



FRAUNHOFER INSTITUTE FOR INTEGRATED CIRCUITS IIS

---

## **WHITE PAPER**

HE-AAC Metadata for Digital Broadcasting

---

## Introduction

The normalization of program loudness is a recurring topic of debate in digital broadcasting circles. Varying levels of loudness between different channels and even within the course of one program continue to give rise to complaints, with many viewers feeling that their overall enjoyment of content is being compromised[1].

These issues have been addressed for example in the great effort of the EBU group PLOUD. In the USA, the CALM act is a popular achievement on the move towards loudness normalization of program content.

In this context, audio-specific metadata plays a significant role in controlling program loudness, dynamic range and downmixing of the emitted audio bitstream.

Having been adopted in TV broadcasting standards such as DVB, ISDB-Tb and ATSC M/H, the MPEG-4 High Efficiency Advanced Audio Coding (HE-AAC) is already in use – or soon will be – by many leading national TV broadcasting services worldwide (e.g., major operators in Brazil, Norway, Sweden, Denmark, UK, New Zealand and Israel). With the HE-AAC audio codec now being a popular choice for digital broadcasting services worldwide, the importance of broadcast-related metadata as a standardized feature of this codec becomes even more acute when the requirements of program loudness normalization are taken into account.

# Concepts of Loudness Normalization

## ***Transition from Analog to Digital Television***

Traditional analog television systems have very limited dynamic range capability and peak audio levels must be tightly controlled to prevent over-modulation of the analog transmitter. Legacy audio production typically relies on VU or PPM meters for level monitoring and broadcasters use compressor/limiters to reduce the dynamic range to fit the analog modulation system. In these systems, loudness control is a byproduct of the careful control of audio levels and limited dynamic range. The television receiver reproduces what it receives and this provides reasonably acceptable quality for the listener, within the limitations imposed by the system. Digital audio production systems and NICAM audio coding increase the available dynamic range, but do not significantly change the audio monitoring and processing systems used by the broadcaster.

Digital television (DTV) broadcasting enables much higher video and audio quality. In particular, digital audio coding systems enable the possibility that audio can be coded for emission with the full dynamic range of the original production, without the constraints on peak level and limited dynamic range needed for analog modulation. However, when loudness is not managed correctly, the result may be that DTV consumers experience unacceptable variations in level. There may be annoying jumps at program, commercials, and interstitial junctions, and also loudness variations between channels.

## ***Loudness Measurement and Control***

Loudness is the perceived strength of an audio signal as heard by the human ear. It is a psycho-acoustical parameter, which cannot be measured with VU or PPM meters, and perceptual loudness measurement equipment is needed to determine the loudness of any particular item of program content. Fortunately, with recommendation BS.1770<sup>1</sup>, the ITU-R has provided a standardized method for the measurement of loudness and true peak level.

Following on from this significant development, a number of further specifications, recommended practices, and guidelines have been produced to address the implementation of effective loudness control. Examples include the Advanced Television Systems Committee (ATSC) Recommended Practice A/85<sup>2</sup> and the European Broadcasting Union (EBU) R128<sup>3</sup>. The EBU has also published related guidelines for production<sup>4</sup> and distribution systems<sup>5</sup>. In the USA, this problem has attracted the attention of legislators and implementation of the so-called CALM Act<sup>6</sup> will require broadcasters and other television program distributors to bring the loudness of commercials in line with that of other program content.

---

<sup>1</sup> ITU-R Recommendation BS.1770: Algorithms to measure audio program loudness and true-peak level

<sup>2</sup> ATSC Recommended Practice A/85 – Techniques for Establishing and Maintaining Audio Loudness for Digital Television

<sup>3</sup> EBU Technical Recommendation R 128 'Loudness normalisation and permitted maximum level of audio signals' (2010)

<sup>4</sup> EBU Tech Doc 3343 'Practical guidelines for Production and Implementation in accordance with EBU R 128' (2011)

<sup>5</sup> EBU Tech Doc 3344 'Practical Guidelines for Distribution Systems in accordance with EBU R 128' (2011)

