AES-Seminarium: Så fungerar MP3

Heiko Purnhagen
Dolby Sweden AB
www.dolby.com

AES-Seminarium: Så fungerar MP3

Heiko Purnhagen

2015-03-17

AES-Seminarium: Så fungerar MP3

Födseln av filformatet mp3 blev starten till en total omstrukturering av musikvärlden, där musik nu kunde skickas över internet. Bitströmmen för komprimerade format som mp3 och aac är blott en tiondel av bitströmmen för CD-formatet. I detta seminarium tittar vi på hur komprimeringen går till och vilken information i musikströmmen som har gått förlorad i processen.

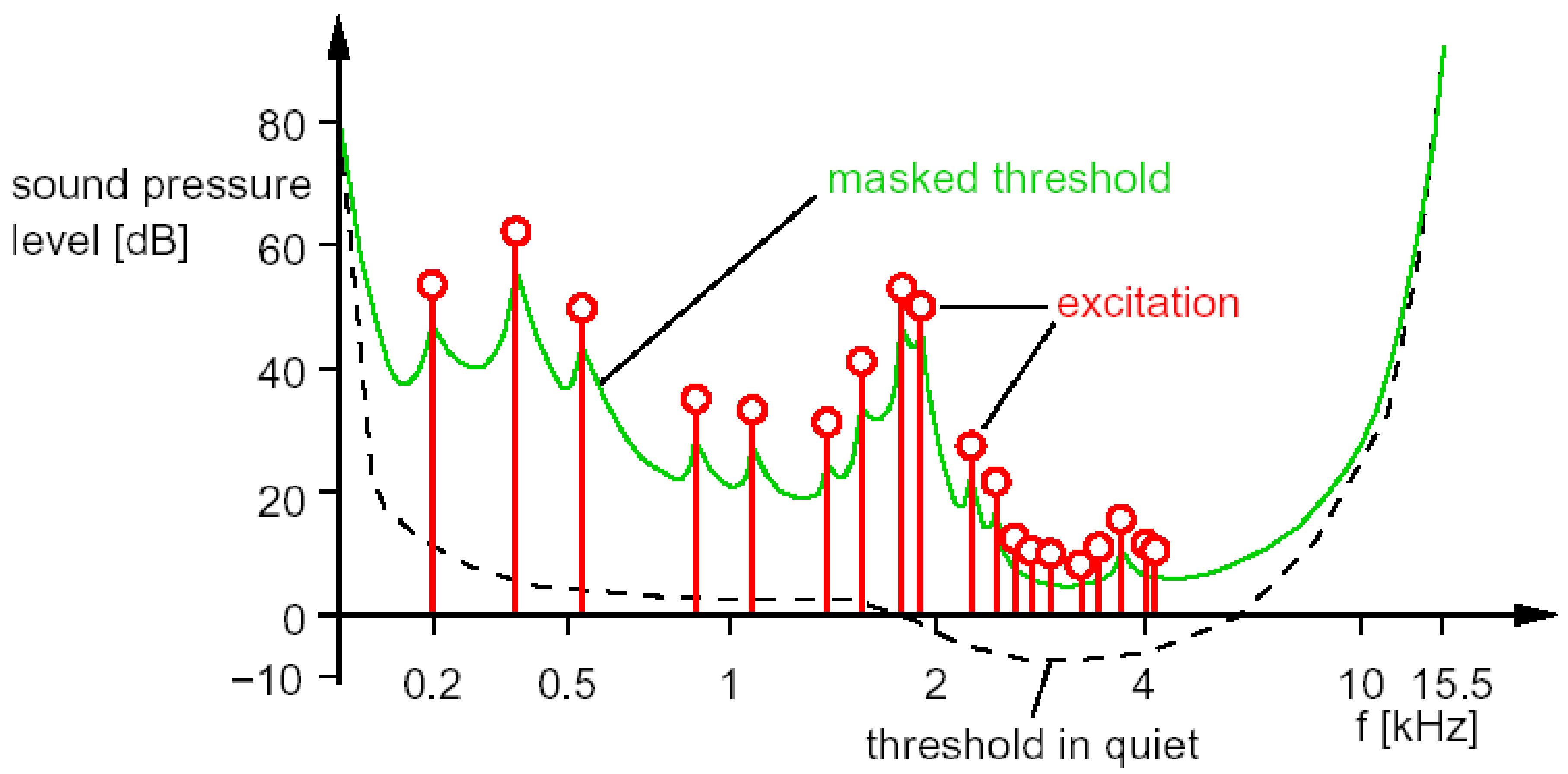
- MP3
  - MPEG-1 Audio, Layer III   (ISO/IEC 11172-3, published 1992)
  - ISO/IEC 11172 – Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s
- AAC
  - MPEG-2 Advanced Audio Coding   (ISO/IEC 13818-7, published 1997)

=> Perceptual Audio Coding

Perceptual Audio Coding – Motivation

- Digital audio requires “high” bit rates…
  CD: 16 bit PCM, 44.1 kHz sampling, stereo => 1.4 Mbit/s
- …too high for many applications
  - portable players
  - broadcast/streaming
- Perceptual audio coding needed
  => same (perceived) quality at lower bit rates
  - Exploit signal redundancy
  - Exploit perceptual irrelevance
- Popular codecs introduced 1992 (compression ratio ~1:10)
  - MPEG-1 Layer III (MP3)
  - Dolby Digital (AC-3)

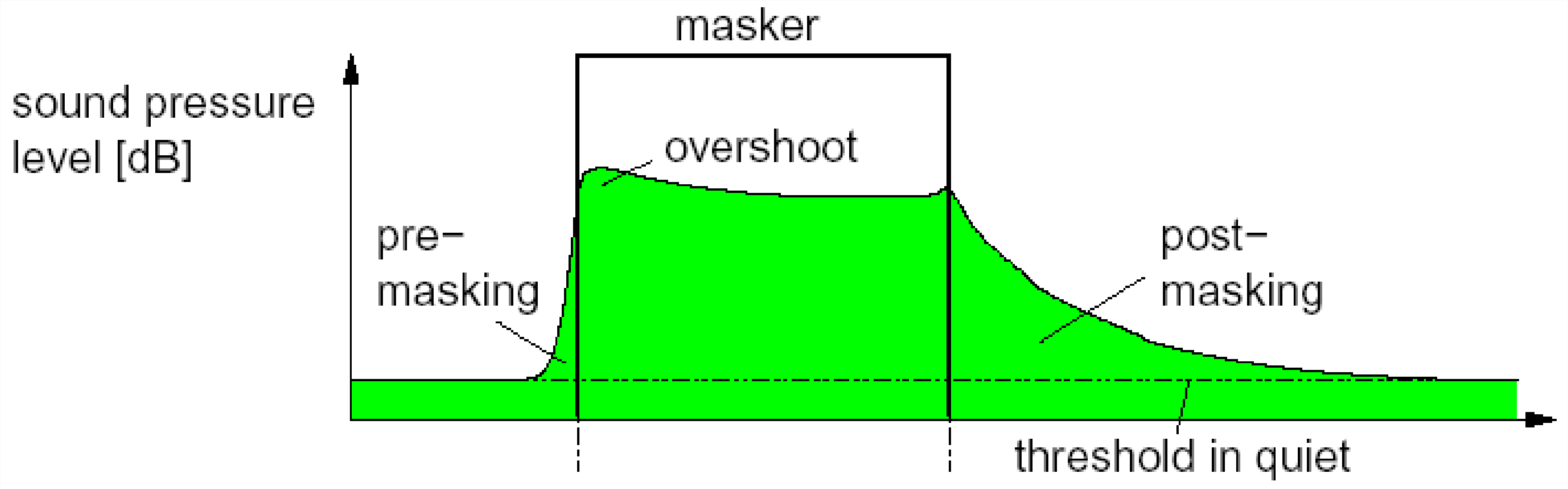
Spectral Masking (simultaneous)



