Prediction and Lossless Audio Coding

Prof. Dr.-Ing. Gerald Schuller

Ilmenau Technical University & Fraunhofer IDMT
Ilmenau, Germany

Use of Redundancy (1)

- For higher correlation between samples ! → higher redundancy
- For „flat" PSD → low redundancy
- ACF (Auto Correlation Function) r_{xx}:

Continuous time:

$$r_{XX}(\tau) = \lim_{T \to \infty} \frac{1}{2\mathrm{T}} \int_{-T}^{T} x(t)\,x(t+\tau)\,dt = E\{x(t)\,x(t+\tau)\}$$

Discrete time:

$$r_{xx}(k) = \sum_{n} x(n)\,x(n+k) = E(x(n)\,x(n+k))$$

Observe:

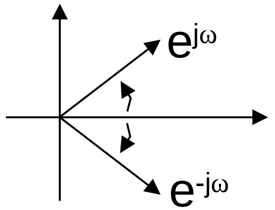
Correlation is convolution with the signal and its time reversed version:

In DFT domain: Multipliplication with its conjugate complex version.

- PSD (Power Spectrum Density):

$$r_{XX}(\tau) \quad \circ\!\!-\!\!\bullet \quad S_{XX}(f) = \int_{-\infty}^{\infty} r_{XX}(\tau)\,e^{-j2\pi f\tau}\,d\tau$$

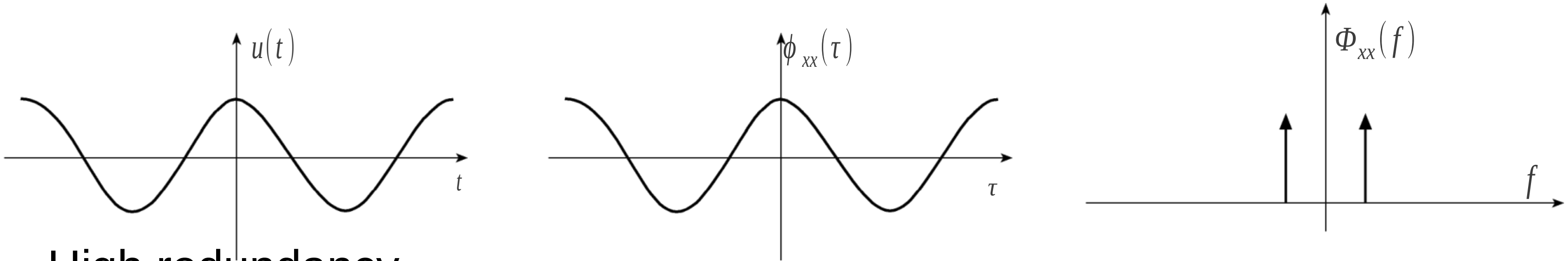
$$S_X(f) \cdot \bar{S}_X(f) = |S_X(f)|^2$$



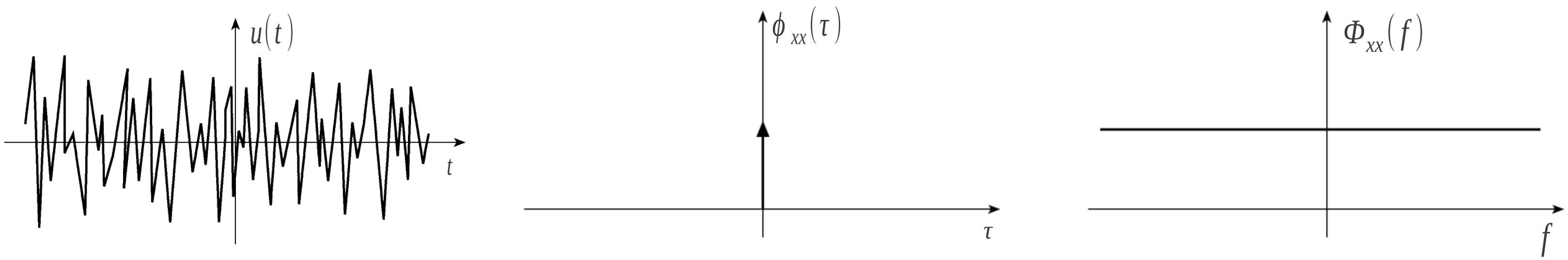
conjugate complex

Use of Redundancy (2)

Signal ACF PSD



High redundancy

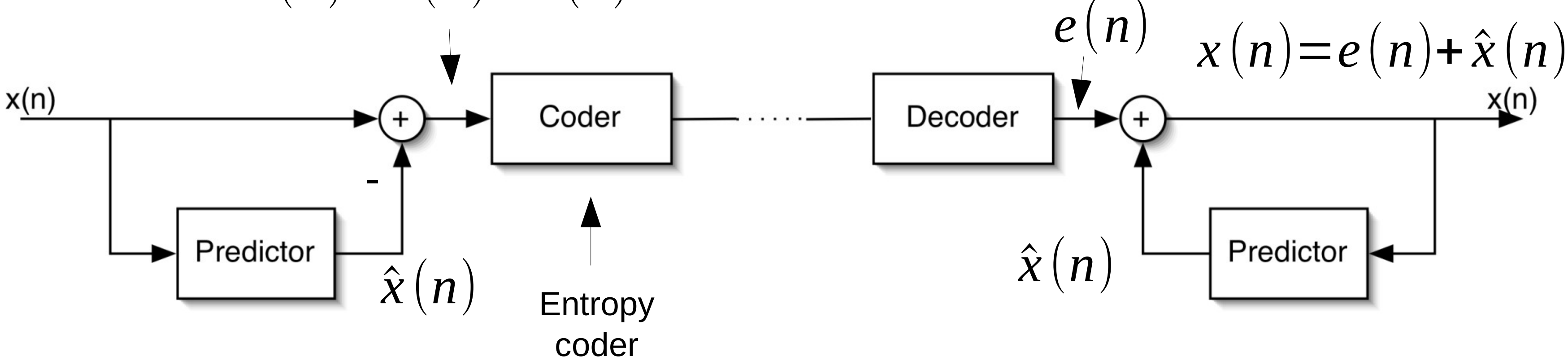


Low redundancy

Predictive Coding

- Use of the correlation of nearby samples
- Method:
  - Prediction of the current sample, using past samples
  - Transmission of the smaller prediction error (smaller code word)

$$e(n) = x(n) - \hat{x}(n)$$



Predictive Coding

- Encoder

prediction error, to be encoded

$$e(n) = x(n) - \hat{x}(n)$$

$$\hat{x}(n) = \sum_{j=1}^{N} h_j \cdot x(n-j)$$

weighted sum of past values

predicted value

predictor- or filter- coefficients

- Decoder receives e(n),

$$x(n) = e(n) + \sum_{j=1}^{N} h_j \cdot x(n-j)$$

error power

- Goal: Minimize the mean squared error $\sigma_e^2 = E\{e^2(n)\}$ by optimizing the filter coefficients h_j

