

Áudio

Profa. Débora Christina Muchaluat Saade

debora@midia.com.uff.br

Áudio

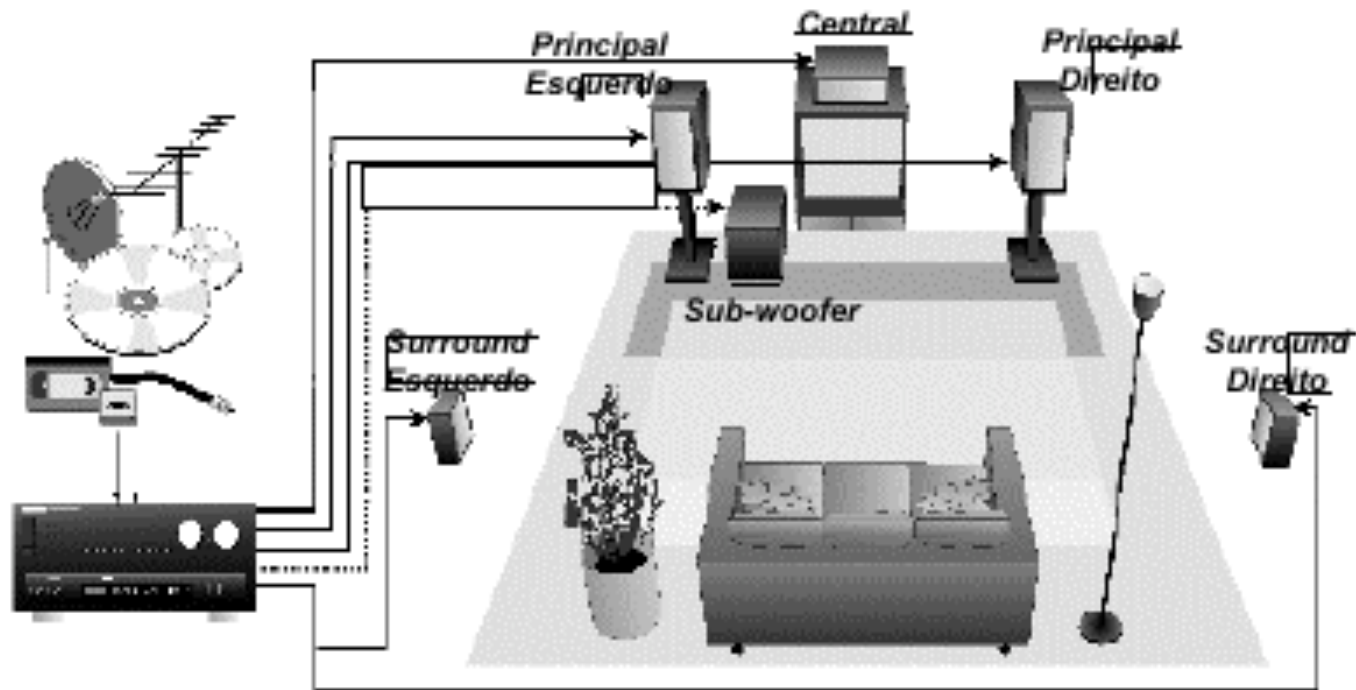
- ✓ **Sinal de Voz**
 - *50Hz a 10KHz*
- ✓ **Música**
 - *15Hz a 20KHz*
 - *Estéreo: 2 canais*
- ✓ **Codificadores de Sinal de Voz**
- ✓ **Codificadores de Áudio Genérico**

MPEG-2 Áudio

✓ MPEG-2 BC:

- *Backward Compatible*
- *Compatível com MPEG-1 (layers I, II e III)*
- *Extensão multicanal – até 6 (5.1) canais (esquerda, direita, centro, 2 canais surround, subwoofer)*
- *Taxas de amostragem menores*
 - 16, 22 e 24 KHz
- *Taxas de bits menores*
 - 32 a 256 Kbps – layer I
 - 8 a 160 Kbps – layers II e III

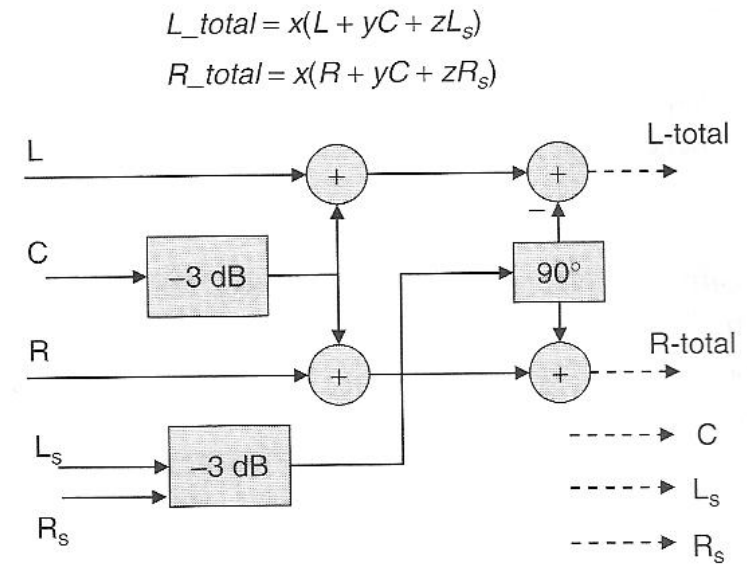
Áudio multicanal



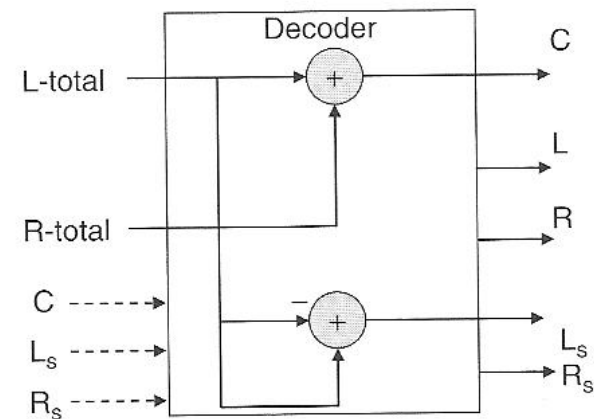
MPEG-2 BC

✓ Permite down-mixing

- *Conversão de 5 canais em 2*



(a)



(b)

MPEG-2 Áudio

✓ MPEG-2 AAC

- *Advanced Audio Coding*
- *Não é compatível com MPEG-1*
- *Bem mais complexo*
- *Amostragem em até 96 KHz*
- *Áudio de altíssima qualidade a 64 Kbps*
- *Adotado como núcleo do codificador geral de áudio do MPEG-4*

