Ogg Vorbis

- Ogg project started 1993 to provide a license-fee free audio coder/ decoder
- Ogg: file transport protocol
- Vorbis: audio coder
  - Psycho-acoustically controlled forward adaptive monolithic codec based on MDCT
  - Inherently variable bit rate coder
  - Provides no framing, synchronization or error protection by itself (therefore use Ogg for file transport, RTP for multicast)
  - Low-complexity decoder, but high memory usage due to non-static probability models
  - Huffman and VQ codebooks are transmitted within bit stream header

Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

Windows Media Audio (WMA)

- Proprietary Audio Coder developed by Microsoft
- Collection of profiles for different applications:
  - WMA 9: most scenarios, backwards compatible to WMA 8, about 20% lower data rate, VBR possible
  - WMA 9 professional: 24 bit/96 kHz audio, 7.1 channels, 128-768 kbps, stereo downmix available
  - WMA 9 voice: speech content at low bit rates (<20 kbps)
  - WMA 9 lossless: compression depending on input audio, used for high-quality archiving purposes

© Fraunhofer IDMT

WMA: main features

- MDCT (or MLT) based
- Multiple numbers of frequency lines (128, 256, 512, 1024, 2048)
- Sinusoidal shaped windows, transition windows and “bridge” windows (“soft” transition between long and short blocks)
- Uniform quantization within scale factor bands
- M/S coding frame-by-frame instead of scale-factor-band-wise
- Bit reservoir available (1-pass and 2-pass coding)

© Fraunhofer IDMT

SOAC and MPEG-H

Prof. Dr.-Ing. Karlheinz Brandenburg

Fraunhofer IDMT & Ilmenau Technical University
Ilmenau, Germany

© Fraunhofer IDMT

SAOC – Spatial Audio Object Coding

Outline

- Introduction
- From Spatial Audio Coding to SAOC
- Audio objects
- SAOC Decoding
- Applications
- Performance Evaluation
- Conclusion

Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

SAOC - Introduction

- Perceptual audio coding for multichannel signals is widely used
  - “Spatial Audio Coding”, for instance MPEG Surround
- Existing Spatial Audio Coders are channel-based
  - Designed for a specific reproduction setup

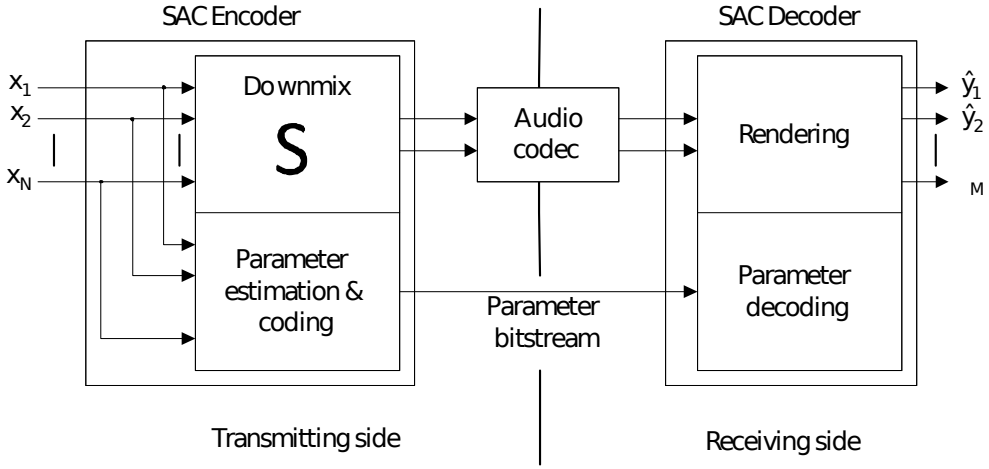
- Spatial Audio Object Coding
  - Continuation of the “Spatial Audio Coding” paradigm
  - Transmit audio objects instead of channel signals
  - ISO/IEC 23003-2:2010 Standard

© Fraunhofer IDMT

SAOC – From Spatial Audio Coding to SAOC

Spatial Audio Coding (e.g., MPEG Surround)

- Channel-oriented
- Downmix (mono or stereo)
- Transmit downmix using standard audio codec (AAC)
- Additional parameter data (parametric coding)
- Output channels for specific reproduction setup
  - 5.1, 7.1

