INTERNATIONAL ORGANISATION FOR STANDARDISATION

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11/N18650**

**July 2019, Gothenburg, CH**

**Title: White paper on MPEG AAC Family of Codecs**

**Source Communication**

**Status: Approved**

# Introduction

The MPEG-4 Audio standard first issued in 1999, ISO/IEC 14496-3:1999 Audio, and it is anticipated that it will be re-issued in a Fifth Edition in 2019. The flagship technology in MPEG‑4 Audio is the Advanced Audio Coding (AAC) family of standards, which has enjoyed extremely widespread deployment (i.e. on more than 10 billion devices).

One reason that the AAC Family of codecs in the MPEG-4 standard is so important in industry is that it has a very wide performance range. At higher bitrates (320 kb/s for a 5.1 channel signal, or 128 kb/s for a stereo signal), it can code all types of signals without any perceivable signal degradation (excellent audio quality). At much lower bitrates (32 kb/s for a stereo signal), it can code music signals with very good audio quality. When extended using the newest MPEG technology, Unified Speech and Audio Coding, it can code any type of signal (i.e. both music and speech) at very low bitrates (24 kb/s for a stereo signal) with very good audio quality. Hence, the AAC family of technology provides every user with a codec for their application’s operating point. This wide operating range and excellent performance across the range is the reason that MPEG‑4 Audio has remained relevant 20 years after it was first issued.

The remainder of this document presents a guide to where the component technologies of the AAC Family of Codecs can be found in the MPEG-4 and MPEG-D standards and how they are used together to build the specific codecs in the AAC Family. To make the guide as unambiguous as possible, reference is made to specific tables and clauses. However, the reader must be aware that MPEG-4 Fifth Edition is still in the final editing process and table and even clause numbering may change.

MPEG‑4 Audio specifies a large number of technologies and taken all together, they may not be relevant to a particular industry segment. Hence, MPEG specifies sub-sets of technologies via *profiles*. An MPEG specification defines a compressed representation, e.g. a bitstream, and also a decoding process, e.g. a decoder. An MPEG profile is used to specify a bitstream as “containing not more than the syntax elements in the profile” and a decoder as “able to decode at least the syntax elements in the profile.” The AAC family of technology is defined by the AAC family of profiles:

* AAC Profile (AAC LC and AAC)
* High Efficiency AAC Profile (HE-AAC)
* High Efficiency AAC v2 Profile (HE-AAC v2)
* Extended High Efficiency AAC Profile (xHE-AAC)
* Low Delay AAC Profile (AAC-LD)
* Low Delay AAC v2 Profile (AAC-ELD)
* High Definition AAC Profile (HD-AAC)

Figure 1, below, shows the AAC family of codecs and how each builds upon the other. USAC, a component of the xHE-AAC Profile is not shown, but will be discussed later in this document.



**Key**

LC Low Complexity

PNS Perceptual Noise Substitution

SBR Spectral Band Replication

PS Parametric Stereo

LD-MPS Low Delay MPEG Surround

SLS Scalable to Lossless Coding

LD Low Delay

Figure 1-- Overview of MPEG AAC audio coding family

Figure 2, below, shows how the AAC, HE-AAC, HE-AAC v2 and xHE-AAC Profiles are nested, with each enclosing profile decoder able to decode all enclosed profile bitstreams.



Figure 2 -- Nested structure of MPEG AAC Profiles

A more technology-oriented overview of the AAC family of codecs can be found in [9]. The MPEG Reference Software which includes all code for the AAC family of codecs can be found at [3] and as a part of ISO/IEC 23003-3:2012, Unified Speech and Audio Coding Second Edition.

# Advanced Audio Coding and AAC Profile

The MPEG-2 Part 7, Advanced Audio Coding (AAC) standard issued in 1997 (ISO/IEC 13818‑7:1997, Advanced Audio Coding) and has three profiles, of which the Low Complexity (AAC LC) Profile was most widely deployed [1]. This was incorporated into MPEG-4 Audio in its first edition (ISO/IEC 14496-3:1999) with an additional coding tool, Perceptual Noise Substitution (PNS), but in every other way it retained compatibility with MPEG-2 AAC LC profile. AAC is able to provide perceptually transparent audio quality at 128 kbit/s for a stereo signal, or 320 kbit/s for a 5.1 channel signal (i.e. as found in digital television).