Masking: test tone (maskee) below threshold is inaudible

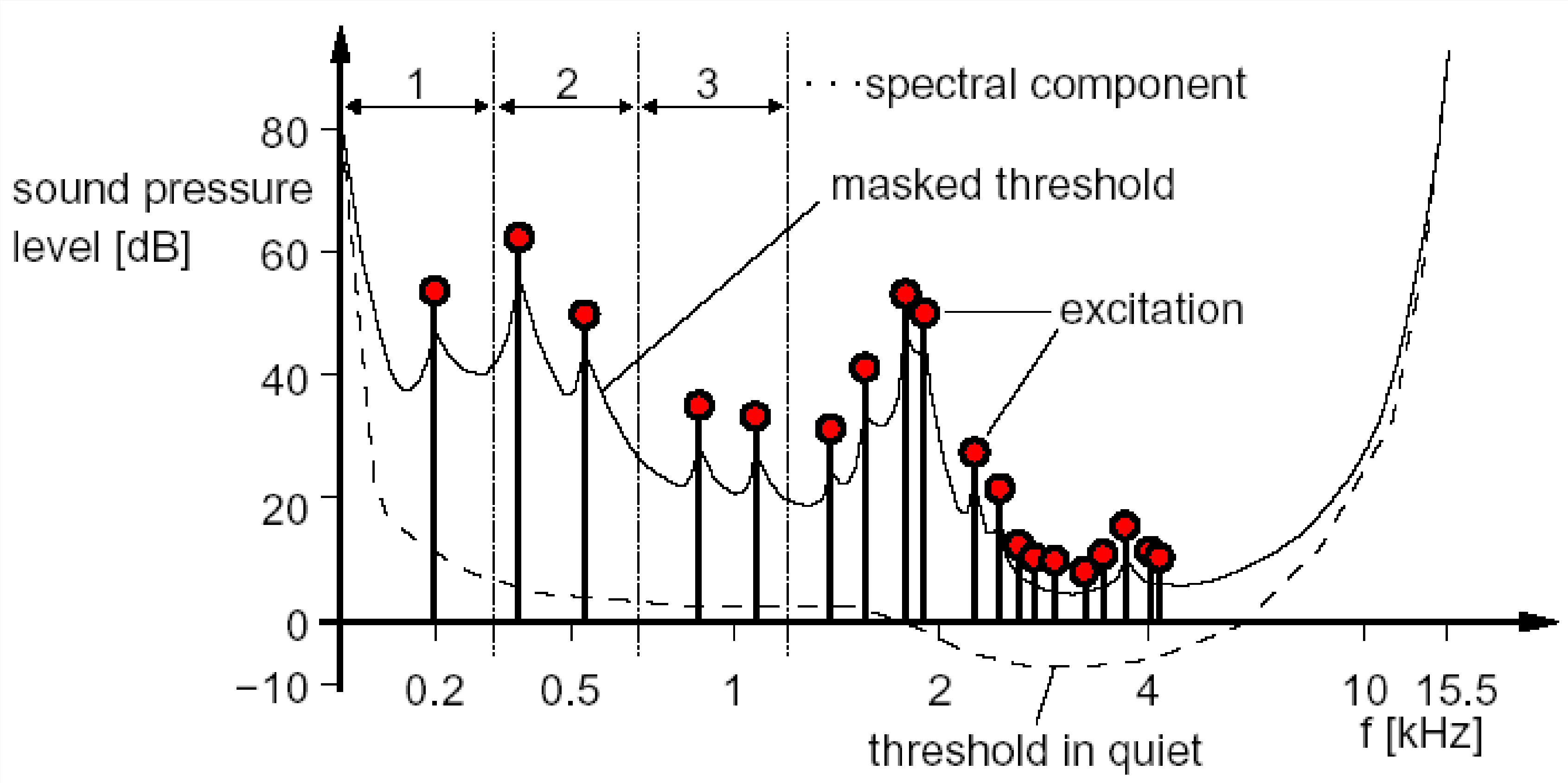
Masked threshold caused by excitation due to masker(s)

Temporal Masking



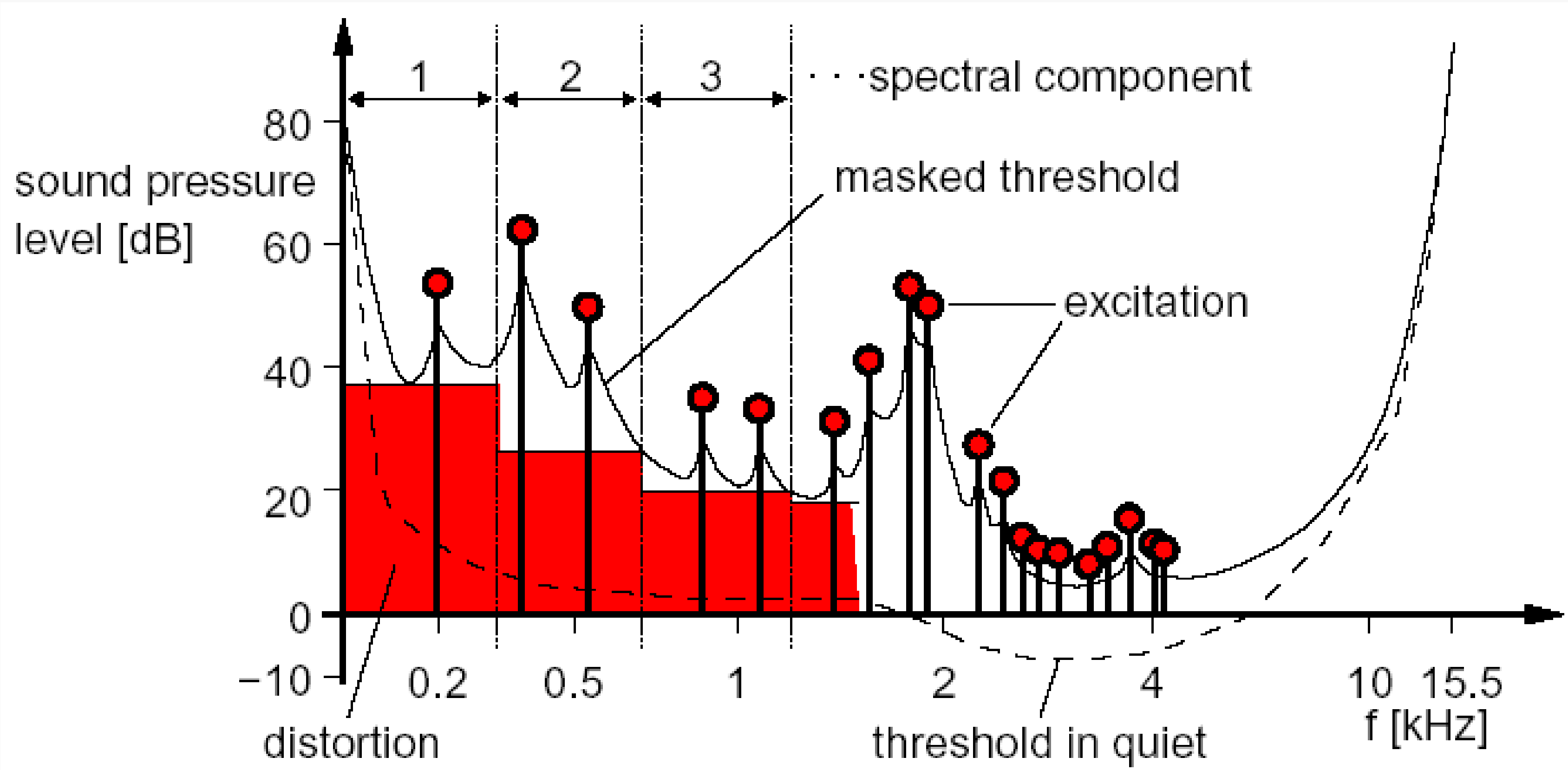
pre-masking: < 5 ms
post-masking: ~ 20 ms