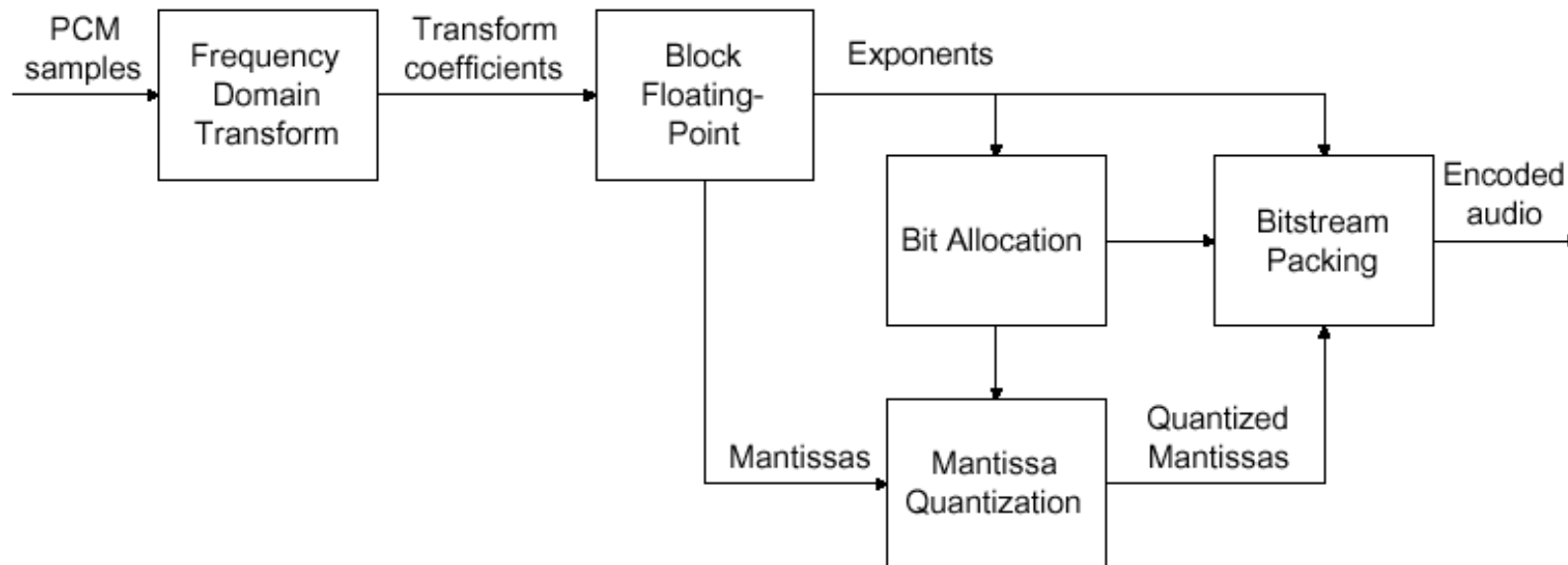
Dolby Digital

- ✓ **www.dolby.com**
- ✓ **Dolby AC-3**
- ✓ **ATSC A/52 – Advanced Television Systems Committee**
- ✓ **Formato usado em DVD e no padrão HDTV norte-americano**
 - *MPEG áudio é opcional em DVD*
- ✓ **Taxas de amostragem de 48, 44.1, 32 KHz**
- ✓ **Taxas de bits de 32 a 640 Kbps**
- ✓ **Até 6 canais (esquerdo, direito, centro, 2 canais surround e subwoofer)**

Dolby AC-3

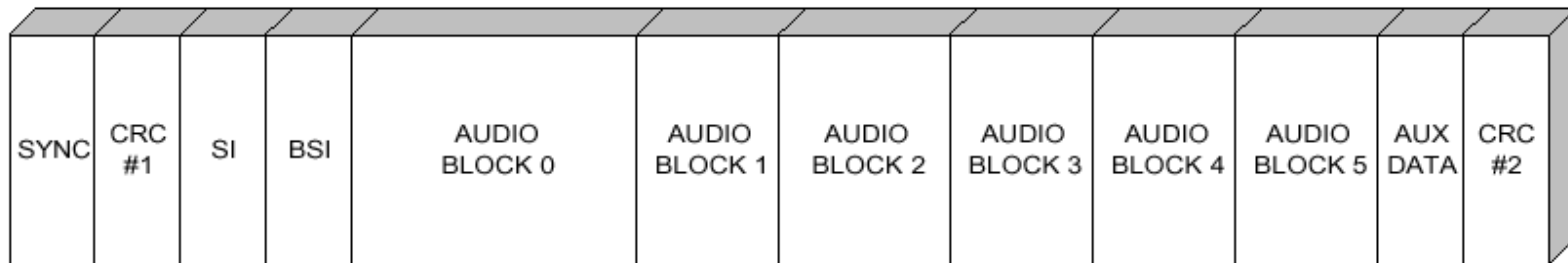
- ✓ **Codificação por sub-bandas**
- ✓ **Codificação perceptual**
- ✓ **MPEG áudio X Dolby AC-3:**
 - ***MPEG***
 - decide a alocação de bits por sub-bandas de acordo com o modelo psicoacústico utilizado (*forward adaptive bit allocation mode*)
 - *Vantagem: decodificador não precisa conhecer o modelo psicoacústico*
 - Informação de alocação de bits deve ser inserida no quadro MPEG-áudio
 - *Desvantagem: desperdiça banda passante com informação de alocação de bits*
 - ***Dolby AC-3***
 - Não passa informação de alocação de bits no quadro AC-3, mas sim alguns parâmetros que permitem ao decodificador inferir o modelo de mascaramento utilizado e a alocação de bits no quadro
 - utiliza notação de ponto flutuante (expoente e mantissa) para representar os coeficientes de frequência

Codificador Dolby AC-3



Dolby AC-3

✓ Quadro AC-3



- ✓ Cada bloco de áudio contém dados comprimidos relativos a 256 novas amostras PCM
 - *Taxa típica - áudio multicanal a 384 Kbps*

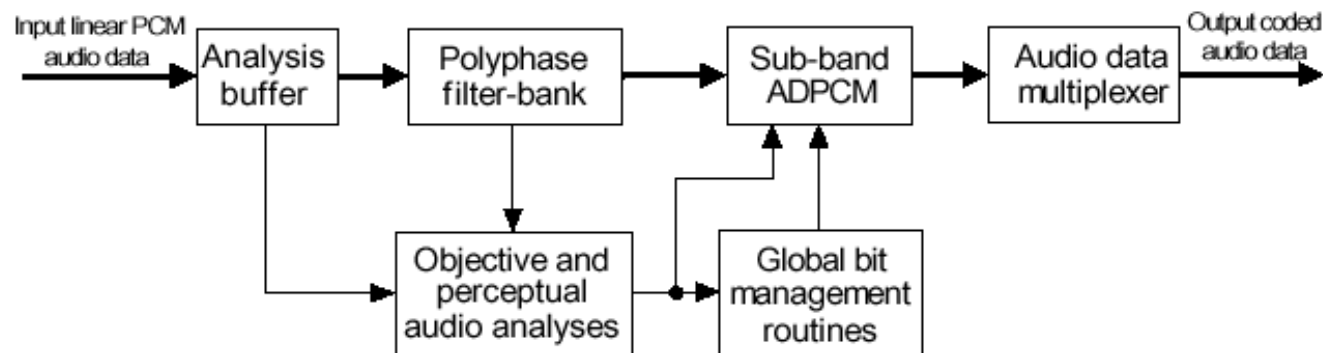
Áudio DTS

- ✓ **Digital Theater Systems (DTS) – www.dtstech.com**
- ✓ **Concorrente do Dolby**
- ✓ **Usado em filmes de cinema**
 - *Jurrassic Park foi o primeiro*
- ✓ **Também permite 6 canais de áudio**
- ✓ **amostragem PCM de até 192KHz com 24 bits por amostra**
- ✓ **256, 512, 1024, 2048 ou 4096 amostras PCM por quadro**
- ✓ **Taxa de 754 Kbps ou 1.5 Mbps**
- ✓ **Outra opção para DVD áudio**

