NEW BIT-PLANE PROBABILITY CALCULATIONS FOR SCALABLE TO LOSSLESS AUDIO CODING

Mahmood Movassagh *Peter Kabal*

Department of Electrical and Computer Engineering
McGill University, Montreal, Canada

ABSTRACT

Considering the properties of the residual signal, core-based bit-plane probabilities are provided for MPEG-4 Audio Scalable to Lossless Coding (SLS), which matches the quantization and coding performed in the core layer. Using the same strategy, new probabilities are obtained to consider the clipping effect in bit-plane coding of an unbounded signal, which is useful for non-core mode of SLS coding. Simulations show that considering the core layer parameters and the clipping effect improve the bit-plane probabilities estimation compared to the existing method.

Index Terms— Scalable Audio Coding, Quantization, Bit-Plane Coding, Arithmetic Coding, AAC, SLS.

1. INTRODUCTION

Bit-rate scalability is a desired feature in multimedia communications. Without the need to re-encode the original signal, it allows for improving the signal's quality as more of a total bit stream becomes available. This way there is no need to store multiple versions of a signal encoded at different bitrates. Scalable coding can be used to provide users with different quality streaming when they have different constraints or when there is a varying channel. It is especially useful in client-server applications where the network nodes are able to drop higher layer packets to satisfy link capacity constraints.

Several scalable coding systems have been proposed, including using wavelet transform [1], bit-plane coding [2, 3], and fine-grain scalable coding [4, 5, 6]. The state of the art MPEG-4 Audio adopts two main scalable audio coding systems [7]. The first one, Scalable Advanced Audio Coding (S-AAC), is based on reconstruction error quantization of AAC coder. The second system is called MPEG-4 scalable-to-lossless (SLS) and was released as a standard audio coding tool in June 2006 [5, 8]. This system has two modes: a perceptual-core mode and a non-core mode. In both modes the input signal is transformed into the frequency domain using IntMDCT, and then the resulting coefficients are encoded. In the perceptual-core mode the coefficients are encoded using AAC perceptual coding which gives a base-quality signal. Then the residue of the coefficients are coded using a specific bit-plane arithmetic coding called bit-plane Golomb code (BPGC) [2, 9]. In the non-core mode, the BPGC is applied directly to the IntMDCT coefficients. There has been research on improving the performance of both S-AAC and SLS, such as [10] and [11].

In this paper we will investigate the statistical properties of the residual signal for the perceptual-core mode of SLS coding and we will show that the bit-plane probabilities used by BPGC are not well matched to the residual signal. We will then propose an alternative approach which depends on the base signal and leads to better estimation of the bit-plane properties of the residual signal. The proposed approach is then easily extended to consider the clipping effect in bit-plane coding of an unbounded signal.

2. AAC AND SLS CODING

In the perceptual-core mode of MPEG-4 SLS coding, the IntMDCT coefficients are quantized in each scalefactor band (SFB) of a data frame using AAC quantization operation given by [12]

$$i_n[l] = \mathrm{sgn}(c_n[l])\mathrm{nint}(|2^{-s[n]/4}c_n[l]|^{3/4} - 0.0946), \quad (1)$$

where $i_n[l]$ is the l^{th} quantized coefficient index in band n, $c_n[l]$ is the corresponding IntMDCT coefficient, and $s[n]$ is the scalefactor used for that band. The nint() and sgn() operations denote the nearest integer and signum functions and 0.0946 is an offset value which is also referred to as the magic number. The scalefactors are obtained using a psychoacoustic model and control the Noise-to-Mask ratios (NMR) in the scalefactor bands. In the next step, the residue of the IntMDCT coefficients is calculated by forming the quantization error. The residual coefficients in each SFB are then coded using BPGC. The subtraction process for obtaining the residual is however different from a typical error mapping: Instead of subtracting the quantized coefficients from the original ones, the difference between an input coefficient and its corresponding quantization interval threshold is considered to be the error. The thresholds are obtained using

$$\mathrm{thr}(i_n[l]) = \begin{cases} \mathrm{sgn}(i_n[l])(2^{s[n]/4}|i_n[l] - 0.4054|^{4/3}), & i \neq 0 \\ 0, & i = 0, \end{cases} \quad (2)$$

where $\mathrm{thr}(i_n[l])$ is the interval threshold for the quantized coefficient index $i_n[l]$ in SFB n. The residual signal $r[l]$ is then obtained as

$$r[k] = \begin{cases} c_n[l] - \lfloor \mathrm{thr}(i_n[l]) \rfloor, & i_n[l] \neq 0 \\ c_n[l], & i_n[l] = 0. \end{cases} \quad (3)$$

BPGC is then applied to this residual signal in each SFB which forms a fine grained SLS coding pattern.

