

The Optimal Choice of Encoding Parameters for MPEG-4 AAC Streamed over Wireless Networks

Claus Bauer
Dolby Laboratories
San Francisco, CA, USA
cb@dolby.com

ABSTRACT

This paper addresses a novel linear programming based approach to optimize the choice of the encoding parameters for the MPEG-4 AAC audio codec. Current techniques solve the encoding problem for MPEG-4 AAC by splitting the encoding procedure into a sequence of smaller problems that are solved heuristically. In contrast, we approach the coding problem as multidimensional optimization problem and achieve large performance gains compared with current methods.

Categories and Subject Descriptors: Discrete Mathematics, G.2: Communications Applications, H.4.3. **General**

Terms: Algorithms. **Keywords:** Discrete Optimization, Communication Systems.

1. INTRODUCTION

In recent years, the delivery of multimedia content over wireless networks has rapidly gained importance. Multimedia applications are expected to be the driving applications for high bandwidth Third Generation Cellular, WiMAX, and WiFi networks. The success of these technologies also depends on the availability of low bit rate audio codecs such as MPEG4-AAC.

The AAC codec exploits perceptual redundancies to achieve transparent perceptual qualities at low bit rates. In the first step, each audio frame is converted from the time to the frequency domain using the MDCT transform. In a second step, the AAC encoder divides the MDCT coefficients into sets of blocks. It either combines all coefficients in a *long block* or divides them into 8 *short blocks*. In general, a long block leads to a higher time domain resolution of the signal whereas a short blocks lead to a higher frequency domain resolution of the signal. If the MDCT coefficients are divided into short blocks, adjacent blocks can be combined into *regions*. Within a long block or within each region, the transform coefficients are combined into scale factor bands (*SFBs*). For each *SFB*, the AAC encoder allocates bits to

encode the transform coefficients. The transform coefficients are quantized using a dynamically chosen quantization step size, and the quantized coefficients are entropy encoded using Huffman Code Books. In addition to the encoded transforms, the encoder transmits the quantizer step size and the Huffman Code Books as side information. The encoder also uses a small number of bits to inform the decoder which blocking/region structure was chosen. Consequently, the total transmission rate from the decoder to the encoder is the sum of the bits needed to encode the transform coefficients, the sum of the bits needed to encode the region/block structure, and the sum of the bits needed to encode the side information.

In summary, we see that the encoder chooses four sets of *encoding parameters*:

1. Block structure.
2. Grouping structure (for short blocks only).
3. Quantization step size.
4. Huffman code books.

To ensure a low bit rate coding, the encoder must choose these four sets of parameters such that the total transmission rate is below a predefined threshold, while ensuring that a predefined objective measure for the perceptual quality of the decoded signal is met.

The most common perceptual quality measures are the average weighted (*ANMR*) and the maximum weighted noise to mask ratio (*MNMR*), which are also called the average and the maximum distortion. We define both notions in section 2. In this paper, we investigate the problem of determining the set of the *encoding parameters* that optimize the *ANMR/MNMR* subject to an upper bound on the permissible transmission rate. Obviously, low transmission rates that do not compromise the quality of the received audio are of strong interest to the delivery of audio content over bandwidth constrained wireless networks. Throughout this paper, we refer to this problem as the *ANMR/MNMR encoding problem*.

The optimal solution of the *ANMR/MNMR* encoding problem is a highly-complex optimization problem because of the interdependencies of the four sets of parameters, which we describe in the following. For a chosen specific blocking/regioning configuration, the encoder calculates the masking curve per long block or per region. The encoder tries to choose a quantization stepsize such that the resulting *ANMR/MNMR* - which per its definition also depends on

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WMuNeP'05, October 13, 2005, Montreal, Quebec, Canada.
Copyright 2005 ACM 1-59593-183-X/05/0010 ...\$5.00.

the masking curve - is small. The choice of the quantization step size in turn influences the choice of the Huffman Code Book used for the entropy encoding of the quantized transform coefficients. Consequently, the choice of the masking curve influences both the choice of the quantization step size and the Huffman Code Book. The complexity of the *ANMR/MNMR* encoding problem is further increased by the fact that the side information for the quantizer step size and the Huffman code books is encoded differentially between adjacent bands because it depends among others on the difference of the values chosen for these parameters at adjacent bands.

Current AAC implementations do not solve the *ANMR/MNMR* encoding problem optimally. Instead of performing a four dimensional optimization over the encoding parameters described above, the existing technologies simplify the *ANMR/MNMR* encoding problem significantly by choosing the different parameter sets sequentially, i.e., these methods neglect the relationship between these parameter. Moreover, in each sequential step, the existing technologies mostly use - with regard to a given metric - non-optimal heuristics to determine the respective parameter set. We give a detailed overview over these technologies in section 3.

In this paper, we develop the first mathematical model of the complete MPEG4-AAC encoding process. Based on this model, the first optimal solution algorithms to both the *ANMR* and the *MNMR* encoding problem are presented. These algorithms allow us to evaluate the performance of existing, non-optimal techniques as described in section 3 in comparison to the optimal results. Finally, we show how the methods developed in this paper could be further developed to design computationally less complex, non-optimal version of the methods presented here that lend themselves to an implementation in future versions of AAC.

In the next section, we give an exact definition of the *ANMR* and *MNMR* problems. We discuss existing algorithms to solve the *ANMR/MNMR* encoding problem in section 3. In section 4, we present a mixed linear programming formulation and solution of the *ANMR/MNMR* encoding problem. In section 5, we use simulation results to evaluate the performance of previous techniques with the optimal solutions presented in this paper. We conclude in section 6.

