C[97]=-0,000247478	C[98]=-0,000277042	C[99]=-0,000307560
C[100]=-0,000339031	C[101]=-0,000371456	C[102]=-0,000404358	C[103]=-0,000438213
C[104]=-0,000472546	C[105]=-0,000507355	C[106]=-0,000542164	C[107]=-0,000576973
C[108]=-0,000611782	C[109]=-0,000646591	C[110]=-0,000680923	C[111]=-0,000714302
C[112]=-0,000747204	C[113]=-0,000779152	C[114]=-0,000809669	C[115]=-0,000838757
C[116]=-0,000866413	C[117]=-0,000891685	C[118]=-0,000915051	C[119]=-0,000935555
C[120]=-0,000954151	C[121]=-0,000968933	C[122]=-0,000980854	C[123]=-0,000989437
C[124]=-0,000994205	C[125]=-0,000995159	C[126]=-0,000991821	C[127]=-0,000983715
C[128]= 0,000971317	C[129]= 0,000953674	C[130]= 0,000930786	C[131]= 0,000902653
C[132]= 0,000868797	C[133]= 0,000829220	C[134]= 0,000783920	C[135]= 0,000731945
C[136]= 0,000674248	C[137]= 0,000610352	C[138]= 0,000539303	C[139]= 0,000462532
C[140]= 0,000378609	C[141]= 0,000288486	C[142]= 0,000191689	C[143]= 0,000088215
C[144]=-0,000021458	C[145]=-0,000137329	C[146]=-0,000259876	C[147]=-0,000388145
C[148]=-0,000522137	C[149]=-0,000661850	C[150]=-0,000806808	C[151]=-0,000956535
C[152]=-0,001111031	C[153]=-0,001269817	C[154]=-0,001432419	C[155]=-0,001597881
C[156]=-0,001766682	C[157]=-0,001937389	C[158]=-0,002110004	C[159]=-0,002283096
C[160]=-0,002457142	C[161]=-0,002630711	C[162]=-0,002803326	C[163]=-0,002974033
C[164]=-0,003141880	C[165]=-0,003306866	C[166]=-0,003467083	C[167]=-0,003622532
C[168]=-0,003771782	C[169]=-0,003914356	C[170]=-0,004048824	C[171]=-0,004174709
C[172]=-0,004290581	C[173]=-0,004395962	C[174]=-0,004489899	C[175]=-0,004570484
C[176]=-0,004638195	C[177]=-0,004691124	C[178]=-0,004728317	C[179]=-0,004748821
C[180]=-0,004752159	C[181]=-0,004737377	C[182]=-0,004703045	C[183]=-0,004649162
C[184]=-0,004573822	C[185]=-0,004477024	C[186]=-0,004357815	C[187]=-0,004215240
C[188]=-0,004049301	C[189]=-0,003858566	C[190]=-0,003643036	C[191]=-0,003401756
C[192]= 0,003134727	C[193]= 0,002841473	C[194]= 0,002521515	C[195]= 0,002174854
C[196]= 0,001800537	C[197]= 0,001399517	C[198]= 0,000971317	C[199]= 0,000515938
C[200]= 0,000033379	C[201]=-0,000475883	C[202]=-0,001011848	C[203]=-0,001573563
C[204]=-0,002161503	C[205]=-0,002774239	C[206]=-0,003411293	C[207]=-0,004072189
C[208]=-0,004756451	C[209]=-0,005462170	C[210]=-0,006189346	C[211]=-0,006937027
C[212]=-0,007703304	C[213]=-0,008487225	C[214]=-0,009287834	C[215]=-0,010103703
C[216]=-0,010933399	C[217]=-0,011775017	C[218]=-0,012627602	C[219]=-0,013489246
C[220]=-0,014358521	C[221]=-0,015233517	C[222]=-0,016112804	C[223]=-0,016994476
C[224]=-0,017876148	C[225]=-0,018756866	C[226]=-0,019634247	C[227]=-0,020506859
C[228]=-0,021372318	C[229]=-0,022228718	C[230]=-0,023074150	C[231]=-0,023907185
C[232]=-0,024725437	C[233]=-0,025527000	C[234]=-0,026310921	C[235]=-0,027073860
C[236]=-0,027815342	C[237]=-0,028532982	C[238]=-0,029224873	C[239]=-0,029890060
C[240]=-0,030526638	C[241]=-0,031132698	C[242]=-0,031706810	C[243]=-0,032248020
C[244]=-0,032754898	C[245]=-0,033225536	C[246]=-0,033659935	C[247]=-0,034055710
C[248]=-0,034412861	C[249]=-0,034730434	C[250]=-0,035007000	C[251]=-0,035242081
C[252]=-0,035435200	C[253]=-0,035586357	C[254]=-0,035694122	C[255]=-0,035758972
C[256]= 0,035780907	C[257]= 0,035758972	C[258]= 0,035694122	C[259]= 0,035586357
C[260]= 0,035435200	C[261]= 0,035242081	C[262]= 0,035007000	C[263]= 0,034730434
C[264]= 0,034412861	C[265]= 0,034055710	C[266]= 0,033659935	C[267]= 0,033225536
C[268]= 0,032754898	C[269]= 0,032248020	C[270]= 0,031706810	C[271]= 0,031132698
C[272]= 0,030526638	C[273]= 0,029890060	C[274]= 0,029224873	C[275]= 0,028532982
C[276]= 0,027815342	C[277]= 0,027073860	C[278]= 0,026310921	C[279]= 0,025527000
C[280]= 0,024725437	C[281]= 0,023907185	C[282]= 0,023074150	C[283]= 0,022228718
C[284]= 0,021372318	C[285]= 0,020506859	C[286]= 0,019634247	C[287]= 0,018756866
C[288]= 0,017876148	C[289]= 0,016994476	C[290]= 0,016112804	C[291]= 0,015233517
C[292]= 0,014358521	C[293]= 0,013489246	C[294]= 0,012627602	C[295]= 0,011775017
C[296]= 0,010933399	C[297]= 0,010103703	C[298]= 0,009287834	C[299]= 0,008487225
C[300]= 0,007703304	C[301]= 0,006937027	C[302]= 0,006189346	C[303]= 0,005462170
C[304]= 0,004756451	C[305]= 0,004072189	C[306]= 0,003411293	C[307]= 0,002774239