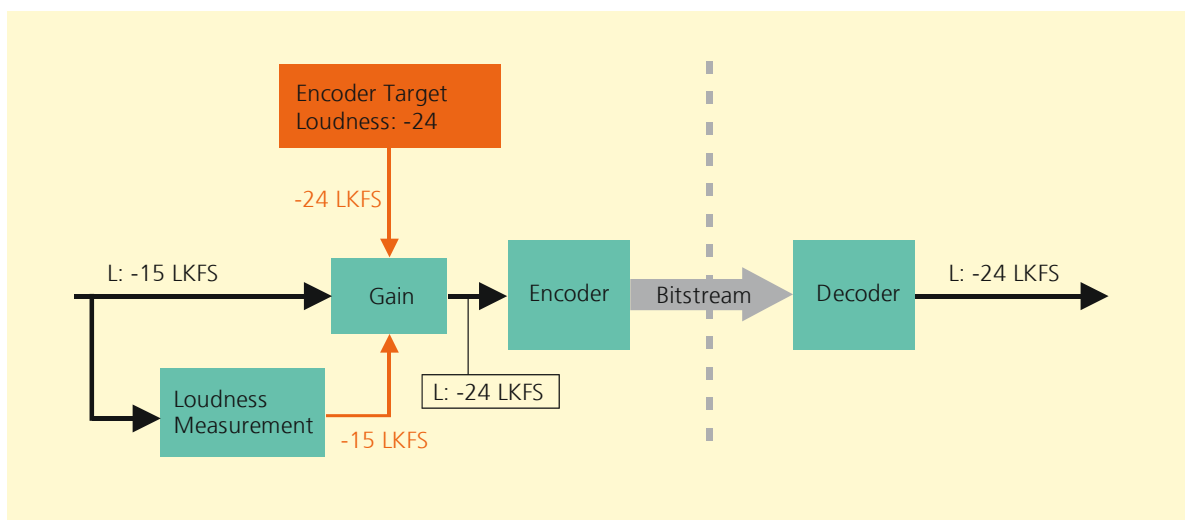
<sup>6</sup> 111th Congress 2009-2010, H.R. 1084: Commercial Advertisement Loudness Mitigation Act

For DTV, there are two basic approaches to controlling loudness variations between different kinds of program content in order to maintain consistent loudness for the consumer. One requires the use of metadata to signal the loudness of content distributed to the consumer and the other approach does not require the use of metadata.

### **Loudness Management without Metadata**

To manage DTV audio loudness without using metadata, a single loudness target value for all content is adopted by the broadcaster. This target loudness is accomplished either manually in production, with the help of loudness meters, or by automatic measurement and correction at some point in the chain prior to encoding. Following this loudness normalization, the audio content can be encoded at the intended loudness for distribution and emission to the consumer.

Figure 1 shows the concept of loudness normalization with measurement and correction before encoding, as it may be accomplished with file-based distribution. In the case shown in the figure, the incoming audio has loudness of -15 LKFS<sup>7</sup> and the encoder target loudness is -24 LKFS, which is also the loudness output by the decoder.



**Figure 1: Loudness normalization before encoding**

DTV receivers for broadcast systems without audio metadata will reproduce the audio as received. Therefore, in addition to the simple gain adjustment as shown in the drawing, prior to encoding by the broadcaster the audio must be processed to control the dynamic range as appropriate for the intended service and receiver characteristics.

Without metadata to guide it, the decoder has no control of loudness or dynamic range. Therefore, in order to avoid consumer complaints, the broadcaster may be obliged to apply significant dynamic range reduction prior to encoding, to suit the characteristics of

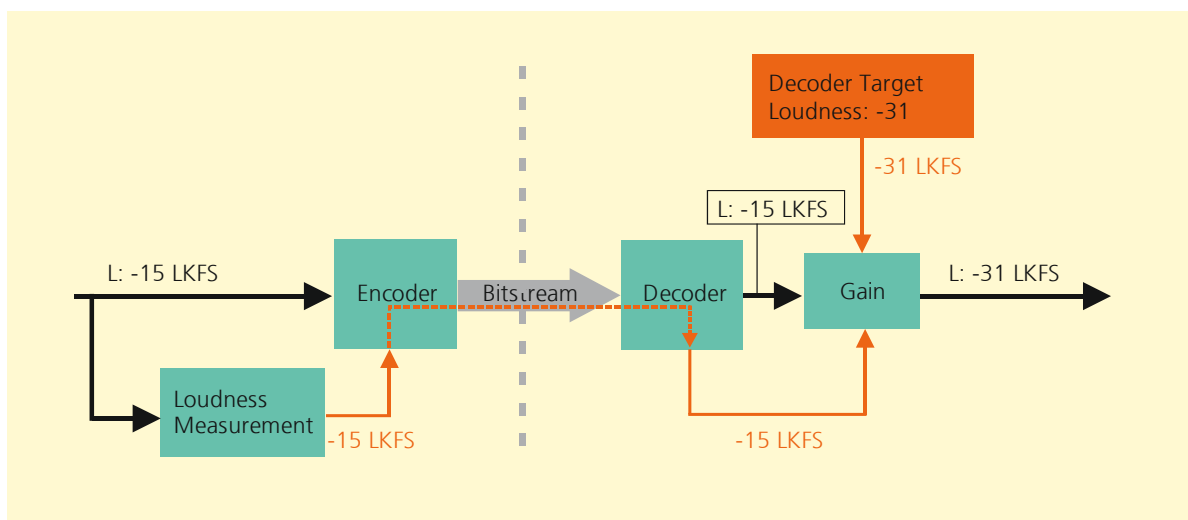
<sup>7</sup>LKFS (Loudness, K-weighted, relative to Full Scale), measured with equipment that implements the algorithm specified by ITU-R BS.1770. A unit of LKFS corresponds to a decibel. The EBU introduced the equivalent measurement unit LUFS (Loudness Units referenced to Full Scale) and extended the ITU measures with the relative measure LU (Loudness Unit).

comparatively low quality television audio systems and, perhaps, noisy listening environments. Unfortunately, consumers with high quality audio systems, such as home theaters, are then denied the full range that their systems can provide. This is the same situation as for analog television and it compromises the quality advantage of digital audio. This situation can be avoided by using audio metadata for loudness management.