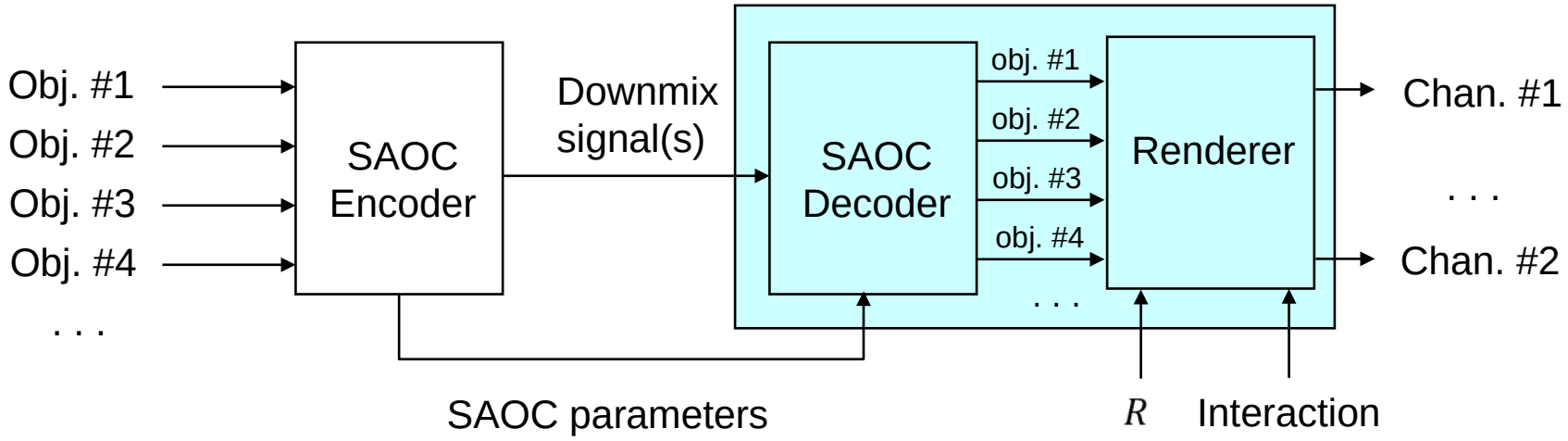


[1] Herre et.al 2012: "MPEG Spatial Audio Object Coding, J. Audio Eng. Soc. 60:9, 2012

© Fraunhofer IDMT

SAOC – Audio Objects

- Audio objects instead of channels
- SAOC encoder: Stereo or mono downmix plus SAOC parameters
- SAOC decoder: Use SAOC parameters to transform downmix into audio objects
- Rendering to loudspeaker configuration (Rendering matrix)



© Fraunhofer IDMT

SAOC – Audio Objects

Advantages of object-based processing

- Coding efficiency: SOAC parameters only a few kbit/s per audio object
- Coding and transmission independent of reproduction setup
- Rendering on arbitrary loudspeaker setups
  - 5.1, 7.1, 10.2, 22.2, Binaural reproduction, Wave field synthesis,...
  - Rendering controllable (real-time user interaction)
- Control over individual audio objects
  - Change gain, equalization, effects, ...

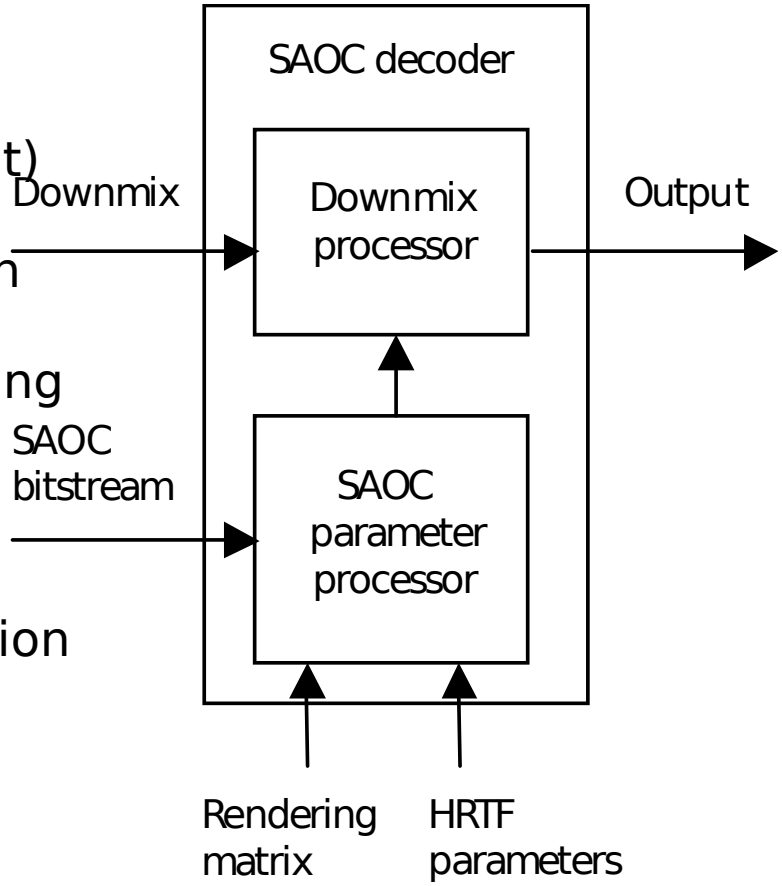
Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

SAOC Decoding Modes

Decoder Processing Mode

- Rendering integrated into decoding (efficient)
- For mono and stereo output, incl. binaural reproduction
- Rendering matrix: realtime control of rendering
- HRTF parameters for binaural
  - Open SAOC interface
  - Enables use of individual HRTFs
  - Efficient parametric representation

