Lecture

Speech and Audio Signal Processing

TECHNISCHE UNIVERSITÄT DARMSTADT

Lecture 3: Audio coding, Part I

Content of the lecture

TECHNISCHE UNIVERSITÄT DARMSTADT

- Audio coding

  Part I:

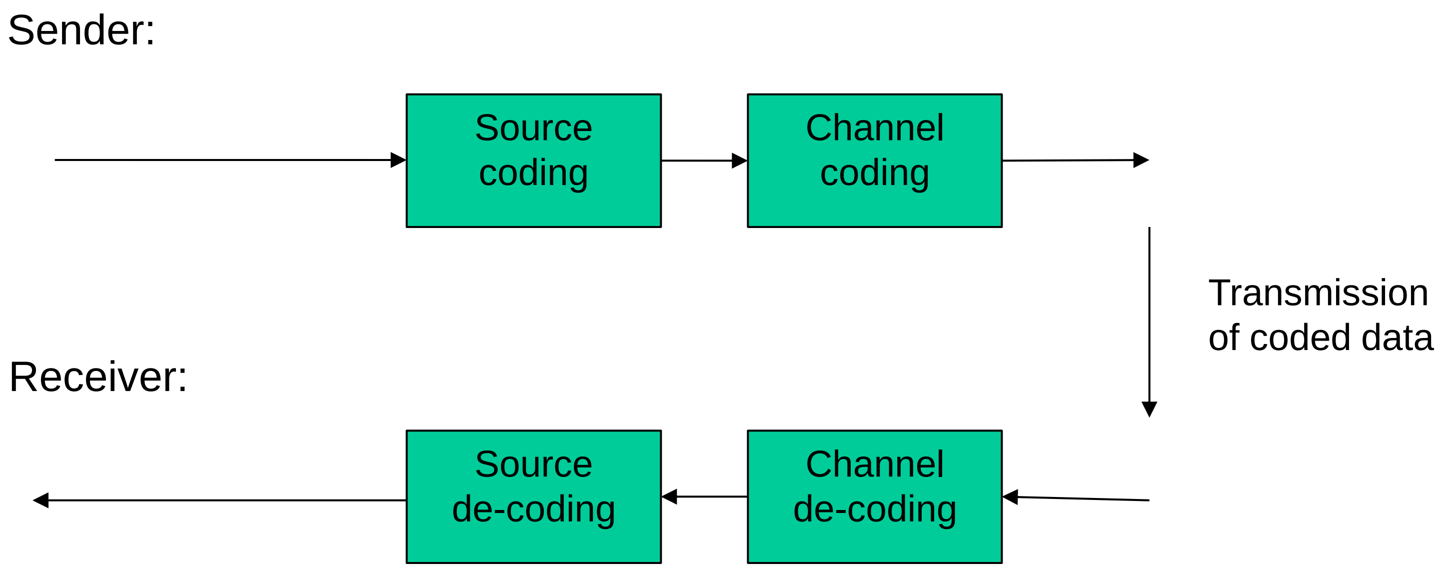
  - Motivation and Principle
  - **Predictive coding:**
    - Signal form coders
  - Audio quality measures

  Part II:

  - Two other types of **predictive coders**:
    - Vocoder and Hybrid coders
  - **Frequency domain / sub-band coders:**
    - MP3 and AAC coders of MPEG2 and MPEG4 standards

  - All including detailed motivation, analysis of coding principles and the descriptions of selected standards.

Coding – Decoding - Principles



Audiocoding is a Source Coding method!

Transmission of samples and quantized audio data

TECHNISCHE UNIVERSITÄT DARMSTADT

Direct PCM (pulse code modulation) coding:

- Quantize each sample by 8 - 16 Bit

- **Telephone speech** => 8 kHz sampling with 8 Bit/sample => 64 kBit / sec (ISDN coding) max freq sinal = 4kHz

- **Wideband speech** => 16 kHz sampling with 8 Bit/sample => 128 kBit / sec

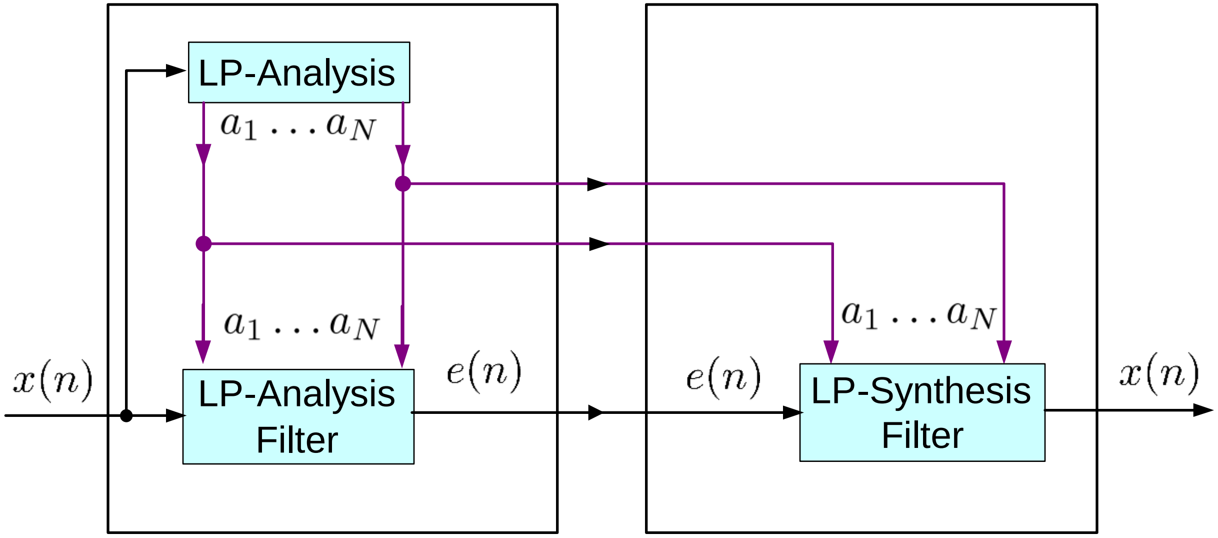
- **Audio data** 16 bit / sample (SNR = 8*16 = 96 dB SNR, i.e., signal to quantization noise) :

1) 16 kHz	sample rate:	256 kBit / sec
2) 22.05 kHz	sample rate:	352.8 kBit / sec
3) 44.1 kHz	sample rate (CD):	705.6 kBit / sec

- => Demand for data reduction for signal transmission and storage!

Model based predictive coding

- General principle: Model based predictive coding



- Typically, three classes of predictive coders:
  - Signal form coder
  - Vocoder
  - Hybrid coder

Main differences:
- Quantization of residual signal e(n)
- Form of the prediction

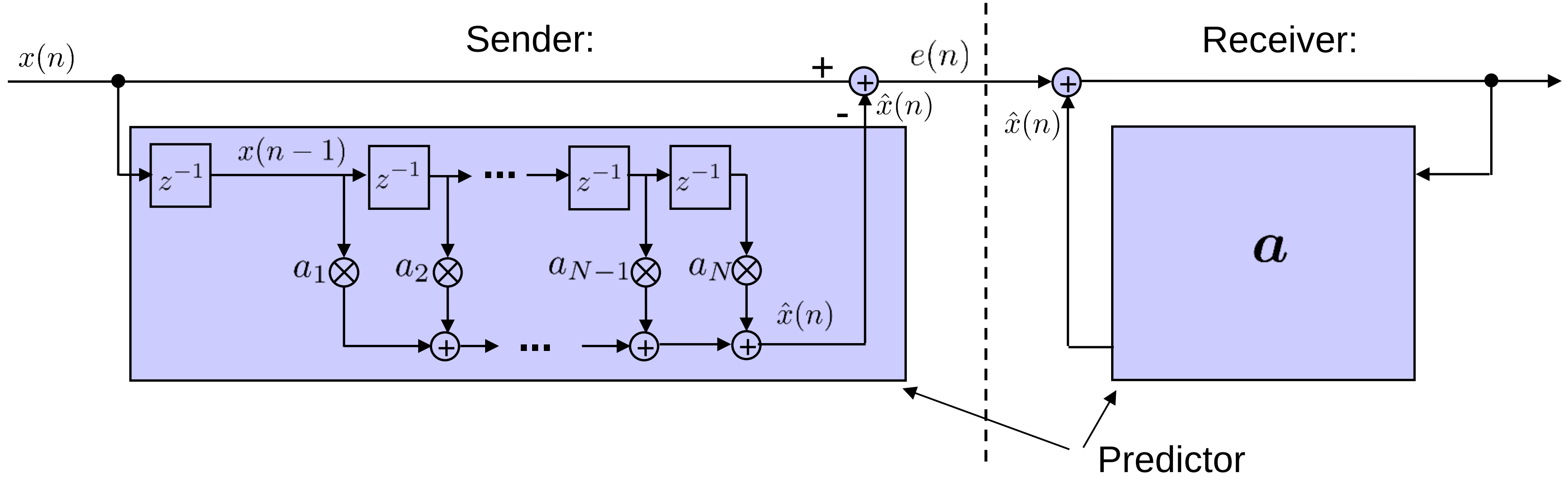
Signal form coder

Differential signal form coding

- High quality audio coders, suitable also for music, data rates > 1.5 Bit / sample

1) Differential pulse code modulation (DPCM):
Classic structure: **open-loop** prediction:

Talvez precise de menos bits para o erro



Open-loop structure requires the transmission of the prediction coefficients $\boldsymbol{a}(n)$ as side information.