## **Loudness Management with Metadata**

Audio metadata is an instruction set carried with the audio essence that describes the audio content and how it should be reproduced by the consumer device at the end of the transmission chain. The audio itself can be encoded to suit the highest quality consumer audio systems and listening environments. When used in distribution, audio metadata may also include parameters relating to control of a downstream emission encoder.

Figure 2 shows the concept of a system using metadata for loudness normalization. The audio is measured by the broadcaster to determine its loudness, in this case -15 LKFS, which is then sent as metadata with the coded audio. The decoder uses the metadata to adjust the audio output level to achieve the decoder target loudness of -31 LKFS.



**Figure 2: Metadata utilized in the encoder/decoder chain**

The metadata sent with the coded audio provides information not only for the loudness of that program, but also for the dynamic range control values, how many channels have been encoded, and how to downmix those channels if necessary. The DTV receiver or set-top box uses the metadata to control the audio output to suit the audio system and listening environment, with consistent loudness and appropriate dynamic range.

A broadcaster may choose to operate a system in which individual items of program content may have different loudness values and the metadata varies from item to item. This is known as “agile” metadata. Alternatively, a broadcaster may choose to adopt a single loudness target for all content, and the loudness metadata can remain fixed at one value to indicate the chosen level of program loudness.

In both cases, the metadata has to be generated for each item of program content and, depending on where this takes place, may have to be conveyed over all stages of contribution and distribution as well as emission. This is described further for HE-AAC metadata in the next section.

## HE-AAC Metadata for Digital Broadcasting

Audio-related metadata for broadcasting systems is generated typically at some point in the content production chain and can be conveyed alongside the coded audio. Three features of metadata are of particular importance and are frequently referred to as the “3 Ds”:

- *Dialogue Normalization* is used to adjust and achieve a constant long-term average level of the main program components across various program materials, e.g., a feature film interspersed by commercials.
- *Dynamic Range Control (DRC)* facilitates control of the final dynamic range of the audio and adjusts compression to suit individual listening requirements.
- *Downmix* maps the channels of a multi-channel signal to the user’s mono or two-channel stereo speaker configuration.

These terms come from the metadata parameters defined for the AC-3 audio codec, used for emission in some digital television systems. They also relate to the Dolby® E<sup>8</sup> audio codec, still widely used in the program broadcast production and contribution chain.

The AAC codec – as well as its derived family members HE-AAC and HE-AAC v2<sup>9</sup> – supports these same features. The naming convention is slightly different: the following table compares the Dolby® nomenclature of the parameters listed above with their equivalents specified for the AAC codec.

### Nomenclature of HE-AAC metadata in comparison to Dolby metadata

	(HE) AAC	(E-)AC-3
<b>Loudness Normalization</b>		
	“Program Reference Level”	“Dialnorm”
<b>Dynamic Range Control</b>		
“Light Compression”	“Dynamic Range Control”	“Line Mode”
“Heavy Compression”	“compression value”	“RF Mode”
<b>Downmix</b>		
	„matrix-mixdown” „Downmixing Levels”	“Downmix”

<sup>8</sup> Dolby and Dolby E are registered trademarks of Dolby Laboratories

<sup>9</sup> Identical metadata is defined for all members of the AAC codec family, which includes HE-AAC and HE-AAC V2, and other versions of the codec, although not necessarily used with all implementations. In this paper, the terms AAC metadata and HE-AAC metadata refer to the same specifications and functionality.

The following paragraphs provide an overview of the specified AAC metadata and its utilization. More detailed specifications, and references to the particular standards where they are defined, are given in *Appendix: AAC Metadata Specifications*.

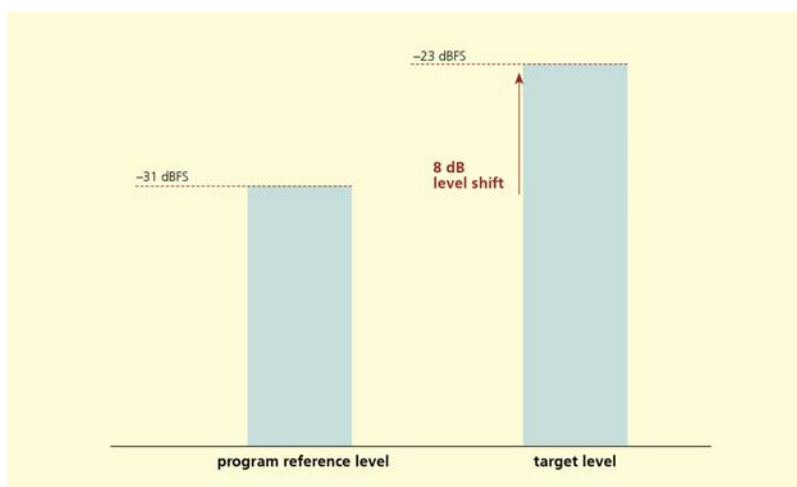
### **Program Reference Level**

This parameter can be used to indicate the loudness of a transmitted program, indicating the program loudness relative to full-scale digital signal level in a range between 0 dBFS and -31.75 dBFS. During decoding, this parameter allows the audio output level to be changed to match a given target level. This parameter is expected to remain constant for the duration of each item of program content.