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Predictive Coding

- Approach: $\frac{\partial \sigma_e^2}{\partial h_j} \overset{!}{=} 0$ Find zero of first derivative with respect to filter the coefficients

$$\sigma_e^2 = E\{(x(n) - \hat{x}(n))^2\}$$

$$\frac{\partial \sigma_e^2}{\partial h_j} = 2E\{(x(n) - \hat{x}(n))\, x(n-j)\},\ j = 1, \ldots, N$$

$$\Rightarrow 0 \overset{!}{=} E\{(x(n) - \hat{x}(n))\, x(n-j)\},\ j = 1, \ldots, N$$

Remember:

$$r_{xx}(k) = E(x(n)\, x(n+k))$$

$$\Rightarrow 0 = r_{xx}(k) - \sum_{j=1}^{N} h_j r_{xx}(k-j),\ k = 0, \ldots, N$$

$$\Rightarrow r_{xx}(k) = \sum_{j=1}^{N} h_j r_{xx}(k-j)$$

Predictive Coding

- With the auto correlation matrix:

$$\underline{\underline{R}}_{XX} = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(N-1) \\ r_{xx}(1) & r_{xx}(0) & & r_{xx}(N-2) \\ \vdots & & \ddots & \vdots \\ r_{xx}(N-1) & r_{xx}(N-2) & \cdots & r_{xx}(0) \end{bmatrix}$$

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Wiener-Hopf-Equation

- We obtain the Wiener-Hopf-Equation in matrix description

$$r_{XX}(k) = \sum_{j=1}^{N} h_j \cdot r_{XX}(k-j)$$

$$\begin{bmatrix} r_{XX}(1) \\ \vdots \\ r_{XX}(N) \end{bmatrix} = \begin{bmatrix} r_{XX}(0) & \cdots & r_{XX}(N-1) \\ \vdots & . & . \\ r_{XX}(N-1) & . & r_{XX}(0) \end{bmatrix} \cdot \underline{h}_{opt}$$

$$\underline{r}_{XX} = \underline{\underline{R}}_{XX} \cdot \underline{h}_{opt}$$

- Vector of optimum filter coefficients:

$$\underline{h}_{opt} = \underline{R}_{XX}^{-1} r_{XX}$$

Reference: Monson H. Hayes: “Statistical Digital Signal Processing and Modelling”, Wiley & Sons.

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Prediction, Orthogonality Principle

- **orthogonality principle**: The prediction error is uncorrelated to the signal (otherwise the prediction could be better) (https://en.wikipedia.org/wiki/Orthogonality_principle)
  - The pred. **error** and the N **past signal** samples are **uncorrelated**, if we have the **optimum prediction coefficients**!

$$E\left(e(n)\cdot x(n-j)\right)=0,\ j=1,\ldots,N$$

- The predicted signal is a linear combination of the past N input samples,

$$\hat{x}(n)=\sum_{j=1}^{N} h_j\cdot x(n-j)$$

- Hence we also get: the predicted signal and the prediction error are uncorrelated,

$$E\left(e(n)\cdot\hat{x}(n)\right)=0$$

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Deriving Wiener-Hopf with Pseudo Inverses (1)

- Input matrix X:

$$\underline{\underline{X}}=\begin{bmatrix} x(N-1) & x(N-2) & \cdots & x(0) \\ x(N) & x(N-1) & & x(1) \\ \vdots & & \ddots & \vdots \\ x(B-1) & x(B-2) & \cdots & x(B-N) \end{bmatrix} \qquad \underline{d}=\begin{bmatrix} x(N) \\ x(N+1) \\ \vdots \\ x(B) \end{bmatrix}$$

- B is a block length for the computation,
- The matrix multiplication $\underline{\underline{X}}\cdot h$ implements the convolution of the predictor
- Solve equation as close as possible to „*d*“ as our desired signal, in a quadratic sense (minimize sum of quadratic error):

$$\underline{\underline{X}}\cdot\underline{h}\approx\underline{d}$$

more equations than unknowns →

Sequence of “next” values

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Deriving Wiener-Hopf with Pseudo Inverses (2)

- Solving the matrix equation with “Moore-Penrose pseudo inverse”,

quadratic matrix $\longrightarrow$ $\left(\underline{\underline{X}}^T \underline{\underline{X}}\right) \cdot \underline{h} = \underline{\underline{X}}^T \cdot \underline{d}$

h which approximates d in quadratic error sense $\longrightarrow$ $\underline{h} = \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \cdot \underline{d}$

$\left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1}$ Converges to autocorr. matr. Inv., $\rightarrow \underline{\underline{R}}_{xx}^{-1}$

$\underline{\underline{X}}^T \cdot \underline{d}$ Cross correlation vector, $\rightarrow \underline{r}_{xx}$

- This results in the Wiener-Hopf-Equation for block size $B \rightarrow \infty$

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Coding Gain

- The prediction error variance/power is

$$\sigma_e^2 = E\left\{\left(x(n)-\hat{x}(n)\right)^2\right\} = E\left\{x^2(n)+\hat{x}^2(n)-2\mathrm{x}(n)\hat{x}(n)\right\}$$

Using the decoder reconstruction equation:

$$x(n)=\hat{x}(n)+e(n)$$

we obtain:

$$\rightarrow E\left(\hat{x}^2(n)\right)=E\left(\hat{x}(n)\cdot\left(x(n)-e(n)\right)\right)=E\left(\hat{x}(n)\cdot x(n)-\hat{x}(n)\cdot e(n)\right)$$

- using the orthogonality principle: $E\left(\hat{x}(n)\cdot e(n)\right)=0$,

we get the substitution $\rightarrow E\left(\hat{x}^2(n)\right)=E\left(\hat{x}(n)\cdot x(n)\right)$

And we can reformulate $\rightarrow \sigma_e^2 = E\left(x^2(n)+\hat{x}^2(n)-2\,x(n)\cdot\hat{x}(n)\right)=$

$$=E\left(x^2(n)\right)+E\left(\hat{x}^2(n)\right)-2\,E\left(x(n)\cdot\hat{x}(n)\right)$$

to $\sigma_e^2 = E\left(x^2(n)\right)-E\left(x(n)\cdot\hat{x}(n)\right)$

Coding Gain

- Now we have

$$\sigma_e^2 = E\left(x^2(n)\right) - E\left(x(n)\cdot\hat{x}(n)\right)$$

And we see that the first term is the signal power,

$$E\left(x^2(n)\right) = \sigma_x^2$$

The second term is

$$E\left(x(n)\cdot\hat{x}(n)\right) = E\left(x(n)\cdot\sum_{j=1}^{N} h_j\cdot x(n-j)\right) = \sum_{j=1}^{N} h_j\cdot r_{XX}(j) = \underline{h}_{opt}^T\cdot\underline{r}_{XX}$$

Since we know that

$$\underline{h}_{opt} = \underline{\underline{R}}_{XX}^{-1}\underline{r}_{XX}$$

We get the result

$$\sigma_e^2 = \sigma_x^2 - \underline{r}_{XX}^T\cdot\underline{\underline{R}}_{XX}^{-T}\cdot\underline{r}_{xx}$$

