APPLICATION BULLETIN

xHE-AAC IN DIGITAL RADIO MONDIALE (DRM)

IMPLEMENTATION GUIDELINES FOR THE REALIZATION OF xHE-AAC IN THE DRM FRAMEWORK

With the adoption of xHE-AAC (Extended High Efficiency Advanced Audio Coding) [1] into the Digital Radio Mondial (DRM) specification [2], this digital radio standard now features a robust and flexible audio codec that adapts to the program content and delivers consistent quality for all content types – even at extremely low data rates. xHE-AAC therefore sets out to enable new services in the DRM framework that have previously not been feasible due to the inherent drawbacks in the previously specified codecs.

The objective of this document is to raise the level of understanding of the DRM system and thus ensure that all technical components work together smoothly in the overall DRM environment.

This document gives a rough outline of the overall system and also provides very detailed and specific guidance on how to best employ xHE-AAC in DRM. It identifies peculiarities and potential stumbling blocks that are typical to complex technical systems of this magnitude and describes recommended practices on how to deal with them.

Fraunhofer Institute for
Integrated Circuits IIS

Director
Prof. Dr.-Ing. Albert Heuberger
Am Wolfsmantel 33
91058 Erlangen
www.iis.fraunhofer.de

Contact
Matthias Rose
Phone +49 9131 776-6175
matthias.rose@iis.fraunhofer.de

Contact USA
Fraunhofer USA, Inc.
Digital Media Technologies*
Phone +1 408 573 9900
codecs@dmt.fraunhofer.org

Contact China
Toni Fiedler
Phone +86 138 1165 4675
china@iis.fraunhofer.de

Contact Japan
Fahim Nawabi
fahim.nawabi@iis.fraunhofer.de

Contact Korea
Howard 주영주
Phone: +82 2 948 1291
youngju.ju@iis-extern.fraunhofer.de

* Fraunhofer USA Digital Media Technologies, a division of Fraunhofer USA, Inc., promotes and supports the products of Fraunhofer IIS in the U. S.

CONTENTS

1. INTRODUCTION

The DRM Broadcasting system has been designed by broadcasters, for broadcasters, but with the active assistance and participation of both transmitter and receiver manufacturers as well as other interested parties (for example, regulatory bodies and R&D organizations). It has been designed specifically as a high quality digital replacement for current analogue radio broadcasting in the AM and FM/VHF bands; as such it can be operated with the same channeling and spectrum allocations as currently employed and is designed to fit in with the existing AM and FM broadcast band plans. [3]

DRM on short, medium and long wave for broadcasting bands up to 30 MHz (called ‚DRM30') facilitates large coverage areas and low power consumption. The DRM standard for broadcast frequencies above 30MHz (called ‚DRM+') uses the same audio coding, data services, multiplexing and signaling schemes as DRM30, but introduces an additional transmission mode optimized for those bands. [3]

The DRM system has adopted the latest audio coding technology from MPEG, xHE-AAC (Extended High Efficiency Advanced Audio Coding). xHE-AAC can operate from only 6 kbit/s per channel, and handles both speech and general-purpose audio content equally well. In addition, DRM still provides AAC audio coding to continue support for on-air services. Existing DRM broadcasts using AAC can migrate to xHE-AAC by upgrading the audio coding library of content servers. [3]

2. TECHNICAL BACKGROUND

2.1 DRM General System – Multiplexing Mechanism

A modern digital broadcast system like DRM allows the transmission of additional digitally encoded information (text, pictures, etc.) along with the main audio program signals. This requires a multiplexing mechanism to combine all these different types of content (audio, multimedia and other data).

The DRM multiplex consists of three different components:

1) Main Service Channel (MSC)
The MSC contains the data for the services (audio and multimedia applications).