The target level is the intended output level of the decoder. It is a modifiable, external parameter given to the decoder according to the receiver requirement specifications of the related broadcast system. The required level change  $g$  to match the desired target level results from the difference between the target level  $L_{target}$  and program reference level  $L_{ref}$ .

$$g = L_{target} - L_{ref}$$

Figure 3 illustrates the utilization for a Program Reference Level of -31 dBFS and a given target level of -23 dBFS.



**Figure 3: Utilization of the Program Reference Level to match a given target level**

### **Dynamic Range Control**

In order to adjust the dynamic range to complement listening requirements, HE-AAC is able to convey specific parameters to be applied in the receiver. The application of dynamic range compression is a user-selectable feature of the decoder. However, dynamic range compression is carried out automatically in the case of a downmix in order to avoid overload of the output signal. There are two compression modes available: one for 'light' and one for 'heavy' compression.

## Light Compression

The values are inserted into the bitstream for each frame according to the desired compression characteristic. The decoder may apply the gains to the spectral data in conjunction with the program reference level prior to frame decoding in order to achieve a reduction of the dynamic range.

This parameter is equivalent to Dolby®'s "Line mode" compression, while the application of dynamic range compression is a user-selectable feature of the decoder.

HE-AAC is capable of relaying the compression in a frequency-selective manner (multi-band compression).

## Heavy Compression

Some listening environments require a very limited dynamic range. Accordingly, the so-called 'midnight mode' of AV receivers applies a strong dynamic compression, for instance to allow intelligibility of the movie dialogue at low playback volumes and simultaneously restricting the volume of special effects such as explosions. The heavy compression metadata included in the bitstream guides the process, which is the equivalent of Dolby®'s "RF mode" compression.

## Downmix/Matrix-Mixdown

In order to represent a multi-channel signal on a two-channel stereo or mono loudspeaker set-up, the HE-AAC decoder applies a downmix. Related metadata parameters contain information about the downmix coefficients for the Center channel and Surround Channels and thereby allow a certain level of control of the resulting downmix.

## HE-AAC Metadata in Practice

### *The Broadcast Chain*

Depending on the work-flow scenario, a broadcaster may choose different methods to obtain the metadata that will be embedded into the emitted HE-AAC bitstream.

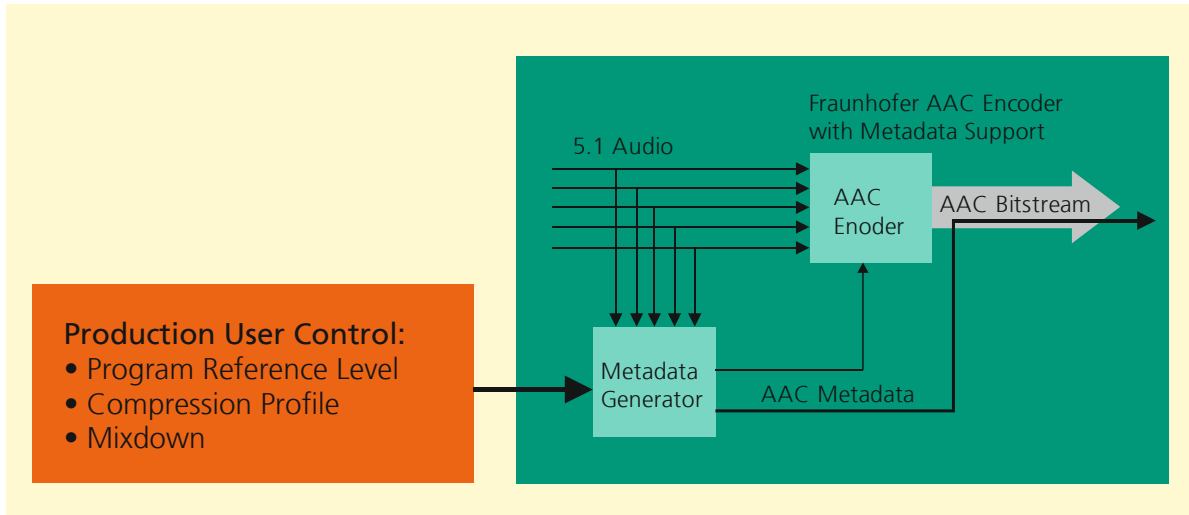
One possible route is for metadata to be supplied to the encoder as input parameters. For example, a broadcaster may select the desired metadata settings individually or choose from a set of pre-defined metadata profiles according to the type of program content. Alternatively, the metadata may originate from earlier stages of the production chain and be carried alongside the audio signal by different interfaces such as Dolby® E, HD-SDI<sup>10</sup>, SDI<sup>11</sup> or AES-41. Similarly, file-based production work-flows employ file formats like the Broadcast Wave File (BWF) or the Material Exchange Format (MXF). Due to the introduction of loudness measurement and normalization, the utilization of loudness-specific metadata based on the ITU or EBU recommendations for loudness increases in importance. Accordingly, these can serve as an

