The MPEG-4 General Audio Coder

Bernhard Grill
Fraunhofer Institute for Integrated Circuits (IIS)

Fraunhofer Institut
Integrierte Schaltungen

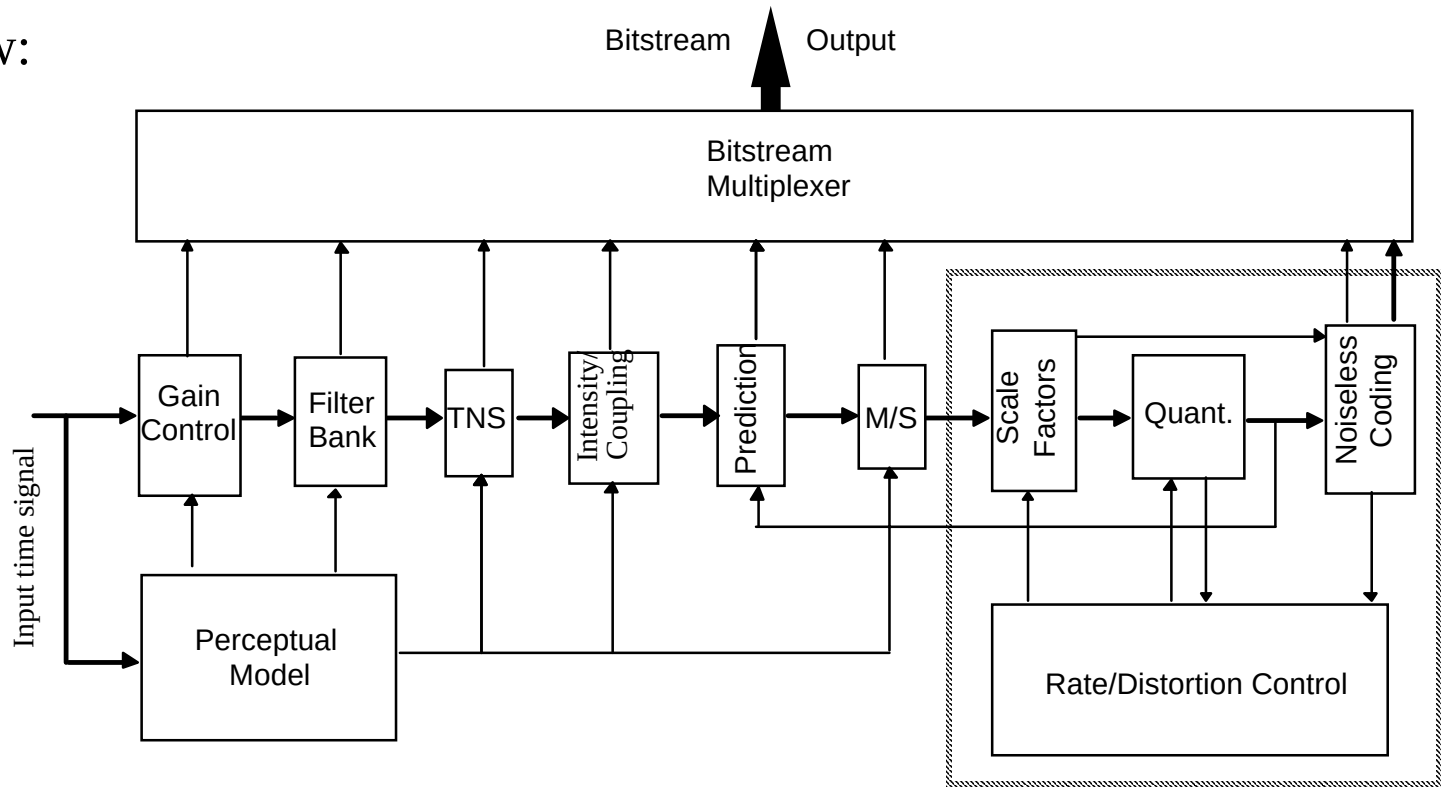
Outline

- MPEG-2 Advanced Audio Coding (AAC)
- MPEG-4 Extensions:
  - Perceptual Noise Substitution (PNS)
  - Long Term Prediction
  - TwinVQ Coding Core
- The MPEG-4 Scalable General Audio Coder
- Results of Listening Tests
- Demonstration of a Real-Time Player

