

TRADEOFF BETWEEN COMPLEXITY AND MEMORY SIZE IN THE 3GPP ENHANCED AACPLUS DECODER: SPEED-CONSCIOUS AND MEMORY-CONSCIOUS DECODERS ON A 16-BIT FIXED-POINT DSP

Osamu Shimada, Toshiyuki Nomura, Akihiko Sugiyama and Masahiro Serizawa

Media and Information Research Laboratories, NEC Corporation
1753 Shimonumabe, Nakahara-ku, Kawasaki, 216-8666, Japan
phone: + (81) 44 431 7513, fax: + (81) 44 431 7590, email: o-shimada@bx.jp.nec.com
web: www.nec.com

Abstract

This paper investigates tradeoff between complexity and memory size in the 3GPP enhanced aacPlus decoder based on 16-bit fixed-point DSP implementation. In order to investigate this tradeoff, the speed- and the-memory conscious decoders are implemented. The maximum number of operations for the implemented speed-conscious decoder is 29.3 million cycles per second (MCPS) for a 32 kb/s bitstream. The maximum number of operations for the memory-conscious decoder, where 70% of the data are allocated to an external memory area, increases by 5.7 MCPS (19%) for the bitstream. The investigation of this tradeoff provides an actual relationship between the computational complexity and the internal memory size of the 3GPP enhanced aacPlus decoder. The implemented decoders enable music download and streaming services on next-generation mobile terminals.

1. INTRODUCTION

The 3rd generation partnerships project (3GPP) [1] has already standardized many technical specifications for 3G mobile systems. In these specifications, highly efficient speech and audio codecs have been included. 3GPP enhanced aacPlus [2] is a newly standardized audio codec for mobile applications. It enables high quality stereo audio compression at a low bitrate in the range of 24–32 kb/s thanks to integration of MPEG-4 parametric stereo (PS) [3][4] into MPEG-4 high efficiency advanced audio coding (HE-AAC) [5].

When 3GPP enhanced aacPlus decoder is implemented on a DSP for music download and streaming services on mobile terminals, there is a tradeoff between low computational complexity and low internal memory size. However, this tradeoff in the 3GPP enhanced aacPlus decoder has not been reported.

This paper investigates tradeoff between complexity and memory size in the 3GPP enhanced aacPlus decoder based on 16-bit fixed-point DSP implementation. In Section 2, the 3GPP enhanced aacPlus algorithm is explained in reference to the HE-AAC algorithm. Section 3 demonstrates computational complexity of the implemented decoder with and without an external memory area on the DSP in order to investigate this tradeoff.

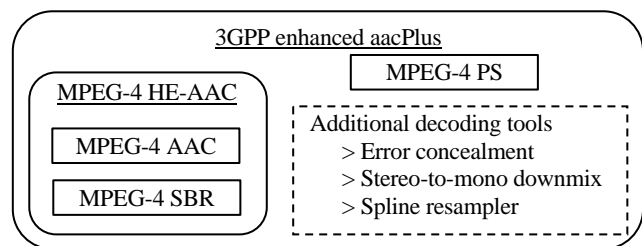


Fig. 1. Framework of 3GPP enhanced aacPlus

2. 3GPP enhanced aacPlus

Figure 1 shows a framework of 3GPP enhanced aacPlus. It is a standard audio codec integrating MPEG-4 PS into HE-AAC. PS enables audio coding at a low bitrate for stereo signals by efficiently representing stereo information. Sound quality of 3GPP enhanced aacPlus at 24 kb/s is statistically equivalent to that of HE-AAC at 32 kb/s [6]. Additional decoding tools are specified in 3GPP surrounded by dashed lines in Fig. 1. They are error concealment tools, stereo-to-mono downmix tools and a spline resampler tool.

2.1 MPEG-4 HE-AAC standard

MPEG-4 HE-AAC is a standard extending MPEG-4 advanced audio coding (AAC) [7] by incorporating spectral band replication (SBR) [5][8]. Figure 2 depicts an encoder and a decoder block diagram of HE-AAC shown in Fig. 1. An SBR encoder and a decoder are a pre- and a post-processor combined with an AAC encoder and a decoder, respectively. At the encoder in Fig. 2 (a), the SBR encoder splits an input signal into low- and high-frequency components. The low-frequency components are used to synthesize a band-limited signal, which is encoded as low-frequency information at the AAC encoder. The high-frequency components are used to calculate a spectral envelope in the SBR encoder. The envelope and other SBR parameters are encoded at a low bitrate as high frequency information. The low- and the high-frequency information are multiplexed into the output bitstream.