Demonstration: Masked Threshold



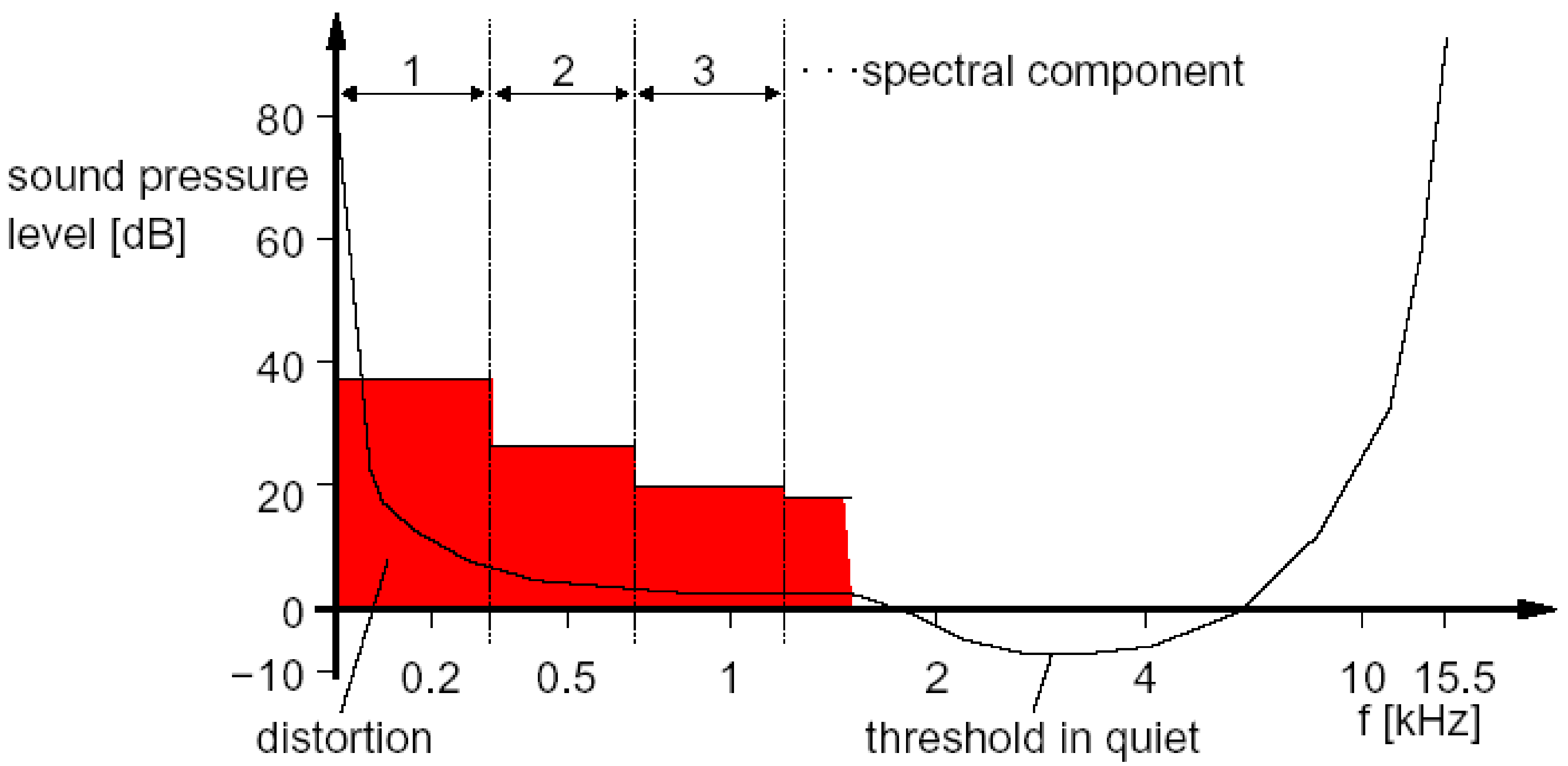
Original signal

Demonstration: Masked Threshold



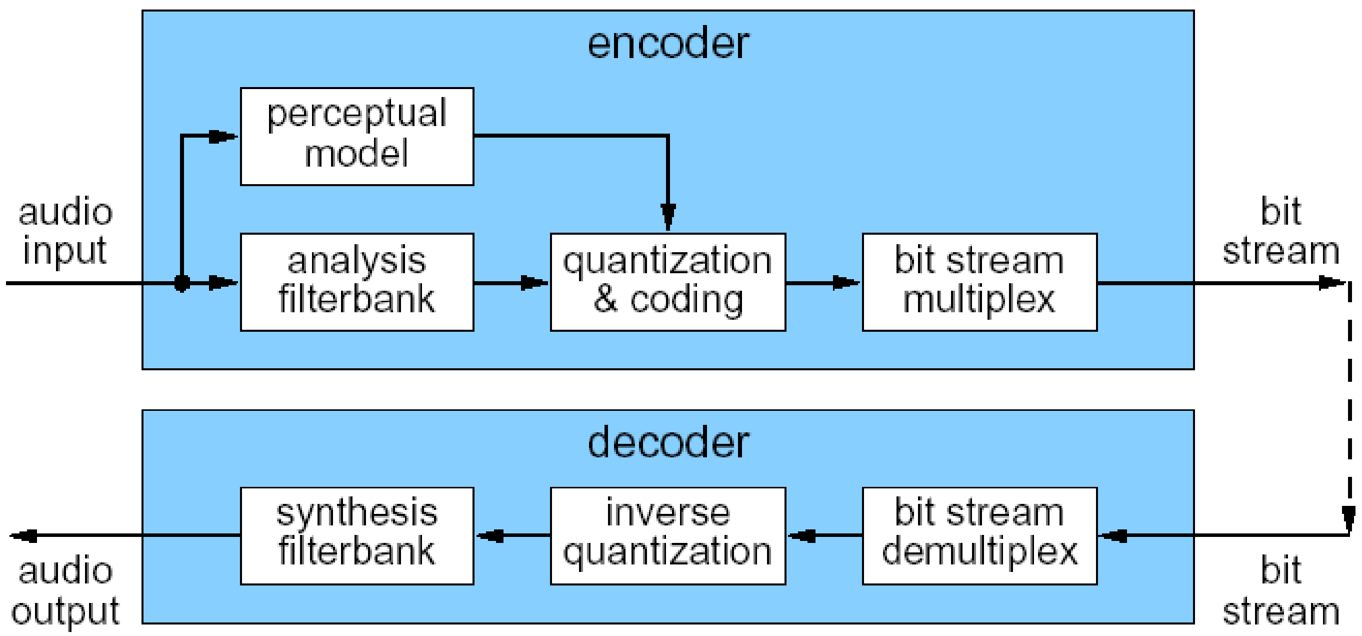
Original signal plus noise at masked threshold   (23 dB SNR)

Demonstration: Masked Threshold



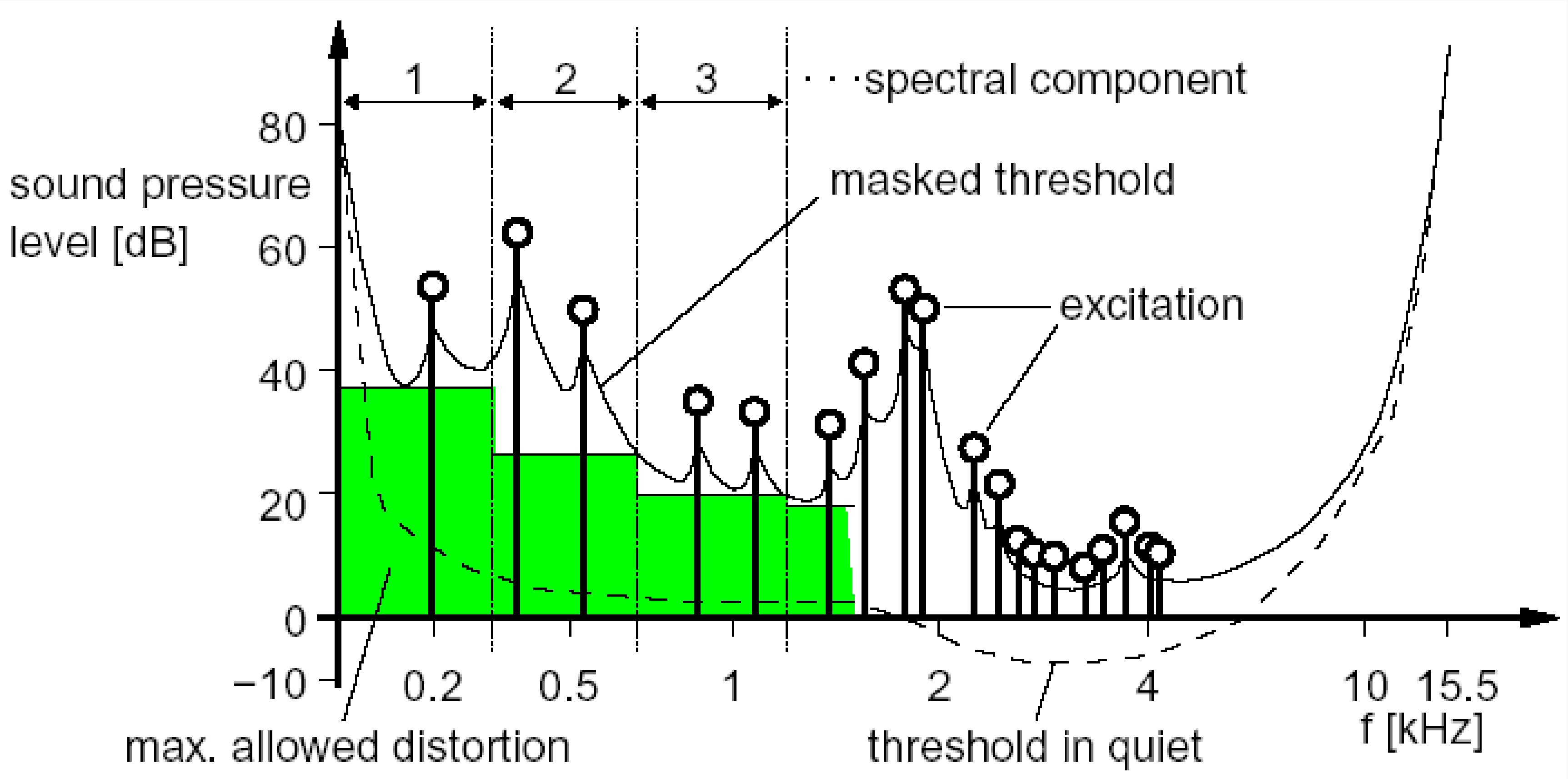
Noise at masked threshold without original signal

Subband/Transform Audio Codec



- Popular “waveform codecs”: MP3, Dolby Digital, AAC, …
- Analysis filterbank/transform enables…
  - …redundancy reduction (spectral decomposition)
  - …time- and frequency selective quantization