---

<sup>10</sup> SMPTE 292M with SMPTE 2020

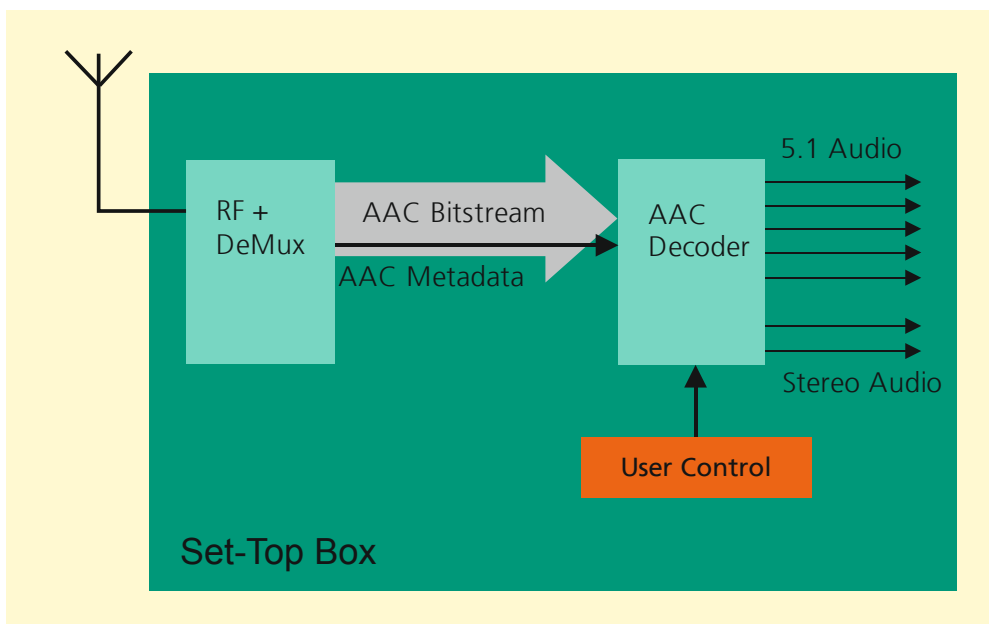
<sup>11</sup> SMPTE 259M with SMPTE 2020

alternative basis to generate the metadata for an encoder. For example, the measured program loudness is practically equivalent to the program reference level parameter.



**Figure 4: HE-AAC encoder with metadata fed as encoder input parameters**

At the receiver end of the chain, the HE-AAC audio bitstream is demultiplexed from the transport stream, and the audio channels are decoded to PCM by an HE-AAC decoder. The embedded audio metadata parameters are extracted and applied during decoding. If necessary, the HE-AAC decoder also applies a downmix of the 5.1 surround audio signal depending on the user's audio system and loudspeaker configuration.

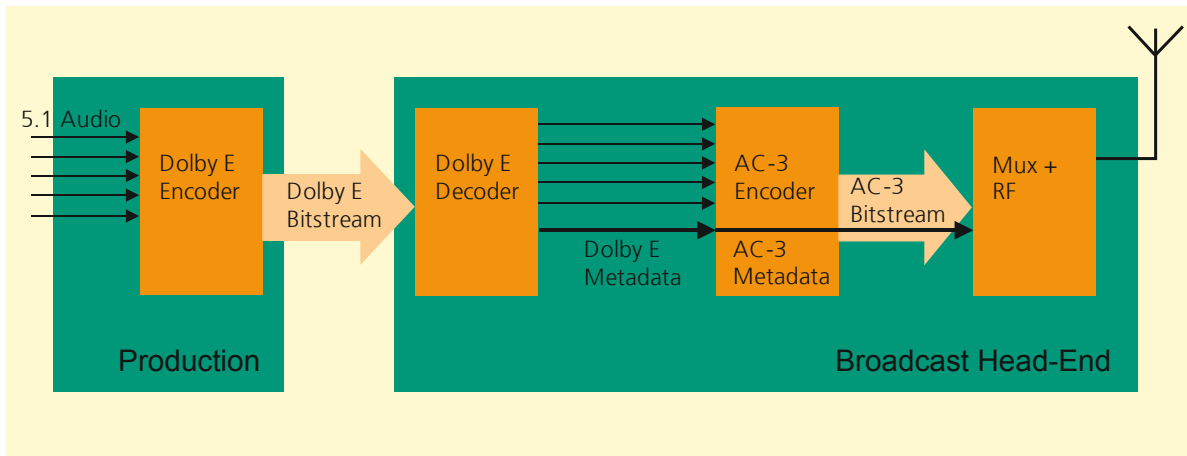


**Figure 5: Receiver device with metadata-capable HE-AAC Decoder**

HE-AAC bitstreams incorporating metadata are compatible with legacy HE-AAC decoders without metadata support. These implementations ignore the metadata information and play back the HE-AAC audio stream without applying the embedded metadata.

