Audio Engineering Society

Convention Paper 6183

Presented at the 117th Convention
2004 October 28–31 San Francisco, CA, USA

MPEG-4 Scalable to Lossless Audio Coding

Rongshan Yu[1], Ralf Geiger[2], Susanto Rahardja[1], Juergen Herre[3], Xiao Lin[1], and Haibin Huang[1]

[1] Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore
{rsyu,rsusanto,linxiao,hhuang}@i2r.a-star.edu.sg

[2] Fraunhofer IDMT, Ilmenau, Germany
ggr@idmt.fraunhofer.de

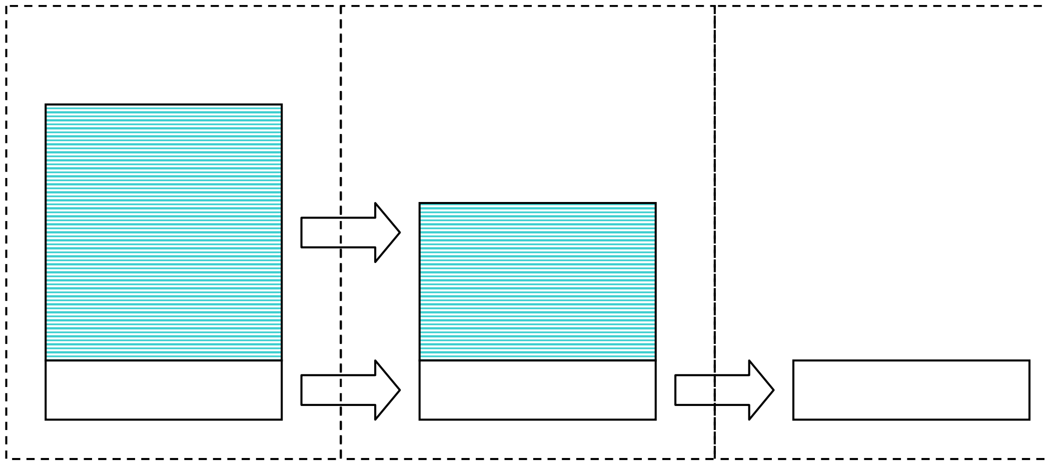
[3] Fraunhofer IIS, Erlangen, Germany
hrr@iis.fraunhofer.de

