Prediction and Lossless Audio Coding

Prof. Dr.-Ing. Karlheinz Brandenburg

Fraunhofer IDMT & Ilmenau Technical University
Ilmenau, Germany

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

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Use of Redundancy (1)

- For higher correlation between samples ! → higher redundancy
- For „flat“ PSD → low redundancy
- ACF (Auto Correlation Function):

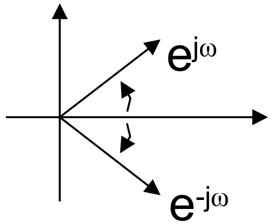
$$r_{XX}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t)\,x(t+\tau)\,dt = E\{x(t)\,x(t+\tau)\}$$

- PSD (Power Spectrum Density):

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

$$\circ\!\!-\!\!\bullet\quad S_X(f) \cdot \bar{S}_X(f) = |S_X(f)|^2$$

impulse response



conjugate complex

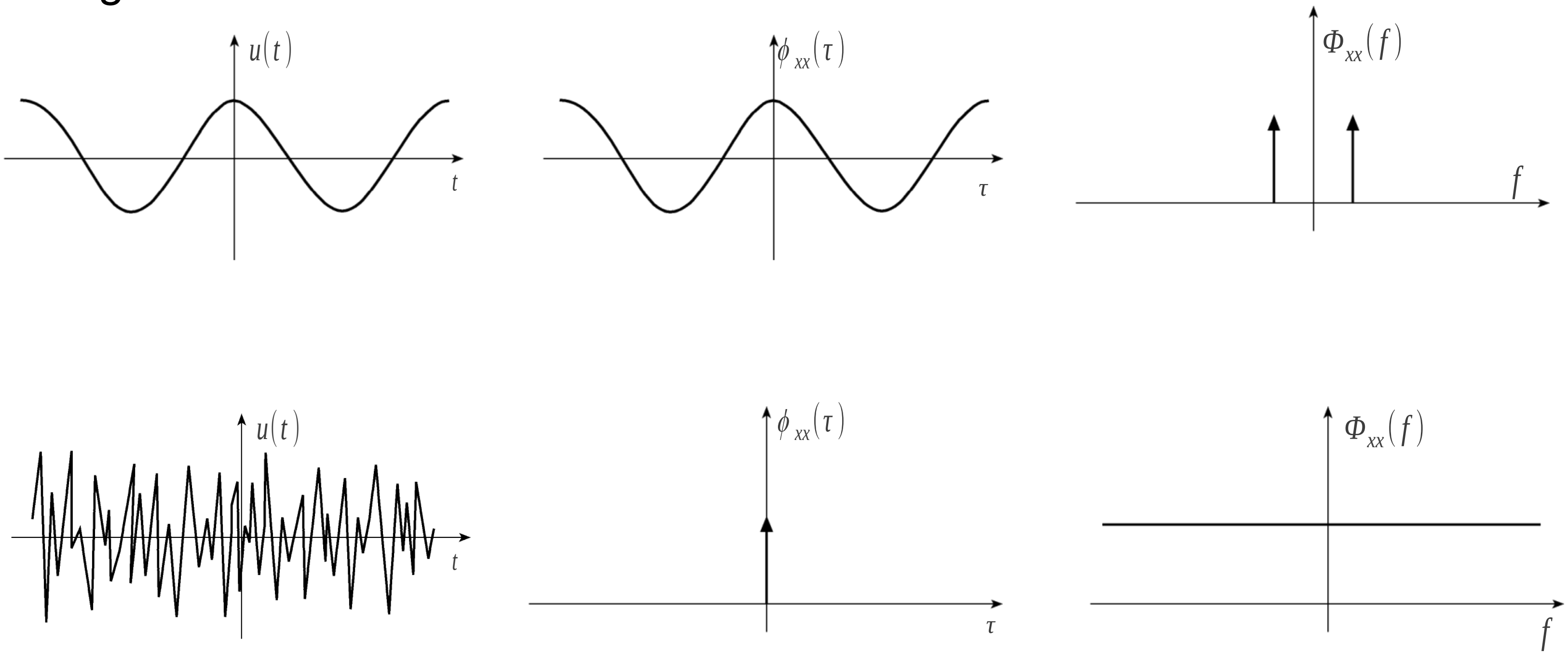
Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Use of Redundancy (2)

Signal | ACF | PSD



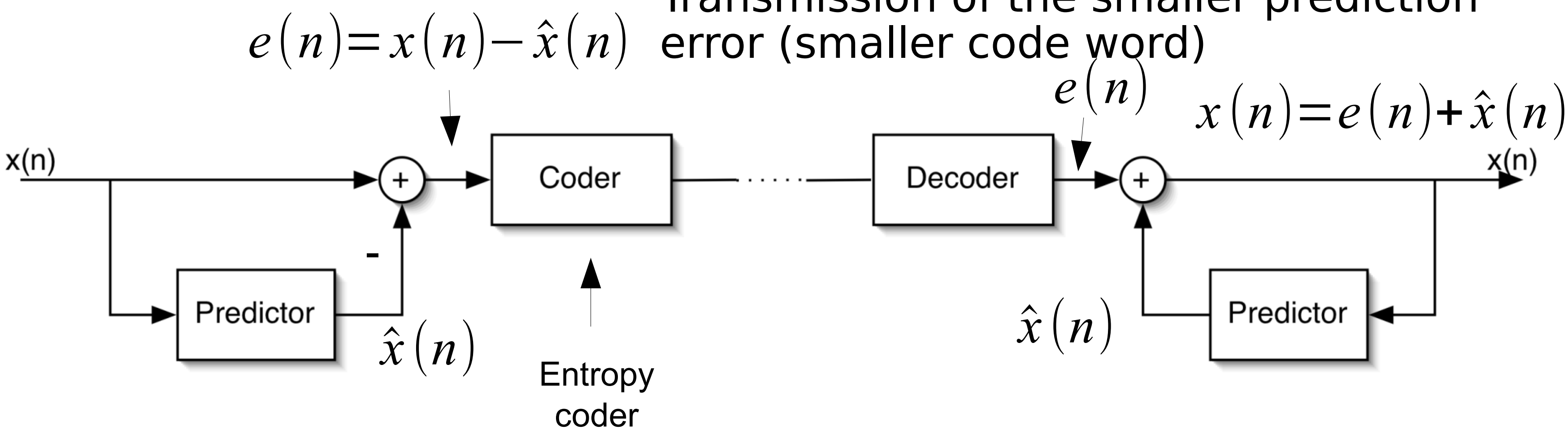
Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Predictive Coding (1)