2. PROBLEM DEFINITION

We first describe the encoding procedure in AAC. The AAC encoder converts the time domain signal into the spectral domain using the modified discrete cosine transform (MDCT). As not all 1024 MDCT coefficients are relevant for the perception of the decoded signal by the human ear, not all coefficients are encoded individually. The perceptually relevant MDCT coefficients are divided into blocks, regions, and scale factor bands (*SFBs*). Now, we explain these divisions in detail. The encoder decides if the transform coefficients are encoded in a *long* block structure or in *short* block structure. In a long block structure, the perceptually relevant coefficients are divided among typically $T = 49$ *SFBs*. In a short block structure, the coefficients are initially divided into 8 blocks which have the approximate same number of coefficients. Adjacent blocks are then grouped into regions where a region consists of between 1 and 8 adjacent blocks. Thus, there are $2^7 = 128$ possible choices of regions. Each region contains $R = 15$ *SFBs*

such that the total number of *SFBs* is Rq , $1 \leq q \leq 8$, where q is the number of regions. We note for further usage that

$$R < T < 8R. \quad (1)$$

In order to simplify our terminology, whenever in this document we talk about a specific blocking and region structure chosen by the encoder, we will talk about a *configuration*. We number the configurations g from $g = 0$ to $g = G + 1$ where the configurations $g = 0, \dots, G = 127$ correspond to specific groupings of short blocks and $g = G + 1 = 128$ corresponds to a long block configuration.

Within either a long block or a region, in case a short block configuration is chosen, the encoder chooses the scale factors and the Huffman Code Books for each *SFB*.

All coefficients belonging to the same *SFB* are encoded using the same quantization step size. The quantizer step size of each band is controlled by a scale factor (*SF*) selected from a range of M_1 *SFs*. Also, within a *SFB*, the quantized coefficients are entropy encoded using the same Huffman Code Book (HCB) selected from M_2 *HCBs*. Typically, $M_1 = 60$ and $M_2 = 12$. The *SFs* and *HCBs* are transmitted as side information.

We define s_i as the *SF* value and h_i as the *HCB* value for the i^{th} scale factor band in the frame. We assume that the s_i and the h_i only take integer values and require $1 \leq s_i \leq M_1$, and $1 \leq h_i \leq M_2$, $\forall i$, $1 \leq i \leq N$. Both the s_i and h_i are indexes into sets of pre-determined Scale Factors and Huffman Code Books, respectively. Finally, we define vectors the $S = \{s_1, \dots, s_N\}$ and $H = \{h_1, \dots, h_N\}$.

The noise to mask ratio (*NMR*) [6], which is the ratio of the quantization noise to the masking threshold [12], is the most widely used objective measure for the evaluation of an audio signal. Two types of *NMR* are widely used [9]: the Average *NMR* (*ANMR*) and the Maximum *NMR* (*MNMR*). In order to define both *NMR* types analytically, we introduce the following notations: We define by $d(s_i)$ the quantization error of the i -th scale factor band if the i -th scale factor is chosen equal to s_i . $w_{g,i}$ denotes the weight of the i -th scale factor band in the g -th configuration which is defined as the inverse of the masking threshold (see [12]) of the i -th band. The *ANMR* is then analytically expressed as the sum

$$ANMR(S, g) : = \frac{1}{N} \sum_{i=1}^N w_{g,i} d(s_i), \quad (2)$$

where N is the number of *SFBs*, i.e., $N \in \{T, Rq\}$, $1 \leq q \leq 8$. Using the same notation, the *MNMR* is expressed as

$$MNMR(S, g) : = \max_{1 \leq i \leq N} w_{g,i} d(s_i). \quad (3)$$

In the following, we derive an analytic expression for the transmission rate. We first introduce the notation i_g : We say $i \in i_g \Leftrightarrow$ if $(i = 1 \vee g = G + 1) \wedge (SFB_i \text{ is the first } SFB \text{ in a region of the configuration } g \vee g \leq G)$. Obviously, if $g \leq G + 1$, then all $i \equiv 1(\text{mod } R)$ belong to i_g . The transmission rate consists of five parts:

- The amount of bits required to inform the decoder if a long or short block configuration is chosen does not depend on the actually chosen configuration. Thus, it has a non-variable impact on the transmission rate and for this reason we do not consider it further in this paper.

- If a short block configuration is chosen, the number of bits B required to inform the decoder about the chosen regions is proportional to the number of chosen regions q minus one i.e., $B = c(q - 1)$, for a given constant c . If a long block configuration is chosen, then $B = 0$.
- Let $Q_i(s_i, h_i, g)$ be the bits required for the g -th configuration to encode the quantized transform coefficient of the i th band with the SF value chosen as s_i and the HCB value chosen as h_i . We note that the function $Q_i(s_i, h_i, g)$ is also a function of the actual signal X , i.e. $Q_i(s_i, h_i, g) := Q_{X,i}(s_i, h_i, g)$. In general, for any two different signals X and Y , $Q_{X,i}(a, b, g) \neq Q_{Y,i}(a, b, g)$. In the sequel, we consider a fixed signal only and therefore omit the index X .
- The function $F(s_{i-1}, s_i)$ gives the number of bits required to specify the SF for the i th band. As for $i \notin 1_g$, the SF are encoded differentially, we note that $F(s_{i-1}, s_i) := F(s_i - s_{i-1})$. The index s_i for $i \in 1_g$ is not encoded differentially, but with its actual value. We denote this value as $\tilde{F}(s_i)$.
- $G_i(H, g)$ represents the number of bits needed to encode the HCB value of the i -th band in the g -th configuration. If $i \notin 1_g$ and $h_i = h_{i-1}$, then there is $G_i(H, g) = 0$. If $h_i \neq h_{i-1}$, then the value of $G_i(H, g)$ depends on the length of the HCB section starting at SFB_i . A HCB section is defined as a sequence of SFB_i , $a \leq i \leq b$, such that $h_{a-1} \neq h_a$, $h_a = h_{a+1} = \dots = h_b$, and $h_b \neq h_{b+1}$ where possibly $a = b$. For a long block configuration, if $a = 1$ or/and $b + 1 = T$ the first or/and third condition do not apply. For short block configurations, the notion of a HCB section is only defined per region, i.e., HCB sections cannot overlap regions. If the last SFB of a given section and the first SFB of the next section have the same HCB value, then this is not considered as a part of one HCB section, but one says that an HCB of the current section terminates at the last SFB of this section and that a new HCB section starts at the first SFB of the next region. For a short block configuration, the conditions $h_{a-1} \neq h_a$ and $h_b \neq h_{b+1}$ do not apply if a is the first SFB of a region and $b + 1$ is the last SFB of a region respectively. We say that the HCB section defined in this way starts at SFB_a , terminates at SFB_b , and has length $l(a) = b - a + 1$.

