Digital Speech and Audio Coding

Mathew Magimai Doss and Petr Motlicek

Idiap Research Institute, Martigny

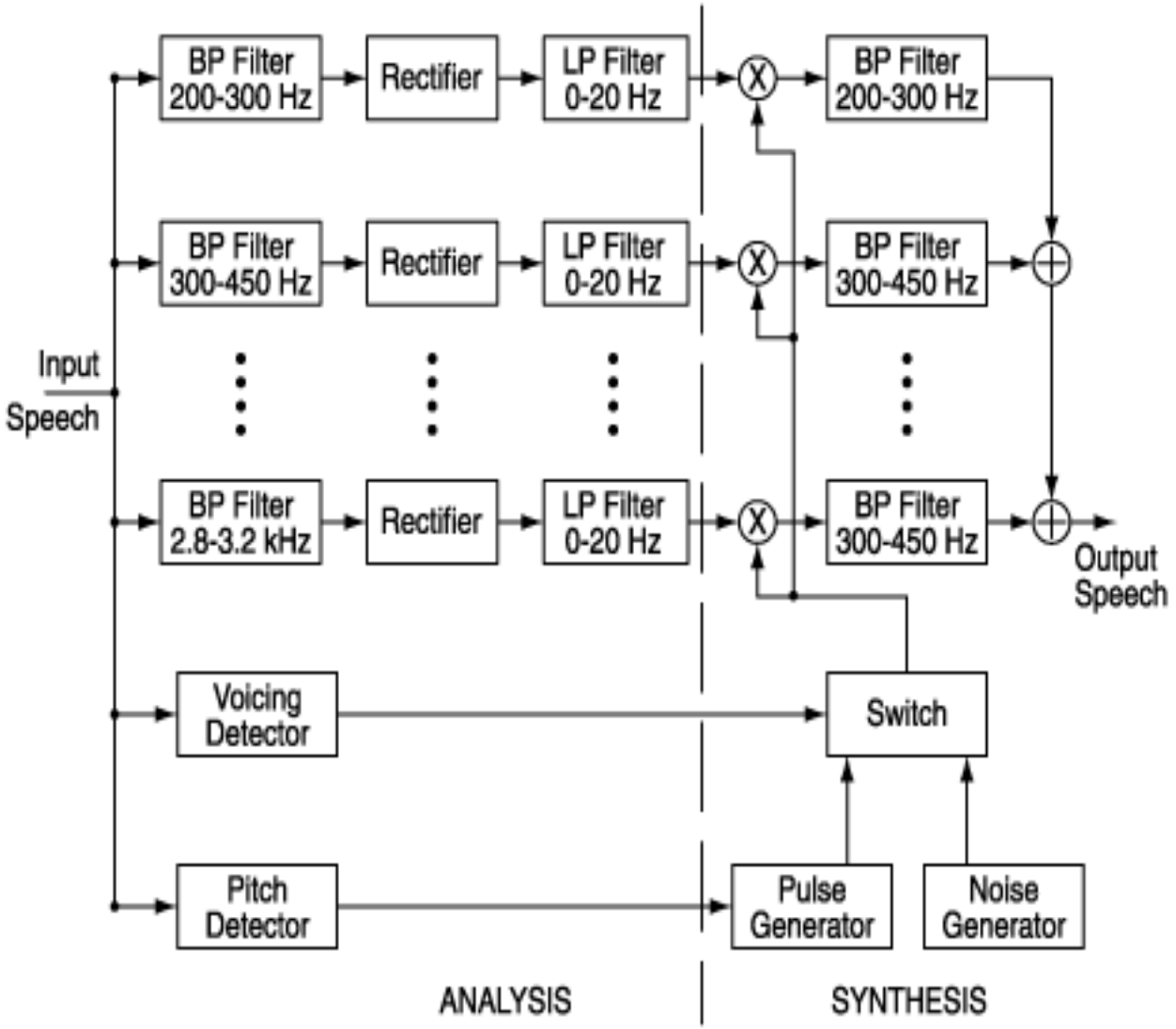
http://www.idiap.ch/

Ecole Polytechnique Fédérale de Lausanne, Switzerland

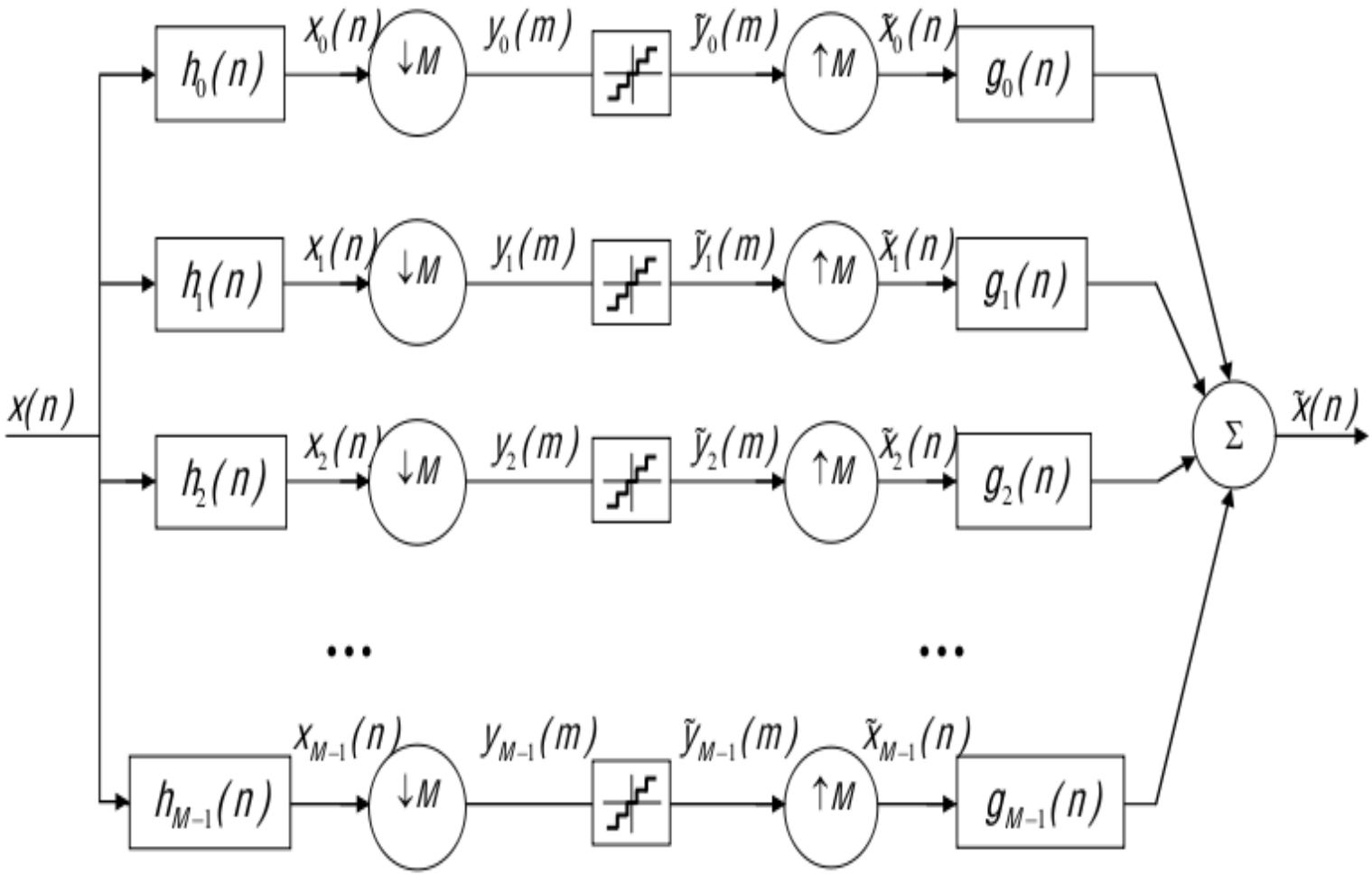
Today's Course

- Subband coding (recap)
