

# Semi-parametric audio coding - Today and beyond

Jonas Engdegård, Senior Research Engineer  
Coding Technologies

2007-07-05

- About Coding Technologies
- Perceptual Audio Coding Overview
- Semi-Parametric Audio Coding Tools
  - Spectral Band Replication (SBR)
  - Parametric Stereo (PS)
  - MPEG Surround (MPS)
- Future Trends in Audio Coding
- Summary

# About Coding Technologies

---

- 1997: Coding Technologies founded
- 2000: Merged with spin-off from Fraunhofer IIS (“Home of mp3”), Germany
- 2007: 62 employees, 4 office locations
  - Nuremberg, Germany
  - Stockholm, Sweden
  - Beijing, China
  - San Francisco, USA
- <http://www.codingtechnologies.com>

# About Coding Technologies

*What are we famous for? What are we doing?*



“Successor“ of mp3...



Part of 12 international standards  
In 200 million mobile phones






New multi-channel format

# Perceptual Audio Coding – Motivation



- Digital Audio (e.g. Compact Disc)  
PCM: 44.1 kHz, 16 bit, stereo -> **1.4 Mbit/s**
  - Perceptual Audio Coding  
...is a compact audio representation, exploiting:
    - Signal redundancy  
e.g. transform coding (“MDCT”)
    - Perceptual irrelevance  
masking effect (“psychoacoustic model”)
- ➔ “transparent” quality at **128 kbit/s**  
compression ratio ~1:12

# Perceptual Audio Coding – Limitations

- Audio Demonstration

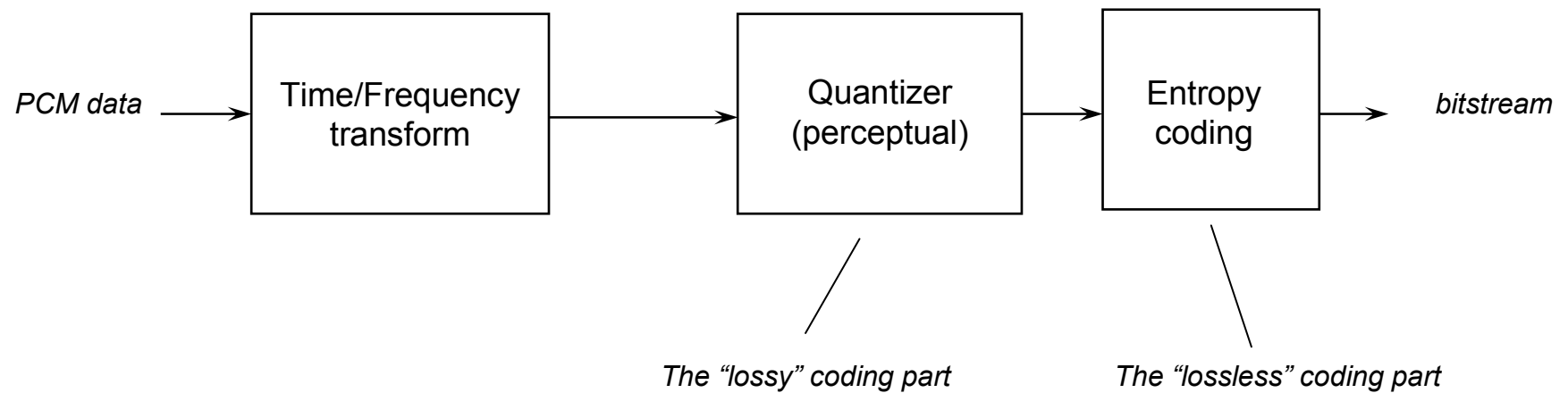
- Original PCM stereo 1.4 Mbit/s 
  - AAC stereo 128 kbit/s 1:12 
  - Difference signal (23 dB SNR) 

- But ...

- AAC stereo 48 kbit/s 1:32 
  - AAC mono 24 kbit/s 1:64 

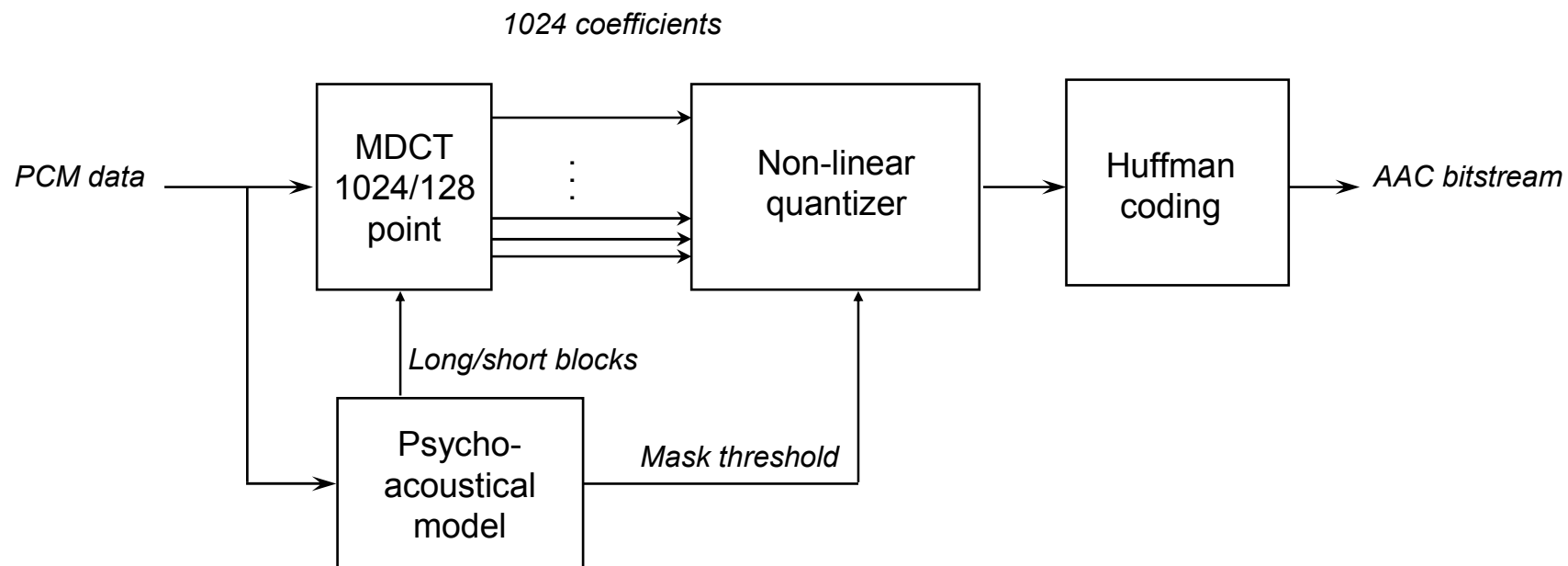
# How traditional Audio Coding works...

Encoder



# How traditional Audio Coding works...

The AAC encoder (slightly simplified)

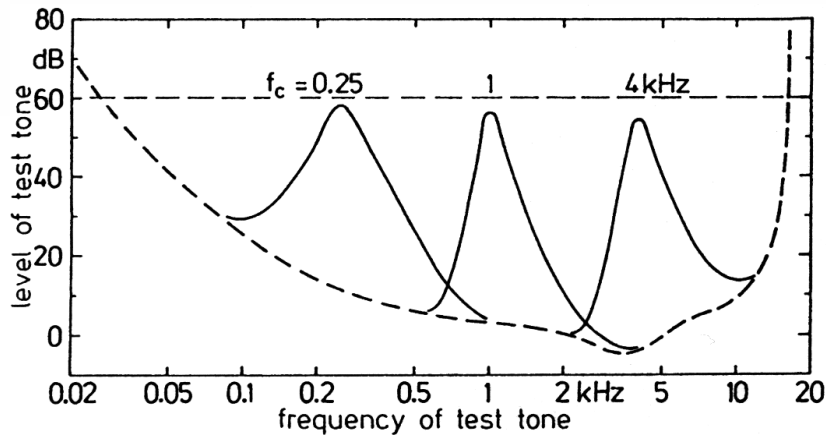


The Modified Discrete Cosine Transform (MDCT) can switch between long (1024 points) and short (128 points) transform sizes.