Áudio DTS

✓ Codificador DTS

- *Codificação por Transformadas e Sub-bandas*
- *Codificação Perceptual*
- *Codificação ADPCM*
- *Codificação por Entropia (VBR)*



Codificação de Áudio AAC

✓ MPEG-2 AAC

✓ MPEG-4 HE-AAC

✓ Referências:

- *MPEG-4 High-Efficiency AAC Coding, Jürgen Herre e Martin Dietz, IEEE Signal Processing Magazine (137), Maio 2008.*
- *The MPEG-4 Book, Fernando Pereira e Touradj Ebrahimi, IMSC Press, Prentice-Hall, 2003.*
- *Audio Signal Processing and Coding, Andreas Spanias, Ted Painter e Venkatraman Atti, John Wiley & Sons, 2007.*

MPEG-2 AAC

- ✓ **MPEG-2 NBC (non-backward compatible)**
- ✓ **AAC - Advanced Audio Coding**
- ✓ **Publicado em 1997**
- ✓ **MPEG-2 BC codificava áudio multicanal a taxas maiores que 640 Kbps**
- ✓ **Surgiu da necessidade de codificar audio multicanal a taxas menores que 384 Kbps**
 - *Multicanal – 320 Kbps*
 - *Estéreo – 128 Kbps*
- ✓ **Taxas de amostragem variam de 8 a 96 KHz**
 - *8, 11, 12, 16, 22, 24, 32, 44, 48, 64, 88, 96 KHz*
- ✓ **Permite até 48 canais de áudio**
- ✓ **Adotado como núcleo do codificador de áudio MPEG-4**

MPEG-2 AAC

✓ Perfis:

- *LC – Low complexity*
 - Não utiliza predição inter-quadros
- *Main*
- *SSR – Scalable Sampling Rate*
 - Codificação escalável
 - Adiciona ferramenta *gain control*

MPEG-2 AAC

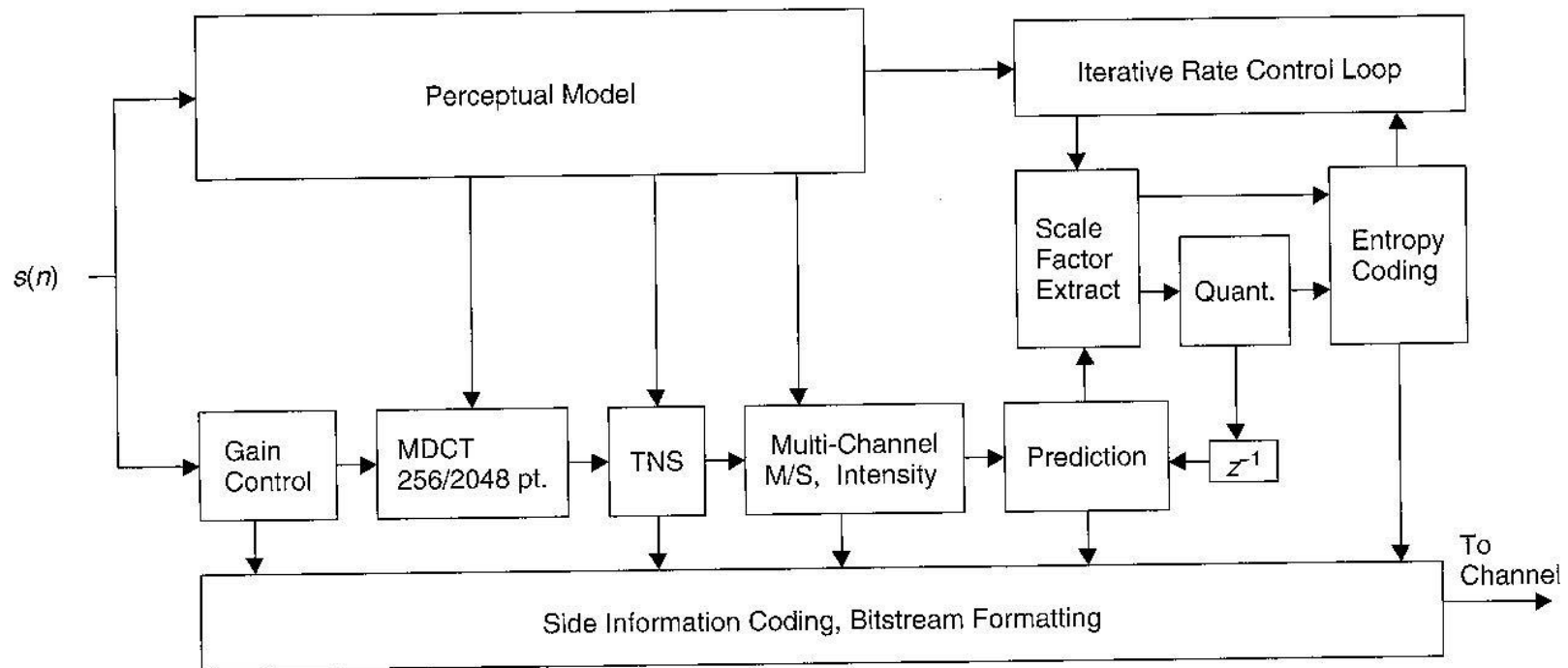
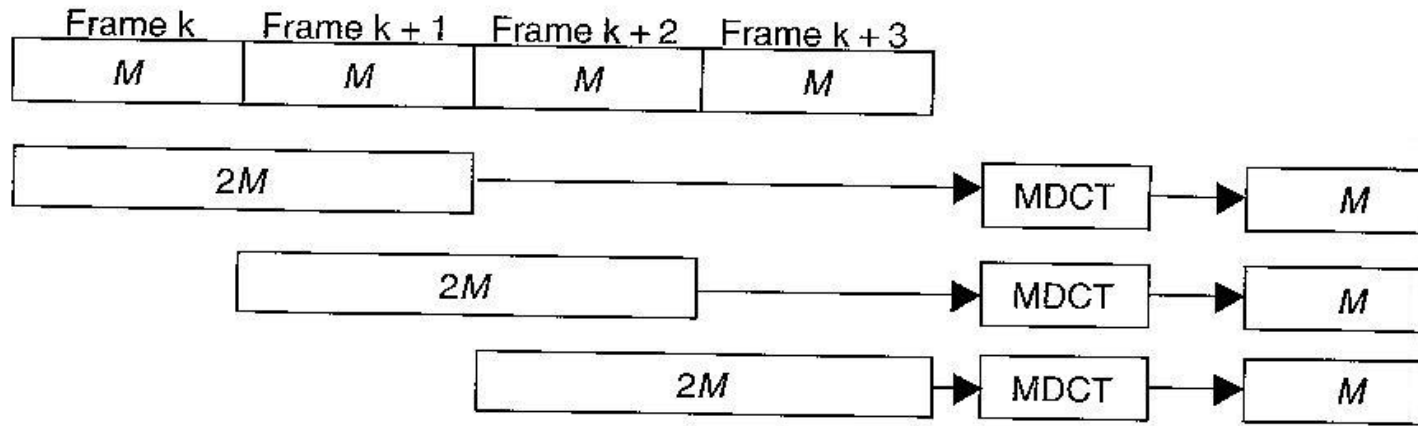


Figure 10.10. ISO/IEC IS13818-7 (MPEG-2 NBC/AAC) encoder ([Bosi96]).