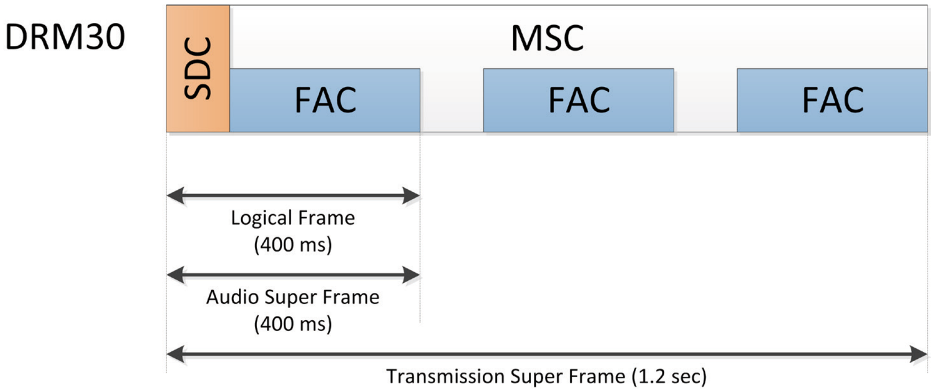
2) Fast Access Channel (FAC)
The FAC provides information on the channel parameters (audio/data stream count, spectrum, signal encoding) as well as service parameters to allow for fast scanning.

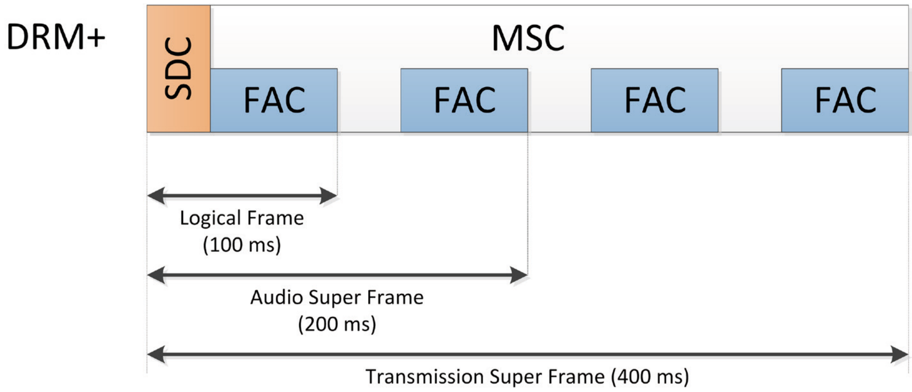
3) Service Description Channel (SDC)
The SDC gives information on how to decode the main service channel. Besides service-related parameters like program label, country of origin and other parameters it also provides all necessary information required for decoding the audio streams.

These three multiplex components are treated as three independent logical channels.

The multiplexing mechanism of DRM utilizes a transmission super frame structure. A transmission super frame has a duration of 1.2 seconds (DRM30) or 400 ms (DRM+).

Each transmission super frame is divided into logical frames of 400 ms (DRM30) or 100 ms (DRM+) length each.





Multiplex structure of DRM30 and DRM+ transmission schemes.
Note the different timescales for DRM30 and DRM+.

The SDC information is not sent within every logical frame but only once per DRM transmission super frame (1.2 sec / 400 ms). The total amount of data may require more than a single SDC block to be sent. Therefore it may take the duration of multiple transmission super frames until the receiver has collected all the information required to playback and decode the DRM multiplex.

In case the MSC stream contains encoded audio content, a matching of logical frames to audio super frames has to be conducted according to:

- DRM30: One audio super frame (ASF) matches one logical frame, i.e. 400ms
- DRM+: One audio super frame spans two successive logical frames, i.e. 200ms

2.2 xHE-AAC

2.2.1 General

xHE-AAC is an update of the well-known HE-AACv2 standard [4], which has been employed in digital radio systems and TV broadcast systems worldwide. It is therefore the latest addition to the AAC family of codecs.

xHE-AAC is particularly powerful at encoding any audio content – including plain speech – at very low bit rates (down to around 6 kbit/s per channel). This is noteworthy because so far the DRM system has required the implementation of multiple audio codecs to encode general audio content (such as music) and plain speech equally well: two dedicated codecs for coding of speech – and only speech – and another codec for coding any other program content.