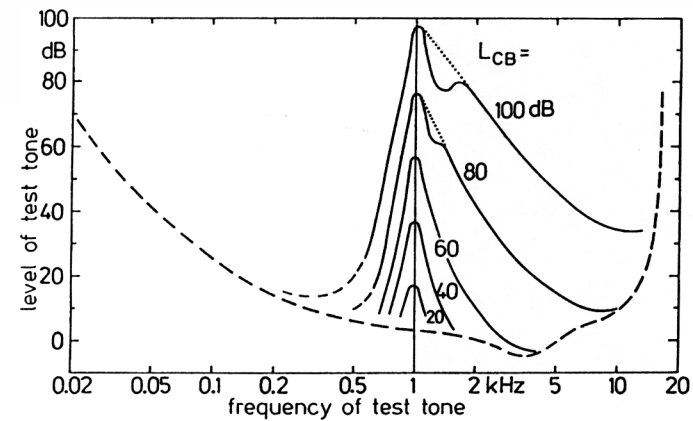


# Why traditional Audio Coding works...

Source: E. Zwicker, H.Fastel: Psychoacoustics, Springer 1998, ISBN 3-540-65063-6



**Masking at different frequencies**



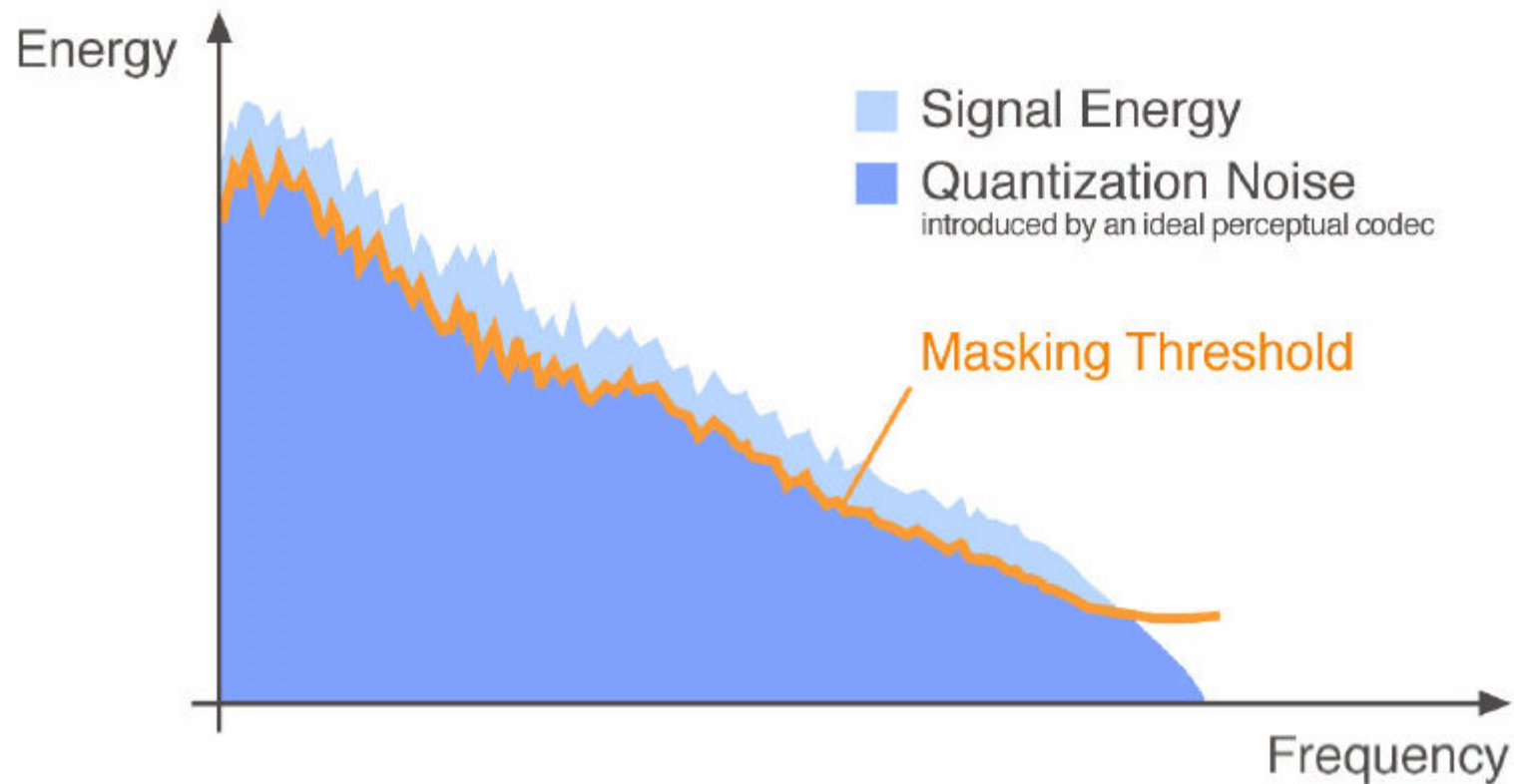
**Masking at different Levels**

Experiment: Masking of tones by critical band wide narrow band noise

- Masking threshold shape dependent on
  - Frequency
  - Lower/upper slope
  - Center frequency of noise
  - Level of narrow band noise

# Why traditional Audio Coding works...

... at the “nominal” bitrate



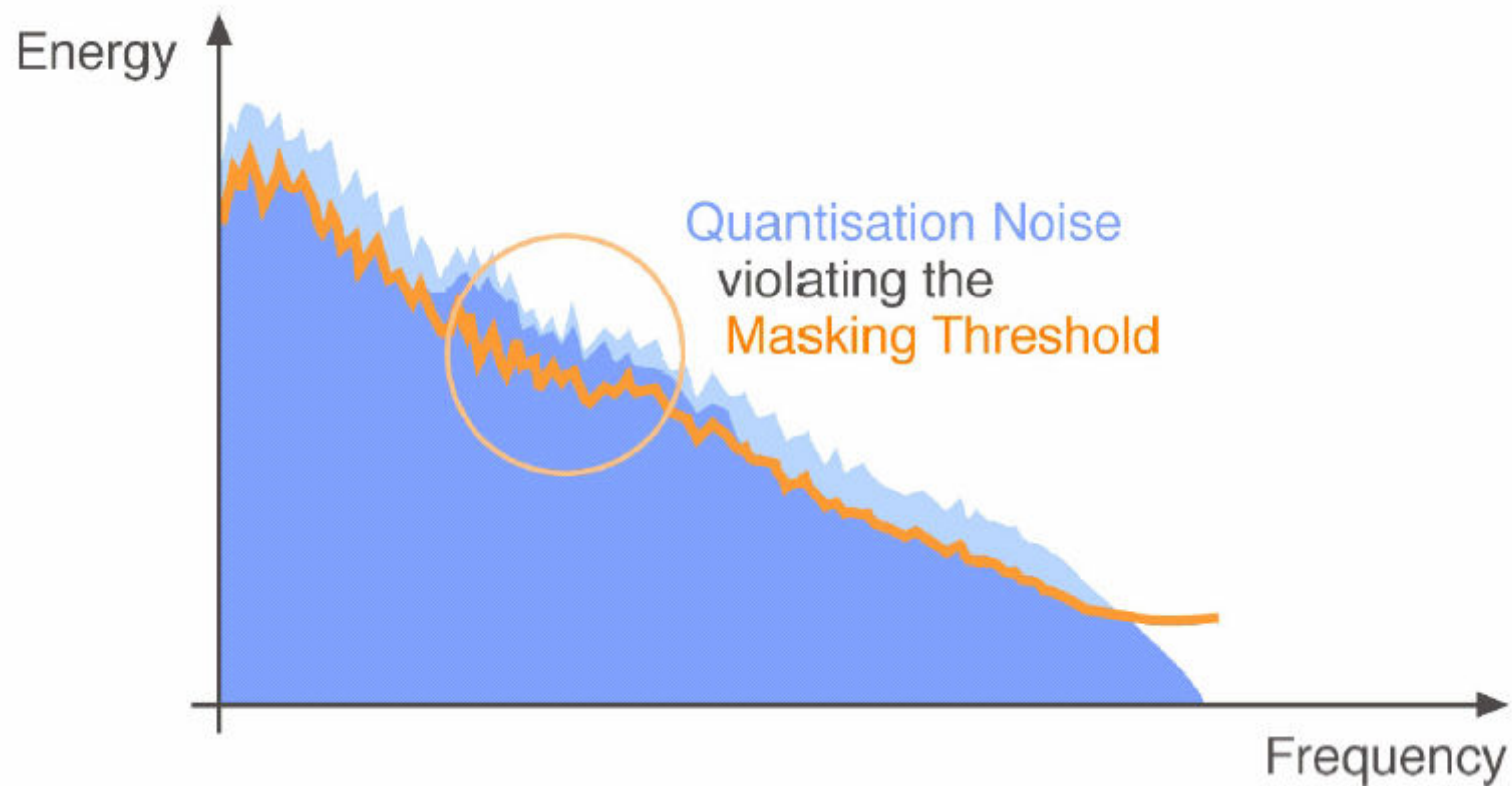
## Semi-parametric audio coding tools

---

- Parametric bandwidth extension methods
  - Spectral Band Replication (SBR) (MPEG 2003)
  
- Parametric channel extension methods
  - Parametric Stereo (PS) (MPEG 2004)
  - MPEG Surround (MPS) (MPEG 2006)

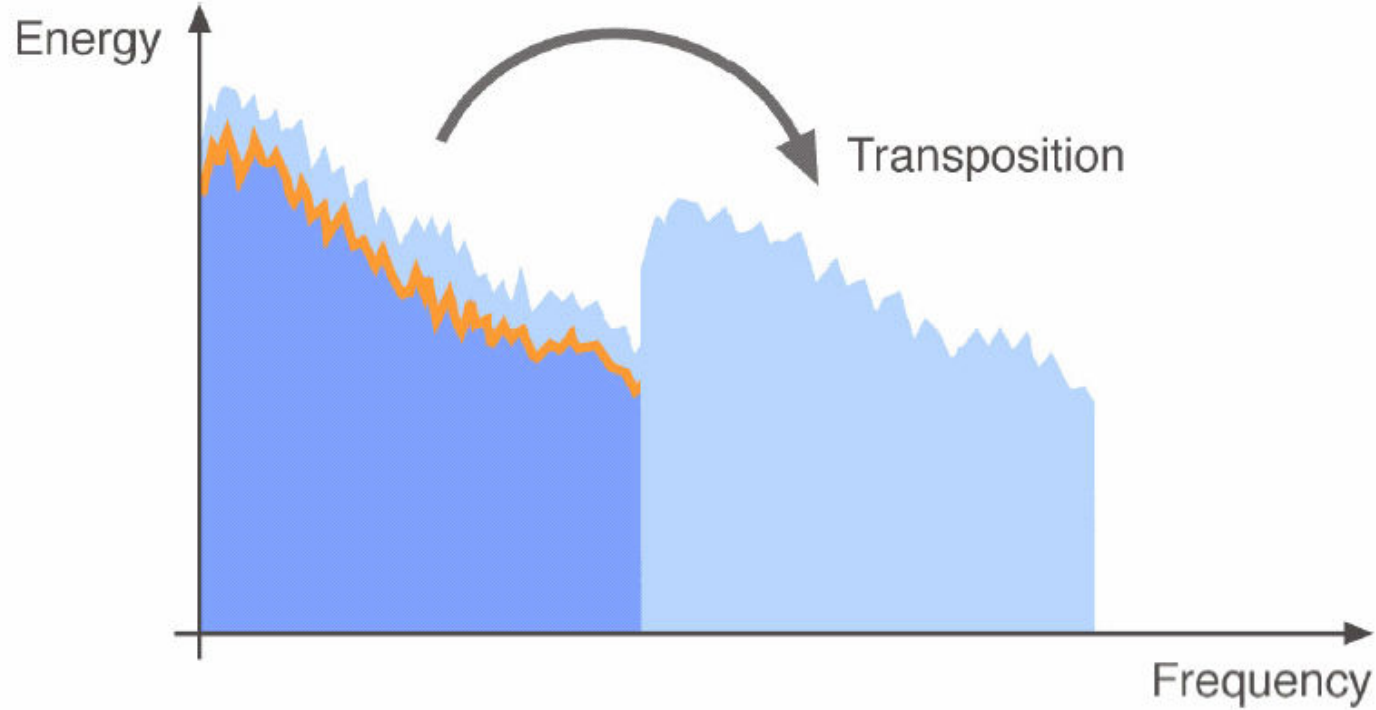
# Why traditional Audio Coding fails ...

... if the bitrate is further reduced



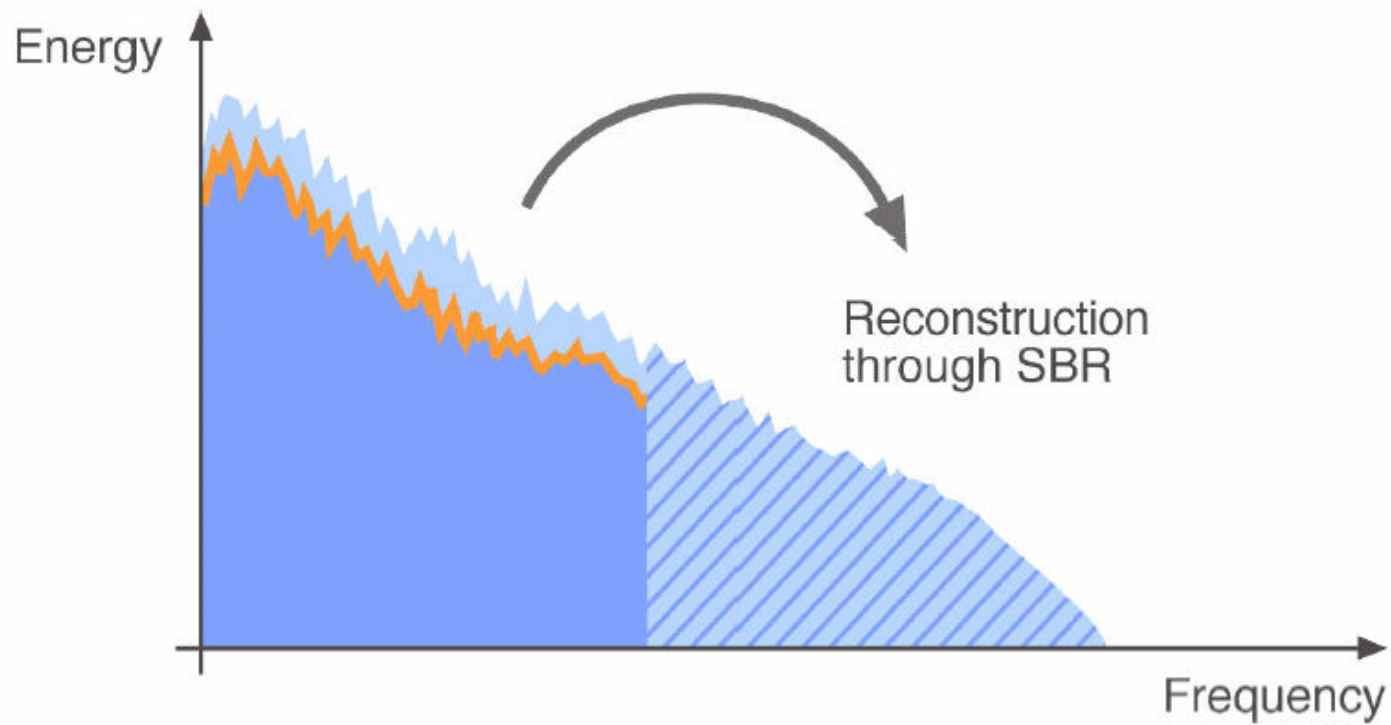
# SBR principle (1)

The high frequencies are reconstructed ...