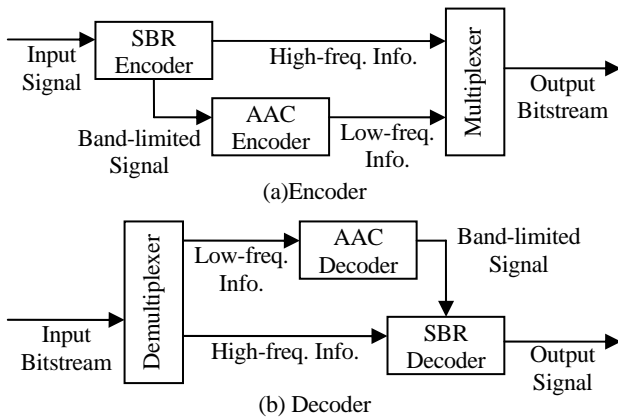


Fig. 2. Block diagrams of the HE-AAC encoder and decoder.

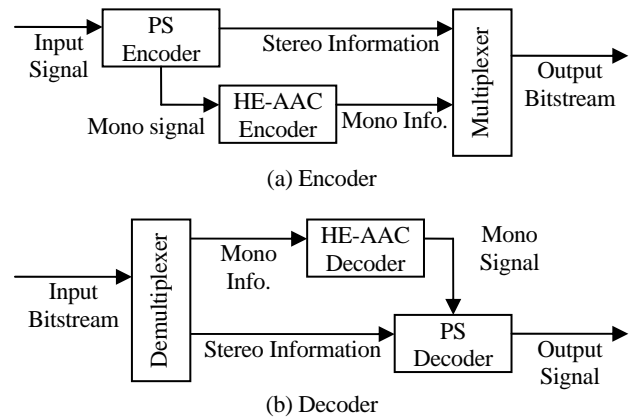


Fig. 3. Block diagrams of the 3GPP enhanced aacPlus encoder and decoder.

At the decoder shown in Fig. 2 (b), the bitstream demultiplexer separates the input bitstream into the low- and the high-frequency information. The AAC decoder generates the band-limited signal from the low-frequency information. The high-frequency information is used to expand the bandwidth of the band-limited signal in order to improve sound quality at the SBR decoder. The expansion process operates in two modes, namely, high quality SBR (HQ-SBR) and low power SBR (LP-SBR).

LP-SBR has been developed specially for mobile terminals to reduce battery consumption. It employs real-valued operations instead of complex-valued operations used in HQ-SBR. However, the real-valued operations cause several problems that seriously deteriorate sound quality. For improving sound quality in real-valued operations, LP-SBR incorporates additional modules [9] with low computational complexity. As a result, The HE-AAC decoder with LP-SBR achieves 30 % lower computational complexity than that with HQ-SBR without statistical difference in sound quality [10].

2.2 3GPP enhanced aacPlus standard

Figure 3 shows block diagrams of the 3GPP enhanced aacPlus encoder and decoder. At the encoder in Fig. 3 (a), the PS encoder separates the stereo signal into stereo information and a mono signal that is obtained by down-mixing the stereo signal. The stereo information consists mainly of the inter-channel intensity difference (IID) and the inter-channel coherence (ICC). IID and ICC represent the power difference and the correlation between the left and the right signals, respectively. The mono signal is encoded as mono information by the HE-AAC encoder. The stereo information and the mono information are multiplexed into the output bitstream.

At the decoder in Fig. 3 (b), the input bitstream is demultiplexed into the mono and the stereo information. The HE-AAC decoder generates the mono signal by decoding the mono information. The decoded mono signal is extended to the stereo signal using the stereo information in the PS decoder.

