



Audio Engineering Society Convention Paper

Presented at the 115th Convention
2003 October 10–13 New York, NY, USA

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

A closer look into MPEG-4 High Efficiency AAC

Martin Wolters¹, Kristofer Kjörling², Daniel Himm¹, Heiko Purnhagen²

¹*Coding Technologies, Deutschherrnstr. 15–19, 90429 Nürnberg, Germany*

²*Coding Technologies, Döbelnsgatan 64, 11352 Stockholm, Sweden*

Correspondence should be addressed to Martin Wolters (wol@codingtechnologies.com)

ABSTRACT

MPEG Spectral Band Replication (SBR) is the newest compression technology available as part of the MPEG standards. It is combined with MPEG Advanced Audio Coding (AAC) and improves coding efficiency by more than 30%. The resulting scheme is called High-Efficiency AAC (HE AAC).

This presentation explains MPEG SBR and its integration into the existing MPEG-4 bitstream format. The SBR technology itself as well as the implications on systems based on MPEG-4 technology are described. Signaling through MPEG-4 Systems and other transport formats is introduced and typical applications and usage scenarios are listed.

1 INTRODUCTION

MPEG-4 High Efficiency AAC (HE AAC) is not a replacement for AAC, but rather a superset which extends the reach of high-quality MPEG-4 Audio to much lower bitrates. High Efficiency AAC decoders will decode both plain AAC and the enhanced AAC plus SBR. The result

is a backward compatible extension of the standard. Listening tests show that MPEG-4 HE AAC offers a significant benefit over proprietary codecs and over AAC without extensions, and place it clearly as the most efficient audio codec in existence. Coding Technologies and its customers have been using this technology for a number

of years under the *aacPlus* brand name.

The technical specifications of HE AAC are now available as a final draft amendment (FDAM) [1] and official standardization is expected in September 2003. These specifications enable manufacturers to work on final implementations and integrations. This paper provides an implementation guideline and provides an overview of the technologies and system integration aspects of HE AAC.

2 INTRODUCTION INTO MPEG-4

2.1 MPEG-4: A leading multimedia framework

In the context of evolving multimedia applications, new demands for efficient and flexible representation of audiovisual content have arisen. Besides high coding efficiency required to cope with the limited bandwidth of the internet or in mobile communication, new functionality like flexible access to coded data and manipulation by the recipient is also desired. To address these requirements and develop inter-operable solutions, ISO/IEC started its MPEG-4 standardization activities “Coding of audiovisual objects” [2], [3], [4]. Part 3 of the MPEG-4 standard (Audio) [5] provides tools for coding of natural and synthetic audio objects and composition of such objects into an “audio scene.” Similarly, Part 2 (Visual) [6] provides tools for coding of natural video and synthetic 2D and 3D objects, while Part 1 (Systems) [7] covers general tools for scene description, synchronization of objects, and multiplexing for transport and storage.

2.2 MPEG-4 Audio Coding

MPEG-4 is based on the notion that the audio part of the audiovisual scene presented at the receiver is composed of one or more so-called *audio objects*. Different audio compression tools (codecs) are available to enable an efficiently coded representation of the audio object(s) in a scene. Natural audio objects, such as recorded speech and music, can be coded at bitrates typically ranging from 2 kbit/s (for narrowband speech) to 64 kbit/s/ch (for CD quality music) using parametric speech coding (HVXC), CELP-based speech coding, parametric audio coding (HILN) or transform-based general audio coding (AAC, TwinVQ). The natural audio and speech coding tools support bitrate scalability, also known as embedded coding. In addition, the parametric coding tools also provide speed and pitch change functionality in the decoder. Synthetic audio objects can be represented using

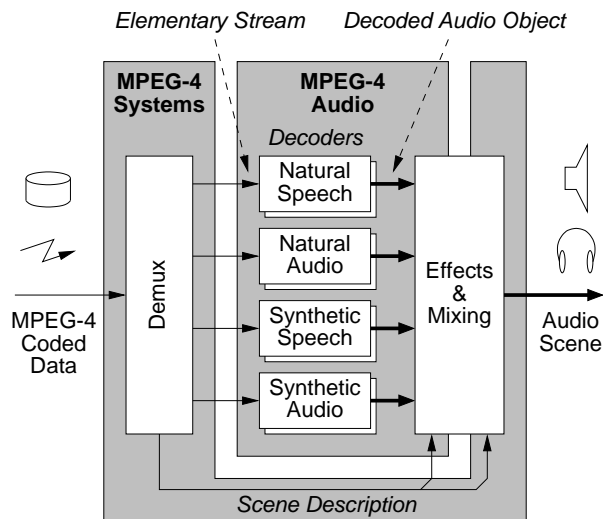


Figure 1: Block diagram of a complete MPEG-4 Audio decoder.

a Text-To-Speech Interface (TTSI) or the Structured Audio (SA) synthesis tools. Other uses of the SA tools are adding effects, like reverberation, and mixing different audio objects to compose the final “audio scene” that is presented to the listener.

Fig. 1 shows the block diagram of a complete MPEG-4 Audio decoder. It includes the decoding tools for the audio objects defined in Part 3 (Audio) of the MPEG-4 Standard [5] as well as bitstream demultiplexing and scene composition defined in the Systems part [7].

The information needed to decode an audio object at the decoder is conveyed by means of a so-called *elementary stream* (ES), as shown in Fig. 2. In case of bitrate scalable configurations, a base-layer ES and one or more enhancement-layers ES(s) are used. The initial ES_Descriptor contains an AudioSpecificConfig element, which carries the *audio object type* (AOT) and further information required to instantiate the required audio decoder. Using the ES_ID, the ES_Descriptor points to the stream of the actual compressed audio data, which is a sequence of so-called *access units* (AU), i.e., decodable bitstream frames.

A common audio object type is AAC LC, which indicates that MPEG-4 Advanced Audio Coding (AAC) in its low-complexity (LC) version is used. This is the primary general audio coder in MPEG-4, which is a compatible extension of the MPEG-2 AAC transform-based audio coder [8]. The principles of perceptual audio cod-

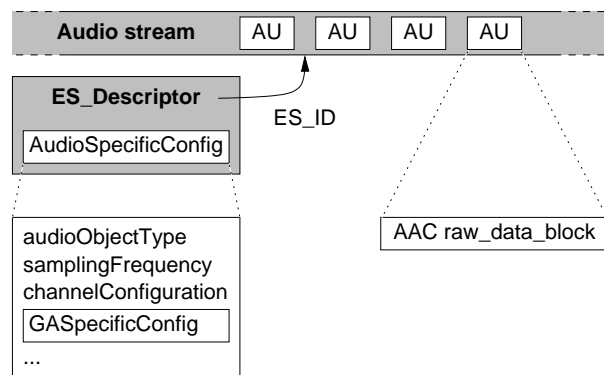


Figure 2: MPEG-4 Elementary Stream (ES) conveying an audio object.

ing and the specifics of the AAC are well documented in the literature, e.g. [9], [10], [11], [12], and hence not reviewed in this presentation.

2.3 Storage and streaming of MPEG-4

In order to allow flexible yet interoperable utilization of MPEG-4 coding technology for the largest possible range of application scenarios, MPEG-4 is designed as a transport-agnostic standard with an abstract interface to carry the compressed data, the so-called *sync layer* (SL). Basically, the SL adds mechanisms like time stamps for AUs to allow synchronization of several ESs, such as an audio and a video stream.