MPEG-4 AAC Profile is specified in MPEG-4 Audio ISO/IEC 14496-3 Table 1.3 — “Audio Profiles Definition,” and contains “AAC LC” which is Audio Object Type ID 2. The components of AAC LC are specified in ISO/IEC 14496-3 Table 1.1 — “Audio Object Type definition based on Tools/Modules,” which includes all MPEG-2 AAC LC Profile tools plus PNS. MPEG profiles have levels, which specify varying capability and complexity. For AAC Profile, these are given in ISO/IEC 14496-3 Table 1.10 — “Levels for the AAC Profile.”

AAC is specified in ISO/IEC 14496-3 clause 4, “General Audio Coding (GA) – AAC, TwinVQ, BSAC,” where AAC is a subset of MPEG‑4 General Audio coding. The AAC bitstream syntax is structured as configuration elements and elements that encode a specific time interval of audio (i.e. an audio “block”). The top-level configuration element is specified in ISO/IEC 14496-3 clause 4.4.1, “Decoder configuration (GASpecificConfig),” while the top-level audio block element is specified in ISO/IEC 14496-3 clause 4.4.2, “GA bitstream payloads.”

# Transport of MPEG-4 Audio streams

AAC Profile bitstreams may be stored in an MPEG-4 File (ISO/IEC 14496-12:2015 ISO base media file format and ISO/IEC 14496-14:2018 MP4 file format). In this case, the AAC bitstream information is supported by additional MPEG-4 Systems information, which is described in ISO/IEC 14496-3 subclause 1.6, “Interface to ISO/IEC 14496-1 (MPEG-4 Systems).” Specifically, the AAC configuration information is carried in an AudioSpecificConfig(), as shown in ISO/IEC 14496-3 Table 1.19 —"Syntax of AudioSpecificConfig().” The Profile and associate Level of a bitstream is signaled as in ISO/IEC 14496-3 Table 1.17 – “audioProfileLevelIndication values.” Example MPEG-4 reference software for carrying AAC Profile bitstreams in MPEG-4 File Format can be found in the “audio\_example” folder in the MPEG-4 File Format reference software distribution (at https://github.com/MPEGGroup/isobmff in IsoLib directory) [2].

Alternatively, an AAC Profile bitstream can be carried in a “stand-alone” multiplex, as specified in ISO/IEC 14496-3 clause 1.7, “MPEG-4 Audio transport stream,” which defines mechanisms to transport MPEG‑4 Audio streams without using ISO/IEC 14496-1 (MPEG-4 Systems) e.g. for use in audio-only applications. Two such mechanisms are specified:

* Low Overhead Audio Transport Multiplex (LATM) is used when the transmission channel provides frame synchronization.
* Low Overhead Audio Stream (LOAS) is used when the transmission channel does not provide frame synchronization.

This is illustrated in Figure 1.3, “Concept of MPEG-4 Audio Transport,” found in ISO/IEC 14496-3 subclause 1.7.1.

Each coded block, or “access unit” in MPEG-4 Systems terminology, of AAC is independently decodable, and so is a random-access point as long as the required configuration information has been received. When encapsulated in the LATM or LOAS transport formats, every random access point must have available the StreamMuxConfig() information, which carries essential, but quasi-static, information such as sampling rate and channel configuration. Periodic transmission of this information allows for random access to the stream. StreamMuxConfig(), which carries AudioSpecificConfig(), must be present at a random access point.

# High-Efficiency Audio Coding (HE-AAC) and HE-AAC Profile

The MPEG-4 High Efficiency Advanced Audio Coding (HE-AAC) technology was standardized in 2003, as amendment 1 to MPEG-4 (ISO/IEC 14496-3:2001/Amd 1:2003 Bandwidth extension). HE-AAC was built upon AAC, in that it has AAC as a core compression engine and has the bandwidth extension or spectral bandwidth replication (SBR) tool as a pre- (encoding) and post- (decoding) processor [4]. Addition of the SBR tool enabled transmission of a stereo signal at 48 kbit/s with very good audio quality. Since AAC was already widely deployed, this permitted extending this base to HE-AAC by only adding the SBR tool to existing AAC implementations.