C[308]= 0,002161503	C[309]= 0,001573563	C[310]= 0,001011848	C[311]= 0,000475883
C[312]=-0,000033379	C[313]=-0,000515938	C[314]=-0,000971317	C[315]=-0,001399517
C[316]=-0,001800537	C[317]=-0,002174854	C[318]=-0,002521515	C[319]=-0,002841473
C[320]= 0,003134727	C[321]= 0,003401756	C[322]= 0,003643036	C[323]= 0,003858566
C[324]= 0,004049301	C[325]= 0,004215240	C[326]= 0,004357815	C[327]= 0,004477024
C[328]= 0,004573822	C[329]= 0,004649162	C[330]= 0,004703045	C[331]= 0,004737377
C[332]= 0,004752159	C[333]= 0,004748821	C[334]= 0,004728317	C[335]= 0,004691124
C[336]= 0,004638195	C[337]= 0,004570484	C[338]= 0,004489899	C[339]= 0,004395962
C[340]= 0,004290581	C[341]= 0,004174709	C[342]= 0,004048824	C[343]= 0,003914356
C[344]= 0,003771782	C[345]= 0,003622532	C[346]= 0,003467083	C[347]= 0,003306866
C[348]= 0,003141880	C[349]= 0,002974033	C[350]= 0,002803326	C[351]= 0,002630711
C[352]= 0,002457142	C[353]= 0,002283096	C[354]= 0,002110004	C[355]= 0,001937389
C[356]= 0,001766682	C[357]= 0,001597881	C[358]= 0,001432419	C[359]= 0,001269817
C[360]= 0,001111031	C[361]= 0,000956535	C[362]= 0,000806808	C[363]= 0,000661850
C[364]= 0,000522137	C[365]= 0,000388145	C[366]= 0,000259876	C[367]= 0,000137329
C[368]= 0,000021458	C[369]=-0,000088215	C[370]=-0,000191689	C[371]=-0,000288486
C[372]=-0,000378609	C[373]=-0,000462532	C[374]=-0,000539303	C[375]=-0,000610352
C[376]=-0,000674248	C[377]=-0,000731945	C[378]=-0,000783920	C[379]=-0,000829220
C[380]=-0,000868797	C[381]=-0,000902653	C[382]=-0,000930786	C[383]=-0,000953674
C[384]= 0,000971317	C[385]= 0,000983715	C[386]= 0,000991821	C[387]= 0,000995159
C[388]= 0,000994205	C[389]= 0,000989437	C[390]= 0,000980854	C[391]= 0,000968933
C[392]= 0,000954151	C[393]= 0,000935555	C[394]= 0,000915051	C[395]= 0,000891685
C[396]= 0,000866413	C[397]= 0,000838757	C[398]= 0,000809669	C[399]= 0,000779152
C[400]= 0,000747204	C[401]= 0,000714302	C[402]= 0,000680923	C[403]= 0,000646591
C[404]= 0,000611782	C[405]= 0,000576973	C[406]= 0,000542164	C[407]= 0,000507355
C[408]= 0,000472546	C[409]= 0,000438213	C[410]= 0,000404358	C[411]= 0,000371456
C[412]= 0,000339031	C[413]= 0,000307560	C[414]= 0,000277042	C[415]= 0,000247478
C[416]= 0,000218868	C[417]= 0,000191212	C[418]= 0,000165462	C[419]= 0,000140190
C[420]= 0,000116348	C[421]= 0,000093937	C[422]= 0,000072956	C[423]= 0,000052929
C[424]= 0,000034332	C[425]= 0,000017166	C[426]= 0,000000954	C[427]=-0,000013828
C[428]=-0,000027180	C[429]=-0,000039577	C[430]=-0,000050545	C[431]=-0,000060558
C[432]=-0,000069618	C[433]=-0,000077724	C[434]=-0,000084400	C[435]=-0,000090122
C[436]=-0,000095367	C[437]=-0,000099182	C[438]=-0,000102520	C[439]=-0,000105381
C[440]=-0,000106812	C[441]=-0,000108242	C[442]=-0,000108719	C[443]=-0,000108719
C[444]=-0,000108242	C[445]=-0,000107288	C[446]=-0,000105858	C[447]=-0,000103951
C[448]= 0,000101566	C[449]= 0,000099182	C[450]= 0,000096321	C[451]= 0,000093460
C[452]= 0,000090599	C[453]= 0,000087261	C[454]= 0,000083923	C[455]= 0,000080585
C[456]= 0,000076771	C[457]= 0,000073433	C[458]= 0,000070095	C[459]= 0,000066280
C[460]= 0,000062943	C[461]= 0,000059605	C[462]= 0,000055790	C[463]= 0,000052929
C[464]= 0,000049591	C[465]= 0,000046253	C[466]= 0,000043392	C[467]= 0,000040531
C[468]= 0,000037670	C[469]= 0,000034809	C[470]= 0,000032425	C[471]= 0,000030041
C[472]= 0,000027657	C[473]= 0,000025272	C[474]= 0,000023365	C[475]= 0,000021458
C[476]= 0,000019550	C[477]= 0,000018120	C[478]= 0,000016689	C[479]= 0,000014782
C[480]= 0,000013828	C[481]= 0,000012398	C[482]= 0,000011444	C[483]= 0,000010014
C[484]= 0,000009060	C[485]= 0,000008106	C[486]= 0,000007629	C[487]= 0,000006676
C[488]= 0,000006199	C[489]= 0,000005245	C[490]= 0,000004768	C[491]= 0,000004292
C[492]= 0,000003815	C[493]= 0,000003338	C[494]= 0,000003338	C[495]= 0,000002861
C[496]= 0,000002384	C[497]= 0,000002384	C[498]= 0,000001907	C[499]= 0,000001907
C[500]= 0,000001431	C[501]= 0,000001431	C[502]= 0,000000954	C[503]= 0,000000954
C[504]= 0,000000954	C[505]= 0,000000954	C[506]= 0,000000477	C[507]= 0,000000477
C[508]= 0,000000477	C[509]= 0,000000477	C[510]= 0,000000477	C[511]= 0,000000477

C.2 Psychoacoustic model

For each frame, corresponding to 1 152 input samples, with a duration of 24 ms at a sampling frequency (f_s) of 48 kHz or 48 ms at a sampling frequency (f_s) of 24 kHz, a bit allocation needs to be determined. The bit allocation of the 32 sub-bands should be calculated on the basis of the signal-to-mask ratios of all the sub-bands. Therefore, for each sub-band the maximum signal level and the minimum masking threshold in dB should be determined. The minimum masking threshold is derived from an Fast Fourier Transform (FFT) of the input PCM signal, followed by a psychoacoustic model calculation.

The FFT in parallel with the sub-band filter compensates for the lack of spectral selectivity obtained at low frequencies by the sub-band filter bank. This technique provides both a sufficient time resolution for the coded audio signal (Polyphase filter with optimized window for minimal pre-echoes) and a sufficient spectral resolution for the calculation of the masking thresholds.

The frequencies and levels of aliasing distortions can be calculated. This is necessary for calculating a minimum bit rate for those sub-bands which need some bits to cancel the aliasing components in the decoder. The additional complexity to calculate the better frequency resolution is necessary only in the encoder, and introduces no additional delay or complexity in the decoder.

The calculation of the signal-to-mask-ratio (SMR) is based on the following steps:

Step 1: calculation of the FFT for time to frequency conversion;

Step 2: determination of the sound pressure level in dB in each sub-band;

Step 3: determination of the threshold in quiet (absolute threshold);

Step 4: finding of the tonal (more sinusoid-like) and non-tonal (more noise-like) components of the audio signal;

Step 5: decimation of the maskers, to obtain only the relevant maskers;

Step 6: calculation of the individual masking thresholds;

Step 7: determination of the global masking threshold;

Step 8: determination of the minimum masking threshold in each sub-band;

Step 9: calculation of the signal-to-mask ratio in each sub-band.

