Audio (Music) Compression

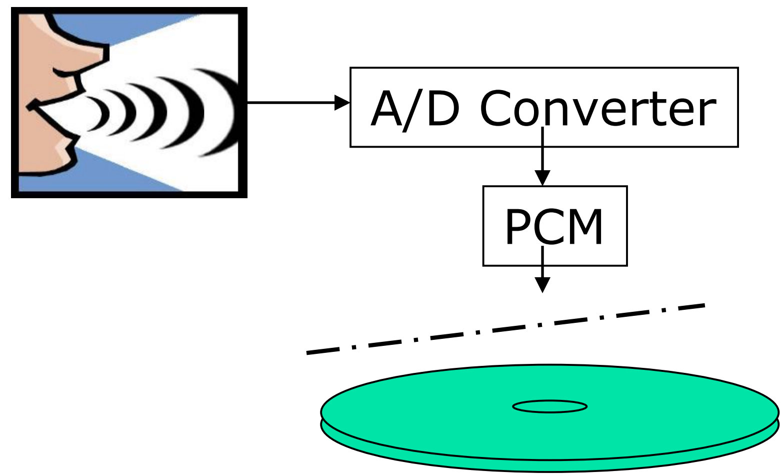
EE443: Design & Application of DSP

Spring 2016

UNIVERSITY OF WASHINGTON
ELECTRICAL ENGINEERING

Digital Audio Recording

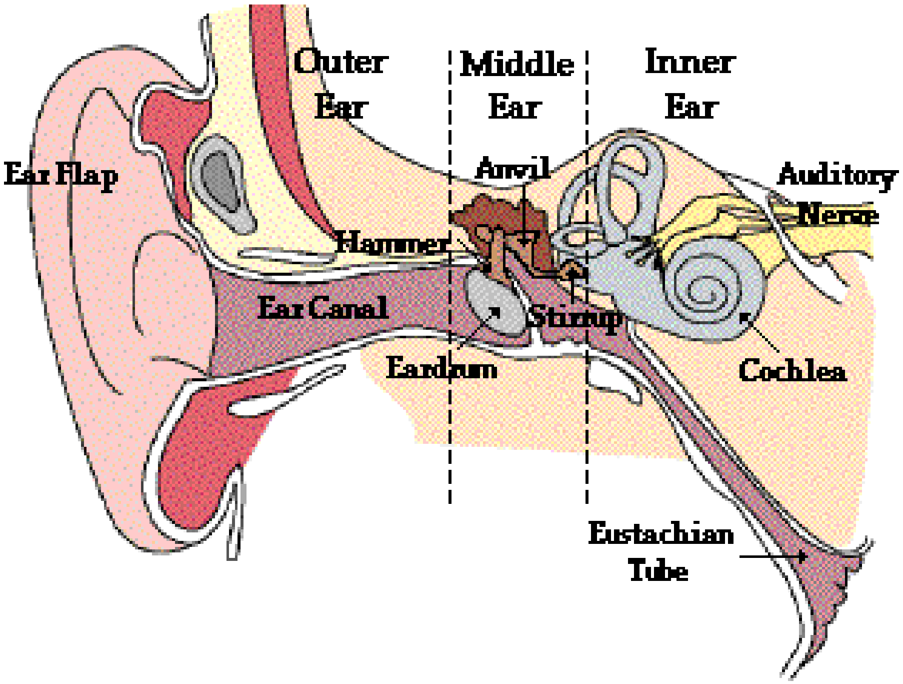
- Philips, 1978, one hour of stereo music



$$44100(samples/\sec)\times 2(channels)\times 2(bytes/sample)\times 3600(\sec/hour) = 635(MBytes, audio\ CD\ capacity)$$

How Human Perceive Sound

- ***Outer Ear:*** directs sounds through the ear canal towards the eardrum.
- ***Middle Ear:*** transforms sound pressure waves into mechanical movement on three small bones.
- ***Inner Ear:*** houses the "cochlea," which converts the middle ear's mechanical movements to basilar membrane movement, and eventually into the firing of auditory neurons which, in turn, send electrical signals to the brain.



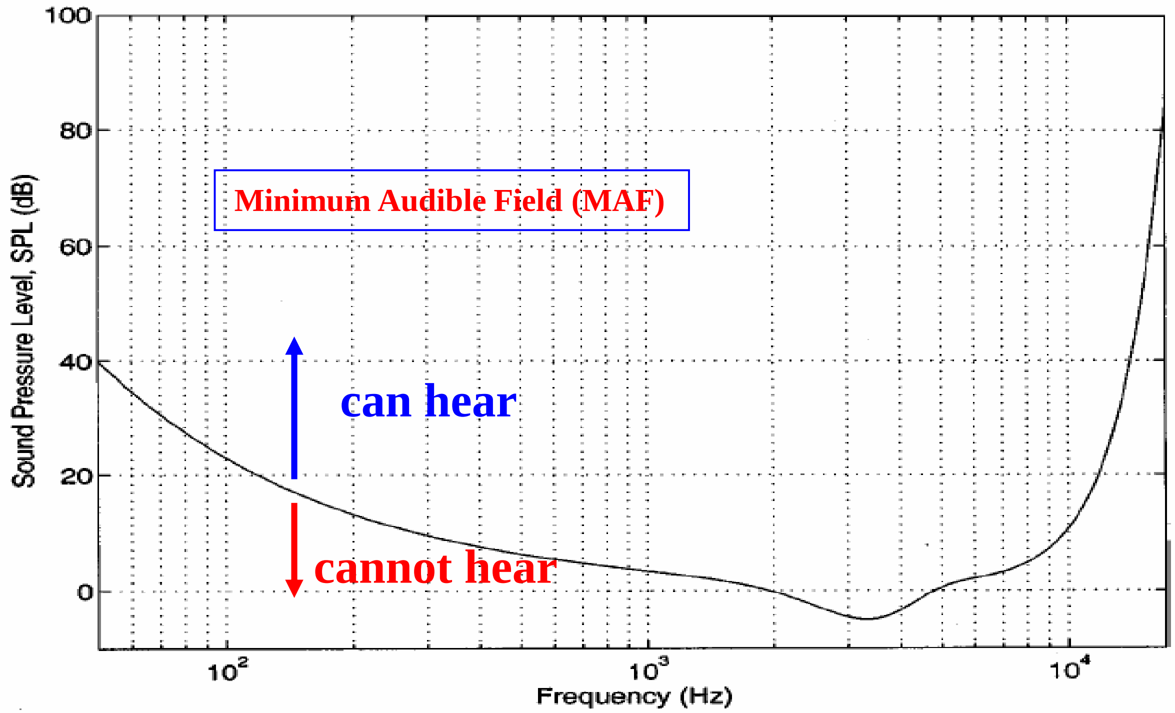
Human Psychoacoustics

- Hearing Sensitivity: Put a person in a quiet room. Raise the level of an 1 kHz tone until it is just barely audible. Vary the frequency and plot these auditory thresholds.
- Frequency Masking: how a loud tone affects neighboring frequency tones in human perception
- Temporal Masking: how a loud tone affects subsequent tones in human perception

Hearing Sensitivity

$$SPL = 20\log_{10}\frac{P}{P_0} \text{ (dB)}$$

$$P_0 = 20\times10^{-6} \text{ pascals}$$



- Human perception of sound is a function of frequency (critical bands) and signal strength
- Minimum Audible Field (MAF): the threshold of hearing

Critical Bands and Barks

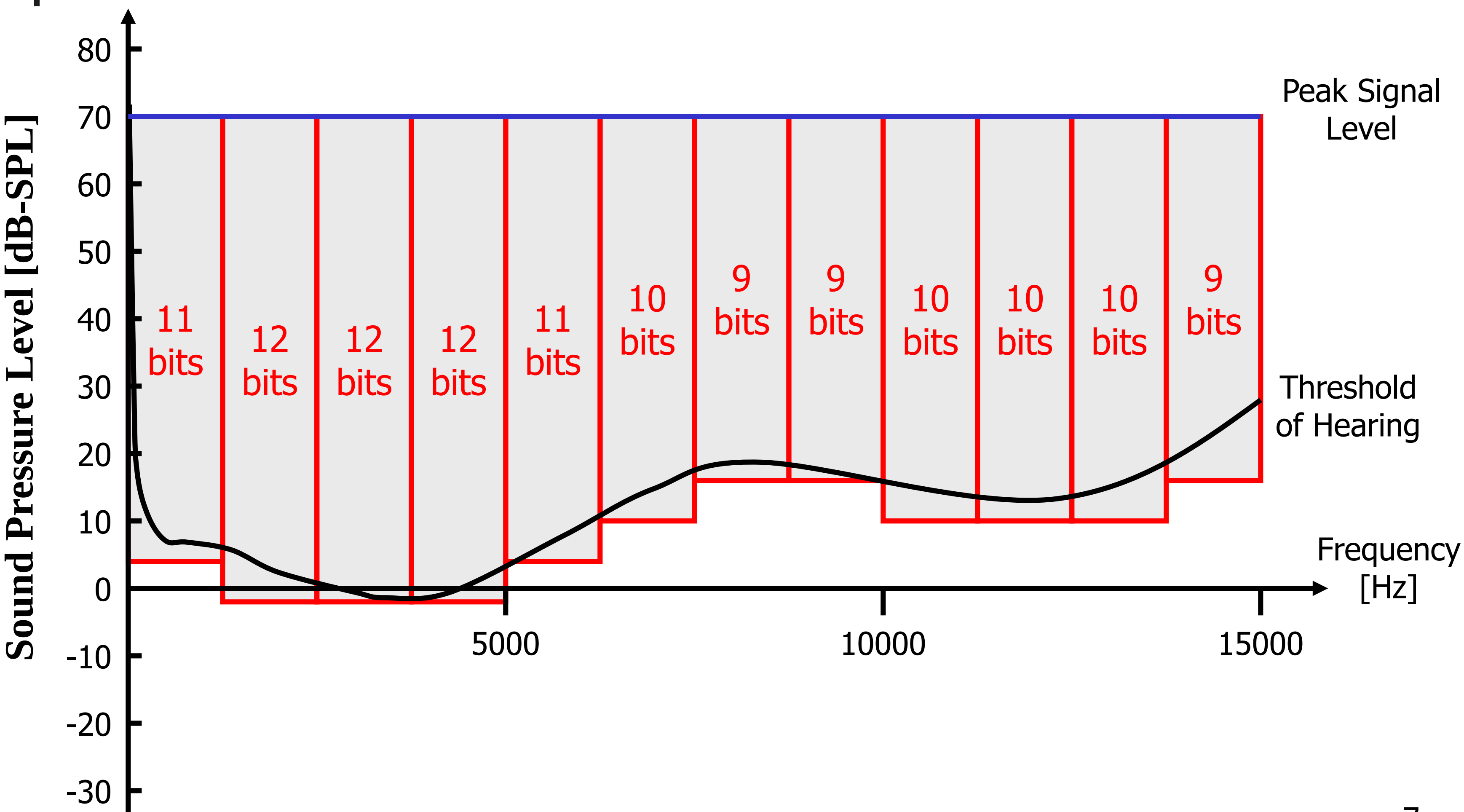
- The frequencies within a critical band are similar in terms of the ear's perception, and are processed separately from other critical bands.
- The bands have width less than 100 Hz at the lowest audible frequencies, and more than 4 KHz at the high end.
- Approximately, the audio frequency range can be partitioned into 25 critical bands.
- 1 Bark = the width of one critical band
- For linear frequencies *f* less than 500 Hz