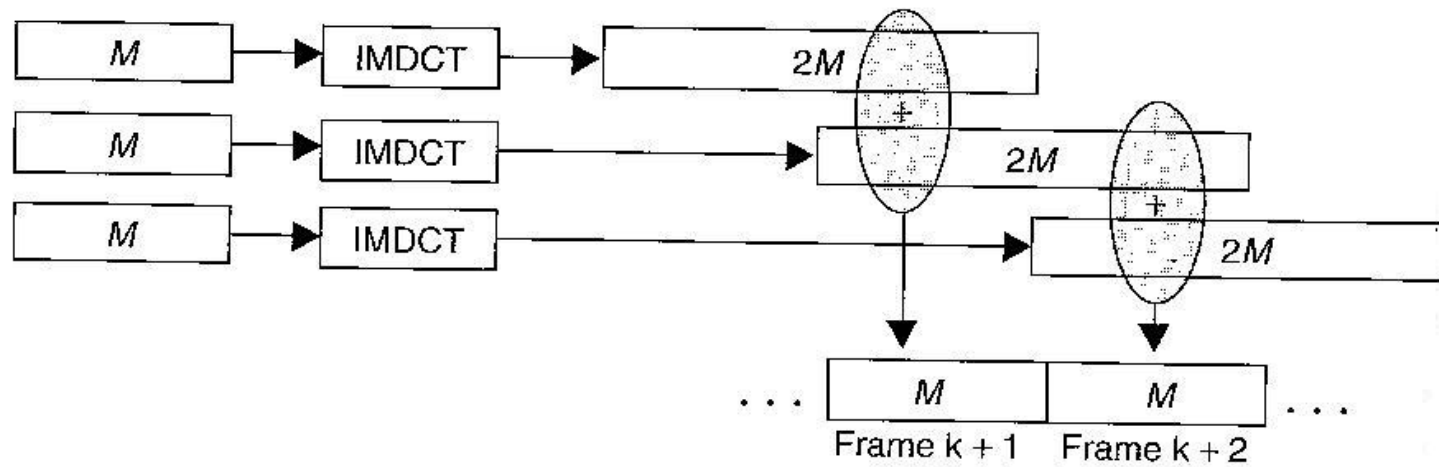
Transformada

- ✓ **MDCT – Modified Discrete Cosine Transform**
- ✓ **Aplicada a janelas de 2048 amostras com 1024 novas amostras (ou 8 conjuntos de 256)**
 - *Produzem 1024 coeficientes na frequência*
 - *1024 sub-bandas com resolução de 23,4Hz (para taxa de amostragem de 48KHz)*

Transformada



(a)



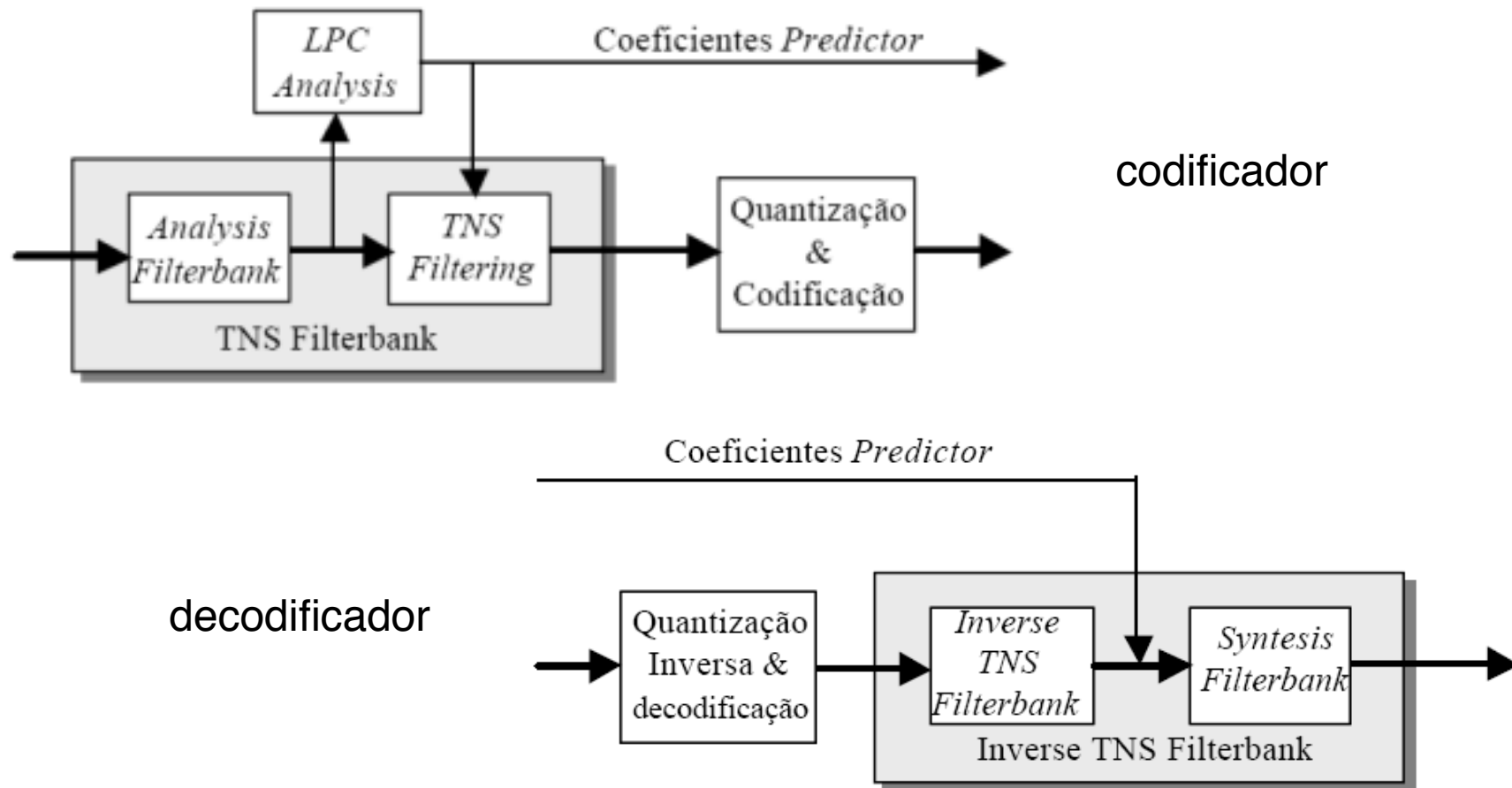
(h)

TNS

- ✓ **TNS – Temporal Noise Shaping**
- ✓ **Novo conceito em codificação perceptual de áudio**
- ✓ **Técnicas de LPC (Linear predictive coding) são aplicadas no domínio da frequência**
- ✓ **Coeficientes espectrais são analisados e o ruído é modelado e representado por parâmetros LPC**
- ✓ **Técnica usada em todos os perfis**

TNS

✓ TNS – Temporal Noise Shaping



Prediction

- ✓ **Predição dos coeficientes espectrais do quadro atual é feita com base nos coeficientes do quadro anterior**
 - *Backward prediction*
 - *Somente para coeficientes de frequência menores que 16KHz*
- ✓ **Economia de bits é conseguida codificando apenas o resíduo (sinal diferença ou sinal de erro)**
- ✓ **Só o perfil Main utiliza esta técnica**

Quantização e Alocação de bits

- ✓ **Processo iterativo até atingir taxa de bits desejada**
- ✓ **Modifica o quantum do fator de escala de cada sub-banda em incrementos de 1.5 dB e obtém códigos de Huffman para fator de escala e coeficientes quantizados até a taxa de bits ser atingida**