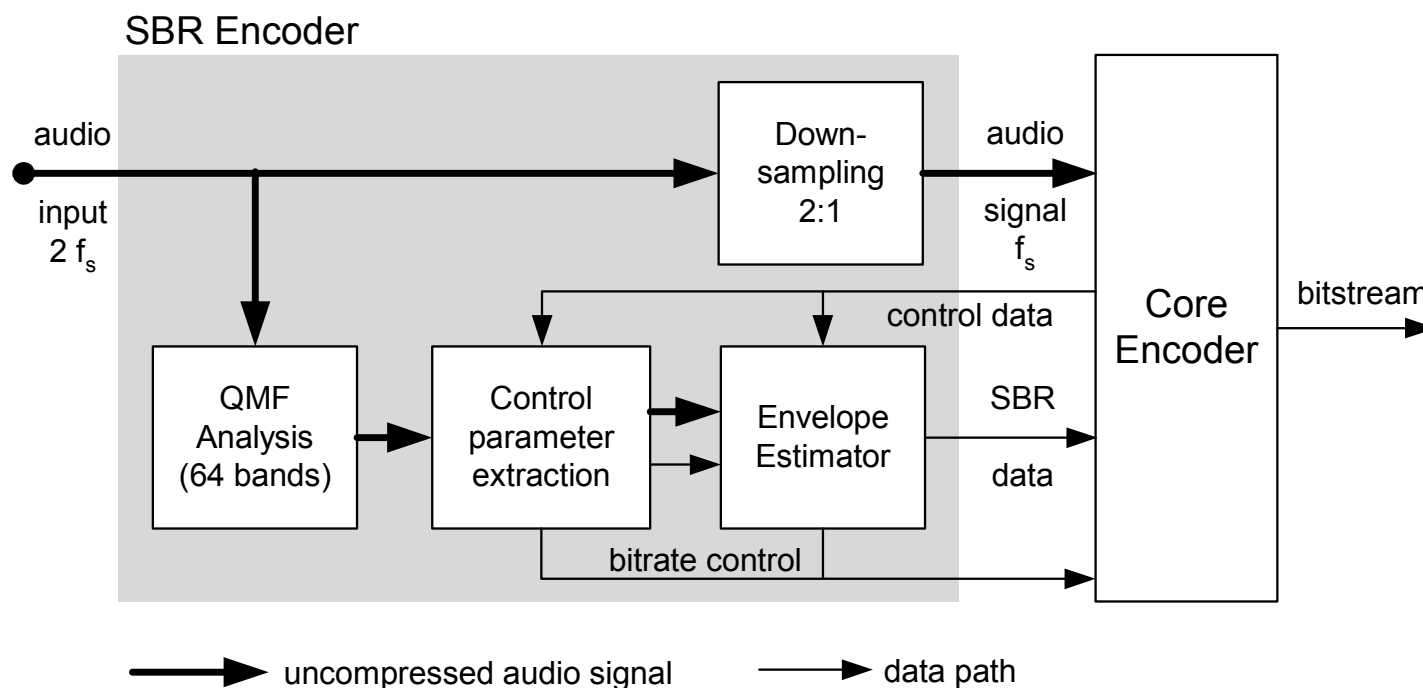


# SBR principle (2)

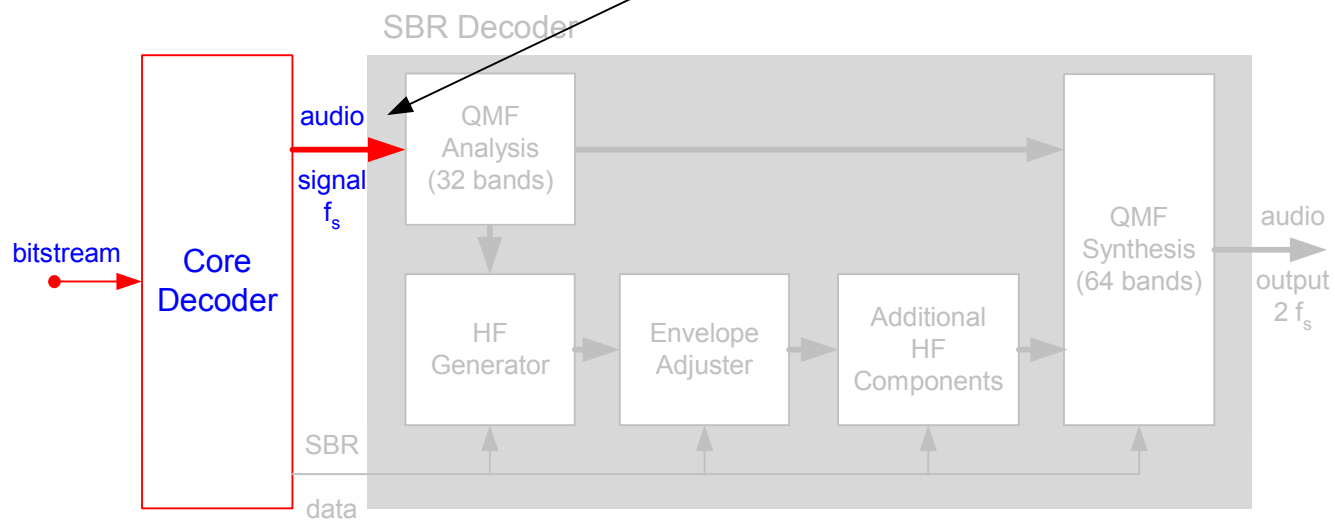
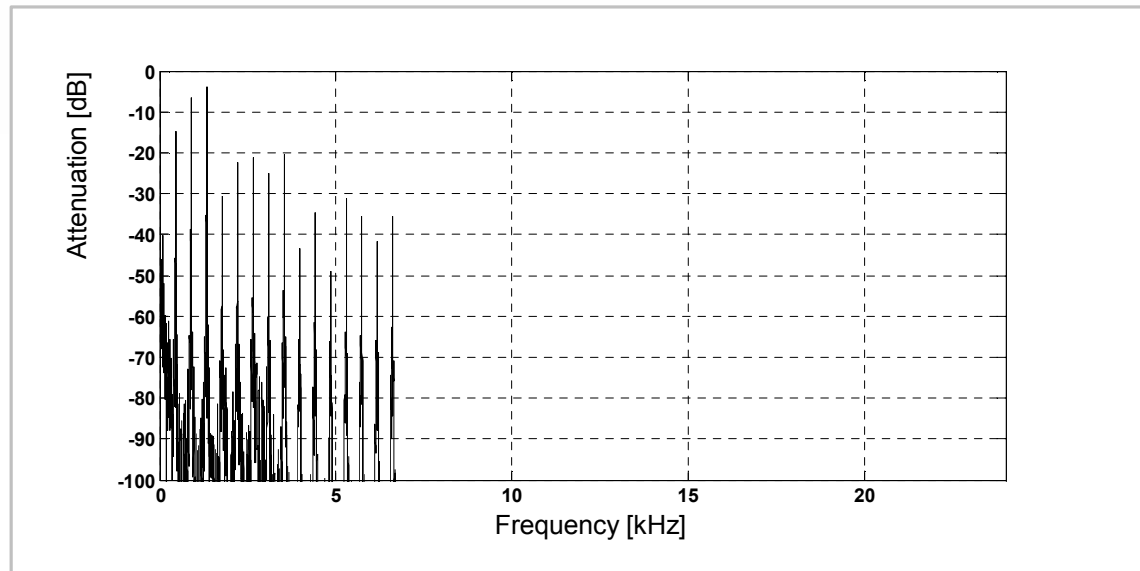
... and adjusted



# SBR Encoder

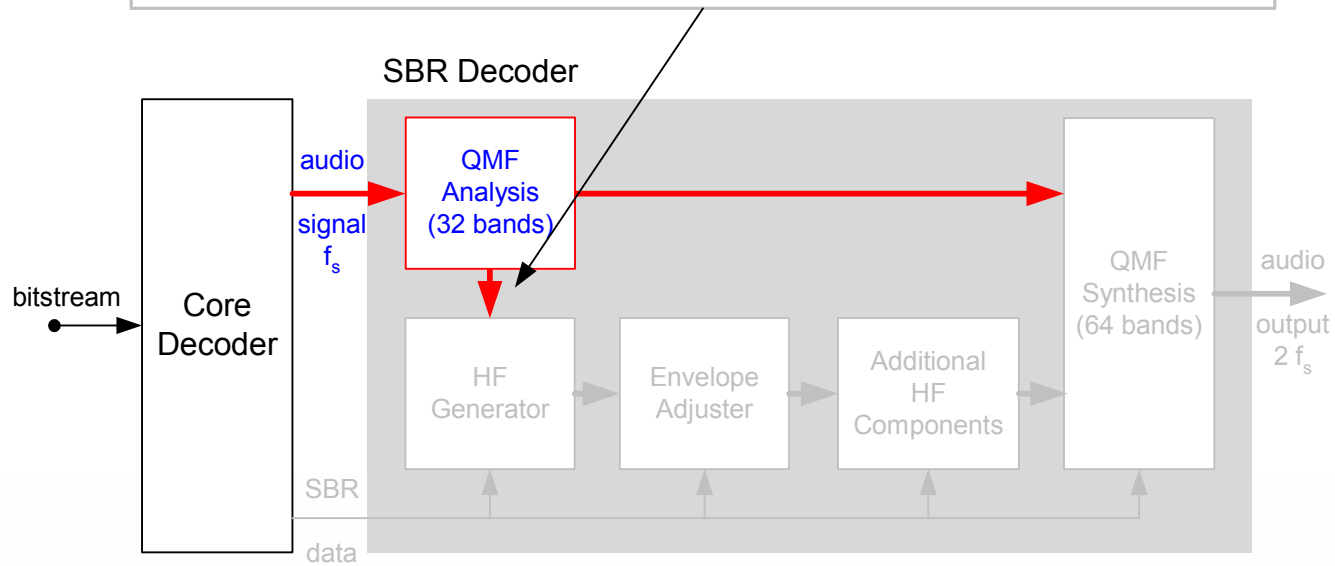
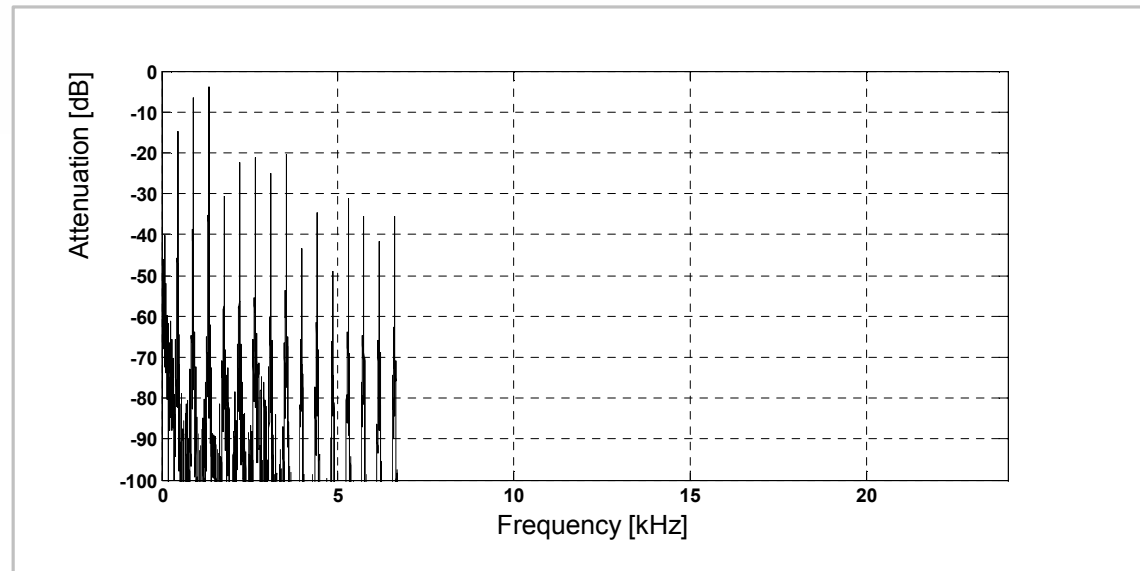