Depending upon the requirements of the application and the characteristics of the transmission channel, different schemes can be used to store or convey MPEG-4 content (Fig. 3). MPEG defined the MP4 file format [7], [13] for storage, which can hold all data available at the SL interface. When MPEG-4 content is carried over a transmission channel, various techniques for multiplexing, synchronization, and for establishing a connection can be appropriate. For inherently packetized channels, like those using the internet protocol (IP), no syncwords are needed and AUs can often directly be mapped to packets. A fixed framing might be available from the modulation scheme for radio channels, but random bit errors are possible. Serial channels might need some syncword mechanisms to establish basic synchronization. If the MPEG-4 content is carried in more than one ES, these streams need to be multiplexed, which can be done by the MPEG-4 FlexMux tool [7], [13].

To carry simple, audio-only MPEG-4 content with low

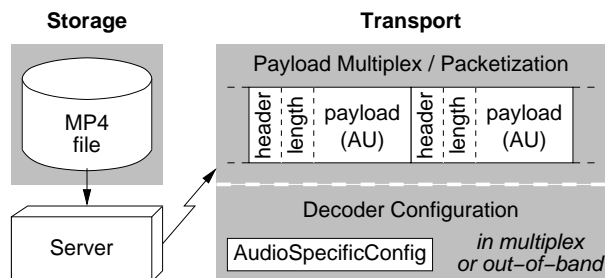


Figure 3: Storage and transmission of MPEG-4 Audio object with single elementary stream (ES).

overhead, the Part 3 (Audio) of the MPEG-4 standard defines a *Low-overhead MPEG-4 Audio Transport Multiplex* (LATM) to multiplex several MPEG-4 Audio payloads and `AudioSpecificConfig()` elements [5, subclause 1.7]. Based on LATM, it also defines a self-synchronized syntax of an MPEG-4 Audio transport stream which is called *Low Overhead Audio Stream* (LOAS).

A simple scheme to carry MPEG-4 over IP based on the realtime transport protocol (RTP) [14] is defined in RFC 3016 [15]. It is a subset of the general framework defined in Part 8 (MPEG-4 over IP) of the MPEG-4 standard [16]. RFC 3016 uses LATM for audio and allows for conveying the `AudioSpecificConfig` either in-band using LATM or out-of-band, i.e., as a MIME parameter conveyed in the Session Description Protocol (SDP).

Other schemes to convey MPEG-4 content include MPEG-4 over MPEG-2 transport streams, as defined in an amendment to the MPEG-2 Systems standard, or proprietary schemes which provide an SL compatible interface.

MPEG-4 AAC LC coded data can also be stored using the *Audio Data Interchange Format* (ADIF) or transmitted using an *Audio Data Transport Stream* (ADTS), see Fig. 4. Both these formats are defined in the MPEG-2 AAC standard [8] but are only informative in MPEG-4, i.e., an MPEG-4 decoder is not required to support ADIF or ADTS.

2.4 Recent Developments: High-Efficiency Audio Coding becomes part of MPEG-4

Standardization of the technology submitted in response to the initial MPEG-4 Call for Proposals in July 1995 was carried out using the core experiment procedure and finalized in two steps: Version 1 in October 1998 [17]

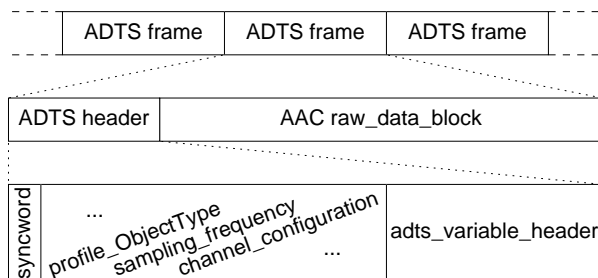


Figure 4: MPEG-2 Audio Data Transport Stream (ADTS).

and Version 2 in December 1999 [18]. In 2001, both versions and their corrigenda were merged to form a 2nd edition of the standard [5].

In July 2000, MPEG issued a Call for Evidence [19] in order to find out whether there are new developments which may provide improvements over the existing MPEG-4 standard. Based on the responses received, a Call for Proposals [20] for new tools for audio coding was issued in January 2001. It specifically asked for technology in the field of MPEG-4 compatible bandwidth extension and in the field of high quality parametric coding.

Coding Technologies submitted its Spectral Band Replication (SBR) techniques used in combination with AAC as bandwidth extension tool. Based on a listening test, it was selected as initial reference model (RM0) and refined in the course of MPEG's core experiment procedure. In March 2003, the standardization of the MPEG-4 SBR tool was finalized [1].

3 AT THE HEART OF HE AAC: MPEG SBR

3.1 Fundamental Concept: High Frequency Reconstruction by Spectral Band Replication

The basic principles of SBR have been elaborated on in several papers [21], [22], [23], [24]. For the convenience of the reader a short review is given below, and some aspects not previously discussed are explained in more detail.

3.1.1 The SBR encoder

The SBR principle stipulates that the missing high frequency region of a lowpass filtered signal can be recovered based on the existing lowpass signal and a small

amount of control data. The required control data is estimated in the encoder given the original wide-band signal. The HE AAC codec is a dual rate system, where the underlying AAC encoder/decoder is operated at half the sampling rate of the SBR encoder/decoder. The basic principle of the HE AAC encoder is depicted in Fig. 5.

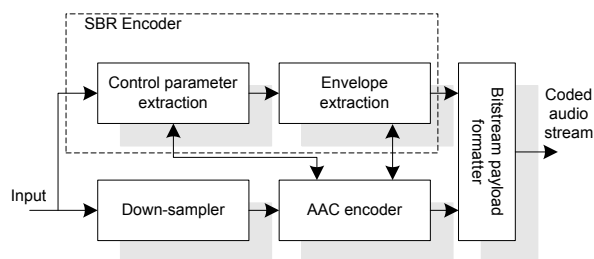


Figure 5: The HE AAC encoder.

In the SBR encoder, where the wide band signal is available, control parameters are estimated in order to ensure that the high frequency reconstruction results in a reconstructed highband that is perceptually as similar as possible to the original highband. The majority of the control data is used for a spectral envelope representation. The spectral envelope information has varying time and frequency resolution to be able to control the SBR process as good as possible, with as little bitrate overhead as possible. The other control data mainly strives to control the tonal-to-noise ratio of the highband. Fig. 6, Fig. 7, and Fig. 8, illustrate some of the characteristics of the control data.

In Fig. 6 the spectrum of the original signal at a given point in time is displayed at the top and the spectrum of the HF generated highband at the same point in time is displayed below. As is evident from the figure, the patching algorithm used to regenerate the highband was not successful in generating all the strong tonal components of the original highband. The missing sinusoids in the regenerated highband are identified in the encoder and a parametric representation of them is incorporated in the SBR control data.

In Fig. 7 the spectra of the original and the HF generated signal is once again displayed albeit at a different point in time. Here it is evident that the tonal-to-noise ratio of the regenerated highband is not similar to that of the original. The highband of the original signal constitutes mainly of noise, whilst the highband of the regenerated signal is very tonal. This information is also incorporated into the

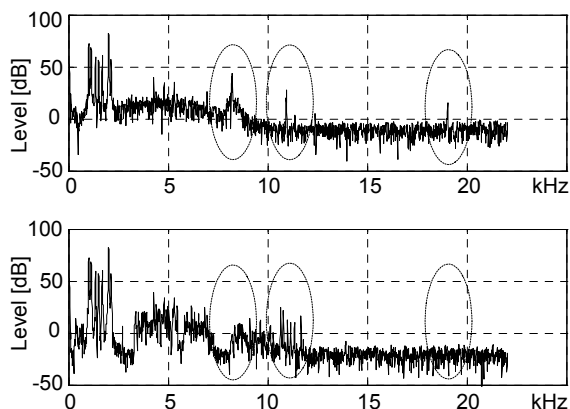


Figure 6: The spectrum of the original signal (top) and the HF generated signal without any additional signal components (bottom). The basic patching algorithm has not regenerated the strong tonal components of the original highband.