  frequency in barks: $b = f/100$

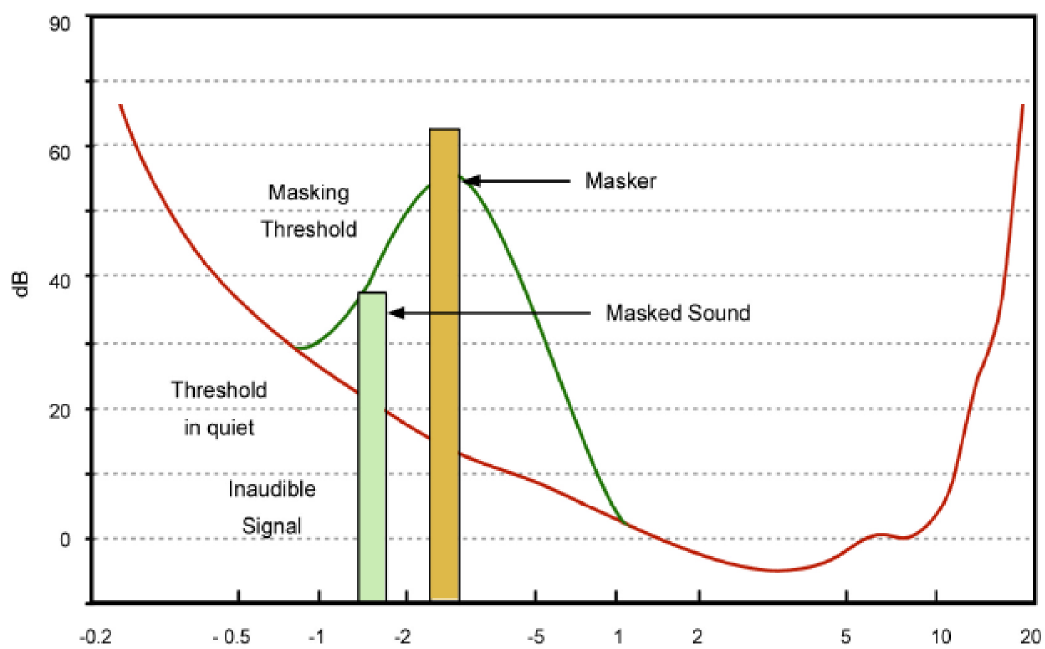
- For linear frequencies *f* greater than 500 Hz

  frequency in barks: $b = 9 + 4\log_2(f/1000)$

Hearing Sensitivity: Quantization under MAF



The Role of Frequency Masking in our Hearing



Signal

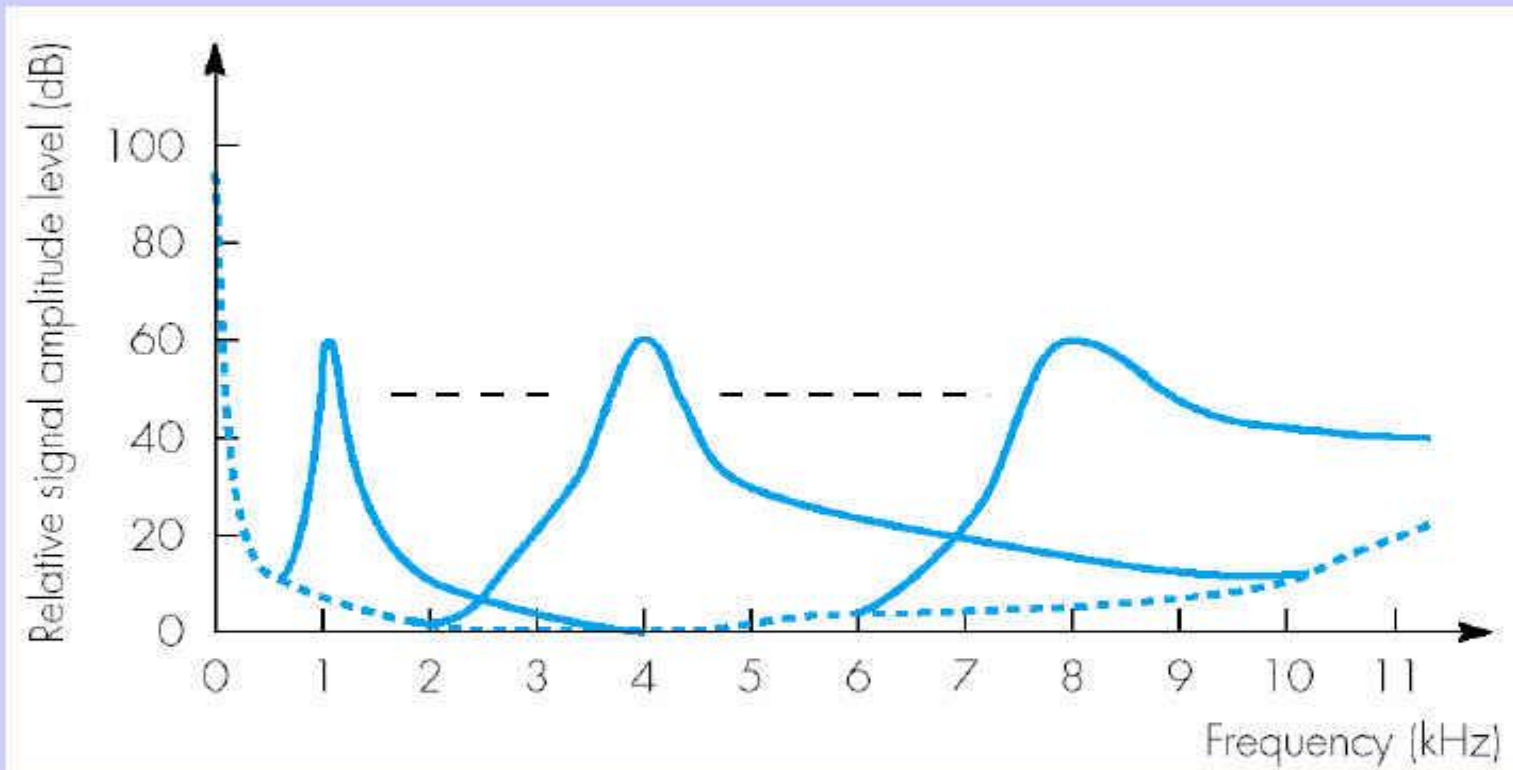
Noise

Signal + Noise
(SNR = 24 dB)