For a long block configuration, i.e., $g = G + 1$, there holds according to [5] for $1 \leq i \leq T$,

$$G_i(H, G + 1) = \begin{cases} 0, & \text{if } i \notin 1_g, h_i = h_{i-1}, \\ 9, & \text{if } i \in 1_g, l(h_i) \leq 30, \text{ or } i \notin 1_g, \\ & h_i \neq h_{i-1}, l(h_i) \leq 30, \\ 13, & \text{if } i \in 1_g, l(h_i) \geq 31, \text{ or } i \notin 1_g, \\ & h_i \neq h_{i-1}, l(h_i) \geq 31. \end{cases}$$

In the case of short block configurations, i.e., $g \leq G$, $G_i(H, g)$ is defined in [5] as follows:

$$G_i(H, g) = \begin{cases} 0, & \text{if } i \notin 1_g, h_i = h_{i+1}, \\ 7, & \text{if } i \in 1_g, l(h_i) \leq 6 \text{ or } i \notin 1_g, \\ & h_i \neq h_{i-1}, l(h_i) \leq 6, \\ 10, & \text{if } i \in 1_g \text{ or } i \notin 1_g, 7 \leq l(h_i) \leq 13, \\ & h_i \neq h_{i-1}, 7 \leq l(h_i) \leq 13, \\ 13, & \text{if } i \in 1_g \text{ or } i \notin 1_g, l(h_i) \geq 14, \\ & h_i \neq h_{i-1}, l(h_i) \geq 14. \end{cases}$$

- The definitions of the functions $F(\cdot)$, $\tilde{F}(\cdot)$, and $G(\cdot)$ show that for $i \in 1_g$ the contribution of SFB_i to the rate is calculated differently from the contributions of the SFB_i with $i \notin 1_g$. In particular, for $i \notin 1_g$, the contribution of the SFB_i depends also on the difference of adjacent h_i and s_i values which is not true for $i \in 1_g$. Thus, in order to simplify the notation in the following sections, we redefine for $i \in 1_g$:

$$Q_i(s_i, h_i, g) \rightarrow Q_i(s_i, h_i, g) + \tilde{F}(s_i) + G_i(H, g).$$

We use these definitions to define the transmission rate $R(S, H, g, B)$ as follows:

$$R(S, H, g, B) = \sum_{i=1}^N Q_i(s_i, h_i, g) + \sum_{\substack{i=1 \\ i \notin 1_g}}^N \left(F(s_i - s_{i-1}) + G_i(H, g) \right) + B, \quad (4)$$

where $N \in \{T, Rq\}$, $1 \leq q \leq 8$ is the number of bands of the configuration g . For a given rate threshold R_t and a distortion measure $NMR(S, g) \in \{ANMR(S, g), MNMR(S, g)\}$, the $ANMR/MNMR$ encoding problem is then defined as follows:

$$\begin{aligned} \text{Minimize} \quad & NMR(S, g) \\ \text{such that} \quad & R(S, H, g, B) \leq R_t. \end{aligned} \quad (5)$$

3. PREVIOUS WORK

Previous research has not investigated the MPEG4-AAC encoding problem as a four dimensional optimization problem, but has chosen the four sets of parameters described in section 1 sequentially. This significantly simplifies the problem as the interdependencies between the 4 parameter sets are neglected.

A typical AAC implementation chooses the four sets of parameters in the following order: First, the encoder decides if a long or a short block configuration is chosen. In case a short block configuration is chosen, the short blocks are grouped into regions. In the last step, the SFs and the $HCBs$ are chosen either in an iterative procedure or jointly.

Any AAC implementation that determines a configuration g before deciding the values of the SFs and $HCBs$ faces the following dilemma. Because the encoder does not know the actual SF and HCB values when choosing a configuration g , it must use an estimate for both the distortion and side cost expected to be induced by the SFs and $HCBs$ to be chosen for the configuration g . We now describe the technologies commonly used to decide the long block length and the grouping structure and show how they use the *Perceptual Entropy* and the *Energy level of the transform coefficients* as measures for the expected distortion, respectively.

In [11], a method to decide between a long and a short block structure is presented. This technology uses the concept of *Perceptual Entropy (PE)* introduced in [7]. PE is defined as a measure of perceptually relevant information contained in an audio signal. Expressed in bits per sample, PE represents a theoretical limit on the compressibility of a particular signal and thus gives an indication on how much bits would be needed to encode the signal without introducing any distortion. An explicit calculation of PE has been

given in [10]:

$$PE = \frac{1}{M} \sum_{i=1}^N \sum_{b_i=bl_i}^{bh_i} \left[\log_2 \left(2 \left[\frac{Re(t_{b_i})}{\sqrt{6/w_i k_i}} \right] + 1 \right) + \log_2 \left(2 \left[\frac{Im(t_{b_i})}{\sqrt{6/w_i k_i}} \right] + 1 \right) \right], \quad (7)$$

where

- i index of critical band,
- bl_i upper bound of band i ,
- bh_i lower bound of band i ,
- k_i number of transform components in band i ,
- b_i , index of the transform coefficients in the i^{th} critical band,
- t_{b_i} , transform coefficient in the i^{th} critical band,

where for a fixed configuration g , w_i is the inverse masking threshold as defined in section 2, and M is the number of quantized coefficients. The formula (7) shows that the perceptual entropy depends on the chosen *SFB* structure and thus - by our discussion in section 2 - depends on whether a long or a short block configuration is chosen. By its definition, the *PE* is well correlated with the bit rate in a way that a lower *PE* suggests a lower *NMR* at a low bit rate cost. According to the method proposed in [11], the *PE* is calculated for both a long block configuration and a short block configuration consisting of 8 regions. If the long block configuration has a lower *PE*, the long block configuration is chosen. Otherwise a short block configuration is chosen. As current AAC implementations decide between a long or short block configuration before deciding the region structure of the short blocks, the approach in [11] does not take into account the actual grouping of the short blocks and calculates the *PE* assuming that each region consists of one block. Thus, for a short block configuration, the actual *PE* chosen in the next encoding step by a *grouping algorithm* most likely different from the *PE* calculated for the specific grouping consisting of 8 regions.