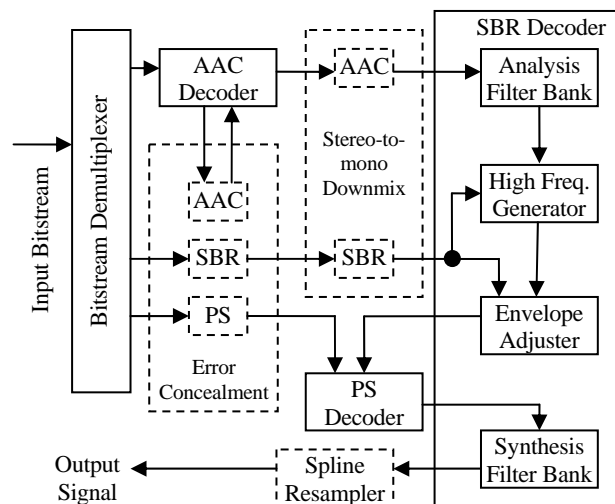


Fig. 4. Block diagram of the 3GPP enhanced aacPlus decoder.

In the specification of 3GPP enhanced aacPlus1, two modes are standardized for stereo signals. At a bitrate higher than 36 kb/s, the encoder is the same as the HE-AAC encoder illustrated in Fig. 2 (a). In this case, the decoder operates in the LP-SBR mode to reduce computational complexity. On the other hand, the framework illustrated in Fig.3 (a) is used to encode stereo signals at lower than or equal to 36 kb/s for improving sound quality.

2.3 3GPP enhanced aacPlus decoder

Figure 4 shows a detailed block diagram of the 3GPP enhanced aacPlus decoder. The input bitstream is demultiplexed into the AAC data, the SBR data and the PS data. The AAC decoder generates the mono signal from the AAC data. An analysis filter bank splits the mono signal into mul-

¹ Both the encoder and the decoder are standardized different from the MPEG standard that specifies only the decoder.

multiple low frequency subband signals. High frequency subband signals are generated in a high frequency generator by copying the low-frequency subband signals. Energy of the generated high frequency subband signals is adjusted in an envelope adjuster using the SBR data. The low and high frequency subband signals in the mono channel are provided for the PS decoder. It generates decorrelated subband signals that have no correlation with the mono subband signals. By mixing the mono and the decorrelated subband signals with a ratio indicated by IID and ICC, stereo subband signals are obtained. A synthesis filter bank synthesizes the stereo signal from these stereo subband signals.

The PS decoder utilizes complex-valued operations for reducing aliasing problems caused by modifying low-frequency subband signals. This is because the human auditory system is sensitive to distortions in the low frequency. The SBR decoder used with the PS decoder also utilizes complex-valued operations, since the SBR and the PS decoders are connected with the complex-valued subband signals.

Three modules surrounded by dashed lines in Fig. 4 are the additional modules specified in 3GPP. The error concealment tools improve sound quality in case of a frame loss. The AAC error concealment tool requires the status of the previous frame and that of the next frame. Consequently, the delay of the decoder increases by one frame. The SBR and the PS error concealment tools generate parameters from those in the previous frames. The stereo-to-mono downmix tools are used for monaural terminals. The SBR data are mixed from stereo to mono in the parameter domain. In this case, the SBR decoder works as the mono decoder in order to reduce computational complexity. The spline resampler tool decimates the output signal for terminals with a narrowband D-to-A converter. This tool converts the sampling frequency from 22.05 or 24 kHz to 8 or 16 kHz.

3. DSP IMPLEMENTATION

For the mobile terminals, it is essential to implement its decoder on an LSI with low power consumption. This is equivalent to DSP implementation with a least number of operations per second. Furthermore, an internal memory area of the DSP is usually shared with other applications such as a video decoder and an acoustic post processing. Therefore, it is of great importance to minimize the internal memory size of the decoder. The balance of the speed and the memory should be determined by mobile terminal manufacturers or their applications.