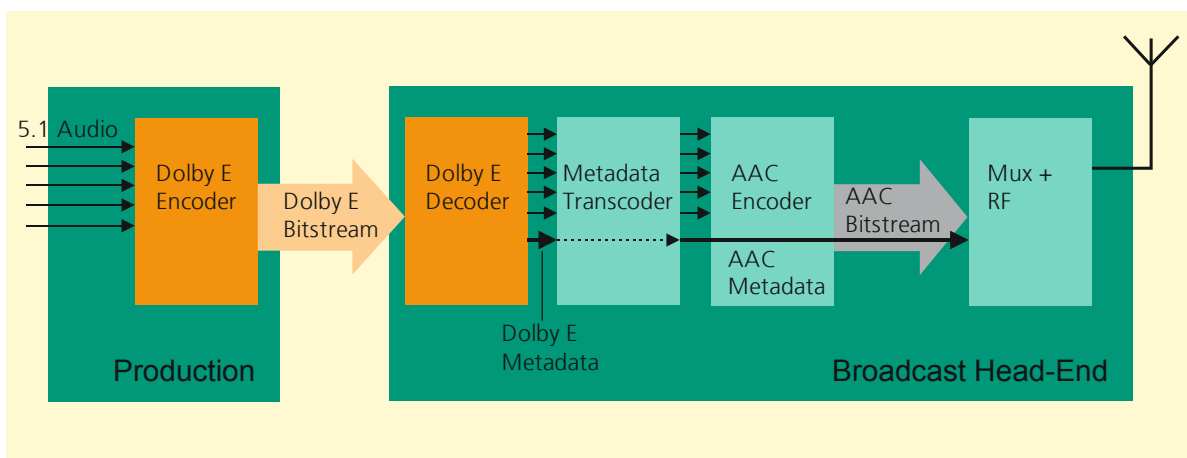
## Transcoding of Metadata in the Broadcast Environment

Many broadcasters use metadata with the Dolby E format in contribution or production. For transmission, this metadata must be converted to equivalent metadata for the AC-3 or HE-AAC codecs. For broadcasting systems which use AC-3, the desired parts of the metadata will be embedded into AC-3 bitstream by AC-3 encoders.



**Figure 6: DVB broadcasting chain with transcoding of an arbitrary program provided by Dolby E into AC-3**

For broadcasting systems which use Dolby® E and existing Dolby® E metadata, the desired parts of the metadata from the Dolby® E system can be preserved and mapped to HE-AAC bitstream elements. The multi-channel audio signal is coded in HE-AAC and carries the corresponding audio metadata derived from the Dolby® E system. In order to convey the supplied metadata parameters alongside the emission codecs such as AC-3 or HE-AAC, they have to be transcoded appropriately[2]. Transcoding of the dialogue normalization and downmix parameters is straightforward while the conversion of DRC metadata has to take into account different frame-sizes of the codecs[2].



**Figure 7: DVB broadcasting chain with transcoding of an arbitrary program provided by Dolby E into HE-AAC**

## ***Transcoding of Metadata in the Home***

In the home environment, different solutions can be considered to handle the HE-AAC bitstream at its metadata:

1. The HE-AAC bitstream can be directly decoded while applying the accompanying metadata in integrated receiver/decoders (IRD) or set-top boxes (STB). While an IRD will utilize the internal loudspeakers in most cases, the audio signal decoded in an STB can be conveyed through the S/P DIF or HDMI interface in order to feed an AV receiver (AVR) or TV set. While the HDMI interface is multi-channel capable, the S/P DIF interface is limited to two-channel stereo in the PCM domain.
2. The HE-AAC bitstream received by an STB can be passed-through to the AVR or TV via the S/P DIF or HDMI interface.<sup>12</sup> This possibility is currently limited by the lack of support by legacy AVRs and therefore not yet provided by receiver requirement specifications.
3. Accordingly, the backwards-compatible approach chosen by several receiver requirement specifications is to transcode the received HE-AAC bitstream into the legacy formats DTS or AC-3 when a multi-channel audio signal is to be conveyed from an STB to an AVR.

In the latter case, the HE-AAC metadata can either be applied during transcoding or be transcoded into the correspondent AC-3 metadata in order to guide the process of decoding the AC-3 bitstream in the sink device. For transcoding of metadata at this side of the chain, the same considerations are valid as described above for transcoding at the encoder side.