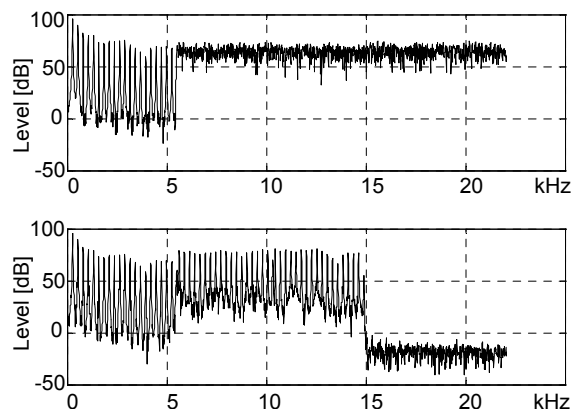


Figure 7: The spectrum of the original signal (top) and the SBR signal without any additional signal components (bottom). The basic high frequency reconstruction algorithm has not maintained the original tonal-to-noise ratio.

SBR control data so that the decoder by means of inverse filtering and noise addition can achieve a highband with a tonal-to-noise ratio similar to that of the original.

Finally, in Fig. 8, a spectrogram of the original signal is displayed with a superimposed time/frequency grid of the spectral envelope data transmitted to the decoder. Here it is evident that the time/frequency resolution of the spectral envelope varies over time, giving a higher time resolution for transient passages and a higher frequency resolution for stationary passages.

The bitrate of the control data varies depending on encoder tuning, but is in general somewhere in the region of 1-3 kbit/s per audio channel. This is far lower than the bitrate that would be required to code the highband with any conventional wave-form coding algorithm. The SBR data format is indicated in Fig. 13. A header flag indicates that an SBR header part is present. The SBR header part contains fundamental information such as SBR frequency range as well as control signals that do not require frequent changes. The `sbr_data` part can be subdivided into side info and raw data, where side info is defined as signals needed to decode the raw data and some decoder tuning signals. Raw data consists of Huffman coded envelope scalefactors and noise floor estimates.

3.1.2 The SBR decoder

The SBR enhanced decoder can be roughly divided into the modules depicted in Fig. 9.

All SBR processing is done in the QMF domain. Hence, the output from the underlying AAC decoder is firstly analyzed with a 32 channel QMF filterbank. Secondly, the HF generator module recreates the highband by patching QMF subbands from the existing lowband to the high band. Furthermore inverse filtering is done on a per QMF subband basis, based on the control data obtained from the bitstream. The envelope adjuster modifies the spectral envelope of the regenerated highband, and adds additional components such as noise and sinusoids, all according to the control data in the bitstream. Since all operations are done in the QMF domain the final step of the decoder is a QMF synthesis to retain a time-domain signal. Given that the QMF analysis is done on 32 QMF subbands for 1024 time-domain samples, and the high frequency reconstruction results in 64 QMF subbands upon which the synthesis is done producing 2048 time-domain samples, an up-sampling by a factor of two is obtained.

The following Fig. 10 displays the spectrum of the SBR signal at different stages in the SBR process. In the upper figure the spectrum of the signal after the high frequency reconstruction, but prior to the envelope adjustment is displayed. In the middle figure the spectrum of the output signal is displayed, and in the bottom figure the spectrum of the original signal is displayed for reference.

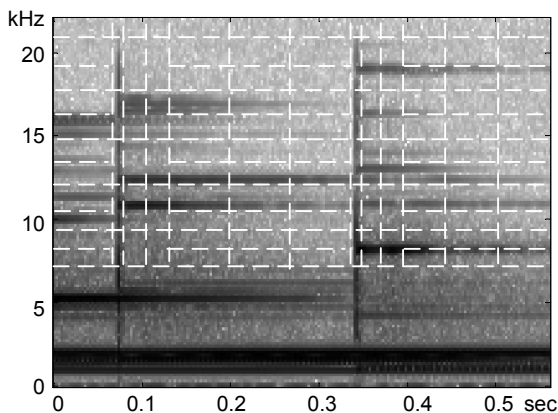


Figure 8: The spectrogram of a part of the input signal with the envelope time/frequency grid superimposed.

3.2 Low power SBR

Platforms with heavy constraints on the computational complexity such as portable devices can use Low power SBR. The main difference between High Quality SBR and Low power SBR is how the data is represented during the SBR process. High Quality SBR uses a complex representation of the subband samples and subsequently all calculations are computed with complex values. Low power SBR simply requires a real-valued representation and hence the computational complexity is heavily reduced.

The complex-valued filterbank is designed to prevent aliasing even in cases where modifications are performed [25]. However, the real-valued filterbank does not have that property. Therefore, an additional algorithm is introduced to minimize the aliasing occurring due to the real-valued filterbank. The aliasing reduction algorithm is devised to avoid introducing aliasing for strong tonal components.

The properties of the real-valued filterbank result in mirroring components (aliasing) in adjacent subbands if the adjacent subbands are given different gain-values in the envelope adjuster. This can occur if adjacent subbands are modified independently from each other. Hence, it is the objective of the aliasing reduction algorithm to identify the subbands where strong aliasing will be introduced if the subbands are modified independently. This is accomplished by observing the reflection coefficients obtained for every subband by first order linear predic-

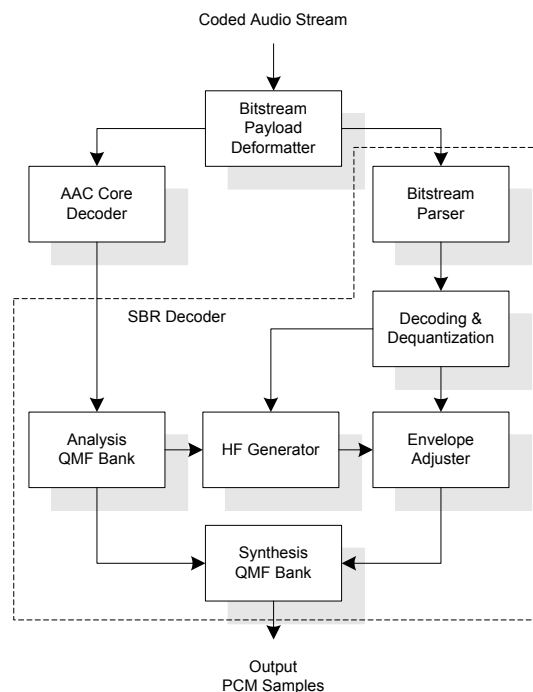


Figure 9: The HE AAC decoder.

tion. A sign is introduced for every subband indicating whether a strong tonal component is situated in the upper (a positive sign) or lower (a negative sign) part of the subband. The aliasing reduction algorithm identifies the cases where the signs of two adjacent subbands are opposite, the sign being positive for the lower subband and negative for the higher subband. These subbands are grouped together, meaning that they will not be modified independently, and hence no aliasing will be introduced by the envelope adjustment in the real-valued filterbank for this signal. However, there are many signals where there are no strong tonal components that can be identified by the above outlined algorithm, but where aliasing is introduced nevertheless. This aliasing is audible when compared to the High Quality SBR system, but not as offensive as the aliasing that would occur for a strong tonal component if the aliasing reduction algorithm was not in place.