Padrão MPEG-4

- ✓ **Objetos de mídia podem ser de origem natural ou sintética**
 - *Áudio (voz ou música), vídeo, objetos 2D e 3D*
- ✓ **Composições de objetos formam cenas audiovisuais**
- ✓ **Interatividade com o usuário**
- ✓ **Transmissão sobre qualquer tipo de rede**

Codificação de áudio MPEG-4

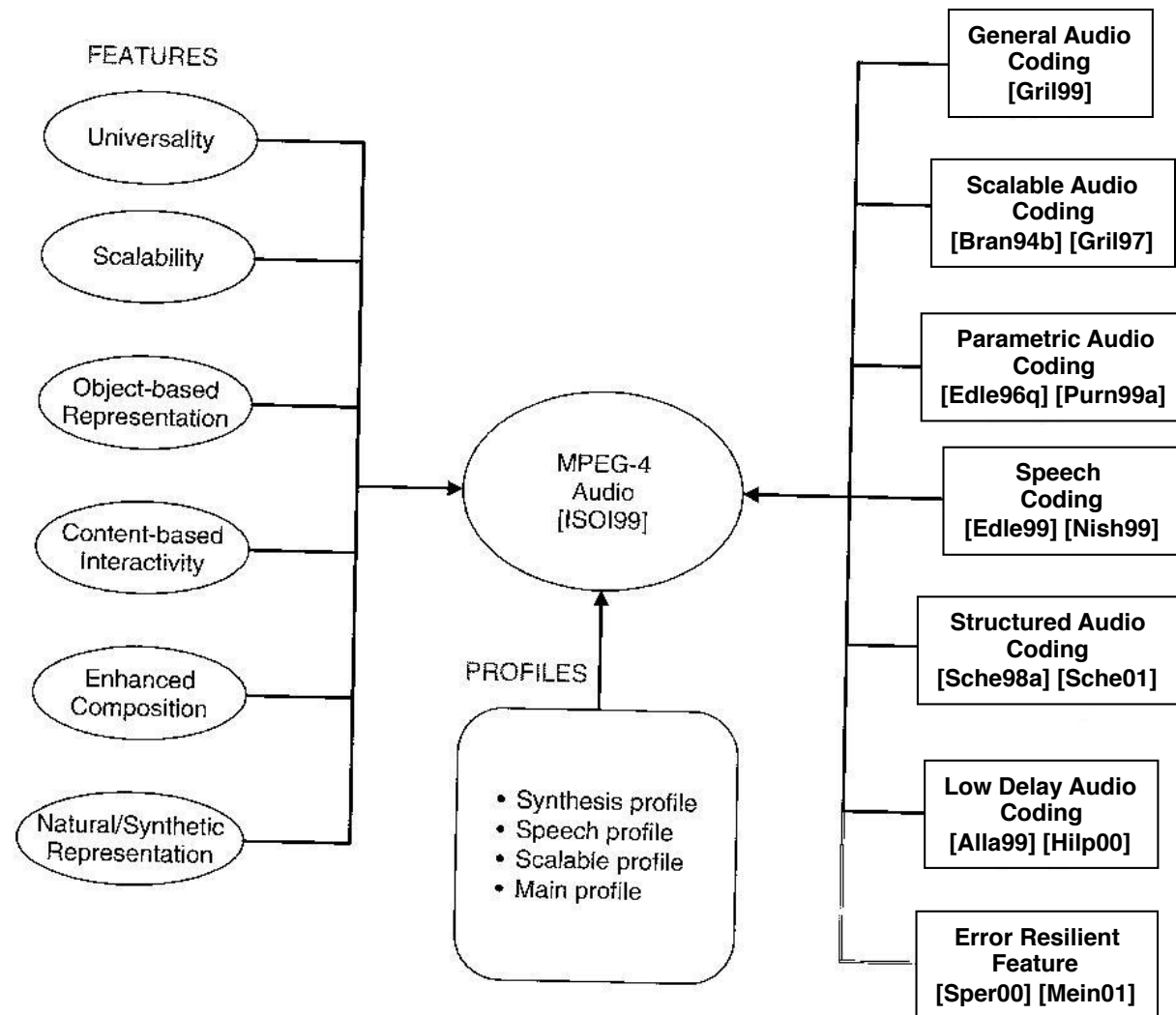


Figure 10.12. An overview of the MPEG-4 audio coder.

MPEG-4 GAC

✓ MPEG-4 GAC – generic audio coder

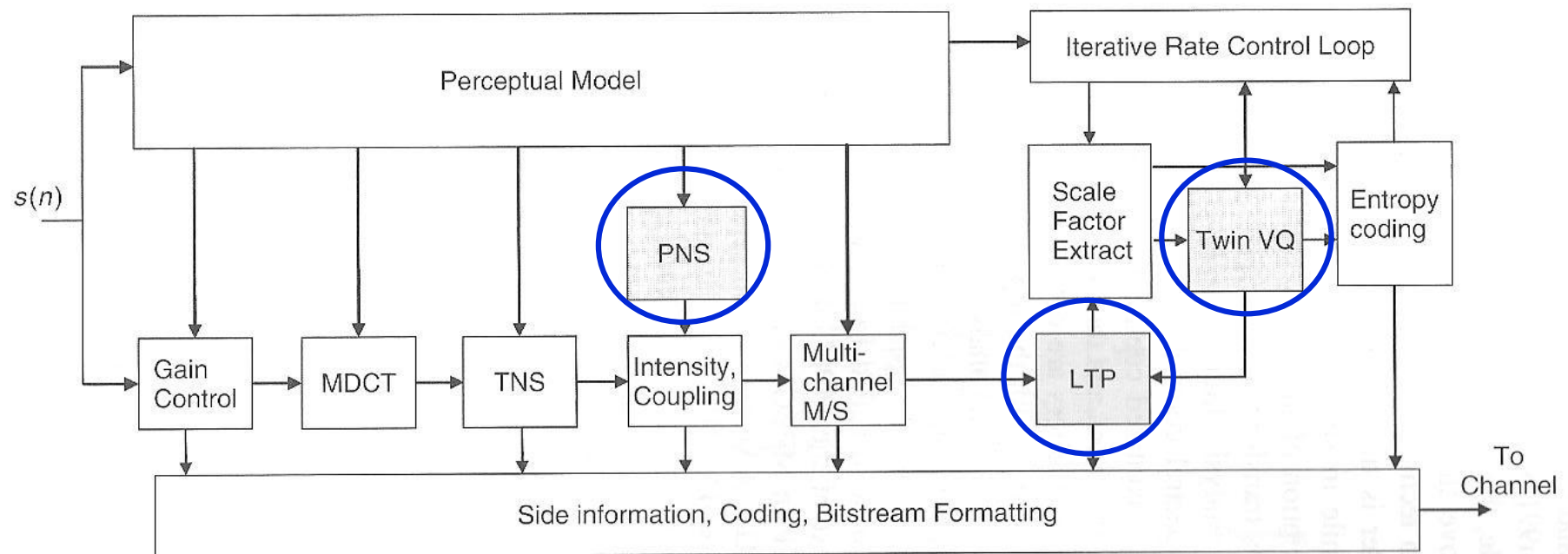


Figure 10.14. MPEG-4 GA coder.