Psychoacoustic Model controls quantization



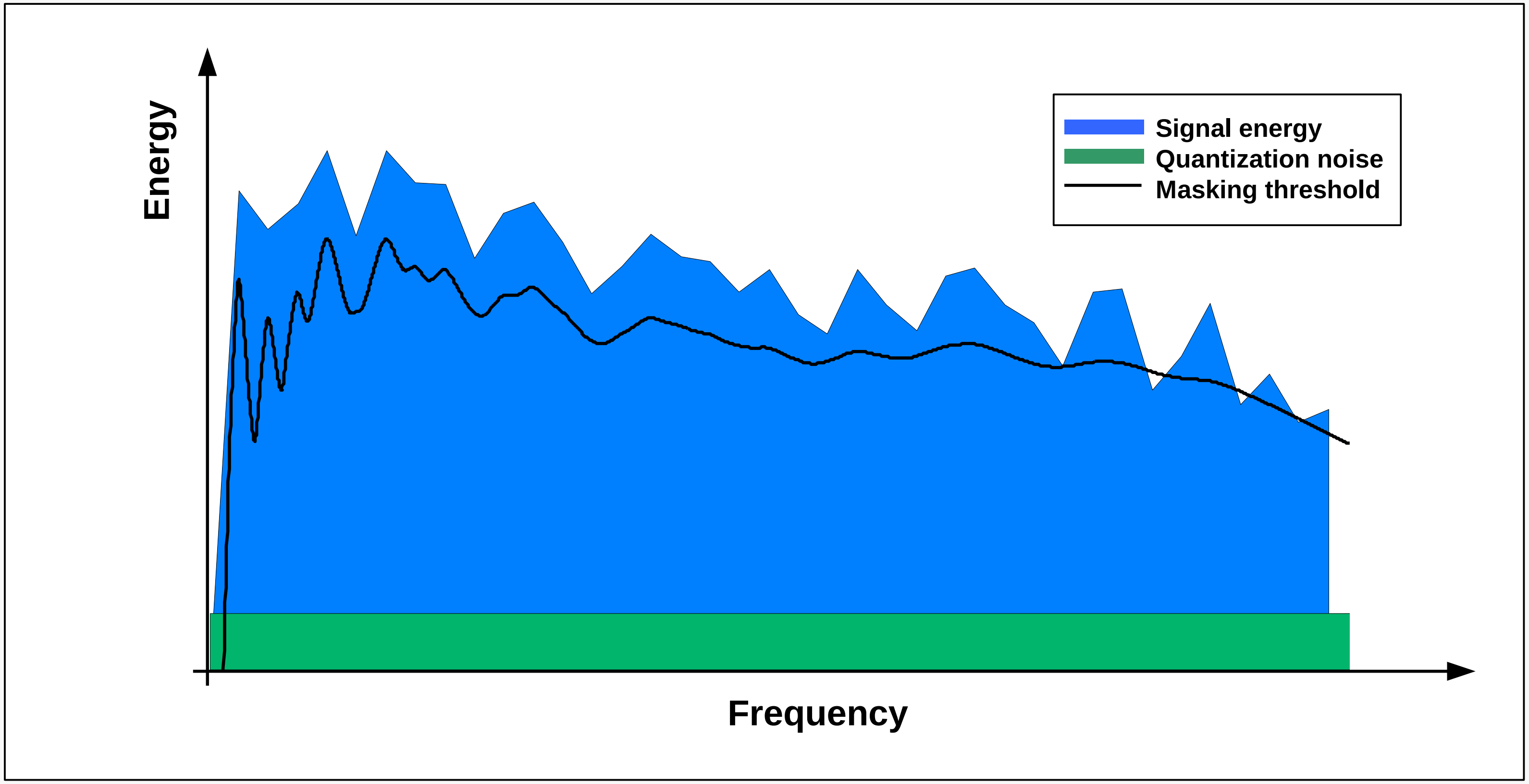
Goal: quantization error (distortion) below signal-dependent time and frequency variant masked threshold

Subband/Transform Audio Coding – Limitations

- Audio Demonstration
  - Original PCM stereo 1.5 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