3.3 Down-sampled SBR

Since the transform size in AAC is fixed to 1024 for long blocks and 128 for short blocks, it becomes evident that choosing the optimum sampling rate for a certain bitrate

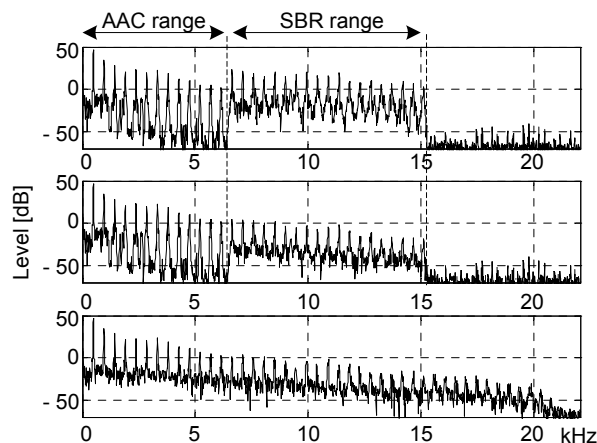


Figure 10: The spectrum of the audio signal after the HF generator (top), the spectrum of the audio signal after envelope adjustment (middle), and the spectrum of the original signal (bottom).

is crucial. In general, in the low-mid bitrate region, i.e., 20 - 32 kbit/s per audio channel, a sampling rate in the 22.05 - 32 kHz region is the best choice since the frequency resolution at the lower sampling rate improves the coding gain, compared to using a higher sampling rate. However the lower sampling rate provides a slower time response of AAC, and at higher bitrates the best trade-off between time and frequency resolution for the AAC codec may be obtained by using a higher sampling-rate. Furthermore, the low sampling rates naturally limits the highest frequency that can be covered by AAC.

The HE AAC codec is always operated in a dual rate mode. This has mainly two advantages. Firstly, as outlined above, for lower bitrates it is generally beneficial to operate the core coder at a lower sampling rate than that of the original signal. Secondly, the HE AAC decoder can always assume that if SBR data is found in the bitstream the SBR Tool should be operated at twice the sampling rate of the core coder.

There are however, scenarios where it is beneficial to have a system that, from the outside, displays the behavior of a single-rate system. If the system is operated at a higher bitrate, the core coder may be operated at the sampling rate of the original signal, e.g. 44.1 kHz. However, since the HE AAC codec is always a dual rate system, this implies that the signal has to be upsampled in the encoder prior to encoding. The result of this upsampling is that the subsequent output from the decoder will

have a sampling rate twice that of the input, in this case 88.2 kHz. In order to circumvent situations in which the output signals have a sampling rate twice that of the input, a down-sampling capability is desirable on the decoder side.

The down-sampling tool is integrated as a part of the synthesis QMF bank in the decoder. Hence, the SBR decoder operates as normal with the exception of the synthesis filterbank, which only performs a 32 subbands synthesis instead of the normal 64 subbands synthesis. The 32 subbands synthesis is done on the 32 lowest (in frequency) subbands out of the 64 available subbands, and hence a down-sampling by a factor of two is obtained, or rather the inherent upsampling by a factor of two is avoided. By using the down-sampled SBR for the above outlined use-case scenario, the output sampling rate can be assured to be the same as that of the input signal even though the AAC coder is operated at the sampling frequency of the original signal and the HE AAC codec is a dual rate system.

Furthermore, by introducing the down-sampling capability, it is possible for hand-held devices to use down-sampled SBR in order to reduce cost of D/A conversion, and also reduce computational complexity by the reduced number of synthesis subbands in the QMF synthesis. In this case the decoder might even generate an output sampling rate lower than the one of the original signal.

3.4 Test results

During the course of standardization within MPEG several listening tests have been conducted in order to assure the quality of the new algorithm. In Fig. 11 and Fig. 12 below, listening test results are displayed for SBR enhanced AAC and MPEG-4 AAC without SBR. The tests were done at the T-Nova test-site in Berlin, using the MUSHRA test method, and the items used were the twelve items commonly used within MPEG (castanets, pitchpipe, glockenspiel, plucked strings, speech, orchestra and pop music).

From Fig. 11, where MPEG-4 AAC at 24 and 30 kbit/s mono is compared with MPEG HE AAC at 24 kbit/s, it is evident that the new HE AAC profile decoder outperforms the MPEG-4 AAC by a wide margin. The absolute performance of the HE AAC profile codec is very good at this low bitrate. From Fig. 12, where MPEG-4 AAC at 48 and 60 kbit/s stereo is compared with MPEG HE AAC

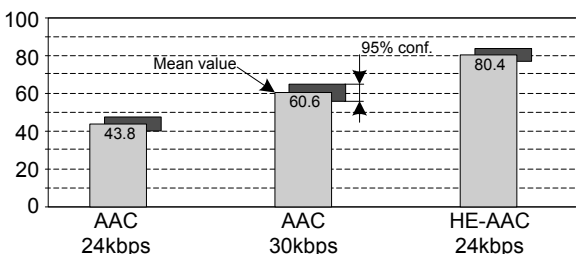


Figure 11: Listening test results for MPEG-4 AAC compared with MPEG-4 HE AAC, mono.

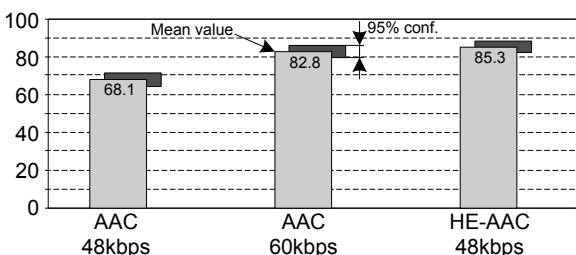


Figure 12: Listening test results for MPEG-4 AAC compared with MPEG-4 HE AAC, stereo.

at 48 kbit/s, it is evident that the new HE AAC profile decoder once again outperforms the MPEG-4 AAC.

Clearly, based on the above test results, HE AAC profile codec has overtaken the leading position in audio-compression efficiency from the previous technology leader MPEG-4 AAC.

4 INTEGRATING MPEG SBR INTO THE MPEG-4 FRAMEWORK

4.1 Embedding SBR data into AAC: Extension payload

SBR data is embedded into the AAC bitstream by means of the `extension_payload()` element, as shown in Fig. 13, and can therefore be combined with various AAC object type flavours (for example AAC LC, AAC Scalable, ER AAC LC). Two types of SBR extension data can be signalled through the `extension_type` field of the `extension_payload()`: `EXT_SBR_DATA` and `EXT_SBR_DATA_CRC`. The latter includes a 10 bit CRC in addition to the SBR data.

SBR can also be used in combination with a bitrate scalable system:

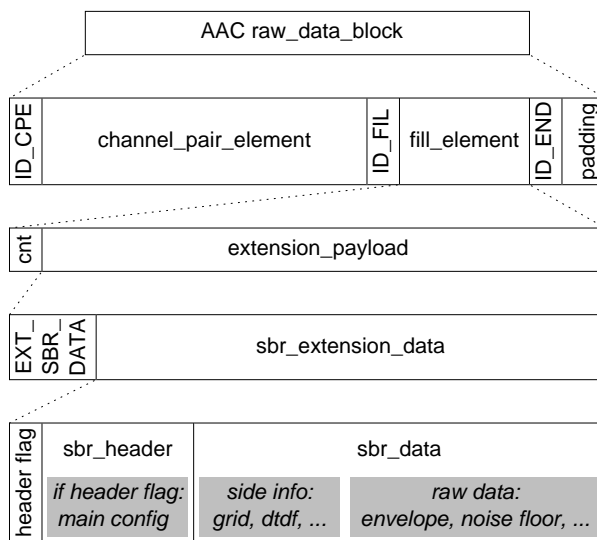


Figure 13: Conveying SBR data as extension payload of an AAC bit-stream frame (this example: Stereo AAC LC).