For the case of a short block configuration, algorithms to determine the grouping of the short blocks into regions are proposed in [4]. The idea of these algorithms is to combine adjacent short blocks into regions, if the transform coefficients in the blocks have similar average or maximum energy, i.e. absolute value. This approach is motivated by the fact that coefficients with same absolute value tend to have a similar masking curve. In turn, following our discussion of the relation between masking curve and the choice of the quantization step size and Huffman Code Books in section 1, a similar masking curve implies that the coefficients can be encoded with the same quantization step size and a Huffman Code books while achieving a low *ANMR* at a low bit cost.

In [4], a metric is defined that measures the cost of joining the adjacent blocks as a function of the expected distortion and the bits required to transmit the grouping structure. Several optimal and non-optimal ways to find the best grouping scheme with regard to the described metric are described.

For a given configuration g , the *SFs* and *HCBs* have to be chosen. The Two Loop Search (TLS) [3] is the most common procedure to solve this optimization problem. The

TLS uses a heuristic approach, that neglects the inter-band dependencies of the side information and thus simplifies the problem significantly by optimizing the two encoding parameters quantizer step size and Huffman Code Books *independently* for each band. This leads to an increased total transmission rate and/or a decrease in the quality of the decoded signal in relation to a predefined quality measure.

In [1], for the first time a *joint* optimization of the encoding parameters of all bands has been proposed. The problem of determining the encoding parameters is modeled as a Trellis Search. The *ANMR* problem is solved using an iterative Viterbi search through a trellis. The cost function includes a Lagrangian multiplier based approach that penalizes any violation of the target rate. For the *MNMR* problem, an iterative Viterbi search that does not use a Lagrangian multiplier is used. It always provides an optimal solution.

In [2], two algorithms that determine the optimal choice of the *SF* and *HCB* for a given configuration in the *ANMR* problem are proposed. They are based on a mixed integer linear and a dynamic programming solution, respectively. For the *MNMR* problem, an optimal algorithm based on a mixed linear integer programming formulation is shown.

4. AN OPTIMAL SOLUTION BASED ON MIXED INTEGER LINEAR PROGRAMMING

In this section, we propose a method to solve the MPEG4-AAC encoding problem (5) optimally as a mixed integer linear program (MILP).

4.1 Preliminary concepts and variables

We introduce additional notation. We see from (1) that the maximum number of bands for any configuration g is $8R$. In the model we develop in this section, we assume the existence of $8R$ *SFBs* for any configuration g , but we divide the $8R$ *SFBs* in *valid SFBs* which are part of the configuration g and *invalid SFBs*, which are not a part of the configuration g and which we only introduce for modeling purposes. For the model, the invalid *SFBs* will make a zero contribution to *NMR(S, g)* and *R(S, H, g, B)* in (5) and (6). For $g = G+1$, the first T *SFBs* are valid, whereas the last $8R-T$ bands are invalid. For short block configurations, we denote by SB_j , $1 \leq j \leq 8$, the 8 short blocks and assign *R SFBs* to each SB_j . We introduce the binary variables u_j , $0 \leq j \leq 7$, where always $u_0 = 1$, and say that the short blocks SB_j and SB_{j+1} belong to the same region if $u_j = 0$, $1 \leq j \leq 7$. Thus, there is a one-to-one relation between all short block configurations g and all possible choices of the set u_j , $1 \leq j \leq 7$. We call the band SB_j valid if $u_j = 1$. Thus, for a short block configuration with s , $1 \leq s \leq 8$ regions, we have s short blocks SB_{a_j} , $1 \leq j \leq s$, $1 \leq a_1 < \dots < a_j < \dots < a_s \leq 8$ with valid bands and the *SFBs* of SB_{a_j} correspond to the *SFB* of the j -th of the s regions. The other $8-s$ *SBs* are not valid. For any configuration g , we denote the set of valid bands as N_g .

In the next step, we transform the functions $G_i(H, g)$ into a set of other function which can be more easily integrated in a linear programming formulation.

First, we consider the difference a of the *HCB* values of two adjacent *SFBs* $a = h_i - h_{i-1}$. For the long block configuration, we define a function $G_L(a)$ and a variable x_1 as

follows:

$$G_L(a) = \begin{cases} 0, & a = 0, \\ 9, & a \neq 0. \end{cases}$$

$$x_1 = \begin{cases} 1, & \text{if } g = G + 1 \text{ and a } HCB \text{ section} \\ & \text{of length } \geq 31 \text{ exists} \\ & \text{between } SFB_1 \text{ and } SFB_T. \\ 0, & \text{else.} \end{cases}$$

Here and in the following, *between a and b* is to be understood as including *a* and *b*. We note that if no *HCB* section of length ≥ 31 exists, then $G_i(H, 128) = G_L(h_i - h_{i-1})$ for all $2 \leq i \leq T$. As $T = 49$, at most one *HCB* section of length ≥ 31 exists. Thus, we see that $G_i(H, 128) = G_L(h_i - h_{i-1})$ for all SFB_i except the SFB_i where the *HCB* section starts. For this SFB_i , there is $G_i(H, 128) = G_L(h_i - h_{i-1}) + 4x_1$.