# Audio Decoder with SBR

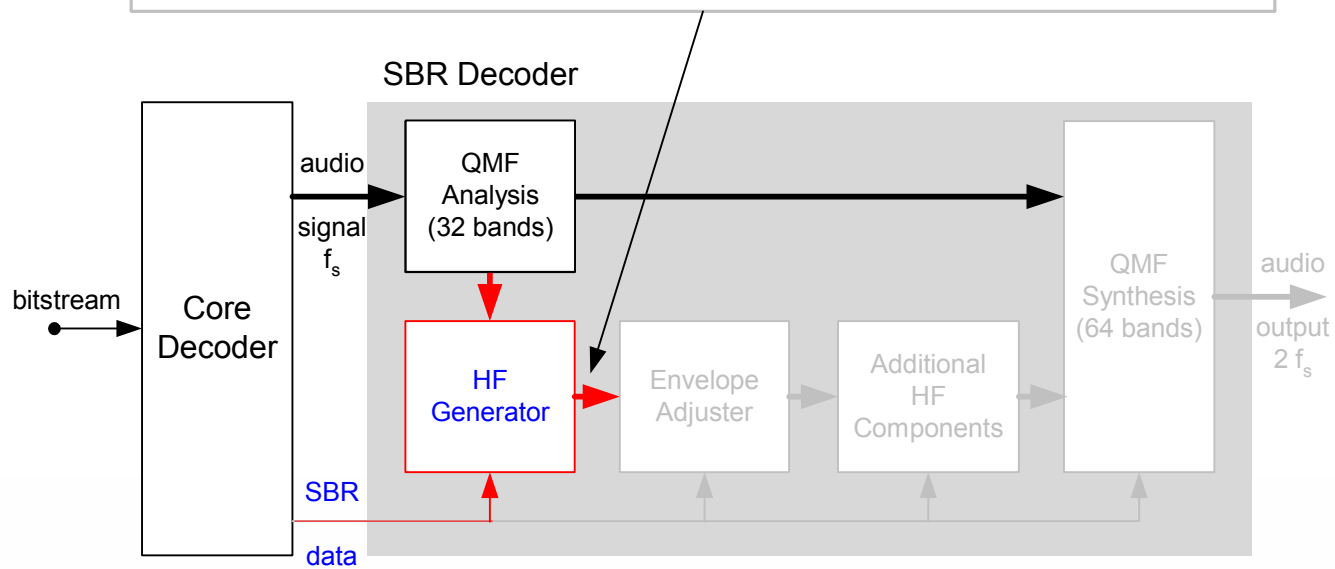
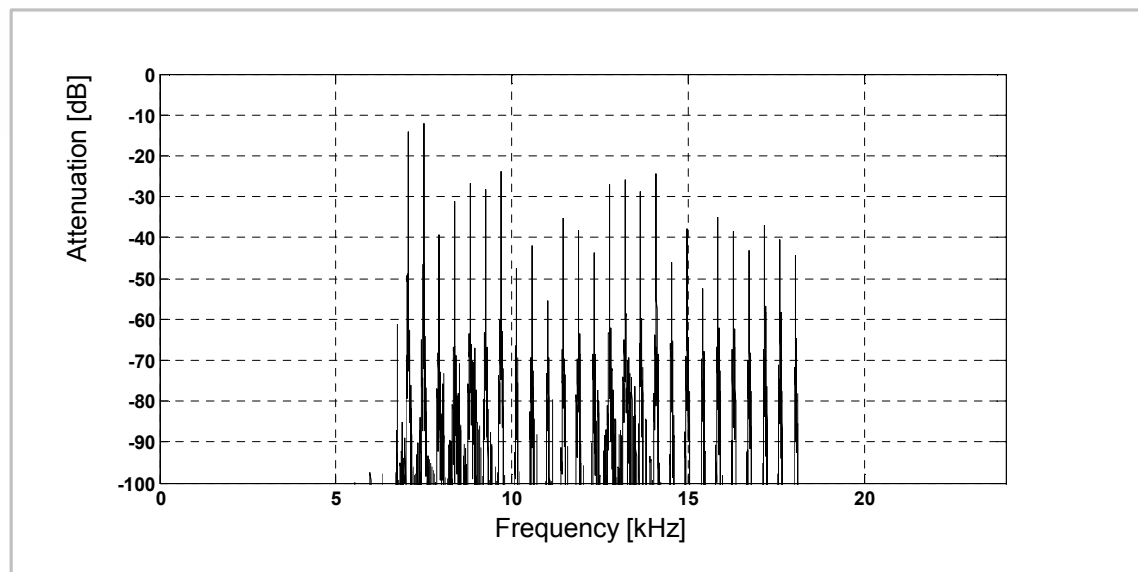




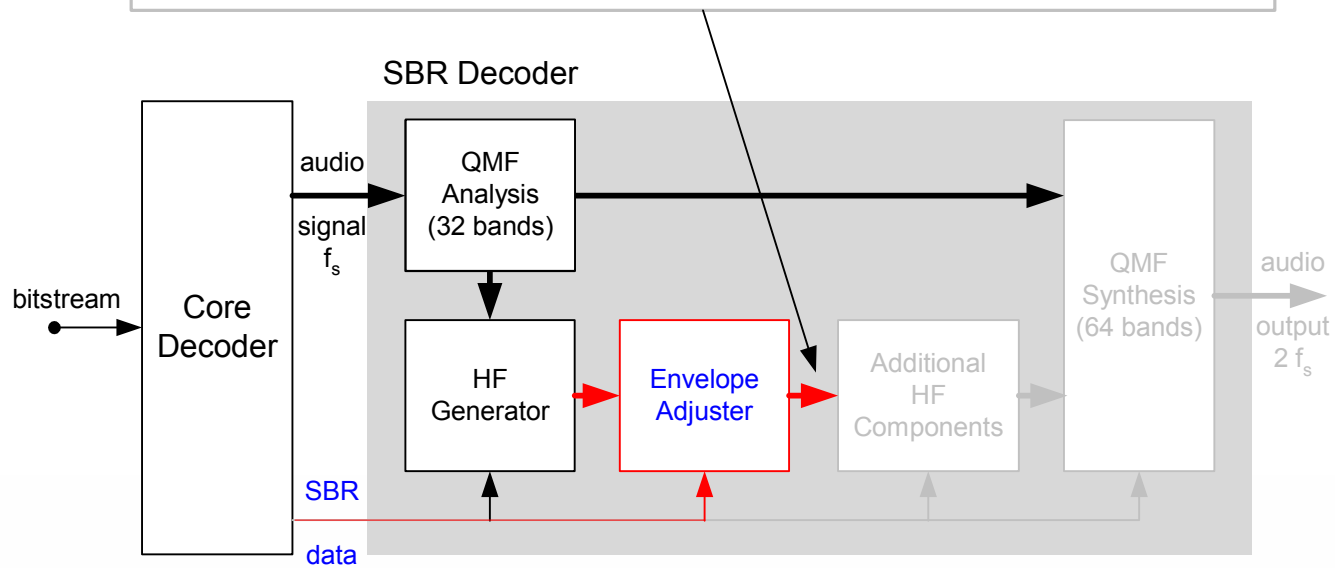
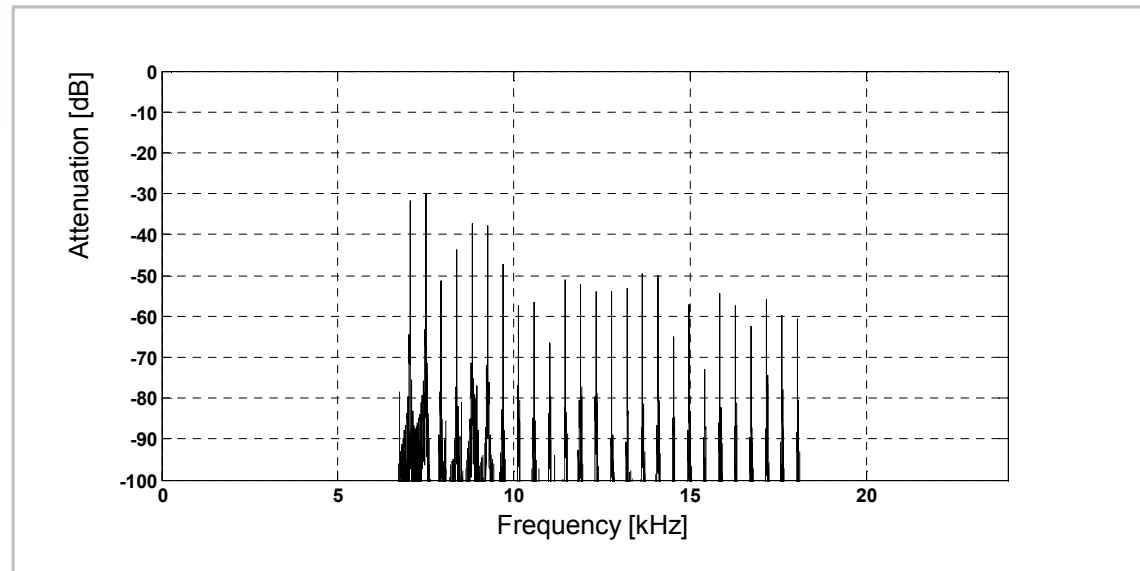
# Audio Decoder with SBR