Subband/Transform Audio Coding – Limitations

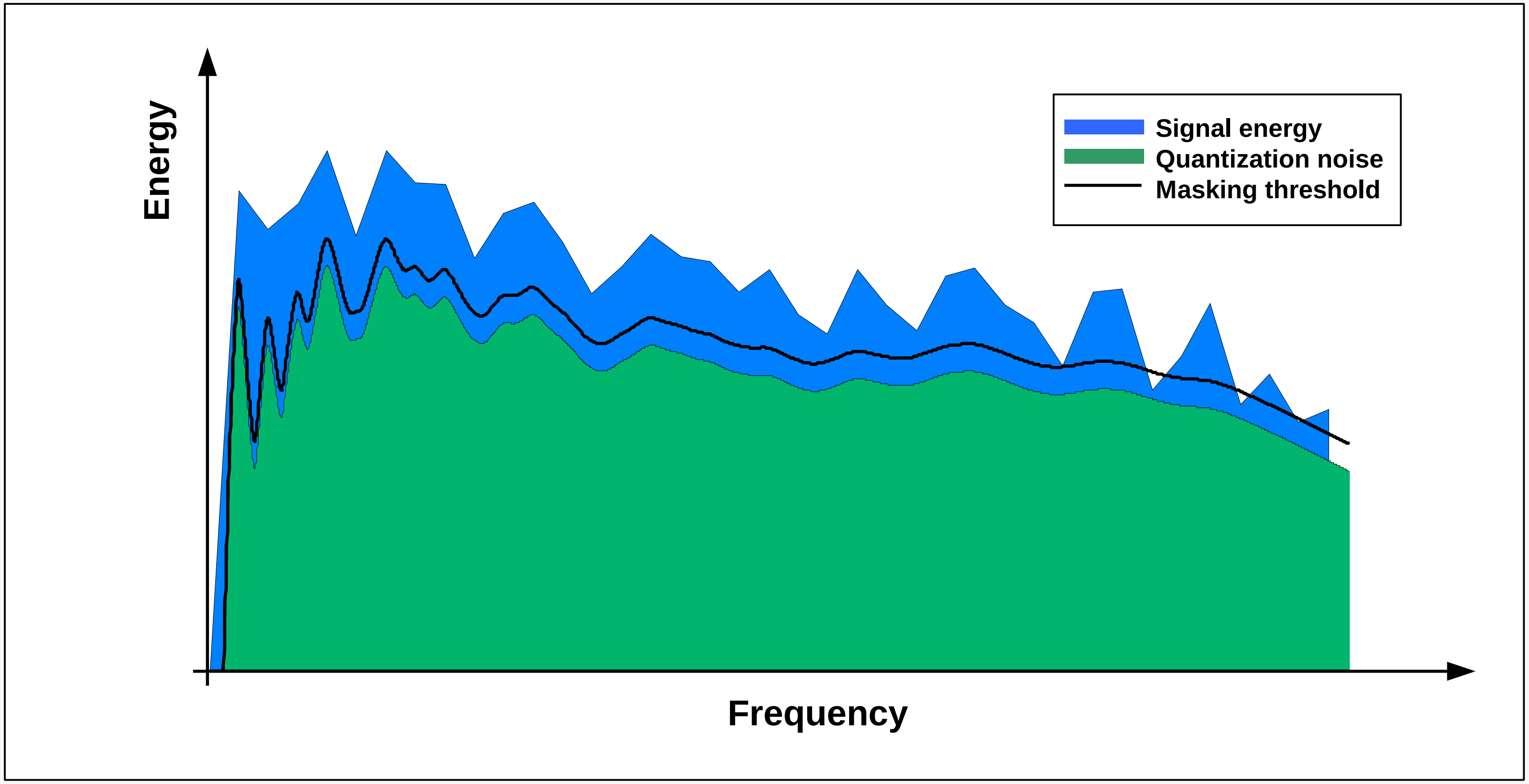
16 bit PCM



Visible blue area indicates bit rate

Subband/Transform Audio Coding – Limitations

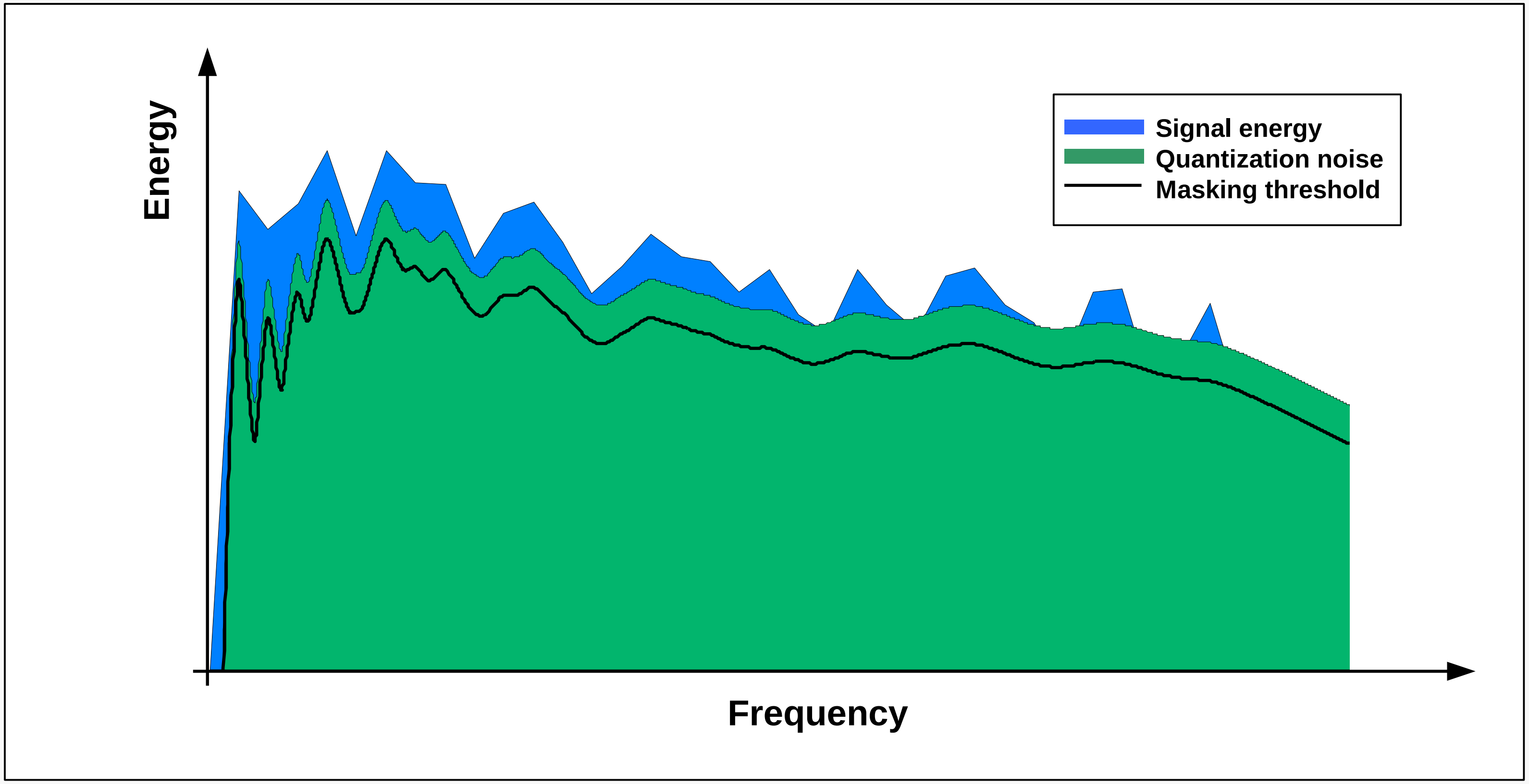
bit rate sufficiently high



Visible blue area indicates bit rate

Subband/Transform Audio Coding – Limitations

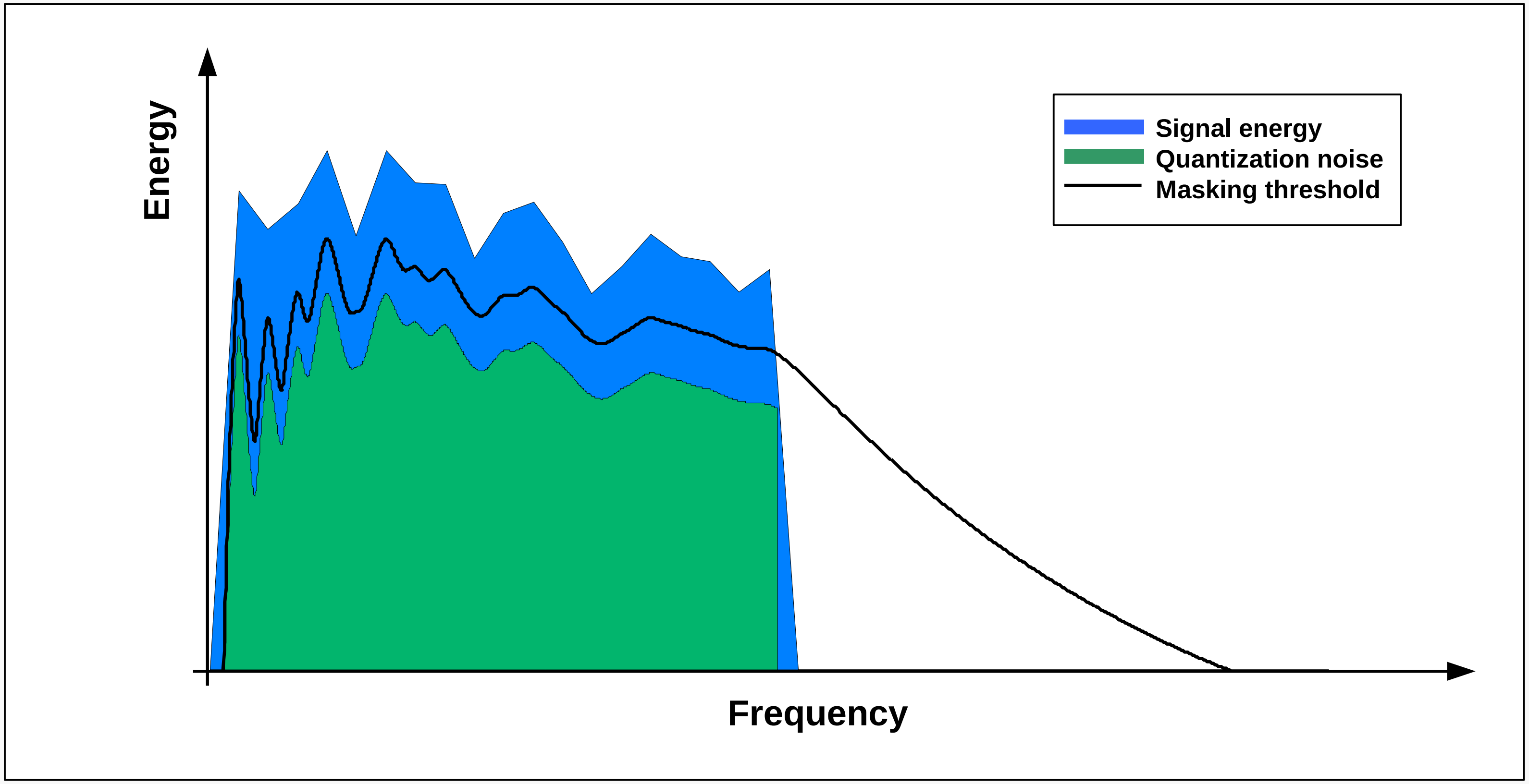
bit rate too low -> quantization noise audible



Visible blue area indicates bit rate

Subband/Transform Audio Coding – Limitations

- -> limit audio bandwidth to reduce coding artifacts



Visible blue area indicates bit rate

Spectral Band Replication (SBR)

- Problem
  - Conventional coding of high frequency band “expensive” (needs many bits)

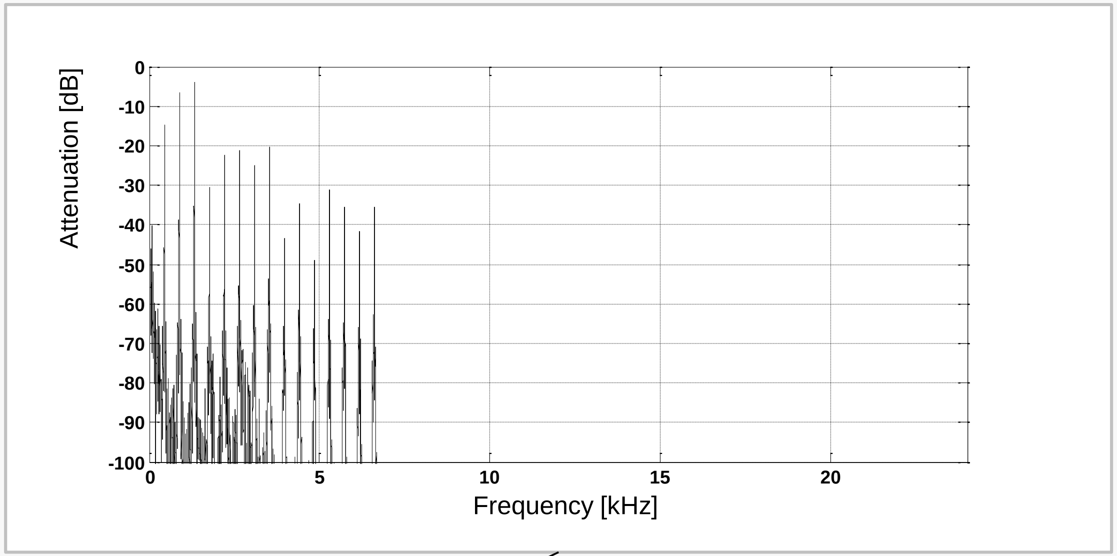
  => limited audio bandwidth at low bit rates