- 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)



Predictive Coding (2)

- Encoder

prediction error, to be encoded

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

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

predicted value

weighted sum of past values

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 by optimizing the filter coefficients h_j

$$\sigma_e^2 = E\{e^2(n)\}$$

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

Predictive Coding (3)

- Approach: $\frac{\partial \sigma_e^2}{\partial a_j} \overset{!}{=} 0$

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

Orthogonality Principle:

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

Eq. 1:

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

$$\Rightarrow 0 = r_{XX}(k) - \sum_{j=1}^{N} a_j r_{XX}(k-j), \quad r_{XX}(k) = E\left\{ x(n) x(n-k) \right\}$$

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

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

Predictive Coding (4)

Prediction error multiplied with the past signal itself is zero

- **orthogonality principle**: for the optimum coefficients the expectation (average) of the error is zero, hence the prediction error is said to be „orthogonal“ to the input signal (Eq. 1), meaning:
  - The pred. **error** and the N **past signal** samples are **uncorrelated**, if we have the **optimum prediction coefficients**!

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

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

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

- we also get: $E\left(e(n)\cdot \hat{x}(n)\right)=0$

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Predictive Coding (5)

- 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:

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

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

Deriving Wiener-Hopf with Pseudo Inverses (1)

- Input matrix X:

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

- Solve equation as close as possible to „*d*“ as our desired signal, in a quadratic sense (minimize sum of quadratic error):

more equations than unknowns $\longrightarrow$ $\underline{\underline{X}} \cdot \underline{h} \approx \underline{d}$

Sequence of “next” values

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

Deriving Wiener-Hopf with Pseudo Inverses (2)

- Solving the matrix equation with pseudo inverse of the input matrix $\underline{\underline{X}}^T$

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 $\longrightarrow$ $\underline{h} = \left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \cdot \underline{d}$

$\left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1}$ ACF estimation matrix inverse, $\rightarrow \underline{\underline{R}}_{xx}^{-1}$

$\underline{\underline{X}}^T \cdot \underline{d}$ 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 (1)

- The prediction error variance/power is

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

Using the decoder reconstruction equation:

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

we obtain:

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

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

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

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

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

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

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Coding Gain (2)

- 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

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

We get the result

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

Coding Gain (2)

