A Low Complexity Design of Psycho-Acoustic Model for MPEG-2/4 Advanced Audio Coding

Shih-Way Huang, Tsung-Han Tsai, Member, IEEE, Liang-Gee Chen, Fellow, IEEE

Abstract —*The paper presents a new low complexity design of Psycho-Acoustic Model (PAM), which is the key technology for MPEG-2/4 Advanced Audio Coding (AAC) encoding. The real-time constraint of MPEG AAC leads to a heavy computational bottleneck on today's portable devices. To overcome this problem, design analysis and optimization of PAM are addressed at the algorithmic level. First, the dominant calculation of spreading function is replaced with small look-up tables. Second, Modified-Discrete-Cosine-Transform-based (MDCT-based) PAM with block type decision in frequency domain is proposed to reduce the overall AAC encoder complexity and maintain the quality. This technique is different from the traditional methods. The computational complexity of PAM is reduced by 82% in total. The proposed design could be implemented in a real-time MPEG-2/4 AAC stereo encoder at Low Complexity profile and bitrate128 kb/s with computational complexity of less than 20 MOPS while maintaining CD quality*[1].

Index Terms — Audio encoder, Low complexity, MPEG Advanced Audio Coding, Psycho-acoustic Model.

I. INTRODUCTION

Portable electronic devices such as smart mobile phones, digital cameras, PDA, and digital audio devices with audio playback (players) and recording (recorders) have been attractive particularly because of the prevalence of MP3 audio files. MP3 is the popular name of MPEG-1 audio [1] and MPEG-2 audio [2] Layer III. Moreover, the so-called MP3's successor, MPEG-2 Advanced Audio Coding (AAC), finalized as an international standard in 1997 [3], [4], was developed to achieve a higher quality than that of MP3. Later, MPEG-4, finished in 1998, integrated AAC as general audio coding [5]. AAC follows the same basic coding paradigm as Layer III but uses new coding tools to improve quality at low bit-rates. Compared with Layer III, AAC reaches the same sound quality as MP3 at about 70% of the bitrate [6].

When AAC is implemented on portable devices, real-time and low power designs are still challenges. In conventional approaches, real-time AAC encoders were achieved on PC-based [7]-[10] or DSP-based [11] platforms. They constructed AAC encoding on floating-point programmable processors of large computational ability and thus yielded large power consumption for portable devices. Takamizawa *et al*. proposed a good implementation on a low power DSP-based AAC encoder [12]. However, due to limited memory capacity, this technique did not support block-switching mode and would degrade the quality. Gayer *et al*. presented that the average processing power requirement for an unoptimized AAC encoder is about 100 MHz on various commercial fixed-point programmable processors [13]. It is obvious that the processing power requirement is still a challenge for portable devices.

The AAC encoder is composed of several coding tools. Among the coding tools, Quantization Loop (Q Loop) and Psycho-Acoustic Model (PAM) dominate the computational complexity [14]. Because the Q Loop in AAC follows the same algorithms as those in MP3, there are already many studies on the reduction in iterations for the Q Loop [15]-[17]. Liu *et al*. proposed that only 10% of the original iterations are achieved by the fast and approximated algorithms [15]. As for PAM, which plays a very important role in AAC encoding, not only determines sound quality, but also exerts a great influence on the computational complexity. Oh *et al*. tried to eliminate PAM to simplify encoder of MPEG-1/2 Layer III [17]. However, it could only provide high sound quality at the expense of higher bitrate than 256 kb/s. Takamizawa *et al*. [8] presented that the complex Fast Fourier Transform (FFT) spectrums in PAM can be replaced by the original Modified Discrete Cosine Transform (MDCT) spectrums in the filterbank to reduce the complexity of whole encoder; [10], [12] also adopted the similar idea. Nevertheless, there is still a problem of the block type decision left, which is essential to an AAC encoder and highly correlated with audio quality for transient signal coding [4]. In [8], they solved it in time domain but did not specify the technical details. In [10], they solved it in time domain by calculating the difference between neighboring subblocks and admitted that tuning efforts were required to maintain the quality. However, [18] shows that block type decision purely in time domain does not perform so reliably. In [12], they did not implement the block type decision.

This paper presents a new low complexity design of PAM at the algorithmic level. We propose a look-up table method to replace the calculation of spreading function, which is computationally dominant in PAM. This method reduced effectively the computational complexity, and the table size could be optimized to only about one-third of the original size. Moreover, referring to [8], we adopted the idea to replace complex-FFT computations in PAM with MDCT computations. The improvement was that our approach

[1] Shih-Way Huang and Liang-Gee Chen are with the Graduate Institute of Electronics Engineering, Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, R.O.C. (email: shihway@video.ee.ntu.edu.tw, lgchen@video.ee.ntu.edu.tw)

Tsung-Han Tsai is with Department of Electrical Engineering, National Central University, Chung-Li, Taiwan, R.O.C. (email: han@ee.ncu.edu.tw)

Contributed Paper
Manuscript received September 3, 2004

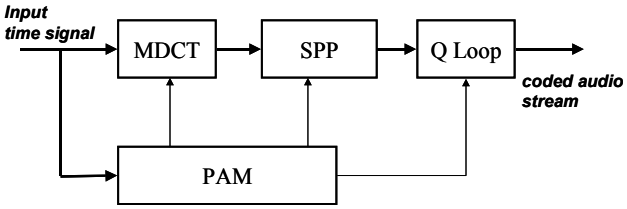

Fig. 1. Block diagram of an AAC encoder

employed block type decision by Perceptual Entropy (PE) in frequency domain instead of time domain to maintain the sound quality. Therefore, only two transforms could be achieved compared with the original three transforms, one for the filterbank and two for PAM. With these methods, the computational complexity of PAM was reduced by 82%, and that of AAC encoding was reduced accordingly. The results lead to a real-time MPEG-2/4 Low Complexity profile stereo encoder at 128 kb/s below 20 MOPS with CD quality maintained. The proposed methods are suitable for implementation in general programmable processors, DSP, and dedicated hardwares.