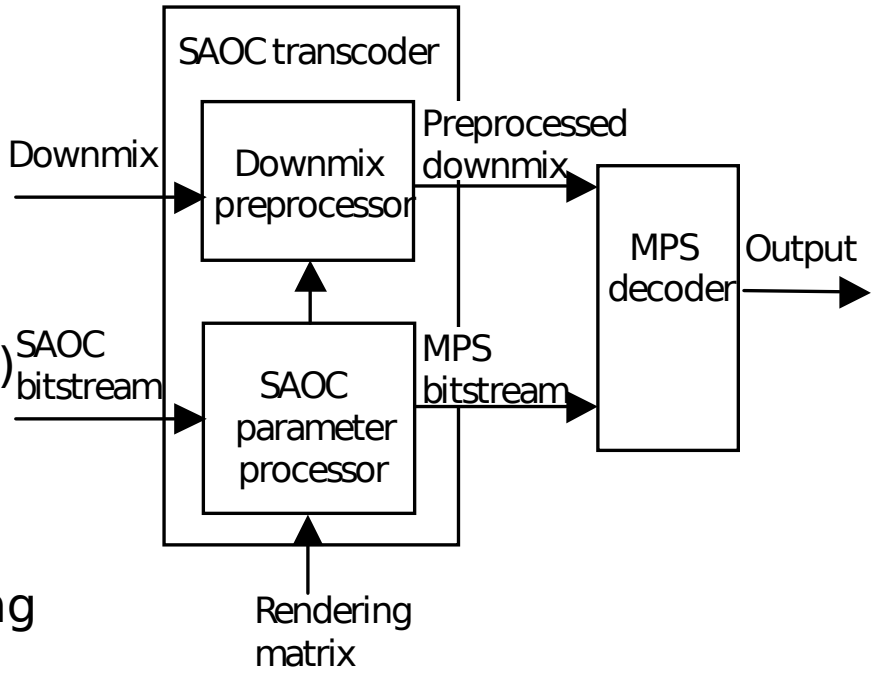


© Fraunhofer IDMT

SAOC Decoding Modes

Transcoder Processing Mode

- For multichannel output (MPEG Surround - MPS)
- SAOC encoder works as transcoder
  - Transcoding of SAOC parameters to MPS bitstream
  - Adjustment of downmix panning (only for stereo downmix)
  - Highly Efficient
    - Operates in transform domain
    - Avoid unnecessary (de)quantization and decoding



Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

SAOC Bitstream

- Contains parametric description of audio objects: SAOC parameters
- Typical: 2-3 kbit/s per audio object (plus 3 kbit/s per audio scene)
- SAOC bitstream embedded in ancillary data of core audio coder
  - Enables backward compatibility
- Parameters transmitted in flexible time/frequency grid
  - Adaptation to bitrate demands and/or signal characteristics
  - Same time/frequency grid as in MPEG Surround
    - Lossless, efficient transcoding

© Fraunhofer IDMT

SAOC Parameters

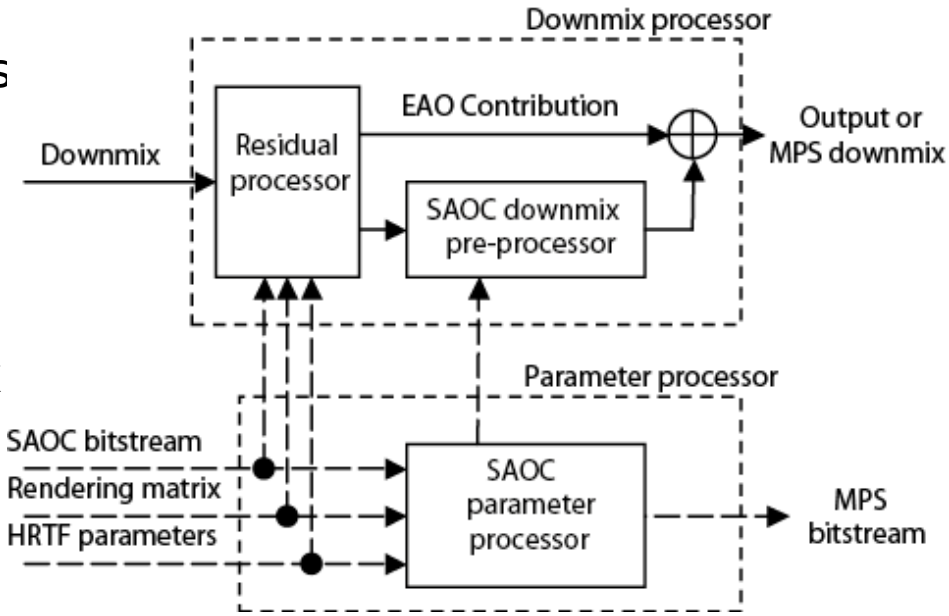
- Object Level Differences (OLD): Level relative to loudest object
- Inter-Object Cross Correlations (IOC): Similarity between pairs of objects
- Downmix Gain (DMG): Gains used in the downmix of individual objects
- Object Energies (NRG): Absolute energy of loudest object. Optional, enables merging of multiple SAOC streams

Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

SAOC – Enhanced Audio Objects

- Allow arbitrary attenuation or amplification of objects
  - Karaoke
  - Solo voices,…
- SAOC bitstream contains residual signal
- Reconstruction from downmix and residual
- Efficient transmission of residual signal (AAC)



Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

SAOC - Applications

Interactive Remix / Karaoke

- Interactive remixes
- Equalization, room simulation,... for individual objects
  - For channel-based formats, only applicable to whole scene
- Modification of specific audio objects (instruments, voices,...)
- Karaoke, vocal solo
  - Suppress main voice or background music
  - Advantageous: Enhanced Audio Objects
- Future extensions of digital broadcasting
  - Clean-audio dialogs
  - Additional objects for interactivity