MPEG-4 Audio is inherently scalable. If, for example, a transmission uses an error-prone channel with limited bandwidth, an audio stream consisting of a small base layer and a larger extension layer provides a robust solution. Strong error protection on the base layer (adding only little overhead to the overall bit-rate) makes sure there is always a signal, even with difficult reception. The extension layer (with little error protection) and base layer together give excellent quality in normal conditions. Any errors lead only to a subtle degradation of quality but never in a total interruption of the audio stream. [26]

Scalable AAC example As an example Fig. 14 shows a possible three layer scalable system without the use of the new SBR tool, similar to the one used in the 1998 MPEG-4 Audio verification test [27]. The base layer includes a 20 kbit/s single channel AAC AOT with a sampling rate of 24 kHz, which covers an audio bandwidth of typically up to 8 kHz. The second layer implements a mono/stereo scalability and results in a stereo signal with 24 kHz sampling rate and 8 kHz bandwidth. Such a layer typically requires another 16 kbit/s. The third layer extends the bandwidth of both channels to about 12 kHz and requires approx. another 16 kbit/s. Depending on the

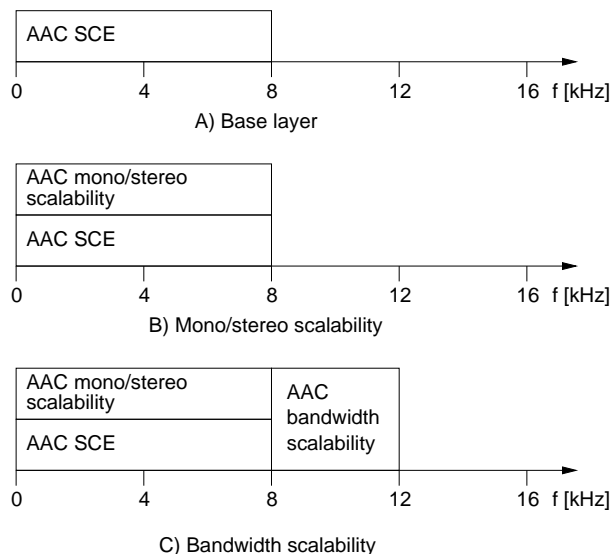


Figure 14: Example of a scalable system without the SBR tool.

quality of the transmission channel, a decoder might be able to decode one or more layers. Specifically when switching between the second and third layer, the change in bandwidth will be clearly noticeable and unpleasant to the listener.

Scalable HE AAC example With MPEG SBR a new tool is standardized that significantly increases the quality of such a scalable system. As indicated in Fig. 15, the SBR tool allows for maintaining a constant bandwidth of the audio signal of typically 16 kHz. Therefore the perceived audio quality in each layer is increased and the artifacts when switching between bandwidth scalable layers of the AAC core codec are reduced. Note, that the bandwidth scalability of the AAC core now becomes a crossover frequency scalability between the AAC and SBR algorithms while the overall bandwidth remains constant. HE AAC also allows for reducing the bitrate in each of the layers while maintaining the same audio quality. This is achieved by lowering the crossover frequency between the AAC and SBR algorithms. In real-world applications often a two-layer scalable configuration is used. Mono \rightarrow stereo scalability and crossover scalability are then combined into a single enhancement layer.

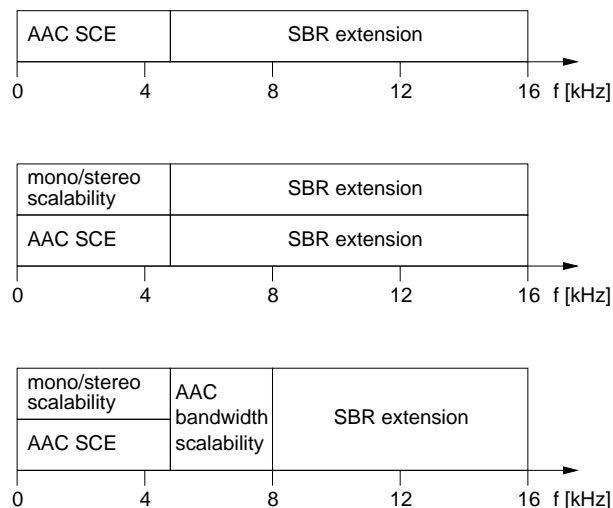


Figure 15: Example of a scalable system with the SBR tool. Contrary to a scalable system without SBR, it provides a constant bandwidth and requires lower bitrates by reducing the crossover frequency between the AAC and SBR cores.

Embedding SBR data into a scalable system The scalable SBR data is embedded into the MPEG-4 stream in the same way as for non-scalable SBR data elements, by means of using the `extension_payload()`. For core coder bandwidth scalability, the SBR data is transmitted in the lowest AAC core coder layer and the SBR data covers the largest SBR frequency range used in the scalable system. In case mono/stereo scalability is chosen, the lowest stereo layer also carries an SBR data element. Note, that in the example above there is no additional SBR data being transmitted as part of the third layer. The SBR decoder will automatically adapt to the new crossover frequency used by the AAC core codec.

4.2 Methods for signaling SBR

The combination of AAC and SBR provides a significant increase of audio compression efficiency. At the same time it enables compatibility with existing AAC-only decoders. An AAC-only decoder will play only the AAC part of an HE AAC bitstream resulting in lower audio quality. There are several ways to signal the presence of SBR data:

1. **implicit signaling:** if SBR extension elements (`EXT_SBR_DATA` or `EXT_SBR_DATA_CRC`) are

detected in the bitstream, this implicitly means that SBR data is present. This mode provides easy backward compatibility with AAC-only decoders but can introduce challenges when operating the decoder in a complex system such as an embedded device. The decoder needs to parse the payload at least partially in order to detect SBR. Only then can the output sampling rate be determined.

2. **explicit signaling:** the presence of SBR data is signaled by means of the AOT SBR in the `AudioSpecificConfig()`. This permits to convey configuration data specific to the SBR decoder, which includes separate specifications of the sampling rates for the SBR and AAC decoders. These specifications are also used to implicitly signal the down-sampling mode described in 3.3. If the sampling rates for the SBR and AAC decoders are identical, the down-sampled SBR tool is used. Two types of explicit signaling are available:

- (a) **hierarchical signaling:** if the first AOT is signaled as SBR, a second AOT is signaled which indicates the underlying AOT, e.g. AAC LC. This is a non backward compatible signaling method.
- (b) **backward compatible signaling:** the extensionAOT is signaled at the end of the `AudioSpecificConfig()`. This signaling method can only be used in systems that convey the length of the `AudioSpecificConfig()`. Because of this restriction, backward compatible explicit signaling can for example not be used with most LATM configurations. This mode does allow to explicitly signal the absence of SBR data as well and thus to explicitly signal the absence of implicit signaling. This circumvents the challenges that can occur when the decoder needs to check for implicit signaling.

To achieve backward compatibility with existing AAC-only decoders, a profile containing AAC (except HE AAC Profile) should be indicated on MPEG-4 Systems level (see Section 4.3) and either signaling method 1 or 2b shall be used. Method 1 can also be used in context with MPEG-2 AAC, where the MPEG-4 `AudioSpecificConfig()` is not available.

4.3 New Profiles & Levels

MPEG-4 provides a large and rich set of tools for the coding of audio objects. In order to allow effective implementations of the standard, subsets of the tool set have been identified that can be used for specific applications. The function of these subsets, called “Profiles,” is to limit the tool set a conforming decoder must implement. For each of these Profiles, one or more Levels have been specified, thus restricting the computational complexity.