The rest of this paper is organized as follows. In Section II and Section III, the analyses of AAC encoding algorithms and the dominant PAM are given. It shows the optimization of PAM is necessary and critical. In Section IV, our optimization methods are proposed to reduce the computational complexity of PAM. Experiments and simulation results are given in Section V. Finally, we summarize our major contributions.

II. Overview of AAC Algorithms

In this section, the AAC algorithms are briefly reviewed. Fig. 1 shows the block diagrams of an AAC encoder. PAM calculates a masking threshold, which is the maximum distortion energy masked by the signal energy for each coding partition. Moreover, it also calculates a block type, which determines the block length used in other parts of the encoding algorithms. Meanwhile, Modified Discrete Cosine Transform (MDCT) transforms input audio samples in time domain into spectrums in frequency domain. The frequency spectrums are transferred to Spectral Processing (SPP) that includes some tools such as Temporal Noise Shaping (TNS) and joint coding to remove redundancies. The spectrums are then non-uniformly quantized based on the masking threshold and available number of bits to minimize the audible quantization error in the Q Loop.

In order to realize the computational complexity of an AAC encoder, we analyzed the AAC LC stereo encoding algorithms in Millions Operations Per Second (MOPS) at the output bitrate of 128 kb/s and 64 kb/s with CD-quality input files of sampling rate 44.1 kHz. In our analytical model, we assumed that every arithmetic operation such as addition, subtraction, multiplication, division, logarithmic operation, trigonometric function, power operation, etc. is defined as one single operation. It will avoid differences among various implementations and simplify the analyses of complex algorithms in AAC. Because MDCT is structuralized, we analyzed it in windowing, pre-twiddle operations, complex-FFT, and post-twiddle operations according to [19]. TNS and joint coding algorithms were estimated directly from the algorithms. As for the non-uniform quantization and noiseless coding, we first analyzed the operations in one inner iteration and one outer iteration using the reference software of ISO/IEC 14496-3 [7]. Then we calculated the numbers of inner and outer iterations, multiplied the operations of the two iterations respectively, and thus estimated the operations in the Q Loop. We observed that the numbers of inner and outer iterations were much dependent on the audio files and compression bitrate. Similarly, PAM was analyzed and was explained in the following section.

Table I
Computational Complexity Analyses of the AAC LC Stereo Encoder

	Bitrate 128 kb/s		Bitrate 64 kb/s	
	Complexity (MOPS)	%	Complexity (MOPS)	%
MDCT	2.1	2.3%	2.1	1.4%
PAM	51.8	57.3%	51.8	35.1%
Q Loop	33.0	36.5%	90.1	61.1%
SPP	3.5	3.9%	3.5	2.4%
Total	90.4	100.0%	147.5	100.0%

The result of computational complexity analyses is displayed in Table I, where PAM and Q Loop dominate up to 90%. Because PAM is not affected by the bitrate as Q loop is, the complexity of PAM is in larger proportion at higher bitrate. As mentioned in the introduction, there are many papers working on the reduction of Q Loop. Accordingly, PAM becomes the bottleneck and requires optimization especially when Q Loop is optimized using the previous techniques.

III. PAM Algorithms and Analyses

In this section, the computational behaviors of PAM are reviewed and the reason why it dominates the AAC is explained. According to [5], Fig. 2 illustrates the block diagram of PAM in 13 steps. Fig. 3 shows the details of the dashed block in Fig. 2. In Steps 1-2, PAM normalizes the time-domain samples as input and transforms them into frequency-domain spectrums of real part *r(w)* and imaginary part *i(w)* by FFT. Real-part spectrums are used to calculate the partitioned energy and imaginary-part spectrums are used to calculate the weighted unpredictability measure *c(b)* in Steps 3-4. In Step 5, both partitioned energy and unpredictability are convolved with the spreading function in order to estimate the effects across the partitioned bands. Tonality index is estimated in Step 6 to indicate whether a signal is tonal-like. Signal-to-Noise Ratio (SNR) is calculated in Step 7 and the masking partitioned energy threshold *nb(b)* is calculated in

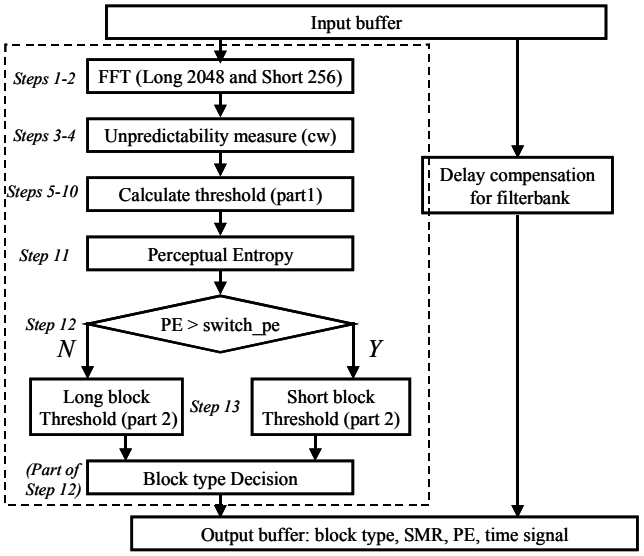

Fig. 2. Block diagram of the PAM in [5]