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Coding Gain

- Prediction error

$$\sigma_e^2 = \sigma_x^2 - \underline{r}_{XX}^T \underline{\underline{R}}_{XX}^{-1} \underline{r}_{XX}$$

Time domain

From Book “Jayant, Noll”:

$$\lim_{N \to \infty} \sigma_e^2 = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \log S_{XX}\left(e^{j\omega}\right) d\omega\right]$$

Frequency domain

$$\Rightarrow \quad \frac{1}{2}\log\left(\sigma_e^2\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log S_{XX}\left(e^{j\omega}\right) d\omega$$

number of bits for predictive coding

can be viewed as number of bits for subband coding

- Comparable to bits for subband coding
- Coding gain depends on this “Spectral Flatness Measure”.

they are equal
→ amount of redundancy is given by signal (not by method)

Reference: “Digital Coding of Waveforms”, Jayant, Noll, Prentice-Hall, 1984

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Predictive Coding – Subband Coding

- Reduce redundancy in input signal

- **Redundancy** in input signal is **independent of method**
  - Predictive coding and subband coding will achieve **same results for N** $\rightarrow$ **∞**
  - Different properties result for finite N

- Example:
  - few sinusoids $\rightarrow$ better prediction with finite N
  - narrowband noise $\rightarrow$ better subband coding with finite N

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Lossless Coding

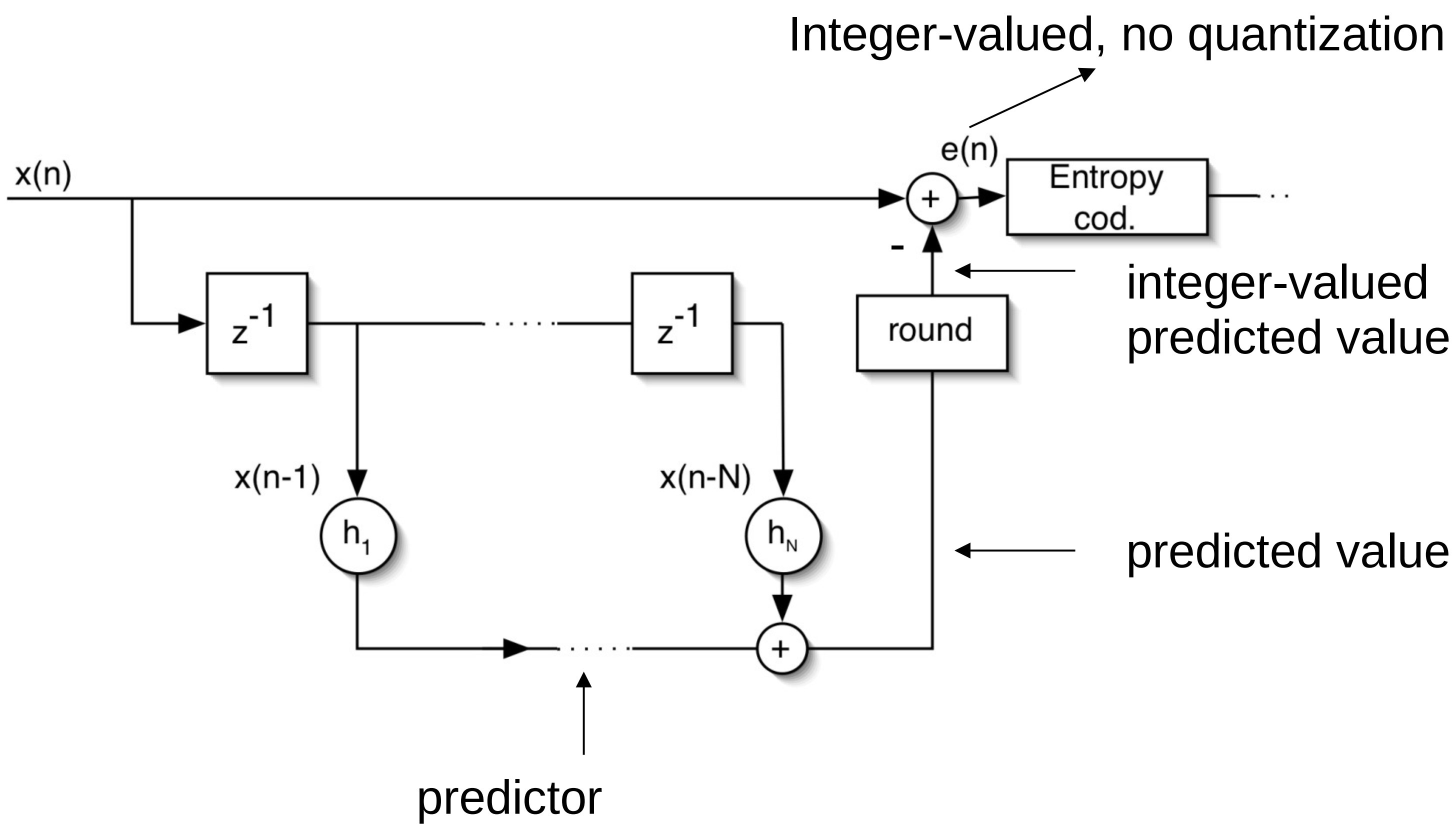
- Definition:
  - the decoded and original signal are **bit identical / integer identical**

- original signal:
  - integer valued audio samples
- lossless coding **only removes redundancy, no psychoacoustics or irrelevancy removal** is done

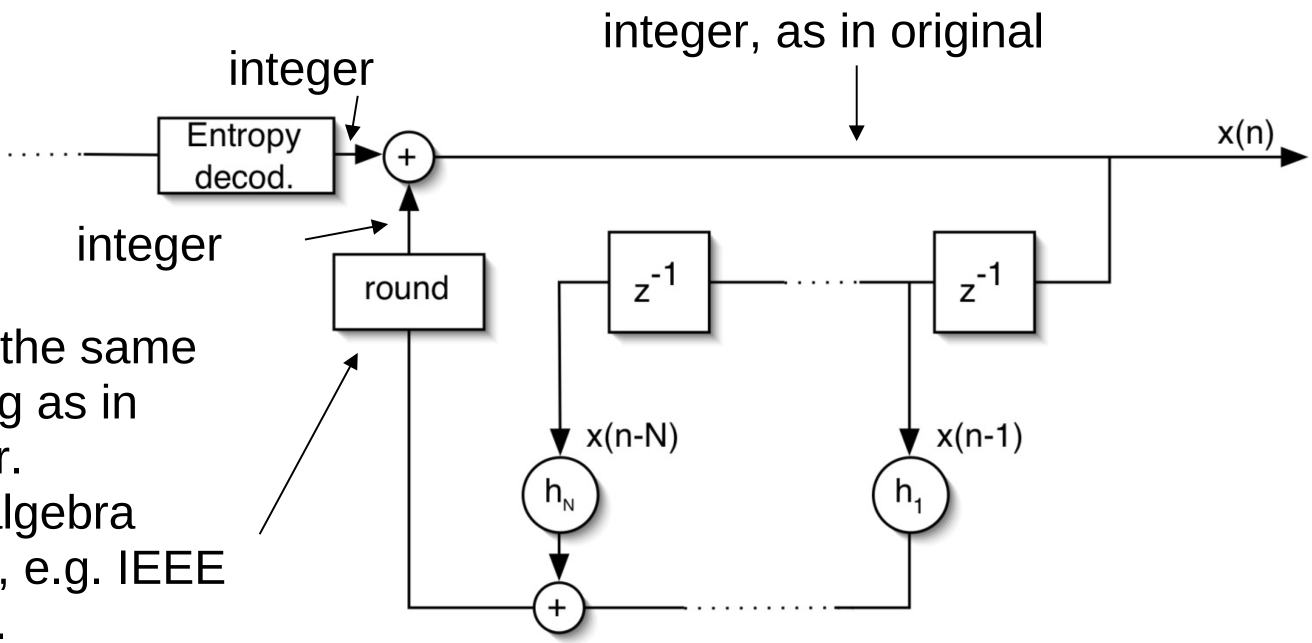
- prediction is convenient for lossless compression
  - integer to integer prediction
  - prediction error can easily be made integer valued
  - inverse prediction results in original integers!