- MPEG coding

Channel Vocoder



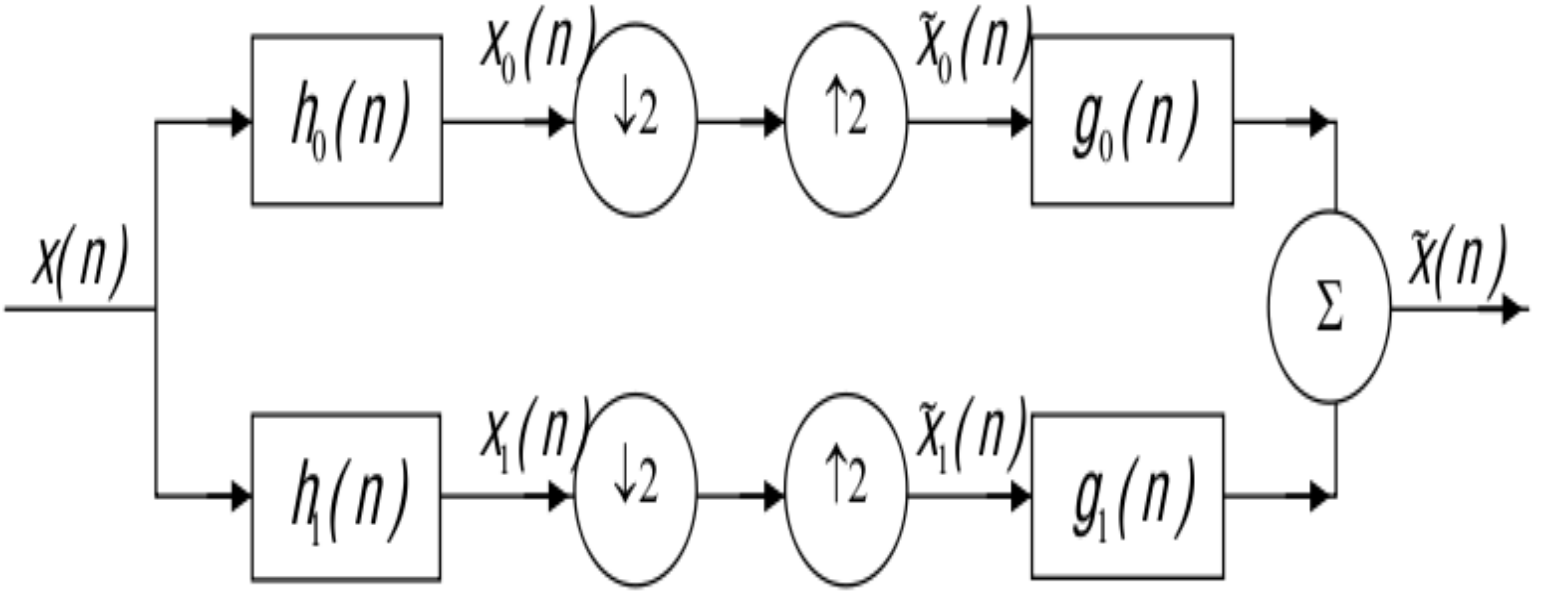
Subband Coding



Subband Coding

Perfect reconstruction

$$\tilde{x}(n) = x(n)$$

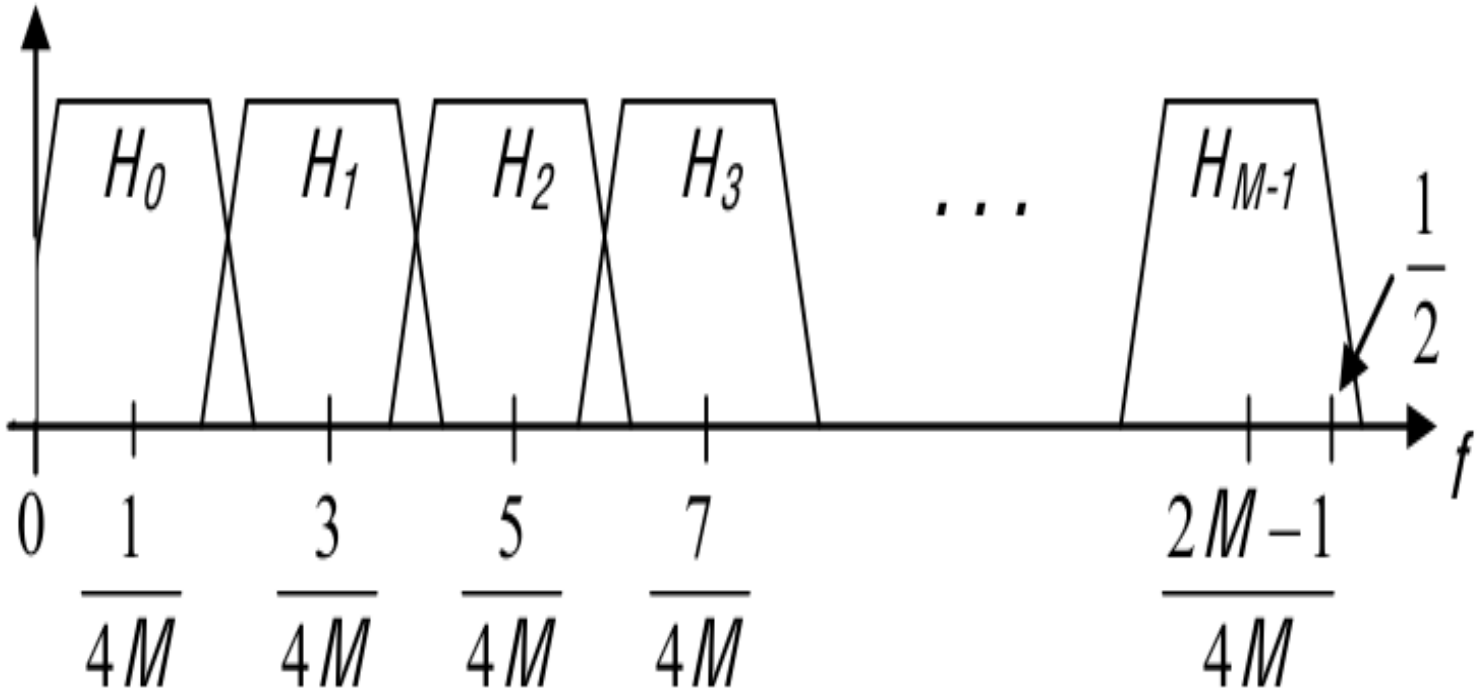


Subband Coding

- Remove distortion due to phase and amplitude responses of the filters
- Remove aliasing distortion