TECHNISCHE UNIVERSITÄT DARMSTADT

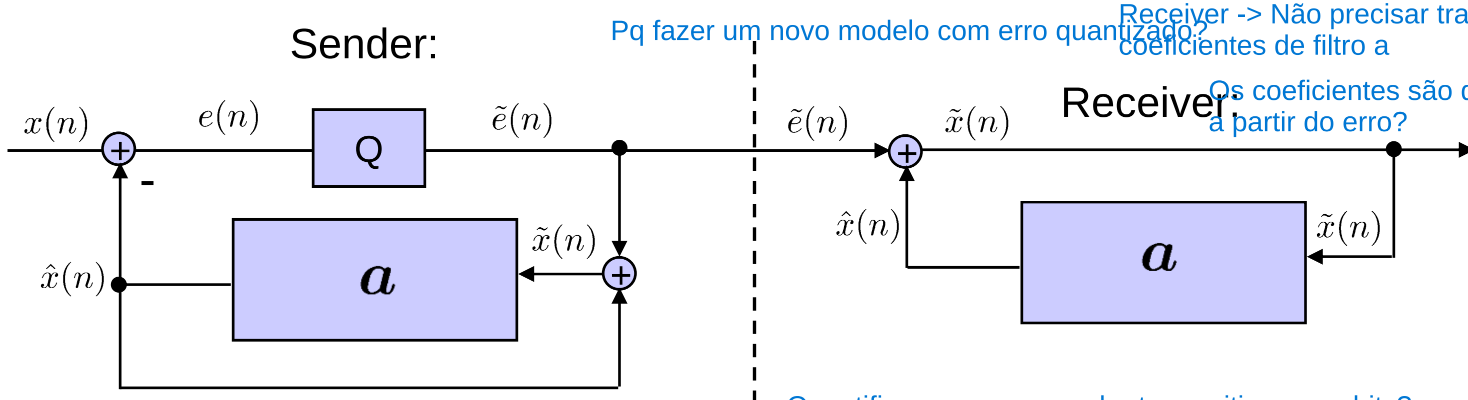
1) Differential pulse code modulation (DPCM):

Alternative structure: **closed-loop** prediction:

Ter o mesmo modelo no Sender e Receiver -> Não precisar transmitir os coeficientes de filtro a

Pq fazer um novo modelo com erro quantizado?

Os coeficientes são determinados a partir do erro?



Quantificar o erro pra poder transmitir menos bits?
Mas o sinal x não vai ser quantizado por natureza?

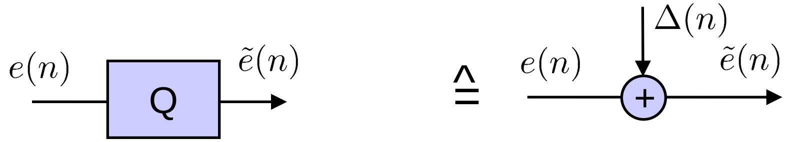
a) Without quantization: open-loop and closed loop structures are identical:

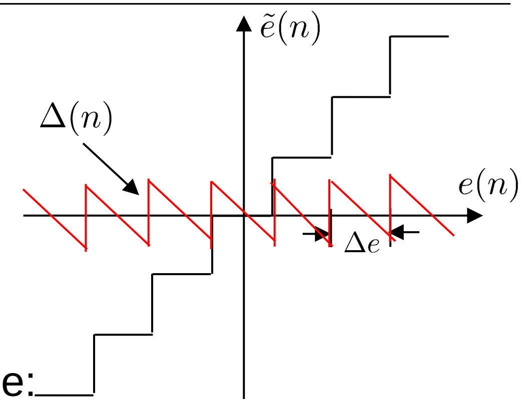
$$\tilde{x}(n) = \hat{x}(n) + e(n) = \hat{x}(n) + (x(n) - \hat{x}(n)) = x(n)$$

b) No transmission of side information is necessary since the quantization block is placed such that both predictor units (sender / receiver) work on the same signal.

Quantization noise analysis

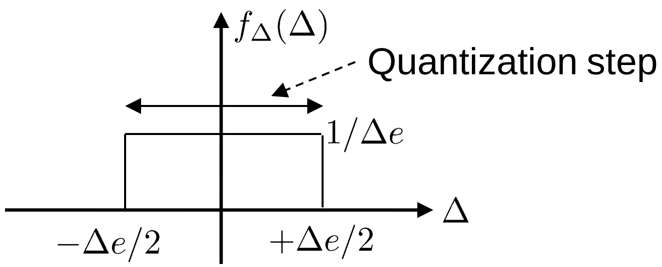
- Quantizer can be modeled by additive noise:



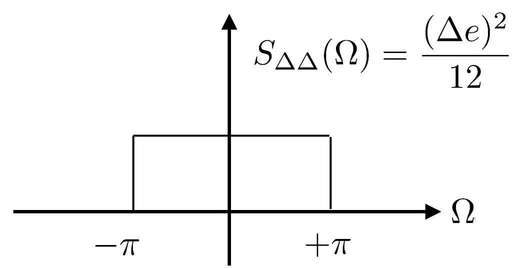


- Model for quantization noise: Equally distributed, white noise:

Equal distribution:



White noise:



Noise power:

$$\sigma^2_{\Delta e} = \int_{-\Delta e/2}^{-\Delta e/2} f_{\Delta}(\Delta)\Delta^2 d\Delta = \frac{(\Delta e)^2}{12}$$

Quantization noise analysis

- Quantization SNR analysis:

max. amplitude

$$SNR_{dB} = 10\,\log_{10}\left(\frac{S_{ee}(\Omega)}{S_{\Delta\Delta}(\Omega)}\right)$$

with: $S_{\Delta\Delta}(\Omega) = \dfrac{(\Delta e)^2}{12}$ $\quad \Delta e = \dfrac{2\,e_{\max}}{2^W}$

$S_{ee}(\Omega) = K\,e_{\max}^2$ $\quad W$: number of Bit

$K \leq 1$

degree of saturation

Pode ser usado como um limiar para o projeto

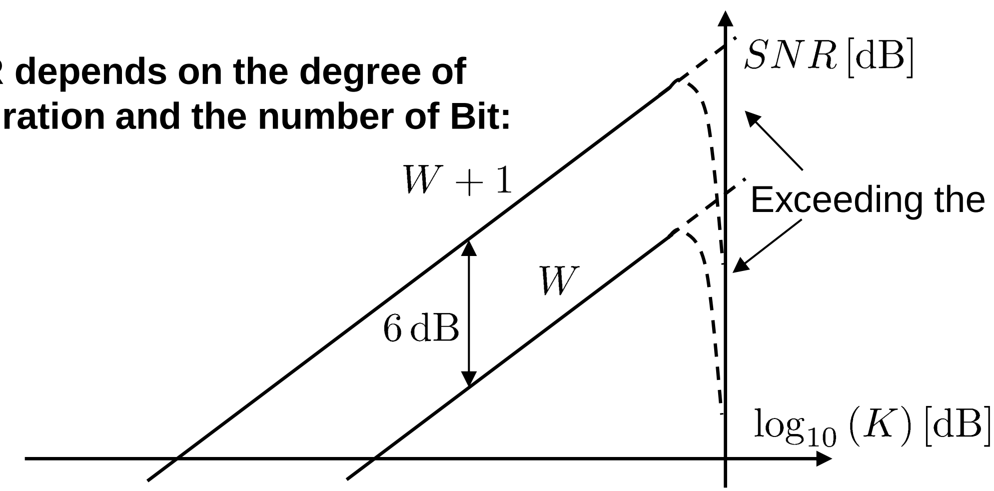
results in:

$$SNR_{dB} = 10\,\log_{10}\left(\frac{K\,e_{\max}^2}{\frac{(\Delta e)^2}{12}}\right) = 10\,\log_{10}\left(\frac{K\,e_{\max}^2\,12}{\frac{(2\,e_{\max})^2}{2^{2W}}}\right)$$

$$= 10\,\log_{10}\left(3\,K\,2^{2W}\right) = 6.02\,W + 10\,\log_{10}\left(3\,K\right)$$

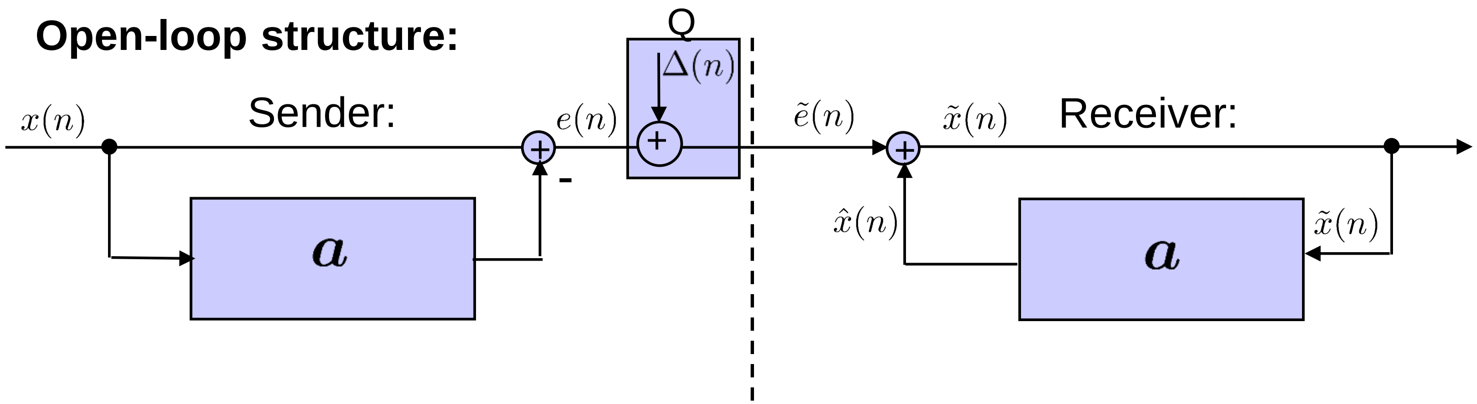
Se vc definir sua quantização no emax, vc precisa ter uma boa noção de quanto vai ser o seu emax antes de construir o seu sistema