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Predictive Lossless Encoder



Predictive Lossless Decoder



exactly the same rounding as in encoder.
Same algebra needed, e.g. IEEE defined.
example: rounding of 0.5 needs to be the same

Approaches to Predictive Coding

- How to **adapt h_j** for real world signals
  - Wiener-Hopf for a **block of a certain length**

    → transmit **h_j as side info** (most freeware lossless audio coders)

    long blocksize: good for low side info

    short blocksize: good for signal adaptation

    -This is called "Linear Predictive Coding" (LPC)

    - For speech coding usually blocks of 20 ms
  - This approach is taken for the **speech coding part** of
    - MPEG-Universal Speech and Audio Coding (**USAC**)

      (its audio coding part uses the AAC tools)
    - 3GPP Enhanced Voice Services (**EVS**) standard
    - The **AMR** (Adaptive Multi Rate) codec.
  - **Python example**: python3 lpcexample.py

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Approaches to Predictive Coding

References:

- 3GPP: https://en.wikipedia.org/wiki/Enhanced_Voice_Services
- MPEG-USAC: https://en.wikipedia.org/wiki/Unified_Speech_and_Audio_Coding
- ITU AMR: https://en.wikipedia.org/wiki/Adaptive_Multi-Rate_audio_codec

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Approaches to Predictive Coding

- **LMS (Least Mean Squares)-Method**: **Online update** derived from "**Stochastic Gradient Descent**" minimization of the prediction error.

  Normalized LMS update formula:

$$h_j(n+1)=h_j(n)+\frac{x(n)-\hat{x}(n)}{a+\lambda\sigma_x^2}x(n-j)$$

  → **no side info, no blocks necessary**

This is called Adaptive Differential Pulse Code Modulation (ADPCM)

It is used e.g. in the G.726, G.722, and G.722.2 ITU-T speech coding standards.

Python example: python3 lmsquantexample.py

Approaches to Predictive Coding

References:

- ITU G.726:
- https://en.wikipedia.org/wiki/G.726
- https://www.itu.int/rec/T-REC-G.726/en

- ITU G.722:
- https://en.wikipedia.org/wiki/G.722
- https://www.itu.int/rec/T-REC-G.722/en
- ITU G.722.2:
- https://www.itu.int/rec/T-REC-G.722.2/en

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

References/Literature:

- Lossless Compression of Digital Audio

  H.Mat, R. Schafer

  IEEE Signal Processing Magazine

  July 2001

  http://ieeexplore.ieee.org

- Perceptual Coding Using Adaptive Pre- and Post-Filters and Lossless Compression

  G. Schuller et al.

  IEEE Trans. On Speech and Audio Signal Processing

  Sept 2002

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Lossless Audio Coding with Filter Banks

- Perceptual audio codecs: usually based on filter banks

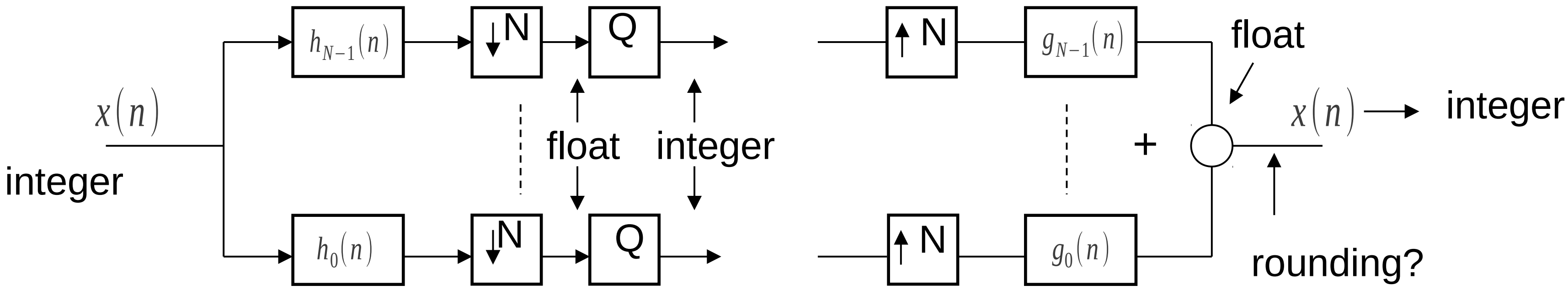
- Lossless audio codecs: usually based on prediction

- Lossless audio coding using filter banks?

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Lossless Audio Coding with Filter Banks

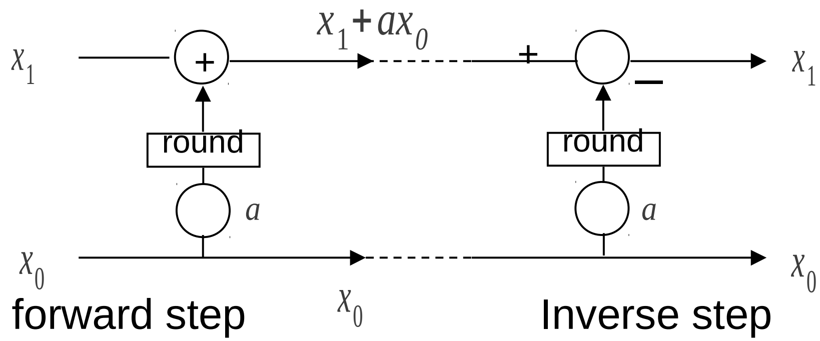
- Problem: Input values integer, output values not integer
- Possible solution: add quantizer



- Drawback of this quantization
  - destroys perfect reconstruction
  - has to be very fine or error in time domain has to be coded additionally

Lifting Scheme (aka „Ladder Network“)