3. BPGC USING ARITHMETIC CODING

In BPGC, first the binary representation of the residual coefficients are obtained. Then each bit-plane of the coefficients, from MSB to LSB, are coded using arithmetic coding [5]. The arithmetic coding uses a probability assignment rule that is derived from the statistical properties of an exponentially distributed source [2]. Closed forms have been obtained in [2] for bit-plane probabilities of a Laplacian source which were adopted in BPGC [9]. In [5] it was shown that the IntMDCT coefficients can be well modeled by a Laplace distribution. Based on this assumption, it is assumed that the residual signal can be approximately considered an exponential source. Therefore, the probabilities obtained for Laplacian source are used for the residual signal as well.

For each SFB the standard deviation (or λ parameter) of the residual signal is estimated by the mean of the signal. The Lambda parameters are then used in the closed-forms obtained in [2] to give the probabilities. The bit-planes are scanned from MSB to LSB and coded using arithmetic coding. In the case of non-core SLS coding, where there is no perceptual core, the bit-plane coding is applied directly to the input IntMDCT coefficients instead of the residual signal. In both cases if all the bit-planes are received by the decoder, lossless quality is achieved. However, by truncating the bit-planes a fine grained scalable lossy-to-lossless (SLS) coding is obtained.

In the next section we will discuss the properties of the residual signal and we will address some issues regarding the bit-plane probabilities used in BPGC.

4. STATISTICAL PROPERTIES OF THE RESIDUAL SIGNAL

Consider Fig. 1. A specific quantization interval was shown for $t_i \leq x < t_i + 1$, where t_i is the beginning threshold of that interval. The pdf of a Laplacian source is given by

$$f_X(x) = \frac{1}{2}\lambda e^{-\lambda|x|}, \quad (4)$$

where $\lambda = \sqrt{2}/\sigma$ and σ is the standard deviation of the signal. In BPGC the absolute value of the residual signal is coded and there is a sign bit which is sent separately. In fact, the one sided non-negative signal is considered which has a pdf in the

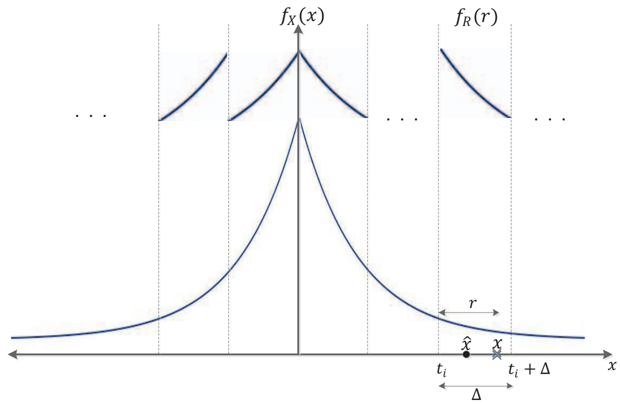

Fig. 1. pdf properties of the residual signal

form of

$$\begin{cases} f_X(x) = \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x<0 \end{cases} \quad (5)$$

Probability of such an exponential signal to be in a specific interval $p_i = p(t_i \leq x < t_{i+1})$ can be obtained by

$$p_i = (e^{-\lambda t_i} - e^{-\lambda t_{i+1}}). \quad (6)$$

Considering $\Delta = t_{i+1} - t_i$, we have

$$p_i = e^{-\lambda t_i}(1 - e^{-\lambda\Delta}). \quad (7)$$

The residual for this interval is $r = x - t_i$ and we have

$$f_X(x|t_i \leq x < t_{i+1}) = \frac{f_X(x)}{p_i} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda t_i}(1 - e^{-\lambda\Delta})}. \quad (8)$$

The conditional pdf of the residual in this specific interval is

$$f_R(r|t_i \leq x < t_{i+1}) = \frac{f_X(x|t_i \leq x < t_{i+1})}{r'(x)}\bigg|_{x=t_i+r} = \frac{\lambda e^{-\lambda r}}{1 - e^{-\lambda\Delta}}. \quad (9)$$