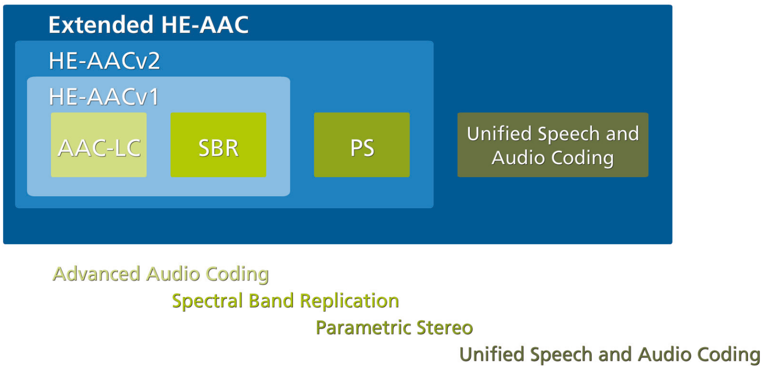
With its characteristic of being equally suitable for speech and music, xHE-AAC frees the broadcasters from having to choose and configure the codec best suited to the current content.

In addition to simplifying the selection of the optimal codec for broadcasters, xHE-AAC as the single solution eases the implementation for receiver manufacturers.

2.2.2 Codec Evolution

In terms of technical evolution xHE-AAC is the successor of the well-known and proven HE-AACv2 audio codec. xHE-AAC adds technology that addresses mainly the efficient coding of speech and audio signal components at very low bit rates. Further modifications, e.g. to the stereo and entropy coding modules, improve the audio quality at higher rates too.

AAC family profile hierarchy



A novelty relative to HE-AACv2 is the fact that the configuration of the codec is handled entirely inside the codec. Whereas HE-AACv2 required knowledge of the internal functionality of bandwidth extension and parametric coding modules, xHE-AAC requires only the declaration of bit rate and channel configuration at initialization time. All internal coding parameters are automatically set to guarantee optimum coding efficiency at maximal achievable audio quality.

2.2.3 Essential Coding Parameters

In order to be able to achieve maximum coding efficiency and audio quality, xHE-AAC in DRM adjusts certain parameters that allow the coding algorithm to work at an optimum operation mode. Some parameters may have an effect on the way the codec interacts with other elements of the DRM framework audio system, most notably these are audio processing frame size (number of audio samples that are processed as a block by encoder and decoder, a.k.a. granule length) and the audio sampling rate (number of audio samples per second).

Possible audio frame lengths for xHE-AAC

Number of samples per audio processing frame (per channel), `granuleLength`
768
1024
2048
4096

Possible audio sampling rates for xHE-AAC

Audio Sampling Rate (samples per second), `sampleRate`
9600
12000
16000
19200
24000
32000
38400
48000

2.2.4 Dynamic Bit Allocation and the „Bit Reservoir“

One inherent principle that modern audio codecs apply when processing arbitrary audio content is the idea of dynamic bit allocation.

Some Background:

Audio coders operate on a block-by-block basis, i.e. a fixed number of audio samples (fixed duration of audio content, typically several 10 ms) are processed as a whole. One processing block of audio signal is called an „audio processing frame“ or just „audio frame“. The compressed representation of an encoded audio frame is typically called an audio access unit (AU).

Further, it is important to understand that specific audio frames may be more difficult to encode than others. „Difficult“ in this context means that more bits are required for encoding the frame at a certain level of quality. As an obvious example it can be anticipated that the encoding of silence is less demanding than that of a violin playing an intricate melody.

The audio encoder in this case uses a lower number of bits for encoding the less-demanding audio frames. The result is a bit stream that exhibits certain variability in the momentary bit rate. The long term average bit rate, however, stays constant.

In other words the encoder can „save“ bits when coding a less demanding audio frame and can „spend“ a larger-than-average number of bits when coding more challenging audio frames. The concept of „saving“ and „spending“ bits led to the figurative term of „bit reservoir“, which means the pool of bits that can be shifted between audio frames depending on momentary encoder bit demand.

In the context of systems which build on constant bit rate (CBR) transmission channels the use of the bit reservoir allows the combining of the advantages of a constant rate channel with the benefits of variable bit rate audio coding.