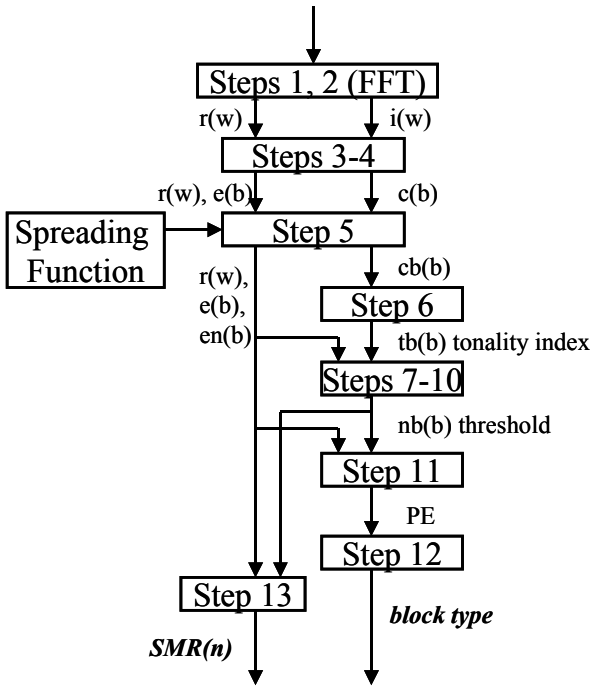

Fig. 3. Detailed block diagram of the PAM from the 13 steps in [5]

Steps 8-10 to estimate the masking curve. Perceptual Entropy (PE) is calculated in Steps 11-12 to determine the block type [20] used in the MDCT, Q loop, and SPP. Block type decision requires detecting whether there is a transient signal in the frame. Finally, Signal-to-Mask Ratio (SMR) is computed in Step 13 as output. *w*, *b*, and *n* indicate indices in the spectral line domain, the threshold calculation partition domain, and the coder scalefactor band domain, respectively.

The results of PAM analyses were shown in Table II. We observed that Step 2, Step 4, and Step 5 were critical algorithms with high computational complexity. Step 2 includes windowing input samples, calculating a standard forward 1024-point complex-FFT for LONG block type and eight 128-point forward complex-FFTs for SHORT block type, and then transforming the magnitude and phase parts of complex-FFT outputs into polar representation. Step 4 calculates the weighted unpredictability measure *c(b)*. Step 2 and Step 4 involve various special functions such as arctangent, sine/cosine, and square root; therefore, they are hard to implement. Step 5 includes the calculation of spreading function and convolutions, and it was the most critical algorithm. Therefore, we optimized the three steps to reduce the computational complexity of PAM. The details are specified in the following sections.

IV. Design and Optimization

Because algorithm-inherent dissipation can be reduced by operation reduction, we propose two methods to simplify the PAM algorithms according to the concept.

A. Method 1: Look-up Table Method for Spreading Function

In Step 5, the calculation of the spreading function, shown in Fig. 4, includes a series of complex operations such as square roots, power of tens, squares, and division. The spreading function is repeatedly calculated within one frame because both partitioned energy and unpredictability are convolved with the spreading function in order to estimate the effects across the partitioned bands. Moreover, it was repeatedly calculated in every frame. For example, if a 10-second music is encoded into a CD-quality audio file, the spreading function is calculated for 10 (seconds) x 44100/1024 (frames per second) x 2 (stereo) x 70 x 70 (square of partition bands) times, assuming only block type LONG is used, i.e., about 4.2-million-times calculation is required. This was the reason why Step 5 involved large operations in computational complexity.

From the definition of spreading function, we found that the results of spreading function were simply affected by bark values, and these bark values were simply affected by sampling rates and the block types used. In other words, the calculation of spreading function was only dependent on sampling rates and the block types used. Hence, we reduced the calculation of spreading functions by replacing it with

Table II
Computational Complexity Analyses of the PAM

Step	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Complexity (MOPS)	0.09	6.36	0.71	3.00	40.28	0.10	0.14	0.10	0.03	0.10	0.04	0.00	0.78	51.75
%	0.2%	12.3%	1.4%	5.8%	77.8%	0.2%	0.3%	0.2%	0.1%	0.2%	0.1%	0.0%	1.5%	100.0%

TABLE III
REDUCTION RATE OF COMPUTATIONAL COMPLEXITY

	Original	By Method 1		By Method 2		By Method 1 and Method 2	
	[7],[9], [11],[13]	-		-		Proposed	
	Complexity (MOPS)	Complexity (MOPS)	Reduction Rate	Complexity (MOPS)	Reduction Rate	Complexity (MOPS)	Reduction Rate
PAM Steps 2-4	10.1	-	-	4.8	52.8%	4.8	52.8%
PAM Step 5	40.3	7.5	81.4%	36.1	10.4%	3.3	91.8%
PAM total steps	51.8	19.0	63.3%	42.3	18.3%	9.5	81.7%
AAC whole system	90.4	57.6	36.3%	78.8	12.8%	46.0	49.1%

```
Spreading function (bark value i, bark value j)
{
if (j>=i)   tmpx = 3.0*(j-i)
     else   tmpx = 1.5*(j-i)
tmpz = 8 * minimum(((tmpx-0.5)^2-2(tempx-0.5)),0)
tmpy = 15.811389 + 7.5(tmpx+0.474) – 17.5(1.0+(tmpx+0.474)^2)^0.5
if(tmpy < -100)   then return 0
     else   return 10^((tmpz+tmpy)/10)
}
```

Fig. 4. The calculation of the spreading function

fixed coefficients and a look-up table method. This approach could easily be implemented in dedicated logics, ROM or RAM. Table III shows the reduction rate via the look-table method (Method 1). Step 5 was reduced dramatically by 81.4% and yielded a 63.3% reduction in PAM. AAC, accordingly, was reduced by 36.3%. In addition, the table size, which represented the sum of square of partitioned bands for both LONG and SHORT block types, was considered. Because most of the coefficients in the table were zeroes, the table was sparse and its size could be further decreased. Table IV is an example of table size for sampling rate at 44100 Hz. Only about one-third of the original size is required if the table is optimized. Therefore, we proposed a linear-array technique to optimize the table with little overhead, and the detail is in [21].