The following gives further details on the above steps.

Step 1: FFT Analysis.

The masking threshold is derived from an estimate of the power density spectrum that is calculated by a 1 024-point FFT. The FFT is calculated directly from the input PCM signal, windowed by a Hann window.

For a coincidence in time between the bit allocation and the corresponding sub-band samples, the PCM-samples entering the FFT should be delayed:

1) the delay of the analysis sub-band filter is 256 samples, corresponding to 5,3 ms at 48 kHz sampling frequency (i.e. f_s = 48 kHz), or 10,67 ms at 24 kHz sampling frequency (i.e. f_s = 24 kHz). A window shift of 256 samples will compensate for the delay in the analysis sub-band filter;

2) the Hann window should coincide with the sub-band samples of the frame. An additional window shift of minus 64 samples will compensate for the delay in the Hann window.

Technical data of the FFT:

- transform length *N*: 1 024 samples;
- Window size: 21,3 ms for f_s = 48 kHz, or 42,67 ms for f_s = 24 kHz;
- Frequency resolution: 46,875 Hz for f_s = 48 kHz, or 23,438 Hz for f_s = 24 kHz;
- Hann window, *h(i):* $h(i) = \sqrt{8/3} \times 0,5 \times \{1 - \cos[2 \times \pi \times i/N]\} \quad 0 \leq i \leq N-1$;
- power density spectrum *X(k)*:

$$X(k) = 10 \times \log_{10} \left| \frac{1}{N} \sum_{l=0}^{N-1} h(l) \times s(l) \times e^{(-j \times k \times l \times 2 \times \pi/N)} \right|^2 dB \quad 0 \leq k \leq N/2$$

where *s(l)* is the input signal.

A normalization to the reference level of 96 dB SPL (Sound Pressure Level) should be done in such a way that the maximum value corresponds to 96 dB.

Step 2: Determination of the Sound Pressure Level.

The SPL L_{sb} in sub-band n should be computed by:

$$L_{sb}(n) = \mathrm{MAX}[\ X(k),\ 20 \times \log_{10}(scf_{max}(n) \times 32\,768) - 10\ ]\ \mathrm{dB}$$

$$X(k) \text{ in sub-band } n$$

where *X(k)* is the Sound Pressure Level of the spectral line with index *k* of the FFT with the maximum amplitude in the frequency range corresponding to sub-band *n*. The expression $scf_{max}(n)$ is the maximum of the three Scale Factors of sub-band *n* within a frame. The *"-10 dB"* term corrects for the difference between peak and rms level. The Sound Pressure Level $L_{sb}(n)$ is computed for every sub-band *n*.

The following alternative method of calculating $L_{sb}(n)$ offers a potential for better encoder performance, but this technique has not been subjected to a formal audio quality test.

The alternative SPL $L_{sb}(n)$ in sub-band *n* should be computed by:

$$L_{sb}(n) = MAX[\ X_{spl}(n),\ 20 \times log_{10}(scf_{max}(n) \times 32\,768) - 10\ ]\ dB;$$

with:

$$X_{spl}(n) = 10\log\Big(\sum_{k(n)}^{k(n+1)} 10^{X(k)/10}\Big)\ \mathrm{dB},$$

where $k(n) = n \times N/64$

and $X_{spl}(n)$ is the alternative Sound Pressure Level corresponding to sub-band n.

Step 3: Considering the threshold in quiet.

The threshold in quiet $LT_q(k)$, also called absolute threshold, is available in table C.2 for the sampling frequency of 48 kHz and in table C.3 for the sampling frequency of 24 kHz. Values are available for each sample in the frequency domain where the masking threshold is calculated.

An offset depending on the overall bit rate should be used for the absolute threshold. This offset is -12 dB for bit rates ≥ 96 kbit/s and 0 dB for bit rates < 96 kbit/s per channel.

Step 4: Finding of tonal and non-tonal components.

The tonality of a masking component has an influence on the masking threshold. For this reason, it is worthwhile to discriminate between tonal and non-tonal components. For calculating the global masking threshold the tonal and the non-tonal components are derived from the FFT spectrum.

This step should start with the determination of local maxima, continued by extracting tonal components (sinusoids) and calculating the intensity of the non-tonal components within a bandwidth of a Critical band. The boundaries of the Critical bands are given in table C.4 for 48 kHz sampling frequency and in table C.5 for 24 kHz sampling frequency.

The bandwidth of the Critical bands varies with the centre frequency with a bandwidth of about only 0,1 kHz at low frequencies and with a bandwidth of about 4 kHz at high frequencies. It is known from psychoacoustic experiments that the ear has a better frequency resolution in the lower than in the higher frequency region. To determine if a local maximum may be a tonal component a frequency range *df* around the local maximum should be examined. For f_s = 48 kHz, the frequency range *df* is given by:

$$df = 93{,}75\ Hz \qquad 0{,}0\ kHz < f \leq 3{,}0\ kHz;$$

$$df = 140{,}63\ Hz \qquad 3{,}0\ kHz < f \leq 6{,}0\ kHz;$$

$$df = 281{,}25\ Hz \qquad 6{,}0\ kHz < f \leq 12{,}0\ kHz;$$

$df = 562{,}50$ Hz $\quad 12{,}0$ kHz $< f \leq 24{,}0$ kHz.

For f_s = 24 kHz sampling frequency, the frequency range *df* is given by:

$df = 93{,}75$ Hz $\quad 0{,}0$ kHz $< f \leq 3{,}0$ kHz;

$df = 140{,}63$ Hz $\quad 3{,}0$ kHz $< f \leq 6{,}0$ kHz;

$df = 281{,}25$ Hz $\quad 6{,}0$ kHz $< f \leq 11{,}25$ kHz.

To make lists of the spectral lines *X(k)* that are tonal or non-tonal, the following three operations are performed:

a) Labelling of local maxima:

- a spectral line *X(k)* is labelled as a local maximum if:

$$X(k) > X(k-1) \quad \text{and} \quad X(k) \geq X(k+1).$$

b) Listing of tonal components and calculation of the Sound Pressure Level:

- a local maximum is put in the list of tonal components if:

$$X(k) - X(k+j) \geq 7 \text{ dB},$$

where *j* is chosen according to:

$j = -2, +2$	for $2 < k < 63$	and $f_S = 48$ kHz;
$j = -3, -2, +2, +3$	for $63 \leq k < 127$	and $f_S = 48$ kHz;
$j = -4, +4$	for $4 < k < 127$	and $f_S = 24$ kHz;
$j = -6, ...,-2, +2, ...,+6$	for $127 \leq k < 255$	and $f_S = 48$ kHz or 24 kHz;
$j = -12, ...,-2, +2, ...,+12$	for $255 \leq k \leq 500$	and $f_S = 48$ kHz or 24 kHz.