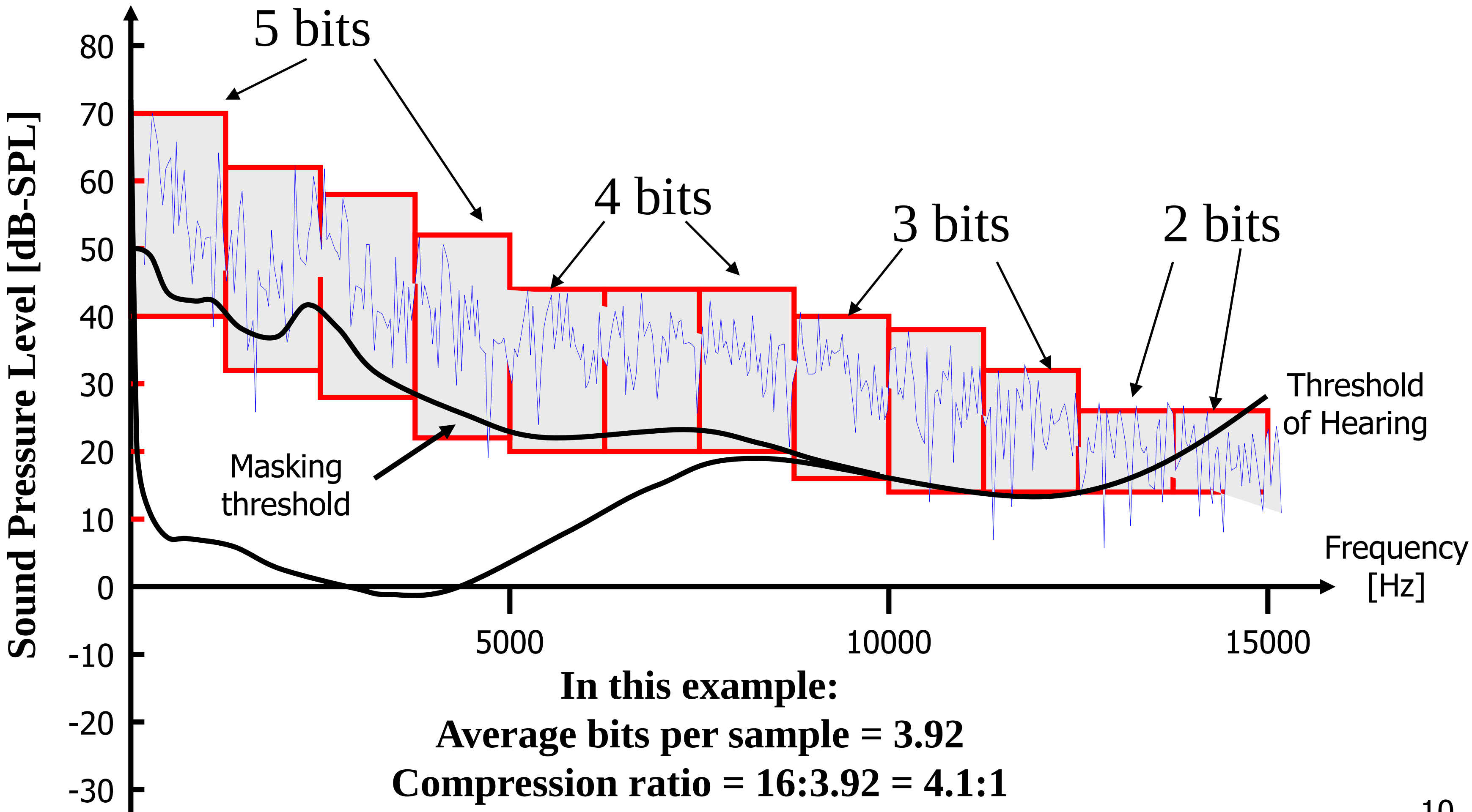
- Do our ear's frequency channels interfere with each other? This effect is called "frequency masking"

Frequency Masking (cont)

- Repeat for various frequencies of masking tones



Saving More Bits with Frequency Masking



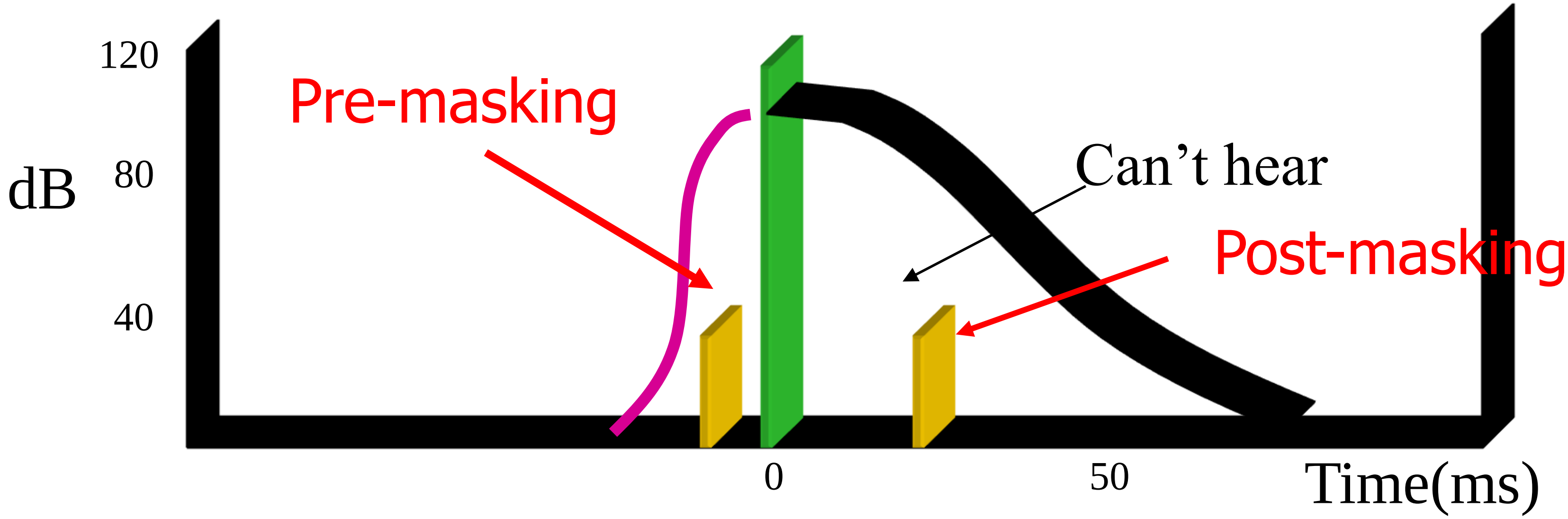
In this example:

Average bits per sample = 3.92

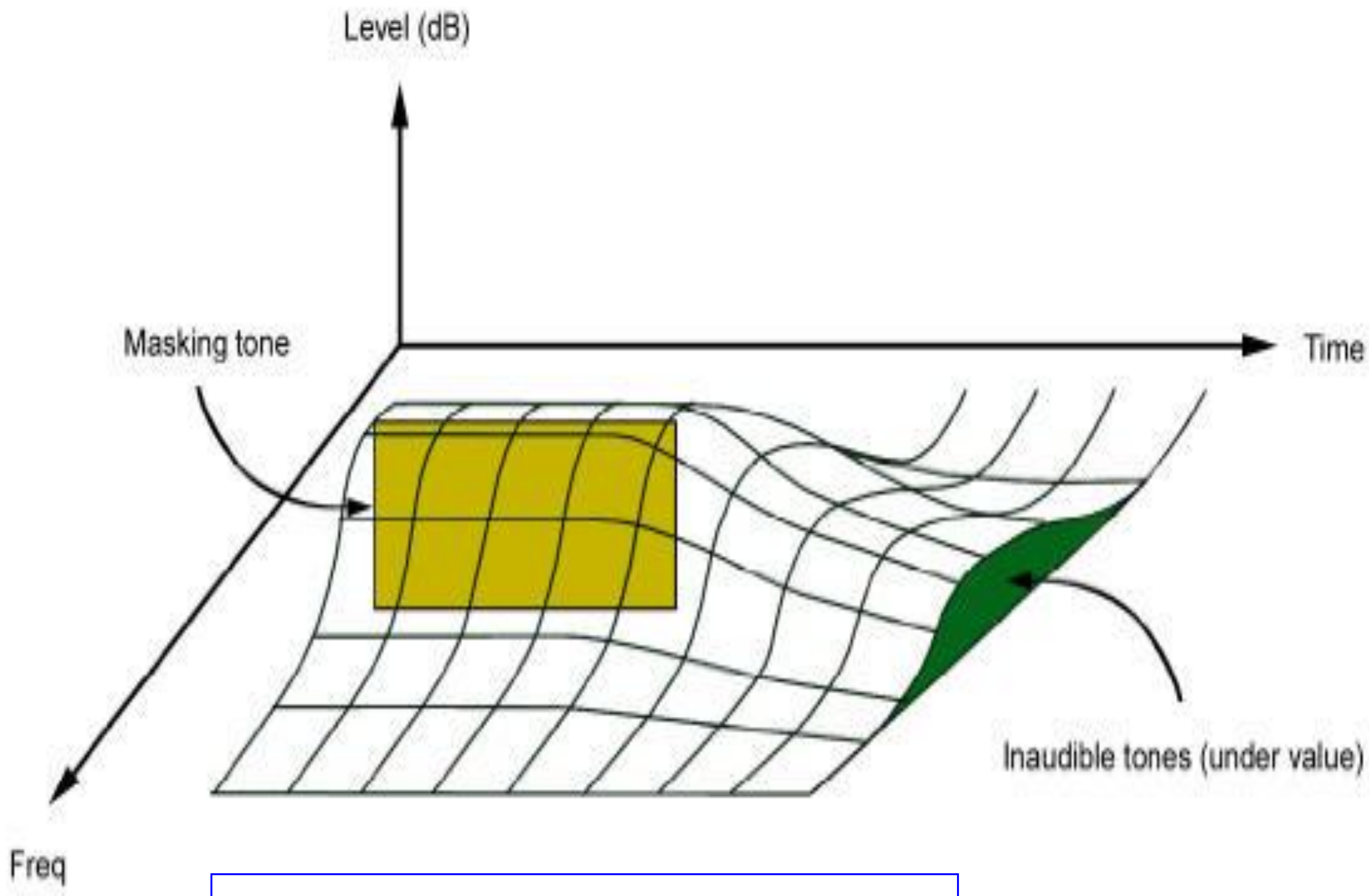
Compression ratio = 16:3.92 = 4.1:1

Temporal Masking

- May occur when two sounds appear within a small interval of time.
- The duration within which pre- masking applies is significantly less than one tenth of that of the post- masking which is in the order of 50 to 200 ms.