HE-AAC is typically deployed as the HE-AAC Profile, which is specified in ISO/IEC 14496-3 Table 1.3 — “Audio Profiles Definition,” as “High Efficiency AAC Profile”, which contains Audio Object Type ID 2 (AAC LC) and ID 5 (SBR tool). The High Efficiency AAC Profile is a superset of the AAC Profile. Levels of HE-AAC Profile, are given in ISO/IEC 14496-3 Table 1.11 — “Levels for the High Efficiency AAC Profile.”

HE-AAC is specified in ISO/IEC 14496-3 subclause 4.6.18, “SBR tool.” The HE-AAC bitstream syntax is a backward-compatible extension of the AAC bitstream: AAC Profile decoders can decode a HE-AAC Profile bitstream, but will produce only a half-bandwidth output signal. The additional HE-AAC Profile information needed to decode the full-bandwidth signal is stored in the AAC Fill\_element() (ISO/IEC 14496-3 Table 4.13 — “Syntax of fill\_element()”) via the “extension\_payload()” mechanism whose syntax is in ISO/IEC 14496-3 Table 4.59 — “Syntax of extension\_payload().” The extension is sbr\_extension\_data() whose syntax is given in ISO/IEC 14496-3 subclause 4.4.2.8, “Payloads for the audio object type SBR” and ISO/IEC 14496-3 Table 4.67 — “Syntax of sbr\_extension\_data().” From this table it can be seen that required HE-AAC data is in sbr\_header() and sbr\_data(), in ISO/IEC 14496-3 Table 4.68 – “Syntax of sbr\_header()” and ISO/IEC 14496-3 Table 4.69 – “Syntax of sbr\_data().”

To allow for a full decoder setup at a random access point a sbr\_header() must be present. Periodic transmission of this information allows for random access to the stream.

# Parametric Stereo and HE-AAC v2 Profile

MPEG standardized the Parametric Coding tools in 2004 as amendment 2 to MPEG-4 (ISO/IEC 14496-3:2001/Amd 2:2004 Parametric coding for high-quality audio). While the core parametric coder did not enjoy wide adoption, it had a *Parametric Stereo* (PS) tool that permitted very efficient coding of a stereo music signal as a coded mono signal plus a small amount of side-information [5]. A HE-AAC v2 Profile bitstream is a superset of a HE-AAC Profile bitstream, which is a superset of an AAC Profile bistream. Hence, a HE-AAC v2 decoder can decode AAC LC, HE-AAC and HE-AAC v2 bit streams. Because of this backwards compatibility, the HE-AAC v2 decoder is perhaps the most widely deployed MPEG Audio decoder.

The combination of HE-AAC and PS enabled transmission of a stereo signal at 32 kbit/s with very good audio quality. The PS tool was combined with HE-AAC in the MPEG‑4 HE-AAC v2 Profile, which contains Audio Object Type IDs 2 (AAC LC), 5 (SBR), and 29 (PS). Levels of HE-AAC v2 Profile, are given in ISO/IEC 14496-3 Table 1.12 — “Levels for the High Efficiency AAC v2 Profile.”

The Parametric Stereo tool is specified in ISO/IEC 14496-3 subclause 8.6.5, “Parametric stereo.” The specification of how to use the parametric stereo tool with the HE-AAC decoder, including additional syntax for “ps\_data()” is given in ISO/IEC 14496-3 Annex 8A, “Combination of the SBR tool with the parametric stereo tool and SBR Enhancements.”

To allow for a full decoder setup at a random access point the enable\_ps\_header bit in ps\_data() must be set. Periodic transmission of the PS configuration information synchronized with sbr\_header() allows for random access to the stream.