SNR depends on the degree of saturation and the number of Bit:

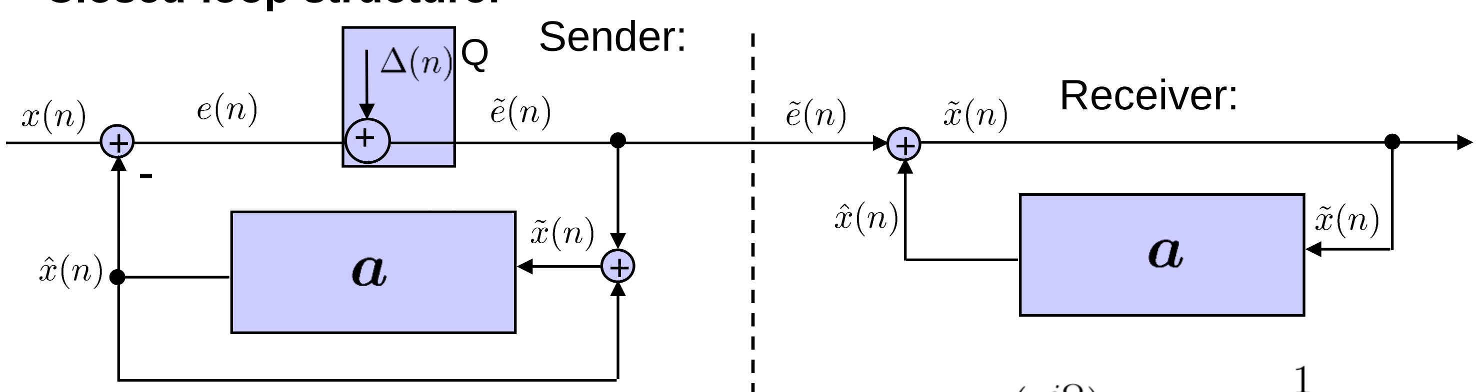


Differences in quantization

- **Open-loop structure:**



- **Closed-loop structure:**

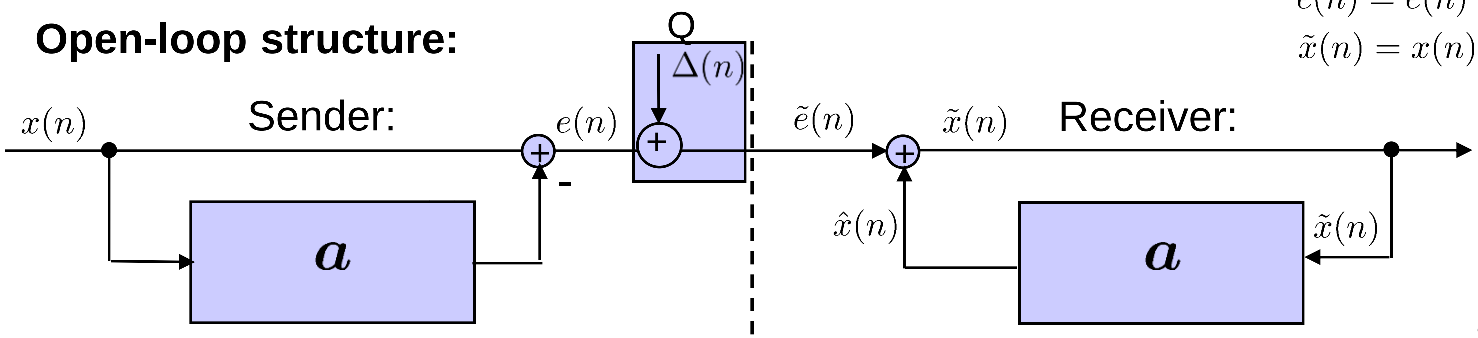


$$H_{\text{inv. PEF}}\left(e^{j\Omega}\right) = \frac{1}{1 - \sum_{i=1}^{N} a_i \, e^{-j\Omega i}}$$

Differences in quantization

$$\tilde{e}(n) = e(n) + \Delta(n)$$
$$\tilde{x}(n) = x(n) + r(n)$$

- **Open-loop structure:**



$$H_{\text{inv. PEF}}\left(e^{j\Omega}\right) = \frac{1}{1 - \sum_{i=1}^{N} a_i\, e^{-j\Omega i}}$$

- Quantization noise at the receiver output: $r(n)$

$$S_{rr}(\Omega) = S_{\Delta\Delta}(\Omega)\,\left|H_{\text{inv. PEF}}\left(e^{j\Omega}\right)\right|^2 = \frac{(\Delta e)^2}{12}\,\left|H_{\text{inv. PEF}}\left(e^{j\Omega}\right)\right|^2$$

$$H_{\text{inv. PEF}}\left(e^{j\Omega}\right) = \frac{1}{1 - A\left(e^{j\Omega}\right)}$$

$$S_{xx}(\Omega) = S_{ee}(\Omega)\,\left|H_{\text{inv. PF}}\left(e^{j\Omega}\right)\right|^2$$

O sinal branco fica colorido com a multiplicação por H_inf

- At the output, the quantization noise is spectrally shaped as the target signal:

  - No gain in signal to quantization noise relation: $\dfrac{S_{xx}(\Omega)}{S_{rr}(\Omega)} = \dfrac{S_{ee}(\Omega)}{S_{\Delta\Delta}(\Omega)}$

  - Identical spectral shapes lead to optimized noise masking

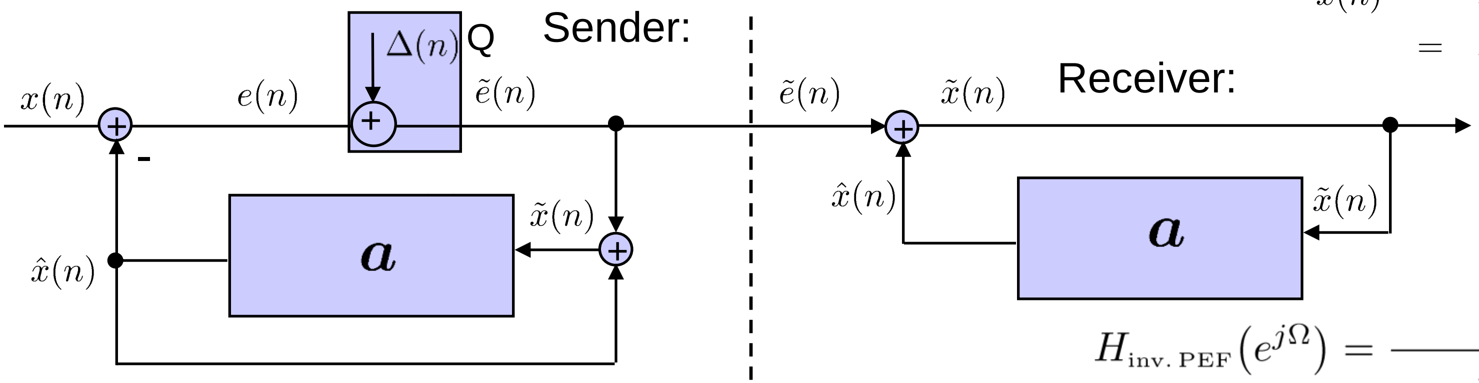
Differences in quantization

Ajuda a não "colorir" o ruido de quantização

- **Closed-loop structure:**

In the closed loop case:

$$\tilde{x}(n) = x(n) + r(n) = x(n) + \Delta(n)$$



$$H_{\text{inv. PEF}}\left(e^{j\Omega}\right) = \frac{1}{1 - \sum_{i=1}^{N} a_i\, e^{-j\Omega i}}$$

- Quantization noise at the receiver output:

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

=> 1) White noise at the receiver output: $\Delta(n)$ $\quad S_{rr}(\Omega) = S_{\Delta\Delta}(\Omega) = \frac{(\Delta e)^2}{12}$

2) No spectral shaping (and masking according to the input signal)

3) SNR gain according to the prediction gain:

$$\frac{S_{xx}(\Omega)}{S_{rr}(\Omega)} = \frac{S_{ee}(\Omega)}{S_{\Delta\Delta}(\Omega)} \left|H_{\text{inv. PEF}}\left(e^{j\Omega}\right)\right|^2 \qquad H_{\text{inv. PEF}}\left(e^{j\Omega}\right) = \frac{1}{1 - A\left(e^{j\Omega}\right)}$$

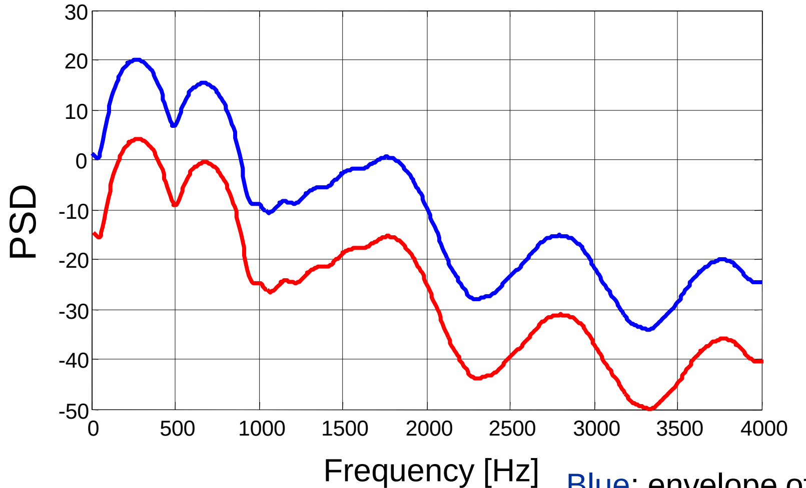
Comparing target and quant. noise PSDs at receiver output

