MPEG-4 General Audio Coding

Jürgen Herre

Fraunhofer Institute for Integrated Circuits (IIS)

Fraunhofer Institut Integrierte Schaltungen

Outline

MPEG-4 General Audio Coding:
The “all-round coder” in MPEG-4 audio

- MPEG-2 Advanced Audio Coding (AAC)
- MPEG-4 Extensions:
  - Perceptual Noise Substitution (PNS)
  - Long Term Prediction
  - TwinVQ Coding Core
- Conclusions

MPEG-2 Advanced Audio Coding (AAC)

History:

- 1994: Official start of AAC development
  Goal: Development of a new powerful state-of-the art multi-channel coder without compatibility constraints
- 1997: AAC International standard (IS)
- 1999: AAC part of the MPEG-4 Standard

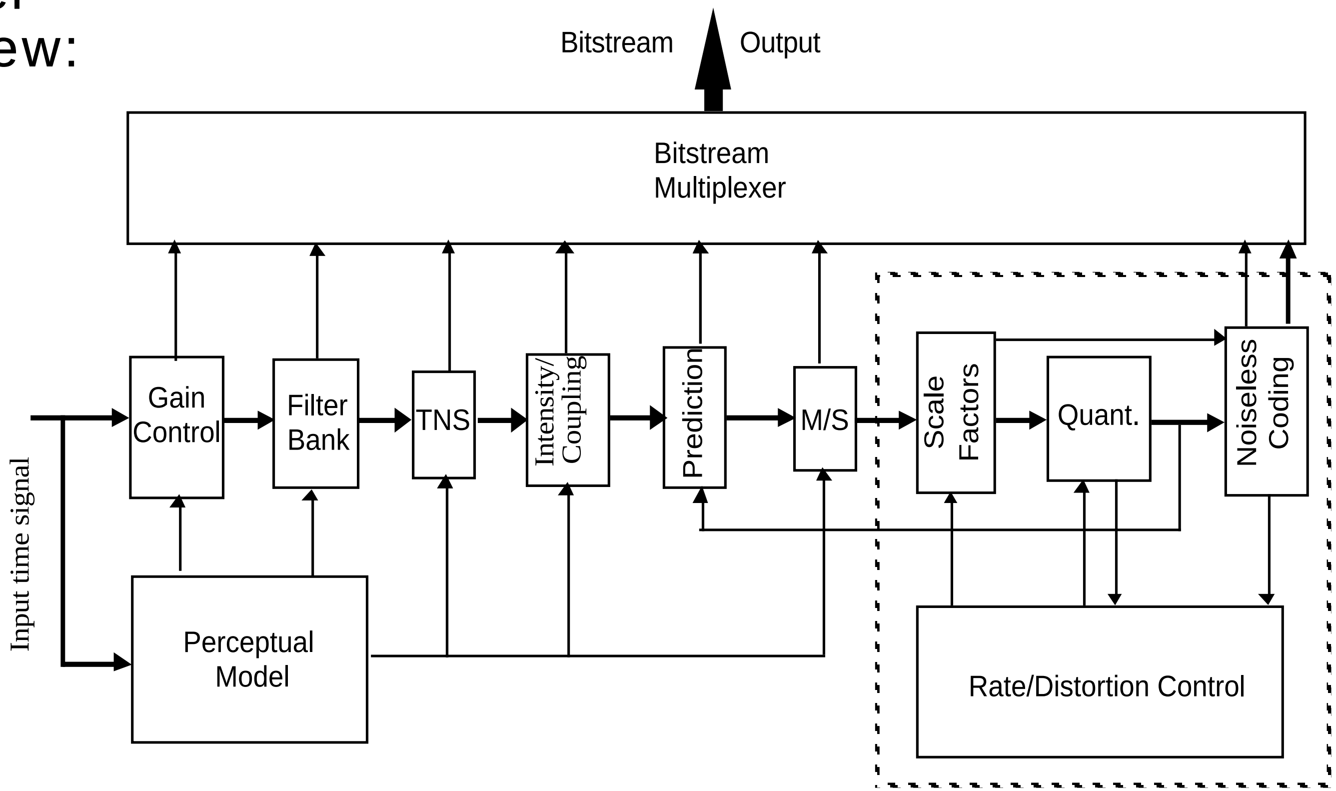
Today:

- Favourite coder for many application areas, like Internet audio, solid state players, ISDN music transmission, High-definition TV (HDTV), satellite & terrestrial digital audio broadcasting, ...