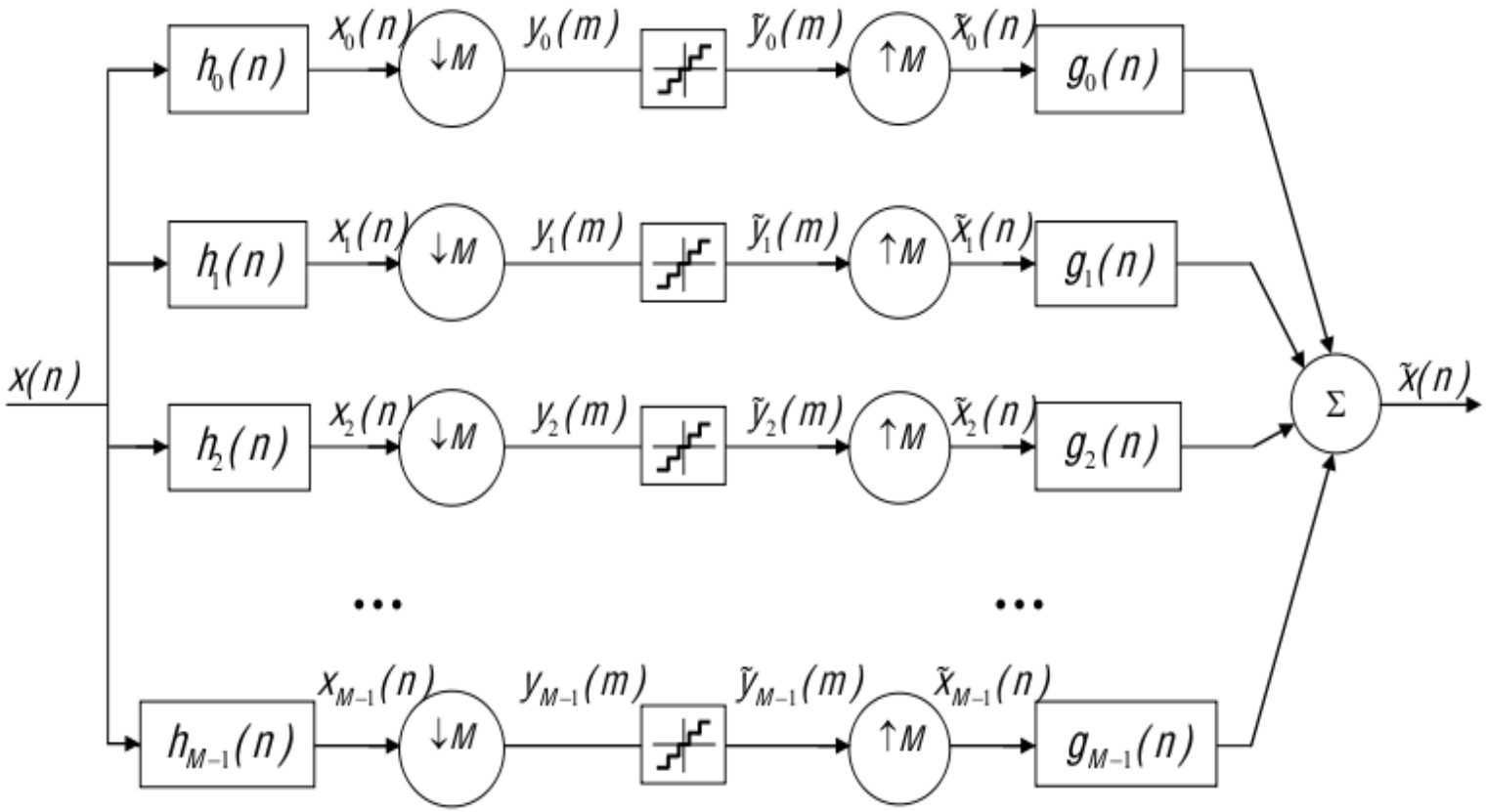
Subband Coding

Pseudo QMF (Polyphase) filters: going from 2 to M bands



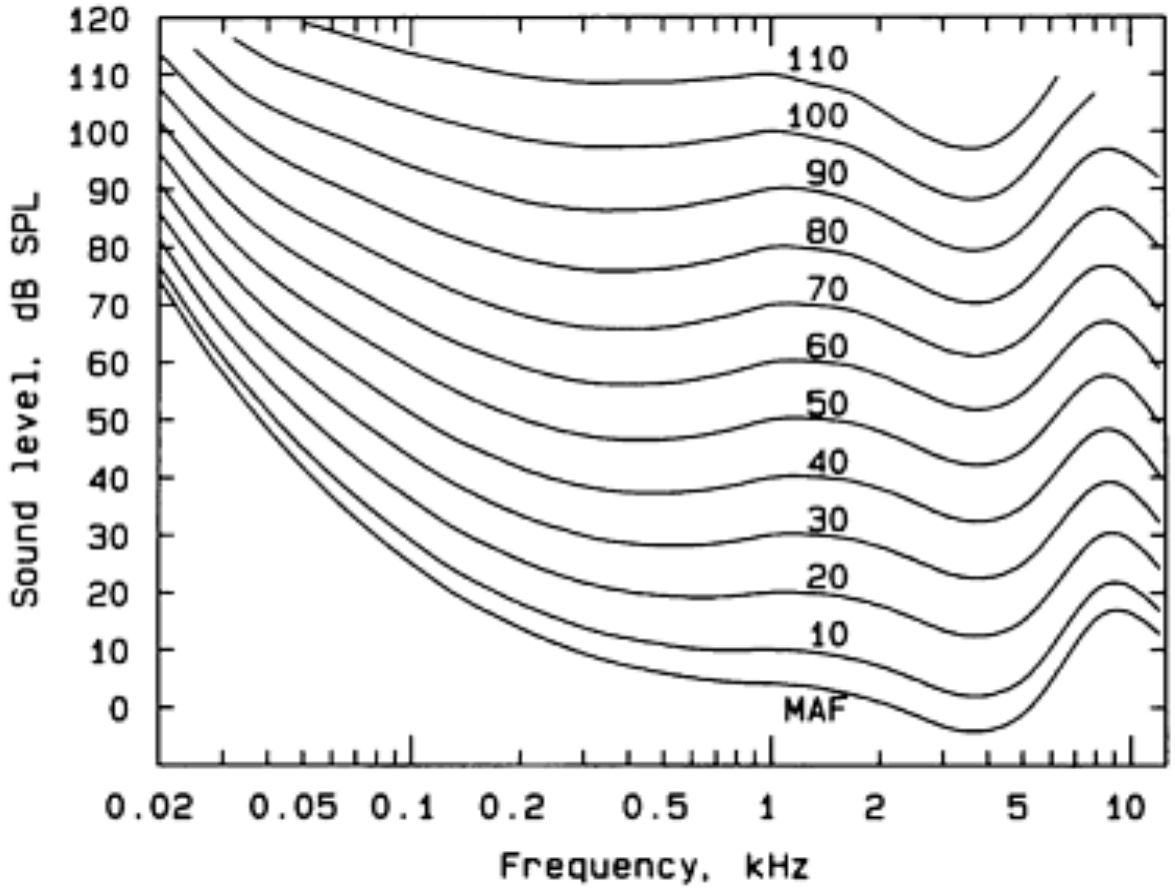
MPEG Audio Coding

Reducing bit rate using psychoacoustic knowledge



Psychoacoustic

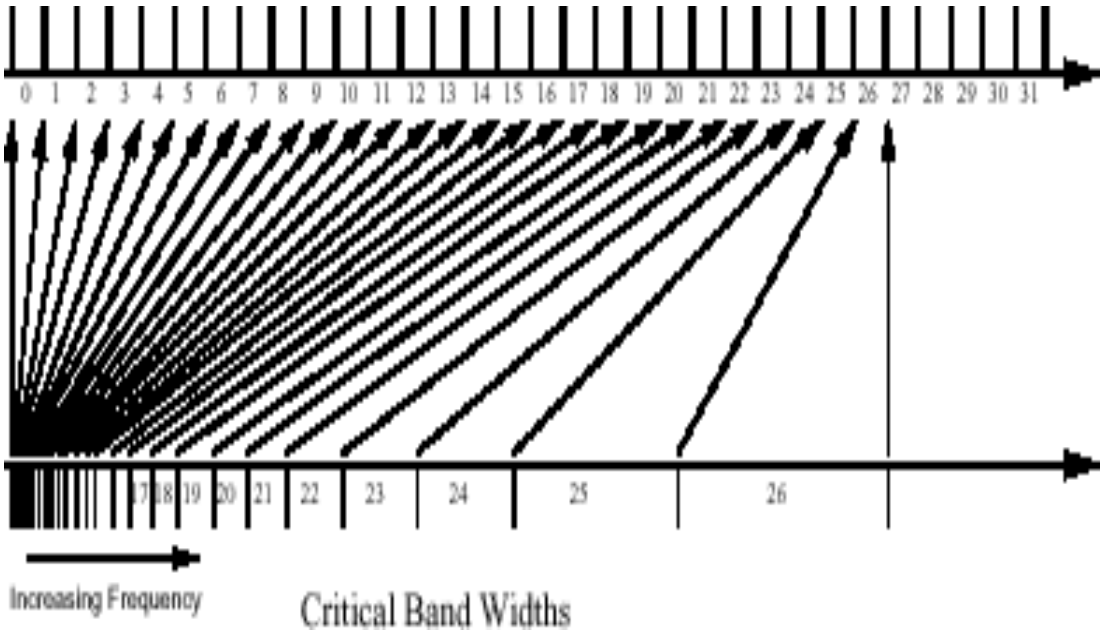
- Equal loudness curve



Psychoacoustic

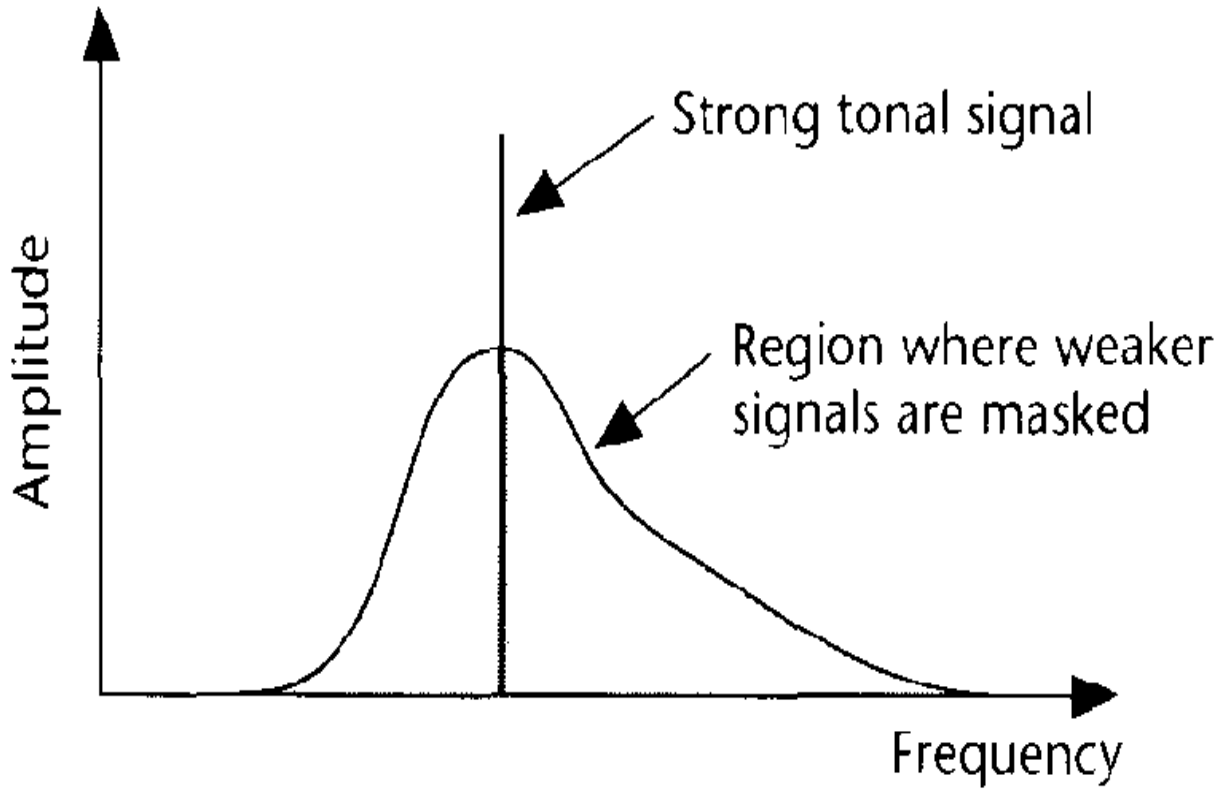
- Critical band

$$\text{Bark} = 13 \cdot \arctan(\frac{0.0076f}{1000}) + 3.5 \cdot \arctan(\frac{f}{7500})^2$$

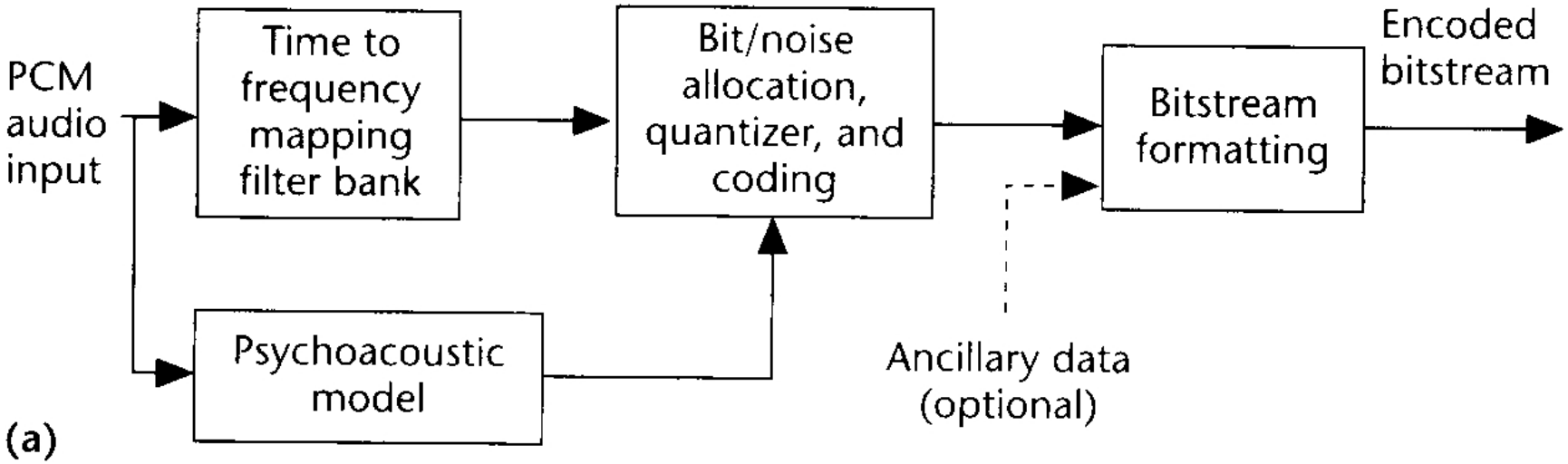


Psychoacoustic