The High Efficiency AAC Profile is introduced as a superset of the AAC Profile. Besides the AOT AAC LC (which is present in the AAC Profile), it includes the AOT SBR. Levels are introduced within these Profiles in such a way, that a decoder supporting the High Efficiency AAC Profile at a given level can decode an AAC Profile stream at the same or lower level.

Level	Max. channels /object	Max. AAC sampling rate, SBR not present [kHz]	Max. AAC sampling rate, SBR present [kHz]	Max. SBR sampling rate, [kHz] (in/out)
1	NA	NA	NA	NA
2	2	48	24	24/48
3	2	48	48	48/48
4	5	48	24/48	48/48
5	5	96	48	48/96

Table 1: Levels within the HE AAC Profile

4.4 HE AAC signaling: Use Scenarios

Contrary to previous sections which described the technical specifications of HE AAC signaling from a MPEG-4 point of view, this section will look at these methods from a system integrator’s point of view. On a high level, it can be distinguished between content providers/creators, decoder manufacturers, and system designers (e.g. closed systems for digital broadcasting etc.). Whereas the first group is in control of the encoding side only, and the second group is in control of the decoder only, the third group has the advantage of being in control over the complete signal chain.

Choosing a format The combination of the different signaling methods, as described in Section 4.2, with

the different transport multiplexes as defined in Section 2.3, results in a large variety of bitstream formats for HE AAC. It should be noted that these different transport formats do not have any effect on the format of AAC raw_data_block. Thus transforming content from one into another transport format is a rather simple process. Therefore, as indicated in Fig. 3, a content provider might store data in the MPEG-4 file format (MP4FF) and then repackage it for the different distribution channels and configurations. On the other hand a decoder manufacturer needs to decide up-front which formats to support. It is conceivable that many MPEG-4 decoder manufacturers will include support for the ADIF and ADTS formats in order to be able to play back the widest possible range of bitstream formats.

4.4.1 HE AAC for content providers

As a content provider/creator, one can decide who will be able to decode the bitstreams. In general, HE AAC is backward compatible to existing AAC-only decoders. However, using decoders without the SBR tool will of course result in lower audio quality. It is possible to use signaling such that only users with new HE AAC decoders can decode the content. This is useful in order to guarantee the full audio quality for all listeners. In such a scenario, backwards compatibility to existing, older, AAC-only decoders is explicitly disabled.

Supporting only HE AAC decoders This can be achieved by indicating the new HE AAC profile. Legacy AAC decoders will not play back such audio content. On the other hand new HE AAC decoders have to use both, AAC and SBR decoder tools.

Supporting legacy MPEG-4 AAC-only decoders There are two signaling methods that will allow even legacy MPEG-4 AAC-only decoders to decode new HE AAC content. See Section 4.2 for details. In either case legacy AAC decoders will skip the SBR data, while new HE AAC decoders use that data to configure and call the SBR tool.

Supporting MPEG-2 and MPEG-4 decoders As long as implicit signaling for the SBR tool and the AAC LC AOT without the PNS tool are used, HE AAC can be formatted such that both MPEG-2 and MPEG-4

decoders can decode the content. MPEG is preparing a new release of the MPEG-2 standard to support this configuration [28]. Thus it is possible to create HE AAC content which can be decoded by MPEG-2 and MPEG-4 AAC decoders. Such HE AAC bitstreams will be decoded by AAC-only and HE AAC decoders, thus guaranteeing the largest number of compatible decoders.

4.4.2 HE AAC for decoder manufacturers

The MPEG standard defines profiles and levels to support interoperability between different encoder and decoder implementations. In general a decoder manufacturer is free to support any subset of tools as required. However, a *MPEG conformant* decoder needs to implement at least the subset of tools specified for a specific profile and level. Ultimately it will be important that the decoder can decode the desired content, which is dependent on the settings used by the content creator. An MPEG conformant encoder will also take into account the different profiles and levels so that interoperability between encoders and decoders is guaranteed. However, one cannot foresee at this point, what kind of levels might be used mostly by content creators.

Computational complexity Specifically for manufacturers of embedded decoders the computational complexity is a main concern. The MPEG standard allows for controlling the computational complexity of a HE AAC decoder by means of the following three methods:

1. In situations where the listening environment is not as demanding, a specific low-power decoder can be used. See Section 3.2 for details on this tool, and Section 5.1 for detailed complexity assessments.
2. In general the SBR decoder operates at double the sampling rate of the AAC core coder. (Dual-rate system). However, in situations where a lower output sampling rate is sufficient (e.g. some portable devices) the SBR tool can run in a down-sampled mode, where the input and output sampling rate are identical. (See Section 3.3 for details.)
3. As the complexity increases with increasing levels it is important to only support the required level. The different levels differ in the maximum number of channels as well as the maximum sampling rates of the AAC core and SBR tool as indicated in Table 1.

Following a few examples of how these options could be combined:

- *Basic configuration for cell phones:* Since the listening environment is not very demanding but computational resources might be extremely limited, a down-sampled, low-power HE AAC decoder supporting level 2 of the HE AAC profile might be suitable.
- *Basic configuration of a portable flash player:* The listening environment might be more demanding than in the case of a cell phone. A high quality HE AAC decoder supporting level 2 of the HE AAC profile might be suitable.
- *Basic configuration of a high-end hard-disk player:* Assuming the player does not support multichannel playback, a high quality HE AAC decoder supporting level 2 and 3 of the HE AAC profile might be suitable. Level 3 will allow to play back high-quality, high-bitrate files that utilize a samplerate of up to 48 kHz for the AAC core coder.
- *Basic configuration of a desktop decoder:* A high-quality HE AAC decoder supporting levels 2 to 5 of the HE AAC profile might be suitable for a state of the art desktop decoder. This will allow play back of up to 5 channels and an output sampling rate of up to 96 kHz.

4.4.3 HE AAC for digital broadcasting

The category of digital broadcasting applications includes proprietary digital radio systems such as XM Satellite Radio [29], digital radio standards such as Digital Radio Mondiale (DRM) [30], Internet-radio applications, or standards such as HDTV, DVD etc. These scenarios are similar in that they describe closed systems where MPEG *conformance* is not of major concern. However, using tools that are described in the MPEG standard and combining them into a new system is not only possible, but is a major concept behind the complete MPEG-4 standard. Although conformance and thus specific profiles and levels are unimportant, computational complexity of the decoder continues to be an area of major concern. So the same concepts as in the previous section also apply here.

A special case is digital broadcasting over noisy channels. For such a situation, a system that can adapt to the

quality of service of the transmission channel is desirable. MPEG-4 introduces scalable audio codecs for such a scenario and the SBR tool supports and enhances this configuration as well. Section 4.1 includes an example for such a scalable system.

5 IMPLEMENTATIONS AND APPLICATIONS

With HE AAC, MPEG once more standardized a leading technology that is suitable for many existing and emerging markets and applications. The coding efficiency is high enough to support multichannel applications at low bitrates. A 5.1 bitstream at 112 kbit/s has been demonstrated and can support areas such as audio/video applications.

MPEG-4 HE AAC is a proven technology which has been widely deployed and is ready for use today. Reference decoder source code will soon be made available through MPEG and optimized source code for both encoders and decoders is available for license from companies such as Coding Technologies. Optimized binary implementations for Win32, Linux, MacOS X, ST Micro, ARM, Motorola, TI C64x & C55x, and Trimedia are presently available and other firmware implementations are being completed.