TECHNISCHE UNIVERSITÄT DARMSTADT

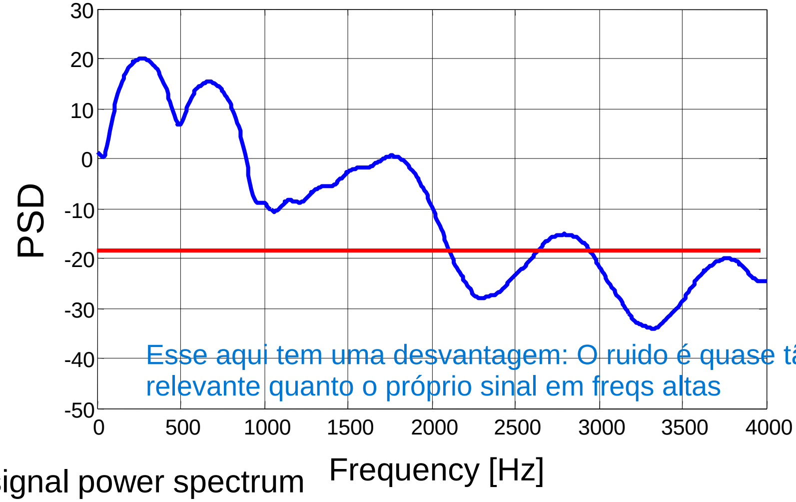
Mas e o contra-ganho? A estrutura closed-loop dificulta o processo de codificação na entrada?

❑ Open-loop structure:

Mas nosso ouvido é mais sensivel em baixas frequencias



❑ Closed-loop structure:



Esse aqui tem uma desvantagem: O ruido é quase tão relevante quanto o próprio sinal em freqs altas

Blue: envelope of input signal power spectrum
Red: quant. noise power spectrum

Quantization noise at the output shaped according to the target signal power spectrum

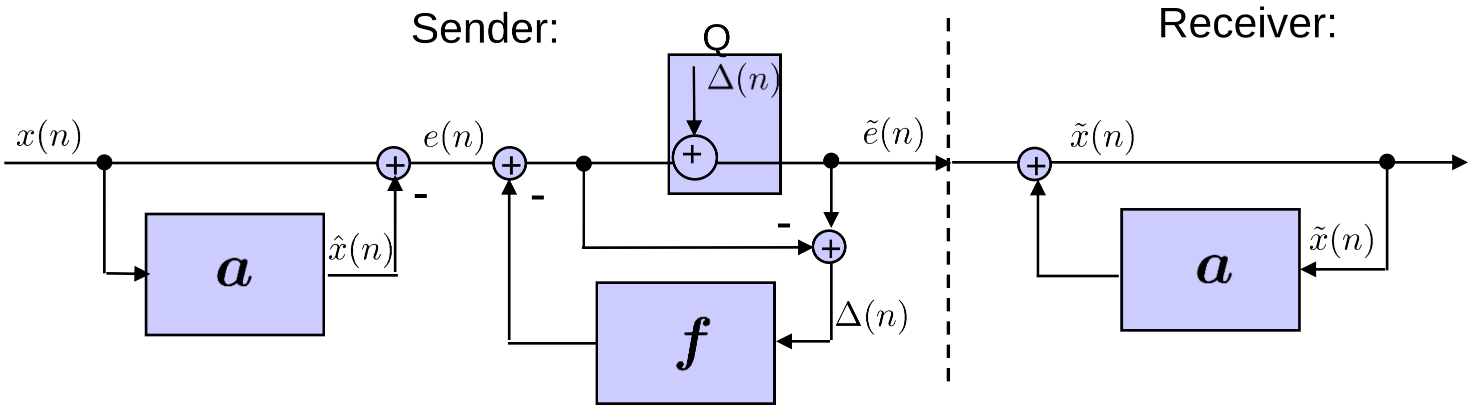
+: good masking
- : no advantage due to prediction

White quantization noise at the output

- : no masking
+ : signal to noise gain according to the prediction gain

Optimized power and shape of quantization noise PSDs

- **Compromized structure between open and closed loop:**



$$E(z) = X(z)\,(1 - A(z))$$

$$\tilde{E}(z) = E(z) + \Delta(z)\,(1 - F(z)) = X(z)\,(1 - A(z)) + \Delta(z)\,(1 - F(z))$$

$$\tilde{X}(z) = X(z) + \underbrace{\Delta(z)\,\frac{1 - F(z)}{1 - A(z)}}_{R(z)}$$

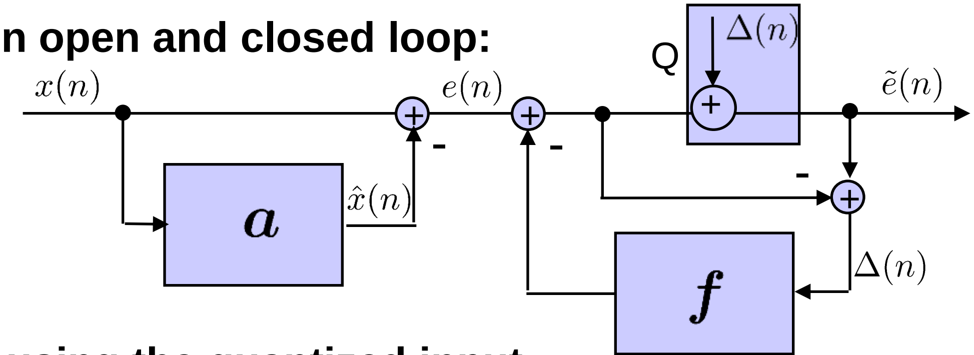
$R(z)$: quantization noise at the output

$F(z) = A(z)$ => White output quantization noise => identical to closed-loop structure

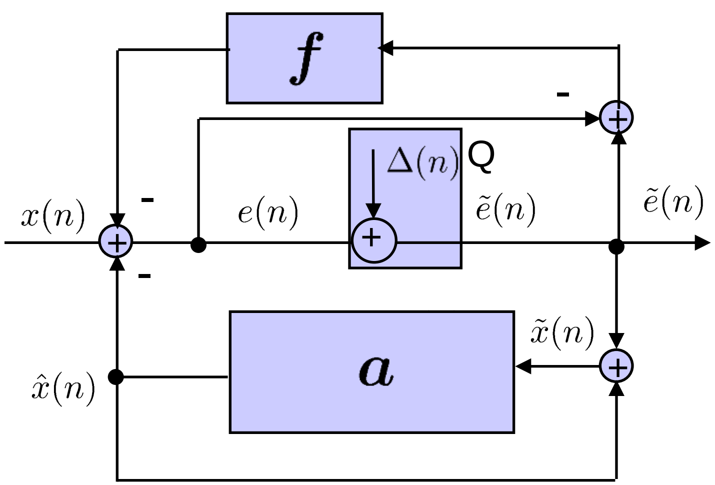
$F(z) = 0$ => Input signal shaped quantization noise => identical to open-loop structure

Optimized power and shape of quantization noise PSDs

- **Compromized structure between open and closed loop:**



- **Comparable structure with a(n) using the quantized input**



$$\tilde{X}(z) = \hat{X}(z) + \tilde{E}(z) = A(z)\,\tilde{X}(z) + \tilde{E}(z) = \frac{\tilde{E}(z)}{1 - A(z)}$$

$$\hat{X}(z) = \tilde{X}(z) - \tilde{E}(z) = \tilde{E}(z)\frac{A(z)}{1 - A(z)}$$

$$E(z) = X(z) - \hat{X}(z) - F(z)\,\Delta(z)$$

$$\boxed{\tilde{X}(z) = X(z) + \Delta(z)[1 - F(z)]}$$

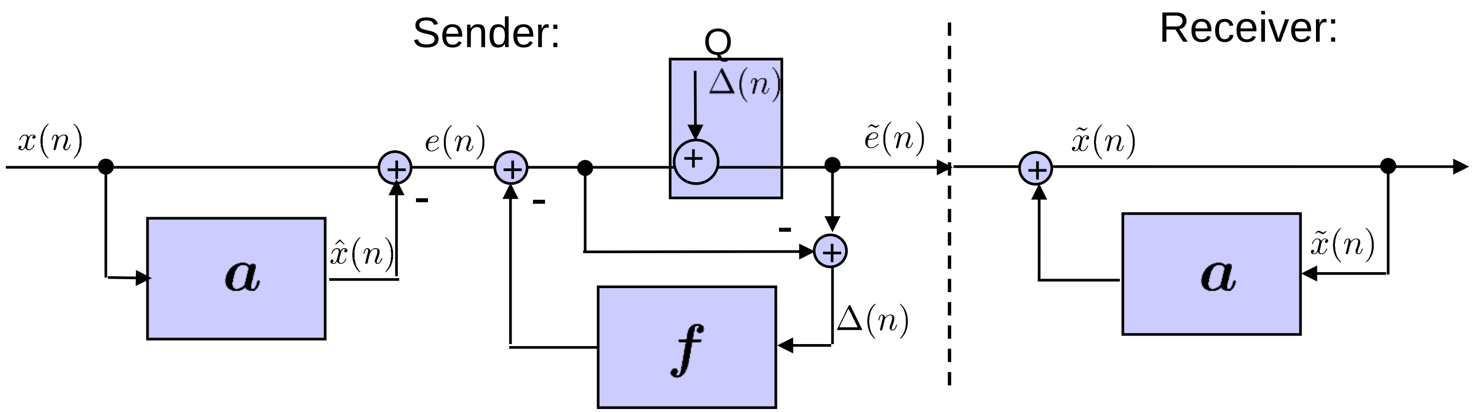
Derivation and more infos: s. Appendix

Choose closed-loop structure: $F(z) = 0$

Choose open-loop structure: $F(z) = \dfrac{-A(z)}{1 - A(z)}$

Optimized power and shape of quantization noise PSDs

- **Compromized structure between open and closed loop:**



- **Continuous switching between both structures:**

$F(z) = A(z/\gamma)$ with: $0 \le \gamma \le 1$

$$\tilde{X}(z) = X(z) + \Delta(z)\,\frac{1 - F(z)}{1 - A(z)}$$

Numerator polynom:

gamma = 1 -> closed-loop

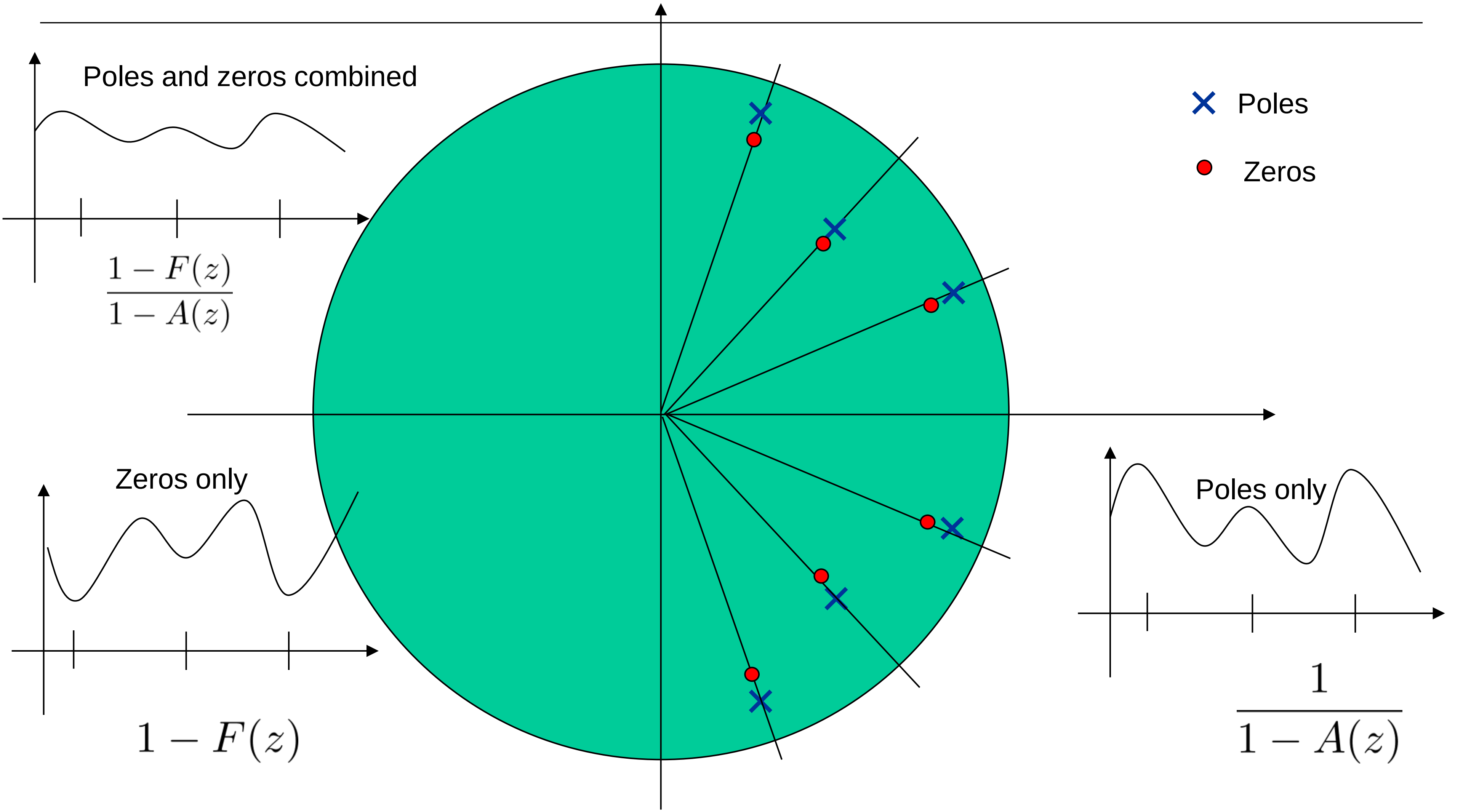
$$1 - F(z) = 1 - A(z/\gamma) = \frac{1}{z^n}\prod_{i=1}^{n}(z - \gamma\, z_{0i})$$ with: $0 \le \gamma \le 1$

γ : shifts the zeros to the origin of the unit-circle

=> some attenuation by the zeros but less than the amplification by the poles

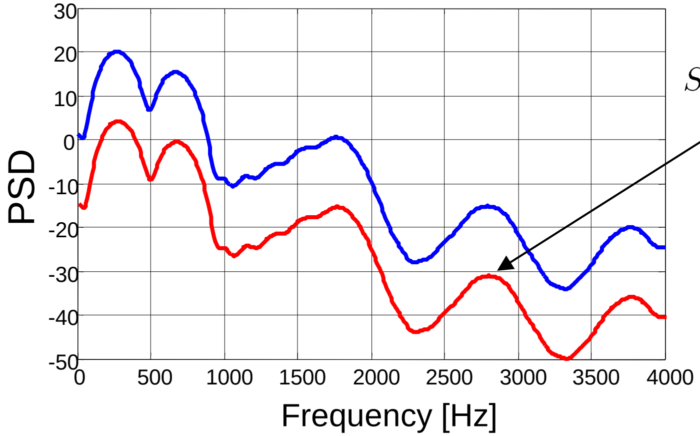
=> no more a flat response but no max. shaping as in the open loop case.

Pole / Zero Filtering



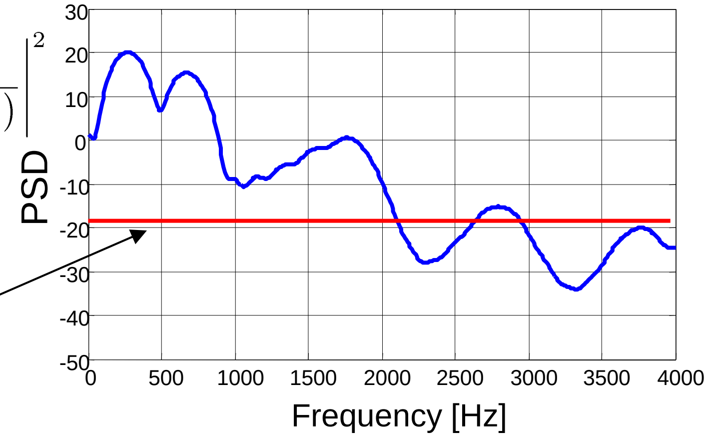
Optimized power and shape of quantization noise PSDs

- **Open-loop structure:**



$$S_{rr}(\Omega) = \frac{(\Delta e)^2}{12} \left| \frac{1}{1 - A\left(e^{j\Omega}\right)} \right|^2$$