- Minimal prediction error

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

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

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

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 SFM

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

Predictive Coding – Subband Coding

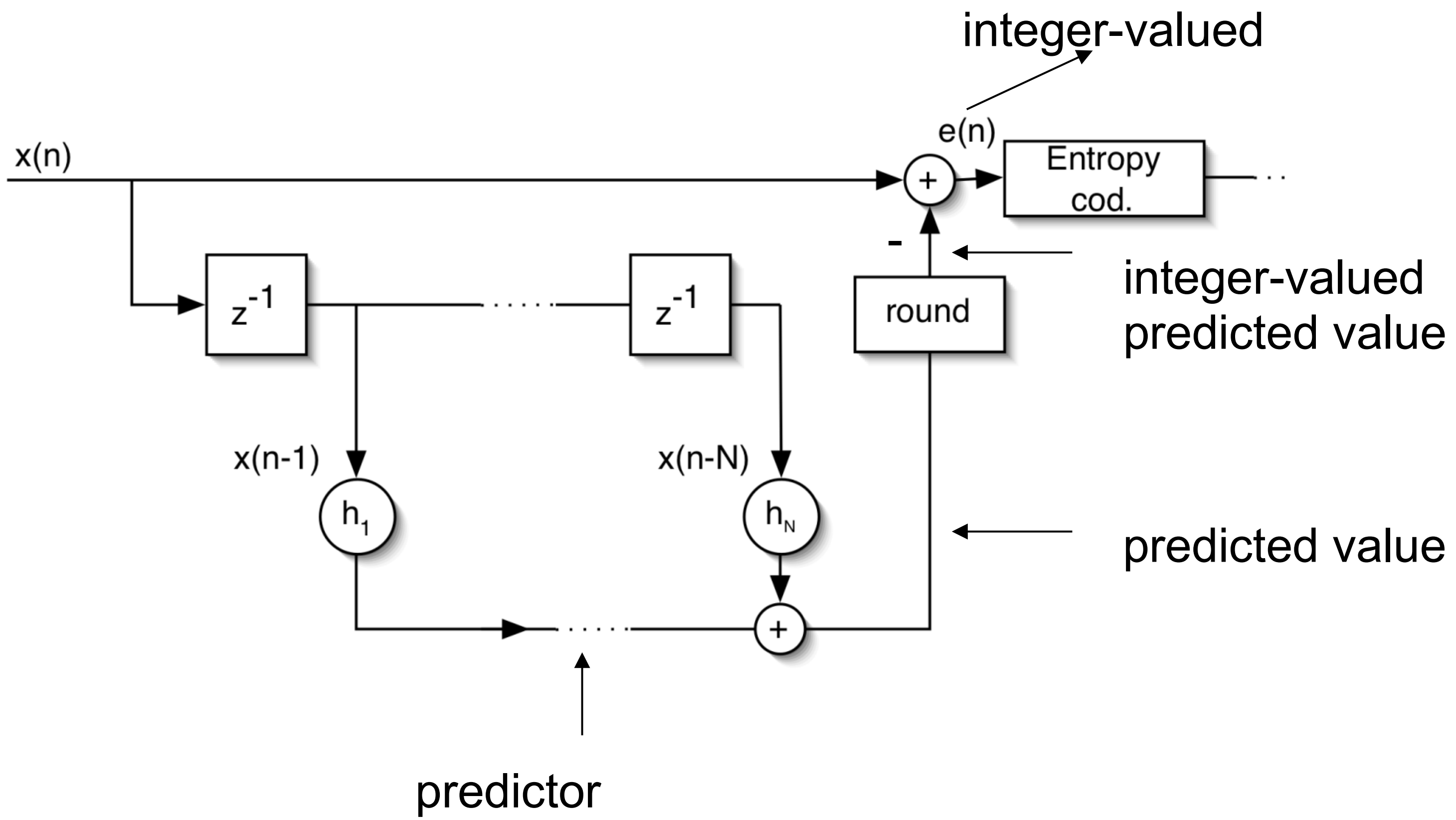
- Reduce redundancy in input signal
- Redundancy in input signal is independent of method
  - Predictive coding and subband coding will achieve same results for $N \rightarrow \infty$
  - Different properties result for finite N
- Example:
  - few sinusoids → better prediction with finite N
  - narrowband noise → better subband coding with finite N

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

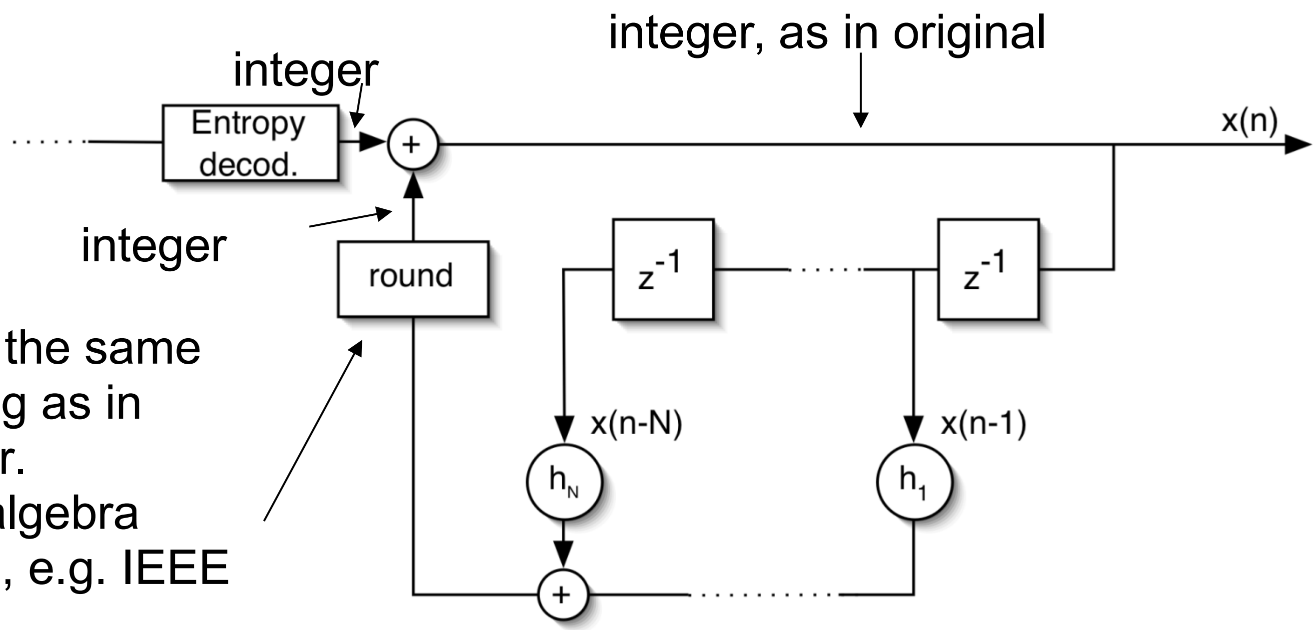
Predictive Encoder



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

TECHNISCHE UNIVERSITÄT ILMENAU
Fraunhofer IDMT

Predictive 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

  - LMS-Method: Online update derived from Wiener-Hopf for h_j based on past samples

    Normalized LMS:

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

    → no side info, no blocks necessary

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

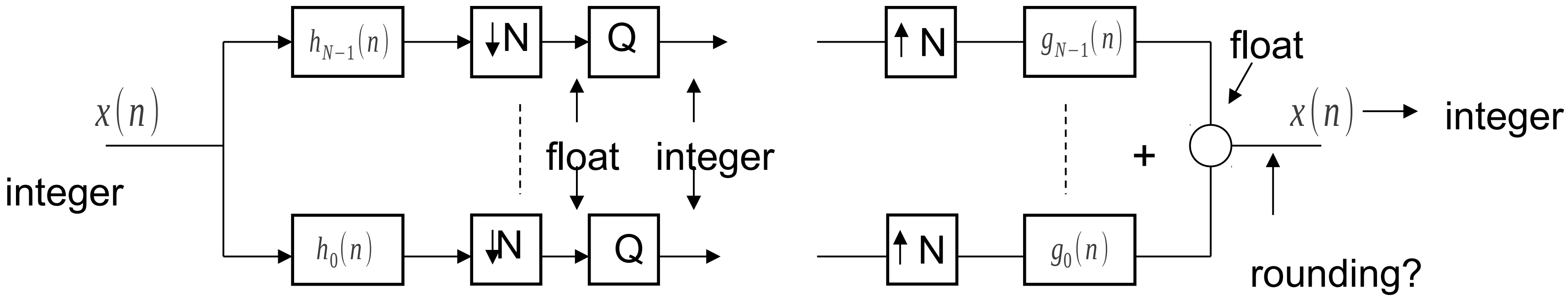
Lossless Audio Coding with Filter Banks

- 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

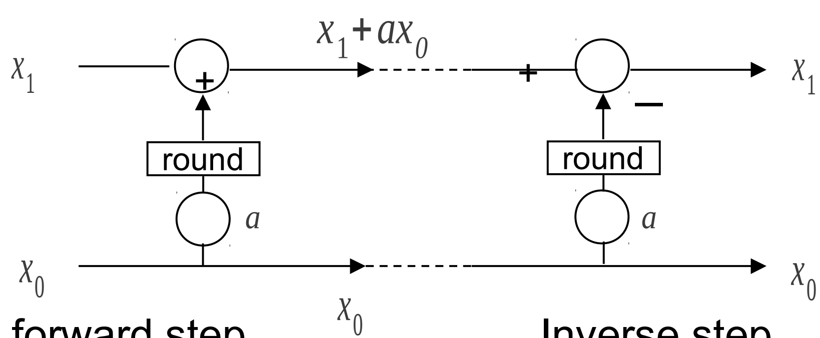
- Problem: Input values integer, output values not integer
- Possible solution: add quantizer



- 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

forward step

Inverse step

$$y_1 = x_1 + round(a * x_0)$$
$$y_0 = x_0$$

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

→ invertible integer-to-integer transform

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

TECHNISCHE UNIVERSITÄT ILMENAU

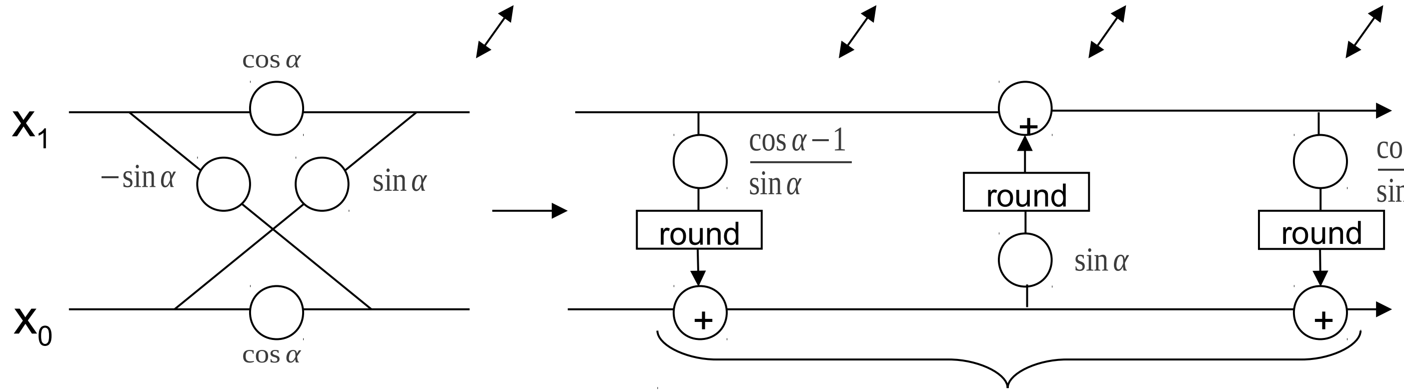
Fraunhofer IDMT

Givens Rotations by Lifting Scheme

- Apply lifting scheme to Givens rotation

- Decomposition:

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



- Result: Invertible integer approximation

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

Application to MDCT

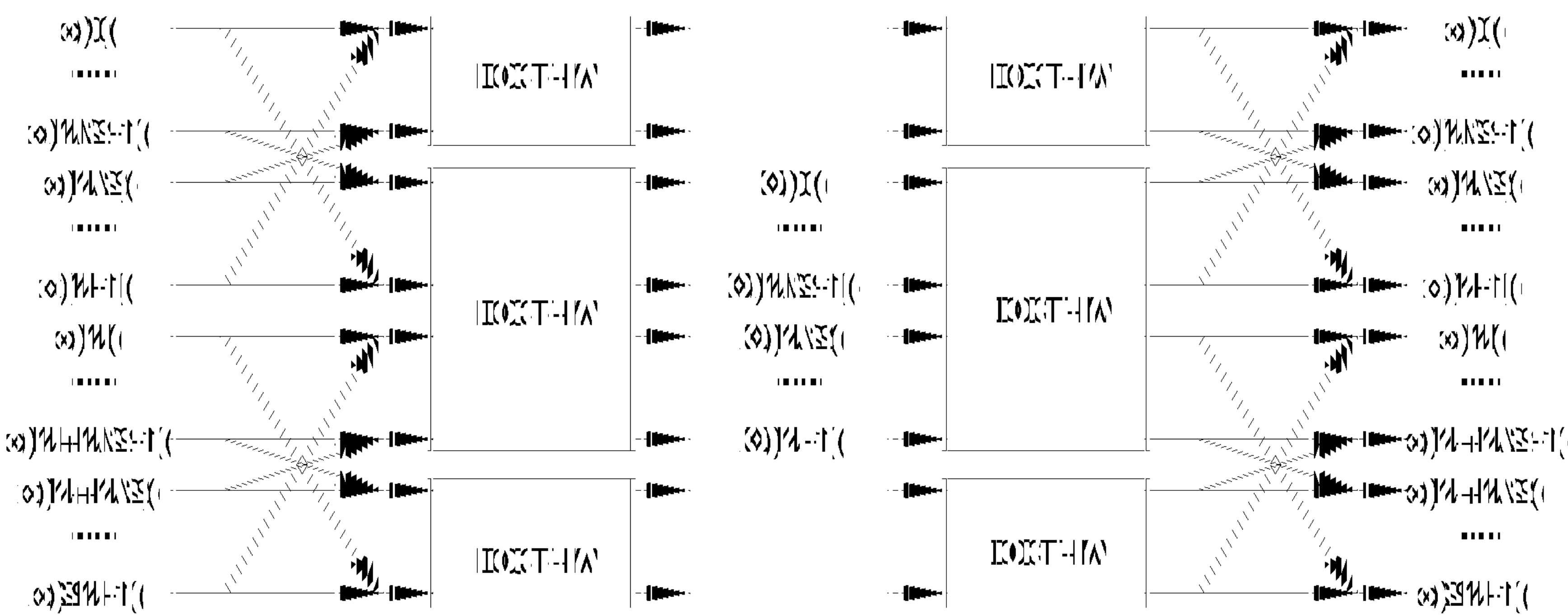
- MDCT can be decomposed into
  - Windowing / Time Domain Aliasing
  - DCT of type IV (DCT-IV)