# Low Delay AAC and Low Delay AAC Profile

In 2005 MPEG standardized in MPEG-4 amendment 1 a low delay version of AAC, called Low Delay AAC or AAC-LD. This specified an alternate transform for AAC that has much lower latency [6]. While a typical AAC encoder/decoder has a one-way latency of perhaps 55 ms (transform delay plus look-ahead processing), AAC-LD achieves a one-way latency of only 20 ms by eliminating “block switching” (and the associated look-ahead processing) and using an innovative low-delay transform. This lower transmission latency permits AAC-LD to be used as a conversational codec, but with a signal bandwidth and perceived quality of a music coder – AAC-LD provides excellent audio quality at 64 kbit/s for a mono signal.

The Low Delay AAC Profile contains Audio Object Type ID 23, Error Resilient Low Delay AAC (ER AAC-LD). The top-level payload syntax is specified in ISO/IEC 14496-3 subclause 4.4.2.9, “Payloads for the audio object type ER AAC ELD” and decoding semantics in ISO/IEC 14496-3 subclause 4.5.2.4, “Payloads for the audio object types ER AAC LC, ER AAC LTP, ER AAC LD, ER AAC ELD and ER AAC scalable.” Levels of Low Delay AAC Profile are given in ISO/IEC 14496-3 Table 1.13 – “Levels for the Low Delay AAC Profile.”

# Enhanced Low Delay AAC and Low Delay AAC v2 Profile

In 2008 MPEG standardized in MPEG-4 amendment 9 the Enhanced Low Delay AAC (AAC-ELD) technology (ISO/IEC 14496-3:2005/Amd 9:2008 Enhanced low delay AAC). AAC-ELD is the core of the Low Delay AAC v2 Profile, which contains Audio Object Type IDs 23 (ER AAC LD), 39 (ER AAC ELD) and 44 (LD MPEG Surround). Here the LD (low delay) MPEG Surround tool provides a function similar to that of the PS tool. The codec provides even greater signal compression than the Low Delay AAC Profile with only a modest increase in latency: it provides excellent audio quality at 48 kbit/s for a mono signal with a one-way latency of only 32 ms. The top-level payload syntax is specified in ISO/IEC 14496-3 subclause 4.4.2.9, “Payloads for the audio object type ER AAC ELD” and the decoding process is specified in ISO/IEC 14496-3 subclause 4.5.2.4, “Payloads for the audio object types ER AAC LC, ER AAC LTP, ER AAC LD, ER AAC ELD and ER AAC scalable.” Levels of Low Delay AAC v2 Profile are given in ISO/IEC 14496-3 Table  1.16 — “Levels for the Low Delay AAC v2 profile.”

# Scalable to Lossless Audio Coding and High Definition AAC Profile

In 2005 MPEG standardized two algorithms for lossless compression of audio, MPEG Audio LosslesS coding (ALS) and Scalable to LosslesS coding (SLS) (ISO/IEC 14496-3:2005/Amd 2:2006 Audio Lossless Coding and ISO/IEC 14496-3:2005/Amd 3:2006 Scalable Lossless Coding). Both provide perfect (i.e. lossless) reconstruction of a standard Compact Disk audio signal with a compression ratio approximately 2:1. However, the SLS codec is unique in that it is built on a variant of AAC that has a bit-exact transform [7]. This permits it to have a variable compression ratio: it can operate similar to AAC and compress a stereo signal to 128 kbit/s (11:1 compression ratio) with excellent quality, and it can increase the coded bitrate (decrease the compression ratio) in a continuous fashion up to lossless reconstruction with a compression ratio of 2:1. The High Definition AAC Profile contains the AAC LC and SLS Audio Objects Types. The SLS tool is specified in ISO/IEC 14496-3 Clause 12, “Scalable lossless coding,” with syntax in ISO/IEC 14496-3 subclause 12.3, “Payloads for the audio object” and decoding semantics in ISO/IEC 14496-3 subclause 12.4, “Semantics.”

# xHE-AAC Profile

In 2012 MPEG standardized Unified Speech and Audio Coding (USAC) (ISO/IEC 23003-3:2012), and issued a Second Edition in 2019. As its name implies, USAC was designed to provide the high performance for both speech and audio (music) signals in one (unified) codec. Service providers do not have to consider what types of content is being compressed for transmission, as USAC provides excellent performance for all content types. In addition, the Extended HE-AAC Profile (xHE-AAC) codec supports an extremely wide bit rate range: starting as low as 12 kbit/s for mono and 24 kbit/s for stereo services, up to 128 kbit/s for stereo and 320 kbit/s for 5.1 channel services. Equally important, the xHE-AAC Profile is built for streaming services and provides reliable service even under the most challenging network conditions. It supports both seamless bit rate switching and random-access points that are fully compatible with MPEG DASH.