If *X(k)* is found to be a tonal component, then the following parameters are listed:

- index number *k* of the spectral line;
- SPL $X_{tm}(k)=10 \times \log_{10} (10^{X(k-1)/10} + 10^{X(k)/10} + 10^{X(k+1)/10})$ dB;
- tonal flag.

Next, all spectral lines within the examined frequency range are set to -∞ dB.

c) Listing of non-tonal components and calculation of the power:

The non-tonal (noise) components are calculated from the remaining spectral lines. To calculate the non-tonal components from these spectral lines *X(k)*, the Critical bands *z(k)* are determined using the table C.4 in the case of 48 kHz sampling frequency coding and table C.5 in the case of 24 kHz sampling frequency coding. For 48 kHz sampling frequency 27 Critical bands are considered (see table C.4). For 24 kHz sampling frequency, 23 Critical bands are considered (see table C.5). Within each Critical band, the power of the spectral lines (remained after the tonal components have been zeroed) are summed to form the Sound Pressure Level of the new non-tonal component $X_{nm}(k)$ corresponding to that Critical band.

The following parameters are listed:

- index number *k* of the spectral line nearest to the geometric mean of the Critical band;
- SPL $X_{nm}(k)$ in dB;
- non-tonal flag.

Step 5: Decimation of tonal and non-tonal masking components.

Decimation is a procedure that is used to reduce the number of maskers which are considered for the calculation of the global masking threshold:

a) Tonal $X_{tm}(k)$ or non-tonal components $X_{nm}(k)$ are considered for the calculation of the masking threshold only if:

$$X_{tm}(k) \geq LT_q(k) \quad or \quad X_{nm}(k) \geq LT_q(k).$$

In this expression, $LT_q(k)$ is the absolute threshold (or threshold in quiet) at the frequency of index k. These values are given in table C.2 for 48 kHz sampling frequency coding and in table C.3 for 24 kHz sampling frequency coding.

b) Decimation of two or more tonal components within a distance of less than 0,5 Bark. The component with the highest power should be kept, and the smaller component(s) should be removed from the list of tonal components. For this operation, a sliding window in the Critical band domain should be used with a width of 0,5 Bark.

In the following, the index j is used to indicate the relevant tonal or non-tonal masking components from the combined decimated list.

Step 6: Calculation of individual masking thresholds.

Of the original 512 frequency domain samples, indexed by k, only a subset of the samples, indexed by i, are considered for the global masking threshold calculation. The samples used are shown in tables C.2 and C.3.

For the frequency lines corresponding to the frequency region which is covered by the first three sub-bands no sub-sampling is used. For the frequency region which is covered by next three sub-bands every second spectral line is considered. For the frequency region corresponding to the next six sub-bands every fourth spectral line is considered. With a sampling frequency of 48 kHz, in the remaining sub-bands every eighth spectral line is considered up to 20 kHz (see also table C.2). With a sampling frequency of 24 kHz, every eighth spectral line is considered for the next 18 sub-bands (see also table C.3). The number of samples, i, in the subsampled frequency domain is 126 with a sampling frequency of 48 kHz, and 132 with a sampling frequency of 24 kHz.

Every tonal and non-tonal component is assigned the value of the index i which most closely corresponds to the frequency of the original spectral line $X(k)$. This index i is given in tables C.2 and C.3.

The individual masking thresholds of both tonal and non-tonal components are given by the following expression:

$$LT_{tm}[z(j),z(i)] = X_{tm}[z(j)] + av_{tm}[z(j)] + vf[z(j),z(i)] \text{ dB};$$

$$LT_{nm}[z(j),z(i)] = X_{nm}[z(j)] + av_{nm}[z(j)] + vf[z(j),z(i)] \text{ dB}.$$

In this formula LT_{tm} and LT_{nm} are the individual masking thresholds at Critical band rate z in Bark of the masking component at the Critical band rate z_m in Bark. The values in dB can be either positive or negative. The term $X_{tm}[z(j)]$ is the Sound Pressure Level of the masking component with the index number j at the corresponding Critical band rate $z(j)$. The term av is called the masking index and vf the masking function of the masking component $X_{tm}[z(j)]$. The masking index av is different for tonal and non-tonal masker (av_{tm} and av_{nm}).

For tonal maskers it is given by:

$$av_{tm} = -1{,}525 - 0{,}275 \times z(j) - 4{,}5 \text{ dB},$$

and for non-tonal maskers:

$$av_{nm} = -1{,}525 - 0{,}175 \times z(j) - 0{,}5 \text{ dB}.$$

The masking function vf of a masker is characterized by different lower and upper slopes, which depend on the distance in Bark $dz = z(i) - z(j)$ to the masker. In this expression i is the index of the spectral line at which the masking function is calculated and j that of the masker. The Critical band rates $z(j)$ and $z(i)$ can be found in tables C.2 and C.3. The masking function, which is the same for tonal and non-tonal maskers, is given by:

$$vf = 17 \times (dz + 1) - (0{,}4 \times X[z(j)] + 6) \text{ dB} \qquad for\ -3 \leq dz < -1 \text{ Bark};$$

$$vf = (0{,}4 \times X[z(j)] + 6) \times dz \text{ dB} \qquad \text{for } -1 \leq dz < 0 \text{ Bark;}$$

$$vf = -17 \times dz \text{ dB} \qquad \text{for } 0 \leq dz < 1 \text{ Bark;}$$

$$vf = -(dz - 1) \times (17 - 0{,}15 \times X[z(j)]) - 17 \text{ dB} \qquad \text{for } 1 \leq dz < 8 \text{ Bark.}$$

In these expressions $X[z(j)]$ is the Sound Pressure Level of the j'th masking component in dB. For reasons of implementation complexity, the masking should no longer considered (LT_{tm} and LT_{nm} are set to -∞ dB outside this range) if $dz < -3$ *Bark*, or $dz \geq 8$ *Bark*.

Step 7: Calculation of the global masking threshold LT_g.

The global masking threshold $LT_g(i)$ at the i'th frequency sample is derived from the upper and lower slopes of the individual masking threshold of each of the *j* tonal and non-tonal maskers, and in addition from the threshold in quiet $LT_q(i)$, which is also given in tables C.2 and C.3. The global masking threshold is found by summing the powers corresponding to the individual masking thresholds and the threshold in quiet.

$$LT_g(i) = 10\log_{10}(10^{LT_q(i)/10} + \sum_{j=1}^{m} 10^{LT_{tm}[z(j),z(i)]/10} + \sum_{j=1}^{n} 10^{LT_{nm}[z(j),z(i)]/10} \quad dB$$

The total number of tonal maskers is given by *m*, and the total number of non-tonal maskers is given by *n*. For a given *i*, the range of *j* can be reduced to just encompass those masking components that are within -8 to +3 Bark from i. Outside of this range LT_{tm} and LT_{nm} are -∞ dB.

Step 8: Determination of the minimum masking threshold.