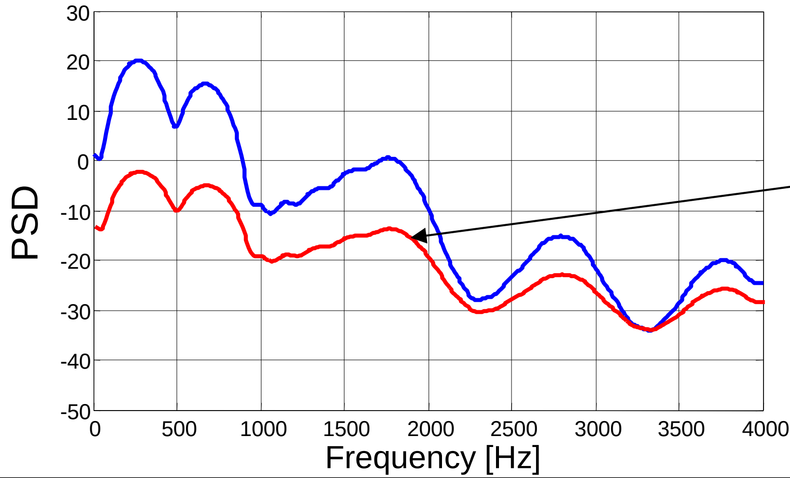
- **Closed-loop structure:**



$$S_{rr}(\Omega) = \frac{(\Delta e)^2}{12}$$

Blue: input signal power spectrum
Red: quant. noise power spectrum

- **Optimized / compromized noise power shape structure:**

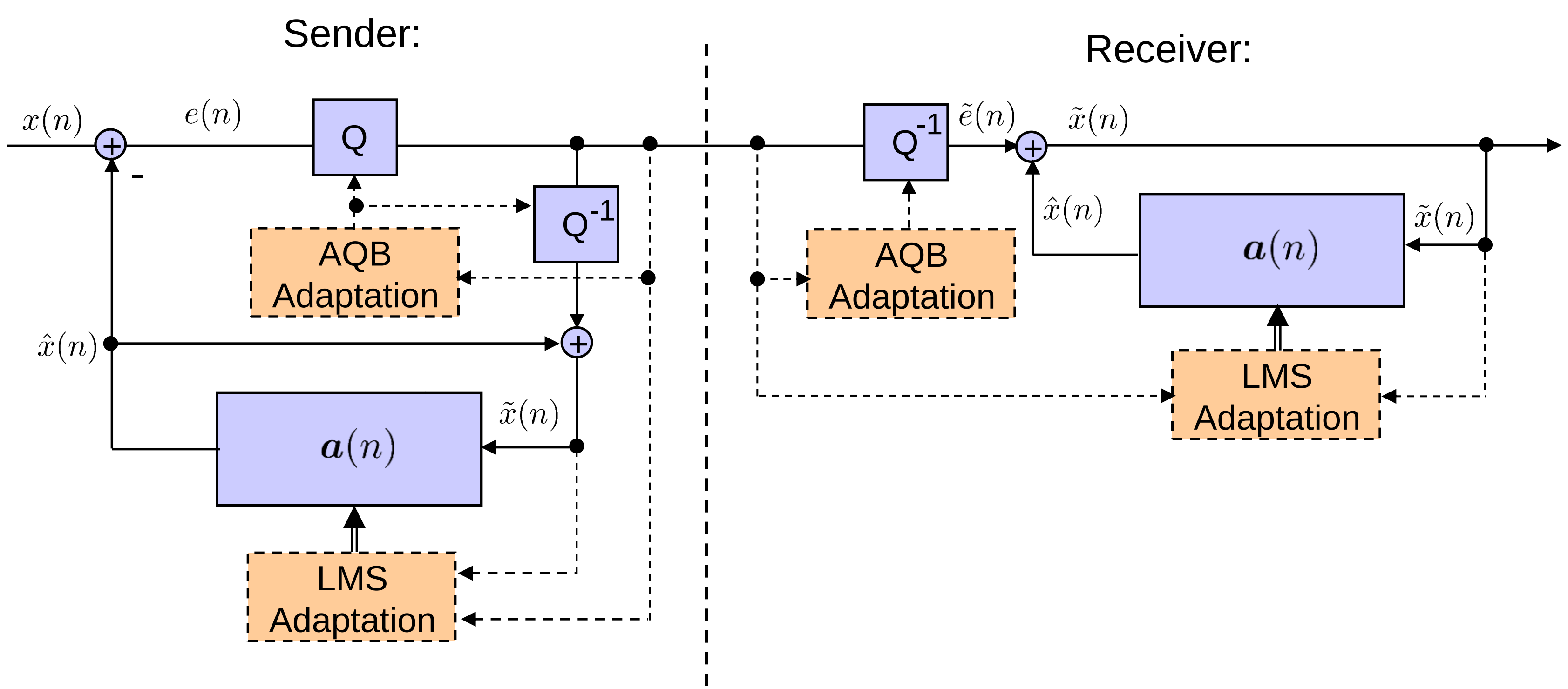


$$S_{rr}(\Omega) = \frac{(\Delta e)^2}{12} \left| \frac{1 - A\left(e^{j\Omega}/\gamma\right)}{1 - A\left(e^{j\Omega}\right)} \right|^2$$

with: $0 \leq \gamma \leq 1$

ADPCM structures

- **ADPCM (adaptive differential PCM) vs. DPCM:**
  So far (DPCM) the quantizer as well as the prediction filter have been assumed to be fix and time-independent.
  In the ADPCM structure both are chosen to be adaptively time dependent:

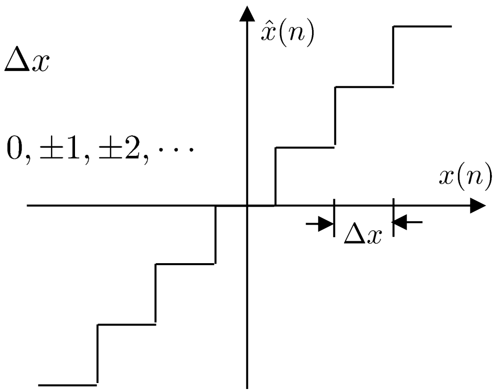


Adaptive quantization

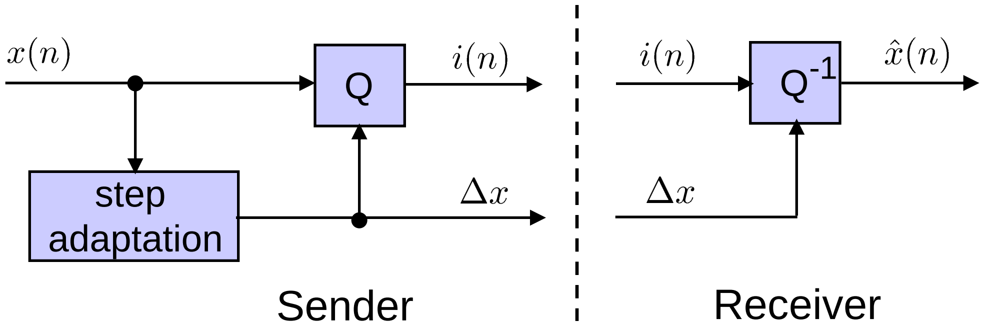
- **Adaptive quantization:**

$\hat{x}(n) = i(n)\,\Delta x$

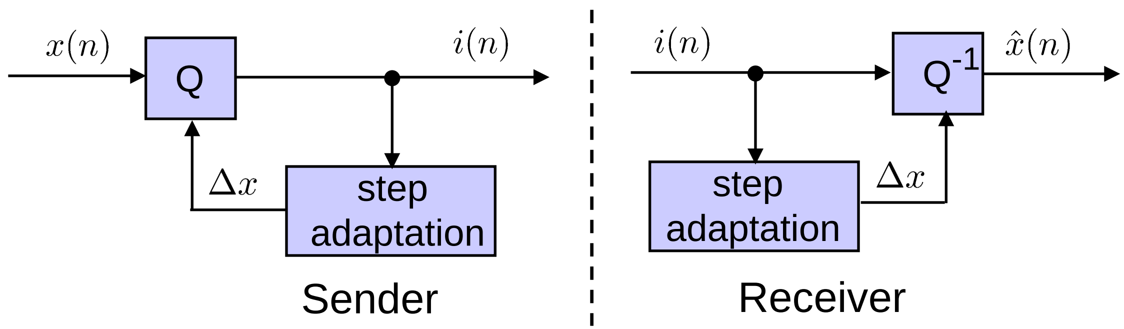
$i(n) = 0, \pm 1, \pm 2, \cdots$



  - AQF: “Adaptive quantization forward”



  - AQB: “Adaptive quantization backward”: no side information (quantization step) necessary



Adaptive quantization

- Quantization step should be chosen proportional to the estimated input signal standard deviation (square root of the power): $\hat{\sigma}_x(n)$

$$\Delta x(n) = c\,\hat{\sigma}_x(n) \quad \text{with: } c = \text{const.} \qquad \hat{x}(n) = i(n)\,\Delta x(n)$$

quantized bit value

- Recursive estimation of the input signal power for the AQB method:

$$\hat{\sigma}_x^2(n) = \alpha\,\hat{\sigma}_x^2(n-1) + (1-\alpha)\,\hat{x}^2(n-1)$$

- Definition of the step multiplicator: $M(n)$

$$\Delta x(n) = M(n-1)\,\Delta x(n-1) \quad \Rightarrow \quad M(n-1) = \frac{\Delta x(n)}{\Delta x(n-1)} = \frac{\hat{\sigma}_x(n)}{\hat{\sigma}_x(n-1)}$$

with:
$$\begin{aligned} M^2(n-1) &= \frac{\hat{\sigma}_x^2(n)}{\hat{\sigma}_x^2(n-1)} = \alpha + (1-\alpha)\,\frac{\hat{x}^2(n-1)}{\hat{\sigma}_x^2(n-1)} \\ &= \alpha + (1-\alpha)\,\frac{i^2(n-1)\,\Delta x^2(n-1)}{\hat{\sigma}_x^2(n-1)} = \alpha + (1-\alpha)\,i^2(n-1)\,c^2 \end{aligned}$$

$$\boxed{\Rightarrow \quad M(n-1) = \sqrt{\alpha + (1-\alpha)\,i^2(n-1)\,c^2}}$$

can be calculated a priori
=> look up table

Adaptive quantization

- Three steps in an adaptive quantization:

1) Determine new quantization step

$$\Delta x(n) = M(n-1)\,\Delta x(n-1)$$

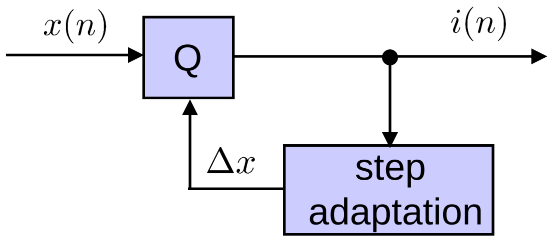
2) Determine quantized value: $i(n)$

$$\hat{x}(n) = i(n)\,\Delta x(n)$$

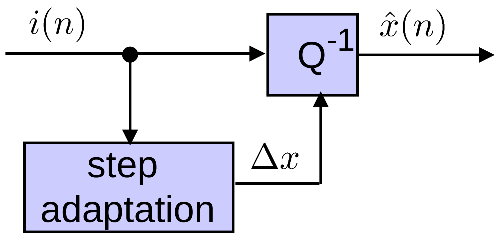
3) Determine next step multiplicator:

$$M(n) = \sqrt{\alpha + (1-\alpha)\,i^2(n)\,c^2}$$

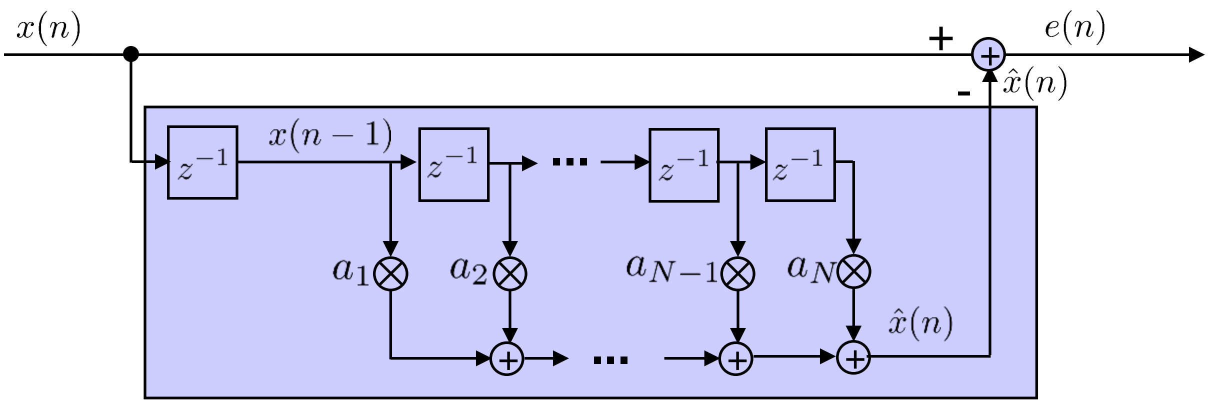
Sender:



Receiver:



LMS adaptation of the predictor



- Direct calculation of the „optimum filter“ (Wiener solution):

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & \cdots & r_{xx}(N-1) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}(N-1) & r_{xx}(N-2) & \cdots & r_{xx}(0) \end{bmatrix}^{-1} \begin{bmatrix} r_{xx}(1) \\ r_{xx}(2) \\ \vdots \\ r_{xx}(N) \end{bmatrix}$$

- Adaptive calculation of the optimum solution (continuous update):

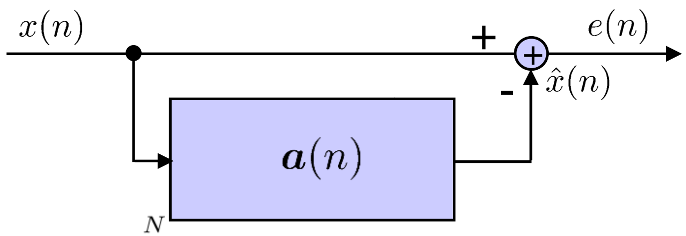
NLMS procedure:

$$\boldsymbol{a}(n+1) \;=\; \boldsymbol{a}(n) + \mu(n)\,\frac{\boldsymbol{x}(n-1)\,e(n)}{\left\|\boldsymbol{x}(n-1)\right\|^2}$$

step-size (between 0 and 1)

All pole / all zero prediction

All pole prediction:



$$H_{\mathrm{PEF}}\left(e^{j\Omega}\right) = 1 - \sum_{i=1}^{N} a_i\, e^{-j\Omega i}$$

$$H_{\mathrm{inv.\,PEF}}\left(e^{j\Omega}\right) = \frac{1}{1 - \sum_{i=1}^{N} a_i\, e^{-j\Omega i}}$$

$$e(n) = x(n) - \sum_{i=1}^{N} a_i\, x(n-i)$$

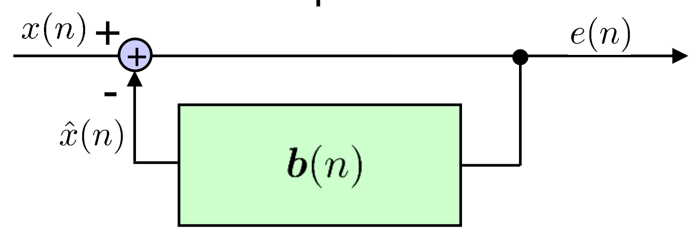
$$\mathrm{E}\left\{e^2(n)\right\} \to \min$$

$$\frac{\partial}{\partial a_j}\mathrm{E}\left\{e^2(n)\right\} = 2\mathrm{E}\left\{e(n)\,\frac{\partial}{\partial a_j}e(n)\right\} \stackrel{!}{=} 0$$

$\mathrm{E}\left\{e(n)\,x(n-j)\right\} \stackrel{!}{=} 0$ for all: $j = 1 \cdots N$

$$\boldsymbol{a}(n+1) \;=\; \boldsymbol{a}(n) + \mu(n)\,\frac{\boldsymbol{x}(n-1)\,e(n)}{\left\|\boldsymbol{x}(n-1)\right\|^2}$$

All zero prediction:



$$H_{\mathrm{PEF}}\left(e^{j\Omega}\right) = \frac{1}{1 + \sum_{i=1}^{M} b_i\, e^{-j\Omega i}}$$

$$H_{\mathrm{inv.\,PEF}}\left(e^{j\Omega}\right) = 1 + \sum_{i=1}^{M} b_i\, e^{-j\Omega i}$$

$$e(n) = x(n) - \sum_{i=1}^{M} b_i\, e(n-i)$$

$$\mathrm{E}\left\{e^2(n)\right\} \to \min$$

$$\frac{\partial}{\partial b_j}\mathrm{E}\left\{e^2(n)\right\} = 2\mathrm{E}\left\{e(n)\,\frac{\partial}{\partial b_j}e(n)\right\} \stackrel{!}{=} 0$$

$\mathrm{E}\left\{e(n)\,e(n-j)\right\} \stackrel{!}{=} 0$ for all: $j = 1 \cdots M$

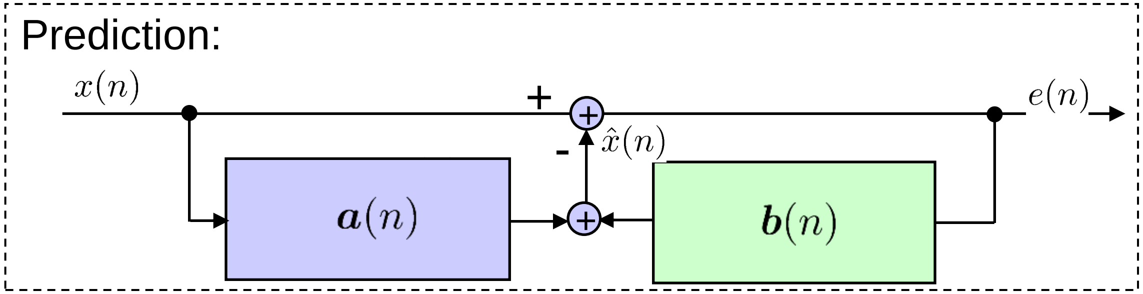
$$\boldsymbol{b}(n+1) \;=\; \beta\,\boldsymbol{b}(n) + \mu(n)\,\frac{\boldsymbol{e}(n-1)\,e(n)}{\left\|\boldsymbol{e}(n-1)\right\|^2}$$

Damping factor!

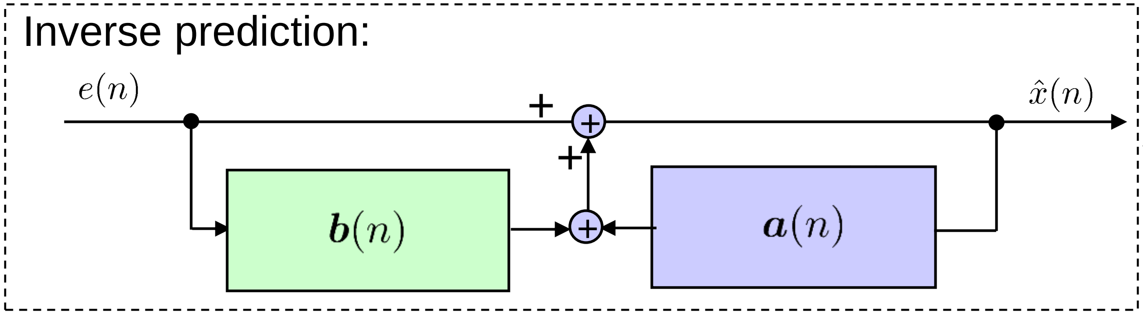
Pole-zero prediction

- Caution!
  The all zero model is an IIR filter which may become instable!
  Therefore, if used, typically only zero-prediction filters up to order 2 are applied. Their stability can be well controlled.

- Pole-zero prediction : $e(n) = x(n) - \sum_{i=1}^{M} b_i\, e(n-i) - \sum_{i=1}^{N} a_i\, x(n-i)$