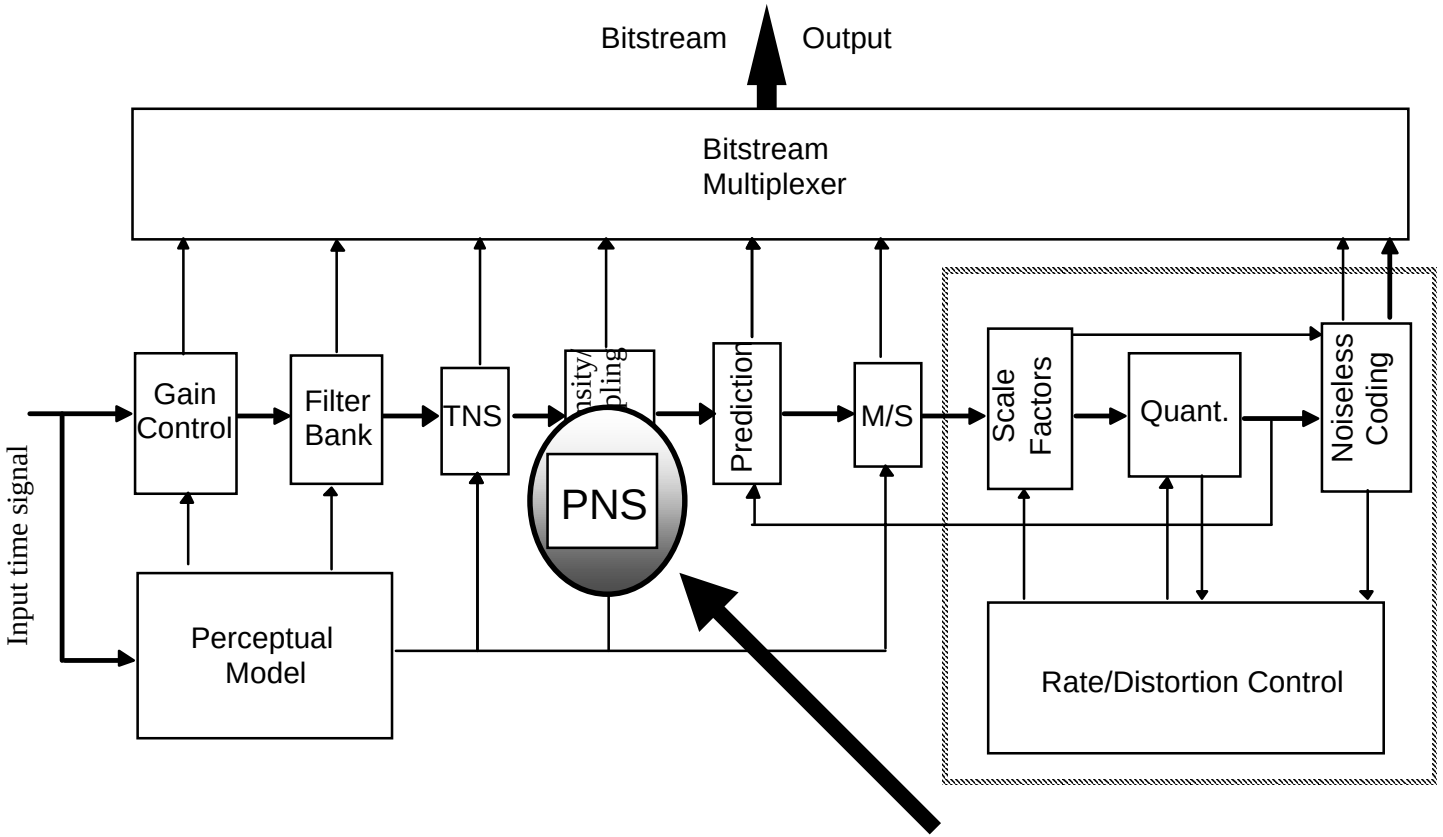
Fraunhofer Institut Integrierte Schaltungen

MPEG-2 AAC

Encoder Overview:



Extension: Perceptual Noise Substitution (PNS)



Perceptual Noise Substitution (2)

Background:

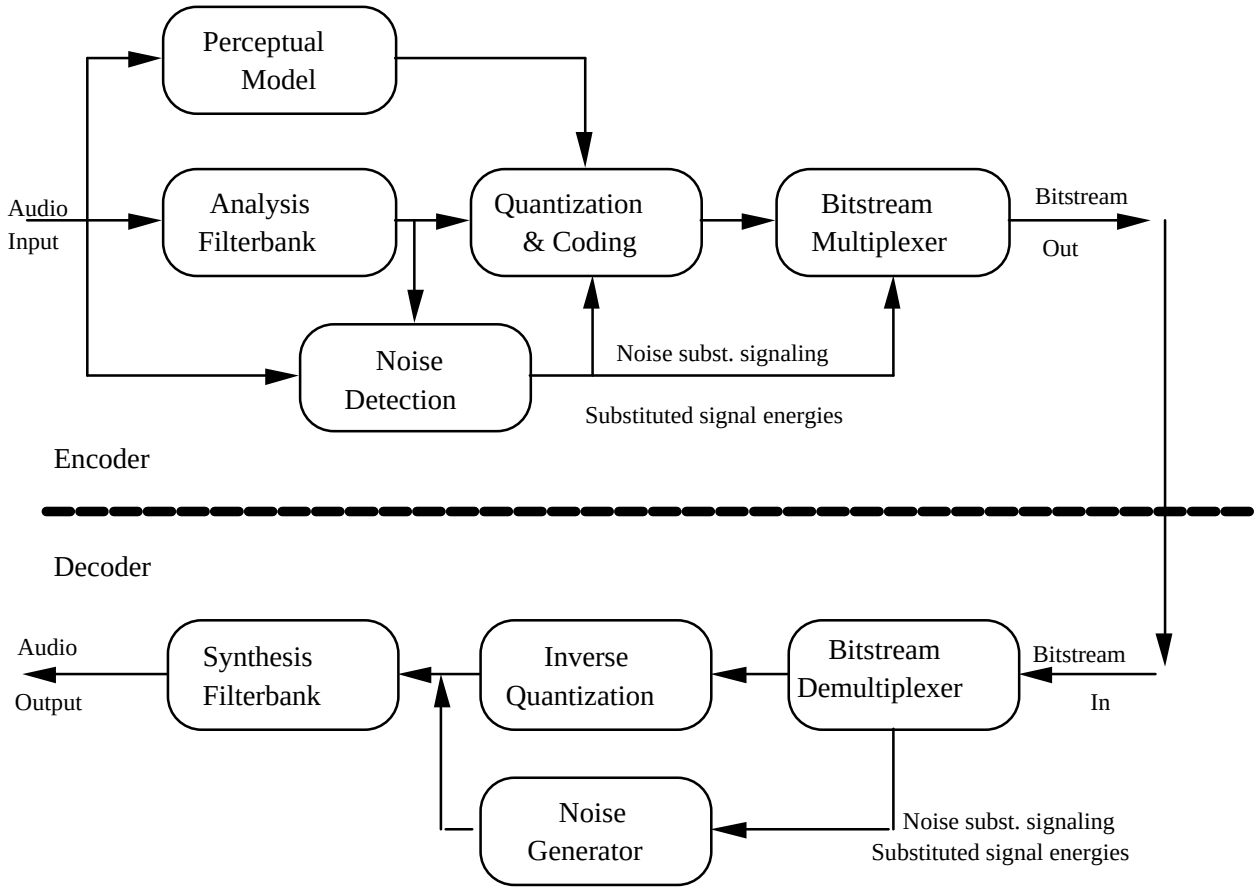
- Parametric coding of noise-like signal components has been used widely e.g. in speech coding