ABSTRACT

1. INTRODUCTION

2. APPLICATION SCENARIOS

2.1. Studio Operations



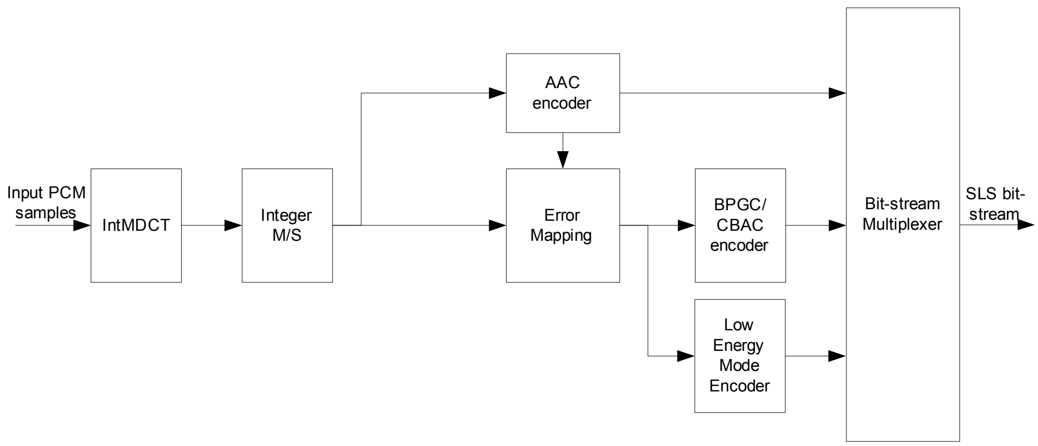
2.2. Archival

2.3. Consumer disc-based delivery

2.4. Internet Delivery of Files

2.5. Streaming Applications

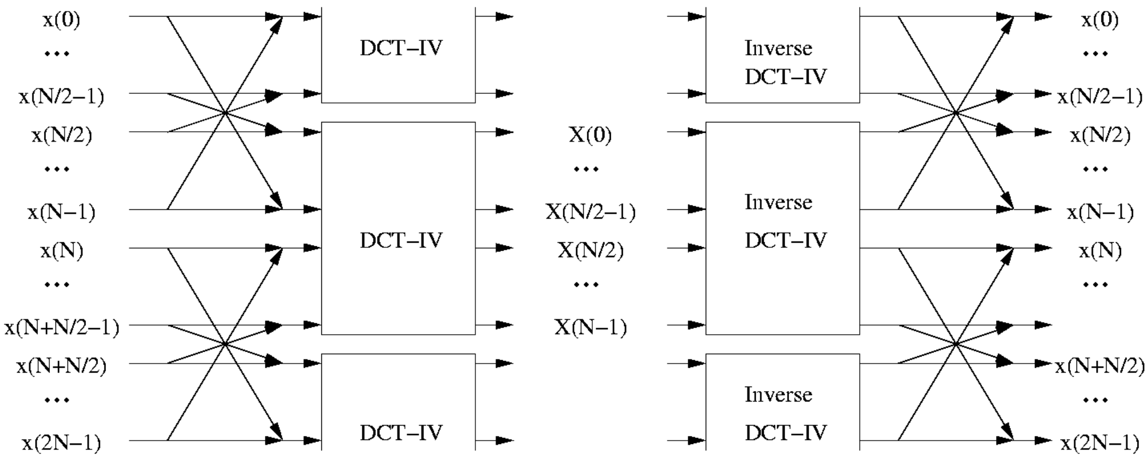
2.6. Broadcast Contribution/Distribution Chain



3. STRUCTURE OF MPEG-4 SLS

4. INTEGER MDCT

4.1. Decomposition of MDCT



$$X\ m\ = \sqrt{\frac{\ }{N}} \sum_{k=}^{N-} w\ k\ x\ k \qquad \frac{k + \ + N \quad m + \ \pi}{N}$$

$m = \quad N -$

-
-

$$\begin{pmatrix} x\ k \\ x\ N - \ - k \end{pmatrix} \mapsto$$

$$\begin{pmatrix} & -\frac{w\ N - \ -k\ -}{w\ k} \\ & \end{pmatrix} \begin{pmatrix} & \\ -w\ k & \end{pmatrix} \begin{pmatrix} & -\frac{w\ N - \ -k\ -}{w\ k} \\ & \end{pmatrix} \begin{pmatrix} x\ k \\ x\ N - \ -k \end{pmatrix}$$

$k = \quad N \quad -$

4.2. Integer DCT-IV

T I

$$\begin{pmatrix} T & \\ & T^{-} \end{pmatrix} = \begin{pmatrix} -I & \\ T^{-} & I \end{pmatrix} \begin{pmatrix} I & -T \\ & I \end{pmatrix} \begin{pmatrix} & I \\ I & T^{-} \end{pmatrix}$$

T T^{-}

$$\begin{pmatrix} I & \\ & -I \end{pmatrix}\begin{pmatrix} I & -\sqrt{\ }DCTIV_N \\ & I \end{pmatrix}\begin{pmatrix} I & \\ -\sqrt{\ }DCTIV_N & I \end{pmatrix}$$

$$\begin{pmatrix} I & -\sqrt{\ }DCTIV_N \\ & I \end{pmatrix}\begin{pmatrix} I & --I \\ & I \end{pmatrix}\begin{pmatrix} I & \\ I & I \end{pmatrix}$$

$$= \begin{pmatrix} -\sqrt{\ }DCTIV_N & -\sqrt{\ }DCTIV_N \\ -\sqrt{\ }DCTIV_N & --\sqrt{\ }DCTIV_N \end{pmatrix}$$

$$= \begin{pmatrix} -\sqrt{\ }\ I & -\sqrt{\ }\ I \\ -\sqrt{\ }\ I & --\sqrt{\ }\ I \end{pmatrix}\begin{pmatrix} DCTIV_N & \\ & DCTIV_N \end{pmatrix}$$

4.3. Noise Shaping

5. ERROR MAPPING

$c[k]$ s

$i[k]$

$$\left|thr(i[k])\right| \le \left|c[k]\right| < \left|thr(i[k])\right| + \Delta(i[k]),\ k \in s,$$