The minimum masking level $LT_{min}(n)$ in sub-band *n* is determined by the following expression:

$$LTmin(n) = MIN[\ LTg(i)\ ] \text{ dB}$$

$$f(i) \text{ in sub-band } n$$

where $f(i)$ is the frequency of the i'th frequency sample. The $f(i)$ are tabulated in tables C.2 and C.3.

A minimum masking level $LT_{min}(n)$ is computed for every sub-band.

Step 9: Calculation of the Signal-to-Mask-Ratio.

The Signal-to-Mask Ratio

$$SMR_{sb}(n) = L_{sb}(n) - LT_{min}(n) \text{ dB}$$

is computed for every sub-band *n*.

Table C.2: Frequencies, critical band rates and absolute threshold for a sampling frequency of 48 kHz

Index number i	Frequency Hz	Critical band rate z	Absolute threshold dB	Index number i	Frequency Hz	Critical band rate z	Absolute threshold dB
1	46,88	0,463	42,10	39	1 828,13	12,518	0,49
2	93,75	0,925	24,17	40	1 875,00	12,684	0,29
3	140,63	1,385	17,47	41	1 921,88	12,845	0,09
4	187,50	1,842	13,87	42	1 968,75	13,002	-0,11
5	234,38	2,295	11,60	43	2 015,63	13,154	-0,32
6	281,25	2,742	10,01	44	2 062,50	13,302	-0,54
7	328,13	3,184	8,84	45	2 109,38	13,446	-0,75
8	375,00	3,618	7,94	46	2 156,25	13,586	-0,97
9	421,88	4,045	7,22	47	2 203,13	13,723	-1,20
10	468,75	4,463	6,62	48	2 250,00	13,855	-1,43
11	515,63	4,872	6,12	49	2 343,75	14,111	-1,88
12	562,50	5,272	5,70	50	2 437,50	14,354	-2,34
13	609,38	5,661	5,33	51	2 531,25	14,585	-2,79
14	656,25	6,041	5,00	52	2 625,00	14,807	-3,22
15	703,13	6,411	4,71	53	2 718,75	15,018	-3,62
16	750,00	6,770	4,45	54	2 812,50	15,221	-3,98
17	796,88	7,119	4,21	55	2 906,25	15,415	-4,30
18	843,75	7,457	4,00	56	3 000,00	15,602	-4,57
19	890,63	7,785	3,79	57	3 093,75	15,783	-4,77
20	937,50	8,103	3,61	58	3 187,50	15,956	-4,91
21	984,38	8,410	3,43	59	3 281,25	16,124	-4,98
22	1 031,25	8,708	3,26	60	3 375,00	16,287	-4,97
23	1 078,13	8,996	3,09	61	3 468,75	16,445	-4,90
24	1 125,00	9,275	2,93	62	3 562,50	16,598	-4,76
25	1 171,88	9,544	2,78	63	3 656,25	16,746	-4,55
26	1 218,75	9,805	2,63	64	3 750,00	16,891	-4,29
27	1 265,63	10,057	2,47	65	3 843,75	17,032	-3,99
28	1 312,50	10,301	2,32	66	3 937,50	17,169	-3,64
29	1 359,38	10,537	2,17	67	4 031,25	17,303	-3,26
30	1 406,25	10,765	2,02	68	4 125,00	17,434	-2,86
31	1 453,13	10,986	1,86	69	4 218,75	17,563	-2,45
32	1 500,00	11,199	1,71	70	4 312,50	17,688	-2,04
33	1 546,88	11,406	1,55	71	4 406,25	17,811	-1,63
34	1 593,75	11,606	1,38	72	4 500,00	17,932	-1,24
35	1 640,63	11,800	1,21	73	4 687,50	18,166	-0,51
36	1 687,50	11,988	1,04	74	4 875,00	18,392	0,12
37	1 734,38	12,170	0,86	75	5 062,50	18,611	0,64
38	1 781,25	12,347	0,67	76	5 250,00	18,823	1,06
77	5 437,50	19,028	1,39	102	11 250,00	22,941	16,54
78	5 625,00	19,226	1,66	103	11 625,00	23,072	18,77
79	5 812,50	19,419	1,88	104	12 000,00	23,195	21,23
80	6 000,00	19,606	2,08	105	12 375,00	23,309	23,94
81	6 187,50	19,788	2,27	106	12 750,00	23,415	26,90
82	6 375,00	19,964	2,46	107	13 125,00	23,515	30,14
83	6 562,50	20,135	2,65	108	13 500,00	23,607	33,67
84	6 750,00	20,300	2,86	109	13 875,00	23,694	37,51
85	6 937,50	20,461	3,09	110	14 250,00	23,775	41,67
86	7 125,00	20,616	3,33	111	14 625,00	23,852	46,17
87	7 312,50	20,766	3,60	112	15 000,00	23,923	51,04
88	7 500,00	20,912	3,89	113	15 375,00	23,991	56,29
89	7 687,50	21,052	4,20	114	15 750,00	24,054	61,94
90	7 875,00	21,188	4,54	115	16 125,00	24,114	68,00
91	8 062,50	21,318	4,91	116	16 500,00	24,171	68,00
92	8 250,00	21,445	5,31	117	16 875,00	24,224	68,00
93	8 437,50	21,567	5,73	118	17 250,00	24,275	68,00
94	8 625,00	21,684	6,18	119	17 625,00	24,322	68,00
95	8 812,50	21,797	6,67	120	18 000,00	24,368	68,00
96	9 000,00	21,906	7,19	121	18 375,00	24,411	68,00
97	9 375,00	22,113	8,33	122	18 750,00	24,452	68,00
98	9 750,00	22,304	9,63	123	19 125,00	24,491	68,00
99	10 125,00	22,482	11,08	124	19 500,00	24,528	68,00
100	10 500,00	22,646	12,71	125	19 875,00	24,564	68,00
101	10 875,00	22,799	14,53	126	20 250,00	24,597	68,00

Table C.3: Frequencies, critical band rates and absolute threshold for a sampling frequency of 24 kHz