MPEG-4:

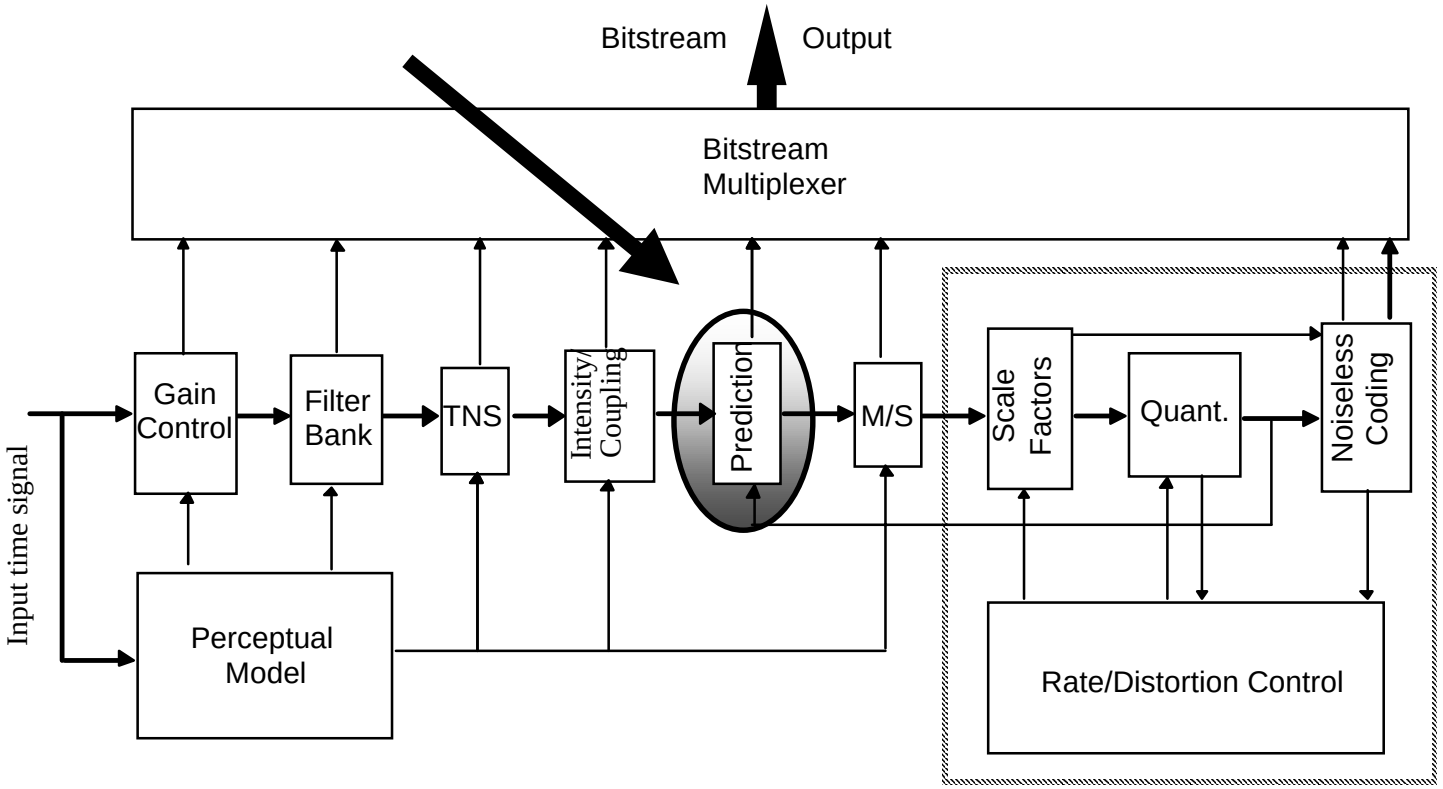
- Perceptual Noise Substitution (PNS) permits a frequency selective parametric coding of noise-like signal components
- Noise-like signal components are detected on a scalefactor band basis
- Corresponding groups of spectral coefficients are excluded from quantization/coding
- Instead, only a "noise substitution flag" plus total power of the substituted band is transmitted in the bitstream
- Decoder inserts pseudo random vectors with desired target power as spectral coefficients

Perceptual Noise Substitution (3)

"Perceptual Noise Substitution" (PNS): Perceptual coder + parametric represent. of noise-like signals



Extension 2: Long Term Prediction



Long Term Prediction (2)

Motivation:

- Tone-like signals require much higher coding precision than noise-like signals (e.g. 20 dB vs. 6 dB)
- Tonal signal components are predictable