© Fraunhofer IDMT

SAOC - Applications

Teleconferencing

- Today: Mainly monophonic reproduction
  - Suboptimal for multi-user scenarios

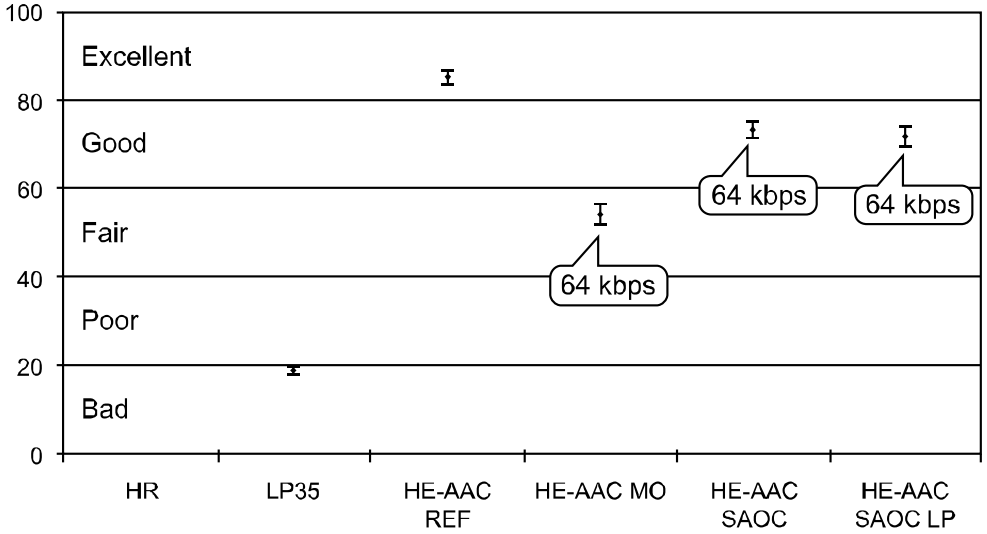
- Key benefits of SAOC
  - Adjustment of individual speaker signals
  - Spatial representation of audio scene
  - Improved intelligibility and listening comfort
  - Match between visual and audio scene
  - Transmission efficiency
  - Backward compatibility

© Fraunhofer IDMT

SAOC – Performance Evaluation

Listening Test – Remix scenario

- Part of MPEG verification tests (5 sites, 125 participants)
- MUSHRA test (ITU BS.1534-1)
- Simulate adjustments to a mix of audio objects
- Core coder: High Efficiency AAC (HE-AAC)



HR: Hidden reference
LP: 3.5 kHz Lowpass
HE-AAC REF: Individual objects, high bitrate
HE-AAC MO: Individual objects, same bitrate
HE-AAC SAOC: standard SAOC
HE-AAC SAOC LP: low power

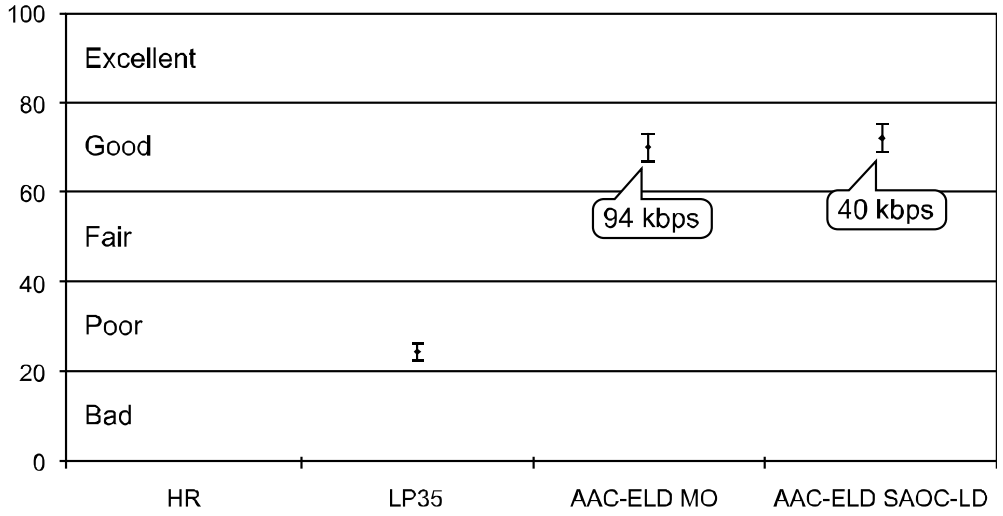
Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

SAOC – Performance Evaluation

Listening Test – Teleconferencing

- Part of MPEG verification tests
- Teleconferencing application: Simulate adjustments of a participant
- Core coder: MPEG-4 Enhanced Low Delay AAC (AAC-ELD)



HR: Hidden reference
LP: 3.5 kHz Lowpass
AAC-ELD MO: Individual objects
AAC-ELD SAOC-LD: low delay

Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

Universal Format for 3D-Audio

MPEG-H

© Fraunhofer IDMT

MPEG-H: Universal Format for 3D-Audio

Recording as:

- HOA (Higher Order Ambisonics)
- Audio-Objects (Track + Metadata)
- SAOC (similar to what we have seen before)

- Or channel based formats (legacy-content)