$thr(i[k])$ $\Delta(i[k])$

$i[k]$

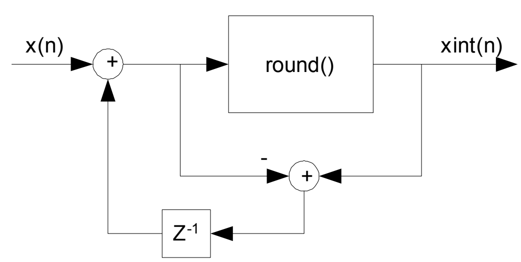
$$thr(i[k]) = \begin{cases} (i[k])\left[\sqrt{\ }^{scale\ \ factor(s)}\left(\left|i[k]\right| - C\right)^{\ /\ }\right] & i[k] \ne \\ & i[k] = \end{cases}$$

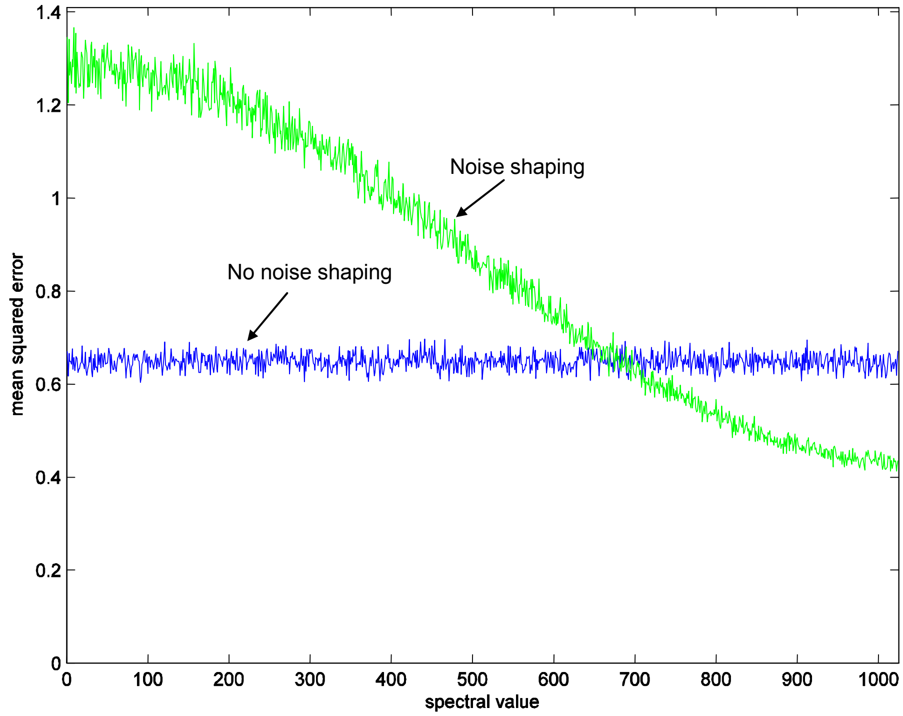
$$\Delta(i[k]) = thr\left(\left|i[k]\right| + \ \right) - thr\left(\left|i[k]\right|\right).$$

$scale\ \ factor[s]$

s

$C =$





$e[k]$

$c[k]$

$e[k]=c[k]-c[k]$

$c[k]=E\{c[k]\ \ i[k]\}$ $E\{\bullet\}$

$e[k]$

$c[k]$

$e[k]$

$$e[k]=\begin{cases} c[k]-\lfloor thr(i[k])\rfloor & i[k]\neq \\ c[k] & i[k]= \end{cases}$$

$k=$.