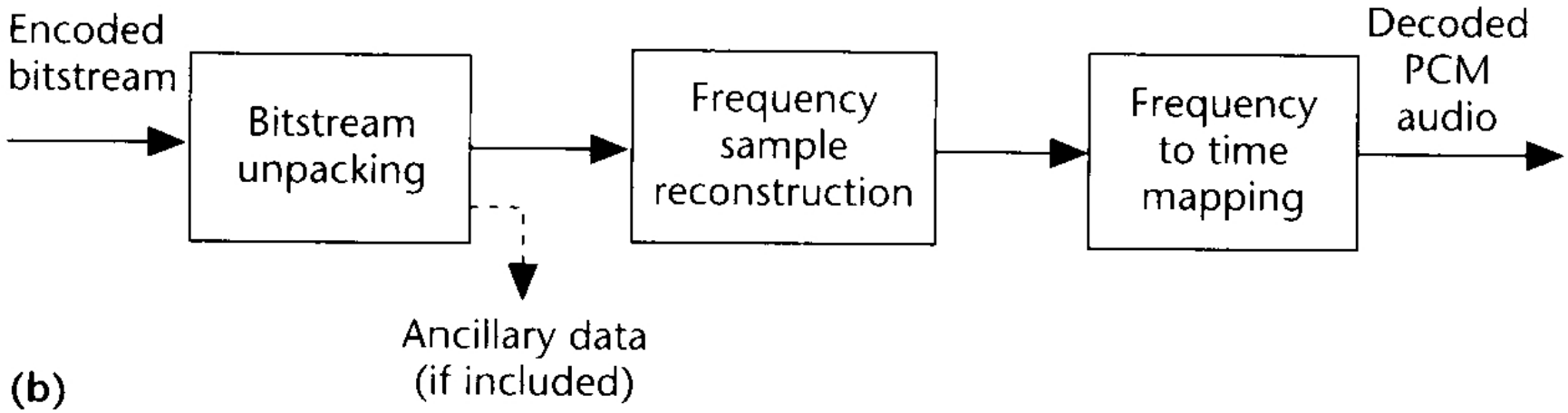
- Masking



MPEG Audio Coding



MPEG Audio Coding



(b)

MPEG Audio Coding

Two key steps

- Estimation of local masking threshold for each subband using the psychoacoustic model
- Bit allocation i.e., number of bits used to quantize each subband based on the local masking threshold

Masking threshold estimation

- Transform short-term audio signal to frequency domain (Fourier transform)
- Critical band integration
- Separation of spectral values into tonal and non-tonal components