The consequences of employing a bit reservoir are manifold. These are the most important aspects to mention:

- The bit reservoir has a maximum limit. This means that the momentary bit rate cannot exceed certain limits (The maximum AU size in bits is limited). Also, an encoder may only produce a limited number of very small AUs in a row.
- The decoder must buffer a certain number of bits of an audio bit stream before decoding of the access units may start. Fortunately, this minimum number of bits is signaled to the decoder as part of the DRM xHE-AAC audio super frame header („bit reservoir level“).
- A DRM xHE-AAC audio super frame (ASF), which contains a constant number of bits, may contain a varying number of encoded audio frames. The number of AUs that start in an ASF is signaled as part of the DRM xHE-AAC audio super frame header („Frame border count“).

2.3 Transport of xHE-AAC

2.3.1 General

With the introduction of xHE-AAC into DRM the transport scheme was revised in order to allow transmission of AUs of varying lengths as described above and to be able to make use of the full potential of the bit reservoir.

Previous transport schemes allowed a varying AU size only within one audio super frame (ASF), but this did not allow for fully flexible bit distribution between frames which led to less than optimal coding efficiency.

The transport scheme employed for xHE-AAC in DRM allows the codec to make full use of the bit reservoir. This is achieved by decoupling the AU frame grid from the ASF grid. In other words, the codec runs fully asynchronous to the underlying transport mechanism.

Each ASF consists of three parts:

- Header
- Payload
- AU directory

The ASF starts with a 2-byte header which contains the following essential information:

Audio Super Frame Header Structure

Frame border count (*b*), 4 bits	Bit reservoir level, 4 bits
Fixed header CRC, 8 bits	

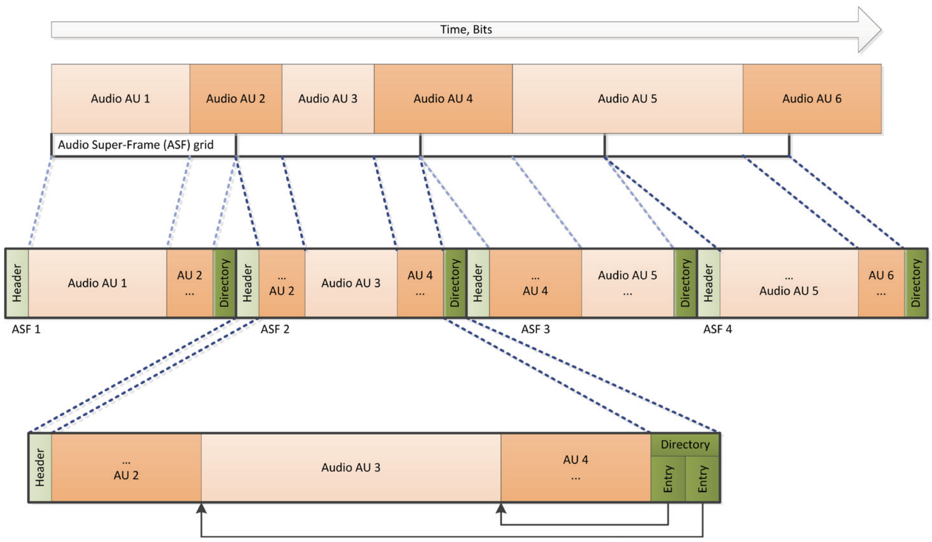
The Frame border count, *b*, indicates to the decoder how many AUs start in a given ASF. This number hence also determines the size of the AU directory which resides at the end of the ASF.

AU Directory Structure

b times	Frame border index, 12 bits	Frame border count (*b*), 4 bits
	Frame border index, 12 bits	Frame border count (*b*), 4 bits
	Frame border index, 12 bits	Frame border count (*b*), 4 bits
	…	…

The Frame border count, *b*, is repeated for reasons of error robustness.

The Frame border index array gives for each AU the 0-based index position of the first byte of the corresponding audio frame in the Payload section of the audio super frame, counted from the beginning of the Payload section.