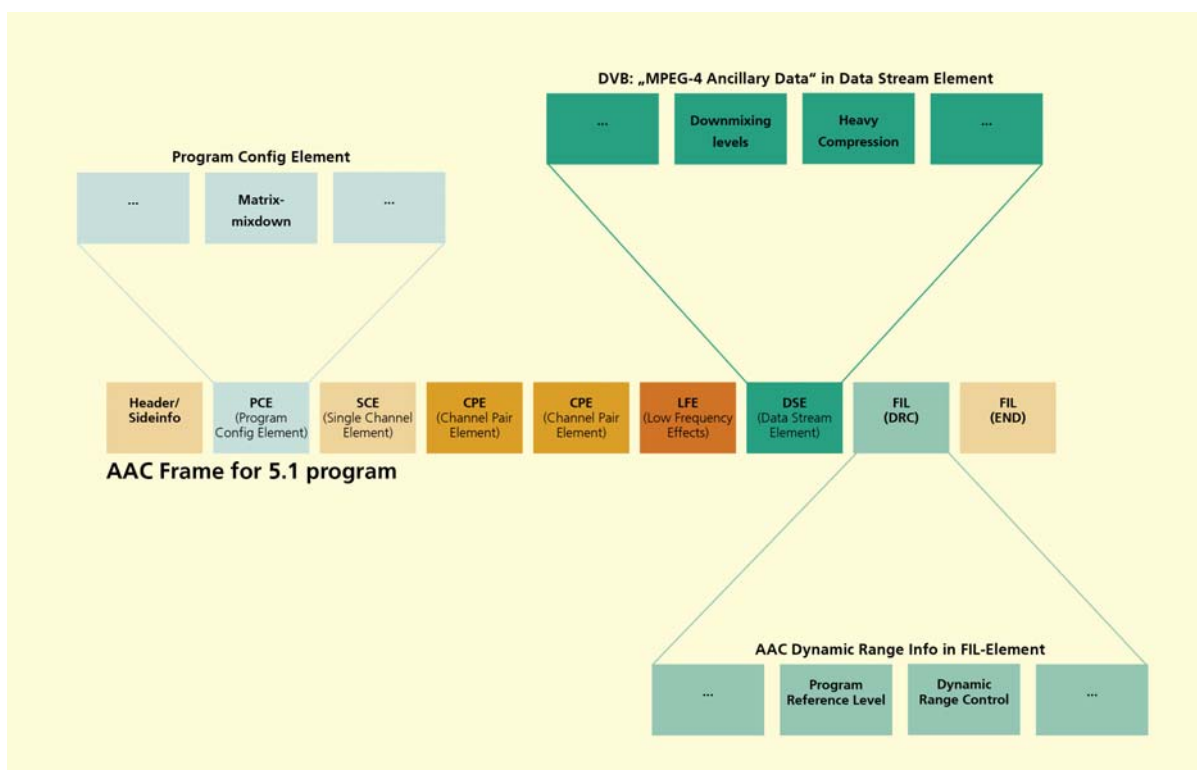
---

<sup>12</sup> The transmission of MPEG-4 AAC bitstreams over S/P DIF has been specified in IEC 61937-11.

## Appendix: AAC Metadata Specifications

### Bitstream Syntax

Figure 8 gives an overview of the MPEG-4 AAC bitstream structure and allocation of the metadata elements. The program reference level, dynamic range settings for light compression and the matrix-mixdown coefficients are defined in ISO/IEC 14496-3 (MPEG 4, part 3: Audio)<sup>13</sup>. Additional metadata for heavy compression gains and higher resolution downmixing levels) may be inserted as ancillary data. Such additional data structures are defined and/or referenced by other standards such as DVB (ETSI TS 101 154)<sup>14</sup> and the Brazilian ISDB-Tb (ABNT NBR 15602-2)<sup>15</sup>.



**Figure 8: Metadata in the AAC bitstream**

Details of the individual metadata elements are set out below, with the equivalent elements for AC-3 shown for comparison.

<sup>13</sup> ISO/IEC 14496-3:2009 – Information technology – Coding of audio-visual objects – Part 3: Audio, Fourth edition 2009

<sup>14</sup> ETSI TS 101 154 V1.9.1 – Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream, Sept. 2009

<sup>15</sup> ABNT NBR 15602-2 – Digital terrestrial television – Video coding, audio coding and multiplexing – Part 2: Audio coding

## Program Reference Level

	AC-3	(HE) AAC
<b>LOUDNESS NORMALIZATION</b>	"Dialnorm"	"Program Reference Level"
<b>Bitstream field name</b>	dialnorm	prog_ref_level
<b>Range</b>	-1 ... -31 dB	0 ... -31.75 dB
<b>Granularity</b>	1 dB	0.25 dB
<b>Bits per Value</b>	5 Bits	7 Bits
<b>Repetition Rate</b>	1 Value per Frame (1536 Samples)	1 Value per Frame (1024/2048 Samples)
<b>Decoder Output Level</b>	-31/-20 dB (fixed) (Line-Mode/RF-Mode)	0 ... -31.75 dB
<b>Usage</b>	Mandatory	Mandatory in several application standards such as DVB and ISDB-Tb

For MPEG-4 AAC, the program reference level is specified as part of the dynamic range information structure in ISO/IEC 14496-3 Subpart 4, Chapter 4.5.2.7. The dynamic range information structure is located in the Fill Element (see Figure 8). It should be noted that in the AAC specification, the term Dynamic Range Control (DRC) is used to refer to the overall structure that includes both program reference level and dynamic range compression parameters.