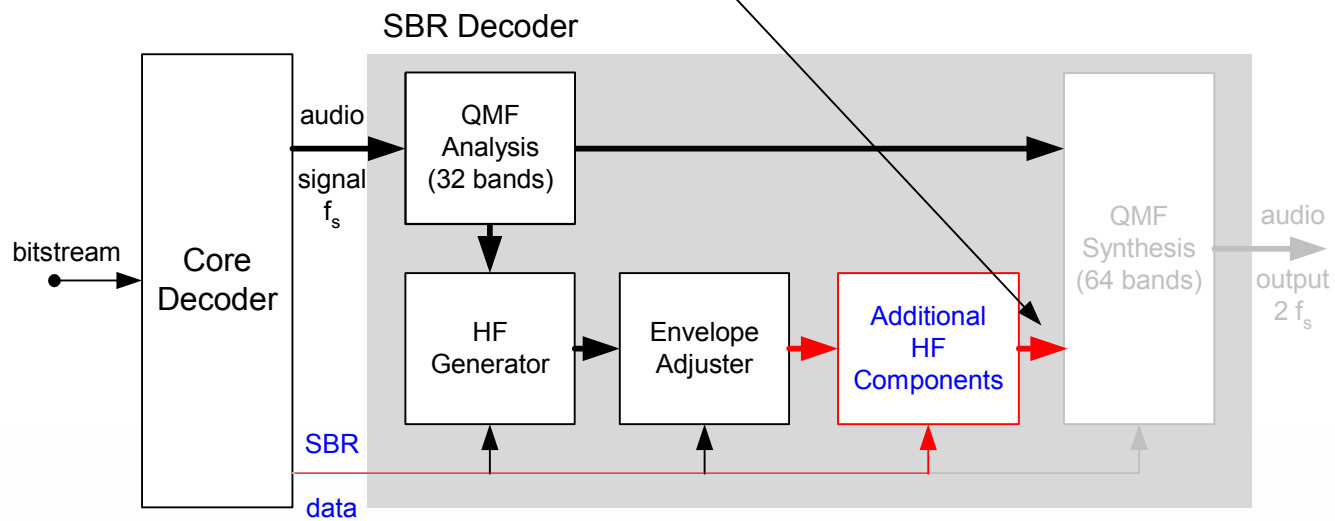
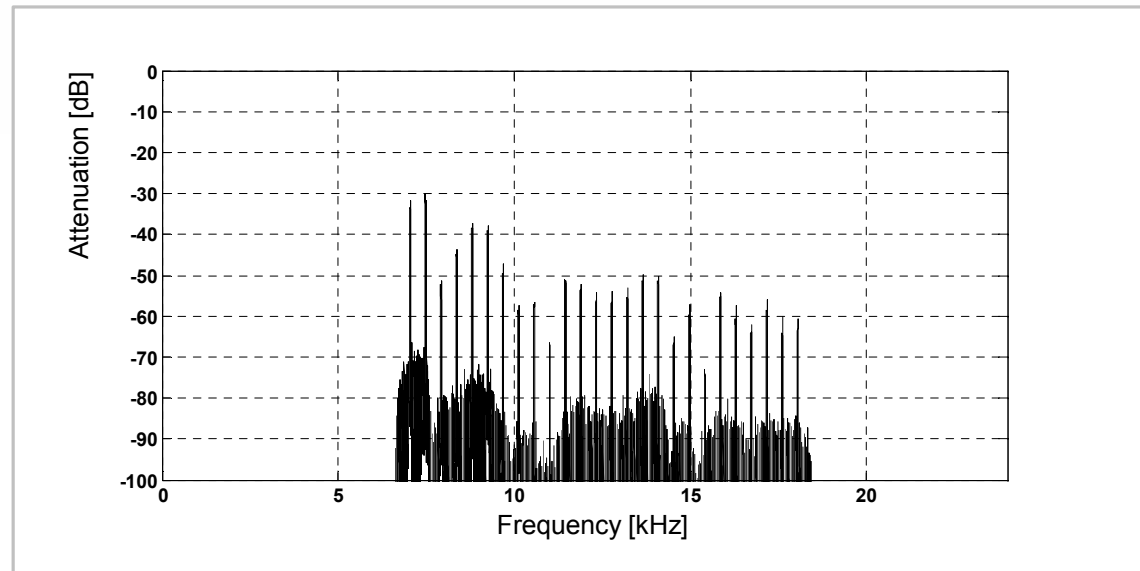
# Audio Decoder with SBR



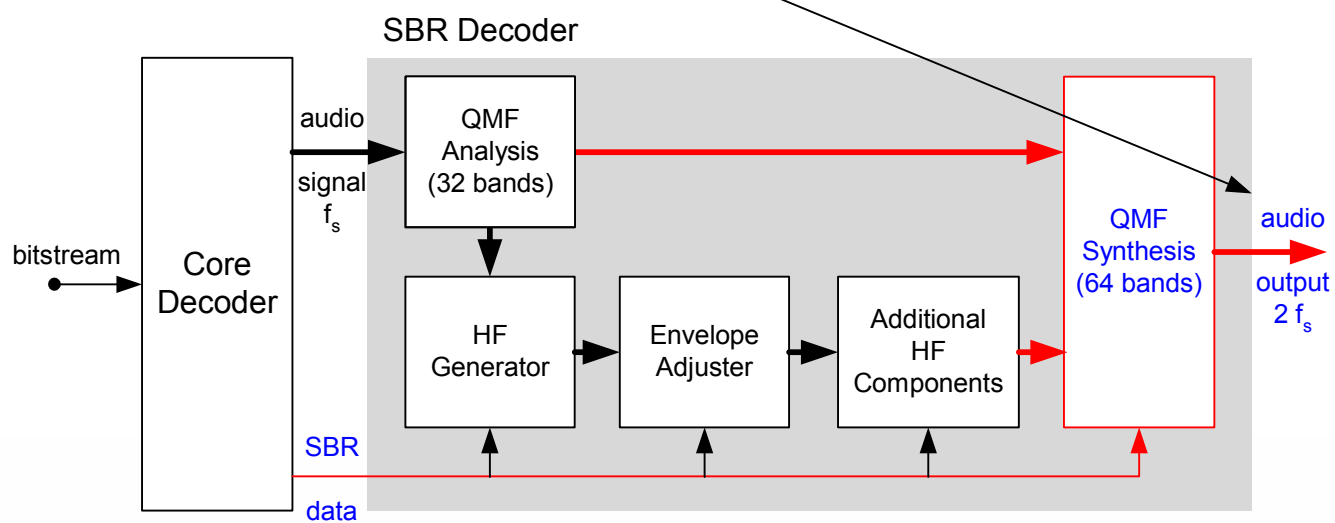
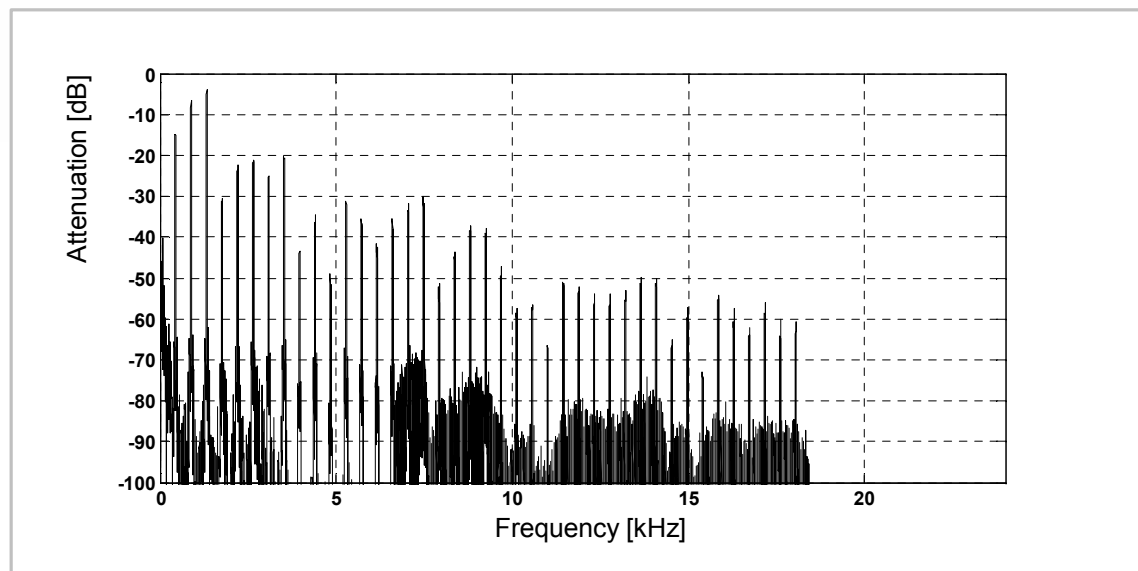
# Audio Decoder with SBR



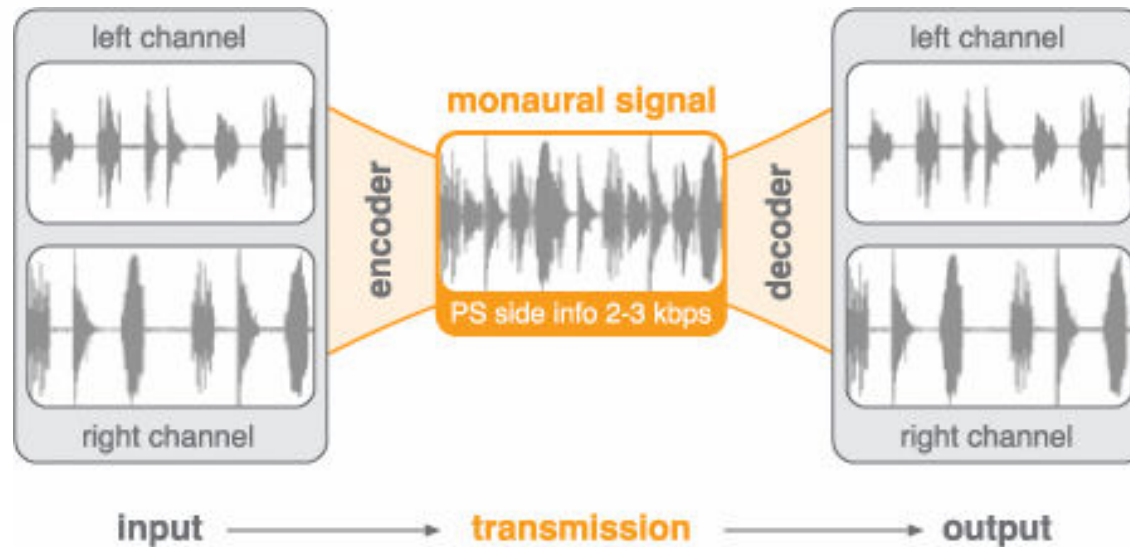
# Audio Decoder with SBR



# Audio Decoder with SBR

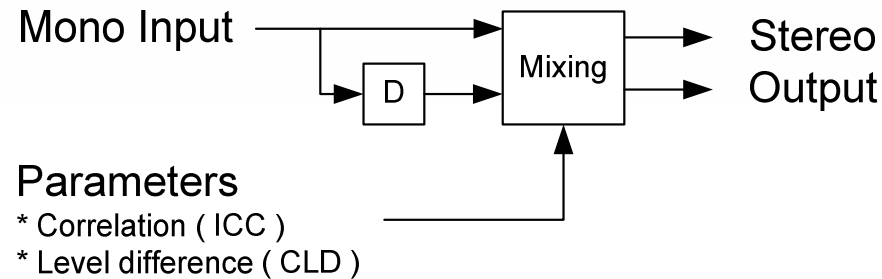


# What is Parametric Stereo (PS)?



- Parametric channel extension method
  - Mono downmix
  - Stereo upmix
- Transmit mono signal + stereo side information
- Reconstruct stereo signal in decoder
- Approx. 2 to 3 kbit/s stereo side information