Combined Frequency & Temporal Masking

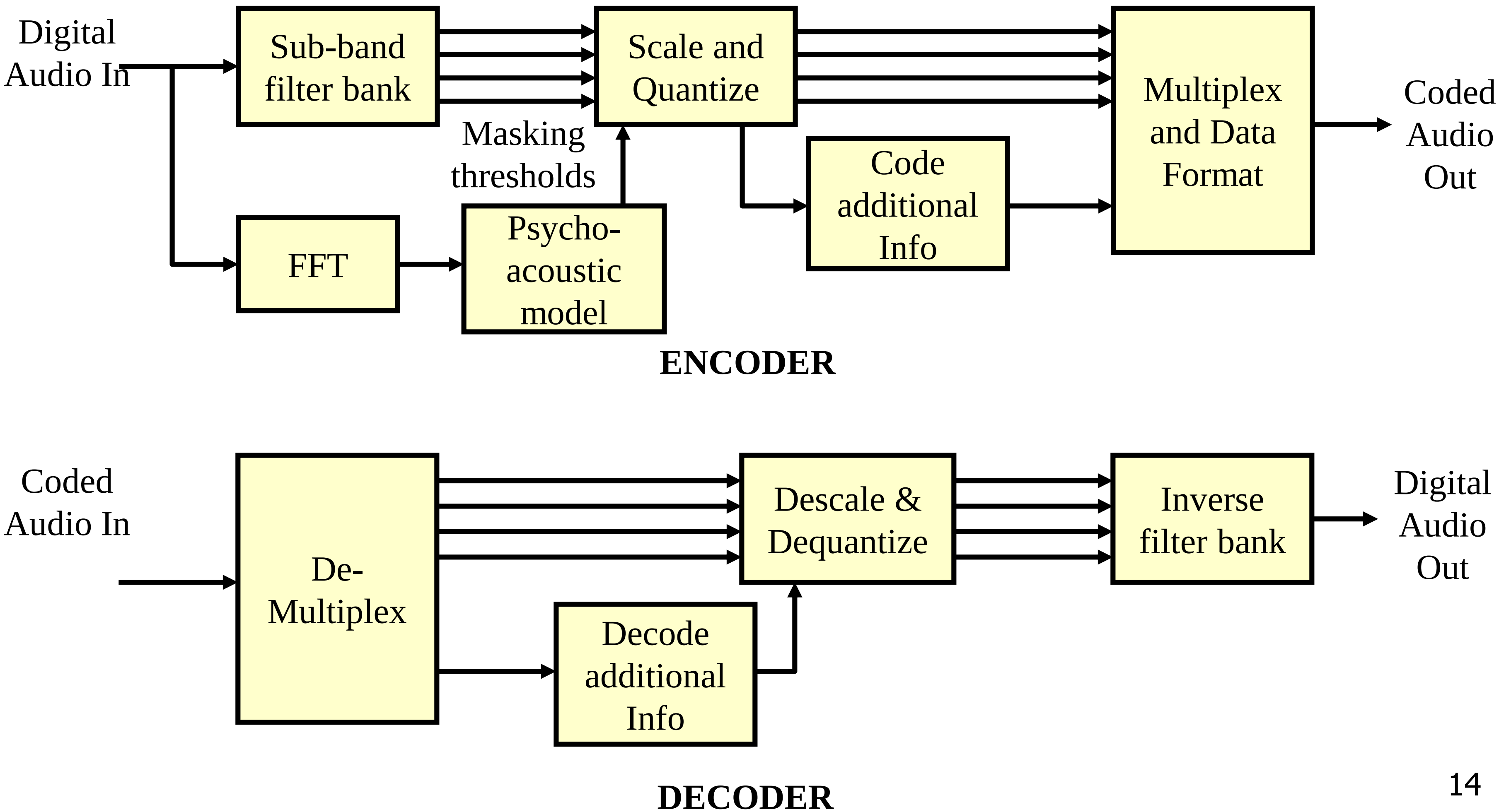


By Steve Church, Telos Systems.

Frame Based PsychoAcoustic Analysis

- The masking analysis and bitrate assignment are prepared and processed in *frames* of a predetermined length (1152 samples, at 44.1 KHz sampling rate) of audio signal
- Each frame contains mostly coded audio, but in addition:
  - The peak level in each frequency sub band
  - The masking level in each sub band
  - The number of bits in each sample in each sub band

A Generic Audio Codec



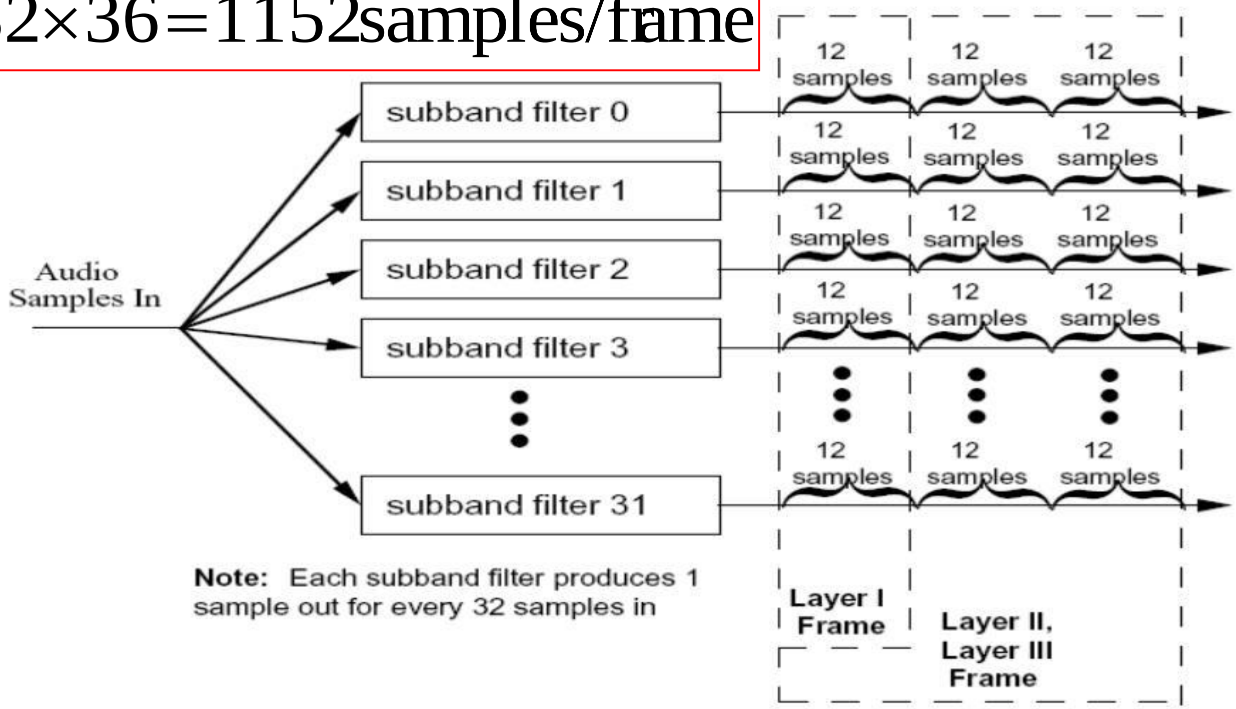
Audio Coding Standards

- MPEG-1 three Audio Layers
  - *MP3* is actually *MPEG-1 Layer 3*.
- Dolby AC3 Audio Coding (5.1 channels, DVD)
- MPEG-2 Audio (Backward Compatible "BC", and Advanced Audio Coding "AAC")
- MPEG-4 Audio (AAC) & High Efficiency Extensions (HE-AAC)
- Microsoft Windows Media 9 Audio (WMA9)
  - multi-channel music distribution and movie sound tracks at broadband rates (e.g., encode 5.1 channels at 128 Kbps in total)

MPEG-1 Audio Coding

- In MPEG audio: 32 equal-width (uniform) frequency subbands
- Use of relatively simple polyphase filter structure.

$\boxed{32 \times 36 = 1152 \text{samples/frame}}$

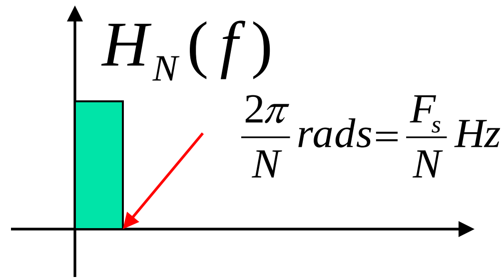


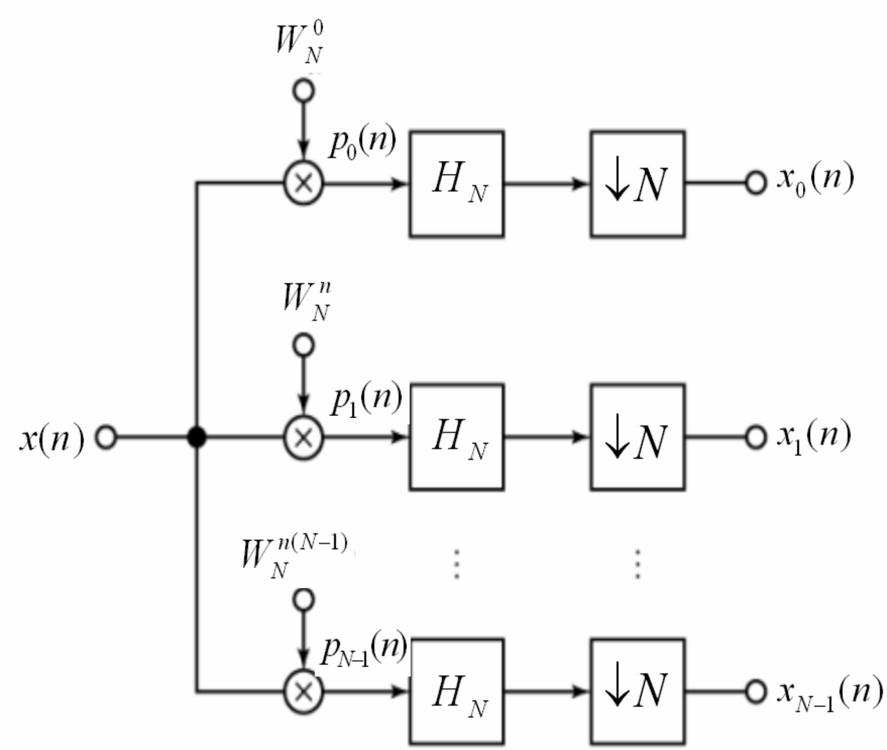
36-point modified DCT before quantization In MPEG-1 layer III (MP3)