TABLE IV
THE SPREADING FUNCTION LOOK-UP TABLE SIZE (EXAMPLE OF SAMPLING RATE 44.1 KHZ, WITH LONG AND SHORT BLOCK TYPE)

	LONG	SHORT	Total	%
Original	4900	1764	6664	100.0%
Optimized non-zeroes	1277	566	1843	27.7%

B. Method 2: MDCT-based PAM with block type decision in frequency domain

While there are two transforms FFTs (of 2048-point LONG block type and eight 256-point SHORT block type) inside the PAM, there is another transform MDCT outside the PAM to transform the input samples into spectrums. Table V details the comparisons between FFT and MDCT. In [8], the MDCT spectrums can replace the complex-FFT spectrums and the tonality can be calculated by Spectral Flatness Measure (SFM) [22] due to the lack of phase-information. Because FFT is replaced, the block type decision is left to implement. If the block type decision is not carefully designed, the quality would degrade. Thus, the combination between block type decision and the flow of MDCT-based PAM is a great problem.

Block type decision is usually performed in time domain or frequency domain. The previous works of MDCT-based PAM

TABLE V
COMPARISONS BETWEEN FFT AND MDCT

	FFT	MDCT
Target	time-to-frequency analyses in PAM	time-to-frequency analyses in the encoder (filterbank)
Input samples	Long: 2048, Short: eight 256 (current and previous frames)	Long: 2048, Short: eight 256 (current and previous frames)
Output spectrums	Long: 1024 real and 1024 imaginary parts	Long: 1024 real parts
	Short: 128 real and 128 imaginary parts	Short: 128 real parts
Windowing	Hann Window	Sine or KBD window
Block switching	none	Long, Start, Short, Stop
Number	2 kernels (Long and Short in parallel)	1 kernel (according to the block type)

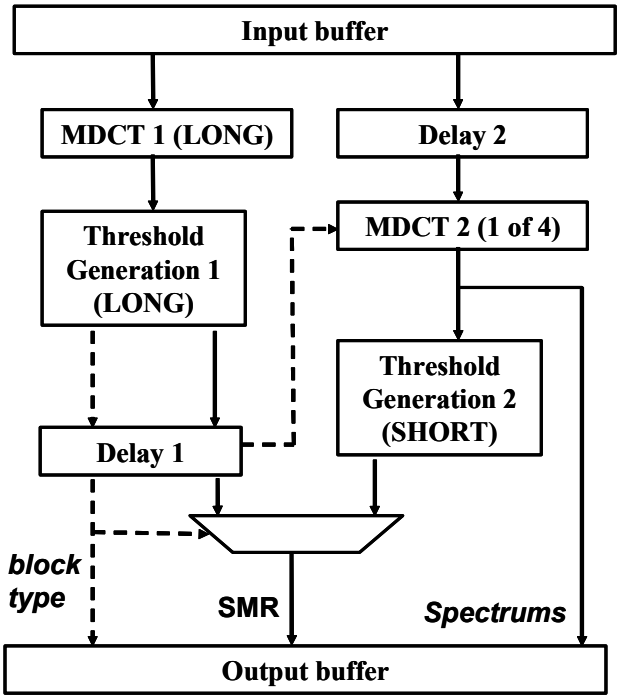

Fig. 5. Block diagram of the proposed MDCT-based PAM

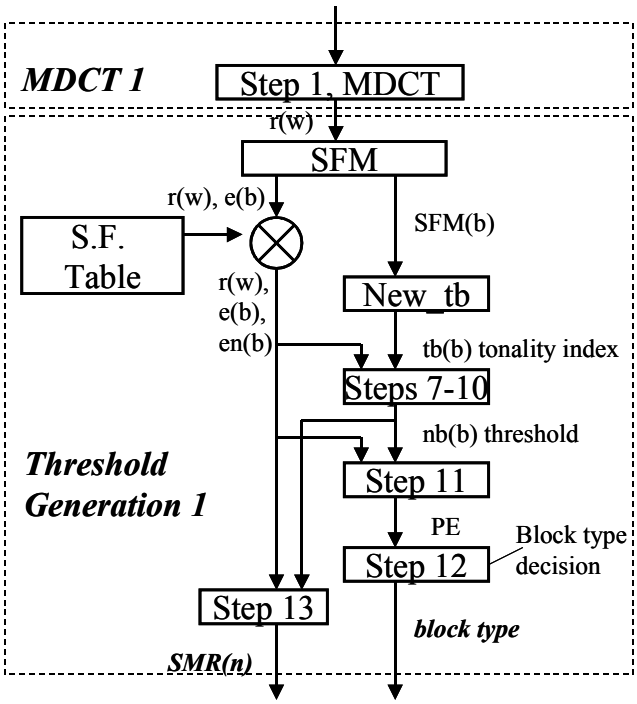

Fig. 6. Block diagram of MDCT 1 and Threshold Generation 1

flow cannot guarantee good quality with the block type decision in time domain alone. Moreover, [18] shows that experimental data suggest that a big surge in PE is always due to pre-echo conditions and the block type switching logic is thus activated. Therefore, it states that the detection in frequency domain by the threshold calculation works more reliably than that in time domain alone using the energy calculation-based methods.