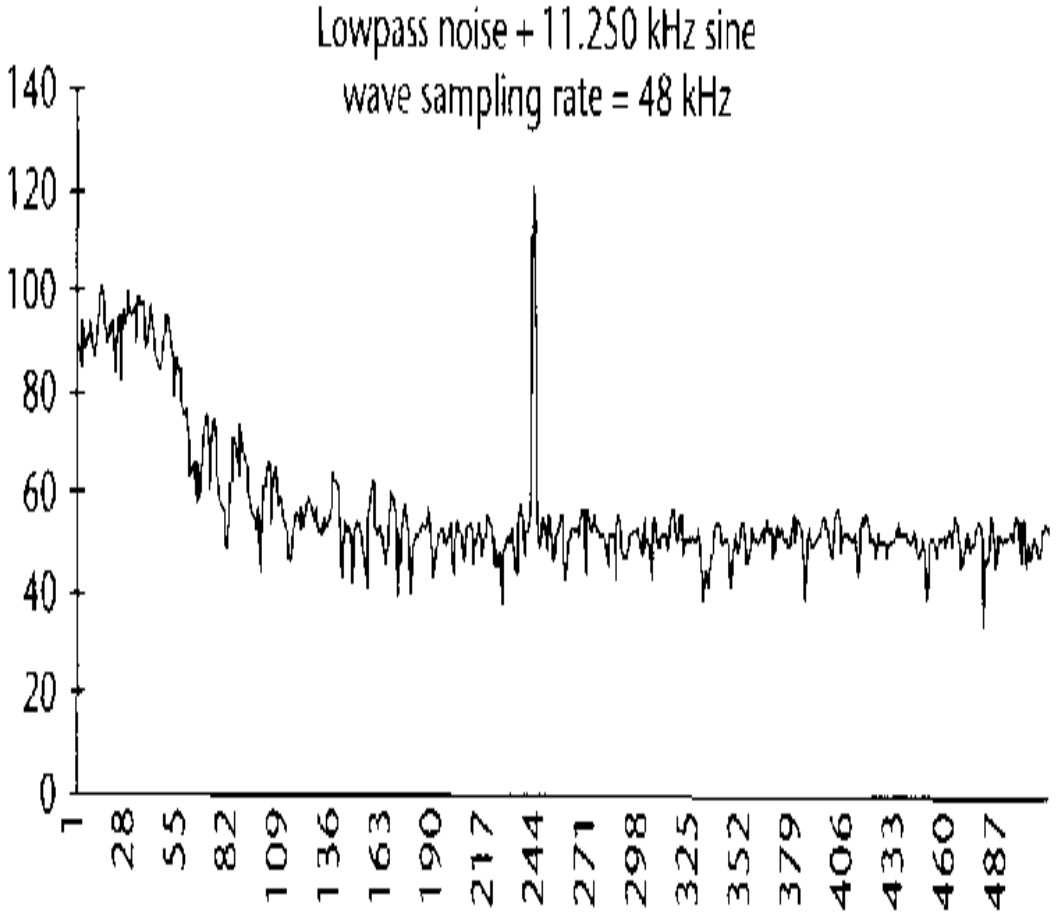
Masking threshold estimation

- Determine noise masking threshold (masking ability of a tone/signal across its neighboring critical bands)
- Find the masking threshold ϕ_k for each subband k,
- Calculating signal-to-mask ratio (SMR) for each subband k

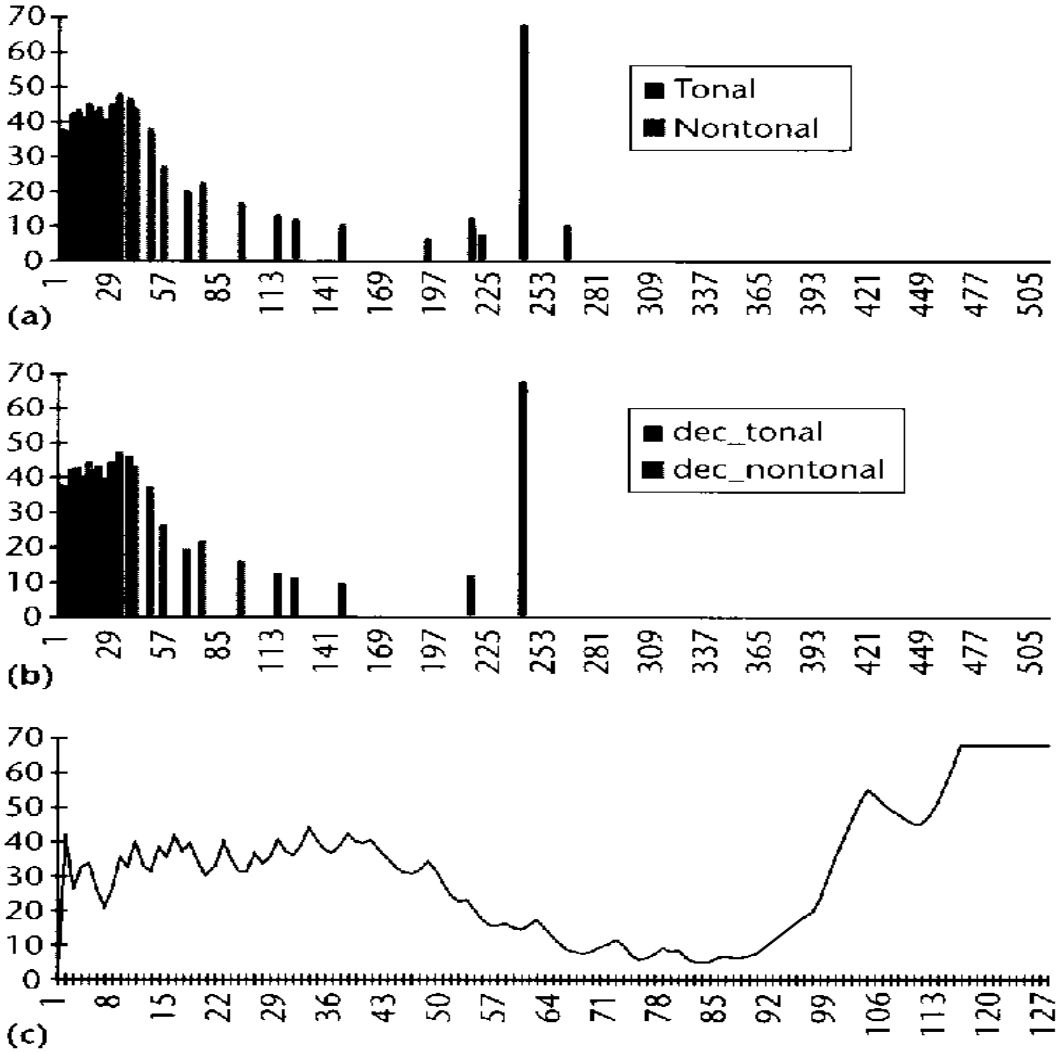
$$\text{SMR}_k = 10 \cdot \log_{10}(\frac{E_k}{\phi_k})$$

E_k, signal energy in the subband.