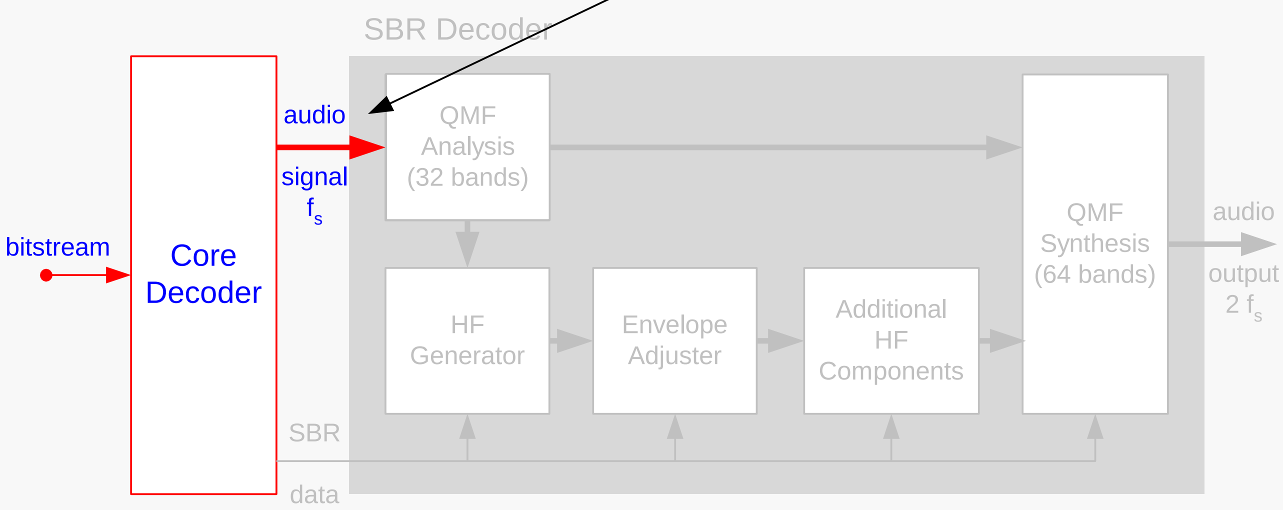
- Approach
  - Reconstruct high frequency band from low frequency band
  - Side information (2 to 3 kbit/s per channel) to control reconstruction

  => Spectral Band Replication (SBR)
  - “Parametric coding” of high frequency band (not “waveform coding”)
  - AAC+SBR: aacPlus / High-Efficiency AAC

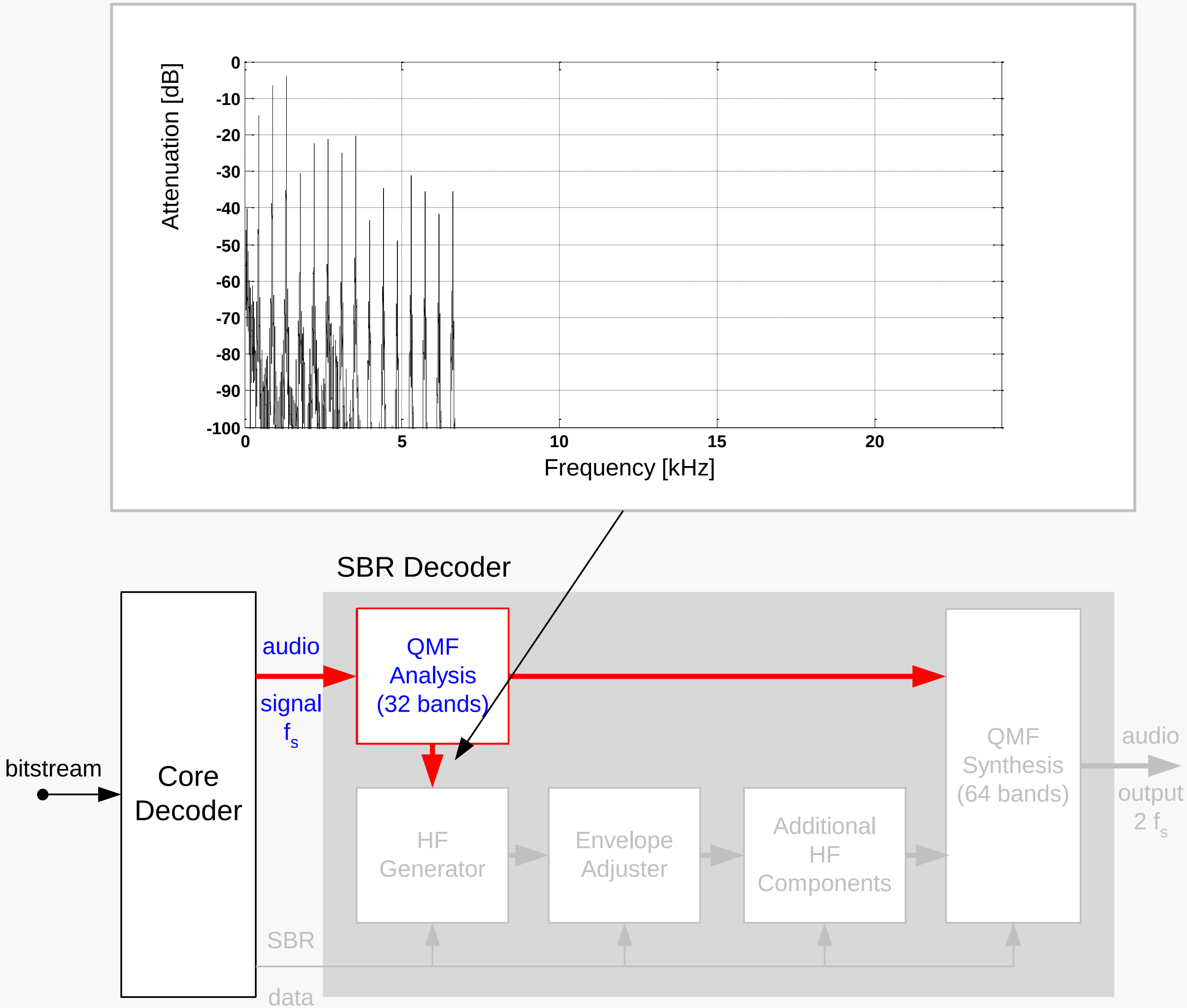
- Developed by Coding Technologies
  - founded 1997 by Lars “Stockis” Liljeryd
  - acquired by Dolby in 2007