In view of this tradeoff, two decoders have been implemented on a 16-bit fixed-point DSP, TMS320VC5510 [11], running at 200 MHz, namely, speed-conscious and memory-conscious implementations. The speed-conscious implementation uses only the internal memory area of the DSP to minimize the number of operations by eliminating the access latency to the external memory area. The memory-conscious implementation minimizes the internal memory size of the decoder at the expense of a slight increase in the number of operations. Both decoders were developed by

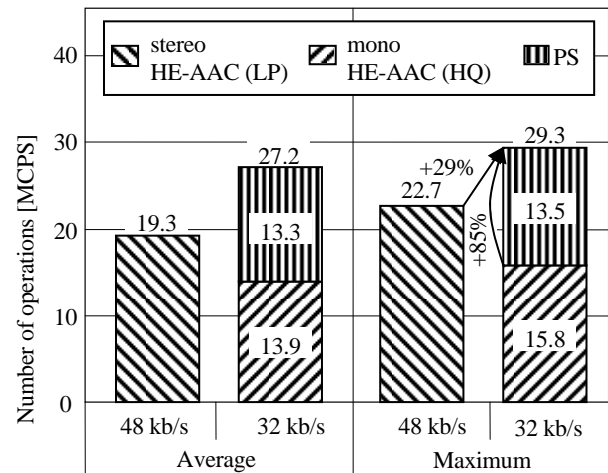


Fig. 5. Computational complexity of the speed-conscious decoder.

Tab. 1. Memory size of the speed-conscious decoder [kWord]

Static RAM	19.0
Scratch RAM	13.5
Constant ROM	19.0
Program ROM	37.4
Total	88.9

incorporating PS and additional tools into the HE-AAC decoder described in [6]. The decoding process has been optimized by maximizing the parallel use of ALUs and MACs, in order to achieve low computational complexity.

3.1 Performance evaluation of speed-conscious decoder

The number of operations was evaluated with two bitstreams encoded at 48 kb/s and 32 kb/s. They were generated, by the 3GPP floating-point reference encoder [12], from a typical stereo music signal sampled at 48 kHz. Complexity of these bitstreams was measured by the 3GPP fixed-point reference decoder [13]. The maximum numbers of operations were 32.2 and 34.6 wMOPS² (weighed million operations per second) for the 48 kb/s and the 32 kb/s bitstreams, respectively.

Figure 5 shows the computational complexity of the implemented speed-conscious decoder with no frame losses. Based on the enhanced aacPlus specification, the 48 kb/s bitstream was decoded by stereo HE-AAC with LP-SBR. PS was used for decoding the 32 kb/s bitstream in combination with mono HE-AAC with HQ-SBR. In this evaluation, the stereo-to-mono downmix and the spline resampler tools in Fig. 4 were disabled. All the data and the instructions code are allocated in the internal memory area on the DSP.

² wMOPS is a measure of computational complexity used in ETSI and ITU-T. This measure is roughly equivalent to MCPS (million cycles per second) for fixed-point DSPs.

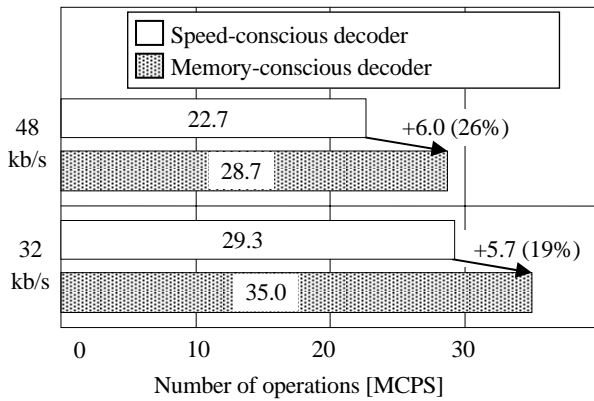


Fig. 6. Complexity comparison between the memory- and the speed-conscious decoders.

The average numbers of operations for the 48 and the 32 kb/s bitstreams are 19.3 and 27.2 MCPS, respectively. 27.2 MCPS consists of 13.9 MCPS for mono HE-AAC with HQ-SBR and 13.3 MCPS for PS. The complexity of PS is about the same as that of mono HE-AAC with HQ-SBR.

The maximum numbers of operations are 23.1 and 33.7 MCPS for the 48 and the 32 kb/s bitstreams, respectively. 33.7 MCPS consists of 16.2 MCPS for mono HE-AAC with HQ-SBR and 17.5 MCPS for PS. Consequently, the 3GPP enhanced aacPlus decoder can be implemented with 85% increase in computational complexity compared to the mono HE-AAC decoder with HQ-SBR. It is 29% more complex than the stereo HE-AAC decoder with LP-SBR.