# Stereo parameter reconstruction







- Adjust mixing to reproduce total energy and

- Channel Level Difference (CLD) :  $q = \frac{E\{|l|^2\}}{E\{|r|^2\}}$

- Correlation (ICC) :  $\rho = \frac{\text{Re}(E\{lr^*\})}{\sqrt{E\{|l|^2\}E\{|r|^2\}}}$

- aacPlus (AAC + SBR + PS)  
-> The most efficient audio codec available today

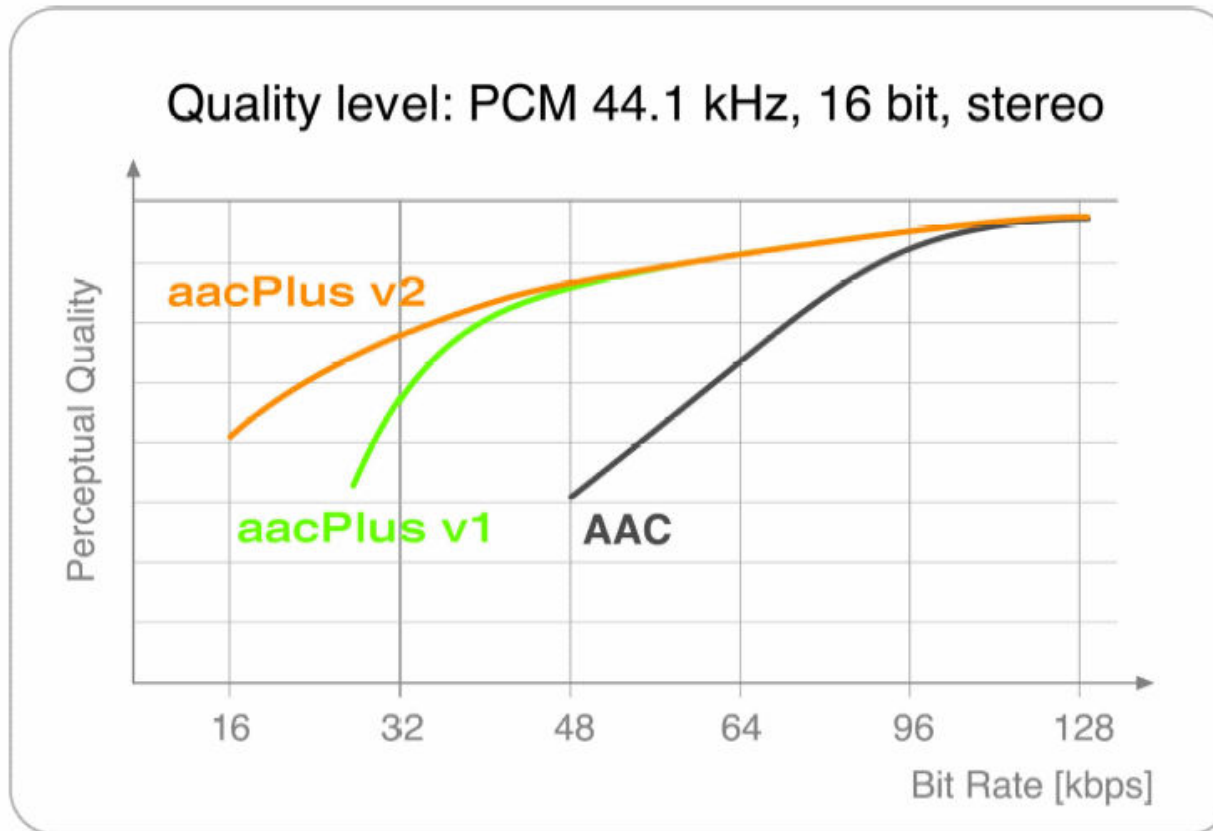
- Audio Demonstration

□ Original PCM stereo	1.4 Mbit/s		
□ AAC stereo	48 kbit/s	1:32	
□ aacPlus stereo	48 kbit/s	1:32	
□ aacPlus (PS)	24 kbit/s	1:64	



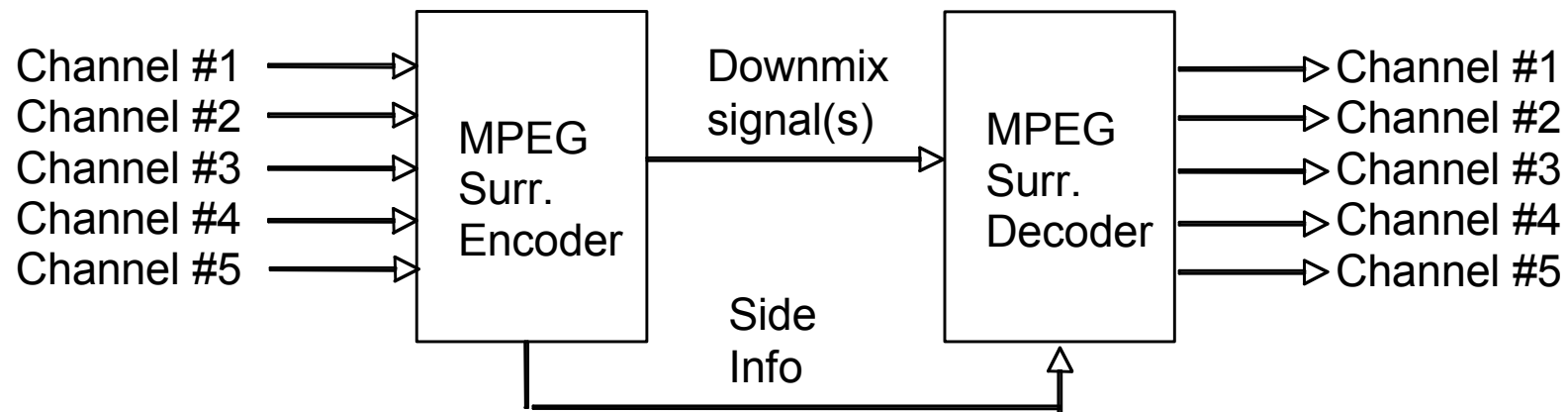
# aacPlus Codec Family

- aacPlus v1 = AAC + SBR
- aacPlus v2 = AAC + SBR + PS

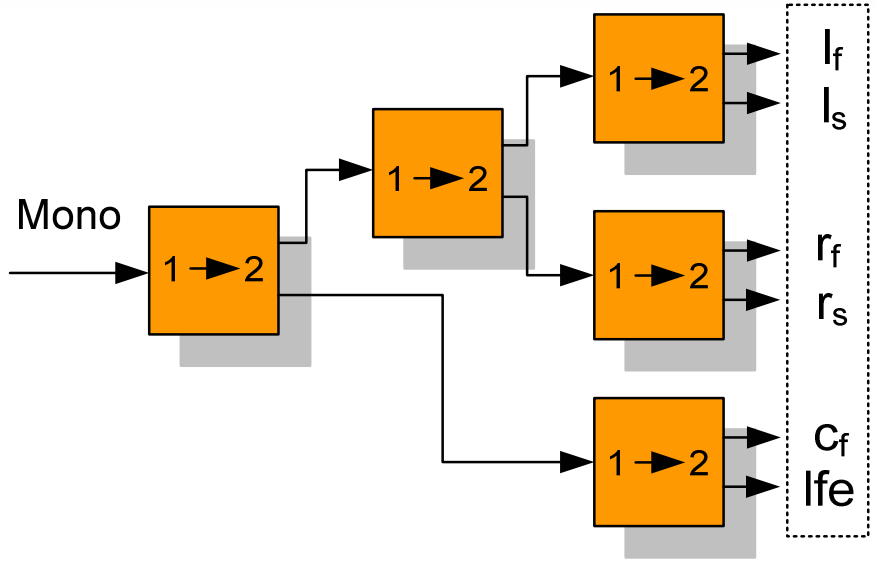


# What is MPEG Surround?

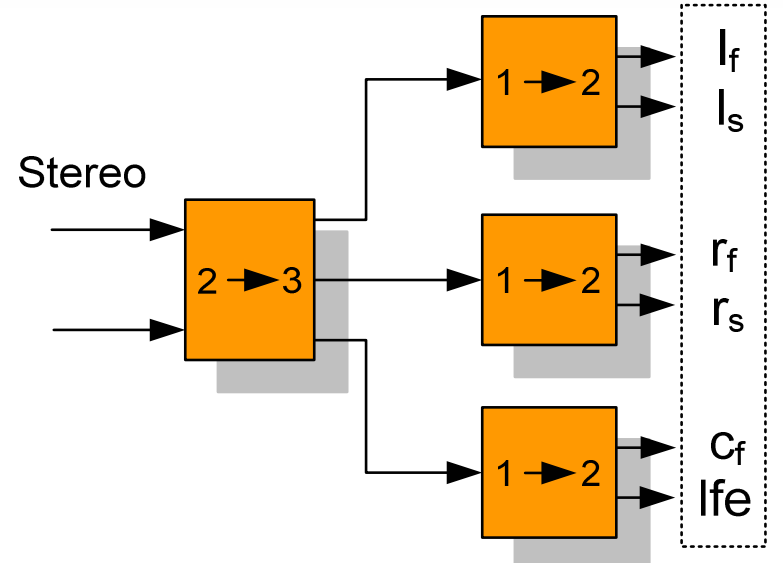
- Parametric channel extension
  - Mono or stereo downmix
  - Upmix to 5.1 (or more...)