$\lfloor\bullet\rfloor$ $\mathbb{R}\to\mathbb{Z}$

$c[k]$

$i[k]\neq$

$e[k]$ $i[k]$

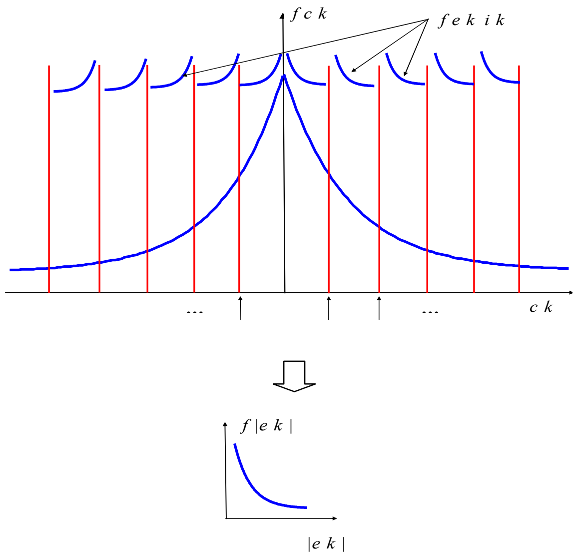
$c[k]$

$e[k]$

6. SCALABLE CODING OF INTMDCT SPECTRAL DATA

6.1. Perceptual Bit-Plane Coding

$\mathbf{e}=\{e[\ ]\ \ \ e[N-\ ]\}$



e

^{T} $T = M$

e

N **e**

$e[k]$ **e**

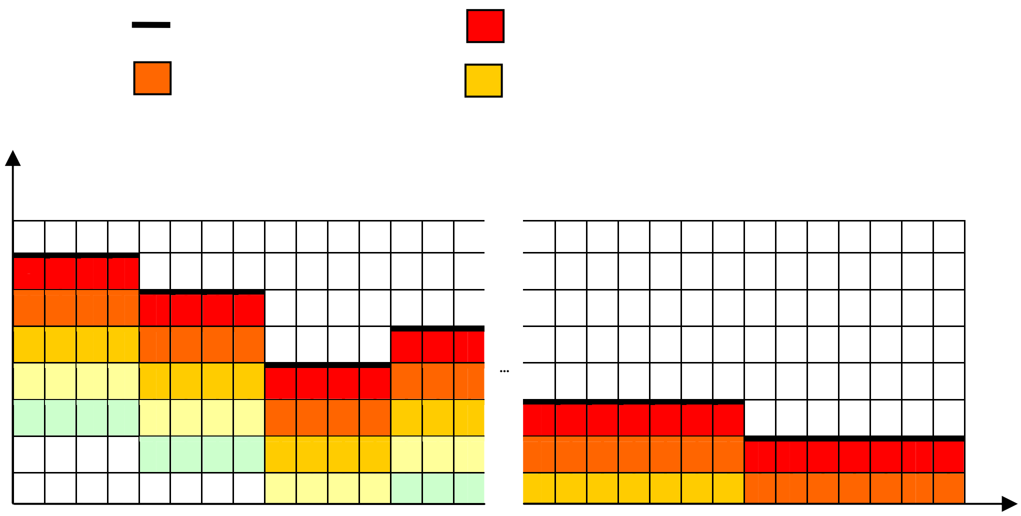
$$e[k] = \quad s[k] - \quad \cdot \sum_{j=}^{M-} b[k \;\; j] \cdot \;^{j} \quad k = \qquad N-$$

$$s[k] \stackrel{\Delta}{=} \begin{cases} & e[k] \geq \\ & e[k] < \end{cases} \quad k = \qquad N-$$