If we had a uniform quantizer, pdf of the residual for such a quantizer would be

$$\begin{aligned} f_R(r) &= \sum_i p_i f_R(r|t_i \leq x < t_{i+1}) = \frac{\lambda e^{-\lambda r}}{1 - e^{-\lambda\Delta}} \sum_i p_i \\ &= \frac{\lambda e^{-\lambda r}}{1 - e^{-\lambda\Delta}}, \end{aligned} \quad (10)$$

which is equal to the conditional pdf obtained for a specific interval due to the memoryless property of the Laplace distribution. But this is not a true assumption because in the AAC core a non-uniform quantizer is used which we will discuss later. Also, it is very important to note that this pdf is not in the typical form of an exponential source (5). The reason is that the residual signal is a signal bounded within the interval ($0 \leq r < \Delta$). Now, let us calculate the mean of the residual signal. The mean can be obtained by

$$\bar{r} = E[r] = \int_0^\Delta r f_R(r)\,dr = \frac{1}{\lambda} - \frac{\Delta}{e^{\lambda\Delta} - 1}. \quad (11)$$

which is also the same for the conditional expectation of an interval of width Δ. It can be seen that the residual mean is different from that of an exponential signal (for which mean$= 1/\lambda$). In a special case where $\Delta \rightarrow \infty$ (unbounded exponential), $\bar{r} = 1/\lambda$. From Equations (10) and (11) it is clear that assuming an exponential source for the residual signal and obtaining the pdf parameters using the statistics (here mean) of the observed residual signal leads to inaccurate bit-plane probabilities: In BPGC the λ (or equivalently L) parameter for the residual is approximated by $1/\bar{r}$ in each SFB. On the other hand, the pdf of the residual is directly dependent on the quantization step sizes. The other issue is that the step sizes are not equal in AAC quantization. Having the interval thresholds from (2) the step sizes can be obtained by

$$\Delta_n(i) = t_n(i+1) - t_n(i), \tag{12}$$

where $\Delta_n(i)$ is the step size of a quantization index i for SFB n and $t_n(i)$ and $t_n(i+1)$ are the corresponding interval thresholds. It can be seen that the the step sizes are dependent not only on the scalefactor parameter used in each SFB, but also on the quantization interval index. Therefore the bit-plane probabilities should be a function of the input source statistics, together with the scalefactors and quantization indices. In the following we will propose a more straightforward approach for calculating the bit-plane probabilities of an exponential source bounded within an arbitary interval. The other advantage of the proposed method is that the clipping effect can be easily considered in the calculations, which is specifically useful for bit-plane coding in the SLS non-core mode.

5. BIT-PLANE PROBABILITY CALCULATIONS

Consider a signal bounded within the interval $0 \leq x \leq \Delta$ (see Fig. 2). We want to create a binary representation for the N most significant non-zero bits of this signal. The MSB of this signal b_1 (we index the bit-planes from MSB to LSB) can be considered to be zero if the signal is $0 \leq x < \Delta/2$, and one if $\Delta/2 \leq x < \Delta$. The probability of $b_1 = 0$ then can be obtained by

$$P_1 = \frac{p(0 \leq x < \Delta/2)}{p(0 \leq x < \Delta)}, \tag{13}$$

and $b_1 = 1$ with the probability of $Q_1 = 1 - P_1$. Note that, while the IntMDCT coefficients are integers, the step size Δ and the residue are not necessarily. Now consider the second bit b_2. The probability of $b_2 = 0$ can be written as (Fig. 2)

$$\begin{aligned} P_2 = &\frac{p(0 \leq x < \Delta/4)}{p(0 \leq x < \Delta/2)} p(0 \leq x < \Delta/2) \\ &+ \frac{p(\Delta/2 \leq x < 3 \times \Delta/4)}{p(\Delta/2 \leq x < \Delta)} p(\Delta/2 \leq x < \Delta). \end{aligned} \tag{14}$$

Due to the memoryless property of the Laplace distribution the two fractions in the above equation are equal. Also, since

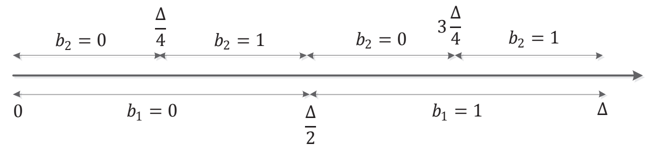

Fig. 2. Obtaining the N most significant non-zero bits of a bounded signal

$p(0 \leq x < \Delta/2) + p(\Delta/2 \leq x < \Delta) = 1$ we get

$$P_2 = \frac{p(0 \leq x < \Delta/4)}{p(0 \leq x < \Delta/2)}. \tag{15}$$

Using the same strategy the kth significant bit can be obtained by

$$P_k = \frac{p(0 \leq x < \Delta/2^k)}{p(0 \leq x < \Delta/2^{k-1})}. \tag{16}$$

Using (7) in (16) gives

$$P_k = \frac{1 - e^{\frac{-\lambda\Delta}{2^k}}}{1 - e^{\frac{-\lambda\Delta}{2^{k-1}}}} = \frac{1}{1 + e^{\frac{-\lambda\Delta}{2^k}}}, \tag{17}$$

and

$$Q_k = 1 - P_k = \frac{1}{1 + e^{\frac{\lambda\Delta}{2^k}}}. \tag{18}$$

Equations (17) and (18) give the bit-plane probabilities for an exponential signal bounded in the interval Δ.