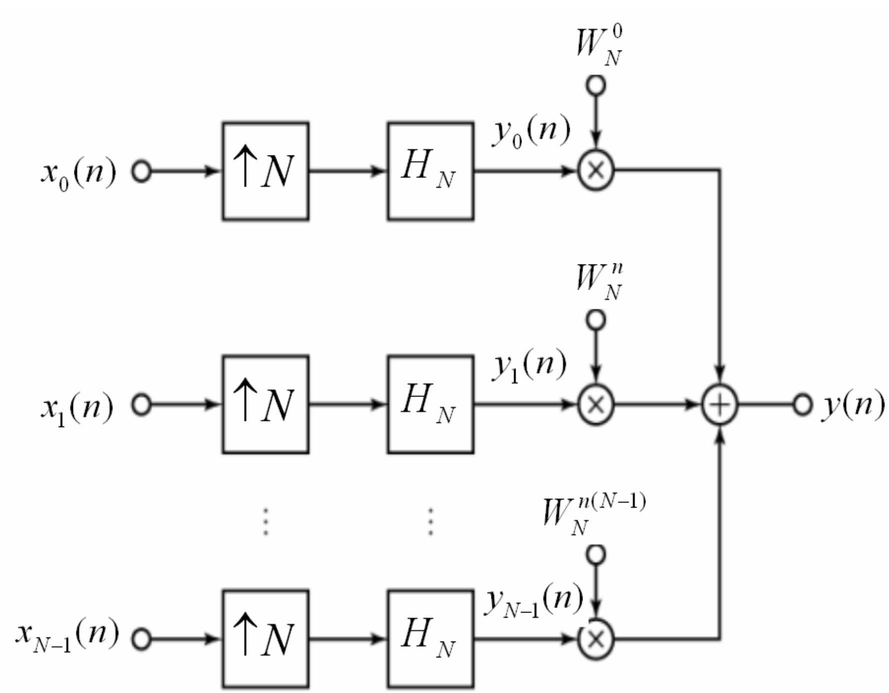
Subband Filtering

$$W_N^k = e^{-j\frac{2\pi k}{N}}$$

shift left the frequency contents

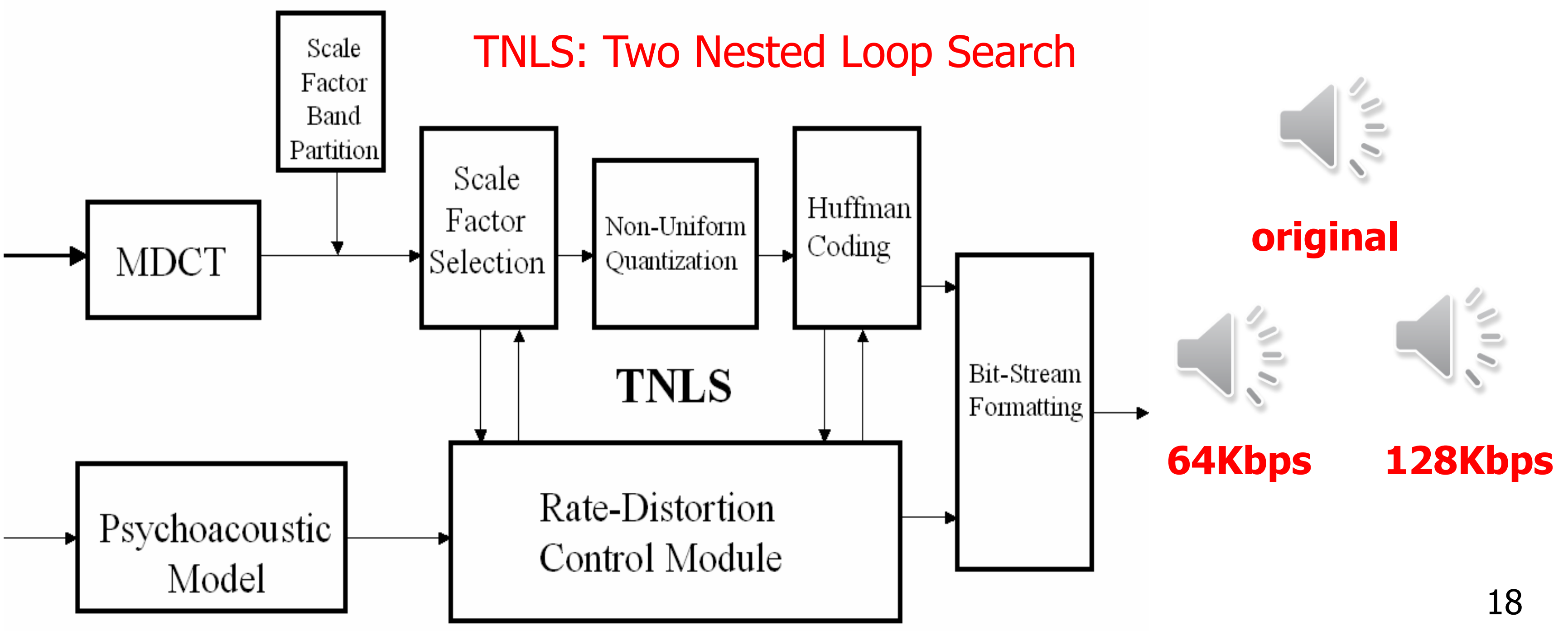






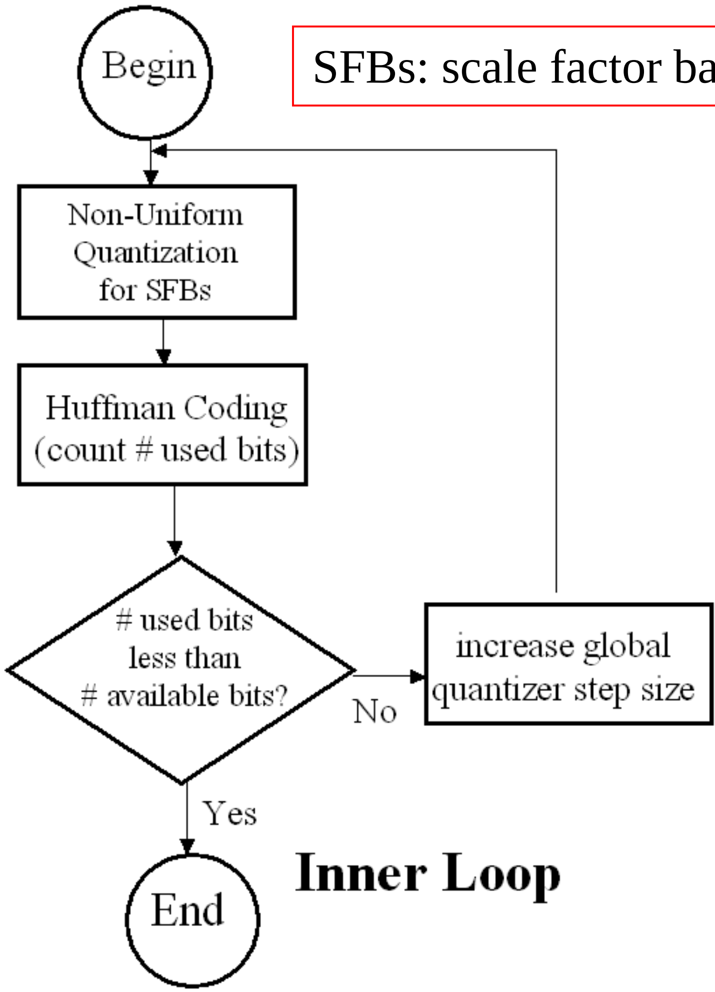
MP3 Bit Allocation

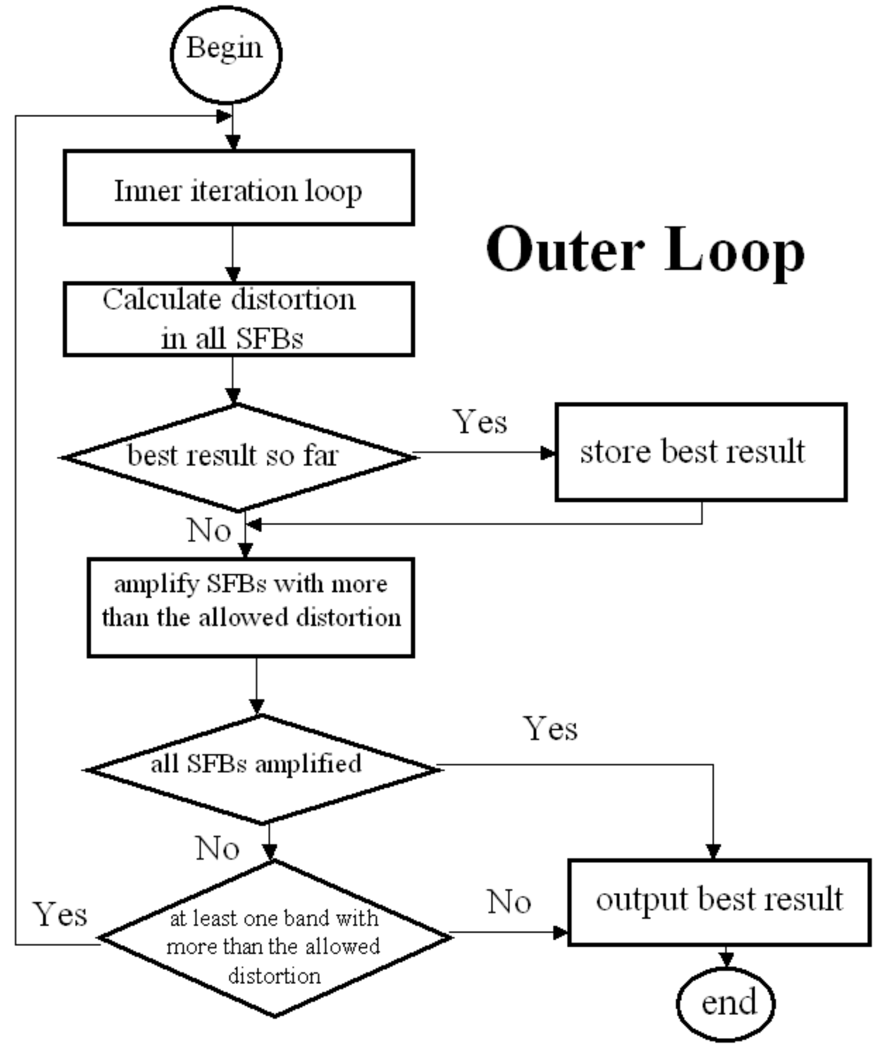
- Raises its MDCT inputs to the 3/4 power before quantization
- The scale factors are weights for ranges of frequency coefficients called scale factor bands (SFBs) – equivalent to change the quantization steps.



Two Nested Loop Search

SFBs: scale factor bands





MS Stereo Coding

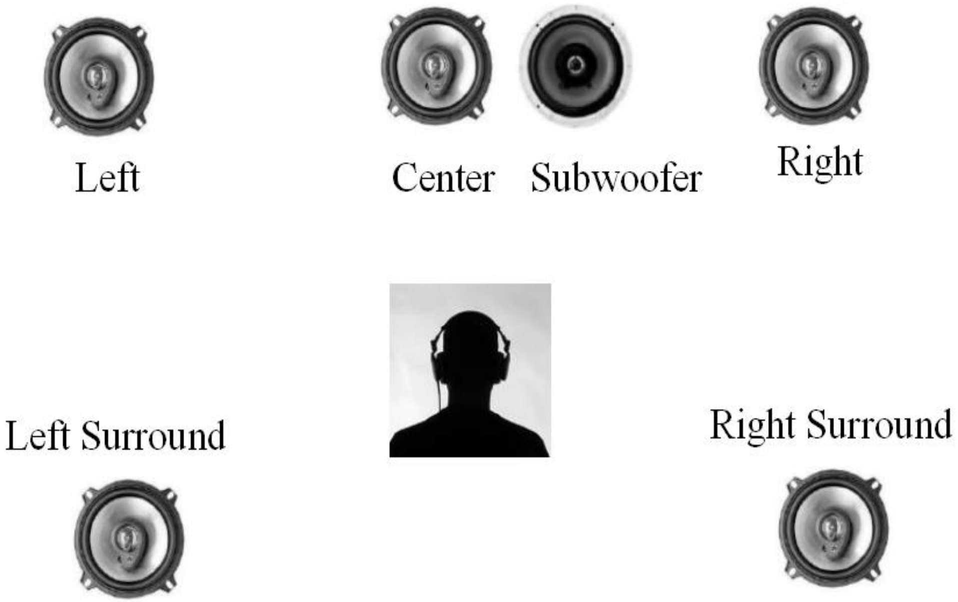
- Encodes the left and right channel signals as middle (sum of left and right) and side (difference of left and right) channels.
- When there is too much difference between the L and R channels, real stereo will be used on that frame.