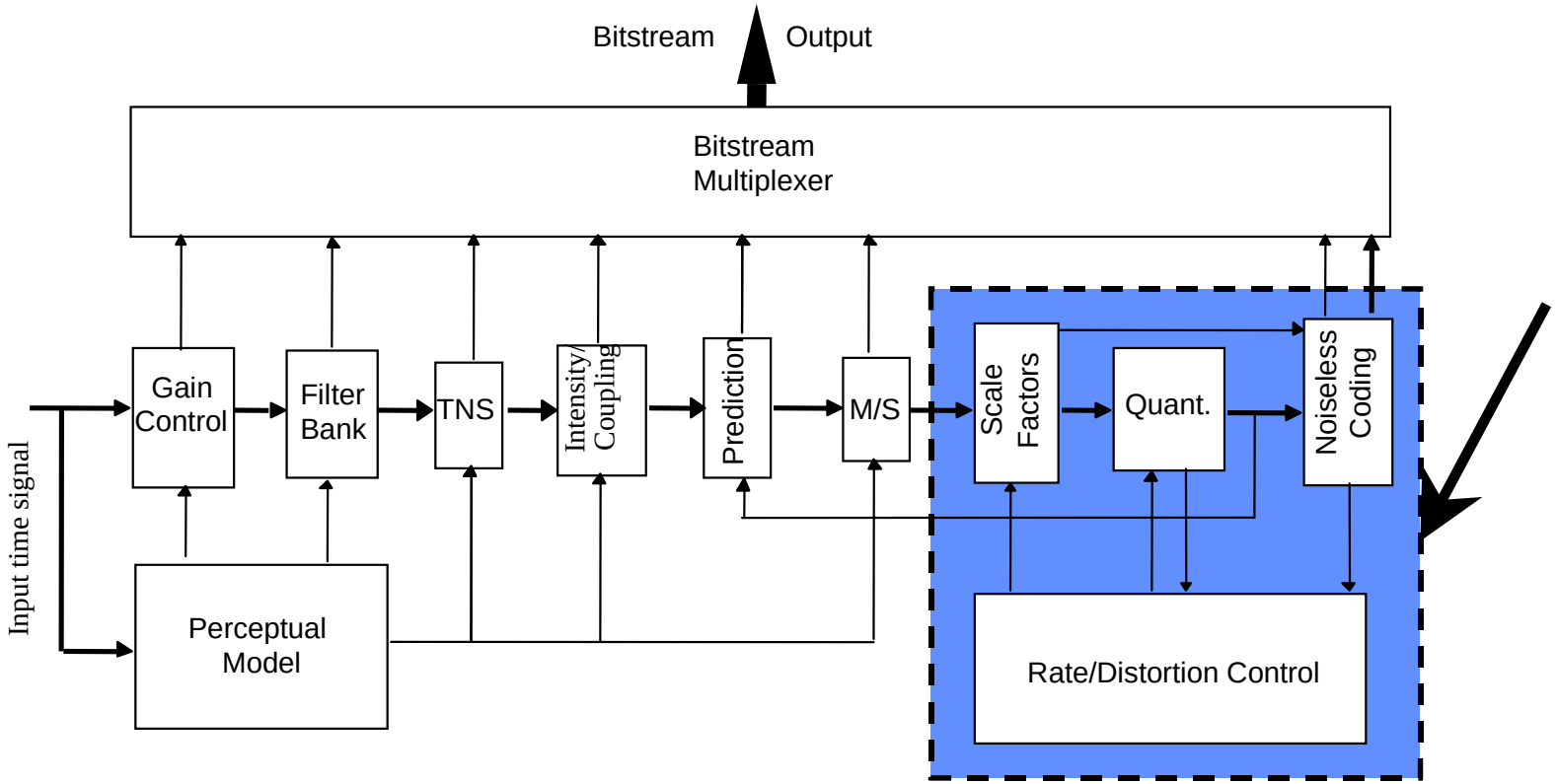
MPEG-2 AAC:

- Prediction of each spectral coefficient with backward adaptive predictor
- High complexity (ca. 50% of decoder computation & RAM)

MPEG-4:

- Long Term Predictor (LTP) as known from speech coding
- Lower complexity: Saving of approx. 50% in terms of computation and memory over MPEG-2 predictors
- Comparable performance to MPEG-2 predictors

Extension 3: Twin-VQ



Transform-Domain Weighted Interleave VQ (2)

Background:

- Audio coding at extremely low bitrates (6-8 kbit/s)
- CELP speech coders do not perform well for music
- 0.5 Bits per frequency line at these data rates !!

MPEG-4:

- *Transform-Domain Weighted Interleave Vector Quantization (TwinVQ)* as alternative coding kernel
  - Vector selection under control of the perceptual model
- Fully integrated into MPEG-4 AAC coding system:
  - Uses same spectral representation as AAC coder
  - Makes use of other MPEG-4 tools (e.g. LTP, TNS, joint stereo coding)

Fraunhofer Institut Integrierte Schaltungen

Transform-Domain Weighted Interleave VQ (3)

Structure:

- Normalization of spectral coefficients:
  - LPC envelope (overall spectral shape)
  - Periodic component coding (harmonic components)
  - Bark-scale envelope coding (additional flattening)

- Vector Quantization (VQ) process:
  - Interleaving of spectral coefficients into new sub-vectors
  - Vector quantization
    (two sets of codebooks, weighted distortion measure allows distortion control by perceptual model)

$\Rightarrow$ no bit/noise allocation or rate control iteration

Scalability

Definition:

- **Capability to decode useful sub-sets of the bitstream**

Types of Scalability:

- **SNR / NMR (Noise to Mask Ratio) Scalability:**
  - "Extension layers improve the SNR/NMR of the coded signal"
- **Audio Bandwidth Scalability:**
  - "Extension layers increase the decodable audio band width"
- **Restriction of Generality:**
  - Very low bit rate core coder optimized for special signals, e.g speech. Additional layers provide good quality for all types of signals.
- **Implementation Complexity:**