$b[k \;\; j] \in \{ \quad \}$ $i =$ k

M **e**

$^{M-} \leq \qquad \{|e[k]|\} < \;^{M} \quad k = \qquad N-$



$$lazy \quad bp = \quad \{L' \in \mathbb{Z} \quad ^{L'+} N \geq A\}$$

$N \qquad A$

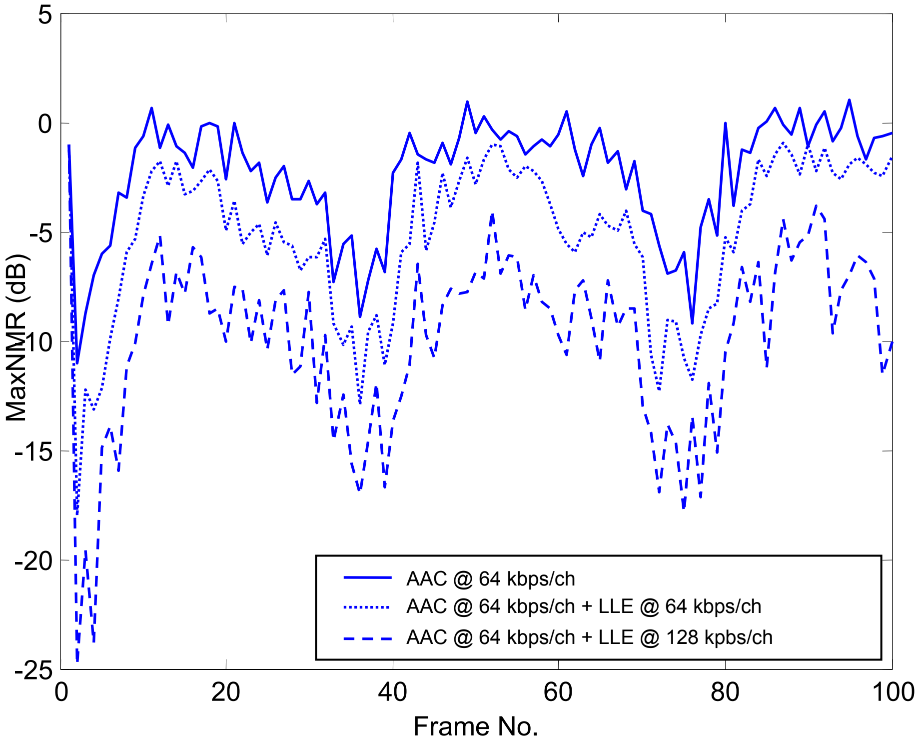
6.2. Bit-Plane Coding with BPGC/CBAC

j

$Q(j)$

$$Q(j) = \begin{cases} \frac{}{+ \quad ^{j-lazy \quad bp}} & j \geq lazy \quad bp \\ - & j < lazy \quad bp \end{cases},$$

$lazy\_bp$



- *frequency band (FB)*

FB

FB *FB*

- *distance to lazy (D2L)*

lazy_bp

$$D\ L = \begin{cases} -j + lazy\ \ bp & j - lazy\ \ bp \geq - \\ & else \end{cases}$$

$DL\ \ \geq$

- *significant state (SS)*

$sig\ \ \ cx\ k\ \ j$

$b\ k\ \ j\ \ \ \ k = \ \ \ \ \ N -$

$$sig\ \ \ cx(k\ \ j) = \{sig(k - \ \ \ j)\ \ sig(k - \ \ \ j)\ \ sig(k + \ \ \ j)\ \ sig(k + \ \ \ j)\}$$

$sig\ \ k\ \ j$

$$sig(k\ j) = \begin{cases} c_j[k] = \\ c_j[k] \neq \end{cases}$$

$sig(k\ j)$ k

$c_j\ k$

$c\ k$ j

$$c_j\ k\ = \sum_{i=j+}^{M-} b\ k\ i\ \cdot\ ^{i} + thr(i[k])$$

$sig\ \ cx\ k\ j$ $sig\ \ cx\ k\ j\ =$

$sig\ \ cx\ k\ j\ =$

$sig\ \ cx(k\ j)$

sig_core

sig_core

$$sig\ \ core(k) = \begin{cases} c[k]\ is\ from\ an\ insignificant\ sfb \\ c[k]\ is\ from\ a\ significant\ sfb \end{cases}$$

$sig\ \ core(k) =$

$$\leq e[i] \leq \Delta(i[k])$$

$\Delta(i[k])$

$$quant\ \ int(k\ j) = \begin{cases} e_j\ k\ +\ ^{j+} \leq \Delta(i[k]) \\ e_j\ k\ +\ ^{j} \leq \Delta(i[k]) < e_j\ k\ +\ ^{j+} \\ \Delta(i[k]) < e_j\ k\ +\ ^{j} \end{cases}$$

$e_j[k]$ $e[k]$

$j+$ $quant\ \ int(k\ j) =$

e

e

e

$$\mathbf{e} = \{e[\ ]\quad e[N-\ ]\}$$

$$e[k] = \begin{cases} (\ s[k] -\ )\left[\sum_{j=T}^{M-} b[k\ j]\ ^{j} + f_k(L\ T)\right] & \sum_{j=T}^{M-} b[k\ j]\ ^{j} \neq \\ & else \end{cases}$$

$k =\quad N-$

$f_k(L\ T)$

$$f(L\ T) = \sum_{j=}^{T-} Q(j)\ ^{j}$$

6.3. Low Energy Mode Coding

$lazy\_bp$

$e[k]$	$b[pos]$
^{M} _ ^{M} _	
pos	

$e[k]$

$\mathbf{b} = \{b[\ ]\ b[\ ]\ \dots\ b[pos]\ \dots\}$

$\mathbf{b}$ pos $e[k]$ $lazy\_bp$

$\{b\ pos\ =\ \} = \quad \{e[k] > pos\ \ e[k] \geq pos\}$

$\leq pos < \ ^{M}$

$b[pos]$

pos $\mathbf{b}$ $lazy\_bp$

pos $lazy\_bp$

7. PERFORMANCE

8. CONCLUSION

9. REFERENCES

Proc. ICASSP