- Goal: **Invertible integer-to-integer transform**
- Principle: Insert quantizer without destroying perfect reconstruction
- Lifting Scheme or Ladder Network:



a: float

$$y_1 = x_1 + round(a * x_0) \qquad x_1' = y_1 - round(a * y_0) = x_1$$

$$y_0 = x_0 \qquad x_0' = y_0 = x_0$$

→ invertible integer-to-integer transform

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

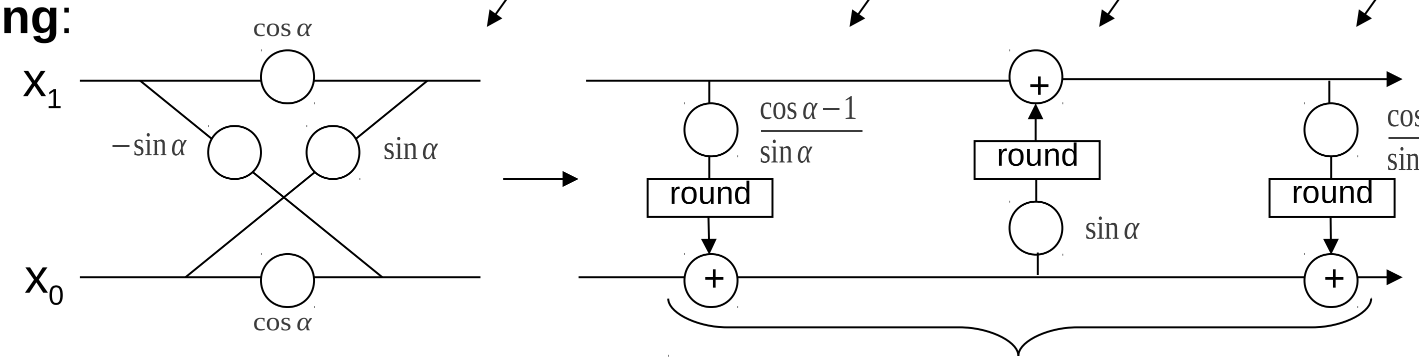
Givens Rotations by Lifting Scheme

- Apply lifting scheme to “Givens rotation” or rotation matrix
- Re-write rotation as product of 3 Lifting matrices:

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} = \begin{bmatrix} 1 & \dfrac{\cos\alpha - 1}{\sin\alpha} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \sin\alpha & 1 \end{bmatrix} \begin{bmatrix} 1 & \dfrac{\cos\alpha - 1}{\sin\alpha} \\ 0 & 1 \end{bmatrix}$$

The same as

Block diagram, where we can now **include rounding**:



- Result: **Invertible integer approximation** of the rotation

Application to MDCT

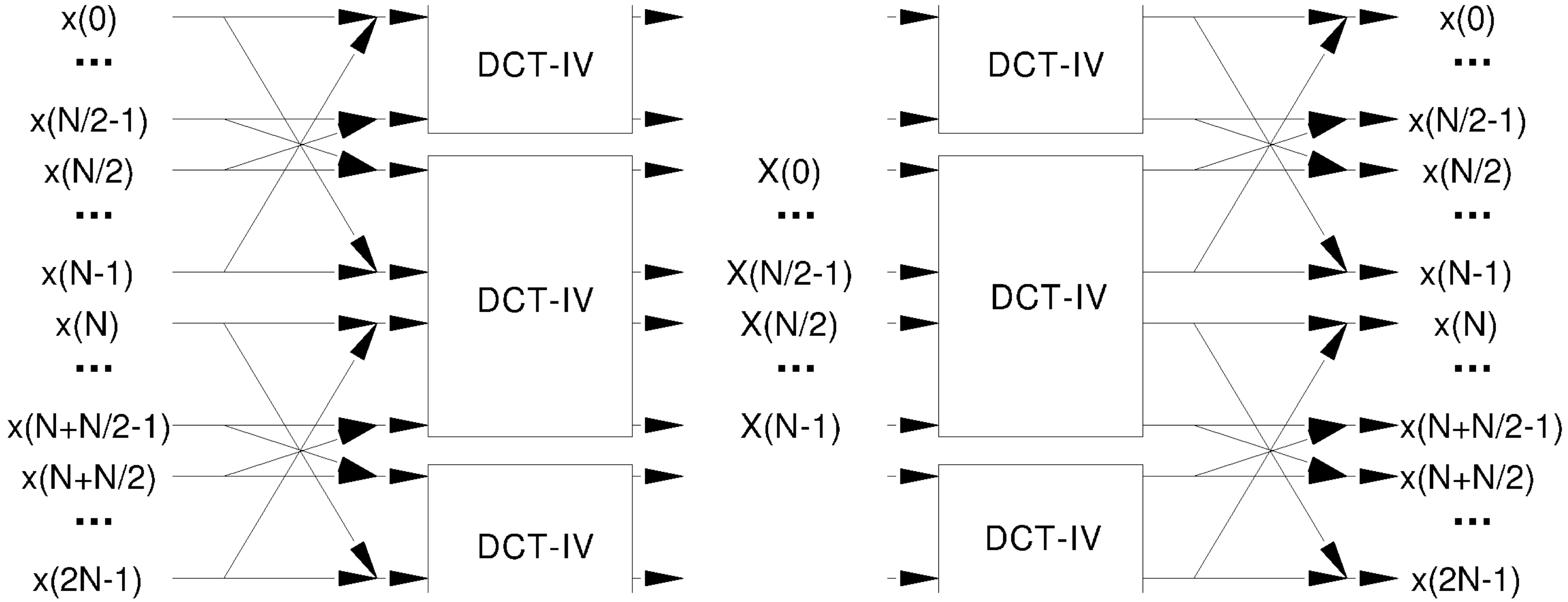
- MDCT can be decomposed into
  - Windowing / Time Domain Aliasing
  - DCT of type IV (DCT-IV)

- Both blocks can be decomposed into Givens rotations

- For DCT-IV: Fast algorithms usually provide such a decomposition

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

MDCT/inverse MDCT by Givens rotations and DCT$_{IV}$



Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU
Fraunhofer IDMT

Integer Modified Discrete Cosine Transform (IntMDCT)

- MDCT can be completely decomposed into Givens rotations
- Apply lifting scheme for each Givens rotation
- Result: Invertible integer approximation of MDCT, called “IntMDCT”