Packing of Audio AUs into the DRM Audio Super Frame (ASF) Transport

2.3.2 Restrictions of the DRM transport system for xHE-AAC payload

Given the ASF header structure described in 2.3.1 with the 4 bit field „Frame border count", it can be deduced that the directory structure can contain a maximum number of 15 entries to indicate 15 AUs starting in a given ASF. This maximum number needs to be taken into account during the encoding process, because in certain sampling rate / frame size combinations and when making use of the bit reservoir mechanism, the theoretical number of starting AUs per ASF can be close to 15 or more. The following paragraphs outline one approach to avoiding this scenario by intelligent monitoring of the current and recent bit usage.

3. TECHNICAL RECOMMENDATIONS FOR XHE-AAC ENCODER IMPLEMENTATIONS

3.1 Embedding xHE-AAC access units in DRM audio super frames

As mentioned in 2.3.2 the number of AU directory entries within an ASF is limited to 15. For certain encoder configurations this limit can be exceeded.

This situation can occur in two cases:

1. The average number of AU per ASF is larger than 15.
2. Within the size variation limits allowed due to the bit reservoir described in 2.2.4 a series of small sized AU can be written. With respect to the ASF this may result in more than 15 AUs in an ASF.

3.1.1 Case 1: Average number AUs per ASF > 15

This is the trivial case in which the (long-term) average number of AUs per ASF is greater than 15. The encoder must reject the configuration as invalid. The average number of AUs per ASF can be calculated simply by dividing the duration of an ASF by the duration of the audio that is represented by one AU, i.e. the duration of one audio granule:

`avgAUsPerASF`	average number of access units per audio super frame
`ASFDuration`	duration of an audio super frame in seconds
`granuleLength`	number of audio samples in one audio processing frame
`sampleRate`	audio sampling rate in Hz (audio samples per second)

The average number of AUs per ASF is then calculated as:

```
avgAUsPerASF  = ASFDuration / (granuleLength / sampleRate)
              = ASFDuration * sampleRate / granuleLength
```

It can be seen that for DRM+ (with 400ms ASF) large sample rates in combination with short audio processing frames may lead to a large average number of AUs per ASF.

If **`avgAUsPerASF`** > 15, then this configuration must be rejected.

3.1.2 Case 2: Natural fluctuation in AUs length

If the average number of AUs is less than but close to 15 then it is possible that due to the momentary bit rate variability introduced by the bit reservoir, a naïve encoding algorithm may produce more than 15 AUs before it reached the ASF size required. For more information please see the section on ‚Dynamic Bit Allocation and the „Bit Reservoir“‘.

In order to prevent this from happening the encoder can pad the 15th AU with so-called „fill bits“ in order to reach the ASF size requirement.

For the vast majority of encoder configurations the upper limit of 15 AU directories per ASF is never exceeded. The mechanism described here is intended to guarantee fail safe operation even in abnormally configured DRM systems.

The approach requires a frame-wise check during the encoding process to see whether 14 AU directories have already been produced for the currently encoded ASF. At that point the encoder can calculate a minimum AU size for the 15th AU.

`minimumAuSize`	minimum size of the 15th access unit within an audio super frame in number of bits
`superFrameSize`	size of an audio super frame in number of bits
`ASFHeaderSize`	size of the audio super frame header (= 16 bits)
`AuPayload`	accumulated size of all AU content already carried in the payload section of this ASF - including any AU data left over from the previous ASF
`DirEntrySize`	size of one entry in the AU directory structure of an ASF (= 16 bits)
`CrcHeaderSize`	size of one AU-CRC which follows each access unit in the audio super frame (= 16 bits)

The minimum AU size is determined according to the following equation:

```
minimumAuSize =        superFrameSize
                       - ASFHeaderSize
                       - AuPayload
                       - (15 * DirEntrySize)
                       - (15 * CrcHeaderSize)
```

A suitable encoder should then take this minimum AU size into account. If towards the end of the encoding process of the currently encoded audio frame the minimum AU size requirement is about to be violated, the encoder can add fill bits to the AU.