For short block configurations $g \leq G$, we define the following functions:

$$G_S(a) = \begin{cases} 0, & a = 0, \\ 7, & a \neq 0. \end{cases}$$

$$x_{2,j} = \begin{cases} 3, & \text{if } u_j = 1, g \leq G, \text{ and a } HCB \text{ section} \\ & \text{of length } \geq 7 \text{ exists between} \\ & SFB_{Rj+1} \text{ and } SFB_{R(j+1)} \\ 0, & \text{else.} \end{cases}$$

$$x_{3,j} = \begin{cases} 3, & \text{if } u_j = 1, g \leq G \text{ and two } HCB \\ & \text{sections of length } \geq 7 \text{ exists} \\ & \text{between } SFB_{Rj+1} \text{ and } SFB_{R(j+1)}, \\ 0, & \text{else.} \end{cases}$$

$$x_{4,j} = \begin{cases} 3, & \text{if } u_j = 1, g \leq G \text{ and a } HCB \text{ section} \\ & \text{of length } \geq 14 \text{ exists} \\ & \text{between } SFB_{Rj+1} \text{ and } SFB_{R(j+1)}, \\ 0, & \text{else.} \end{cases}$$

We see that if $u_j = 0$, then the variables $x_{i,j}$, $2 \leq i \leq 4$, $0 \leq j \leq 7$ are equal to zero which reflects the fact the short block SB_j only contains invalid SFB_i . For $u_j = 1$, if no *HCB* section of length ≥ 7 exists, then $G_i(H, g) = G(h_i - h_{i-1})$ for all $Rj + 2 \leq i \leq R(j + 1)$. In the other case, we note that there exist either one or two *HCB* sections of length between 7 and 13 or there is a section of length ≥ 14 . A case by case analysis shows that for the SFB_i where this (these) section(s) start, there is $G_i(H, g) = G(h_i - h_{i-1}) + r$, where $r = 3 = 3x_{2,j}$ if a (first) *HCB* section of length between 7 and 13 starts at SFB_i , $r = 6 = 3x_{3,j}$ if a second *HCB* section of length between 7 and 13 starts at SFB_i , or $r = 6 = 3(x_{2,j} + x_{4,j})$ if a *HCB* section of length > 13 starts at SFB_i .

Combining these definitions, we see

$$\sum_{\substack{i=1 \\ i \in N_g}}^N G_i(H, g) = \sum_{i=2}^T G_L(h_i - h_{i-1})$$

$$+ \sum_{\substack{i=1 \\ i \in N_g, i \notin 1g}}^{8R} G_L(h_i - h_{i-1}) + 4x_1$$

$$+ \sum_{j=1}^8 3(x_{2,j} + x_{3,j} + x_{4,j}). \quad (8)$$

We introduce additional variables: For a given configuration g , the variables $A_{i,a}$ describe the difference of the values s_i and s_{i-1} . In particular, if SFB_i is a valid band,

the variable $A_{i,a}$ is set equal to 1 if the actual difference between s_i and s_{i-1} is equal to $a - M_1$, and is set equal to zero otherwise. $A_{i,a}$ is equal to zeros for all possible values of a if SFB_i is an invalid *SFB*. Similarly, for the case of a long block configuration, i.e., $g = G + 1$, the variables $B_{L,i,b}$ describe the difference of the values h_i and h_{i-1} , and the variables $B_{S,i,b}$ describe the difference of the values h_i and h_{i-1} for $g \leq G$. We define

$$A_{L,i,b} = \begin{cases} 1, & \text{if } g = G + 1, \text{ and } s_i - s_{i-1} = a - M_1, \\ 0, & \text{else,} \end{cases}$$

$$B_{L,i,b} = \begin{cases} 1, & \text{if } g = G + 1, \text{ and } h_i - h_{i-1} = b - M_2, \\ 0, & \text{else,} \end{cases}$$

$$2 \leq i \leq T, 1 \leq b \leq 2M_2 - 1.$$

$$A_{S,i,b} = \begin{cases} 1, & \text{if } 0 \leq g \leq G, SFB_i \text{ is valid,} \\ & \text{and } h_i - h_{i-1} = b - M_1, \\ 0, & \text{else,} \end{cases}$$

$$B_{S,i,b} = \begin{cases} 1, & \text{if } 0 \leq g \leq G, SFB_i \text{ is valid,} \\ & \text{and } h_i - s_{i-1} = b - M_2, \\ 0, & \text{else,} \end{cases}$$

$$2 \leq i \leq 8R, 1 \leq a \leq 2M_1 - 1,$$

We define the variables $Z_{g,i,a,b} \forall i, 1 \leq i \leq N, \forall a, 1 \leq a \leq M_1 \forall b, 1 \leq b \leq M_2$ that describe which values are taken by the variables s_i and h_i at the i -th stage of the g -th configuration:

$$Z_{g,i,a,b} = \begin{cases} 1, & \text{if } s_i = a, h_i = b, \\ & \text{configuration } g \text{ is chosen,} \\ 0, & \text{else.} \end{cases}$$

Next, we define the variable $e_{i,m}$ as the weighted distortion that occurs at the i -th band when the g -th partition mode is chosen and if $s_i = m$, i.e.,

$$e_{g,i,m} = w_{g,i} d(s_i = m), \forall 1 \leq i \leq 8R, 1 \leq m \leq M_1.$$

Further, in order to avoid negative function arguments we set $F^*(a_2 - a_1) = F(M_1 + a_2 - a_1)$ and $G^*(b_2 - b_1) = G(M_2 + b_2 - b_1)$. Finally, we abbreviate the set of all variables $Z_{g,i,a,b}$ as Z , $A_{L,i,b}$ and $A_{S,i,b}$ as A , the set of all variables $B_{L,i,b}$ and $B_{S,i,b}$ as B , the set of variables u_j as U , and the set of the variables $x_1, x_{2,j}, x_{3,j}, x_{4,j}$ as X .