Index number i	Frequency Hz	Critical band rate z	Absolute threshold dB	Index number i	Frequency Hz	Critical band rate z	Absolute threshold dB
1	23,44	0,232	68,00	39	914,06	7,945	3,70
2	46,88	0,463	42,10	40	937,50	8,103	3,61
3	70,31	0,694	30,43	41	960,94	8,258	3,51
4	93,75	0,925	24,17	42	984,38	8,410	3,43
5	117,19	1,156	20,22	43	1 007,81	8,560	3,34
6	140,63	1,385	17,47	44	1 031,25	8,708	3,26
7	164,06	1,614	15,44	45	1 054,69	8,853	3,17
8	187,50	1,842	13,87	46	1 078,13	8,996	3,09
9	210,94	2,069	12,62	47	1 101,56	9,137	3,01
10	234,38	2,295	11,60	48	1 125,00	9,275	2,93
11	257,81	2,519	10,74	49	1 171,88	9,544	2,78
12	281,25	2,742	10,01	50	1 218,75	9,805	2,63
13	304,69	2,964	9,39	51	1 265,63	10,057	2,47
14	328,13	3,184	8,84	52	1 312,50	10,301	2,32
15	351,56	3,402	8,37	53	1 359,38	10,537	2,17
16	375,00	3,618	7,94	54	1 406,25	10,765	2,02
17	398,44	3,832	7,56	55	1 453,13	10,986	1,86
18	421,88	4,045	7,22	56	1 500,00	11,199	1,71
19	445,31	4,255	6,90	57	1 546,88	11,406	1,55
20	468,75	4,463	6,62	58	1 593,75	11,606	1,38
21	492,19	4,668	6,36	59	1 640,63	11,800	1,21
22	515,63	4,872	6,12	60	1 687,50	11,988	1,04
23	539,06	5,073	5,90	61	1 734,38	12,170	0,86
24	562,50	5,272	5,70	62	1 781,25	12,347	0,67
25	585,94	5,468	5,50	63	1 828,13	12,518	0,49
26	609,38	5,661	5,33	64	1 875,00	12,684	0,29
27	632,81	5,853	5,16	65	1 921,88	12,845	0,09
28	656,25	6,041	5,00	66	1 968,75	13,002	-0,11
29	679,69	6,227	4,85	67	2 015,63	13,154	-0,32
30	703,13	6,411	4,71	68	2 062,50	13,302	-0,54
31	726,56	6,592	4,58	69	2 109,38	13,446	-0,75
32	750,00	6,770	4,45	70	2 156,25	13,586	-0,97
33	773,44	6,946	4,33	71	2 203,13	13,723	-1,20
34	796,88	7,119	4,21	72	2 250,00	13,855	-1,43
35	820,31	7,289	4,10	73	2 343,75	14,111	-1,88
36	843,75	7,457	4,00	74	2 437,50	14,354	-2,34
37	867,19	7,622	3,89	75	2 531,25	14,585	-2,79
38	890,63	7,785	3,79	76	2 625,00	14,807	-3,22
77	2 718,75	15,018	-3,62	105	6 187,50	19,788	2,25
78	2 812,50	15,221	-3,98	106	6 375,00	19,964	2,43
79	2 906,25	15,415	-4,30	107	6 562,50	20,135	2,63
80	3 000,00	15,602	-4,57	108	6 750,00	20,300	2,83
81	3 093,75	15,783	-4,77	109	6 937,50	20,461	3,06
82	3 187,50	15,956	-4,91	110	7 125,00	20,616	3,30
83	3 281,25	16,124	-4,98	111	7 312,50	20,766	3,57
84	3 375,00	16,287	-4,98	112	7 500,00	20,912	3,85
85	3 468,75	16,445	-4,92	113	7 687,50	21,052	4,16
86	3 562,50	16,598	-4,80	114	7 875,00	21,188	4,50
87	3 656,25	16,746	-4,61	115	8 062,50	21,318	4,86
88	3 750,00	16,891	-4,36	116	8 250,00	21,445	5,25
89	3 843,75	17,032	-4,07	117	8 437,50	21,567	5,67
90	3 937,50	17,169	-3,73	118	8 625,00	21,684	6,12
91	4 031,25	17,303	-3,36	119	8 812,50	21,797	6,61
92	4 125,00	17,434	-2,96	120	9 000,00	21,906	7,12
93	4 218,75	17,563	-2,55	121	9 187,50	22,012	7,67
94	4 312,50	17,688	-2,14	122	9 375,00	22,113	8,26
95	4 406,25	17,811	-1,73	123	9 562,50	22,210	8,88
96	4 500,00	17,932	-1,33	124	9 750,00	22,304	9,54
97	4 687,50	18,166	-0,59	125	9 937,50	22,395	10,24
98	4 875,00	18,392	0,05	126	10 125,00	22,482	10,98
99	5 062,50	18,611	0,58	127	10 312,50	22,566	11,77
100	5 250,00	18,823	1,01	128	10 500,00	22,646	12,60
101	5 437,50	19,028	1,36	129	10 687,50	22,724	13,48

Index number i	Frequency Hz	Critical band rate z	Absolute threshold dB	Index number i	Frequency Hz	Critical band rate z	Absolute threshold dB
102	5 625,00	19,226	1,63	130	10 875,00	22,799	14,41
103	5 812,50	19,419	1,86	131	11 062,50	22,871	15,38
104	6 000,00	19,606	2,06	132	11 250,00	22,941	16,41

C.3 Bit allocation procedure

Before adjustment to a fixed bit rate, the number of bits, "*adb*", that are available for coding the sub-band samples and the Scale Factors should be determined. This number can be obtained by subtracting from the total number of available bits "*cb*", the number of bits for bit allocation "*bbal*", and the number of bits "*banc*" for ancillary data:

$$adb = cb - (bbal + banc)$$

The resulting number should be used to code the sub-band samples and the Scale Factors. The principle used in the allocation procedure is minimization of the total Noise-to-Mask Ratio over the DAB audio frame with the constraint that the number of bits used does not exceed the number of bits available for that frame. Use is made of tables 5 to 7, that indicate for every sub-band the number of steps that may be used to quantize the samples. The number of bits that represent these quantized samples can be derived from table 8.

Table C.4: Critical band boundaries for a sampling frequency of 48 kHz

No.	Index of table F&CB	Frequency Hz	Bark z
0	1	46,875	0,463
1	2	93,750	0,925
2	3	140,625	1,385
3	5	234,375	2,295
4	7	328,125	3,184
5	9	421,875	4,045
6	12	562,500	5,272
7	14	656,250	6,041
8	17	796,875	7,119
9	20	937,500	8,103
10	24	1 125,000	9,275
11	27	1 265,625	10,057
12	32	1 500,000	11,199
13	37	1 734,375	12,170
14	42	1 968,750	13,002
15	49	2 343,750	14,111
16	53	2 718,750	15,018
17	59	3 281,250	16,124
18	65	3 843,750	17,032
19	73	4 687,500	18,166
20	77	5 437,500	19,028
21	82	6 375,000	19,964
22	89	7 687,500	21,052
23	97	9 375,000	22,113
24	103	11 625,000	23,072
25	113	15 375,000	23,991
26	126	20 250,000	24,597
NOTE: The frequencies represent the top end of each critical band.			

Table C.5: Critical band boundaries for a sampling frequency of 24 kHz

No.	Index of table F&CB	Frequency Hz	Bark z
0	4	93,75	0,925
1	9	210,94	2,069
2	13	304,69	2,964
3	18	421,88	4,045
4	23	539,06	5,073
5	28	656,25	6,041
6	33	773,44	6,946
7	39	914,06	7,945
8	46	1 078,13	8,996
9	51	1 265,63	10,057
10	55	1 453,13	10,986
11	60	1 687,50	11,988
12	66	1 968,75	13,002
13	73	2 343,75	14,111
14	77	2 718,75	15,018
15	82	3 187,50	15,956
16	89	3 843,75	17,032
17	96	4 500,00	17,932
18	101	5 437,50	19,028
19	106	6 375,00	19,964
20	113	7 687,50	21,052
21	121	9 187,50	22,012
22	132	11 250,00	22,941
NOTE: The frequencies represent the top end of each critical band.			