3.2 Ensuring high quality for sequences of strongly tonal audio

Particularly tonal signal components have shown to be an especially difficult kind of signal to encode. This is particularly true when employing the ACELP coding tool typically used at very low bit rates in xHE-AAC.

In order to ensure a consistent high quality of tonal signals too it is advised to take certain precautions at the encoder.

At low bit rates xHE-AAC typically operates in the so-called linear prediction domain (LPD) mode. The LPD makes use of two different signal representation methods. These are Algebraic Code Excited Linear Prediction (ACELP) and transform coded excitation (TCX). Further, the ACELP is using a long-term prediction (LTP) method which reuses previously reconstructed samples to calculate the following frame. This results in the situation where ACELP is computing the signal including quantization errors, which may be carried over to the next frame. During usual operation the previous computed samples are incorporated with a pitch gain factor less than 1.0, so the error decreases over time. In the very unusual case that the gain factor is larger than 1.0 over a longer period of time the resulting signal may become instable, which would have a negative impact on the quality, possibly leading to signal saturation.

Experience shows that this issue only appears for synthetic, strongly tonal audio signals (e.g. a pure sine wave).

The straightforward solution is to ensure that the encoder frequently employs TCX coding mode, in particular, if there is an increased risk that this issue may arise. Indicators are, for example, a high pitch gain factor over a longer time period.

4. REFERENCES

[1] ISO/IEC 23003-3, MPEG-D Part3, Unified Speech and Audio Coding

[2] ETSI ES 201 980 V4.1.1, Digital Radio Mondiale (DRM); System Specification

[3] DRM Introduction and Implementation Guide, http://www.drm.org/

[4] ISO/IEC 14496-3, MPEG-4 Part 3, Audio

[5] Fraunhofer IIS. (2014, August), Broadcasting – Fraunhofer IIS. [Online]. http://www.iis.fraunhofer.de/en/bf/amm/produkte/digirundfunk.html

5. GLOSSARY

AAC	Advanced Audio Coding
ACELP	Algebraic code exited linear prediction (coding)
AM	Amplitude Modulation
ASF	Audio Super Frame
AU	Access Unit
CRC	Cyclic Redundancy Check
DRM	Digital Radio Mondiale (Not to be confused with Digital Rights Management)
ETSI	European Telecommunications Standards Institute
FAC	Fast Access Channel
FM	Frequency Modulation
ISO	International Organization for Standardization
LPD	Linear Prediction Domain
MPEG	Moving Picture Experts Group
MSC	Main Service Channel
PS	Parametric Stereo
SBR	Spectral Band Replication
SDC	Service Description Channel
TCX	Transform Coded Excitation (coding)
USAC	Unified Speech and Audio Coding
VHF	Very High Frequency (Band)
xHE-AAC	Extended High-Efficiency Advanced Audio Coding

ABOUT FRAUNHOFER IIS

When it comes to advanced audio technologies for the rapidly evolving media world, Fraunhofer IIS stands alone. For more than 25 years, digital audio technology has been the principle focus of the Audio and Multimedia division of Fraunhofer Institute for Integrated Circuits (IIS). From the creation of mp3 and the co-development of AAC to the future of audio entertainment for broadcast, Fraunhofer IIS brings innovations in sound to reality. Today, technologies such as Fraunhofer Cingo for virtual surround sound, Fraunhofer Symphoria for automotive 3D audio, AAC-ELD for telephone calls with CD-like audio quality, and Dialogue Enhancement that allows television viewers to adjust dialogue volume to suit their personal preferences are among the division's most compelling new developments.

Fraunhofer IIS technologies enable more than eight billion devices worldwide. The audio codec software and application-specific customizations are licensed to more than 1,000 companies. The division's mp3 and AAC audio codecs are now ubiquitous in mobile multimedia systems.

Fraunhofer IIS is based in Erlangen, Germany and is an institute of Fraunhofer-Gesellschaft. With 23,000 employees worldwide, Fraunhofer-Gesellschaft is comprised of 67 institutes making it Europe's largest research organization.

For more information, contact Matthias Rose, matthias.rose@iis.fraunhofer.de, or visit www.iis.fraunhofer.de/audio.