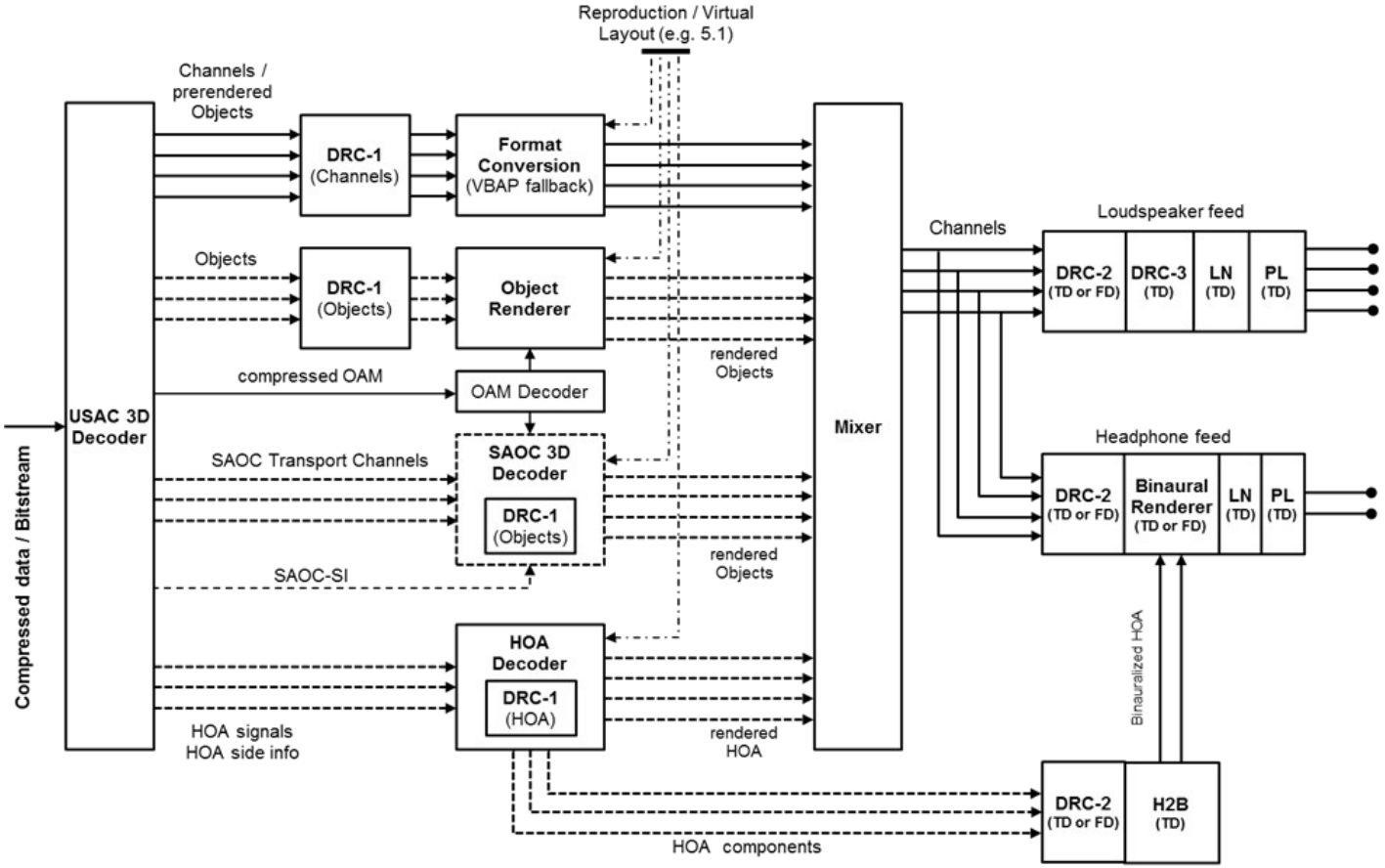
Replay as:

- Audio objects (via VBAP)
- Needs geometry data, metadata and one audio stream each

Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

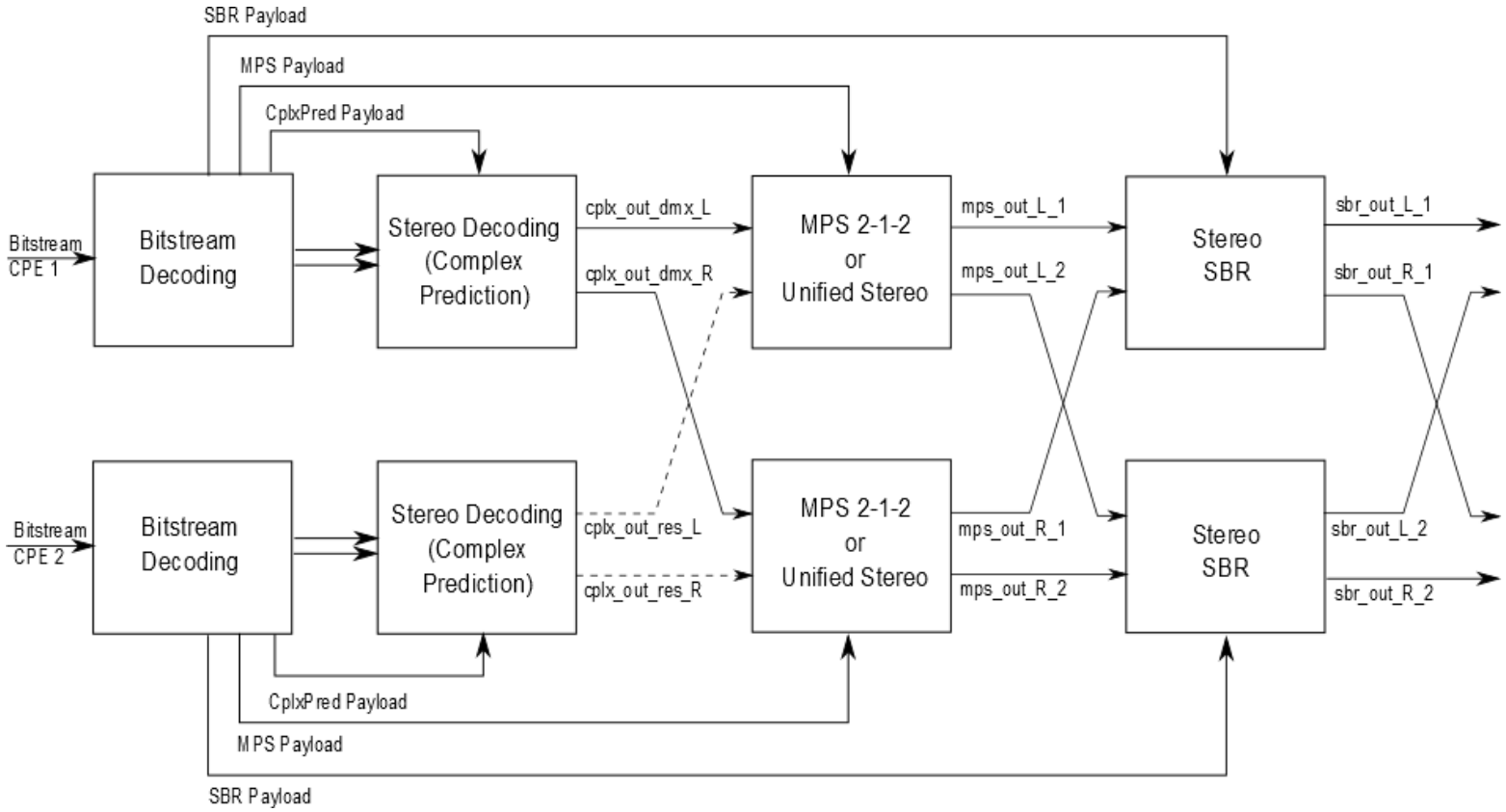
MPEG-H: Block Diagram Decoder



Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

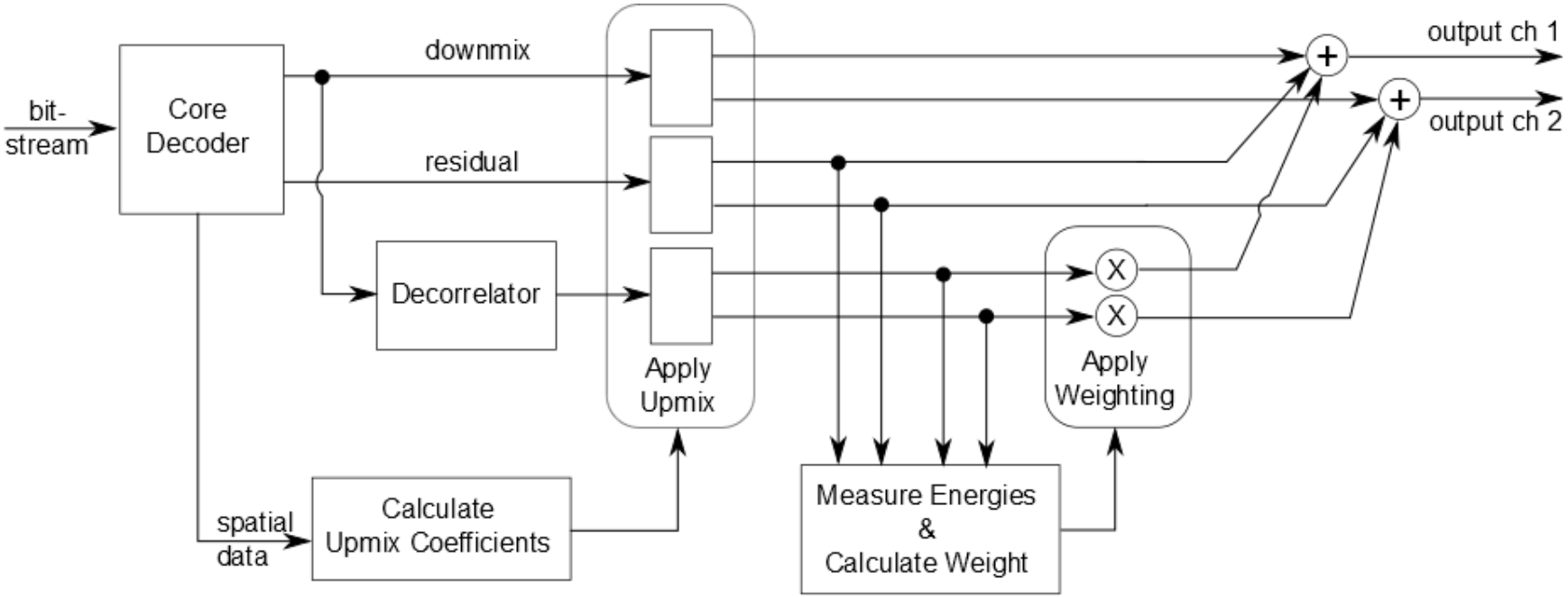
Quad Channel Elements



Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

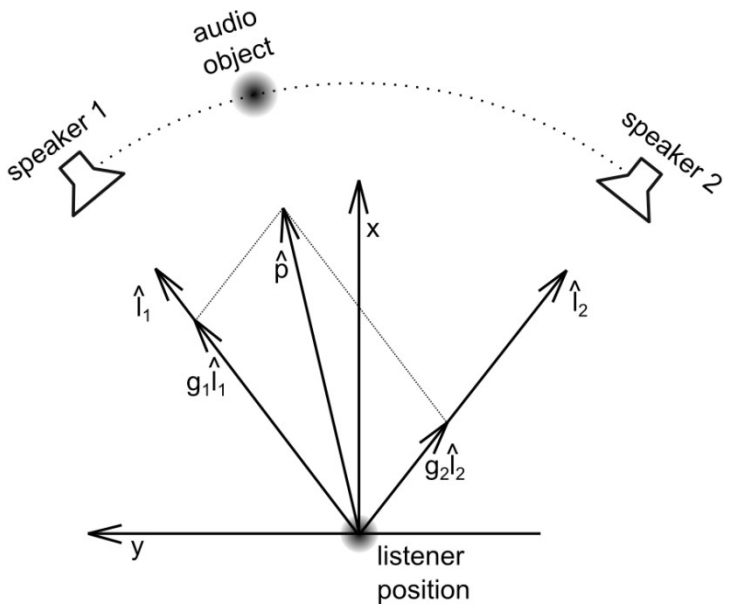
Hybrid Residual Coding



Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

VBAP: Audio object as linear combination of loudspeaker vectors



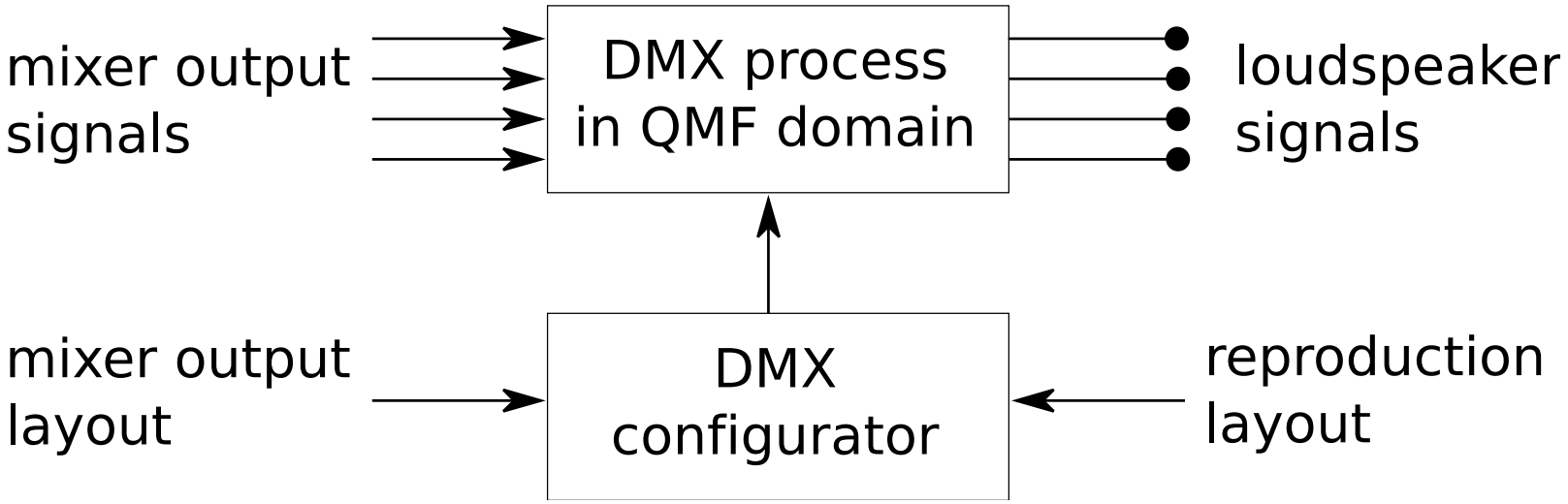
Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Format Converter: Downmix for rendering



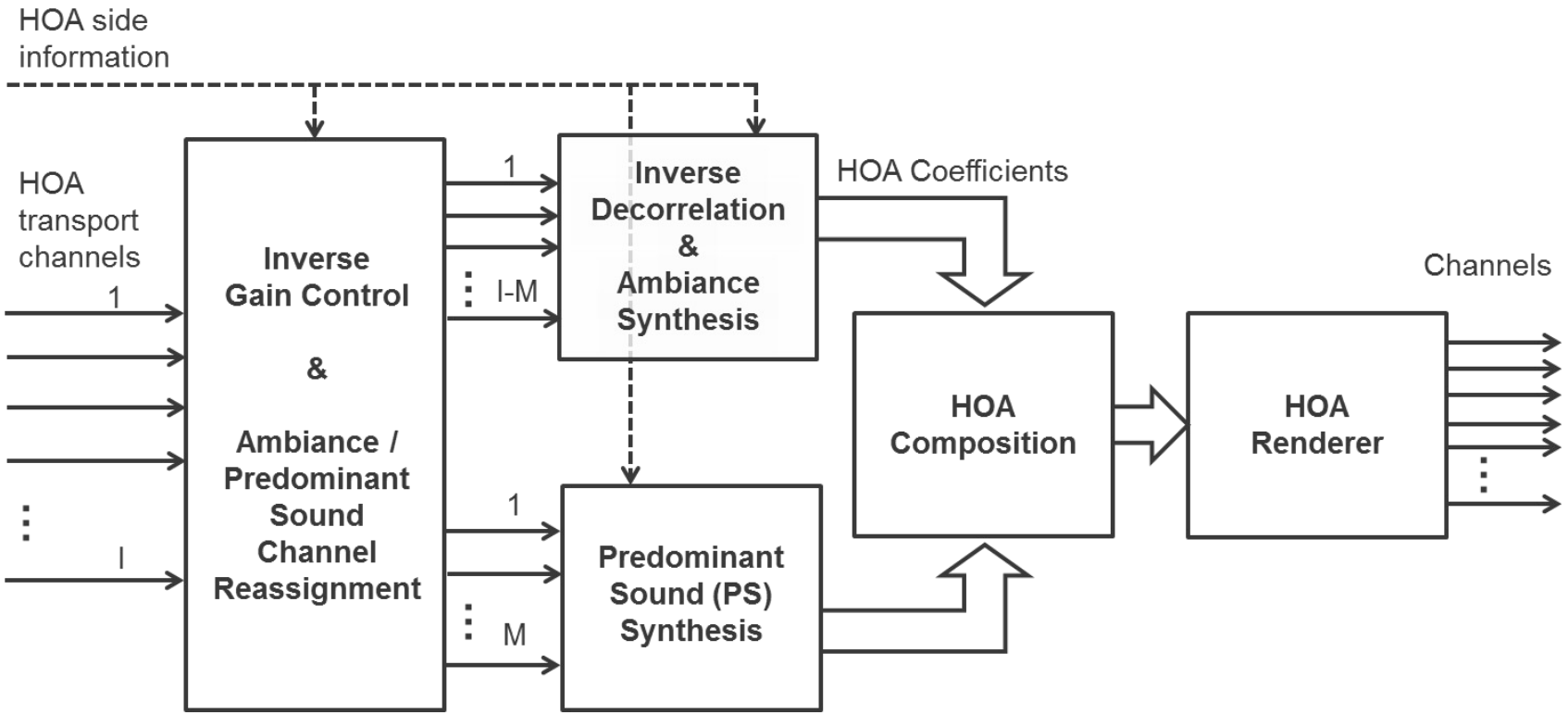
Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