In order to reduce the complexity as well as maintain the sound quality, we decided to employ MDCT-based PAM with the block type decision in frequency domain instead of time domain. The idea is using two MDCTs in parallel. One is MDCT 1 (LONG block type) to generate the thresholds, detect the transients, and decide the block type, whereas the other MDCT 2 is to calculate the MDCT spectrums. MDCT 1 is one-frame ahead of MDCT 2 in sequence because MDCT 1 is used to predict and determine the block type of MDCT 2. Fig. 5 shows our MDCT-based PAM. In addition to the input and output buffers, it is composed of two MDCTs, two Threshold Generations, two delays, one denormalization, and one multiplexer. Compared with the original FFT-based PAM (see Fig. 2), where LONG FFT is a 2048-point and SHORT FFTs are eight 256-point calculations, we omitted eight 256-point SHORT FFTs and replaced the LONG FFT with the LONG MDCT (MDCT 1) and then calculated its corresponding thresholds in Threshold Generation 1. Moreover, the other MDCT (MDCT 2) should be calculated according to either one of the four block types. It can replace the original MDCT outside the PAM. Before the process of SPP and Q Loop, the MDCT 2 spectrums are denormalized because Step 1 normalizes the input time signal. Threshold Generation 2 calculates corresponding thresholds only when SHORT block

TABLE VI
DIFFERENT SCHEMES OF MDCT-BASED PAM

	Transform	Block type decision	Threshold generation	Complexity	Quality degradation	Examples
Original	MDCT (1 of 4) FFT (LONG) FFT (SHORT)	In frequency domain (by PE)	LONG SHORT	High	Reference	[7], [9], [11], [13]
Scheme 1	MDCT (1 of 4) MDCT (LONG)	In frequency domain (by PE)	LONG SHORT	Medium	Little	Proposed Method 2
Scheme 2	MDCT (1 of 4)	In time domain	1 of 4	N/A	Medium	[8], [10]
Scheme 3	MDCT (LONG)	None	LONG	Low	Much	[12]

TABLE VII
COMPUTATIONAL COMPLEXITY OF THE AAC ENCODER AFTER OPTIMIZATION (128 KB/S, LC STEREO)

	Optimized complexity (MOPS)
Proposed PAM	9.5
Q Loop	3.3*
SPP	3.5
Total	16.3

*Q Loop is assumed to be optimized to 1/10 of the original from [15].

type is determined. The output threshold is multiplexed by the block type. Note that the two delays are required because MDCT 1 and Threshold Generation 1 are to predict and determine the block type in MDCT 2 and Threshold Generation 2. Thus, MDCT 1 and Threshold 1 are one-frame ahead of MDCT 2 and Threshold Generation 2 in timing sequence.

Fig. 6 illustrates the MDCT 1 and Threshold Generation 1 in detail. Threshold Generation includes SFM, corresponding calculation of tonality (*New_tb*) [22], the spread of partitioned energy, and Steps 7-13. In Fig. 6, we modified the original steps (see Fig. 3). Steps 2-4 were replaced with MDCT. Because there was no phase-information (imaginary part), Step 5 required some modification that only the partitioned energies were convolved with the spreading functions. Step 6 was also modified where SFM was used to generate the tonality index from the MDCT spectrums. Threshold Generation 2 was the same as Threshold Generation 1 except for omission of Steps 11-12 because it was unnecessary to determine the block type here.

Table VI shows different schemes of MDCT-based PAM, including the previous works and the proposed Method 2. Because the technical details such as the time-domain block type decision of [8] and [10] were not presented, we could not estimate the complexity of their designs. However, we could infer that the proposed MDCT-based PAM with block type in frequency domain had the advantages on quality. Table III shows the reduction of computational complexity with Method 1 and Method 2 individually and totally. In Method 2, the reduction rate is about 18.3% mainly because there is no calculation of imaginary parts of spectrums. By these two methods, PAM was dramatically reduced by 81.7%, and AAC, 49.1%, accordingly. Compared with the previous works, the proposed PAM had much lower computational complexity.

TABLE VIII
TESTED AUDIO FILES AND THEIR CHARACTERISTICS

Audio files	Characteristics
castanets	Castanets
speech	Male speech
velvet	Drum
frer07_1	Electronic tune
trpt21_2	Trumpet
horn23_2	Horn
gspi35_2	Glockenspiel
quar48_1	Quartet
spfe49_1	Female speech in English

Therefore, it is suitable to be implemented in the low complexity real-time AAC stereo encoder. The results in Table VII show that the encoder at 128 kb/s and Low Complexity profile could be achieved below 20 MOPS if the Q loop was assumed as mentioned.

V. EXPERIMENTS AND RESULTS

We induced the significant reduction in computational complexity by means of the proposed PAM. All the experiments were simulated with ISO/IEC 14496-5 reference software [7] on a Pentium IV 2.4 GHz PC platform. The audio files were castanets (considered non-sustained, percussive sound), human speech, velvet (non-sustained, percussive sound), and the others are from Sound Quality Assessment Materials [23] such as artificial signals, single instruments, vocal, and speech. Their characteristics are listed in Table VIII. These audio files were CD-quality sampled at 44100 Hz with 16 bits for stereo channels. The AAC encoder was simulated at 128 kb/s and under Low Complexity Profile.

A. Profiling the Reduction Rate

In order to prove the effect of the proposed PAM, we simulated by profiling the time before and after optimizations of each block in AAC. Table IX shows the results. The reduction rates in different audio files are almost correspondent with the analytical value 81.7% from Table III regardless of the proportion of PAM, and consequently the reduction rate of AAC is high. Note that the reduction rates of AAC in castanets and in velvet are not high because the computational complexity of Q Loop occupies a lot in the file. However, in our assumption, Q Loop could be optimized by other works, but it is beyond our discussion.