- Joint stereo is commonly used in MP3 with a bitrate of 128 Kbps or lower to improve the quality.

MPEG2 Advanced Audio Coding (AAC)

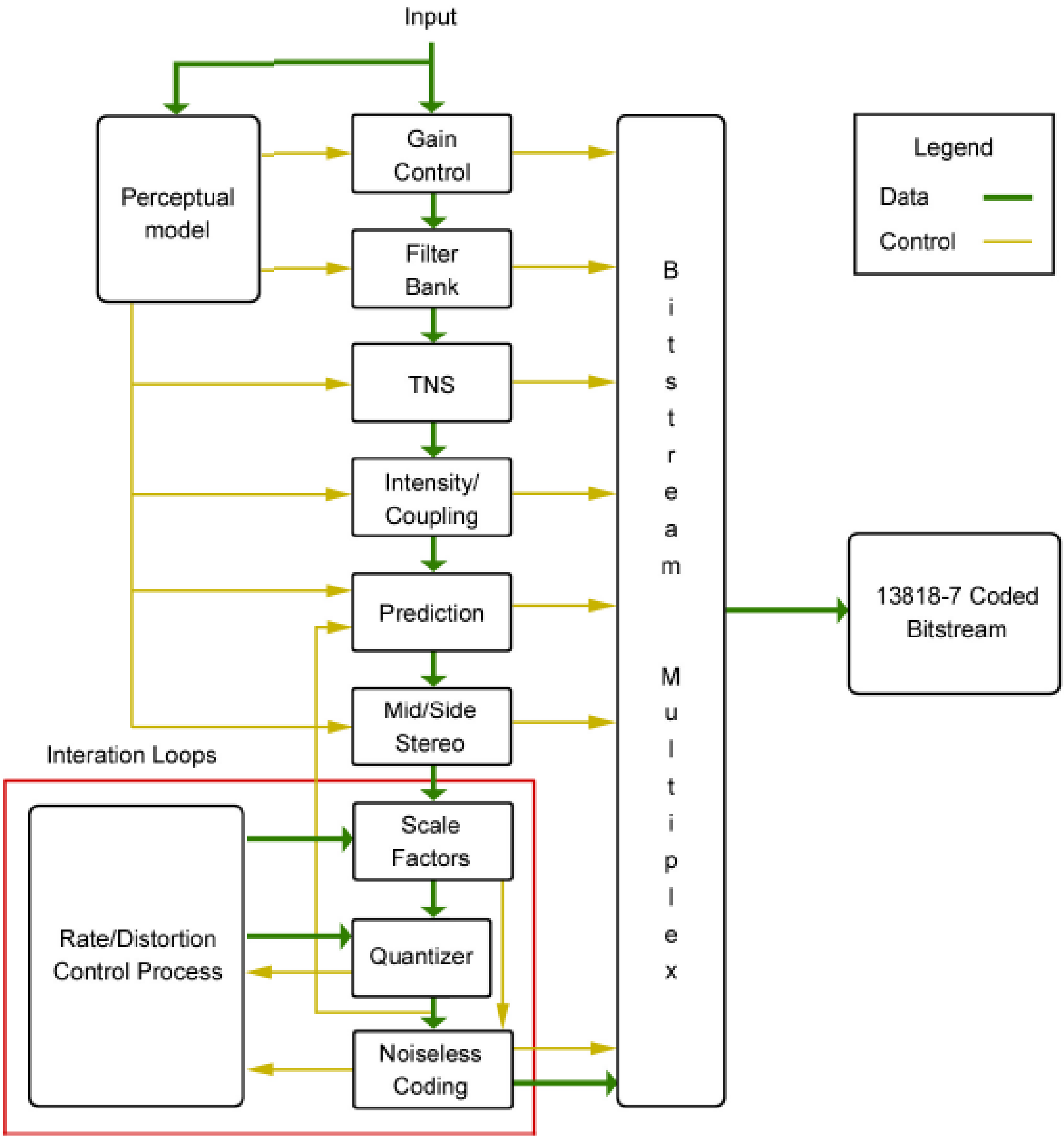
- Developed by the Fraunhofer Institute in conjunction with companies like AT&T, Sony, and Dolby for MPEG2 in 1997.
- Technically, the AAC format can support up to 48 full frequency sound channels, so 5.1 or 7.1 sound is entirely possible. It also supports sample rates up to 96KHz, twice the maximum afforded by MP3 (quality better than AC3). About 32 Kbps/channel
- AAC is the format used for songs downloaded in the popular iTunes Music Store.

Surrounding Sound

- 5 full-bandwidth channels (left, center and right, left surround and right surround, 3-20K Hz) and one subwoofer -- low frequency enhancement (3-120 Hz).