Audio Decoder with SBR

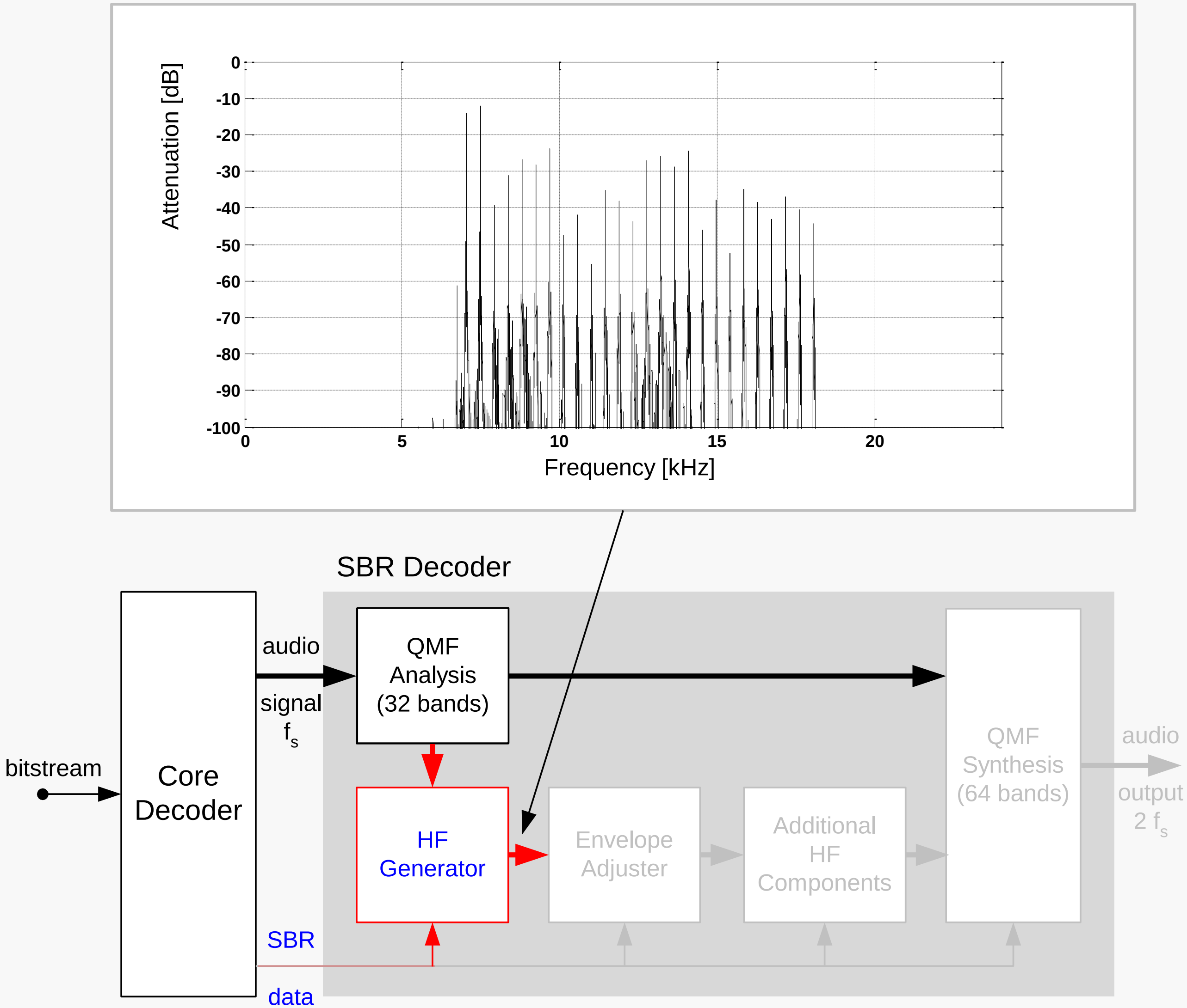




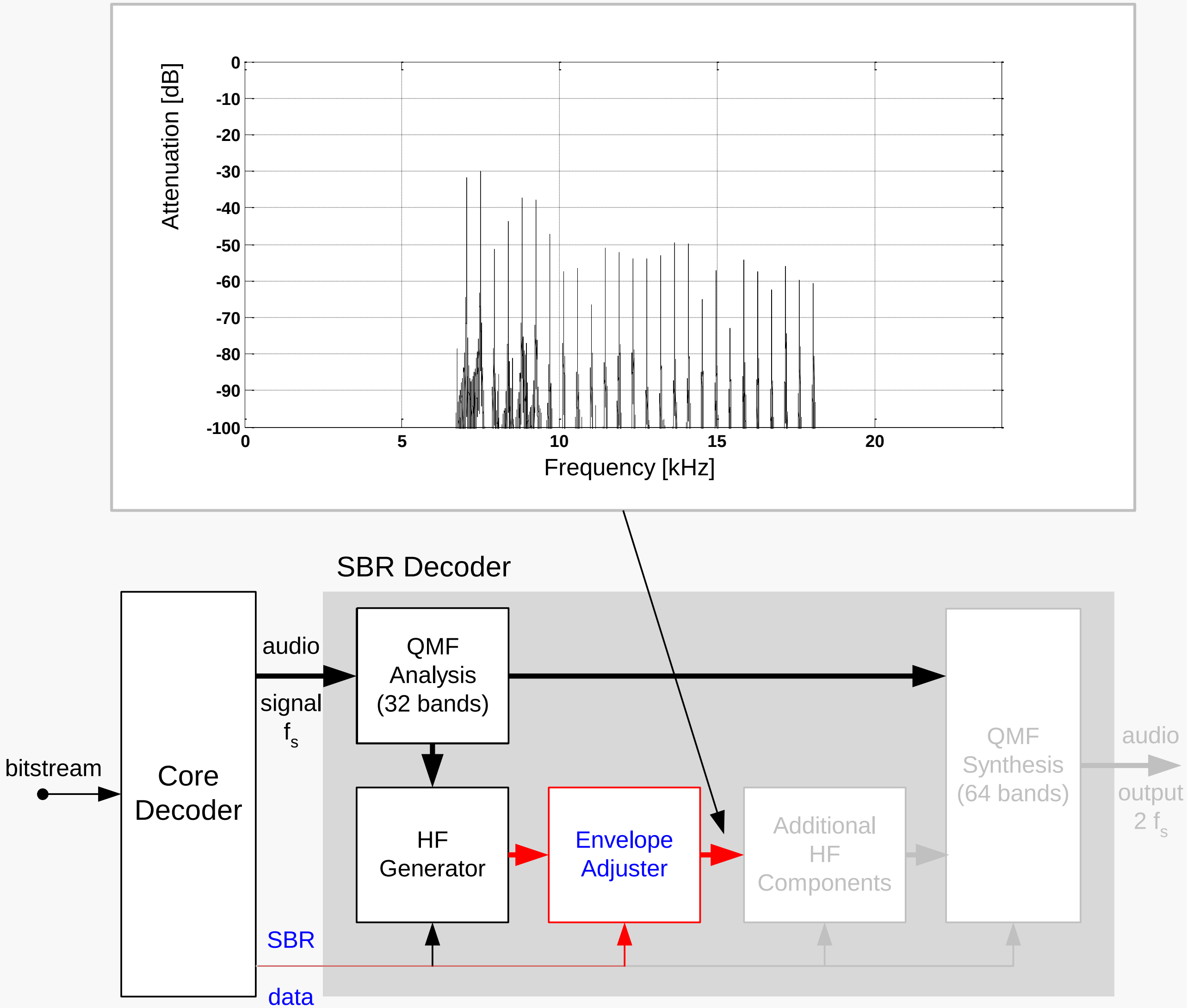
Audio Decoder with SBR



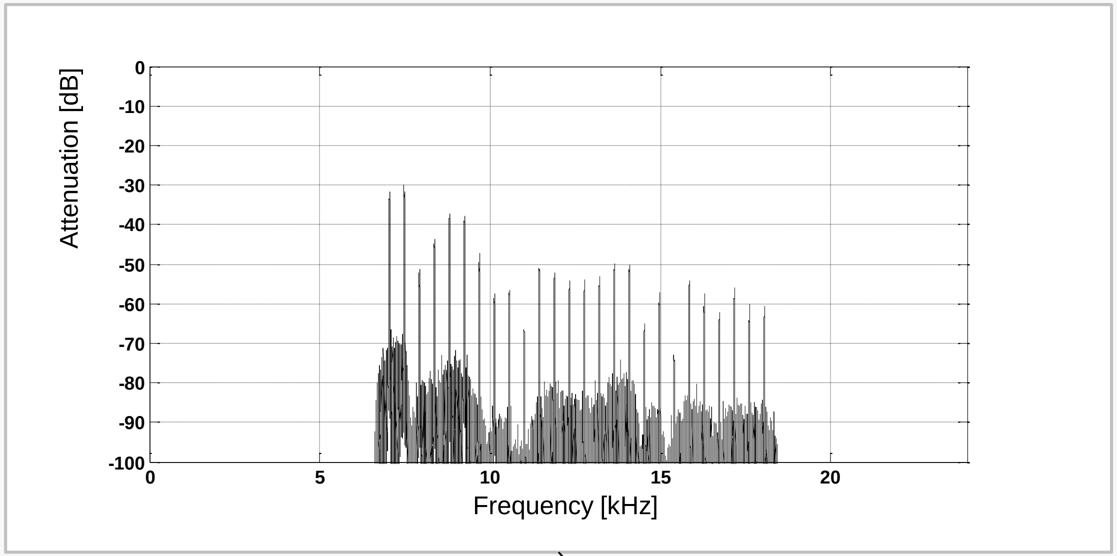
Audio Decoder with SBR

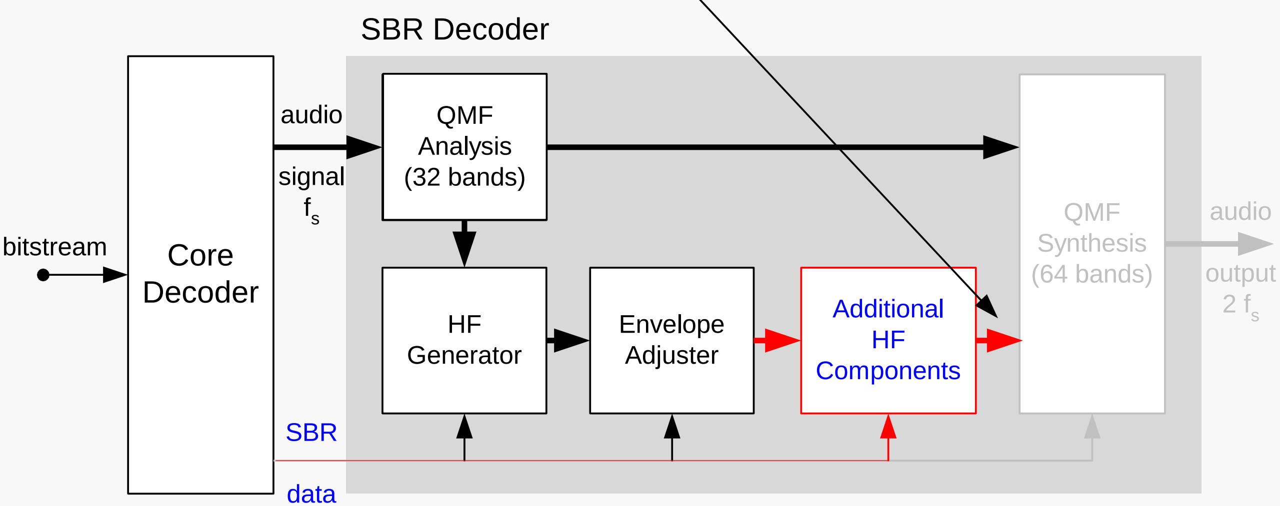


Audio Decoder with SBR

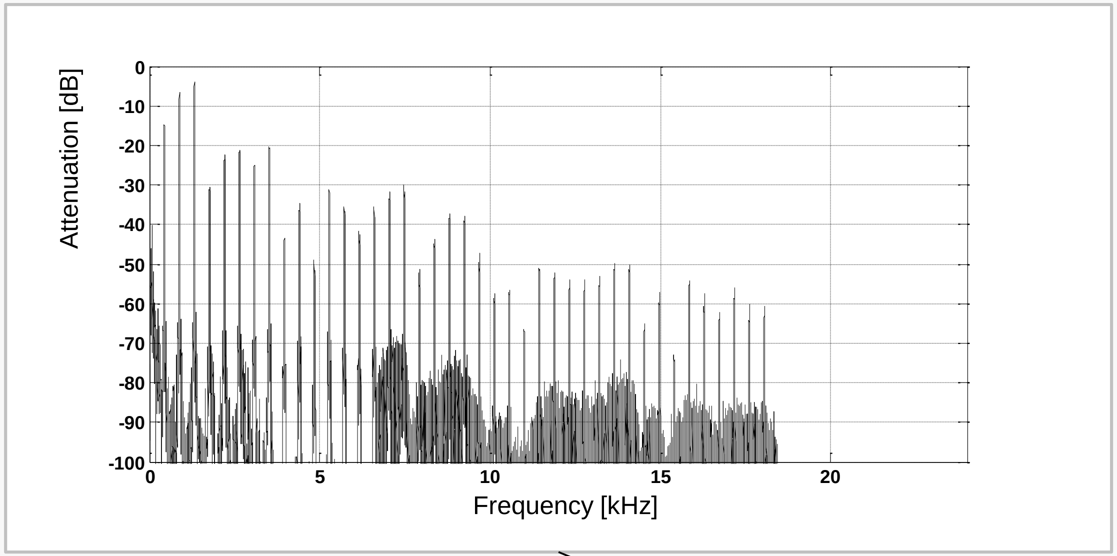


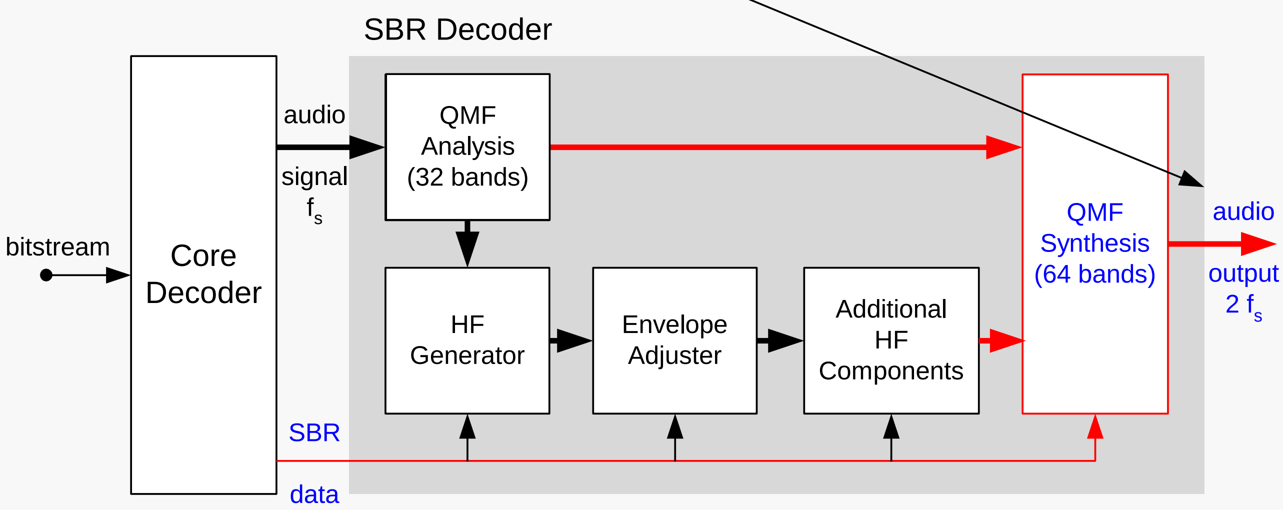
Audio Decoder with SBR





Audio Decoder with SBR

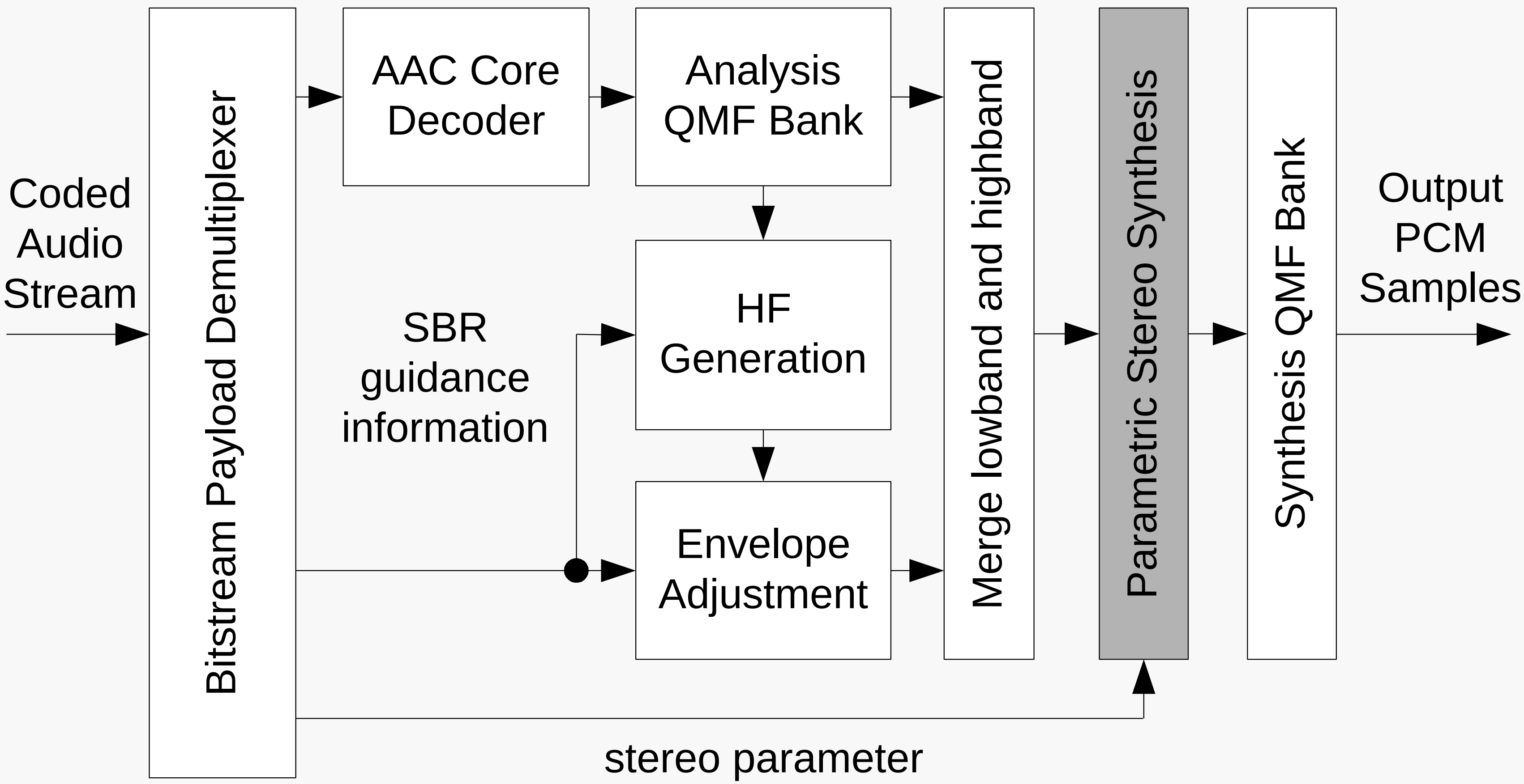




Parametric Stereo (PS)

- Problem
  - Mono coding better than stereo at low bit rates
- Approach
  - Transmit mono signal + stereo side information
  - Reconstruct stereo signal in decoder
- Stereo Parameter (per subband and time slot)
  - Interchannel Intensity Difference (IID) -> “pan”
  - InterChannel Correlation (ICC) -> “ambience”
- Approx. 2 to 3 kbit/s stereo side information
- “Parametric coding” of spatial/stereo image

Combining Parametric Stereo with SBR



Combining waveform and parametric coding

- AAC + SBR + PS => HE-AAC v2
  - MPEG-4 High-Efficiency AAC v2 (ISO/IEC 14496-3, published 2004)
    - Enhanced aacPlus (3GPP Release 6, TS 26.410)

- Audio Demonstration

Original PCM stereo		1.5 Mbit/s		
AAC stereo		48 kbit/s	1:32	
HE-AAC stereo	(AAC+SBR)	48 kbit/s	1:32	
HE-AAC v2 stereo	(AAC+SBR+PS)	24 kbit/s	1:64	

Recent developments

- Combination of speech and audio codecs
  - MPEG
    - USAC: Unified Speech and Audio Coding (ISO/IEC 23003-3)
  - IETF
    - Opus (RFC6716)
  - 3GPP/LTE
    - EVS: Enhanced Voice Services (3GPP Release 12, TS 26.445)
- Immersive sound
  - Dolby Atmos
  - NHK 22.2
  - MPEG-H 3D Audio

Thank you!