Novas ferramentas

✓ PNS (Perceptual Noise Substitution)

- *Perceptual coding + substitute parametric form of noise-like signals*
- *Componentes similares a ruído são identificados e excluídos do sinal*
- *Um flag de substituição de ruído é adicionado juntamente com a potência total desses coeficientes para cada sub-banda*
- *Decodificador regenera o ruído com base na potência informada*

PNS

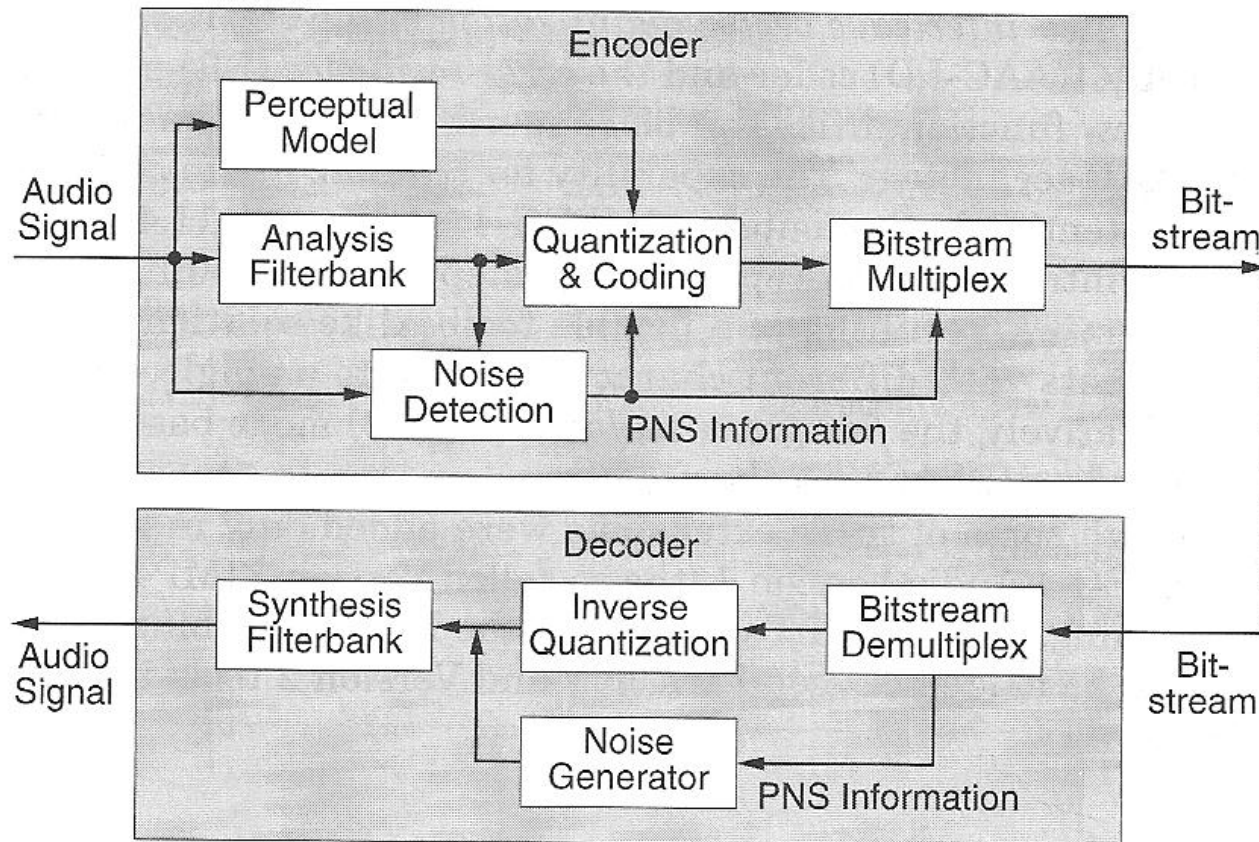


Fig. 11.9 PNS coding principle.

Novas ferramentas

✓ LTP (Long-Term Prediction)

- *Predição no domínio da frequência com base nos quadros anteriores (backward prediction)*
 - Os valores espectrais quantizados são mapeados para o domínio do tempo e reconstruídos
- *Comparando o sinal reconstruído com o original, os parâmetros ótimos para atraso (pitch lag) e amplitude (gain) são determinados para formar o sinal predito*
- *O sinal original e o predito são subtraídos formando o sinal residual (diferença)*
- *Um dos dois (original ou residual) é codificado de forma a minimizar a taxa de bits (frequency selective switch)*
 - Esses parâmetros LTP são transmitidos ao decodificador (forward adaptation)

LTP

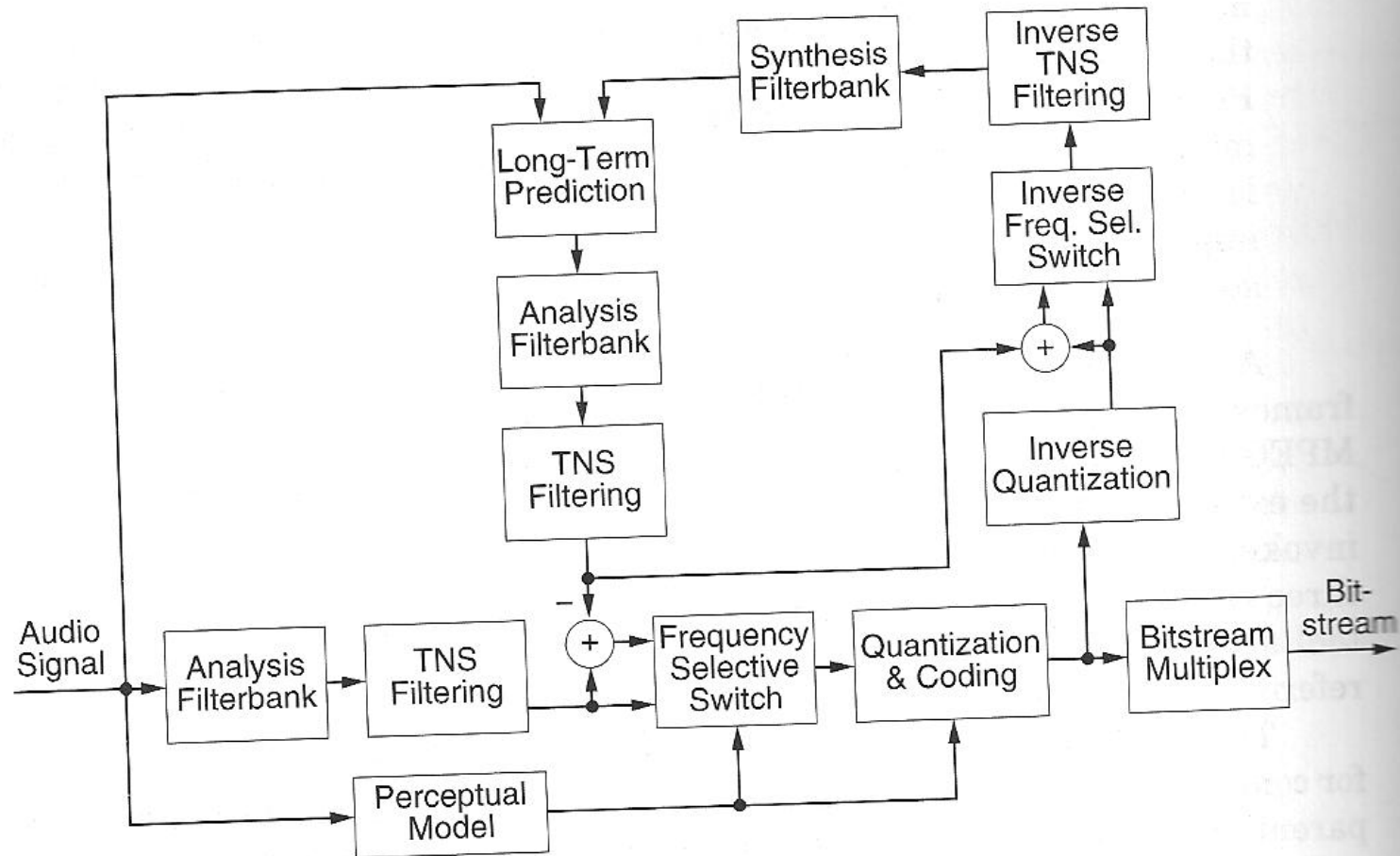


Fig. 11.10 LTP Integration into a T/F coder.