Application examples

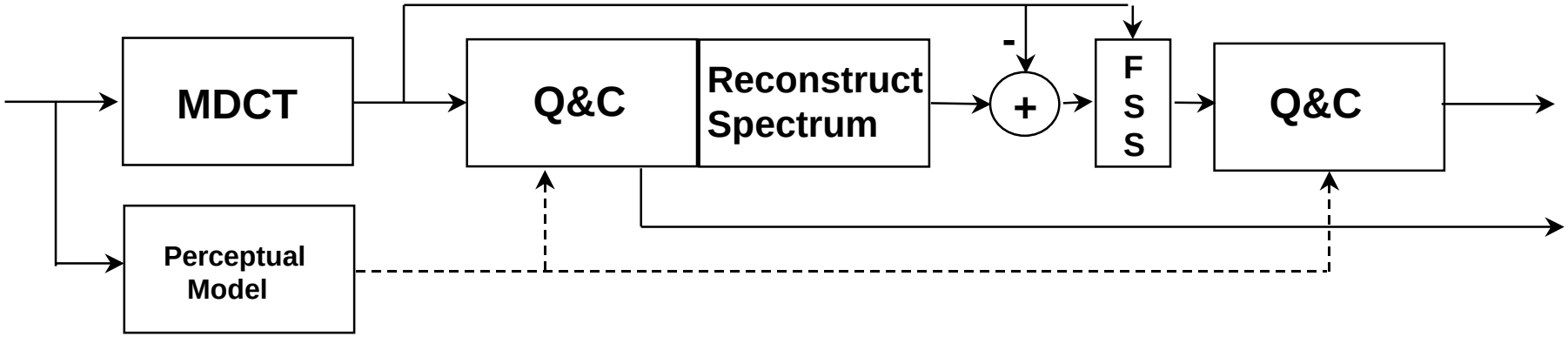
- **Network based (packetized) transmission**
  - Requires routers which know about the importance of a packet
  - Less important (outer layer) packets may be dropped if the available bandwidth decreases

- **Broadcast**
  - The most important (inner layer) packets are transmitted with a better error protection scheme

- **Music data base**
  - High quality content is encoded and stored
  - Access to a lower quality version is possible without recoding to allow for pre-listening with a lower quality

Scalable GA Coder (I)

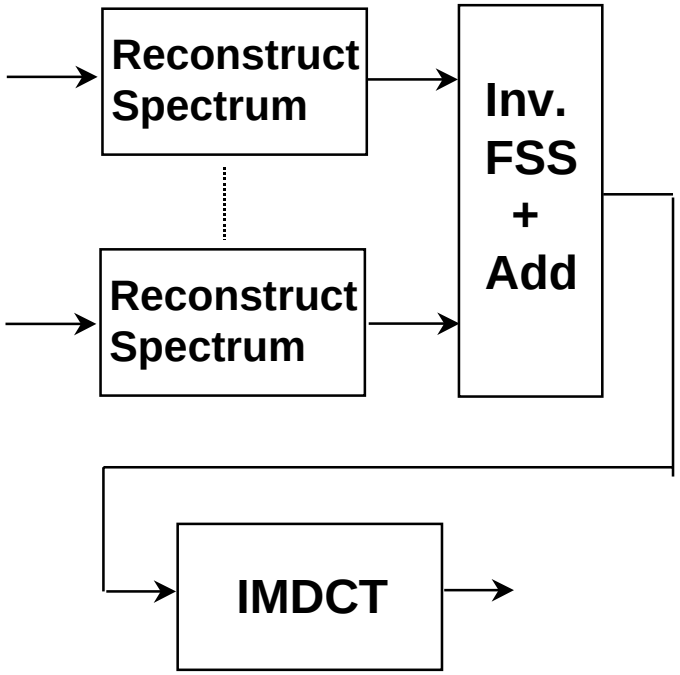
Encoder Block Diagram



- Encoding of the error signal of an AAC or Twin-VQ Quantization and Coding (Q&C) module in a second, or third, or n-th similar quantization module in the frequency domain
- Solutions using only AAC, or only Twin-VQ modules possible
- Additionally, Twin-VQ / AAC combinations defined
- Useful for large enhancement steps of >= 8 kbit/s per step

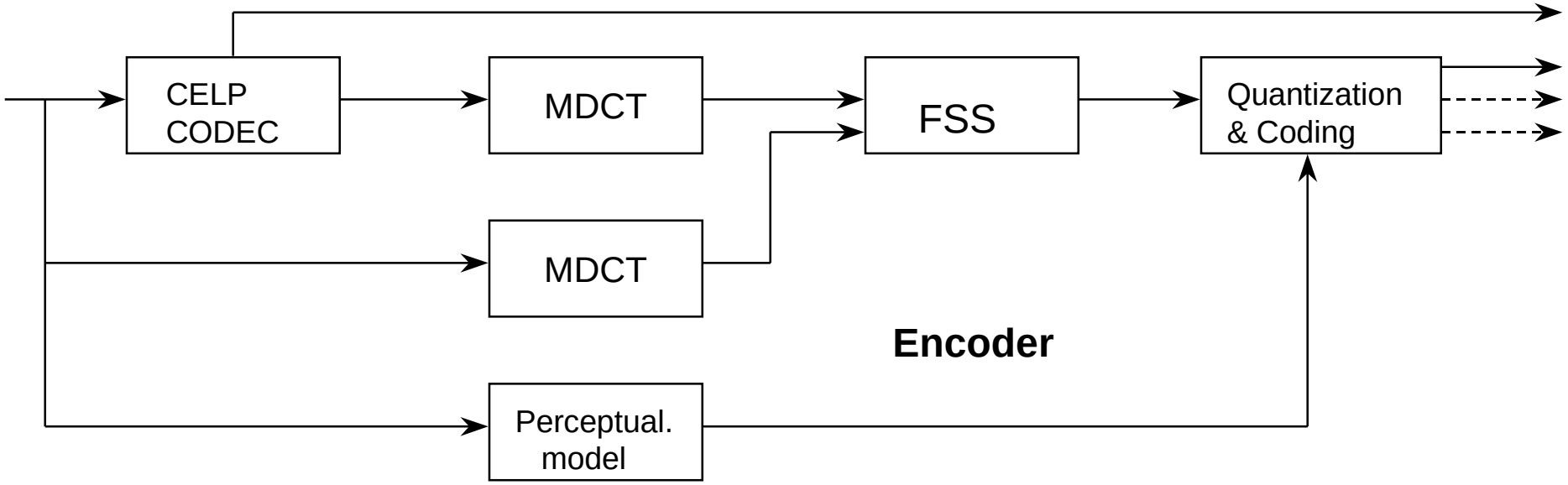
Scalable GA Coder (II)