- 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

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

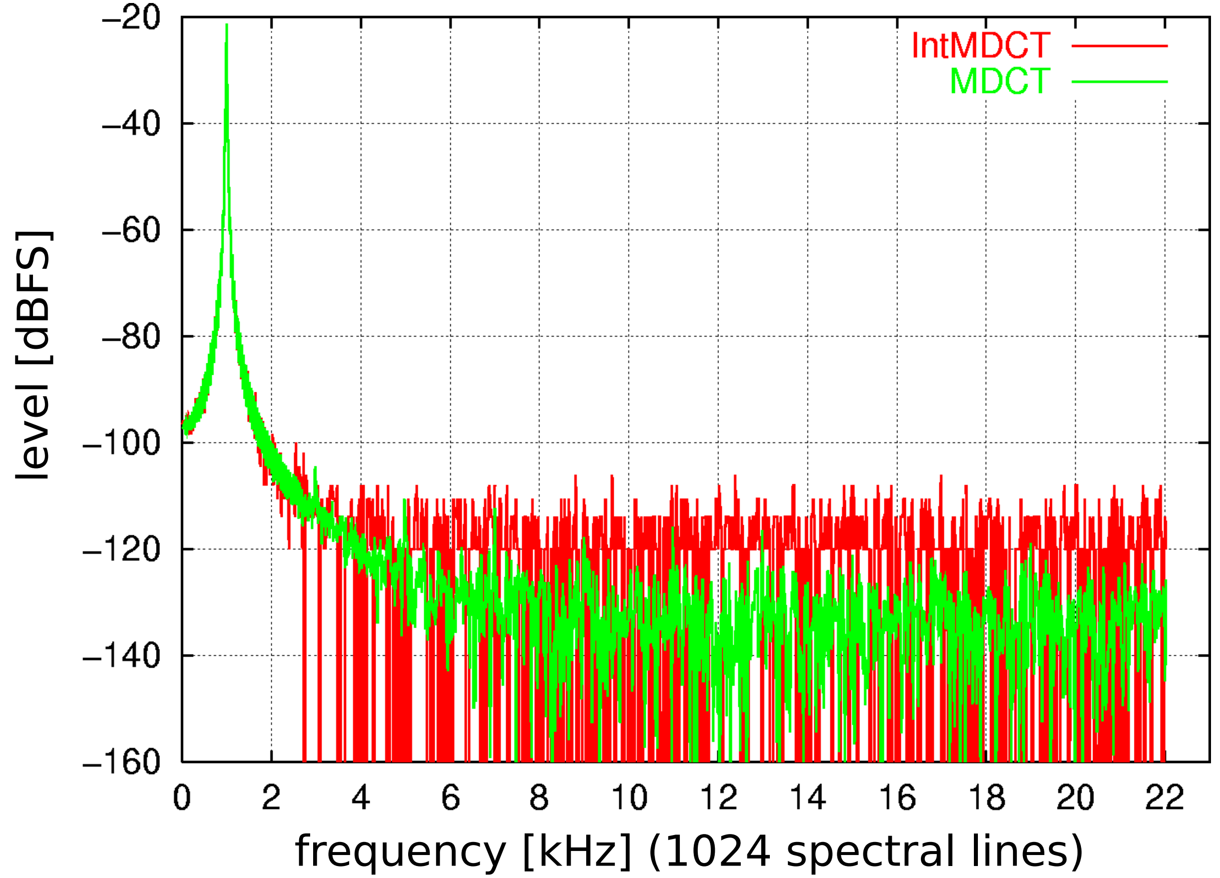
Properties of IntMDCT

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

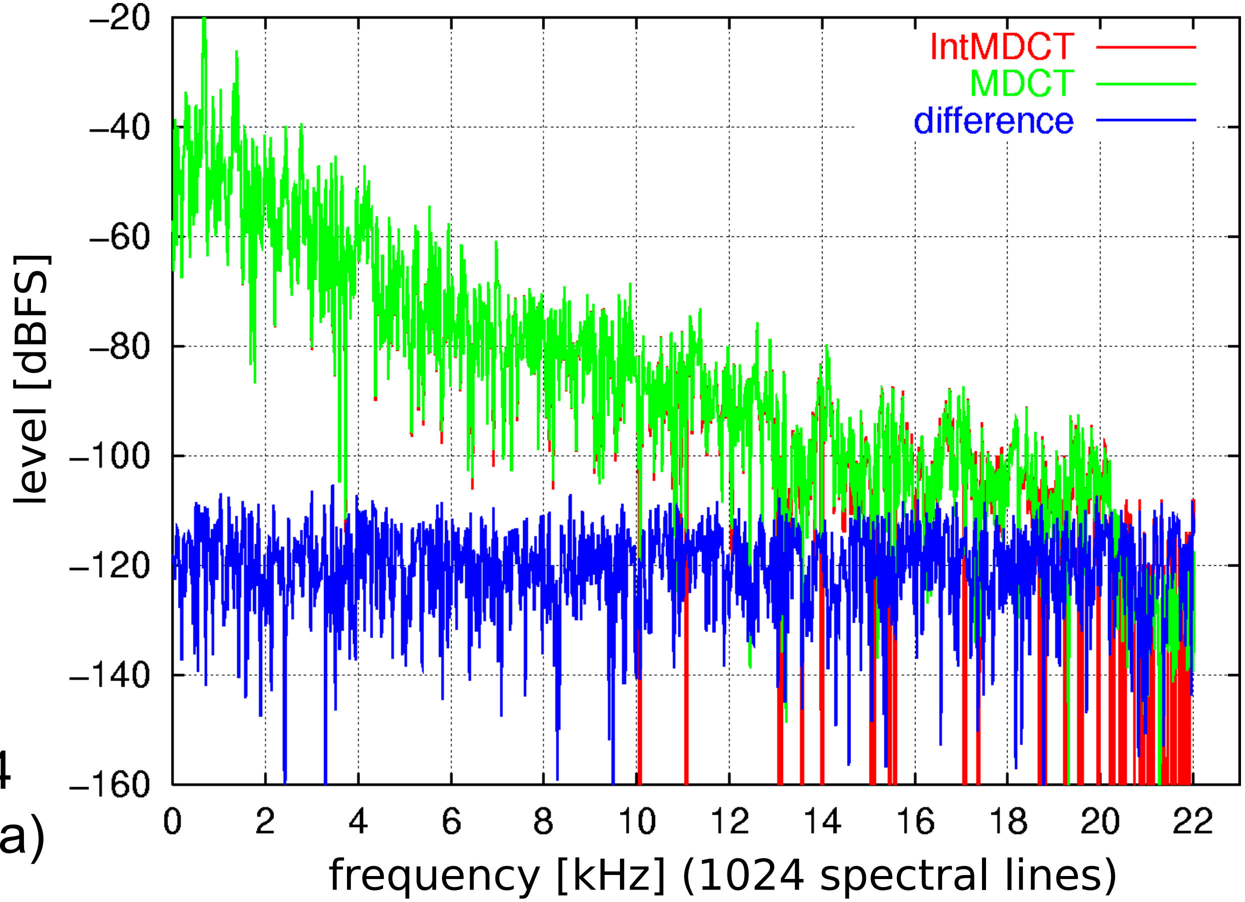
- Allows lossless coding in frequency domain by entropy coding of integer spectral values

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)



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

Recent Improvement: Multi-Dimensional Lifting

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

$\mathrm{DCT}_{IV} \cdot \mathrm{DCT}_{IV} = I$

1024 1024

[left,right]