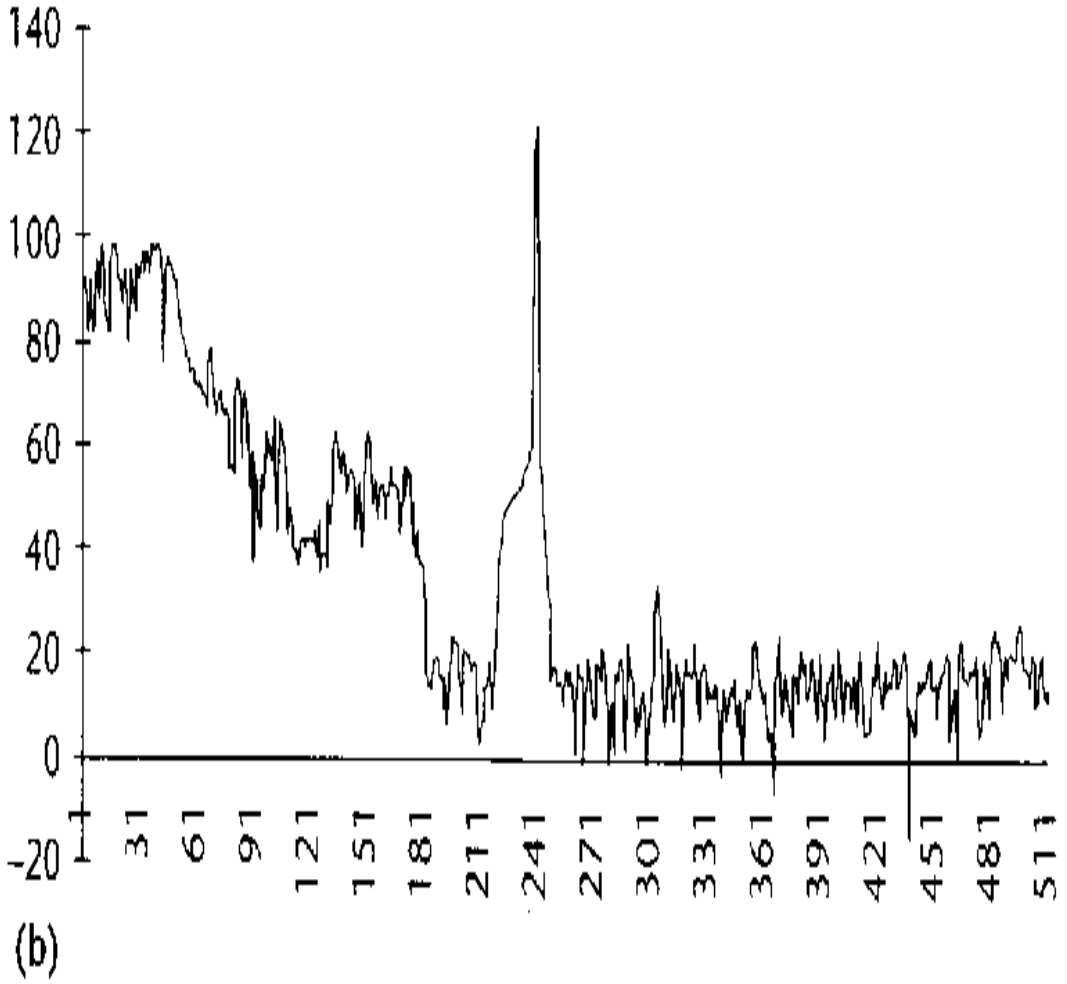
Masking threshold estimation



Masking threshold estimation



Masking threshold estimation



Bit Allocation

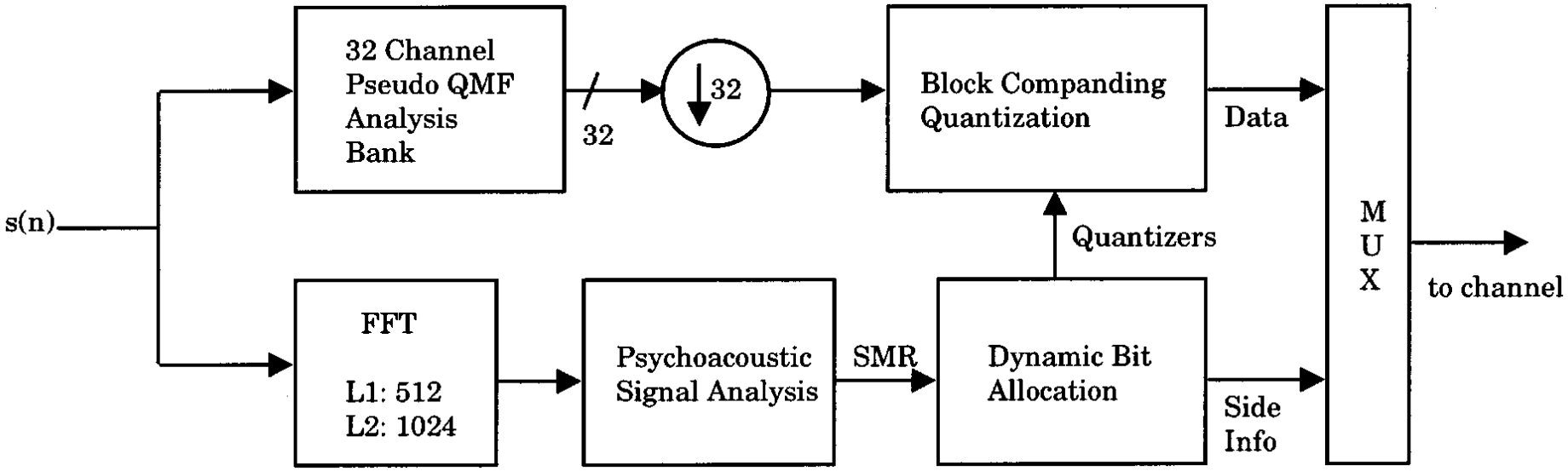
Given the total number of bits and the SNR for each band (supplied by MPEG/audio standards)

- Compute mask-to-noise ratio (MNR) for each subband k

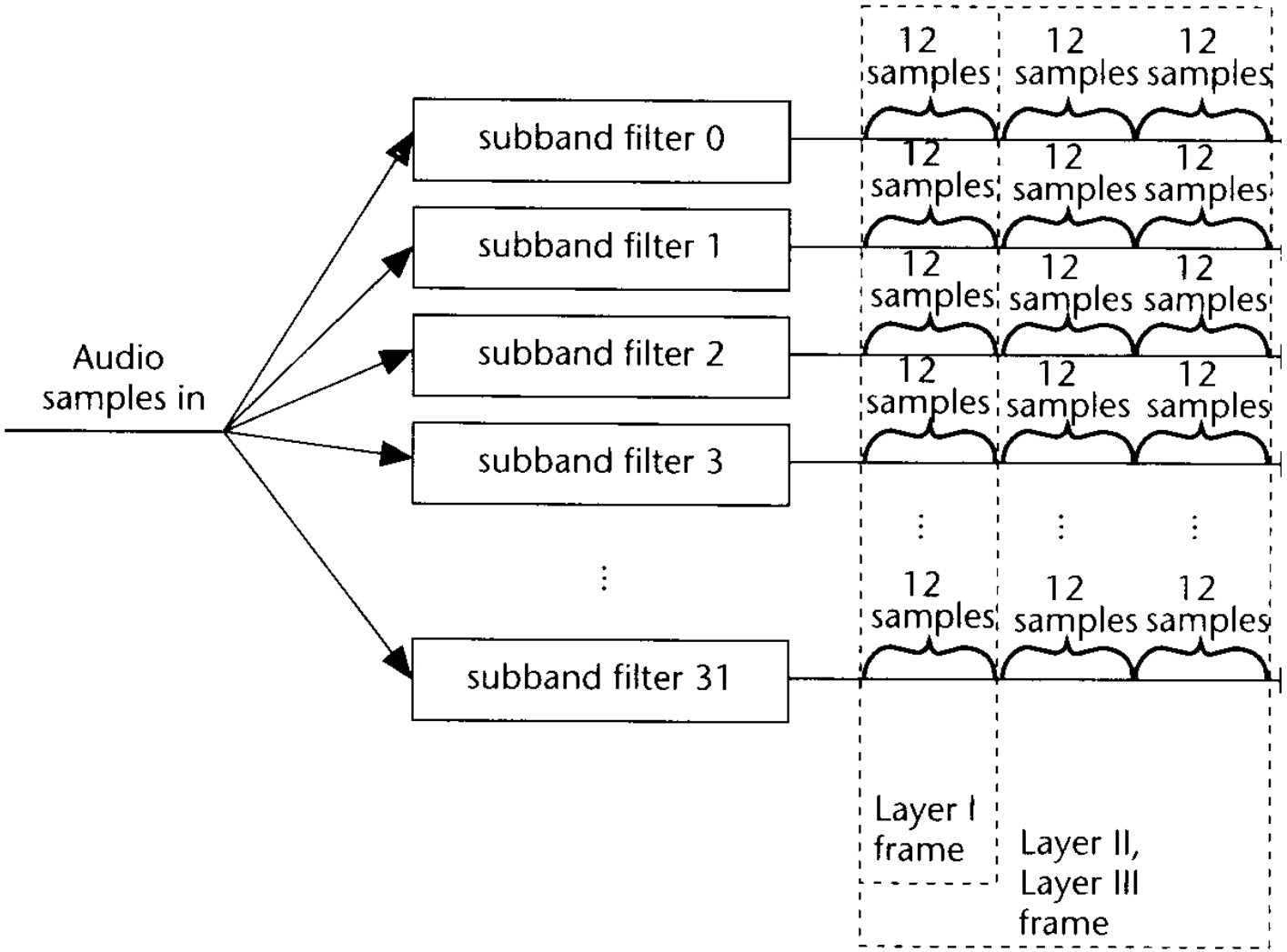
$$MNR = SNR - SMR$$

- Assign sufficient bits to the subband with lowest MNR.
- Apply the criteria to all subbands until all the bits are exhausted.