Fraunhofer Institut
Integrierte Schaltungen

MPEG-2 AAC (2)

Encoder Overview:



Fraunhofer Institut Integrierte Schaltungen

MPEG-2 AAC (3)

Basic features:

- High frequency resolution filterbank-based coder (1024 lines MDCT with 50% overlap)
- 1:8 block switching (8 *128 lines MDCT)
- Non-uniform quantizer
- Noise shaping in half critical bands (scalefactor bands)
- Huffman coding of scalefactors and spectral coefficients

Fraunhofer Institut
Integrierte Schaltungen

MPEG-2 AAC (4)

Advanced coding tools:

- Window shape adaptation
- Temporal noise shaping (TNS)
- Gain control (SRS profile only)
- Backward adaptive prediction

Joint stereo coding tools:

- Mid/Side stereo (MS) per scalefactor band
- Intensity stereo coding between channel pairs
- Coupling channel(s)

Other features:

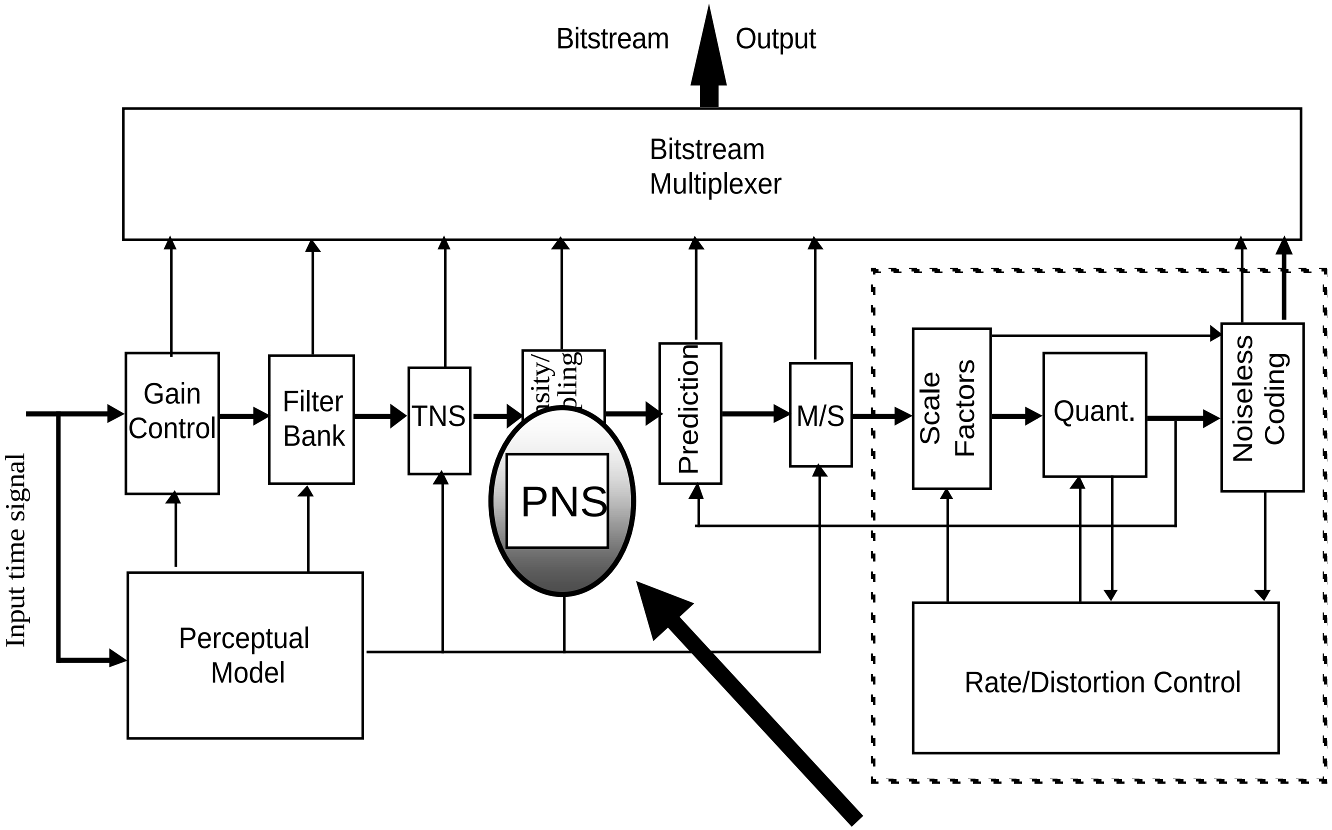
- Flexible bitstream format for up to 48 channels
- Low Frequency Enhancement (LFE) channel(s)