Now let us consider a case where we have an unbounded exponential signal (appropriate for bit-plane coding in non-core SLS coding). In general, for a binary representation of such a signal an infinite number of bits are required. However, in practice the signal is bounded before coding, which means clipping happens. Assume the signal is bounded to be $0 \leq x \leq X_m$, and let us define $P_e = p(x > X_m) = e^{-\lambda X_m}$. Using a similar strategy used in (14) we can obtain

$$\begin{cases} P_k^c = P_k'(1 - P_e) \\ Q_k^c = Q_k'(1 - P_e) + P_e, \end{cases} \tag{19}$$

where the superscript c stands for the clipping consideration and

$$\begin{cases} P_k' = \frac{1}{1 + e^{-\frac{\lambda X_m}{2^k}}} \\ Q_k' = \frac{1}{1 + e^{\frac{\lambda X_m}{2^k}}}. \end{cases} \tag{20}$$

6. BIT-PLANE CODING USING THE OBTAINED PROBABILITIES

We saw in the previous section that, bit-plane coding of the residual signal should be performed considering the quantization parameters used in the core layer. In other words the probabilities are a function of the scalefactors and quantization indices. Figure 3 shows step sizes of the AAC quantizer versus quantization indices (which are in the range of

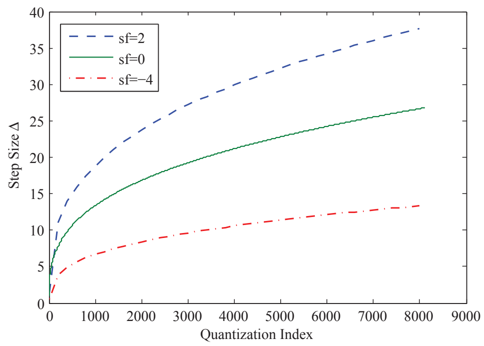

Fig. 3. AAC quantizer step size vs. quantization index for three different scalefactor values.

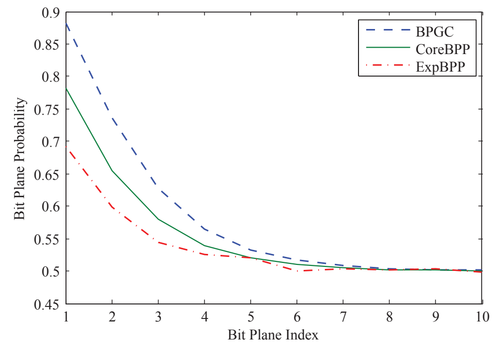

Fig. 4. Bit-Plane probability vs. bit-plane index for the residual signal ($\lambda\Delta = 2.56$)

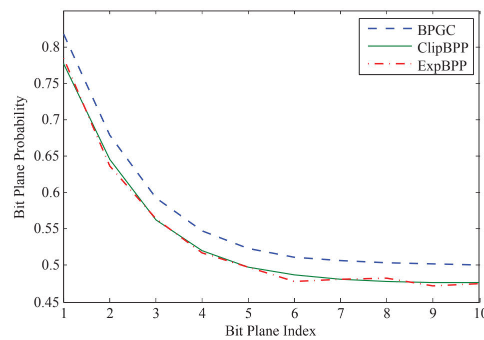

Fig. 5. Bit-Plane probability vs. bit-plane index for the input signal ($\lambda X_m = 3$)

($0 \leq |i| \leq 8191$) for three different scalefactor values. It can be seen that increasing the scalefactor increases the step sizes. Also, due to the compression used in the quantization, for each specific scalefactor the step sizes increase by going to the higher indices. Therefore, the bit-plane probabilities obtained in (17) can be expressed by

$$P_k(s,i) = \frac{1}{1 + e^{-\frac{\lambda\Delta(s,i)}{2^k}}}, \tag{21}$$

where s and i stand for the scalefactor and quantization index respectively. $\Delta(s,i)$ can be obtained from (12) and (2). Having this set of probabilities, the bit-plane coding can be performed using arithmetic coding based on Algorithm 1. This