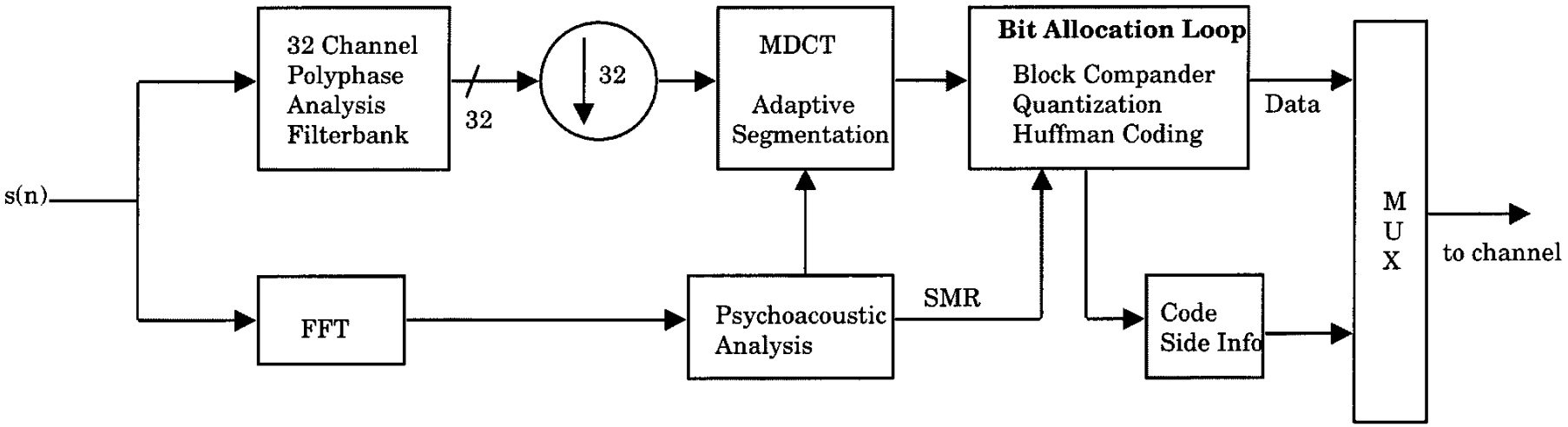
MPEG-1 Layer I and Layer II



MPEG-1 Layer I and Layer II

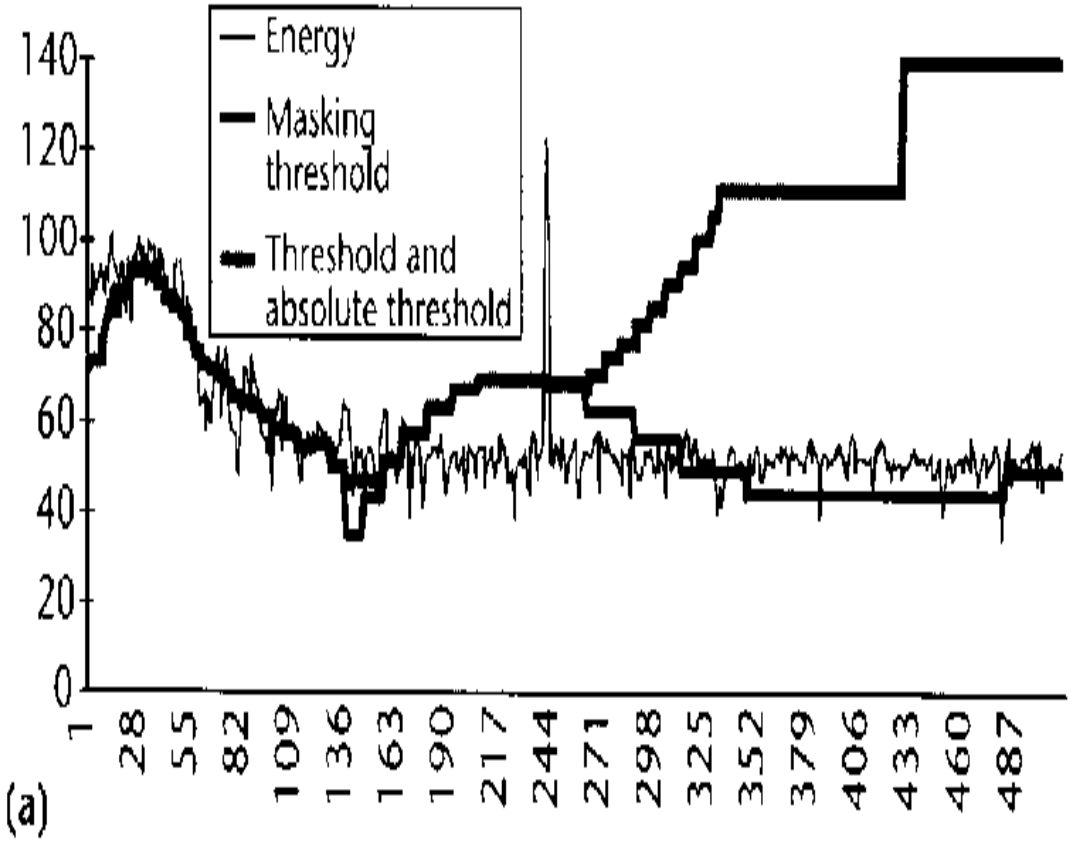


MPEG-1 Layer III (MP3)

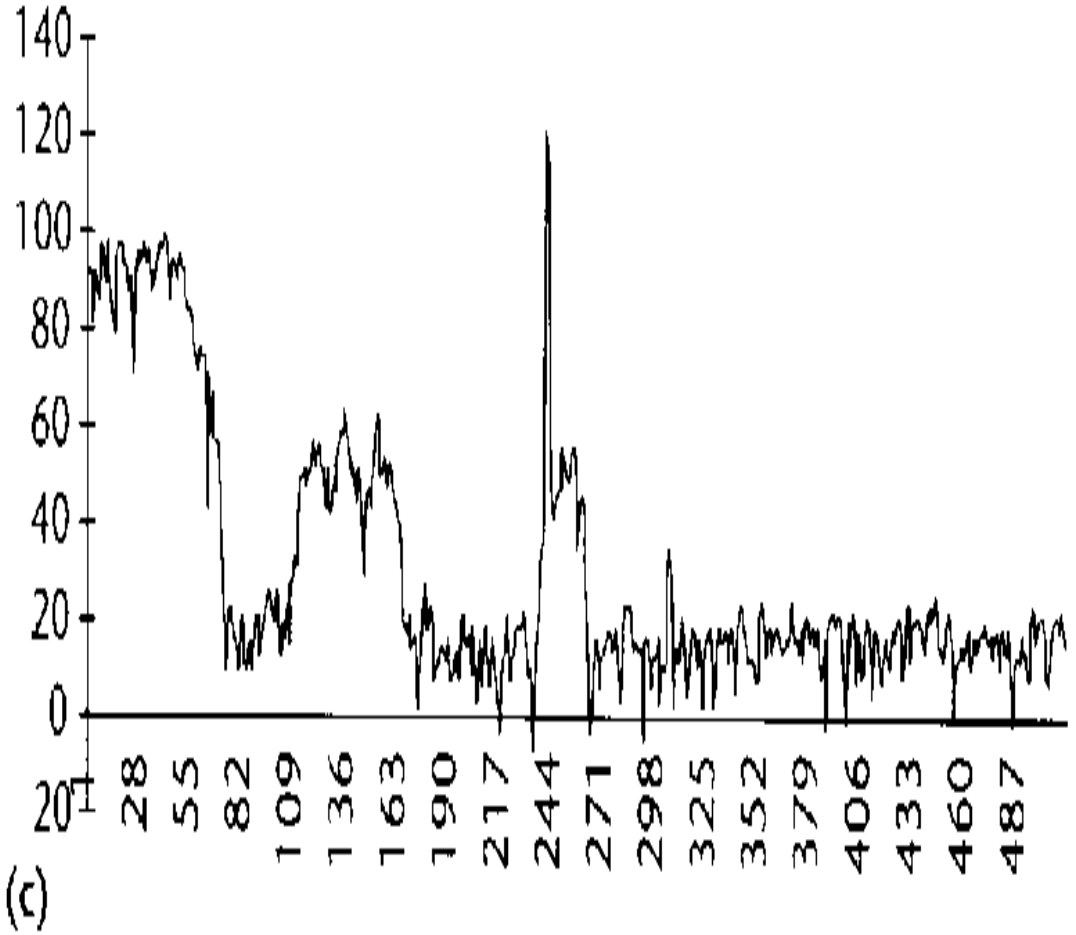


MP3: Psychoacoustic Model 2

Masking threshold estimated using tonality index.