# MPEG Surround - Tree structure

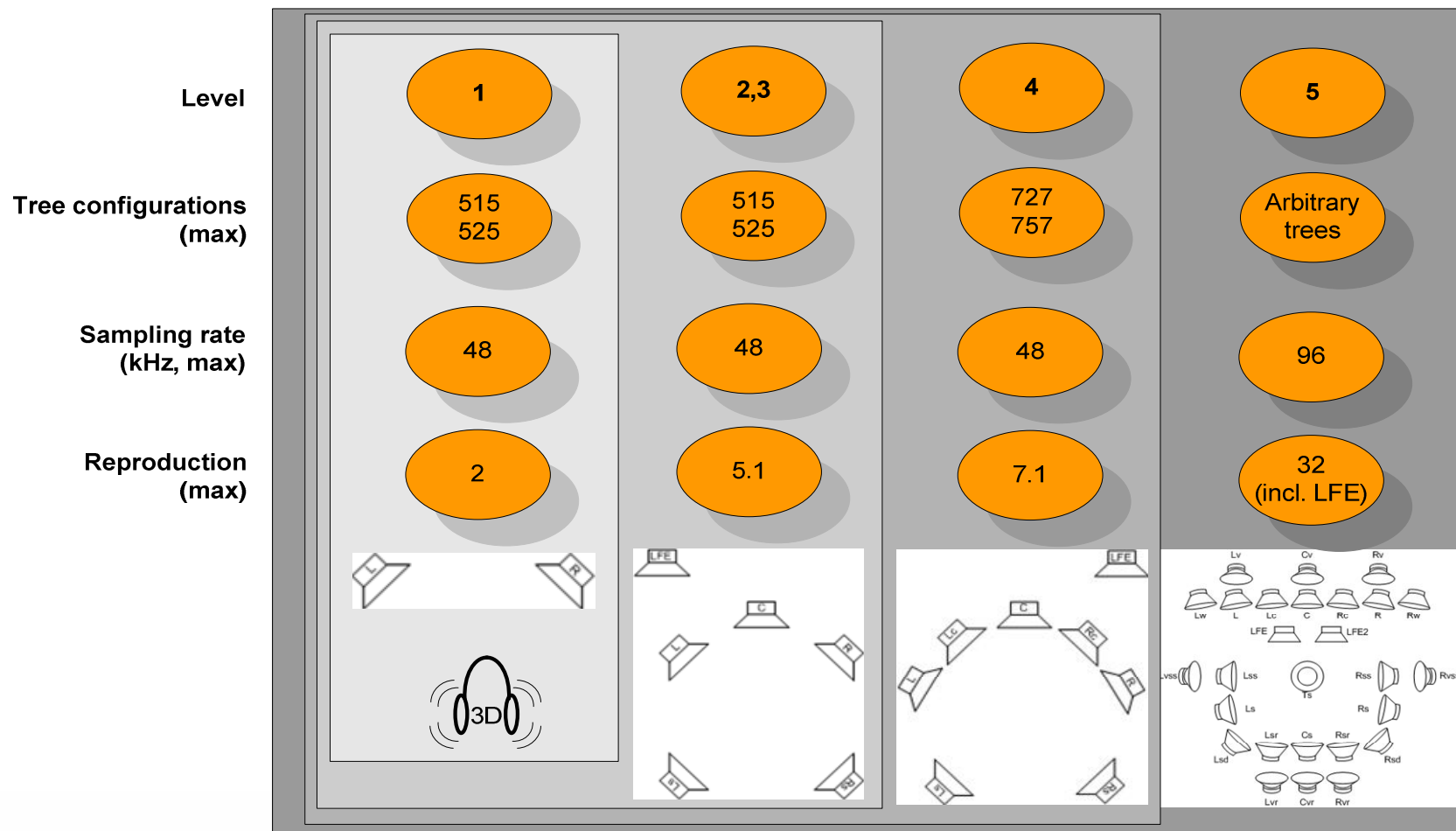


Mono to Surround  
(5.1 – 1 - 5.1)



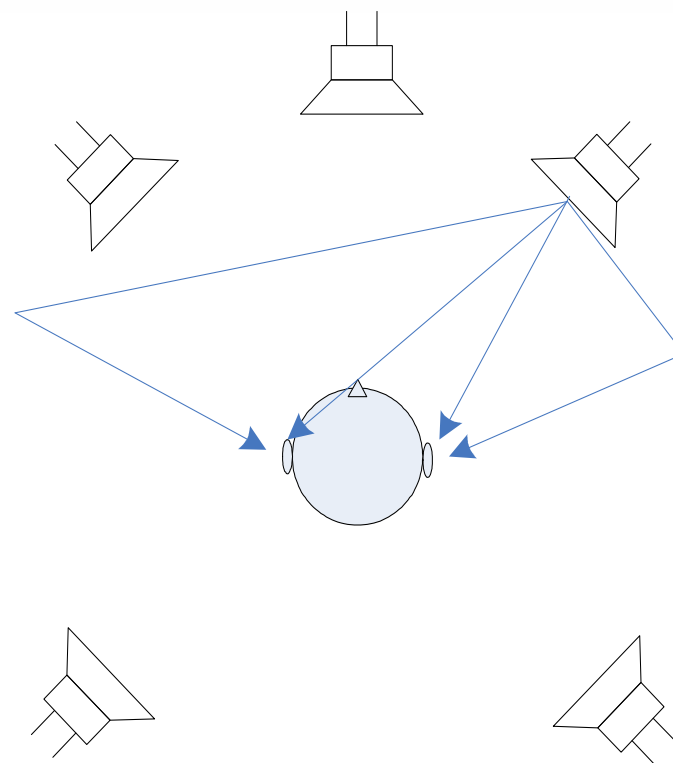
Stereo to Surround  
(5.1 – 2 - 5.1)

# MPEG Surround - Configurations

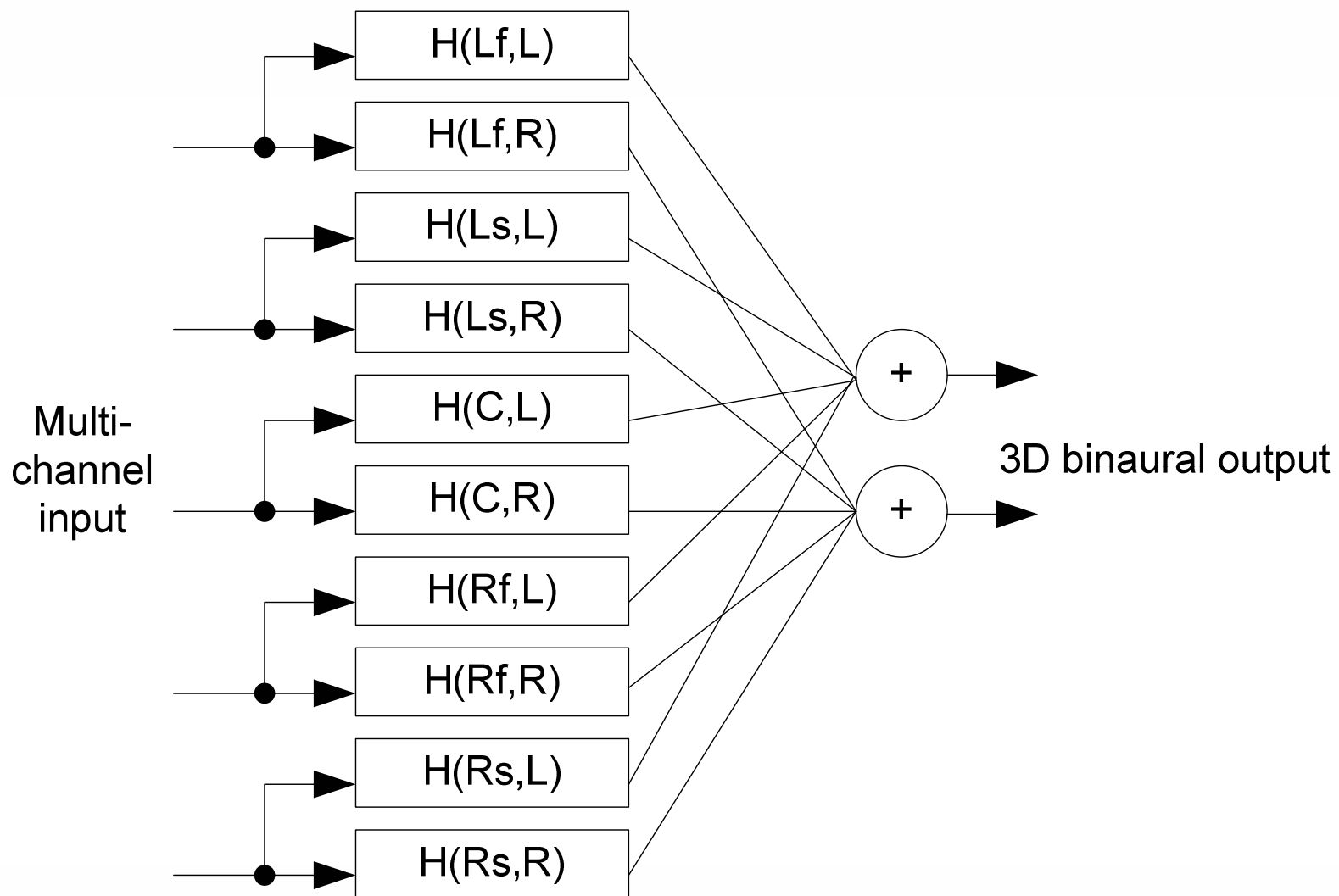


# Binaural spatialization

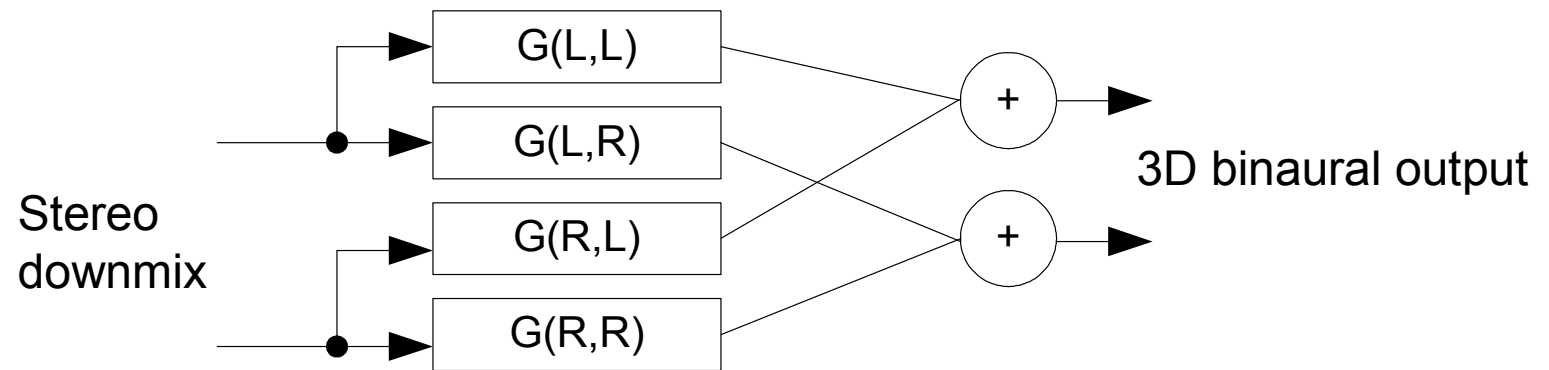
- Goal
  - Surround over headphones
- Approach
  - Impulse response from each speaker to each ear
  - Head-Related Transfer Function (HRTF)
  - Binaural Room Impulse Response (BRIR)



# Binaural synthesis - Traditionally



# Binaural synthesis – MPEG Surround



## Parametric index

---

- Traditional transform codecs are essentially “waveform” coders (mp3, AAC,...).
- There are also pure synthetic/parametric coders (also e.g. MIDI, misc speech synthesis)
- **aacPlus** combines waveform coding (AAC) with parametric coding tools (SBR, PS).