Algorithm 1: Core-based Residual Bit-plane Coding

- Take input IntMDCT coefficients $c_n[l]$.
- Calculate the λ parameter for each scalefactor band n using $\frac{1}{\lambda} = E[|c_n|] = (\sum_l |c_n[l]|)/L_n$.
- Take the scalefactors and quantization indices from the core layer.
- Calculate the bit-plane probabilities using (21).
- Scan from MSB plane to LSB plane and perform arithmetic coding for bit-plane symbols using the obtained probabilities.

is a modified version of the algorithm proposed in [2], where 1) instead of the residual we use the input coefficients to approximate the λ parameter, 2) the new set of core-based probabilities are used instead. A similar algorithm can be used for non-core bit-plane coding of the input coefficients, considering the clipping effect. In this case the line regarding the core parameters should be omitted and equations in (19) should be used for bit-plane probabilities.

7. SIMULATION RESULTS

We compare our proposed core-based method for obtaining the residual signal bit-plane probabilities (referred to as CoreBPP) with the method proposed in [2], which is used by BPGC in SLS coding. We took the IntMDCT coefficients of the same 15 audio files used in [5] as input signal. These coefficients were quantized using (1) for different step sizes. The bit-plane probabilities of the residual signals were computed using the two methods and were compared to what was obtained experimentally (ExpBPP). Experiments showed that for $\lambda\Delta = \sqrt{2}\Delta/\sigma < 10$ the new method leads to clearly better results. In fact, for smaller $\lambda\Delta$ values the difference between the methods becomes greater. Such a comparison is shown in Fig. 4 for a sample case of $\lambda\Delta = 2.56$.

Also, Fig. 5 shows a comparison of bit-plane probabilities of the input signal for $\lambda X_m = 3$ when 1) clipping is considered using (19) and referred to as ClipBPP, 2) without considering clipping [2]. The experiments showed that for $\lambda X_m \leq 4$ the difference between the two methods become considerable.

8. CONCLUSION

An alternative approach was proposed for calculating the bit-plane probabilities of the residual signal for perceptual-core mode of MPEG-4 SLS coding. The new approach matches the quantization performed in the core layer. Therefore, the probabilities are estimated considerably better for the residual signal compared to what used in BPGC. Also, the clipping effect can be easily considered in the proposed method, which is useful for non-core mode of SLS coding.

9. REFERENCES

[1] D. Ning and M. Deriche, "A bitstream scalable audio coder using a hybrid WLPC-wavelet representation," in *Proc. IEEE ICASSP*, Apr. 2003, vol. 5, pp. V–417–20.

[2] H. Huang, H. Shu, and S. Rahardja, "Bit-plane arithmetic coding for Laplacian source," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 3358 –3361.

[3] D. H. Kim, J. H. Kim, and S. W. Kim, "Scalable lossless audio coding based on MPEG-4 BSAC," in *Proc. 113th AES Conv.*, Oct. 2002, Paper Number:5679.

[4] S. Kim, S.H. Park, and Y. B. Kim, "Fine grain scalability in MPEG-4 Audio," in *Proc. 111th AES Conv.*, Nov. 2001, Paper Number:5491.

[5] R. Yu, S. Rahardja, L. Xiao, and C. C. Ko, "A fine granular scalable to lossless audio coder," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1352 –1363, Jul. 2006.

[6] M. Movassagh, J. Thiemann, and P. Kabal, "Joint entropy-scalable coding of audio signals," in *Proc. IEEE ICASSP*, Mar. 2012, pp. 2961 –2964.

[7] *Information Technology—Coding of Audio-Visual Objects—Part 3: Audio—Subpart 4: General Audio Coding (GA)*, ISO/IEC std. ISO/IEC JTC1/SC29, 14496-3:2005, 2005.

[8] *Information Technology—Coding of Audio-Visual Objects—Part 3: Audio—Amd. 3: Scalable Lossless Coding (SLS)*, ISO/IEC std. ISO/IEC JTC1/SC29, 14496-3:2005/Amd.3:2006, 2006.

[9] R. Yu, C.C. Ko, S. Rahardja, and X. Lin, "Bit-plane Golomb coding for sources with Laplacian distributions," in *Proc. IEEE ICASSP*, Apr. 2003, vol. 4, pp. 277–80.

[10] A. Aggarwal, S. L. Regunathan, and K. Rose, "Efficient bit-rate scalability for weighted squared error optimization in audio coding," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1313 –1327, Jul. 2006.

[11] E. Ravelli, V. Melkote, T. Nanjundaswamy, and Kenneth Rose, "Joint optimization of base and enhancement layers in scalable audio coding," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 4, pp. 711–724, Apr. 2013.

[12] 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.