MP3: Psychoacoustic Model 2



MP3: Transform coding

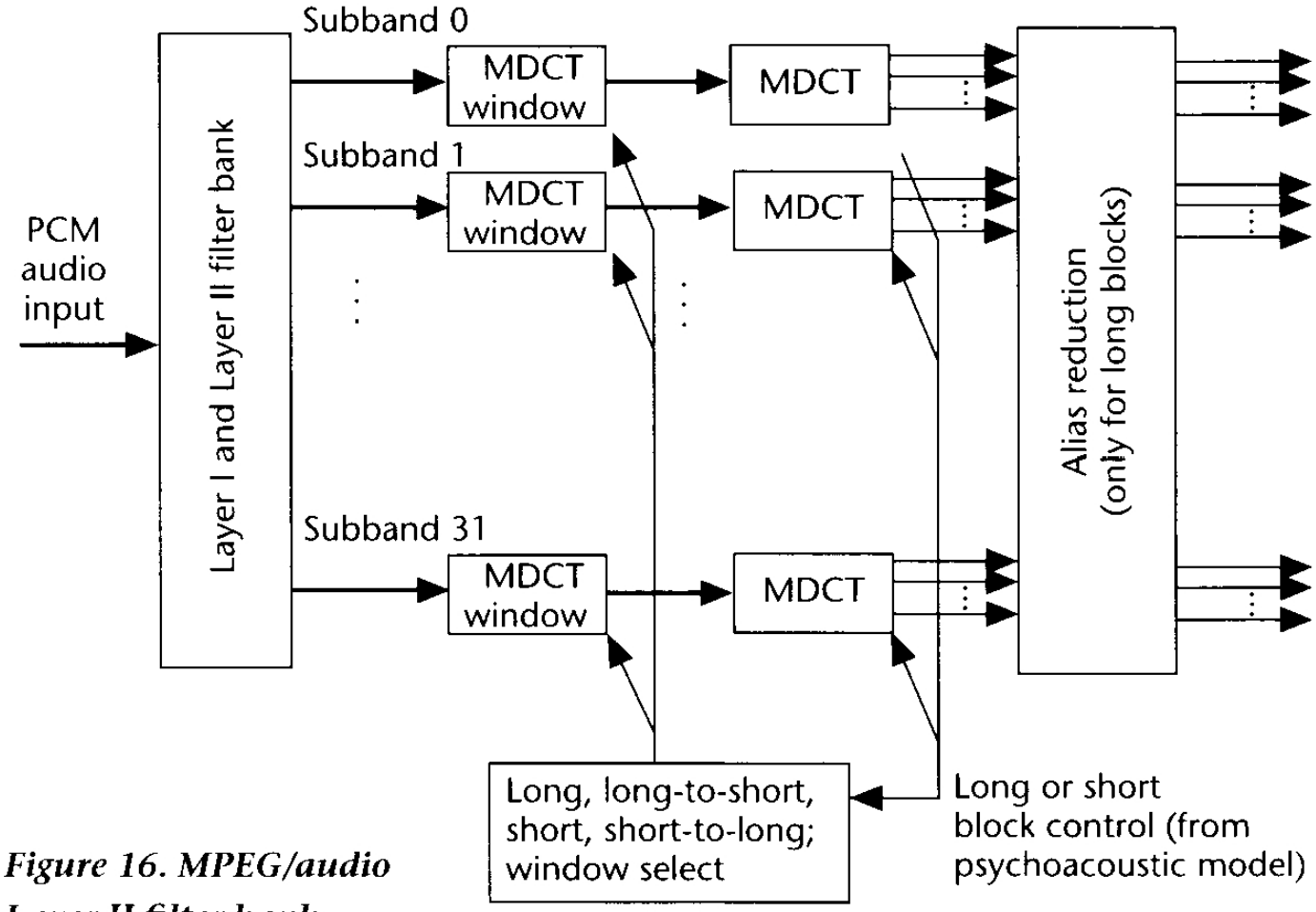
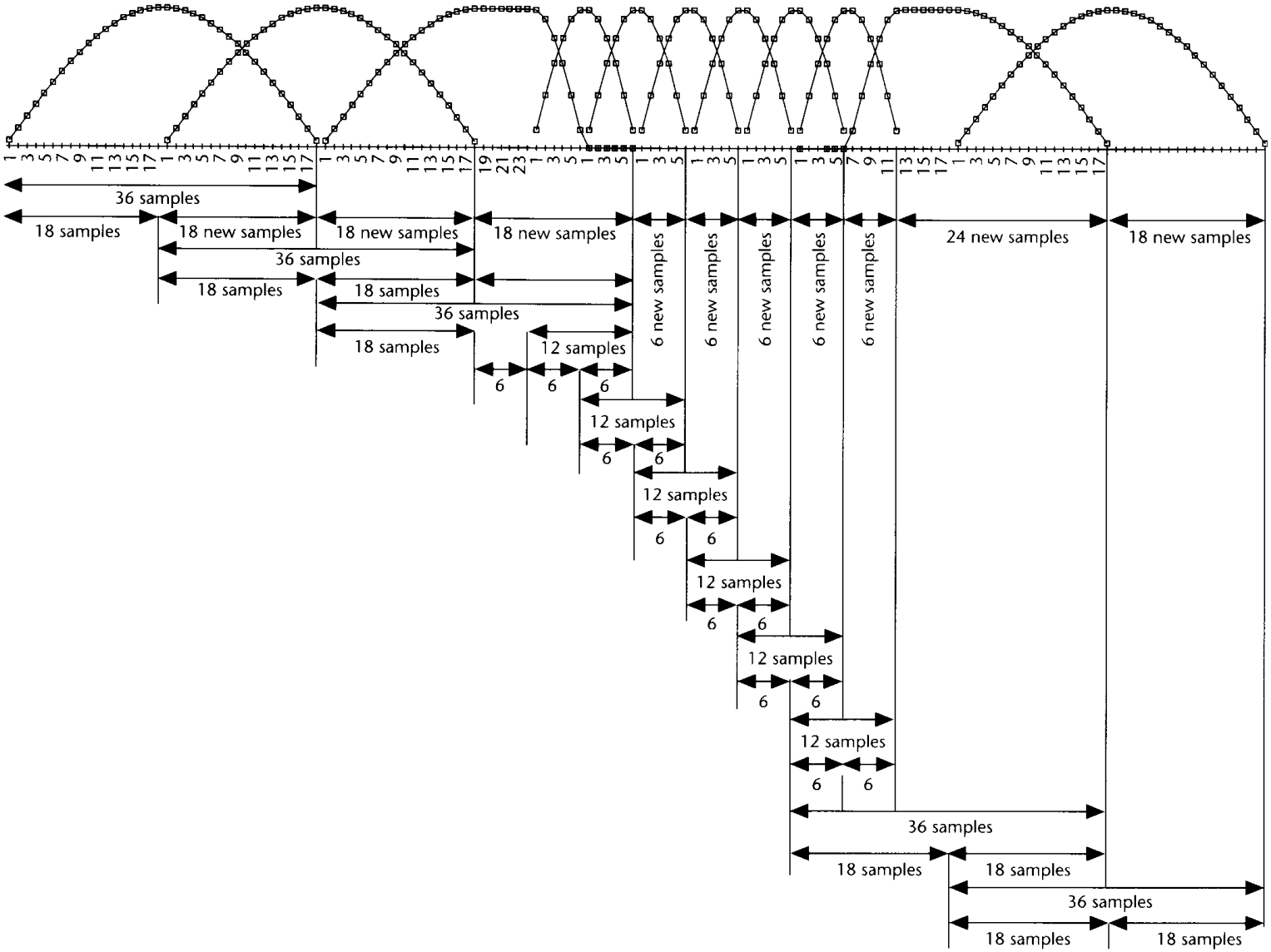


Figure 16. MPEG/audio Layer II filter bank processing (encoder

MP3: Variable Window Length



MP3: Bit allocation

- Nonuniform coding
- Huffman (entropy) coding

Use of bit reservoir.

MPEG-1

- MPEG-1 Layer 1: simple complexity, bit rate above 128 Kbps, compression level 1 : 4, used in Philips Digital Compact Cassette.
- MPEG-1 Layer 2: Intermediate complexity, bit rate around 128 Kbps, compression level 1 : 6-8, used in digital video broadcasting, video CD.
- MPEG-1 Layer 3 (mp3): High complexity, high audio quality for bit rates around 64 Kbps, compression level 1 : 10-12, used for audio transmission over ISDN, internet.