MPEG-2 AAC Performance

Test Results:

- Broadcast quality at 320 kbit/s for 5 channels (better than MPEG-2 Layer II at 640 kbit/s)
- Broadcast quality at 128 kbit/s stereo
- Comparison to other codecs:
  AAC 96 kbit/sec stereo comparable to
  - AC-3 at 160 kbit/s
  - Layer II at 192 kbit/s
  - Layer III at 128 kbit/s
- Very low bitrates (comparison within MPEG):
  AAC best audio coder at bitrates down to 16 kbit/s for mono and stereo

Extension: Perceptual Noise Substitution (PNS)



Perceptual Noise Substitution (1)

Background:

- Parametric coding of signals gives a very compact signal representation
- Parametric coding of noise-like signal components has been used widely e.g. in speech coding
- Can similar techniques be used in perceptual audio coding ?

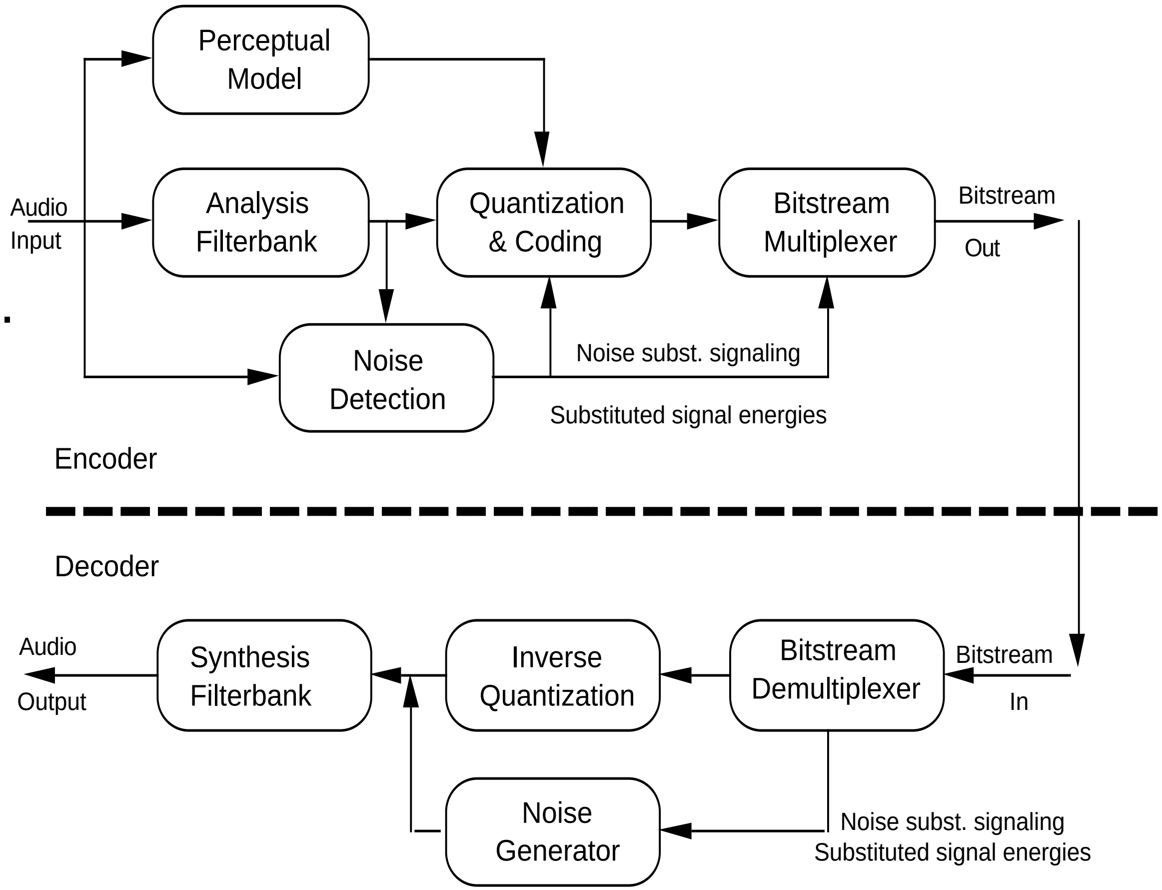
MPEG-4:

- Perceptual Noise Substitution (PNS) permits a frequency selective parametric coding of noise-like signal components

Fraunhofer Institut
Integrierte Schaltungen

Perceptual Noise Substitution (2)

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



Fraunhofer Institut Integrierte Schaltungen

Perceptual Noise Substitution (3)

Principle:

- 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

$\Rightarrow$ Highly compact representation for noise-like spectral components

Fraunhofer Institut
Integrierte Schaltungen

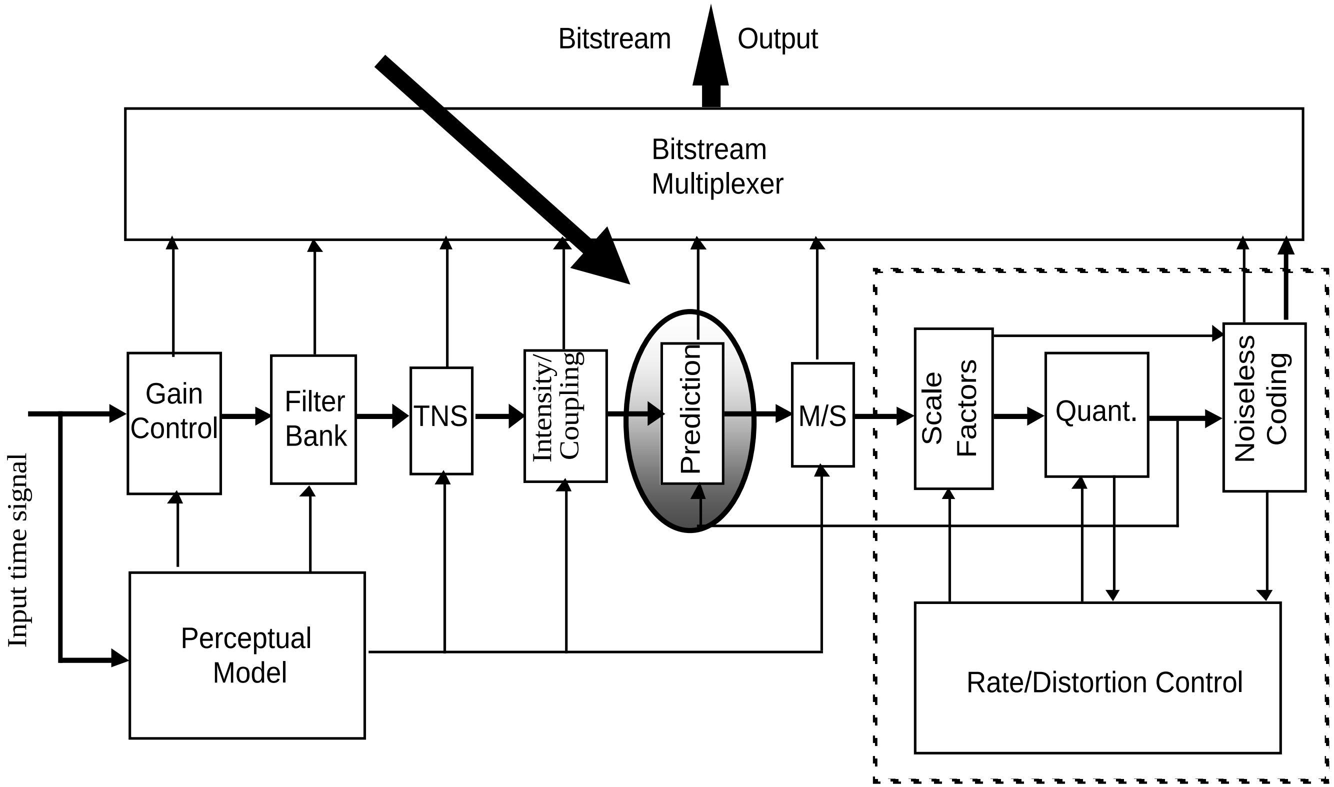
Demonstration: Perceptual Noise Substitution