4.2 MILP formulation for the ANMR problem

For the ANMR distortion measure, we now formulate the MPEG4-AAC encoding problem as follows:

$$\text{Minimize}$$

$$\text{ANMR}(Z) : = \frac{1}{N} \sum_{g=0}^{G+1} \sum_{i=1}^{8R} \sum_{a=1}^{M_1} \sum_{b=1}^{M_2} e_{g,i,a} Z_{g,i,a,b},$$

such that (9) - (40) are satisfied, where

$$R(Z, A, B, g, U, X) \leq R_t, \quad (9)$$

$$\begin{aligned}
R(Z, A, B, g, U, X) &= \sum_{g=0}^{G+1} \sum_{i=1}^{8R} \sum_{a=1}^{M_1} \sum_{b=1}^{M_2} Z_{g,i,a,b} Q_{g,i}(a, b) \\
&+ \sum_{i=2}^T \sum_{a=0}^{2M_1} A_{L,i,a} F^*(a) + \sum_{\substack{i=1 \\ i \notin N_g}}^{8R} \sum_{a=0}^{2M_1} A_{S,i,a} F^*(a) \\
&+ \sum_{i=2}^T \sum_{b=0}^{2M_2} B_{L,i,b} G_L^*(b) + \sum_{\substack{i=1 \\ i \in N_g}}^{8R} \sum_{b=0}^{2M_2} B_{S,i,b} G_S^*(b) \\
&+ 4x_1 + 3 \sum_{j=0}^7 (x_{2,j} + x_{3,j} + x_{4,j}) + c \sum_{j=1}^7 u_j, \quad (10)
\end{aligned}$$

$$\sum_{g=0}^{G+1} v_g = 1, \quad (11)$$

$$\sum_{j=1}^7 u_j 2^{j-1} = \sum_{g=0}^G g v_g, \quad (12)$$

$$\sum_{a=1}^{M_1} \sum_{b=1}^{M_2} Z_{g,i,a,b} = v_g, \quad 0 \leq g \leq G+1, i \in N_g, \quad (13)$$

$$\sum_{a=1}^{M_1} \sum_{b=1}^{M_2} Z_{g,i,a,b} = 0, \quad 0 \leq g \leq G+1, i \notin N_g, \quad (14)$$

$$\sum_{a=1}^{2M_2-1} A_{L,i,a} \leq 1, \quad (15)$$

$$\sum_{b=1}^{2M_2-1} B_{L,i,b} \leq 1, \quad (16)$$

$$\forall i, 2 \leq i \leq T.$$

$$\sum_{a=1}^{2M_1-1} A_{S,i,a} \leq 1, \quad (17)$$

$$\sum_{b=1}^{2M_2-1} B_{S,i,b} \leq 1, \quad (18)$$

$$1 \leq i \leq 8R, i \notin 1_g$$

$$\sum_{i=2}^T \sum_{a=1}^{2M_2-1} A_{L,i,a} = (T-1)v_{G+1}, \quad (19)$$

$$\sum_{i=2}^T \sum_{b=1}^{2M_2-1} B_{L,i,b} = (T-1)v_{G+1}, \quad (20)$$

$$\sum_{i=Rj+2}^{R(j+1)} \sum_{a=1}^{2M_1-1} A_{S,i,a} = (R-1)u_j, \quad (21)$$

$$\sum_{i=Rj+2}^{R(j+1)} \sum_{b=1}^{2M_2-1} B_{S,i,b} = (R-1)u_j, \quad (22)$$

$$0 \leq j \leq 7.$$

$$\begin{aligned}
A_{L,i,a} + 2 - v_{G+1} &\geq \\
\sum_{d_3=1}^{M_2} Z_{G+1,i-1,c,d_3} + \sum_{d_4=1}^{M_2} Z_{G+1,i,c+a-M_1,d_4}, \quad (23)
\end{aligned}$$

$$\begin{aligned}
B_{L,i,b} + 2 - v_{G+1} &\geq \\
\sum_{d_5=1}^{M_1} Z_{G+1,i-1,d_5,d} + \sum_{g=1}^G \sum_{d_6=1}^{M_1} Z_{G+1,i,d_6,d+b-M_2}, \quad (24)
\end{aligned}$$

$$2 \leq i \leq T,$$

$$\begin{aligned}
A_{S,i,a} + 2 - u_j &\geq \\
\sum_{g=1}^G \sum_{d_3=1}^{M_2} Z_{g,i-1,c,d_3} + \sum_{g=1}^G \sum_{d_4=1}^{M_2} Z_{g,i,c+a-M_1,d_4}, \quad (25)
\end{aligned}$$

$$\begin{aligned}
B_{S,i,b} + 2 - u_j &\geq \\
\sum_{g=1}^G \sum_{d_5=1}^{M_1} Z_{g,i-1,d_5,d} + \sum_{g=1}^G \sum_{d_6=1}^{M_1} Z_{g,i,d_6,d+b-M_2}, \quad (26)
\end{aligned}$$

for $0 \leq j \leq 7$, $Rj+2 \leq i \leq R(j+1)$, $1 \leq a \leq 2M_1-1$, $1 \leq b \leq 2M_2-1$, $\max(1, 1+M_1-a) \leq c \leq \min(M_1, 2M_1-a)$, $\max(1, 1+M_2-b) \leq d \leq \min(M_2, 2M_2-b)$.

$$t_k \geq \frac{1}{30} \sum_{i=k}^{k+29} B_{L,i,M_2} - 1, \quad (27)$$

$$Tx_1 \geq \sum_{k=2}^{T-29} t_k, \quad (28)$$

$$g_k \geq \frac{1}{5} \sum_{i=k}^{k+5} B_{S,i,M_2} - 1, \quad (29)$$

$$Rx_{2,j} \geq \sum_{k=Rj+2}^{R(j+1)-6} g_k, \quad (30)$$

$$d_{1,j} \geq \frac{1}{11} \sum_{i=Rj+2}^{R(j+1)} B_{S,i,M_2} - 1, \quad (31)$$

$$e_{1,j} \geq 1 - B_{S,Rj+8,M_2} - B_{S,Rj+9,M_2}, \quad (32)$$

$$e_{2,j} \geq 1 - B_{S,Rj+8,M_2} - B_{S,Rj+15,M_2}, \quad (33)$$

$$e_{3,j} \geq 1 - B_{S,Rj+2,M_2} - B_{S,Rj+9,M_2}, \quad (34)$$

$$d_{2,j} \geq \frac{1}{13} \sum_{i=Rj+2}^{R(j+1)} B_{S,i,M_2} - 1, \quad (35)$$

$$e_{4,j} \geq 2 - B_{S,Rj+8,M_2} - B_{S,Rj+9,M_2}, \quad (36)$$

$$x_{3,j} \geq \frac{1}{3} (d_{1,j} + e_{1,j} + e_{2,j} + e_{3,j}) - 1, \quad (37)$$