2048x2048

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

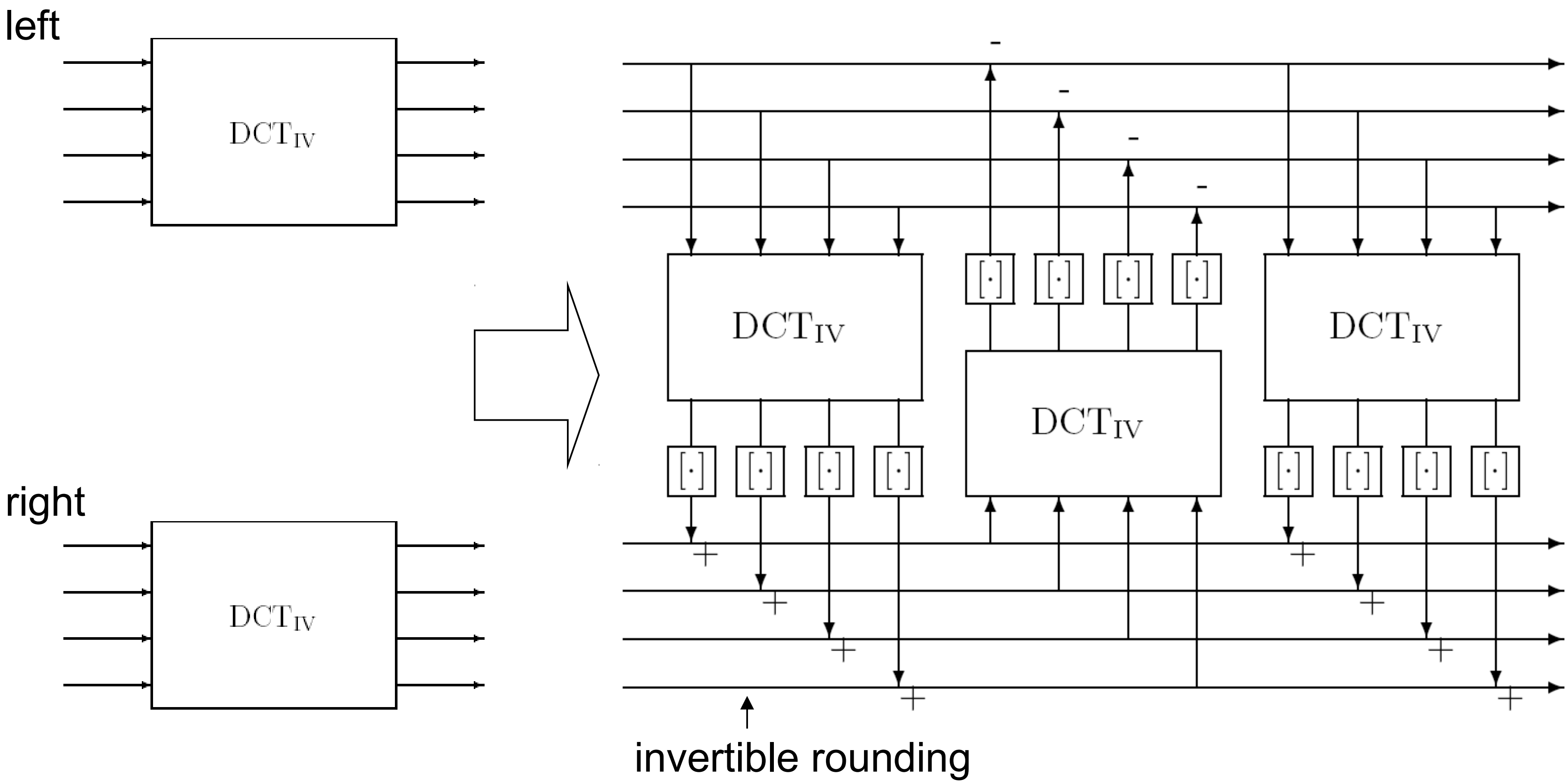
$$\begin{pmatrix} -I_N & 0 \\ \mathrm{DCT}_{\mathrm{IV}} & I_N \end{pmatrix} \begin{pmatrix} I_N & -\mathrm{DCT}_{\mathrm{IV}} \\ 0 & I_N \end{pmatrix} \begin{pmatrix} 0 & I_N \\ I_N & \mathrm{DCT}_{\mathrm{IV}} \end{pmatrix}$$