## Dynamic Range Compression

	AC-3	(HE) AAC
<b>LIGHT COMPRESSION</b>	“Line Mode”	MPEG “Dynamic Range Control”
<b>Bitstream field name</b>	dynrng	dyn_rng_ctl, dyn_rng_sgn
<b>Range</b>	-24 ... +24 dB	-31.75 ... +31.75 dB
<b>Granularity</b>	0.25 dB	0.25 dB
<b>Repetition Rate</b>	6 Values per Frame (1536 Samples)	1 Value per Frame (1024/2048 Samples) + “Interpolation scheme”
<b>HEAVY COMPRESSION</b>	“RF Mode”	DVB “compression value”
<b>Bitstream field name</b>	compr	compression_value
<b>Range</b>	-48 ... +48 dB	-48 ... +48 dB
<b>Granularity</b>	0.5 dB	0.5 dB
<b>Repetition Rate</b>	1 Values per Frame (1536 Samples)	1 Value per Frame (1024/2048 Samples)

The parameters for light compression as described above are covered by the MPEG-4 AAC DRC tool defined in ISO/IEC 14496-3, Chapter 4.5.2.7. The related parameters for dynamic range compression and program reference level are co-located in the dynamic range information structure in the Fill Element. Application standards such as ETSI TS 101 154 require receiver devices to support the DRC tool.

The parameters for heavy compression reside in the MPEG-4 Ancillary Data, which is part of the Data Stream Element (see Figure 8). The signaled 8-bit compression values allow a compression range of +/- 48dB and are applied to the spectral values data prior to reconstruction. Unlike the metadata for light compression, heavy compression is not part of the MPEG specification but is complemented by the specification for DVB in ETSI TS 101 154, Chapter C.5.

## Downmix

	AC-3	(HE) AAC
<b>Coefficients</b>	Center: {-6,-4.5,-3} dB Surround: {-∞,-6,-3} dB	matrix-mixdown: Center: -3dB Surround: {-∞,-6,-3,0} dB  <i>or</i>  Ancillary Data: Center + Surround: {-∞,-9,-7.5,-6,-4.5,-3,-1.5,0} dB

The **matrix-mixdown** method for deriving a stereo or mono signal from 5.1 channels, as specified in ISO/IEC 14496-3, Chapter 4.5.1.2.2, provides a downmix coefficient for the surround channels of a multi-channel signal, while the center channel is specified with a fixed value. The specification contains four possible surround channel values (-3 dB, -6 dB, -9 dB and -∞ dB). The matrix-mixdown parameters are located in the Program Config Element (PCE) (see Figure 8).

ETSI TS 101 154, Chapter C.5 specifies the **downmixing levels** structure for usage in DVB, enabling the transmission of downmix coefficients with a higher resolution than the ISO standard, in a range from 0 dB to -9 dB in 1.5 dB steps. Furthermore, the downmix coefficients can be set to -∞ dB, which results in a complete removal of the related channels. In addition to the surround channels, the structure includes a downmix coefficient for the center channel, which gives even more freedom to control the downmix. The downmix level parameters are conveyed in the Ancillary Data of the Data Stream Element (DSE) (see Figure 8).

## Fraunhofer IIS Implementations

As one of the leading suppliers, the Fraunhofer IIS offer HE-AAC Encoder and Decoder implementations which support the standardized metadata as described above. In addition, Fraunhofer IIS offers dedicated solutions for metadata creation and transcoding. The following list summarizes the available software implementations:

- **HE-AAC Encoder and Decoder** implementations with metadata support are available for various platforms, including optimized implementations for PC as well as for floating- and fixed-point DSP platforms.
- The **metadata generator** as part of the HE-AAC Encoder implementation inserts metadata into the bitstream from the given user-input
- The **metadata transcoder** converts Dolby® E or AC-3 metadata into HE-AAC metadata

For further information and evaluation regarding the available solutions, please refer to <http://www.iis.fraunhofer.de/amm>.

### ABOUT FRAUNHOFER IIS

Fraunhofer IIS, based in Erlangen, Germany, is the home of the Fraunhofer Audio and Multimedia division, which has been working in compressed audio technology for more than 20 years and remains a leading innovator of technologies for cutting-edge multimedia systems. Fraunhofer IIS is universally credited with the development of mp3 and co-development of AAC (Advanced Audio Coding) as well as technologies for the media world of tomorrow, including MPEG Surround, MPEG Spatial Audio Object Coding and the Fraunhofer Audio Communication Engine.

Through the course of more than two decades, Fraunhofer IIS has licensed its audio codec software and application-specific customizations to at least 1,000 companies. Fraunhofer estimates that it has enabled more than 1 billion commercial products worldwide using its mp3, AAC and other media technologies.

The Fraunhofer IIS organization is part of Fraunhofer-Gesellschaft, based in Munich, Germany. Fraunhofer-Gesellschaft is Europe's largest applied research organization and is partly funded by the German government. With nearly 17,000 employees worldwide, Fraunhofer-Gesellschaft is composed of 59 Institutes conducting research in a broad range of research areas. For more information, contact Matthias Rose, [matthias.rose@iis.fraunhofer.de](mailto:matthias.rose@iis.fraunhofer.de), or visit [www.iis.fraunhofer.de/amm](http://www.iis.fraunhofer.de/amm).

## References

- 
- [1] Moerman, Jean Paul: Program Loudness: Nuts & Bolts. 118th AES Convention, 2005
  - [2] Schildbach, Wolfgang; Krauss, Kurt; Rödén, Jonas: Transcoding of Dynamic Range Control Coefficients and Other Metadata into MPEG-4 HE AAC. 123rd AES Convention, 2007