MPEG-2 AAC Encoding

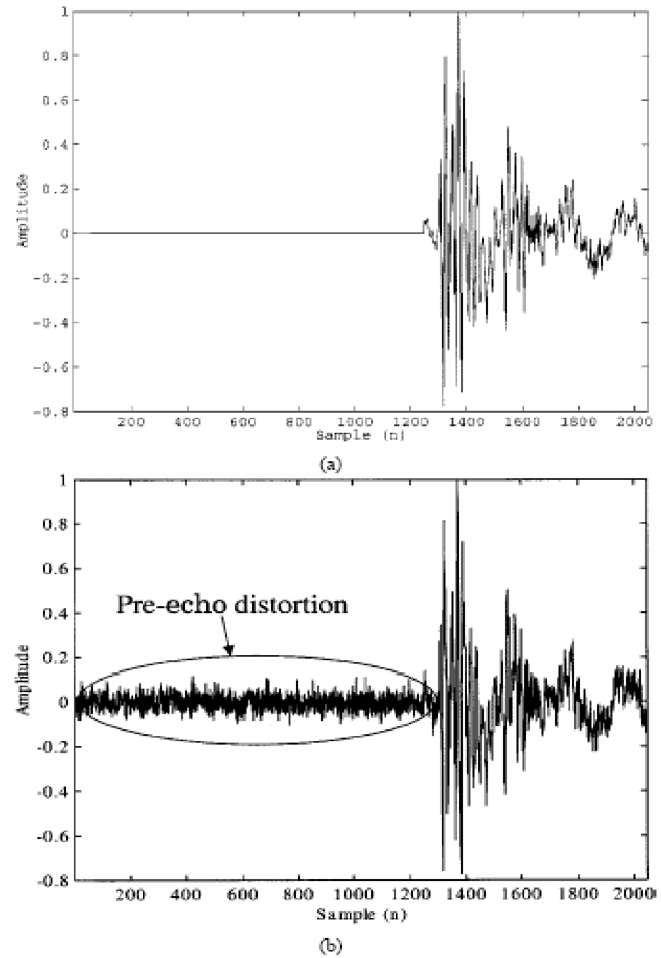


Filter Banks

- In contrast to MP3 which uses subbands and MDCT, AAC uses a plain MDCT, together with the large transform length (8x128=1024 spectral lines per transform with 50% overlap), adapted window shape function and transform block switching.
- AAC's plain MDCT provides better frequency selectivity for the filter bank.
- AAC also uses some short windows, which are smaller than MP3 ones, to provide better transients handling and less pre-echo.

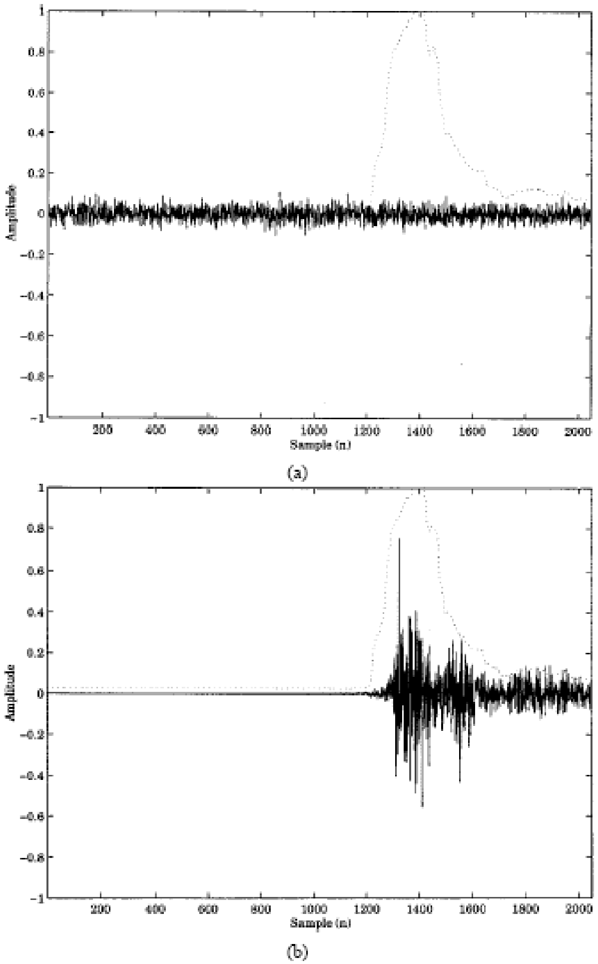
Pre-Echo Distortion

- A sharp attack begins near the end of a transform block immediately following a region of low energy (e.g., audio recordings of percussive instruments)
- A block-based algorithm (e.g., MP3), the inverse transform will spread quantization distortion evenly in time throughout the reconstructed block.



Temporal Noise Shaping

- Frequency Domain Linear Prediction (LP) across frequency (rather than time), since for an impulsive (transient) time signal, it exhibits "tonal" in the frequency domain, i.e., a few sinusoidal components in frequency domain.
- Effectively separates the time-domain waveform into an envelope and temporally flat "excitation", i.e., shapes the quantization noise to follow the input signal's energy envelope.



Other Techniques in MPEG-2 AAC

64Kbps

128Kbps

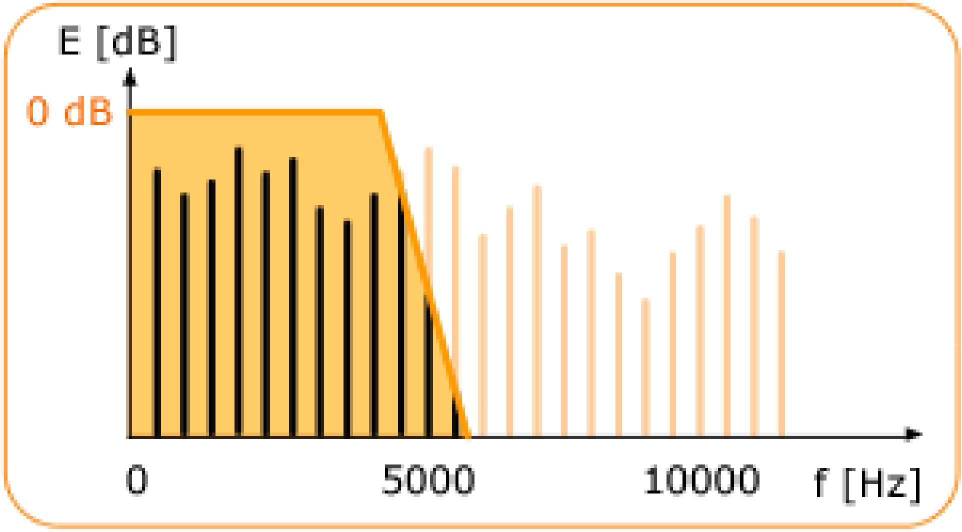
- ***Time Domain Linear Prediction:*** The dual of TNS is that a tonal signal in the time domain has transient peaks in the frequency domain..
  - A second-order backward-adaptive predictor reduces the bit rate for coding subsequent subband samples in a given subband based on the quantized spectrum of the previous frame.
- ***MS Stereo:*** toggle middle/side stereo on a subband basis instead of entire frame basis as in MP3.
- ***Quantization:*** by allowing finer control of quantization resolution, the given bit rate can be used more efficiently.
- ***Huffman Coding:*** One scale factor Huffman codebook and 11 spectrum Huffman codebooks are used in MPEG-2 AAC.

MPEG-4 High Efficiency HE-AAC version 1 (v1)

- A low bit rate codec in the AAC family: a combination of the MPEG-2 AAC LC (Advanced Audio Coding Low Complexity) audio coder and the Spectral Band Replication (SBR) bandwidth expansion technique -- 30% more efficient than MPEG-2 AAC.
- Targeted to medium-quality encoding at bitrates 24 Kbps/channel and higher.
- Also known as AACPlus and can be used in multi-channel operations.
- Not intended as a replacement for AAC LC, but rather its extension, and is proposed mainly for internet, mobile, and broadcasting arenas.

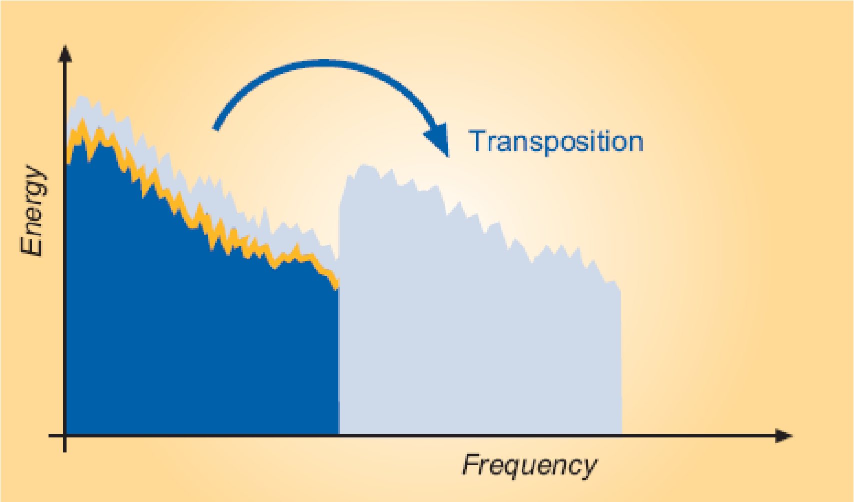
Bandwidth Reduction in Audio Coding

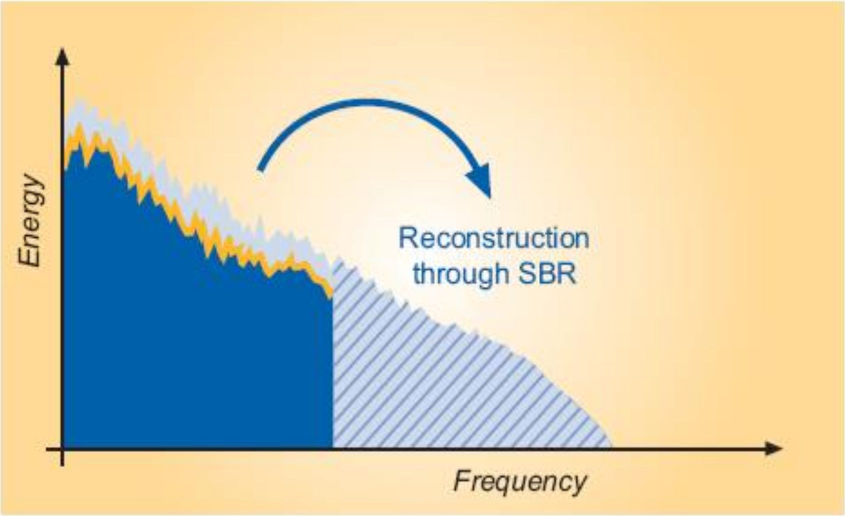
- The low bitrate audio codecs either reduce the audio bandwidth or introduce annoying coding artifacts → SBR
- The 64 Kbps MP3 offers an audio bandwidth of only about 10 KHz and introduce a fair amount of coding artifacts.