The xHE-AAC Profile is specified in ISO/IEC 23003-3, clause 4.5.4 “Extended High Efficiency AAC Profile,” and contains the audio object types 42 (USAC), 5 (SBR), 29 (PS) and 2 (AAC LC) as defined in ISO/IEC 14496-3. The audio object type 42 (USAC) conforms to the Baseline USAC Profile, which is defined in ISO/IEC 23003-3. The components of xHE-AAC Profile are shown in ISO/IEC 23003-3, Table 1 “Summary of the Location of and Normative Reference to the Definitions of all AAC, HE-AAC and USAC Coding Tools as employed in the Extended High Efficiency AAC profile.” The levels for xHE-AAC Profile are show in ISO/IEC 23003-3 Table 3 “Levels for the Extended HE AAC profile.” The USAC technology is specified in ISO/IEC 23003-3.

# References

1. Bosi, Marina; Brandenburg, Karlheinz; Quackenbush, Schuyler; Fielder, Louis; Akagiri, Kenzo; Fuchs, Hendrik; Dietz, Martin; Herre, Jürgen; Davidson, Grant; Oikawa, Yoshiaki, “ISO/IEC MPEG-2 Advanced Audio Coding,” JAES Volume 45 Issue 10 pp. 789-814; October 1997.

1. <https://github.com/MPEGGroup/isobmff> (MPEG-4 File Format Reference Software)
2. <https://wiki.multimedia.cx/index.php/AAC_Reference_Software>
3. Wolters, Martin; Kjorling, Kristofer; Homm, Daniel; Purnhagen, Heiko, “A Closer Look into MPEG-4 High Efficiency AAC,” 115th AES Convention, October 2003, paper 5871.
4. Purnhagen, Heiko; Engdegard, Jonas; Roden, Jonas; Liljeryd, Lars, “Synthetic Ambience in Parametric Stereo Coding,” 116th AES Convention, May 2004, paper 6074.
5. Schnell, Markus; Schmidt, Markus; Jander, Manuel; Albert, Tobias; Geiger, Ralf; Ruoppila, Vesa; Ekstrand, Per; Bernhard, Grill, “MPEG-4 Enhanced Low Delay AAC - A New Standard for High Quality Communication,” 125th AES Convention, October, 2008, paper 7503.
6. Geiger, Ralf; Yu, Rongshan; Herre, Jürgen; Rahardja, Susanto; Kim, Sang-wook; Lin, Xiao; Schmidt, Markus, “ISO/IEC MPEG-4 High-Definition Scalable Advanced Audio Coding,” JAES Volume 55 Issue 1/2 pp. 27-43; February 2007.
7. Neuendorf, Max; Multrus, Markus; Rettelbach, Nikolaus; Fuchs, Guillaume; Robilliard, Julien; Lecomte, Jérémie; Wilde, Stephan; Bayer, Stefan; Disch, Sascha; Helmrich, Christian; Lefebvre, Roch; Gournay, Philippe; Bessette, Bruno; Lapierre, Jimmy; Kjörling, Kristofer; Purnhagen, Heiko; Villemoes, Lars; Oomen, Werner; Schuijers, Erik; Kikuiri, Kei; Chinen, Toru; Norimatsu, Takeshi; Chong, Kok Seng; Oh, Eunmi; Kim, Miyoung; Quackenbush, Schuyler; Grill, Bernhard, “The ISO/MPEG Unified Speech and Audio Coding Standard—Consistent High Quality for All Content Types and at All Bit Rates,” JAES Volume 61 Issue 12 pp. 956-997; December 2013.

1. <https://www.iis.fraunhofer.de/content/dam/iis/de/doc/ame/wp/FraunhoferIIS_White-Paper_AAC-Broadcast-CableTV.pdf>