The allocation procedure is an iterative procedure where, in each iteration step the number of levels of the sub-band that has the greatest benefit is increased.

First the Mask-to-Noise Ratio "*MNR*" for each sub-band should be calculated by subtracting from the Signal-to-Noise-Ratio "*SNR*" the Signal-to-Mask-Ratio "*SMR*":

$$MNR = SNR - SMR$$

The *SNR* can be found in the informative table C.6. The *SMR* is the output of the psychoacoustic model.

Table C.6: Signal-to Noise-Ratios

No. of steps	SNR dB
0	0,00
3	7,00
5	11,00
7	16,00
9	20,84
15	25,28
31	31,59
63	37,75
127	43,84
255	49,89
511	55,93
1 023	61,96
2 047	67,98
4 095	74,01
8 191	80,03
16 383	86,05
32 767	92,01
65 535	98,01

Then zero bits should be allocated to the sub-band samples and the Scale Factors. The number of bits for the sub-band samples *bspl* and the number of bits for the Scale Factors *bscf* are set to zero. Next an iterative procedure should be started. Each iteration loop should contain the following steps:

- determination of the minimal *MNR* of all sub-bands;
- the accuracy of the quantization of the sub-band with the minimal *MNR* should be increased by using the next higher entry in the relevant tables 5 to 7;
- the new *MNR* of this sub-band should be calculated;
- *bspl* should be updated according to the additional number of bits that are necessary. If a non-zero number of bits is assigned to a sub-band for the first time, *bsel* is updated, and *bscf* is updated according to the number of Scale Factors for this sub-band. Then *adb* should be calculated again using the formula:

$$adb = cb-(bbal + bsel + bscf + bspl + banc).$$

The iterative procedure should be repeated as long as *adb* is not less than any possible increase of *bspl*, *bsel* and *bscf* within one loop.

C.4 Bit sensitivity to errors

This part of the annex indicates the sensitivity of individual bits to random errors if application-specific error protection is needed. This sensitivity for each bit is given in table C.7 by a value from 0 to 5, indicating the amount of degradation resulting from one isolated error:

5 catastrophic;

4 very annoying;

3 annoying;

2 slightly annoying;

1 audible;

0 insensitive.

The values are not the results of precise measurements, rather they rely upon knowledge of the coding scheme. They assume that the error detection scheme is not in use. The DAB audio frame header and error check information defined in clauses 5.4.2.3 and 5.4.2.4 are considered to have the highest sensitivity.

Some fields in the DAB audio frame do not have a fixed length. All bits in this fields are rated for error sensitivity, even if not in use.

Table C.7: bit sensitivity of DAB audio frame bits

Parameters	Number of bits	Sensitivity
bit allocation	all bits	5
ScFSI	all bits	5
ScFs	5 (msb)	4
	4	4
	3	4
	2	3
	1	2
	0 (lsb)	1
sub-band samples (note)	8 -16 (msb)	3
	5 - 7	2
	3,4	1
	(lsb) 0 - 2	0
NOTE: According to the bit allocation.		

C.5 Error concealment

A feature of the coded bit stream is the CRC word which provides some error detection facility to the decoder, described in clause B.2. The Hamming distance of this error detection code is d = 4, which allows for the detection of up to 3 single bit errors or for the detection of one error burst of up to 16 bit length. The amount and the position of the protected bits within one encoded DAB audio frame generally depends on the mode and the bit rate.

The CRC word should be used to control an error concealment strategy in order to avoid severe impairments of the reconstructed audio signal due to errors in the most sensitive information.

Some basic techniques may be used for concealment, for instance information substitution, or muting. A simple substitution technique consists, when an erroneous frame occurs, of replacing it by the previous one (if error free).

In addition to the error protection facilities provided by the MPEG-1 and MPEG-2 Audio Layer II (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) audio coding standards, facilities for an error check of the Scale Factors have been provided in an ISO compatible manner. The exact method is described in clause B.3. To avoid audible distortions, evoked by erroneous Scale Factors, the application of a concealment technique, either muting of those Scale Factors where an error was detected, as a rather simple method, or a repetition of the previously received Scale Factors, which did not show an error in the 3 MSbs, as a more advanced method, is recommended.

C.6 Joint stereo coding

The optional joint stereo coding method used is intensity stereo coding. Intensity stereo coding can be used to increase the audio quality and/or reduce the bit rate for stereophonic signals. The gain in bit rate is typically about 10 kbit/s to 30 kbit/s. The additional decoder complexity is negligible. The increase of encoder complexity is small. The encoder and decoder delay is not affected.

Psychoacoustic results indicate that, at high frequencies (above about 2 kHz), the localization of the stereophonic image within a Critical band is determined by the temporal envelope and not by the temporal fine structure of the audio signal.

The basic idea for Intensity stereo coding is that for some sub-bands, instead of transmitting separate left and right sub-band samples, only the sum-signal should be transmitted, but with Scale Factors for both the left and right channels, thus preserving the stereophonic image.

Flow diagrams of a stereo encoder and decoder, including intensity stereo mode, are shown in figures C.2 and C.3. First, an estimation should be made of the necessary bit rate for both left and right channel. If the necessary bit rate exceeds the available bit rate, the necessary bit rate should be decreased by setting a number of sub-bands to Intensity stereo mode. Depending on the bit rate needed, sub-bands:

- 16 to 31;
- 12 to 31;
- 8 to 31; or
- 4 to 31,

can be set to Intensity stereo mode. For the quantization of such combined sub-bands, the higher of the bit allocations for left and right channel should be used.

The left and right sub-band signals of the sub-bands in joint stereo mode should be added. These new sub-band signals should be scaled in the normal way, but the originally determined Scale Factors of the left and right sub-band signals should be transmitted according to the bit stream syntax. Quantization and coding of common sub-band samples, and coding of common bit allocation should be performed in the same way as in independent coding of the left and right channel of a stereophonic programme.