This section will first list the computational requirements of HE AAC decoder implementations. It will be shown, that despite the fact that the quality of HE AAC compared to plain AAC is significantly increased, the computational complexity of HE AAC is still comparable to a plain AAC decoder. In case of the already mentioned Low power decoder, the computational complexity is even identical to that of a plain AAC decoder. Thus HE AAC can indeed be implemented on low-power devices such as cell phones, enabling the delivery of full-band audio signals at a quality level that was not conceivable until today.

Section 5.2 then lists a few examples of applications and use scenarios that are well suited for HE AAC.

5.1 Computational complexity

The computational complexity of decoders is a major factor in deploying codecs in real-world applications. Specifically in mobile devices, computational complexity and memory requirements are directly related to power consumption and implementation cost. HE AAC was specifically designed to minimize these resources.

Since HE AAC adds additional algorithms to the decoding process compared to plain AAC, one might expect a significant increase in computational complexity. However, as described in Section 3.3, the AAC core typically runs at half the output sampling rate. Thus the AAC core only requires about 50 % of the computational complexity compared to an AAC core running at the full sampling rate. The remaining resources are then being used by the additional SBR tool.

The MPEG standard indicates the computational complexity of various profiles, levels, and decoders by means of an abstract assessment based on Processor Complexity Units (PCU) and RAM Complexity Units (RCU). Although these numbers are suitable as a general, platform independent indicator of the computational complexity, a set of real-world numbers for embedded implementations is often a better approach. The numbers published in this paper are derived from implementations based on Coding Technologies Fixed-Point Firmware Reference Code (FFR) [31] [32] which is optimized in terms of memory usage and processing power. This code has been ported to various platforms so that the resulting performance numbers provide a realistic overview of the complexity.

Table 2 indicates the distribution of processing power between the AAC and the SBR decoder for a 24-bit and a 16-bit processor. All numbers given in this document relate to 44.1 kHz stereo operation. To achieve the superior audio quality of HE AAC a data width of at least 20 bits is required. Thus large parts of the SBR decoder need to be implemented using double precision arithmetic on devices with smaller data width, such as common 16-bit devices. Table 2 shows this difference in computational complexity.

Low power decoder As already described in Section 3.2, the MPEG SBR tool is available in two versions: High-quality (HQ) and Low power (LP). The difference only affects the decoding process. The encoder and the bitstream are identical and do not contain any information regarding these two versions. Which decoding process is used can be chosen depending on the application and the need to reduce power consumption and/or chip size. The high-quality version offers better audio quality at the cost of greater computational complexity. It should be used in situations where audio quality is most important. The Low power version results in a reduced audio quality but also requires less computational resources

	AAC	SBR	Total
HE AAC @24 bit	5	3	8
HE AAC @16 bit	9	4	13

Table 3: Distribution of ROM usage in KWords between AAC decoder and SBR decoder

for both memory and complexity. Typical scenarios for the Low power version are mobile devices where complexity and thus power consumption is critical while the listening environment is usually not as demanding as in other applications. Table 2 shows that the computational complexity of the Low power decoder does not exceed the requirements of a plain AAC decoder. Thus better compression technology at the same computational cost is achieved.

Memory requirements Tables 3 and 4 summarize the memory requirements of the different HE AAC decoder implementations. The total data memory required can be as low as 20 KWords depending on the target platform and the chosen decoder implementation.

5.2 Applications and Use scenarios

The unique capability of HE AAC to achieve high quality at very low bitrates not only enhances existing markets but also enables new markets for digital audio. Where transmission bandwidth is constrained, the value of “High-efficiency AAC” is magnified. It not only enhances audio-only services, but also video services like digital TV. By coupling HE AAC with MPEG-4 Video, more bits can be allocated to the video signal without degrading the quality of the audio signal. This is true for mono, stereo, and multichannel applications. Combined with the new MPEG/ITU Advanced Video Coding (AVC) standard (included in MPEG-4 as part 10), even more significant gains in quality are possible.

Licensing Licensing of the SBR technology follows the simple licensing model of MPEG-4 AAC with annual caps for personal computer applications, per-unit fees for hardware devices and no additional fees for electronic music distribution. Licensing for HE AAC consists of two parts with schedules and details available from Coding Technologies [33] for the SBR object types and from Via Licensing [34] for the AAC object types.

	AAC ^a	SBR	Total
HE AAC HQ @ 24-bit	10	15-20	25-30
HE AAC LP @ 24-bit	10	ca. 10	ca. 20
HE AAC HQ @ 16-bit	20	30-40	50-60
HE AAC LP @ 16-bit	20	ca. 20	ca. 40
plain AAC (full sampling rate) @ 24-bit	20	-	20
plain AAC (full sampling rate) @ 16-bit	40	-	40

^a runs at half sampling rate in HE AAC

Table 2: Distribution of computational complexity in MIPS between the AAC and SBR decoder (stereo @ 44.1 kHz f_s). The complexity on 16-bit devices is largely dependent on the number of cycles needed to execute a double precision multiply-accumulate (MAC) instruction. Above estimates assume three cycles per double-precision MAC.

	input	AAC		SBR		output	Total
		static	dynamic	static	dynamic		
HE AAC HQ @24 bit	0.7	2.5	2	4	3	4	≈ 15
HE AAC LP @24 bit	0.7	2.5	2	ca. 3	ca. 1	4	≈ 12
HE AAC HQ @16 bit	0.7	4	4	7	8	4	≈ 24
HE AAC LP @16 bit	0.7	4	4	ca. 6	ca. 4	4	≈ 19

Table 4: Distribution of RAM usage in KWords between the AAC core and SBR (stereo)

Mobile streaming and download There is significant expectation for mobile multimedia services associated with the new 2.5G and 3G mobile service networks. Streaming video and other high-bandwidth services have been showcased as the ideal applications for this new infrastructure. The problem is that for much of the coverage area the peak bandwidth in these networks is generally around 144 kbit/s with individual users sustaining connections of about 40 kbit/s [35]. Delivering quality video over this type of connection is problematic and may not meet the consumer expectation. HE AAC is well suited to solve this problem by being able to provide consumer-grade download and streaming audio services within today's bandwidth. 48 kbit/s HE AAC provides CD-quality stereo programming while at 32 kbit/s, it provides excellent quality stereo programming. These bitrates combined with the growing market for subscription audio services show a strong business opportunity starting in 2003 on into 2004 and beyond.

Digital broadcast via satellite and cable *aacPlus* gained its first commercial success with XM Satellite Radio [29]. It allowed XM Radio to launch the most successful digital radio system in the US to date. *aacPlus* was also selected by the Digital Radio Mondiale [30]

consortium as part of the standard for Shortwave and AM digital radio. This heritage has brought credibility for HE AAC in the open standards world of digital satellite and cable broadcasting. As operators look to enhance their services with more channels or with high-definition, the efficiency of HE AAC gives them more options to either consolidate audio bandwidth to make room for more video or to layer more audio services like multilingual and 5.1 surround. Since Spectral Band Replication (SBR) is being added to the MPEG-2 standard as well, operators have the flexibility to use HE AAC within the MPEG-2 or MPEG-4 context as desired.

Internet distribution Video on demand and subscription content services are continuing to grow on the Internet. In these services, aggregate server bandwidth usage and last-mile bandwidth constraints make it difficult for operators and aggregators to provide high-quality, reliable services to consumers. By cutting the bandwidth requirements for audio nearly in half, MPEG-4 both reduces costs and increases reliability for Internet content services. In addition the flexible signaling of HE AAC within the MPEG-4 framework as described in Section 4.2 allows to control the user's listening experience. The content provider can for example decide if

the new HE AAC bitstreams can be played back on plain AAC decoders as well as on HE AAC decoders. This trade off between backward compatibility and guaranteed audio quality at the decoder can be chosen for individual bitstreams.