$$x_{3,j} \geq d_{2,j} + e_{4,j} - 1, \quad (38)$$

$$w_k \geq \frac{1}{12} \sum_{i=k}^{i+12} B_{S,i,M_1} - 1, \quad (39)$$

$$Rx_{4,j} \geq \sum_{k=Rj+2}^{R(j+1)-12} w_k, \quad (40)$$

for $0 \leq j \leq 7$. The variables Z , X , v_g , u_j , $e_{i,j}$, $d_{i,j}$, g_k , t_k , w_k are required to be integers $\in \{0, 1\}$ and the variables A and B are required to be non-negative real numbers.

Now, we show that this model correctly represents the MPEG4-AAC encoding problem as defined in (5).

The equation (11) shows that exactly one configuration g is chosen by the encoder. For $g \leq G$, the equation (12)

establishes the one-to-one relation between g and a possible choice of the u_j discussed in section 4.1. For a long block configuration, the equations (11) and (12) force all u_j , $j \leq 2$ to vanish which is of no significance for the solution of the MILP because the u_j play no role for a long block configuration (see 4.1.)

We now explain how the variables $Z_{i,a,b}$, $A_{i,a}$, $B_{L,i,a}$, and $B_{S,i,b}$ consistently express the choice of the *SF* and *HCB* values at all *SFBs*. The equations (13) and (14) ensure that for the chosen configuration g and each valid band *SFB_i* exactly one variable $Z_{g,i,a,b}$ is equal to 1, whereas all the other variables are equal to zero, i.e., exactly one *SF* and one *HCB* are chosen per band. For an invalid band *SFB_i*, all $Z_{g,i,a,b} = 0$. For a long block configuration and fixed values of the variables $Z_{G+1,i,a,b}$, the equations (15), (19), and (23) ensure that if $i \notin N_g$ and $s_i - s_{i-1} = a - M_1$, then $A_{L,i,a} = 1$, and all other $A_{L,i,a} = 0$ as required by the definition of $A_{L,i,a}$. The equations (16), (20), and (24) imply the analogous relation for the variables $B_{L,i,b}$.

For $g \leq G$, and fixed values of the variables $Z_{g,i,a,b}$, the equations (17), (21), and (25) ensure that for any $u_j = 1$, and $Rj + 2 \leq i \leq R(j + 1)$, if there is $s_i - s_{i-1} = a - M_1$, then $A_{S,i,a} = 1$ and all other $A_{S,i,a} = 0$ as required by the definition of $A_{S,i,a}$. The equations (18), (22), and (26) imply the same relations for the variables $B_{S,i,b}$. For all invalid bands, the variables A and B are equal to zero.

Finally, we analyze the equations (27) - (40) and compare the values taken by the variables X with their definition in section 4.1. If within a long block configuration at least one *HCB* section of length ≥ 31 exists, the equations (27) and (28) show that $x_1 = 1$ which corresponds to the definition of x_1 in section 4.1. In the second possible case that no long block configuration with a *HCB* section of length ≥ 31 is chosen, the equations (27) and (28) allow x_1 to take any value $\in \{0, 1\}$ which contradicts the definition of $x_1 = 0$ in section 4.1. By its definition in section 4.1, the variable x_1 does not represent any choice of the encoding variables itself, but merely expresses the impact of specific choices of the encoding variables on the rate. If x_1 is chosen equal to 1, then x_1 makes a higher contribution to the rate $R(Z, A, B, X)$ than defined in section 4.1 and thus eventually reduce the solution space of the variables Z, A , and B . To avoid this reduction, the linear program will in the second case always set $x_1 = 0$. Thus, we see that with regard to x_1 , the MILP formulation allows to find the optimal solution of the MPEG4 - AAC encoding problem under the exact constraints formulated in section 4.1. For the discussion of the remaining variables of the set X , the same arguments apply. Thus, we will only show that if their definition in section 4.1 requires them to be equal to 1, then the MILP will force them also to be equal to one, whereas in the case that their definition in section 4.1 requires them to be equal to 0, the MILP allows them to take any value $\in \{0, 1\}$.

The equations (29) and (30) force the variables $x_{2,j}$ to be equal to 1 if at least one *HCB* section of length ≥ 7 exists between the *SFB_{Rj+1}* and *SFB_{R(j+1)}*. If such an *HCB* section does not exist, then $x_{2,j}$ can take any value $\in \{0, 1\}$. The relations (31) - (38) imply that $x_{3,j} = 1$ if two *HCB* sections of length ≥ 7 exists between the *SFB_{Rj+1}* and *SFB_{R(j+1)}*. This can be seen by a case-by-case analysis of all the possible choices of *HCBs* that contain two *HCB* sections of length ≥ 7 . The equations (31) - (34) and (37) model the case when two *HCB* sections of length 7 exist,

where as the equations (35), (36), and (38) model the case when one *HCB* of length and one *HCB* section of length 8 exist. Finally, the equations (39) and (40) imply that $x_{4,j} = 1$ if one *HCB* section of length ≥ 14 exists.

Using the definition of the variables $Z_{g,i,a,b}$ and $e_{g,i,m}$, one sees that the definition of the objective function in (9) corresponds to the definition of the objective function in (2).

We now consider the expression for the rate in (10). The functions $F^*(n)$ and $G^*(n)$ are identical to the function $F(n)$ and $G(n)$ with the argument shifted by M_1 and M_2 , respectively. The shifts have been introduced in order to avoid negative variables in the MILP formulation. Recalling the definitions of the variables $Z_{i,a,b}$, $A_{i,a}$, and $B_{i,b}$, we see that the expressions (4) and (10) are identical. We note that for a short block configuration, there is $q - 1 = \sum_{2 \leq j \leq 8} u_j$, where q is the number of regions, and that for a long block configuration, $\sum_{1 \leq j \leq 7} u_j = 0$. Thus the value of B as defined in

section 2 and used in (4) is correctly expressed in (10).