TABLE IX
SIMULATED REDUCTION RATES BY THE PROPOSED PAM

	castanets	speech	velvet	frer07_1	trpt21_2	horn23_2	gspi35_2	quar48_1	spfe49_1
PAM	82.0%	81.7%	80.6%	77.6%	81.1%	78.5%	80.1%	74.9%	82.5%
AAC	26.0%	40.5%	9.7%	68.2%	68.9%	68.7%	61.6%	43.3%	59.3%

TABLE X
ENCODING AUDIO QUALITY OF DIFFERENT MDCT-BASED PAM SCHEMES WHERE QI STANDS FOR QUALITY IMPROVEMENT

	SHORT (%)	Scheme 1 (Proposed)		Scheme 3	
		QI (ODG)	QI (NMR)	QI (ODG)	QI (NMR)
castanets	28%	-0.04	0.07	-1.05	-1.01
speech	6%	-0.16	-0.03	0.20	0.02
velvet	64%	0.00	-0.72	-0.46	-4.48
frer07_1	0%	0.50	2.75	0.50	2.75
trpt21_2	0%	0.00	0.00	0.00	0.00
horn23_2	0%	0.00	0.00	0.00	0.00
gspi35_2	2%	0.03	0.75	0.12	-0.16
quar48_1	0%	0.03	-0.03	0.03	-0.03
spfe49_1	12%	0.00	-0.07	0.51	0.36

TABLE XI
SIMULATED ENCODING TIME OF THE PROPOSED ENCODER

	Length (s)	Original (s)	Proposed (s)	Reduction Rate
castanets	6.6	9.92	8.59	13.4%
speech	16.7	15.72	11.44	27.2%
velvet	11.9	26.75	26.41	1.3%
frer07_1	35	18.88	7.5	60.3%
trpt21_2	17.8	10.41	4.31	58.6%
horn23_2	25.9	14.59	5.74	60.7%
gspi35_2	19	11.89	5.84	50.9%
quar48_1	23.1	17.98	10.19	43.3%
spfe49_1	19.2	15.13	9.25	38.9%

B. Encoding Quality

Because MDCT-based PAM was used to replace the original FFT-based PAM, the outputs such as SMR and decisions of block type had changed. The encoding quality would change, too. To guarantee the quality of the proposed PAM, we compared it with that of the original PAM. Before this, the assessment of sound quality should be mentioned. In addition to informal tests on subjective listening, we referred to ITU-R Recommendation BS. 1387 [24] as objective audio quality measurement method, where Objective Difference Grade (ODG) is defined. ODG is a measure of quality comparable to the Subjective Difference Grade (SDG). ODG and SDG are both calculated as the difference between the quality rating of the reference and the test signal, which is measured with the five-point scale defined in ITU-R BS.1116 [25]. They have a range of [-4;0] where -4 stands for very annoying difference and 0 stands for imperceptible difference between reference and test signal. Besides ODG, we also adopted Noise-to-Mask-Ratio (NMR), which is a ratio of the noise introduced by the encoder to the allowed masking threshold. Negative NMR represents that the noise is masked. Thus, the more negative NMR stands for the better sound quality. Next, we compared the sound quality of the compressed signal of the original encoder by FFT-based PAM with sound quality of the uncompressed audio in ODG and NMR, represented as ODG (FFT-based) and NMR (FFT-based). In the same way, we obtained ODG (MDCT-based) and NMR (MDCT-based) by replacing FFT-based PAM with MDCT-based PAM. For simplicity, Quality Improvement (QI) is defined to stand for the sound quality improvement by the proposed MDCT-based PAM. The positive value of QI represents that the proposed MDCT-based PAM is better in sound quality, whereas the negative value of QI represents that the proposed MDCT-based PAM is worse in sound quality.

QI (ODG) = ODG (MDCT-based) – ODG (FFT-based)

QI (NMR) = NMR (FFT-based) – NMR (MDCT-based)

According to the different schemes of MDCT-based PAM from Table VI, we compared encoding quality with that achieved by the original FFT-based PAM respectively. We could not compare Scheme 2 in [8] and [10] and just compared Scheme 3, which did not implement block type decision. Table X shows the results, where all the values of ODG and NMR were calculated by EAQUAL [26]. In column SHORT (%), the percentage of determined SHORT block type in each file is estimated by the original FFT-based PAM. It is obvious that the sound quality of the proposed MDCT-based PAM (Scheme 1) and that of the original FFT-based PAM are almost the same in most of the audio files. The sound quality degrades a lot in Scheme 3 when the tested files are dominated by SHORT block types (transient signals) such as in castanets and velvet, as has been expected. Note that in the speech files, Scheme 3 performs better than the others, even the original FFT-based PAM. The reason may be that the block type decision is not very proper for human speech, but this is beyond our discussion. Moreover, Fig. 7 shows the left-channel waveform and spectral views of castanets, one encoded by the original FFT-based PAM and the other by the proposed MDCT-based PAM. In waveform view, the horizontal axis stands for time, and the vertical axis stands for amplitude. In spectral view, the horizontal axis stands for time, and the vertical axis stands for frequency (0 to 20 kHz). Both views show that they are almost the same, too.