Decoder Block Diagram



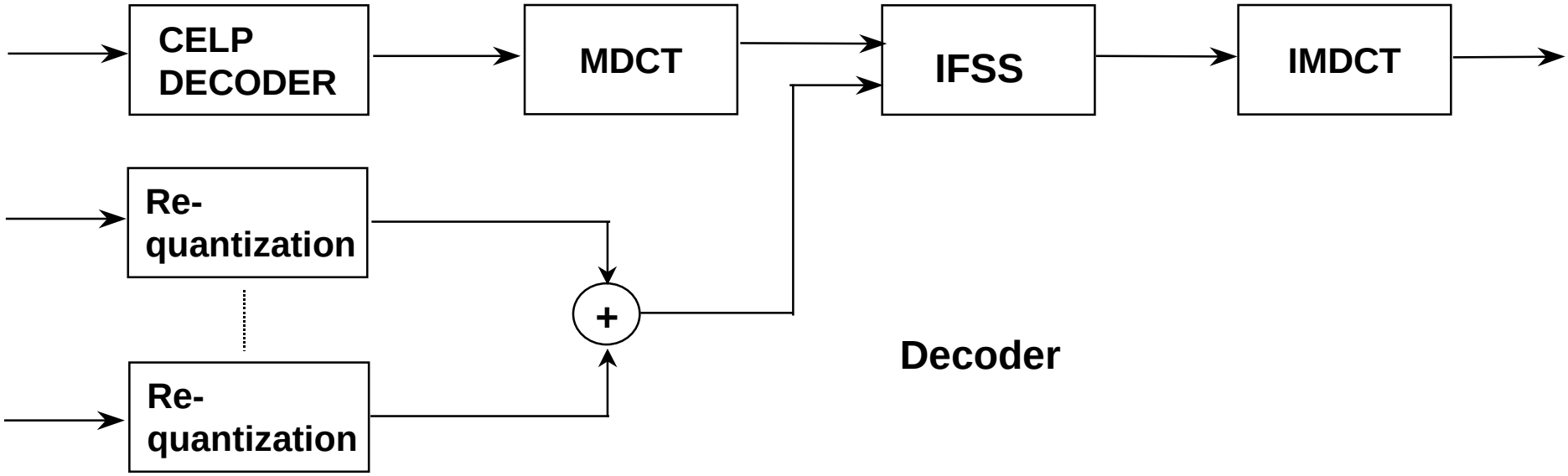
- **Twin-VQ Q&C Modules**
  - 8 kbit/s fixed step size Vector Quantizer (VQ) modules
  - optional 6 kbit/s in first layer
  - first choice for a 6 or 8 kbit/s base layer for the coding of general audio signals
- **AAC Q&C modules**
  - Any step size possible
  - Reasonable step sizes from 8 to >64 kbit/s
  - The same end quality can be achieved as from a single step AAC coder
  - However, a higher bit rate may be required for the same audio quality

Scalable GA Coder : Combination with CELP Coder (I)

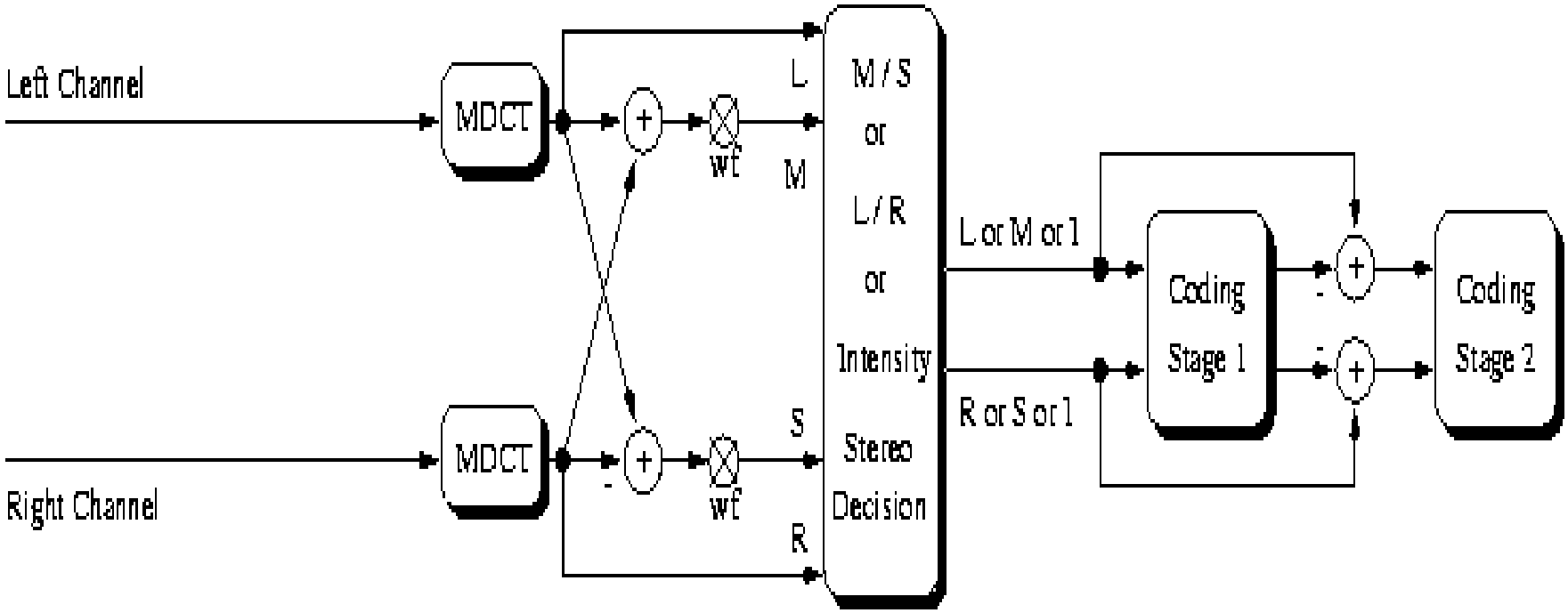


- Very low bitrate core coder (e.g. speech coder)
- Core coder typically operating at a lower sampling frequency
- MDCT used for efficient up-sampling