The memory size of the implemented speed-conscious decoder is shown in Tab. 1. The sizes of static RAM, scratch RAM, constant ROM and program ROM are 19.0, 13.5, 19.0 and 37.4 kWord, respectively. Overall, the memory size is 88.9 kWord. Scratch RAM is shared with other applications, since data in the area are used to decode only the current frame. The size of constant ROM is larger than that in 3GPP reference fixed-point decoder [10]. This is because several data in constant ROM such as huffman codebooks have been modified to achieve low computational complexity. The codebooks are also divided into sub-codebooks to extract quantized values efficiently. Program ROM keeps instruction code to decode bitstreams.

3.2 Performance evaluation of memory-conscious decoder

In order to reduce the internal memory use as well as the increase in computational complexity, some data were moved to the external memory area. These data are not accessed frequently in the decoding process. All the instruction data in program ROM are allocated to the external memory area, since the complexity increase can be avoided by use of the instruction cache. In addition to the moved data mentioned earlier, some areas in static RAM such as filter

Tab. 2. Memory allocation of the memory-conscious decoder [kWord]

	Internal	External	Total
Static RAM	4.8	14.2	19.0
Scratch RAM	12.8	0.7	13.5
Constant ROM	8.0	11.0	19.0
Program ROM	0	37.4	37.4
Overall	25.6	63.3	88.9

buffers were moved to the external memory area. Although these data are accessed frequently, the access is limited to a specific period. To reduce the latency in the memory access, the data are copied to the internal memory area just before the RAM access. After the access, they are stored to the external memory area again.

Table 2 presents the memory allocation of the memory-conscious decoder. The internal memory sizes of static RAM, scratch RAM, constant ROM and program ROM are 4.8, 12.8, 8.0 and 0 kWord, respectively. Since 70% of the data was moved to the external memory area, the internal memory size is reduced from 88.9 kWord to 25.6 kWord. The remaining internal memory area and computing power can be used for other applications.

The computational complexity of the memory-conscious decoder was evaluated in comparison with that of the speed-conscious decoder. The conditions are the same as the earlier evaluation. Figure 6 illustrates the computational complexity of the memory- and the speed-conscious decoders. The maximum numbers of operations for the 48 kb/s and the 32 kb/s bitstreams are 28.7 and 35.0 MCPS, respectively. 35.0 MCPS consists of 19.8 MCPS for mono HE-AAC with HQ-SBR and 15.2 MCPS for PS. The maximum number of operations for the 48 kb/s and the 32 kb/s bitstreams increase by 6.0 and 5.7 MCPS, respectively. While 70% of the data are allocated to the external memory area, the maximum numbers of operations increase only by 26% and 19% for the 48 kb/s and the 32 kb/s streams, respectively.

3.3 Tradeoff between complexity and internal memory size

Figure 7 summarizes the tradeoff between the computational complexity and the internal memory size in the 3GPP enhanced aacPlus decoder for the 32 kb/s bitstream. The area surrounded by dashed oval, which is distributed from top left to bottom right, shows the feasible area of the decoder. The position in the area is determined based on the decoder implementation. The speed-conscious decoder is located at the bottom right, since it is optimized for the computational complexity. The position of the memory-conscious decoder is not the top left, because complexity drastically increases when the internal memory size is further reduced. The balance of the speed and memory size should be determined by the application. For cellular phones, the memory-conscious decoder is preferable, since the decoder is usually used with other applications. On the other hand, the speed-conscious

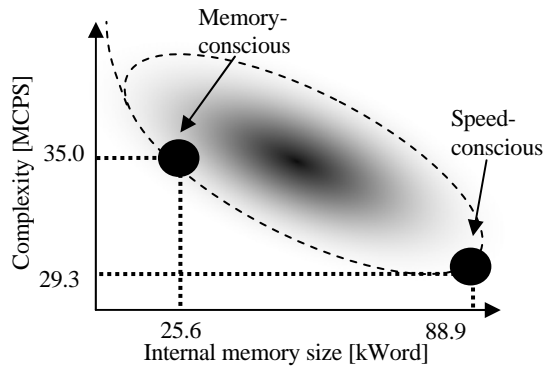


Fig. 7. Tradeoff between complexity and internal memory size for the 32 kb/s bitstream.