C. Encoding Time

In order to obtain the realistic reduction of the proposed PAM for the whole AAC encoder, we considered effective encoding time (run time) for evaluating the reduction. Therefore, we simulated the encoding time of each audio file before and after the optimization. The results are given in Table XI. Column Length means the original length (in seconds) of the test file, column Original means the encoding

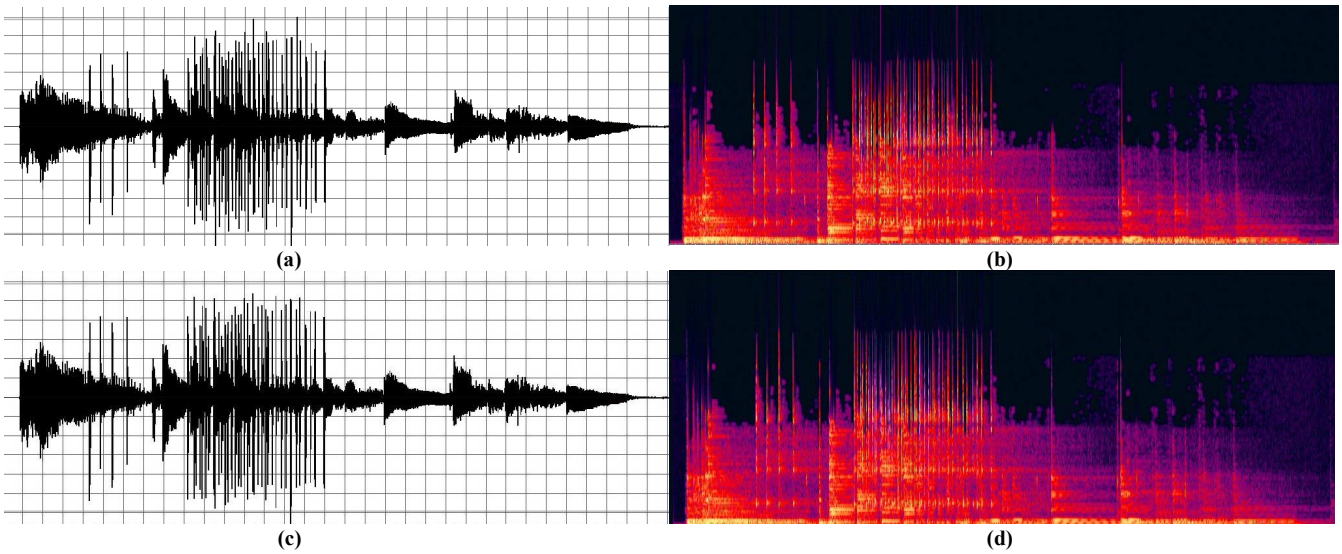

Fig. 7 (a), (b) are the left-channel waveform and spectral views encoded by the original FFT-based PAM, and (c), (d) are the left-channel waveform and spectral views encoded by the proposed MDCT-based PAM. They are almost the same.

time of the original encoder, and column Proposed means the encoding time of the optimized encoder. From the table, the reduction rate is dependent on the proportion of PAM in each file. When the proportion of PAM is larger, the reduction rate of AAC is higher. Note that the reduction rates in castanets and velvet are low because of low proportions of PAM and high proportions of Q Loop.

VI. CONCLUSION

In the paper, we designed a low complexity PAM for the MPEG-2/4 AAC encoder. We first constructed an analytical model to evaluate the computational complexity of AAC encoding algorithms. According to the analytical model, PAM became the most dominant when we assumed that Q Loop was optimized. In order to optimize PAM algorithms, we proposed a look-up table method to replace the calculation of spreading functions. It could dramatically reduce the computational complexity by 63.3% with only about one-third of the original table size. Moreover, referring to [8], we employed MDCT-based PAM to reduce the computational complexity with block type decision in frequency domain instead of time domain with the sound quality maintained. The two methods reduced the PAM complexity by 81.7%. The proposed design leads to a real-time MPEG-2/4 stereo encoder at Low Complexity Profile and bitrate 128 kb/s with the complexity below 20 MOPS while maintaining CD quality. This work can be implemented in general programmable processors, DSP, and dedicated hardwares.

REFERENCES

[1] MPEG. Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s, part 3: Audio, International Standard IS 11172-3, ISO/IEC JTC1/SC29 WG11, 1992.

[2] MPEG. Information Technology – generic coding of moving pictures and associated audio, part 3: Audio, International Standard IS 13818-3, ISO/IEC JTC1/SC29 WG11, 1994.

[3] MPEG. MPEG-2 Advanced Audio Coding, AAC, International Standard IS 13818-7, ISO/IEC JTC1/SC29 WG11, 1997.

[4] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, Y. Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding," *J. Audio Eng. Soc.*, Vol. 45, No. 10, 1997 October.

[5] MPEG. Information technology – Coding of audio-visual objects – Part 3: Audio, International Standard IS 14496-3, ISO/IEC JTC1/SC29 WG11, 1999.

[6] K. Brandenburg, "MP3 and AAC explained," AES 17th International Conference on High Quality Audio Coding, Italy, Sep. 2-5, 1999.

[7] MPEG-4 Version 1 Reference Software (ISO/IEC 14496-5:2000) Available: http://www.tnt.uni-hannover.de/project/mpeg/audio/ftp/

[8] Y. Takamizawa, T. Nomura, M. Ikekawa, "High-quality and processor-efficient implementation of an MPEG-2 AAC encoder," in *Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, Page(s): 985 –988.

[9] D. H. Kim, D. H. Kim, J. H. Chung, "Optimization of MPEG-4 GA AAC on general PC," in *Proceedings of the 44th IEEE 2001 Midwest Symposium on Circuits and Systems*, Vol. 2, pp. 923-925.

[10] I. Dimkoviae, D. Milovanoviae, Z. Bojkoviae, "Fast software implementation of MPEG advanced audio encoder," *2002 14th International Conference on Digital Signal Processing*, Vol. 2, Page(s): 839 –843.

[11] D. Huang, X. Gong, D. Zhou, T. Miki, S. Hotani, "Implementation of the MPEG-4 Advanced Audio Coding encoder on ADSP-21060 SHARC," in *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems*, Vol. 3, page(s): 544 –547.