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU
Fraunhofer IDMT

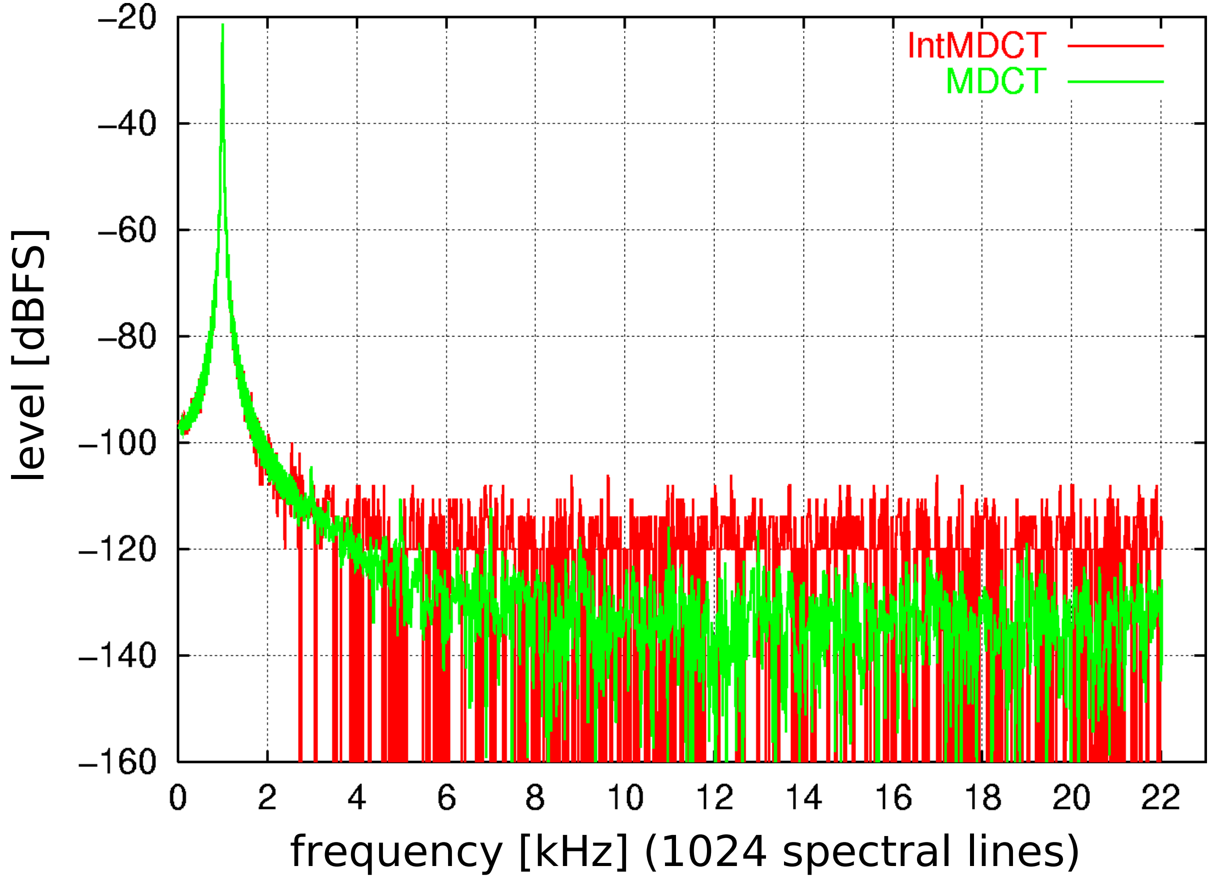
Properties of IntMDCT

- Inherits properties of MDCT
  - perfect reconstruction
  - critical sampling
  - overlapping of blocks
  - good spectral representation of audio signal

- Allows lossless coding in frequency domain by entropy coding of integer spectral values (again, no quantization necessary)

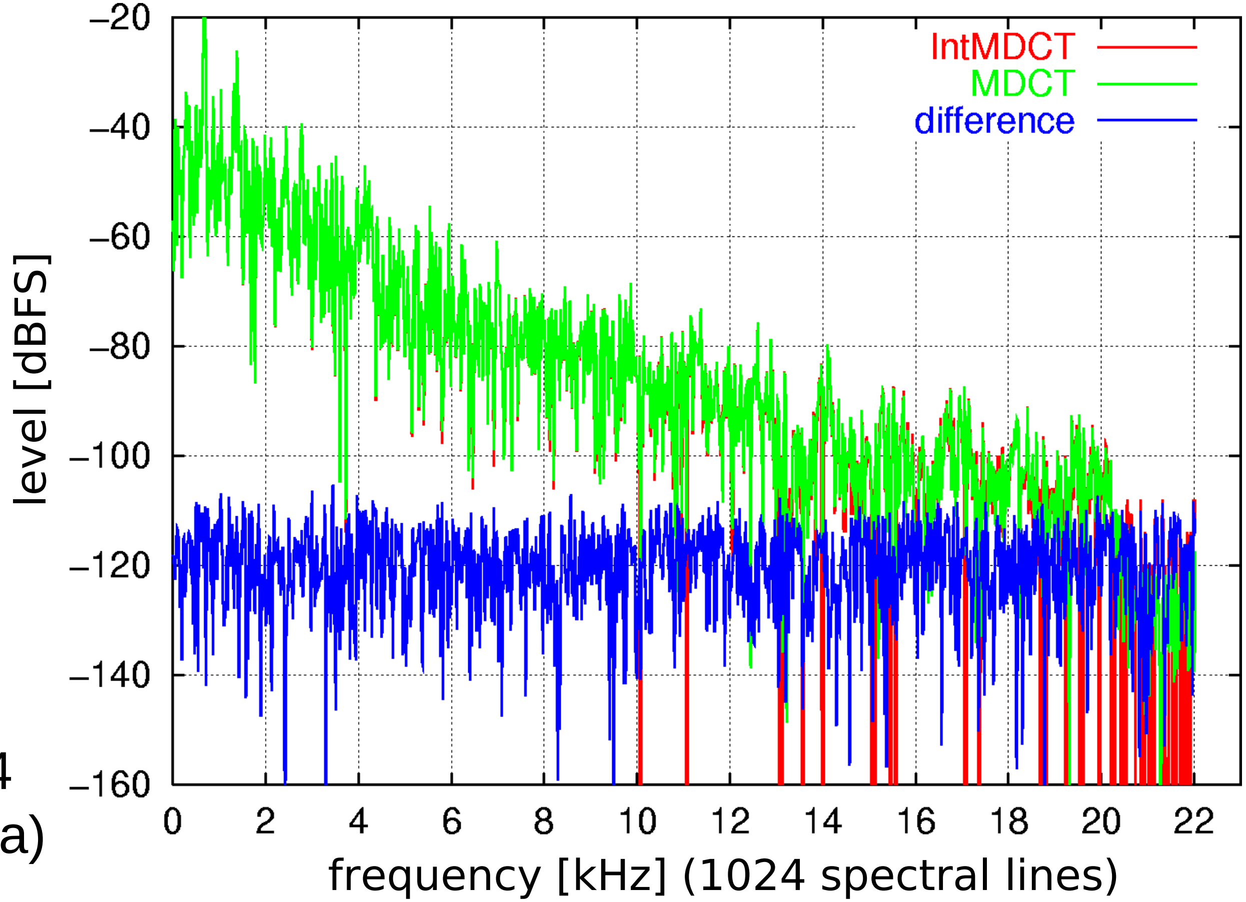
Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

IntMDCT and MDCT of sine wave (1kHz, -20dBFS)



Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

IntMDCT, MDCT and difference values



Item: SQAM, track 64
(Orff: Carmina Burana)

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU
Fraunhofer IDMT

Recent Improvement: Multi-Dimensional Lifting

- Decompose DCT-IV into two DCT-IV of half length
- Further decompose:

1024 1024

[left,right]

2048x2048

$DCT_{IV} \cdot DCT_{IV} = I$

$$\begin{pmatrix} \mathrm{DCT_{IV}} & 0 \\ 0 & \mathrm{DCT_{IV}} \end{pmatrix} =$$

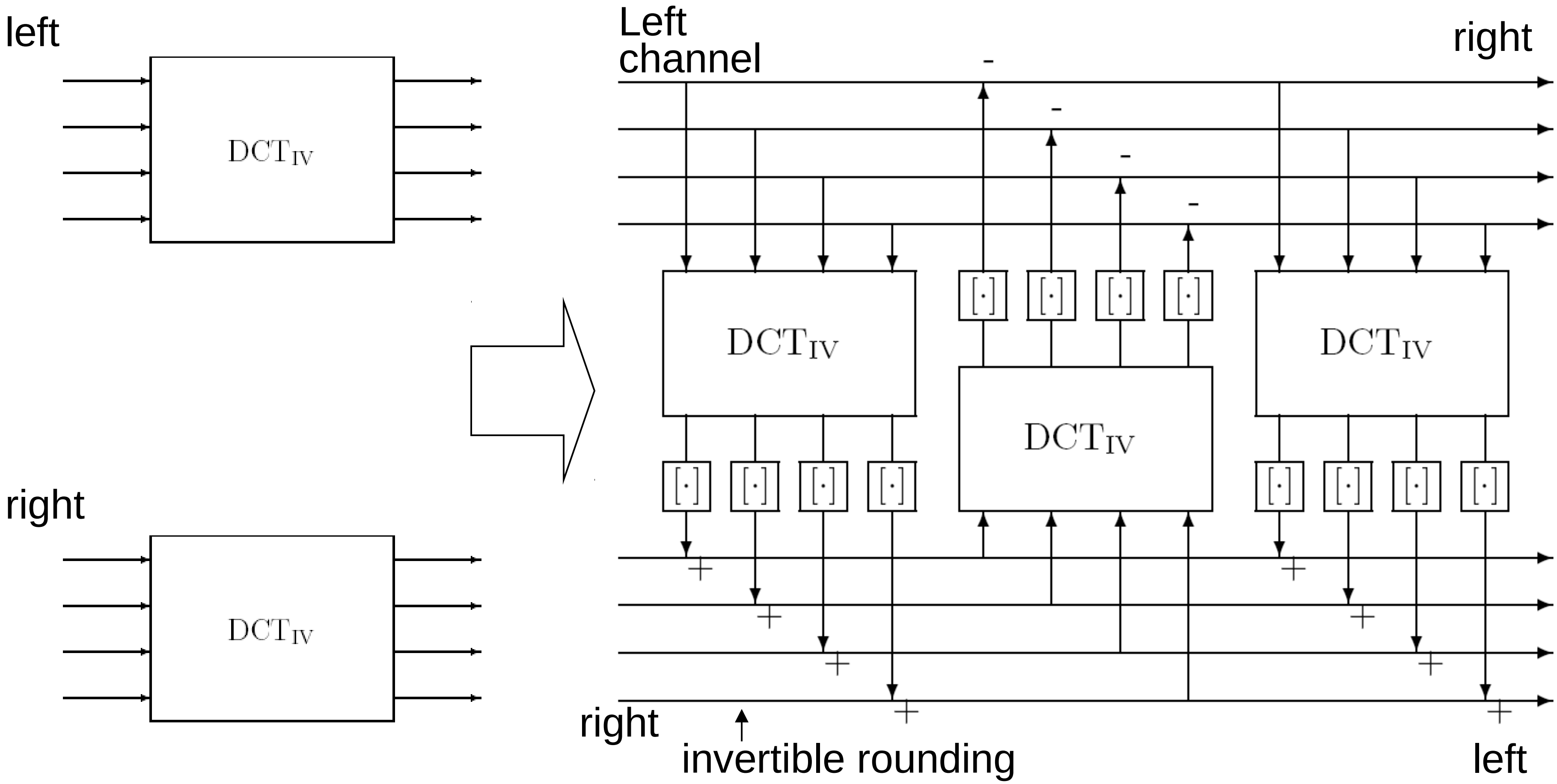
$$\begin{pmatrix} -I_N & 0 \\ \mathrm{DCT_{IV}} & I_N \end{pmatrix} \begin{pmatrix} I_N & -\mathrm{DCT_{IV}} \\ 0 & I_N \end{pmatrix} \begin{pmatrix} 0 & I_N \\ I_N & \mathrm{DCT_{IV}} \end{pmatrix}$$

DCT is not in main signal path any more! → lifting

- Apply lifting scheme to 2x2 **block** matrices instead of 2x2 matrices
- Result: Approximation error reduced from O(Nlog(N)) to O(N)

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Two blocks of DCT-IV by Multi-Dimensional Lifting



Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU

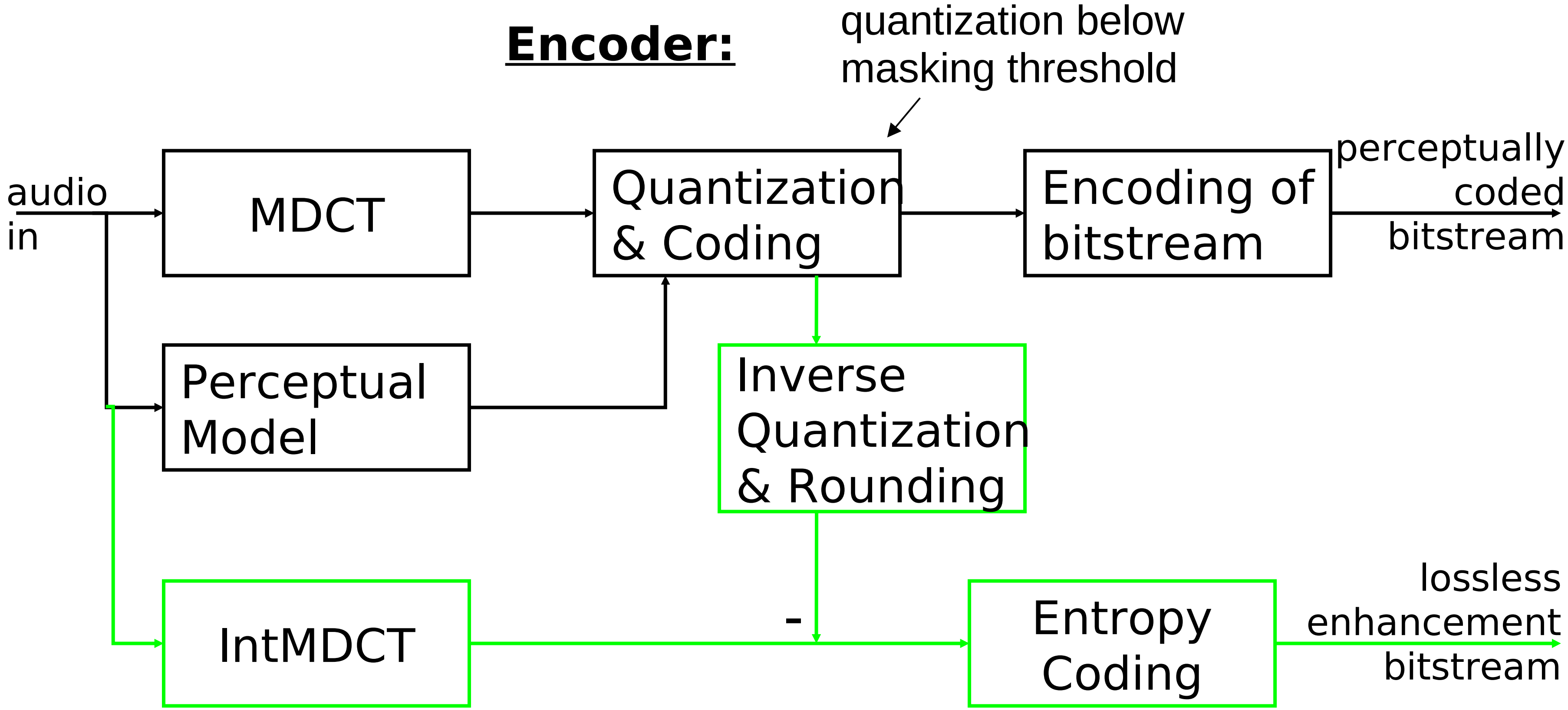
Fraunhofer IDMT

Lossless enhancement of perceptual coder (1)

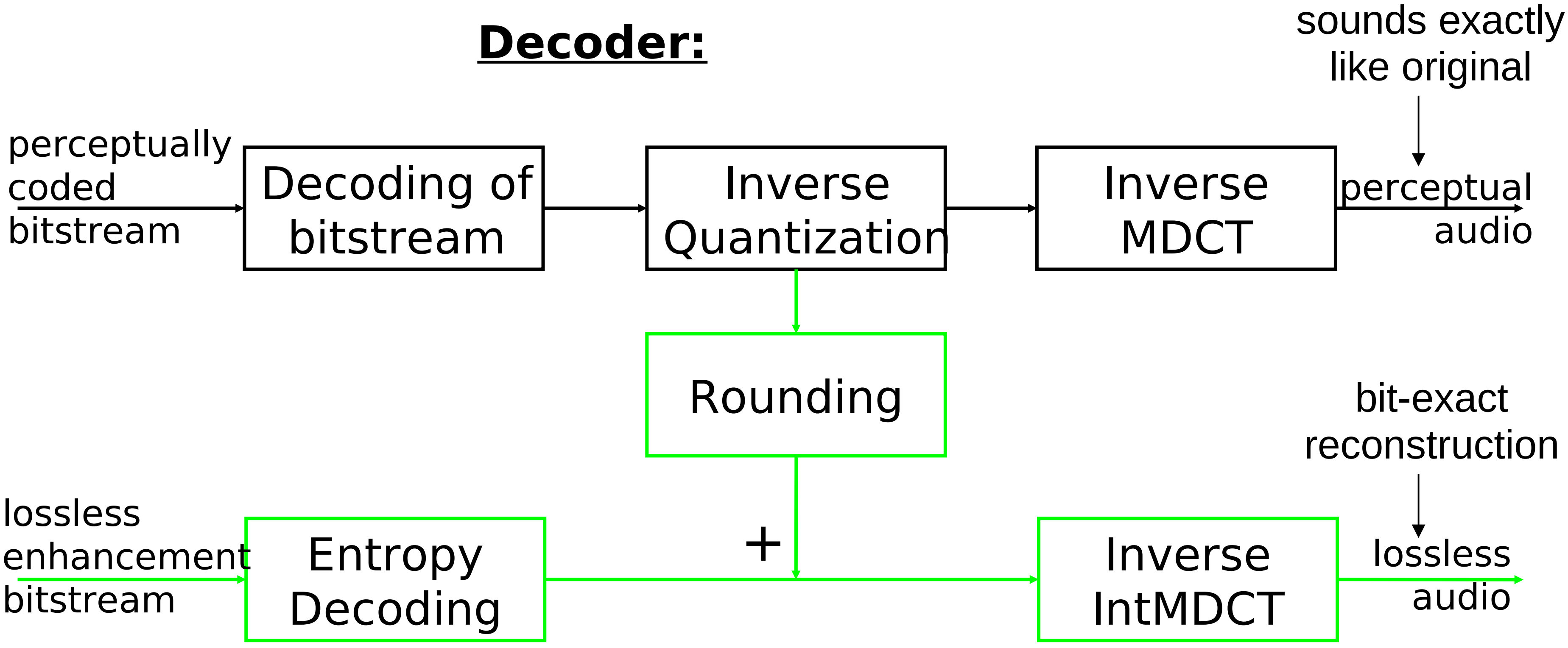
- IntMDCT closely approximates MDCT
- Scalable combination with MDCT-based perceptual codec (e.g. AAC) possible
- Scalable bitstream with two layers allows two stages of decoding
  - Perceptually coded (e.g. AAC @ 128 kBit/s)
  - Lossless (higher, variable bitrate)

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Lossless enhancement of perceptual coder (2)



Lossless enhancement of perceptual coder (3)



Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

Compression Results

Results in bits per sample:

	48 kHz 16 bit	48 kHz 24 bit	96 kHz 24 bit	192 kHz 24 bit
AAC Enhancement	1.3 6.5	1.3 14.4	0.8 11.0	0.5 9.2
AAC + Enhancement	<u>7.8</u>	15.7	11.8	9.7
Lossless-only	7.5	15.3	11.6	9.5
Monkey's Audio 3.97	7.2	15.2	11.5	9.4
Simulcast (AAC + Monkey's Audio)	<u>8.5</u>	16.5	12.3	9.9

Signals: Test set used in MPEG Lossless Audio activities

Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU
Fraunhofer IDMT

Conclusions

- Lossless Audio Coding with filter banks is possible
- Lifting Scheme or Ladder Network is appropriate tool
- IntMDCT allows
  - Efficient lossless audio coding
  - Scalable lossless enhancement of MDCT-based perceptual audio codec (e.g. AAC)

References for IntMDCT:

- Yokotani, Y.; Geiger, R.; Schuller, G.D.T.; Oraintara, S.; Rao, K.R.: "Lossless Audio Coding Using the IntMDCT and Rounding Error Shaping", IEEE Transactions on Audio, Speech, and Language Processing, Volume 14, Issue 6, pp. 2201-2211, November 2006
- R. Geiger, G. Schuller: "Fine Grain Scalable Perceptual and Lossless Audio Coding Based on IntMDCT", IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hong Kong, April 6-10, 2003