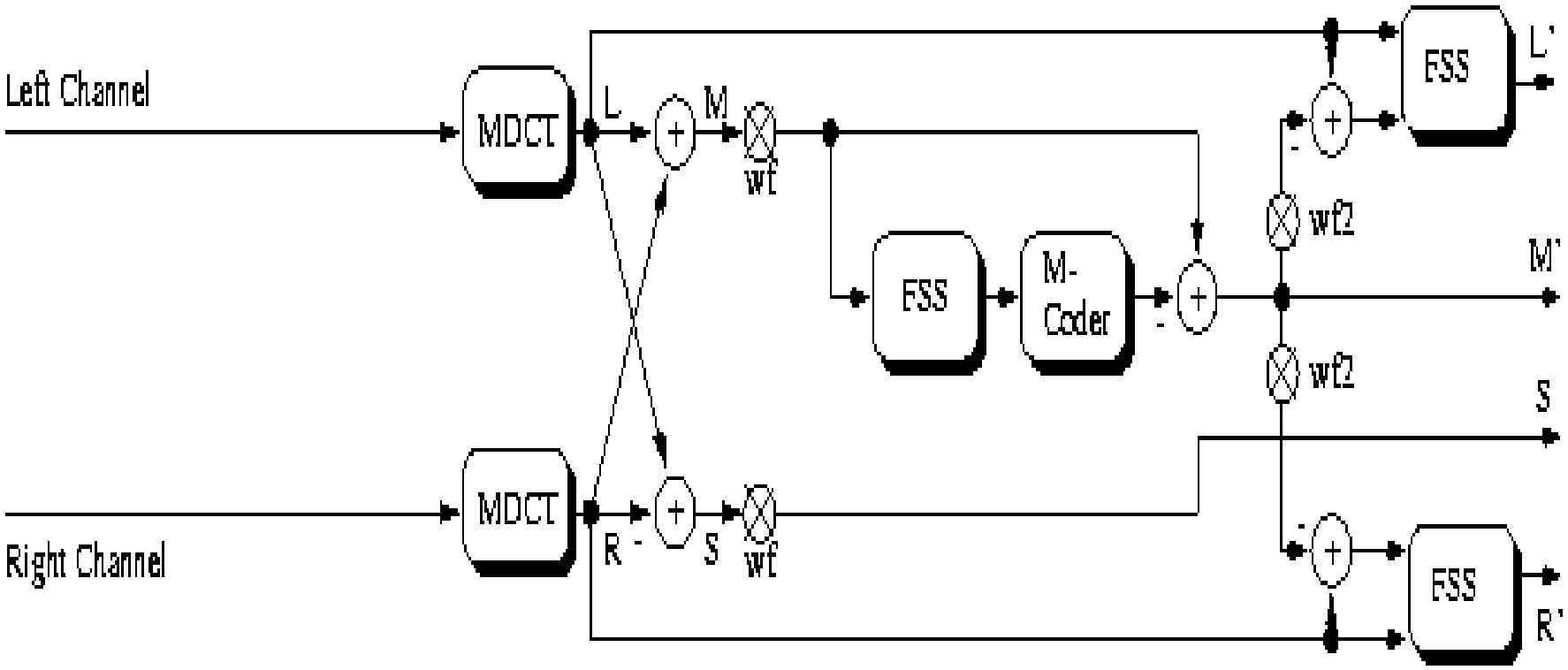
Scalable GA Coder : Combination with Core Coder (II)