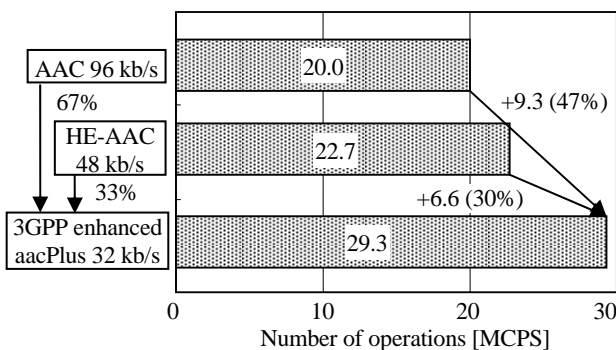


Fig. 8. Complexity comparison among the AAC, the HE-AAC and the 3GPP enhanced aacPlus decoders.

decoder is suitable for music-only players in order to reduce battery consumption. Both of the implemented decoders enable music download and streaming services on next-generation mobile terminals.

3.4 Complexity comparison

Computational complexity of the implemented 3GPP enhanced aacPlus decoder was compared with that of conventional AAC and HE-AAC decoders. The AAC, the HE-AAC and the 3GPP enhanced aacPlus bitstreams were encoded at 96, 48 and 32 kb/s from typical stereo signals sampled at 48 kHz. These bitrates were chosen so that the bandwidth of each decoded stereo signal is about 16 kHz.

Figure 8 shows the comparison result on computational complexity among three decoders. All the data and the instructions code of each decoder are allocated in the internal memory area on the DSP. The maximum numbers of operations are 20.0, 22.7 and 29.3 MCPS for the AAC, the HE-AAC and the 3GPP enhanced aacPlus decoders, respectively. 3GPP enhanced aacPlus achieves 67% bitrate reduction from AAC at the cost of 47% increase in the number of operations. Furthermore, 3GPP enhanced aacPlus achieves 33% bitrate re-

duction from HE-AAC at the cost of 30% increase in computational complexity.

4. CONCLUSIONS

Tradeoff between complexity and memory size in the 3GPP enhanced aacPlus decoder based on 16-bit fixed-point DSP implementation has been investigated. It has been shown that the maximum number of operations for the implemented speed-conscious decoder is 22.7 and 29.3 MCPS for the 48 and 32 kb/s bitstreams. It has also been shown that the maximum number of operations for the memory-conscious decoder, where 70% of the data are allocated to the external memory area, increases only by 6.0 and 5.7 MCPS for the bitstreams, respectively. The investigation of this tradeoff demonstrates the relationship between the computational complexity and the internal memory size of the 3GPP enhanced aacPlus decoder. The implemented decoders enable music download and streaming services on next-generation mobile terminals.

REFERENCES

- [1] <http://www.3gpp.org/>
- [2] 3GPP TS 26.401, "General audio codec audio processing functions; Enhanced aacPlus general audio codec; General description."
- [3] ISO/IEC 14496-3:2001/Amd 2:2004, "Parametric coding for high-quality audio."
- [4] Erik Schuijers, et al., "Low complexity parametric stereo coding," 6073, the 116th AES Convention.
- [5] ISO/IEC 14496-3:2001/Amd 1:2003, "Bandwidth extension."
- [6] ISO/IEC JTC 1/SC 29 /WG 11, N7137, "Listening test report on MPEG-4 High Efficiency AAC v2."
- [7] ISO/IEC 14496-3:2001, "Information technology - Coding of audio-visual objects - Part 3: Audio."
- [8] Martin Dietz, et al., "Spectral band replication, a novel approach in audio coding," 5553, the 112th AES Convention.
- [9] Osamu Shimada, et al., "A Low Power SBR for the MPEG-4 Audio Standards and its DSP implementation," 6048, the 116th AES Convention.
- [10] ISO/IEC JTC1/SC29/WG11, N6009, "Report on the Verification Tests of MPEG-4 High Efficiency AAC."
- [11] Texas Instruments, SPRS076L, "TMS320VC5510/5510A Fixed-Point Digital Signal Processors."
- [12] 3GPP TS 26.410, "General audio codec audio processing functions; Enhanced aacPlus general audio codec; Floating-point ANSI-C code."
- [13] 3GPP TS 26.411, "General audio codec audio processing functions; Enhanced aacPlus general audio codec; Fixed-point ANSI-C code."