Novas ferramentas

✓ TwinVQ

- *Transformed Weighted Interleave Vector Quantization*
- *Controla a distorção da quantização através de um modelo perceptual*
- *Intercalamento permite alocação de bits constante para todos os subvetores*
- *Utiliza quantização vetorial para representar os coeficientes*

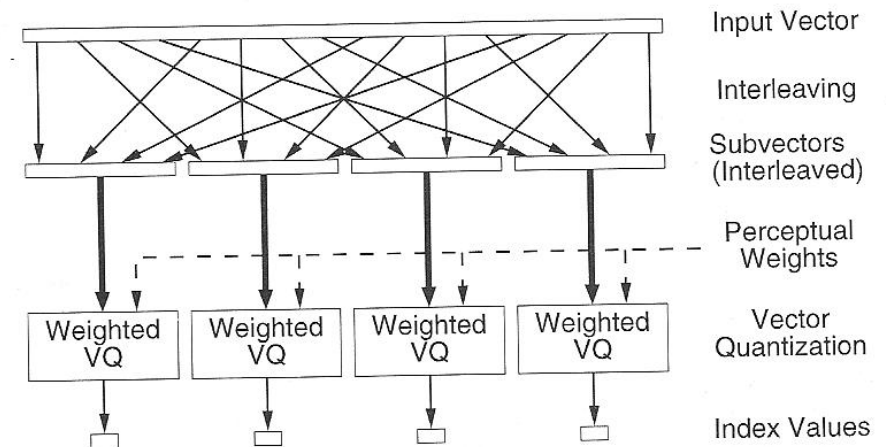


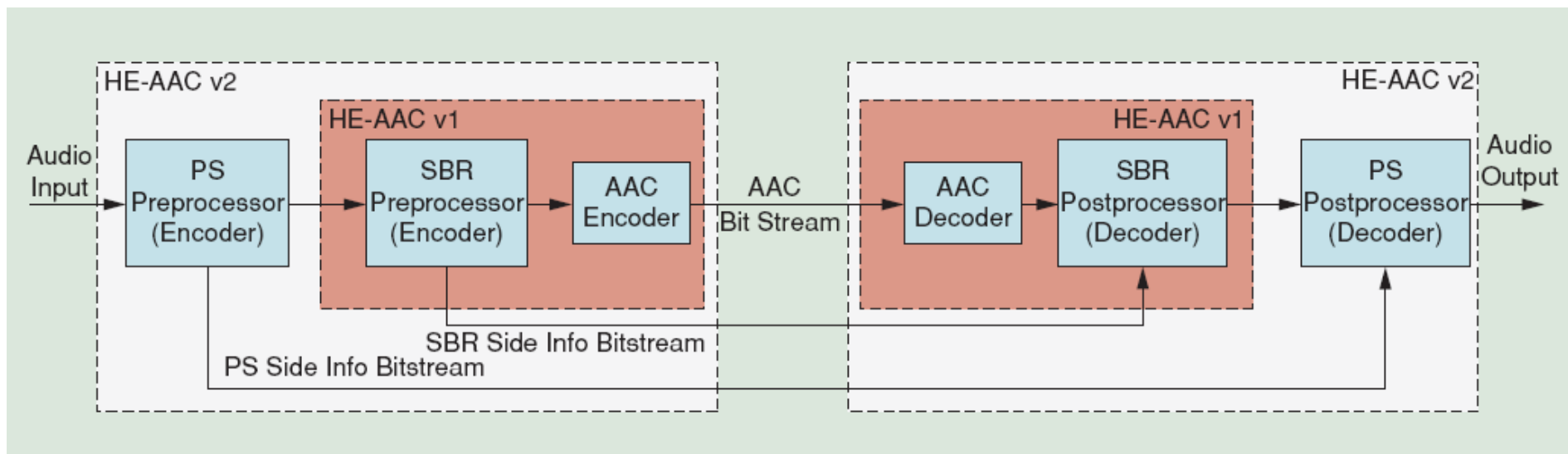
Fig. 11.11 TwinVQ interleaving and VQ encoding.

MPEG-4 AAC

- ✓ **MPEG-4 AAC publicado em 1999**
- ✓ **HE-AAC v.1 publicada em 2003 (AAC + SBR)**
- ✓ **HE-AAC v.2 publicada em 2004 (v.1 + PS)**
- ✓ **Em 2005, é publicada a 3a. Versão do MPEG-4 audio com todas as ferramentas incluídas**
 - *ISO/IEC 14496-3:2005 - Part 3: Audio*
- ✓ **Decodificador v.2 também decodifica v.1 e AAC**

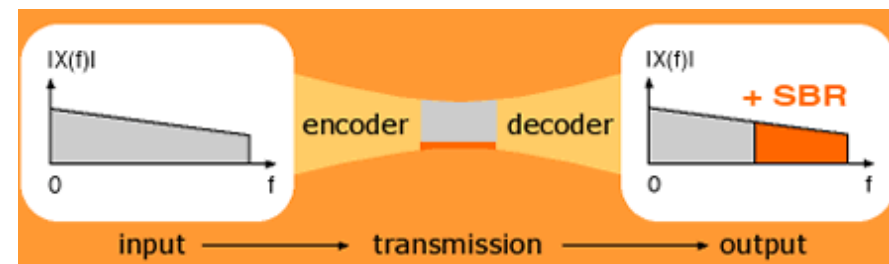
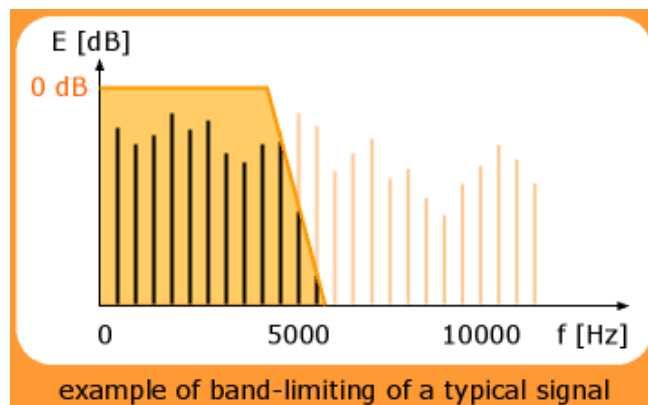
MPEG-4 HE-AAC

- ✓ **SBR (spectral band replication) e PS (parametric stereo) são ferramentas de pré-processamento no codificador e pós-processamento no decodificador**