6 CONCLUSIONS

The flexible integration of the new MPEG SBR tool into the MPEG-4 framework allows to control the backwards compatibility of HE AAC. Audio content can be packaged such that only HE AAC decoders can play back the audio streams or it can be packaged so that legacy AAC decoders can play back the new content as well, albeit at a lower audio quality. As such MPEG SBR integrates perfectly into MPEG-4 and enhances this open toolbox with an important audio coding method.

MPEG-4 also provides several transport protocols that support streaming applications as well as file based storage. The MPEG-4 framework guarantees the easy integration of HE AAC with MPEG video, thus combining state of the art video coding with state of the art audio coding. High quality multichannel and multilingual audio content at low bitrates thus becomes available for deployment in various audio/video application, enabling a whole new market for content storage and distribution.

In addition MPEG standardized for the first time two audio decoders that can both decode the same audio content: A high-quality version as well as a low power version. The availability of both decoders for HE AAC enables the standard to run on the widest possible range of processors in mobile and portable device applications. The flexibility in content creation as well as decoder implementation make HE AAC a highly effective yet generic MPEG-4 profile. This reflects the true spirit of the MPEG-4 standard: Providing a set of state of the art multimedia technologies that can be combined to support all multimedia applications.

7 REFERENCES

- [1] ISO/IEC JTC1/SC29/WG1, "Text of ISO/IEC 14496-3:2001/FDAM1, Bandwidth Extension," ISO/IEC JTC1/SC29/WG11 N5570, Mar. 2003.
- [2] ISO/IEC JTC1/SC29/WG11, "Overview of the MPEG-4 standard," ISO/IEC JTC1/SC29/WG11 N4030, Mar. 2001, available: <http://mpeg.telecomitalia.com/standards/mpeg-4/mpeg-4.htm>.
- [3] ISO/IEC JTC1/SC29/WG11, "Official MPEG Home Page," available: <http://mpeg.telecomitalia.com/>.
- [4] F. Pereira and T. Ebrahimi, Eds., *The MPEG-4 Book*, Prentice Hall, Englewood Cliffs, NJ, US, 2002.
- [5] ISO/IEC, "Coding of audio-visual objects – Part 3: Audio (MPEG-4 Audio, 2nd edition)," ISO/IEC Int. Std. 14496-3:2001, 2001.
- [6] ISO/IEC, "Coding of audio-visual objects – Part 2: Visual (MPEG-4 Visual, 2nd edition)," ISO/IEC Int. Std. 14496-2:2001, 2001.
- [7] ISO/IEC, "Coding of audio-visual objects – Part 1: Systems (MPEG-4 Systems, 2nd edition)," ISO/IEC Int. Std. 14496-1:2001, 1999.
- [8] ISO/IEC, "Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC)," ISO/IEC Int. Std. 13818-7:1997, 1997.
- [9] K. Brandenburg, "Perceptual coding of high quality digital audio," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds., chapter 2, pp. 39–83. Kluwer, Boston, 1998.
- [10] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, no. 4, pp. 451–513, Apr. 2000.
- [11] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding," *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 789–814, Oct. 1997.
- [12] J. Herre and H. Purnhagen, "General audio coding," in *The MPEG-4 Book*, F. Pereira and T. Ebrahimi, Eds., chapter 11. Prentice Hall, Englewood Cliffs, NJ, US, 2002.
- [13] C. Herpel, G. Franceschini, and D. Singer, "Transporting and storing MPEG-4 content," in *The MPEG-4 Book*, F. Pereira and T. Ebrahimi, Eds., chapter 7. Prentice Hall, Englewood Cliffs, NJ, US, 2002.

- [14] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real time applications," IETF RFC 1889, Jan. 1996, available: <http://www.ietf.org/rfc/rfc1889.txt>.
- [15] Y. Kikuchi, T. Nomura, S. Fukunaga, Y. Matsui, and H. Kimata, "RTP payload format for MPEG-4 audio/visual streams," IETF RFC 3016, Nov. 2000, available: <http://www.ietf.org/rfc/rfc3016.txt>.
- [16] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC FDIS 14496-8, Carriage of ISO/IEC 14496 contents over IP networks," ISO/IEC JTC1/SC29/WG11 N4712, Mar. 2002.
- [17] ISO/IEC, "Coding of audio-visual objects – Part 3: Audio (MPEG-4 Audio Version 1)," ISO/IEC Int. Std. 14496-3:1999, 1999.
- [18] ISO/IEC, "Coding of audio-visual objects – Part 3: Audio, AMENDMENT 1: Audio extensions (MPEG-4 Audio Version 2)," ISO/IEC Int. Std. 14496-3:1999/Amd.1:2000, 2000.
- [19] ISO/IEC JTC1/SC29/WG11, "Call for evidence justifying the testing of audio coding technology," ISO/IEC JTC1/SC29/WG11 N3483, July 2000, available: <http://www.tnt.uni-hannover.de/project/mpeg/audio/public/w3483.pdf>.
- [20] ISO/IEC JTC1/SC29/WG11, "Call for proposals for new tools for audio coding," ISO/IEC JTC1/SC29/WG11 N3794, Jan. 2001, available: <http://www.tnt.uni-hannover.de/project/mpeg/audio/public/w3794.pdf>.
- [21] M. Dietz, L. Liljeryd, K. Kjörning, and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in *112th AES Convention, Munich, May 10 - 13, 2002*, Preprint 5553.
- [22] S. Meltzer, F. Henn, and R. Böhm, "SBR enhanced audio codecs for digital broadcasting such as "Digital Radio Mondiale" (DRM)," in *112th AES Convention, Munich, May 10 - 13, 2002*, Preprint 5559.
- [23] T. Ziegler, A. Ehret, P. Ekstrand, and M. Lutzky, "Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm," in *112th AES Convention, Munich, May 10 - 13, 2002*, Preprint 5560.
- [24] A. Ehret, M. Dietz, and K. Kjörning, "State-of-the-Art Audio Coding for Broadcasting and Mobile Applications," in *114th AES Convention, Amsterdam, March 22 - 25, 2003*.
- [25] P. Ekstrand, "Bandwidth Extension of Audio Signals by Spectral Band Replication," in *IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002), Leuven, Belgium, Nov. 15, 2002*.
- [26] MPEG-4 Industry Forum, "MPEG-4 - The Media Standard," Nov. 2002, available: <http://www.mpegif.org/public/documents/vault/m4-out-20027.pdf>.
- [27] ISO/IEC JTC1/SC29/WG11, "MPEG-4 Audio verification test results: Audio on the Internet," ISO/IEC JTC1/SC29/WG11 N2425, Oct. 1998, available: <http://www.tnt.uni-hannover.de/project/mpeg/audio/public/w2425.pdf>.
- [28] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 13818-7:2003/FPDAM 1, Signalling of Bandwidth Extension," ISO/IEC JTC1/SC29/WG11 N5568, Mar. 2003.
- [29] "XM Satellite Radio," WWW page, 2003, <http://www.xmradio.com/>.
- [30] "Digital Radio Mondiale," WWW page, 2003, <http://www.drm.org>.
- [31] Coding Technologies, "Embedded aacPlus Decoder Implementations - Product Sheet," June 2003.
- [32] Coding Technologies, "Embedded HE-AAC: Implementation Guideline," June 2003.
- [33] "Coding Technologies," WWW page, 2003, <http://www.codingtechnologies.com>.
- [34] "Via Licensing," WWW page, 2003, <http://www.vialicensing.com>.
- [35] D. Frerichs, "New MPEG-4 High-efficiency AAC Audio," Apr. 2003, available: <http://www.mpegif.org/public/documents/vault/m4-out-30034.zip>.