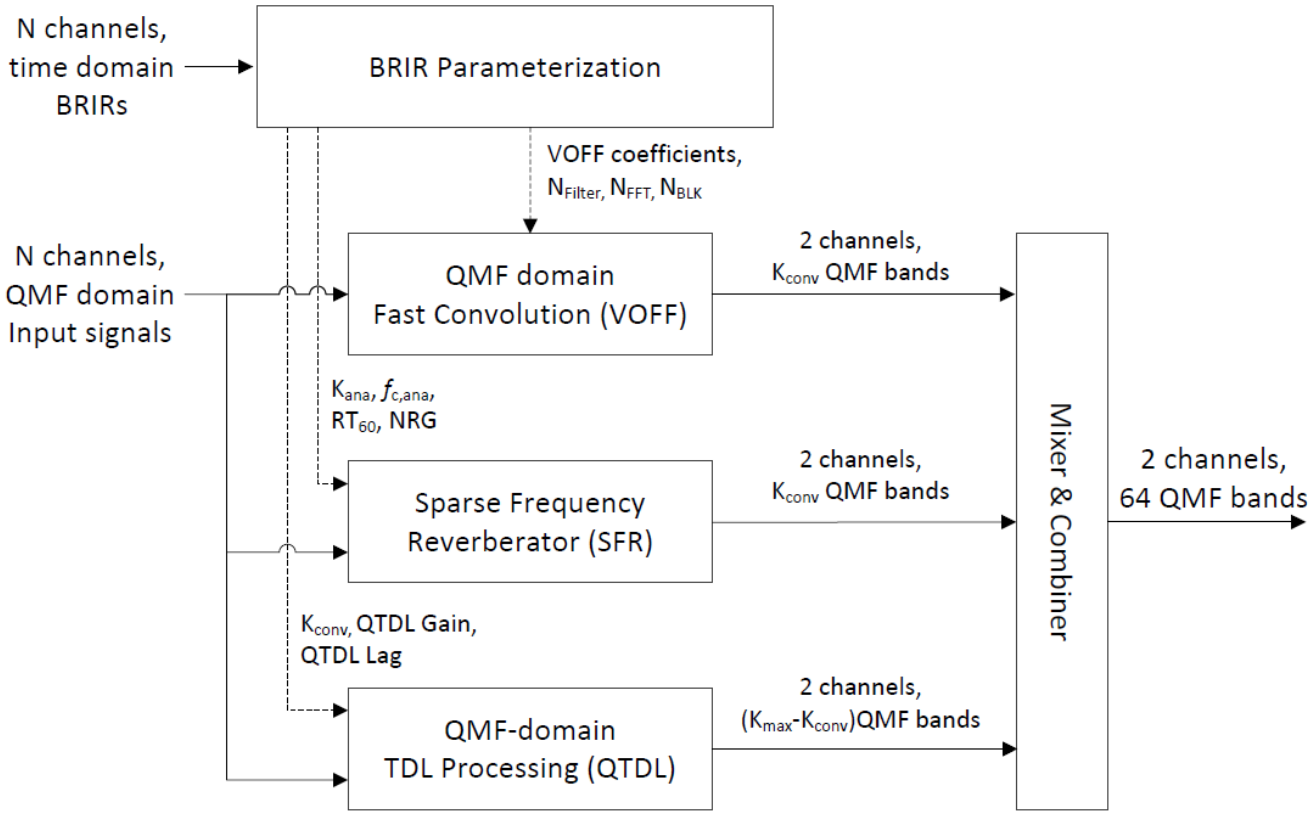
HOA – Decoder



Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT

Binaural reproduction



Prof. Dr.-Ing. K. Brandenburg, karlheinz.brandenburg@tu-ilmenau.de Prof. Dr.-Ing. G. Schuller, gerald.schuller@tu-ilmenau.de

© Fraunhofer IDMT