- Suzan Vega / Tom's Diner:
  w/o PNS   with PNS   PNS parts only
- Orchestral Music :
  w/o PNS   with PNS   PNS parts only
- Contemporary Pop Spot:
  w/o PNS   with PNS   PNS parts only

AAC coder @ 32 kbit/s, fs=48 kHz, bandwidth 14 kHz

Note: This is not "normal" operation mode

Extension: Long Term Prediction



Fraunhofer Institut
Integrierte Schaltungen

Long Term Prediction (1)

Background:

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

$\Rightarrow$ High bit rate necessary to code signals with many tonal components (e.g. Harpsichord, Pitch Pipe)

- Tonal signal components are <u>predictable</u>

$\Rightarrow$ Further quality enhancement by <u>predictive coding</u>

Fraunhofer Institut
Integrierte Schaltungen

Long Term Prediction (2)

MPEG-2 AAC:

- Prediction of each spectral coefficient; backward adaptive 2nd order Lattice predictor
- High complexity (ca. 50% of decoder computation & RAM)

  ⇒ Prohibitive for cost sensitive applications, used in "MPEG-2 Main Profile" only

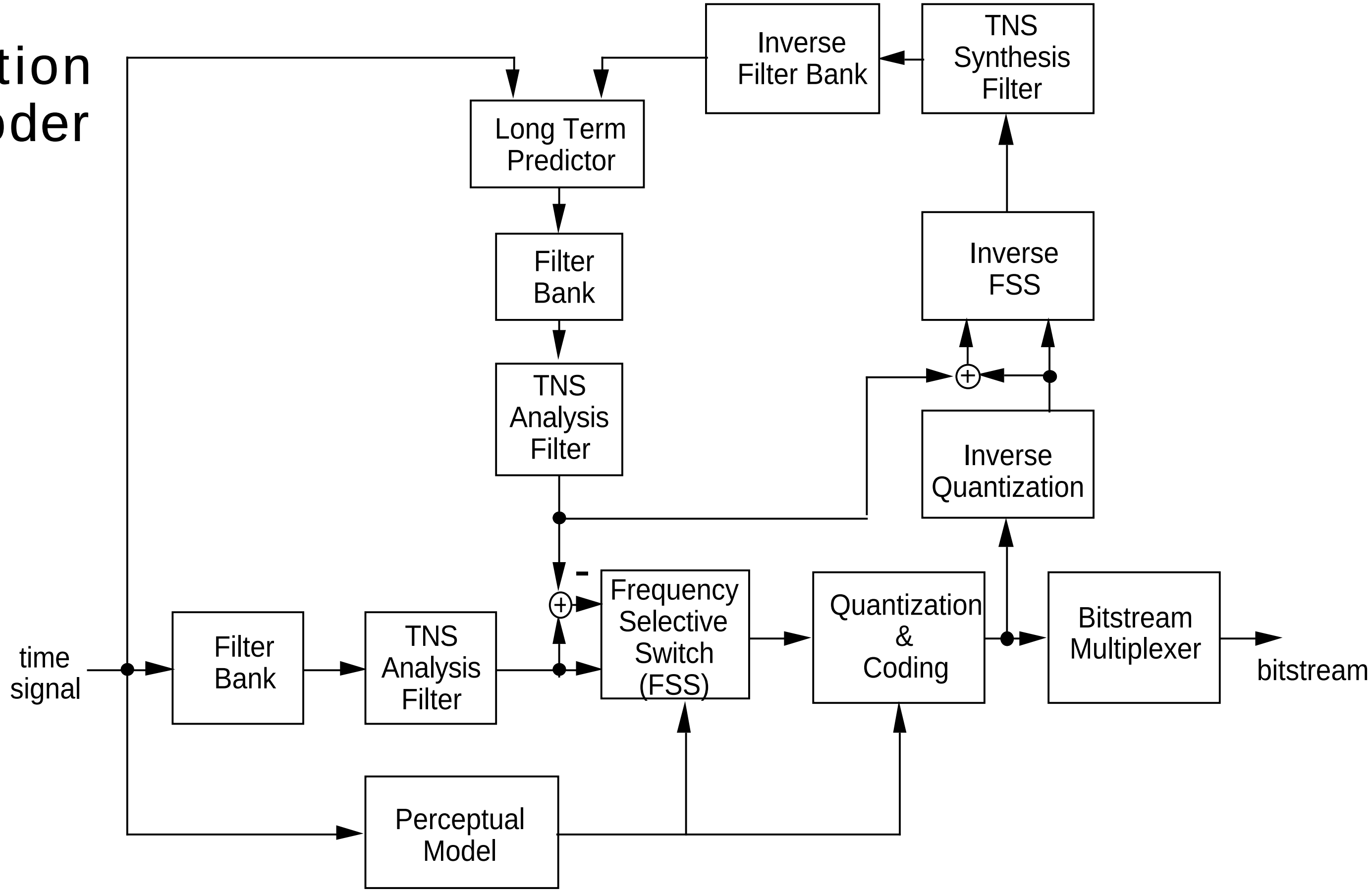
MPEG-4 AAC:

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

Fraunhofer Institut Integrierte Schaltungen

Long Term Prediction (3)

LTP Integration into AAC coder



IIS
Fraunhofer Institut Integrierte Schaltungen

Transform-Domain Weighted Interleave VQ (1)

Background:

- Audio coding at extremely low bitrates (≥ 6 kbit/s)
- CELP speech coders do not perform well for music

MPEG-4:

- *Transform-Domain Weighted Interleave Vector Quantization (TwinVQ)* as additional coding kernel
- 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)
  - Possible core coder for ⇒ MPEG-4 scalable coding

Fraunhofer Institut
Integrierte Schaltungen

Transform-Domain Weighted Interleave VQ (2)

Structure:

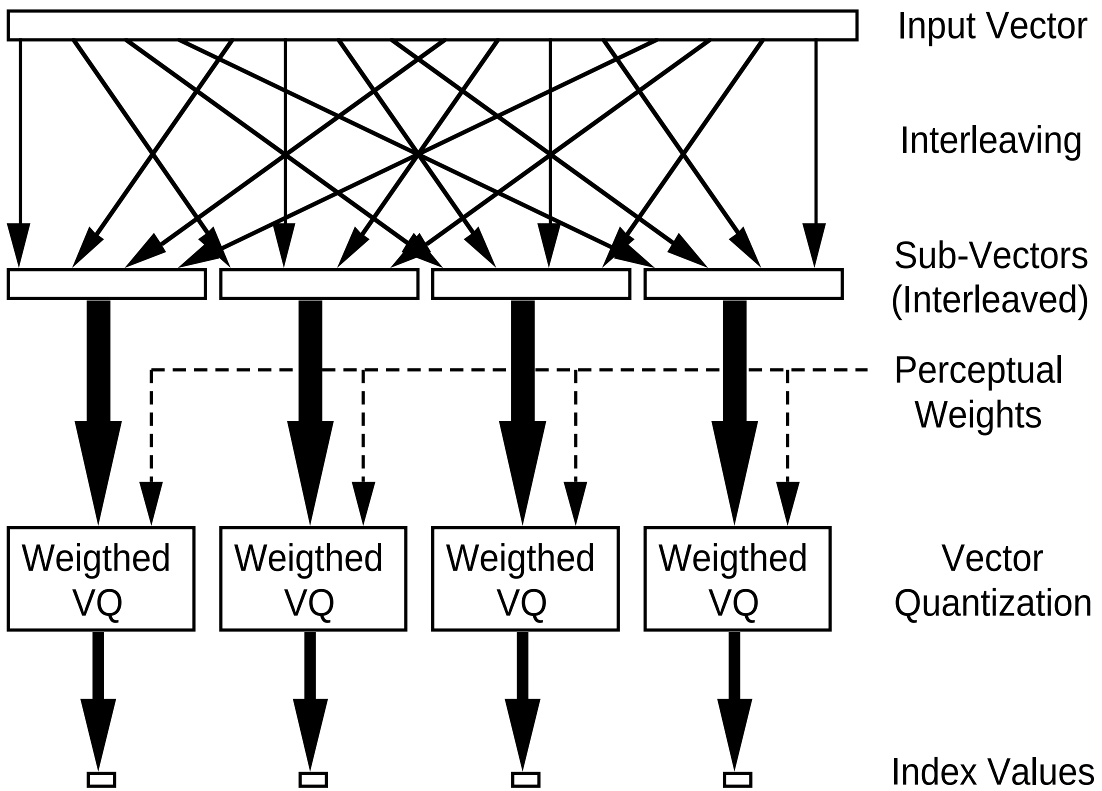
- Normalization of spectral coefficients:
  - LPC envelope (overall spectral shape)
  - Periodic component coding (harmonic components)
  - Bark-scale envelope coding (additional flattening)
- Vector Quantization (VQ) process:
  - Interleaving of spectral coefficients into new sub-vectors
  - Vector quantization
    (two sets of codebooks, weighted distortion measure allows distortion control by perceptual model)

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

IIS
Fraunhofer Institut Integrierte Schaltungen

Transform-Domain Weighted Interleave VQ (3)

Interleave VQ Structure:



Fraunhofer Institut Integrierte Schaltungen

Conclusions

- MPEG-4 General Audio Coding:
  The “all-round” coding system among the MPEG-4 audio schemes, providing a set powerful tools
- Based on MPEG-2 Advanced Audio Coding kernel
- Several enhancements for improved coding efficiency
- Perceptual Noise Substitution (PNS):
  Exploiting noise-like components in the signal
- Long Term Prediction (LTP):
  Taking advantage of very stationary / tonal signals
- TwinVQ coder kernel supports audio coding at extremely low data rates ($\Rightarrow$ MPEG-4 scalable coding)

Fraunhofer Institut Integrierte Schaltungen