Scalable Stereo Coding: Stereo / Stereo

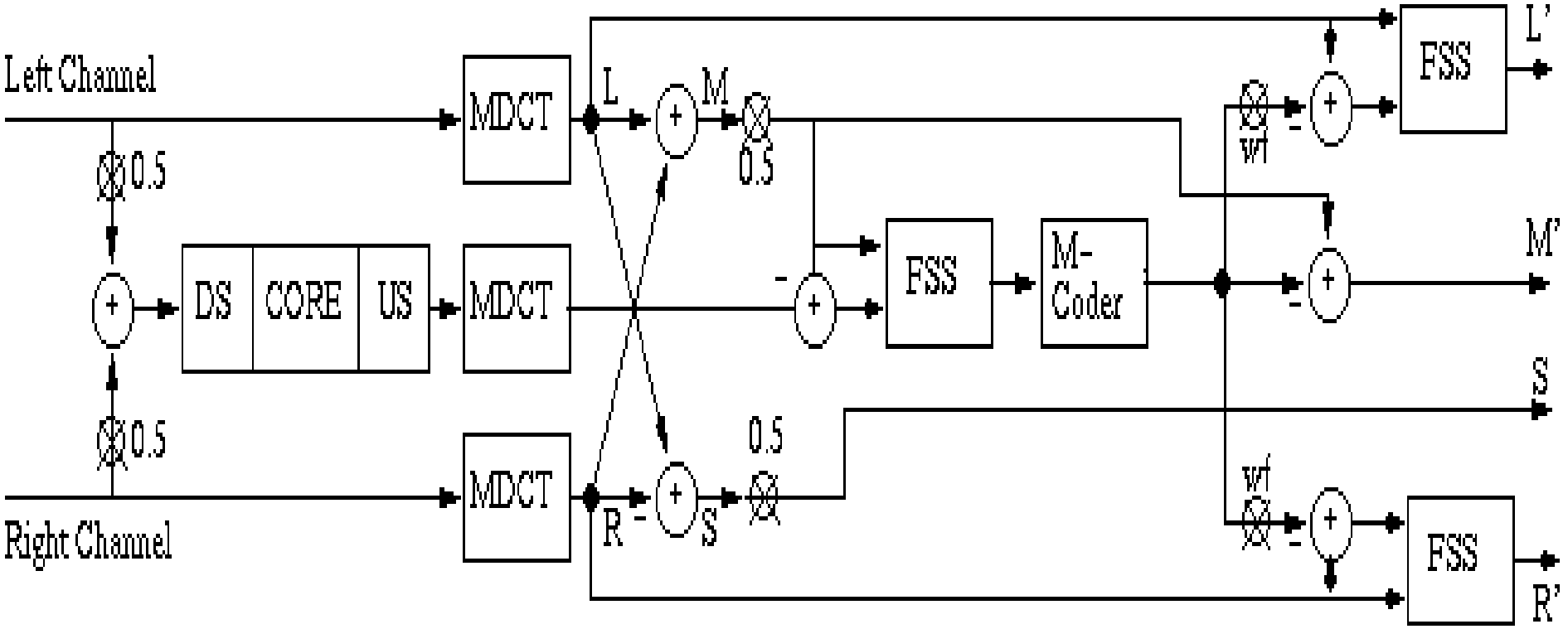


Scalable Stereo Coding: Mono / Stereo



Fraunhofer Institut
Integrierte Schaltungen

Scalable Stereo Coding: Mono Core / Mono GA / Stereo GA



Scalable GA Coder : Typical Configurations

- Some successfully tested mono/mono combinations:

  6 kbit/s CELP + 18 kbit/s AAC
  6 kbit/s TwinVQ + 18 kbit/s AAC
  8 kbit/s TwinVQ + 8 kbit/s TwinVQ

  6 kbit/s CELP + 18 kbit/s + 24 kbit/s AAC

- Mono/stereo combinations

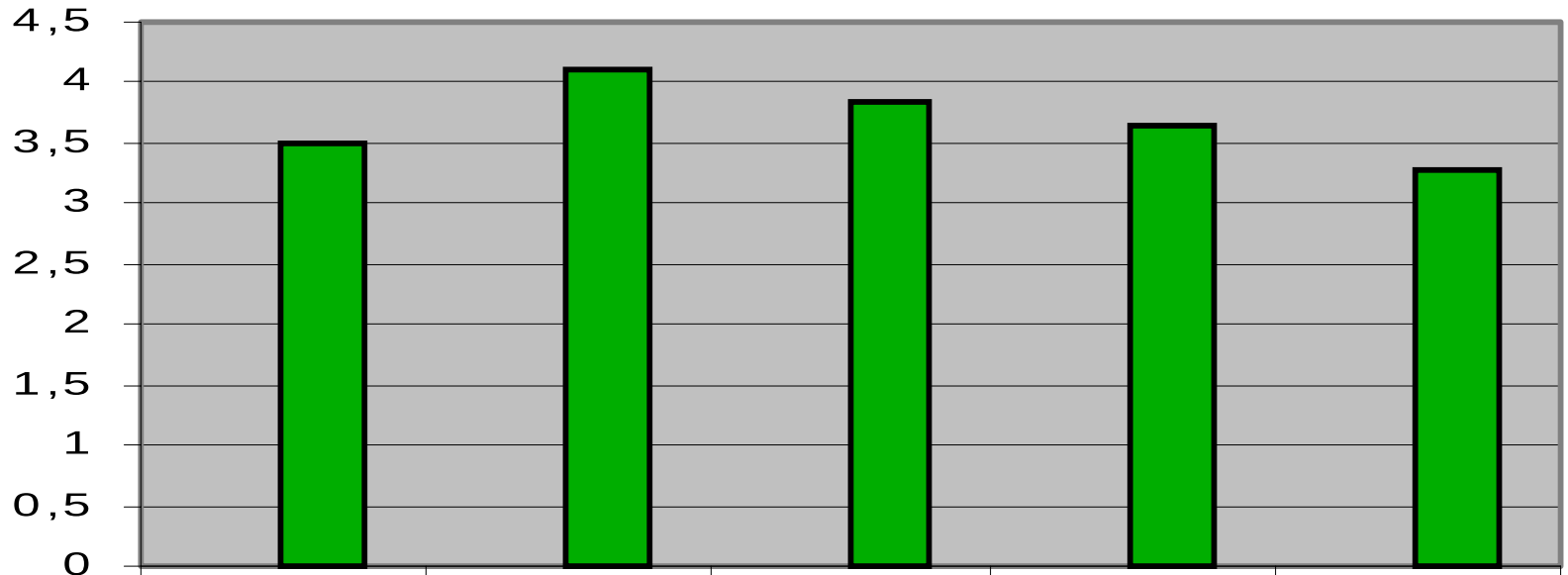
  6 kbit/s mono CELP + 18 kbit/s mono + 24 kbit/s stereo AAC
  24 kbit/s mono + 16 kbit/s stereo + 16 kbit/s stereo AAC
  24 kbit/s mono + 72 kbit/s stereo AAC

- Stereo/stereo combinations

  2 x 6 kbit/s mono CELP + 36 kbit/s stereo AAC

Fraunhofer Institut Integrierte Schaltungen

Results (I) Mono Configurations



	Layer 3 24 kbit/s	AAC 24 kbit/s	CELP+AAC 24 kbit/s	Tw inVQ + AAC 24 kbit/s	AAC 18 kbit/s
Series1					
Series2	3,5	4,1	3,85	3,65	3,27

Fraunhofer Institut Integrierte Schaltungen

Results (II) Mono / Stereo Configuration

Conclusions

- Highest quality coding with proven AAC technology
- PNS, LTP and TwinVQ further enhance the very low bitrate performance
- Mono, Stereo, and Multi-channel Stereo supported
- Bitrate range 6 - ~300 kbit/s per channel at 8 - 96 kHz SR
- Additional flexibility with the scalable coding modes
  - Unique capabilities through the availability of the mono-stereo coding modes
- Overall complexity within the limits of today's hardware

- ==>
- The MPEG-4 GA coder the most versatile audio coding system available today
- Low-Delay and Error Resilience Additions in MPEG-4 Version 2

Fraunhofer Institut Integrierte Schaltungen