4.3 MILP formulation for the MNMR problem

For the *MNMR* distortion measure, we introduce an additional variable p , which in accordance with the definition of *MNMR(S)* in (3) is defined as

$$p = \max_{1 \leq i \leq N} \sum_{g=1}^G w_i d(s_i).$$

We note that $w_i d(s_i) = \sum_{g=1}^G \sum_{a=1}^{M_1} \sum_{b=1}^{M_2} e_{g,i,a} Z_{g,i,a,b}$. Thus, we can model the joint optimization problem (5) as follows:

Minimize p such that

$$\sum_{g=1}^G \sum_{a=1}^{M_1} \sum_{b=1}^{M_2} e_{i,a} Z_{i,a,b} \leq p, \quad \forall i, 1 \leq i \leq 8R$$

and (9) - (40) are satisfied.

As for the *AMNR* MILP model, the variables $Z_{i,a,b}$ are required to be non-negative and the variables $A_{i,a}$ and $B_{i,b}$ are required to be non-negative real numbers.

5. NUMERICAL RESULTS

In this section, we compare our method numerically with two well-known ways of encoding an MPEG-4 AAC bitstream. In order to choose a configuration we apply the methods given in [4] and [11], whereas for the choice of the *SF* and *HCB* we consider both the Two-Loop search [3] and the Trellis [1] described in 3. We note that while the Two Loop Search can be executed in real-time, whereas the Trellis Search cannot be used in a real-time encoder. The optimization procedure presented in this paper cannot be executed in real-time either. For the purpose of the experiments and to reduce the execution time, we have relaxed several of the constraints of the MILP formulation. Thus, the results presented here are not the best possible that can be achieved with our method. However the performance gains achieved with the relaxed constraints are already very significant. We are currently developing simulations to test the unrelaxed formulation of the MILP.

For the *ANMR* problem, Table 1 shows that the MILP solution is on average about 10% better than the Trellis Search

Table 1: Comparison of average ANMR values

Compared optimization techniques	ANMR ratio
Two Loop Search vs. MILP	2.4
Trellis Search vs. MILP	1.1

Table 2: Distortion relative to masking curve

Optimization techniques	Probab. that ANMR is below masking curve
Two Loop Search	57%
Trellis Search	86%
MILP	93%

and more than two times better than the Two Loop Search. We achieved similar results for the MNMR problem.

Whereas the *NMR* is an effective measure of ultimate performance of the encoding process, the objective of perceptual compression is to achieve an *NMR* that is *satisfactory*, regardless of whether it is the best *NMR* possible or not. In the context of audio coding, an *NMR* is widely considered as *satisfactory* if it is below or at least not far above the masking curve. Table II shows that whereas the Two Loop Search achieves an *ANMR* below the masking curve with a probability of 57%, the Trellis Search and the MILP formulation obtain *ANMR* values below the masking curve with a probability of 86% and 93%, respectively. Again, similar results were achieved for the *MNMR* problem.

6. CONCLUSIONS

This paper develops the first complete model of the encoding process for the MPEG-4 AAC audio codec. Previous technique to solve the AAC encoding problem divide the encoding problems into sub-problems which are then solved heuristically. In contrast, we model the encoding process as a four dimensional optimization problem. The numerical experiments show that even a relaxed version of the MILP formulation developed in this paper achieves significant performance gains compared to existing technologies. For future work, we develop a test environment to evaluate a non-relaxed version of the MILP and we investigate ways to reduce the execution speed of the algorithm without compromising the quality of the encoded audio signal.

7. REFERENCES

- [1] Aggarwal, A. et al., *Trellis-Based Optimization of MPEG-4 Advanced Audio Coding*. IEEE Workshop on Speech Coding, September 2000, Wisconsin, USA.
- [2] Bauer, C., Vinton, M., *Joint Optimization of Scale Factors and Huffman Code Books for MPEG-4 AAC*, Proceedings of the IEEE MMSP 2004, Siena, Italy.
- [3] Bosi, M., et al., *ISO/IEC MPEG-2 Advanced Audio Coding*. Journal of the Audio engineering society, Vol. 45, No. 10, 1997, Oct, pp.791 -811.
- [4] Feller, M. Bauer, C.; Vinton, M.; Davidson, G.; *Encoding Blocks of Audio Information Arranged in Frames with Constrained Optimization of Segmenting the Frames into Groups of Blocks*. Pending patent application PCT/US2005/001715.
- [5] ISO/IEC, IS-14496 (Part 3) 2001.
- [6] Beaton, R.J., Beerends, J.G., Keyhl, M., Treurniet, W.C., *Objective perceptual measurement of audio quality*. Collected Papers of Audio Bit - Rate Reduction(1996), Gilchrsit, N., Grewin, C.; AES, pp. 126 - 152.
- [7] Johnston, J.D.; *Estimation of perceptual entropy using noise masking criteria*, ICASSP, 1988, New York, NY, May 1998.
- [8] Johnston, J.D.; *Transform Coding of Audio Signals Using Perceptual Noise Criteria*. IEEE Journal on Selected Areas of Communications, Vol. 6, No. 2, February 1988.
- [9] Najafzadeh, H., Kabal, P., *Perceptual bit allocation for low rate coding of narrowband audio*. ICASSP 2000, Istanbul, Turkey, June 2000.
- [10] Painter, T., Spanias, A., *Perceptual Coding of Digital Audio*. Proc. of the IEEE, Vol. 88, No. 4, April 2000, pp. 451 - 513.
- [11] Princen, J., Johnston, J.D., *Audio coding with signal adaptive filterbanks*. ICASSP 1995, Detroit, May 1995.
- [12] Zwicker, E., Fasthl, H., *Psychoacoustics: Facts and Models*. Second Edition, Springer Publishing House, 1990c1999.