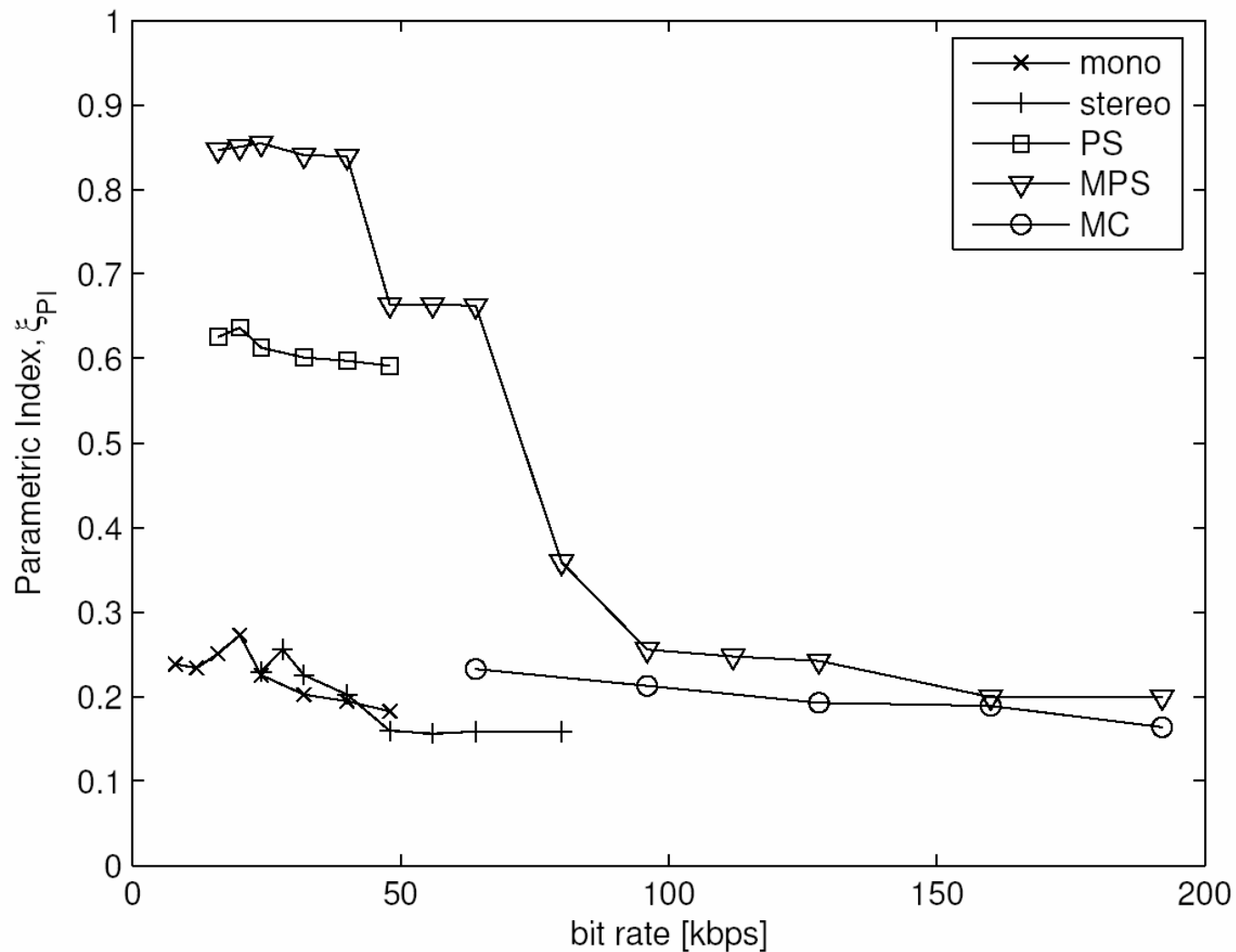
# Parametric index

- The parametric index indicates how much of a codec is parametric:

$$\xi_{PI} = \frac{B_{total} - B_{wave}}{B_{total}}$$

Channel mode	bit rate [kbps]	$m$	$n$	SBR range [Barks]	$\xi_{PI}$
aacPlus mono	24	1	1	17.00–21.94 (5512–14814 Hz)	0.23
aacPlus stereo	48	2	2	18.82–22.39 (7924–16192 Hz)	0.16
aacPlus PS	24	2	1	17.00–21.94 (5512–14814 Hz)	0.61
aacPlus MPS	32	5	1	17.86–22.39 (6546–16192 Hz)	0.84
aacPlus MPS	64	5	2	19.23–22.80 (8613–17571 Hz)	0.66

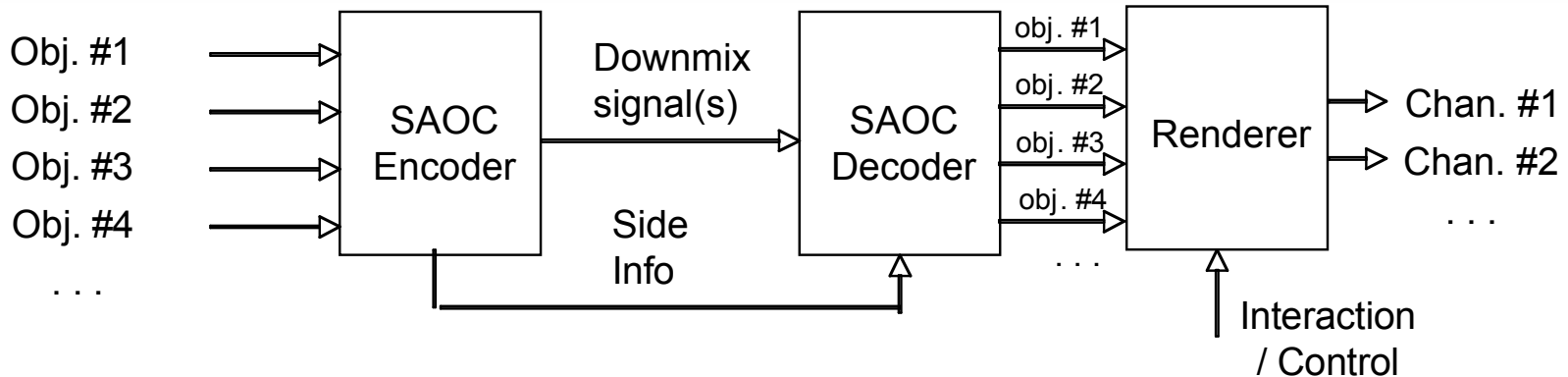
# Parametric index



## Future trends in audio coding

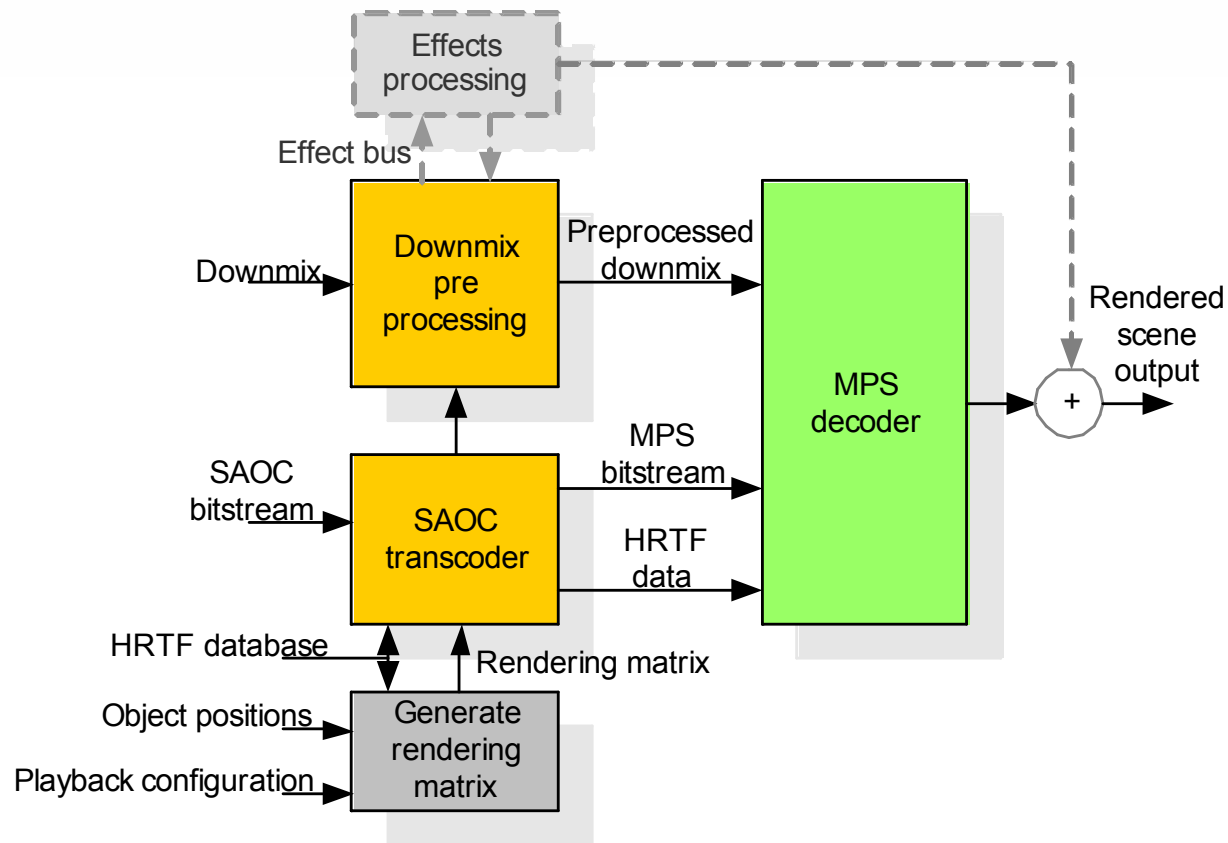
- More and more parametric tools in codecs -
- Audio/music is going mobile -
- Move from stereo towards:
  - multi-channel playback -
  - Binaural/HRTF 3D audio - ...
- User interaction:
  - Modify levels of individual audio objects - ...
  - Personalize your sound - ...

# Spatial Audio Object Coding (SAOC)



- The same base concept as MPEG Surround:
- Replace audio “channels” with audio “objects”
- Add rendering/mixing stage at the end

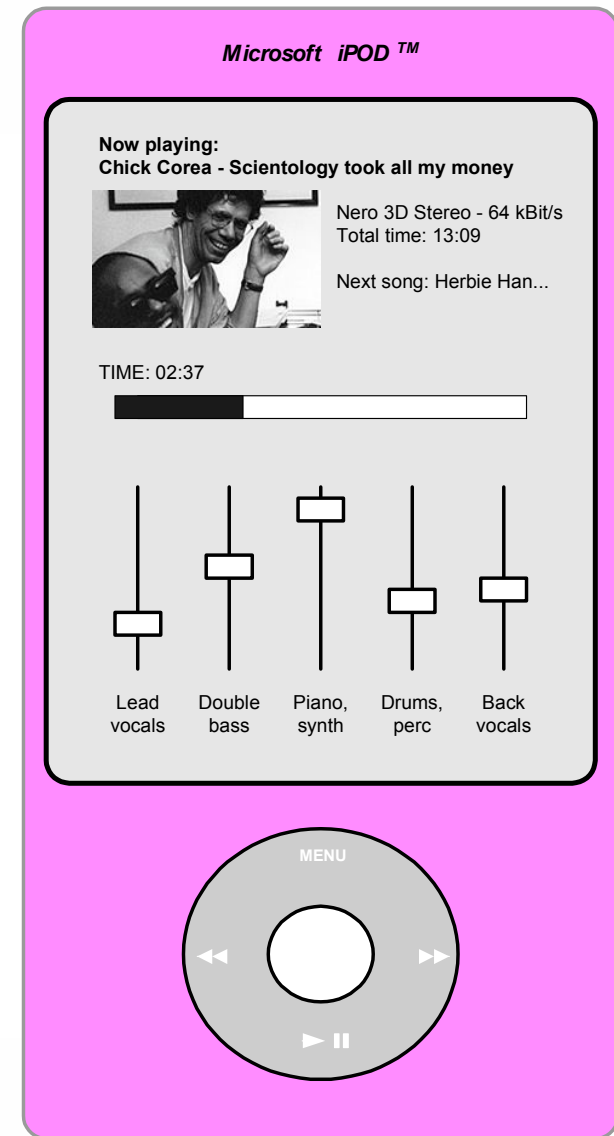
# Spatial Audio Object Coding (SAOC)



SAOC decoder connected to MPEG Surround decoder

# User interaction – do your own remix

- Modify levels of individual audio objects
- Personalize your sound



- Parametric audio coding tools
  - Spectral Band Replication (SBR)
  - Parametric Stereo (PS)
  - MPEG Surround
- aacPlus = AAC + SBR + PS
  - High quality stereo at 24 kbit/s (1:64)
  - Standardized in MPEG-4  
(spec. in 3GPP, DVB, DAB+, DMB, DRM, ...)
- Future trends
  - Mobile, multi-channel / 3D, user-interaction