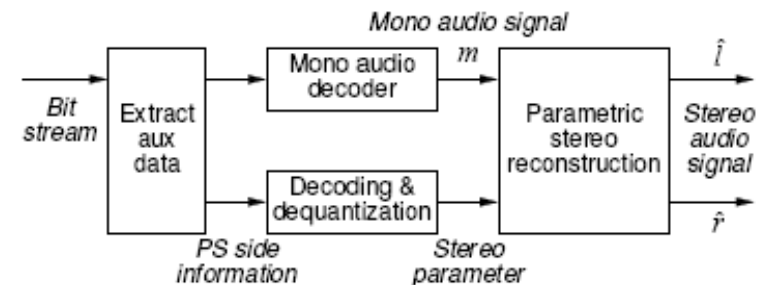
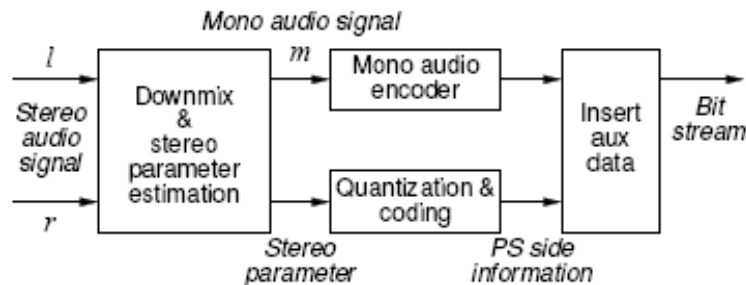


SBR

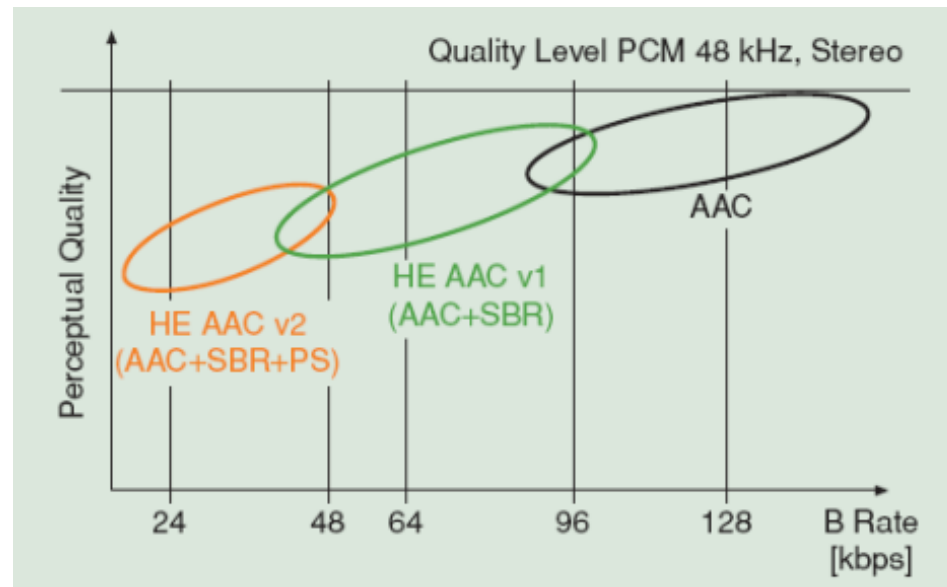
- ✓ **Spectral Band Replication**
- ✓ **Percepção auditiva humana é mais sensível a baixas frequências**
- ✓ **SBR não transmite a parte alta do espectro de frequências, ela é regenerada pelo decodificador com base na transposição das frequências baixas e um conjunto de parâmetros que estimam o envelope espectral**



- ✓ **Parametric Stereo**
- ✓ **Usada somente em áudio estéreo**
- ✓ **Extensão de joint-stereo coding (MPEG-1)**
- ✓ **converte o sinal estéreo em:**
 - *sinal mono e*
 - *um conjunto de parâmetros que definem a composição espacial do sinal*
 - *(inter-channel intensity difference, inter-channel phase difference, inter-channel coherence, overall phase difference)*



MPEG-4 AAC



MPEG-4

✓ Níveis e Perfis

LEVEL	AAC PROFILE			HE-AAC V1 PROFILE			HE-AAC V2 PROFILE			
	MAX. CHANNELS	MAX. SAMPLING RATE [kHz]	MAX. CHANNELS	MAX. AAC SAMPLING RATE, SBR NOT PRESENT [kHz]	MAX. AAC SAMPLING RATE, SBR PRESENT [kHz]	MAX. SBR SAMPLING RATE [kHz]	MAX. CHANNELS	MAX. AAC SAMPLING RATE, SBR NOT PRESENT [kHz]	MAX. AAC SAMPLING RATE, SBR PRESENT [kHz]	MAX. SBR SAMPLING RATE [kHz]
1	2	24	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	2	48	2	48	24	48	2	48	24	48
3	N/A	N/A	2	48	48	48	2	48	48	48
4	5	48	5	48	48	48	5	48	48	48
5	5	96	5	96	48	96	5	96	48	96

Principais níveis:
Nível 2 (dispositivos estéreo) e
Nível 4 (sistemas multicanal – TV Digital)

SBTVD

- ✓ **Taxas de amostragem do sinal de entrada**
 - *32KHz, 44.1KHz, 48KHz*
- ✓ **Quantização**
 - *16 ou 20 bits por amostra*
- ✓ **Mínimo de 1 canal e máximo de 5.1 canais**
 - *Pode ser transmitido simultaneamente em mais de um modo*
 - *Receptores full-seg devem ser capazes de converter áudio multicanal em estéreo*

✓ Perfis e níveis:

- *LC (Low Complexity)*
 - Perfil básico do AAC
 - Níveis L2 e L4
- *HE (high efficiency)*
 - Perfil avançado combinando perfil LC com ferramenta SBR (spectral band replication)
 - Níveis L2 e L4
- *HE + ferramenta PS (parametric stereo)*
 - Nível L2

SBTVD

✓ Serviço full-seg

Parâmetro	Restrição
Mecanismos de transporte permitidos	LATM/LOAS (conforme ISO/IEC 14496-3)
Números de canais recomendados	Mono (1.0), 2 canais (estéreo ou 2.0), ou multicanal (5.1)
Perfis e níveis permitidos	<i>Low complexity</i> AAC: nível 2 (LC-AAC@L2) para dois canais <i>Low complexity</i> AAC: nível 4 (LC-AAC@L4) para multicanal <i>High Efficiency</i> (HE): nível 2 (HE-AAC v1@L2) para dois canais <i>High Efficiency</i> (HE): nível 4 (HE-AAC v1@L4) para multicanal
Taxa máxima de bits permitida	Conforme ISO/IEC 14496-3
Amostras por quadro	frameLengthFlag em GASpecificConfig() deve ter valor 0, indicando que a extensão do quadro deve ser de 1024 amostras para AAC e 2048 quando usando SBR. 960 amostras para AAC (ou 1 920 quando usando SBR) não devem ser utilizadas

SBTVD

✓ Serviço one-seg

Parâmetro	Restrição
Mecanismos de transporte permitidos	LATM/LOAS, conforme ISO/IEC 14496-3
Perfis e níveis permitidos	<i>High efficiency</i> (HE): nível 2 (HE-AAC v2@L2)
Número máximo de canais codificados	2 canais por fluxo de bits (estéreo ou 2 canais monaurais)
Taxa máxima de bits	Conforme ISO/IEC 14496-3