[12] Y. Takamizawa, T. Okumura, T. Nomura, M. Ikekawa, and I. Kuroda, "20mW MPEG-2/4 AAC LC stereo encoder on a 16-bit DSP," Workshop and Exhibition on MPEG-4, San Jose, California, June 25-27 2002.

[13] M. Gayer, M. Lohwasser, M. Lutzky, "Implementing MPEG Advanced Audio Coding and Layer-3 encoders on 32-bit and 16-bit fixed-point processors," presented at the AES 115th Convention, New York, Oct. 10-13, 2003.

[14] T. Tsai, S. Huang, L. Chen, "Design of a low power psychoacoustic model co-processor for MPEG-2/4 AAC LC stereo encoder," in *Proceedings of the 2003 IEEE International Symposium on Circuits and Systems*, Vol. 2, page(s): 552 –555, May 25-28, 2003.

[15] C. Liu, C. Chen, W. Lee, S. Lee, "A fast bit allocation method for MPEG layer III," in *1999 IEEE International Conference on Consumer Electronics*, pages 22-23.

[16] C. Yang, S. Chen, "New static and dynamic search algorithms for fast MP3 bit allocations," in *2003 IEEE International Conference on Multimedia and Expo*, Vol. 1, pages 77-80.
[17] H. Oh, J. Kim, C. Song, Y. Park, D. Youn, "Low power MPEG/audio encoders using simplified psychoacoustic model and fast bit allocation," *IEEE Transactions on Consumer Electronics*, Vol. 47, No. 3, pp. 613 – 621, Aug. 2001.
[18] M. Kahrs, K. Brandenburg, *Applications of digital signal processing to audio and acoustics*. Kluwer Academic Publishers, 1998, p.59.
[19] P. Duhamel, Y. Mahieux, J. P. Petit, "A fast algorithm for the implementation of filter banks based on time domain aliasing cancellation," in *Proceedings of the 1991 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2209-2212.
[20] M. Bosi, Richard E. Goldberg, *Introduction to digital audio coding and standards*. Kluwer Academic Publisher Press, 2003, pp. 295-296.
[21] S. Huang, T. Tsai, L. Chen, "Memory reduction technique of spreading function in MPEG AAC encoder", in Proc. of the 7th Int. Conference on Digital Audio Effects (DAFX-04), Naples, Italy, October 5-8, 2004, submitted for publication
[22] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas in Communications*, Vol. 6, No 2, pp. 314-323, Feb., 1988.
[23] SQAM - Sound Quality Assessment Material. Available: http://www.tnt.uni-hannover.de/project/mpeg/audio/sqam/
[24] ITU-R Recommendation BS. 1387: "Method for objective measurements of perceived audio quality," July 2001.
[25] ITU-R Recommendation BS. 1116, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," International Telecommunication Union, Geneva, Switzerland, 1994.
[26] Available: http://www.mp3-tech.org/programmer/sources/eaqual.tgz

Shih-Way Huang was born in Braunschweig, Germany, in 1975. He received the B.S. and M.S. degrees in Electrical Engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1997 and 1999, respectively. He is currently a Ph.D. student in Department of Electrical Engineering, National Taiwan University. His major research interests include audio coding algorithms and DSP architecture design.

Tsung-Han Tsai was born in Chunghua, Taiwan, R.O.C. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1990, 1994, and 1998 respectively. Dr. Tsai was an Instructor (1994-1998) and an Associate Professor (1998-1999) of the department of electrical engineering at Hwa Hsia College of Technology and Commerce. From 1999 to 2000, he was an Associate Professor of electronic engineering at Fu Jen University. Currently, he is an Assistant Professor in the department of electrical engineering at National Central University. He is also a member of IEEE and Audio Engineering Society (AES). Dr. Tsai has been awarded 8 patents and more than 70 referred papers published in international journals and conferences. His research interests include VLSI signal processing, video/audio coding algorithms, DSP architecture design, wireless communication, and System-On-Chip design.

Liang-Gee Chen (F'01) was born in Yun-Lin, Taiwan, R.O.C., in 1956. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from National Cheng Kung University, Tainan, Taiwan, R.O.C., in 1979, 1981, and 1986, respectively.

In 1988, he joined the Department of Electrical Engineering, National Taiwan University, Taiwan, R.O.C. During 1993–1994, he was a Visiting Consultant in the DSP Research Department, AT&T Bell Labs, Murray Hill, NJ. In 1997, he was a Visiting Scholar of the Department of Electrical Engineering, University of Washington at Seattle. Currently, he is a Professor at National Taiwan University. His current research interests are DSP architecture design, video processor design, and video coding system.

Dr. Chen has served as Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY since 1996, Associate Editor of IEEE TRANSACTIONS ON VLSI SYSTEMS since January 1999, and Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: ANALOG AND DIGITAL SIGNAL PROCESSING. He was the Associate Editor of the Journal of Circuits, Systems and Signal Processing in 1999, served as the Guest Editor of The Journal of VLSI Signal Processing-Systems for Signal, Image and Video Technology, and was the General Chairman of the 7th VLSI Design/CAD Symposium and the 1999 IEEE Workshop on Signal Processing Systems: Design and Implementation. He received the Best Paper Award from the ROC Computer Society in 1990 and 1994, the Long-Term (Acer) Paper Awards annually from 1991 to 1999, the Best Paper Award of the Asia-Pacific Conference on Circuits and Systems in the VLSI design track in 1992, the Annual Paper Award of Chinese Engineer Society in 1993, the Outstanding Research Award from the National Science Council, and the Dragon Excellence Award from Acer, both in 1996. He is currently the elected IEEE Circuits and Systems Distinguished Lecturer for 2001–2002. He is a member of Phi Tan Phi.