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

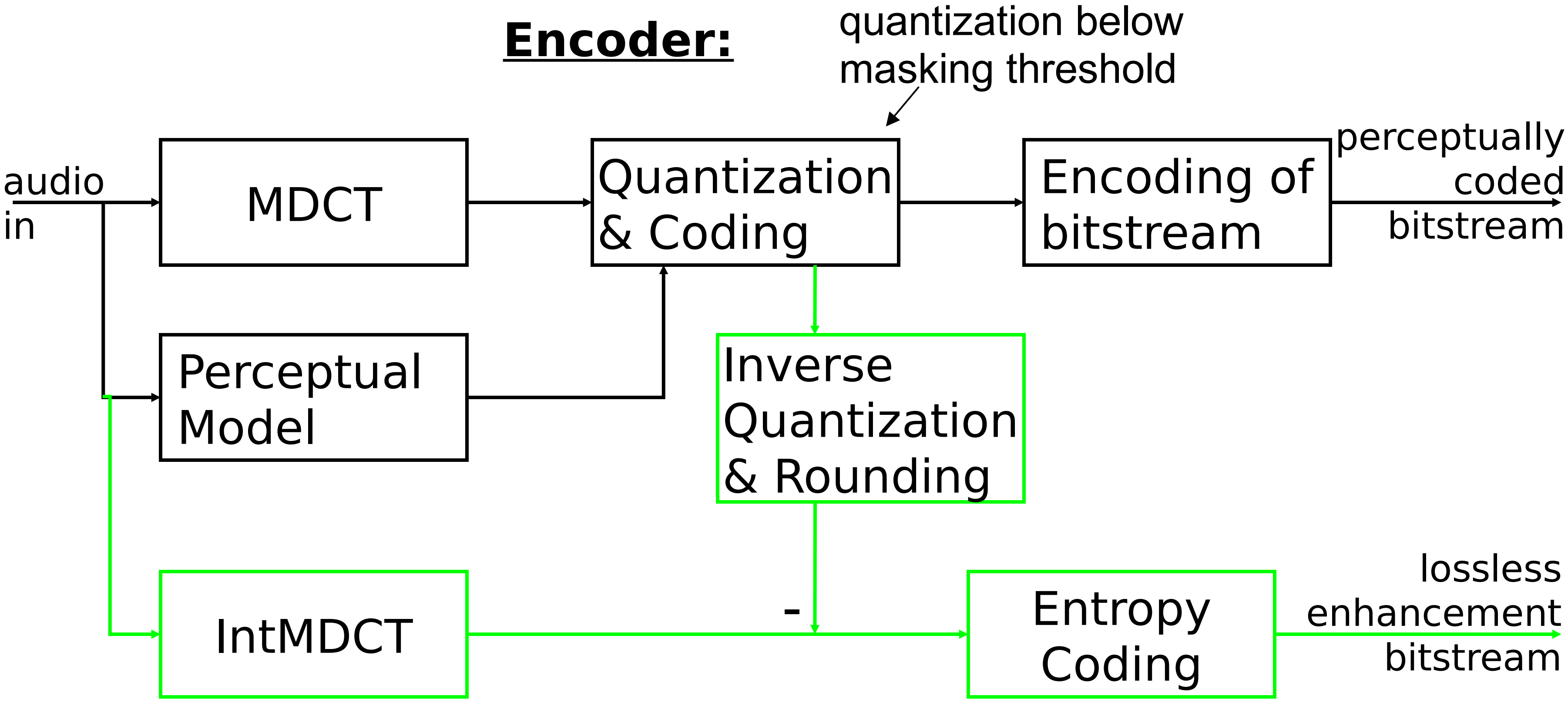


Lossless enhancement of perceptual coder (1)

- 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)



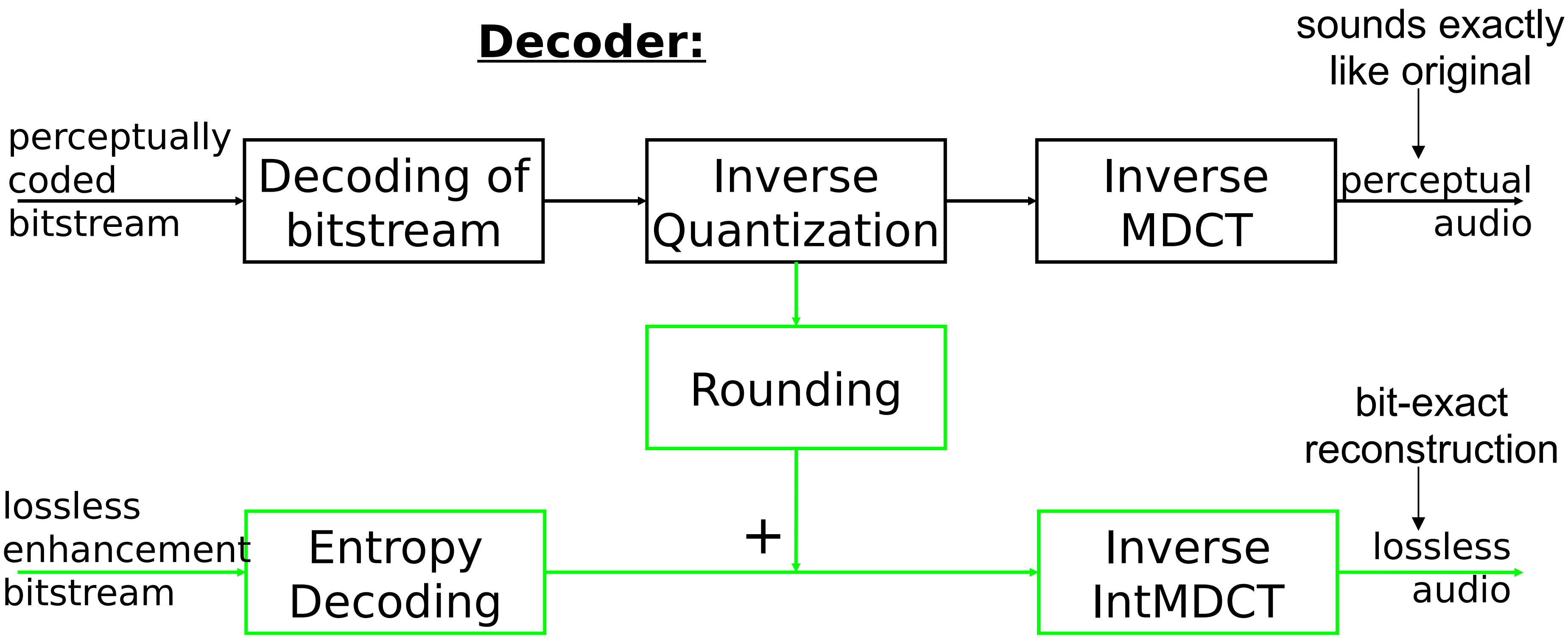
Prof. Dr.-Ing. K. Brandenburg, bdg@idmt.fraunhofer.de Prof. Dr.-Ing. G. Schuller, shl@idmt.fraunhofer.de

TECHNISCHE UNIVERSITÄT ILMENAU

Fraunhofer IDMT

Lossless enhancement of perceptual coder (3)

Decoder:



Compression Results

Results in bits per sample:

	48 kHz 16 bit	48 kHz 24 bit	96 kHz 24 bit	192 kHz 24 bit
AAC	1.3	1.3	0.8	0.5
Enhancement	6.5	14.4	11.0	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 ongoing MPEG Lossless Audio activities

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)

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

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

References:

- 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