example of band-limiting of a typical signal

Spectral Band Replication (SBR)

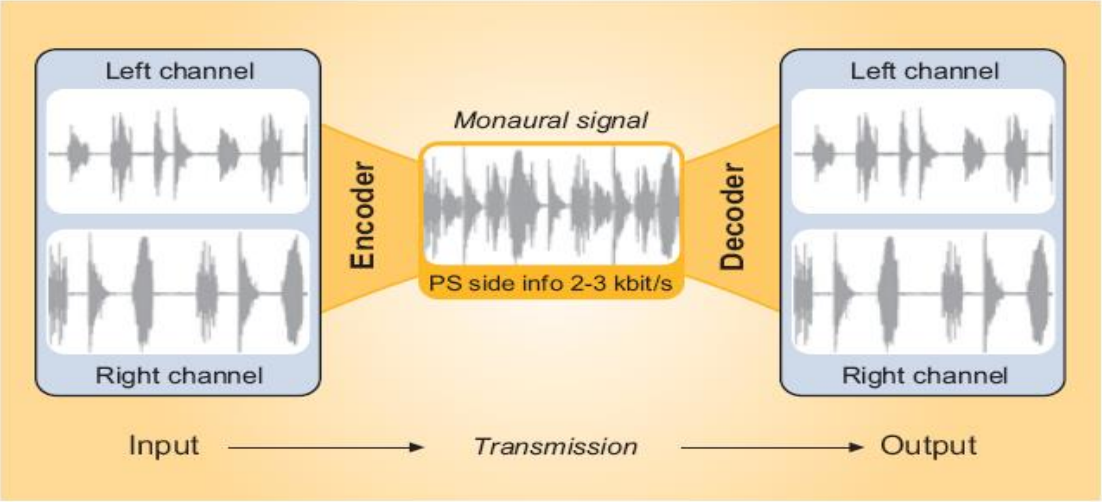




Stereo Bitrate (Kbps)	AAC Frequency Rang (Hz)	SBR Frequency Range (Hz)
20	0-4500	4500-15400
32	0-6800	6800-16900
48	0-8300	8300-16900

MPEG-4 High Efficiency HE-AAC version 2 (v2)

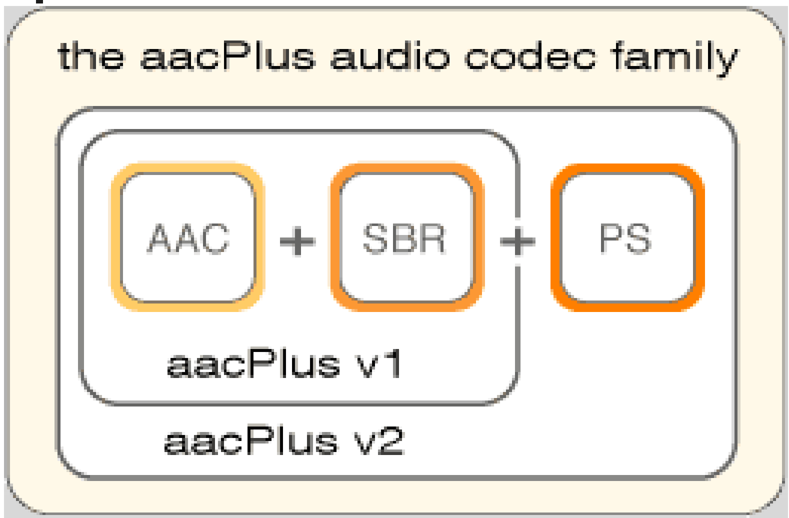
- The extension to HE AAC v1 for extremely low bitrates, such as 32 Kbps for stereo input (i.e., 16 Kbps/channel).
- Using **PS** (Parametric Stereo, one combined mono plus 2-3Kbps side information) technology, HE AAC v2 is nearly 50% more better than HE v1 when used for internet, mobile, broadcasting, and other domains with limited resources.

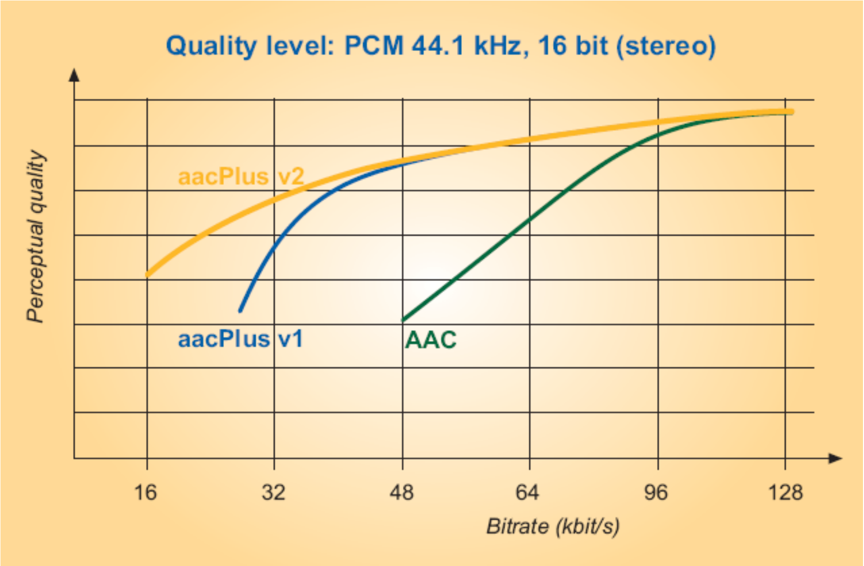


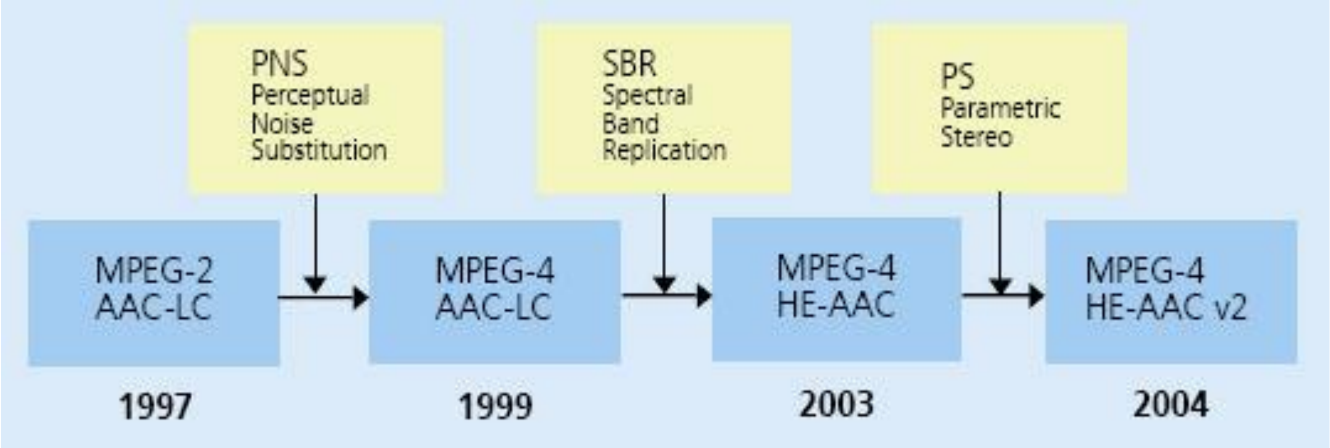
Parameters of PS

- *Inter-channel Intensity Difference (IID),* describing the intensity difference between the channels.
- *Inter-channel Cross-Correlation (ICC),* describing the cross correlation or coherence between the channels. The coherence is measured as the maximum of the cross-correlation as a function of time or phase.
- *Inter-channel Phase Difference (IPD),* describing the phase difference between the channels.

Performance Comparison







Final Project

- Proposal: A 4-page proposal, due May 5th, (email to me)
  - 1. motivation
  - 2. problem formulation
  - 3. proposed tasks and data collection procedures
  - 4. timeline and individual responsibility
  - 5. References
- Final Report and Demo: A 20-minute group presentation of the project on June 1st, demo on June 8^{th}, and A 15-page final project group report, due June 10th
  - 1. Problem formulation
  - 2. Project algorithms and implementation
  - 3. Results and discussions
  - 4. References