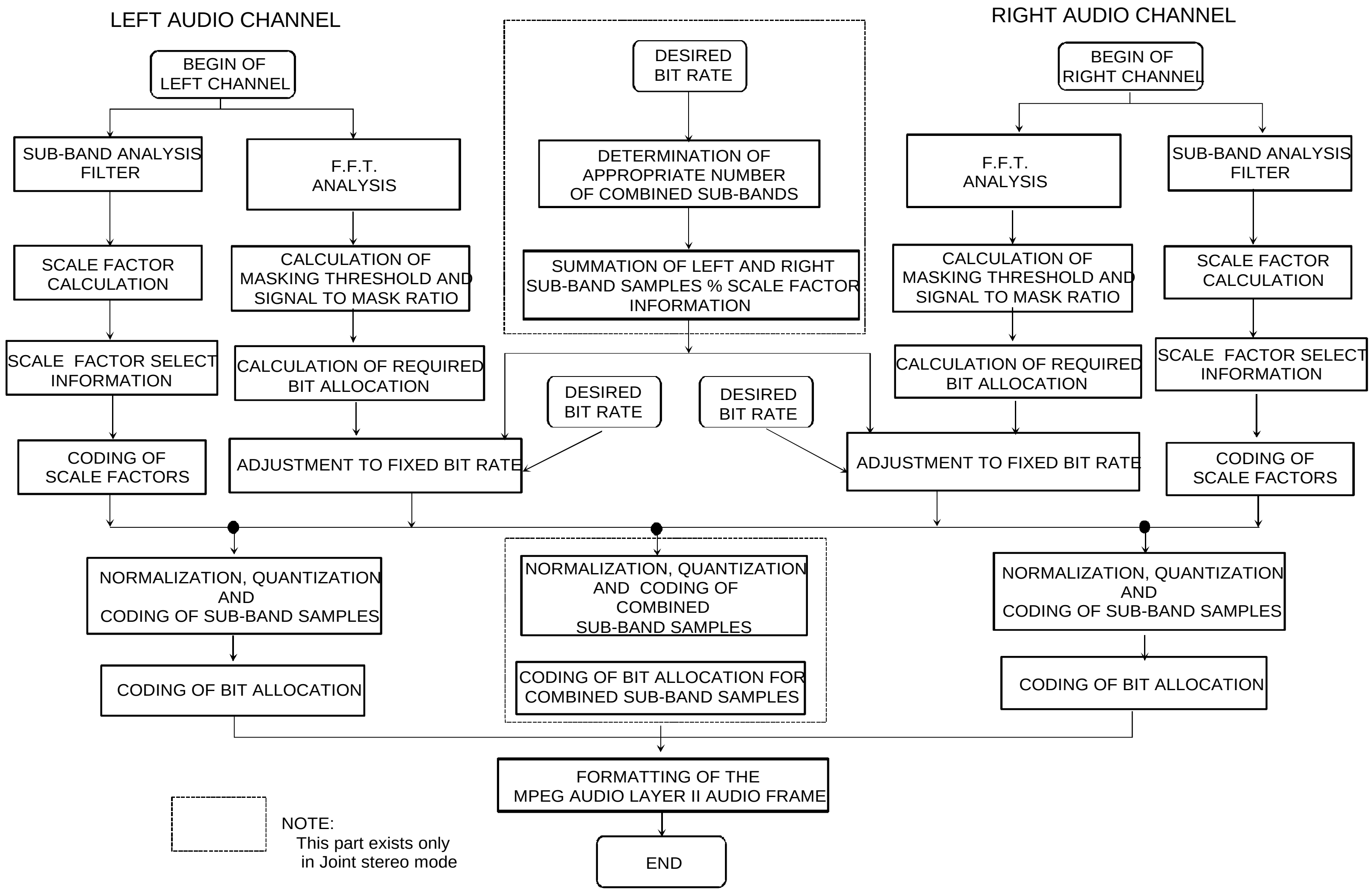

Figure C.2: General MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) Layer II stereo encoder flow chart

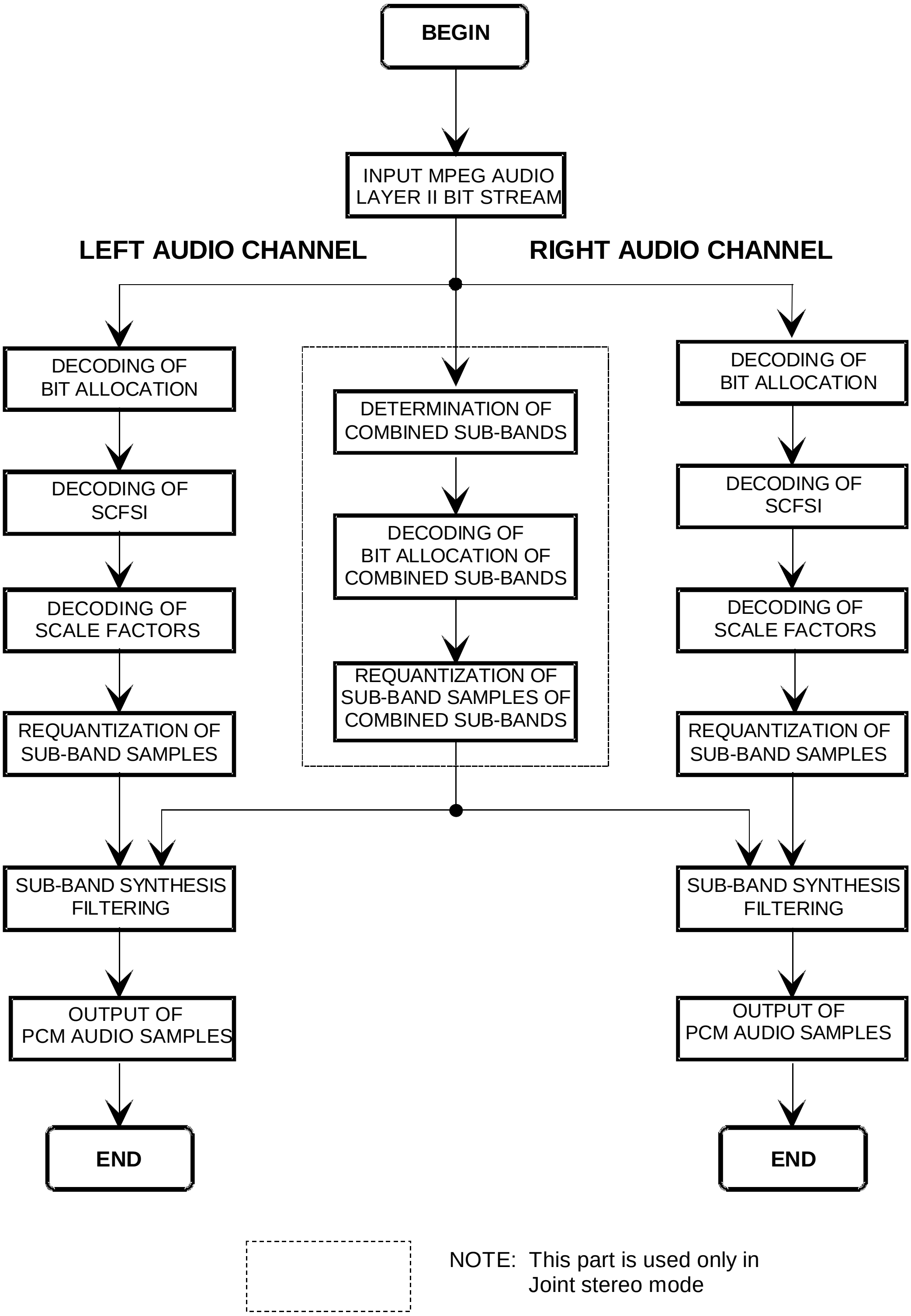

Figure C.3: General MPEG Audio (ISO/IEC 11172-3 [2] and ISO/IEC 13818-3 [4]) Layer II stereo decoder flow chart

History

Document history		
V1.1.1	October 2016	Publication
V1.2.1	September 2019	Publication