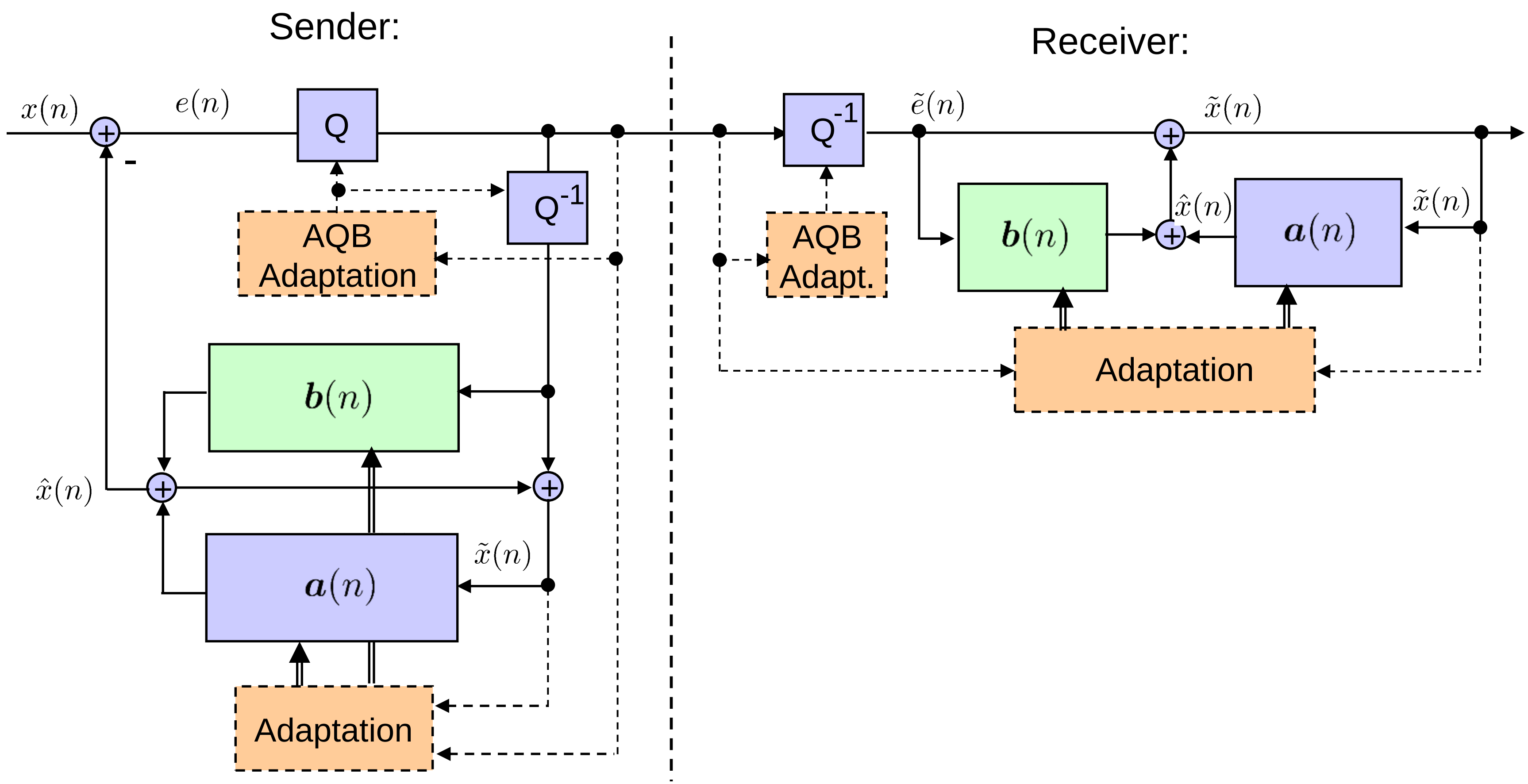
$$H_{\text{PEF}}(e^{j\Omega}) = \frac{1 - \sum_{i=1}^{N} a_i\, e^{-j\Omega i}}{1 + \sum_{i=1}^{M} b_i\, e^{-j\Omega i}}$$



$$H_{\text{inv. PEF}}(e^{j\Omega}) = \frac{1 + \sum_{i=1}^{M} b_i\, e^{-j\Omega i}}{1 - \sum_{i=1}^{N} a_i\, e^{-j\Omega i}}$$

ADPCM structures: The standardized ITU-T G.722 codec

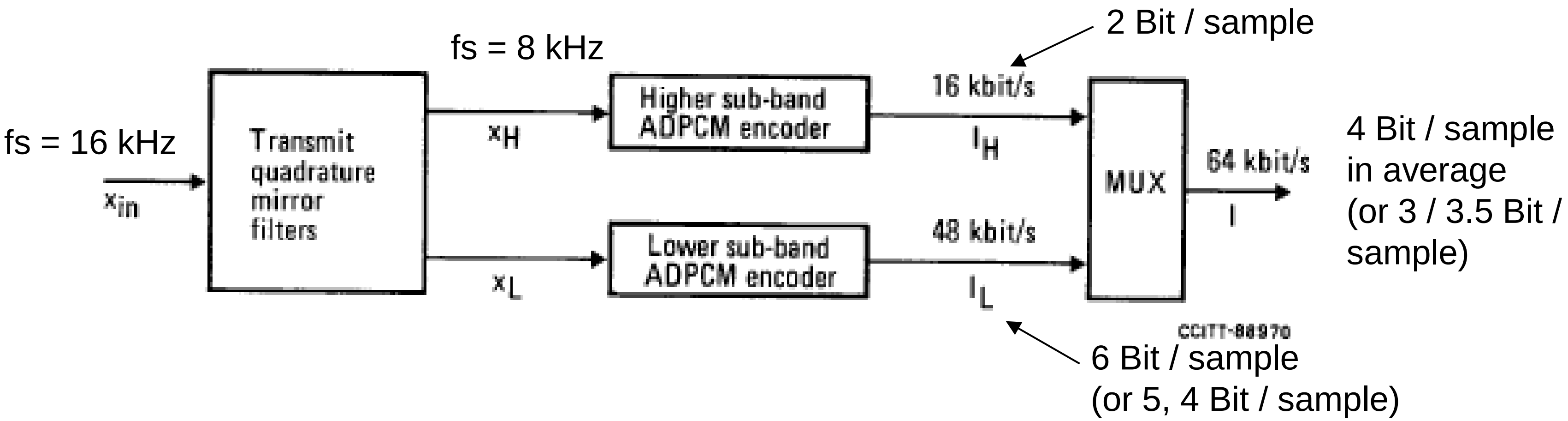
- Adaptive quantization and adaptive pole-zero prediction:



The standardized ITU-T G.722 codec

- G.722: high quality audio coder for speech and music.
- Input signal (sampling rate 16 kHz) split in a high and low-pass component:

  Coding according to the human perception (Bark-Scale!)
  => higher resolution of the low frequencies than for the high frequencies
  => maximizing the perceived audio quality by quantizing the components
  between 0-4 kHz with a better quantization (6 Bit / sample)
  than the components between 4-8 kHz (2 Bit / sample).
- Allows for a rate scaling in three modes:
  64 kBit/s = 48+16 kBit/s; 56 kBit/s = 40+16 kBit/s; 48 kBit/s = 32+16 kBit/s



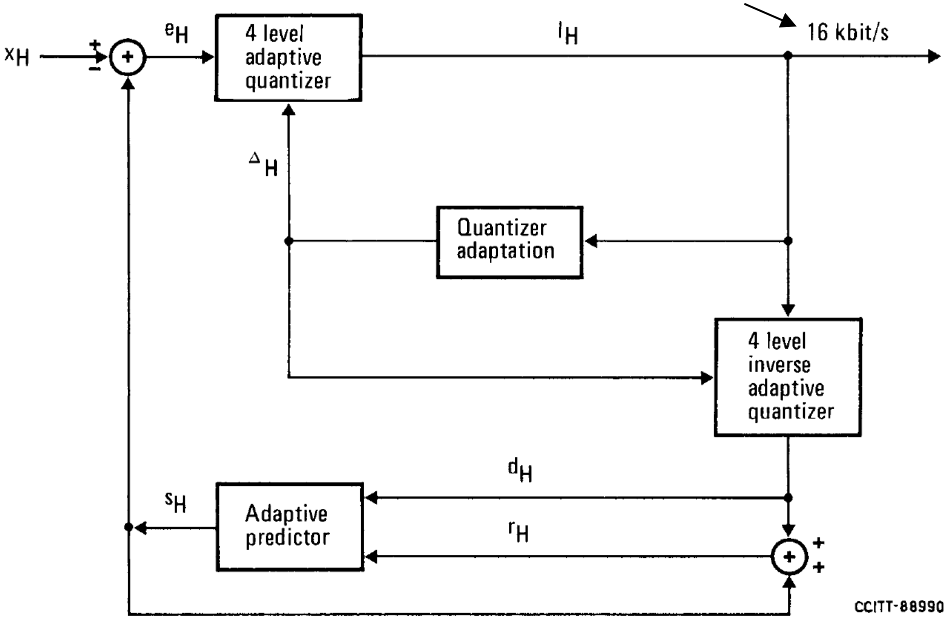
The standardized ITU-T G.722 codec: Coder (Sender)

- The classic ADPCM structure:
  => two Bit quantization of the
  residual signal

- The modified ADPCM structure:
  => up to six Bit quantization of the residual
  signal, predictor running at 4 Bit
  => allows for a data rate down-scale

Higher sub-band

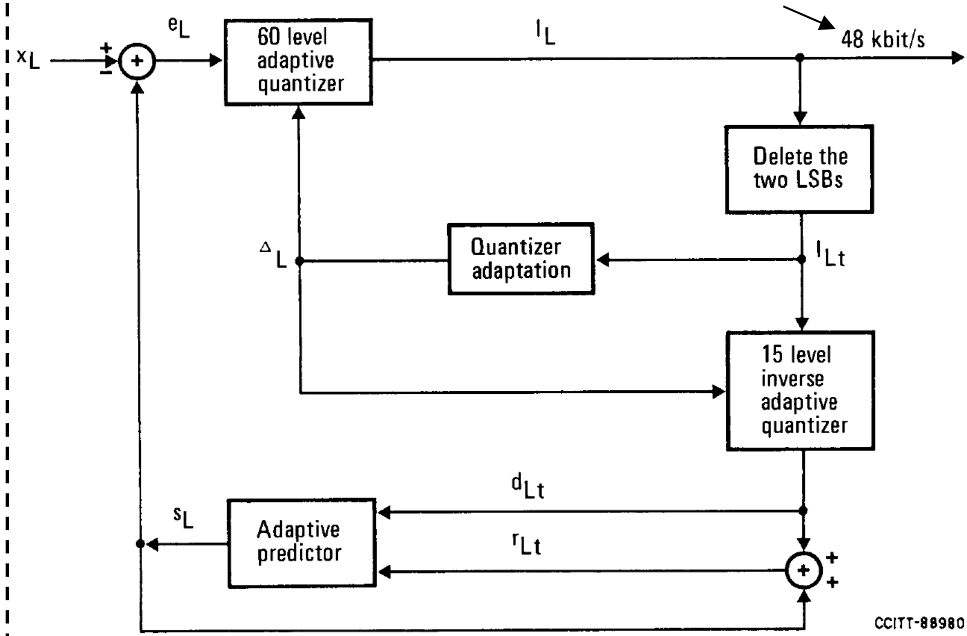
2 Bit / sample



Pole-zero prediction with: $N = 6;\ M = 2$

Lower sub-band

6Bit / sample



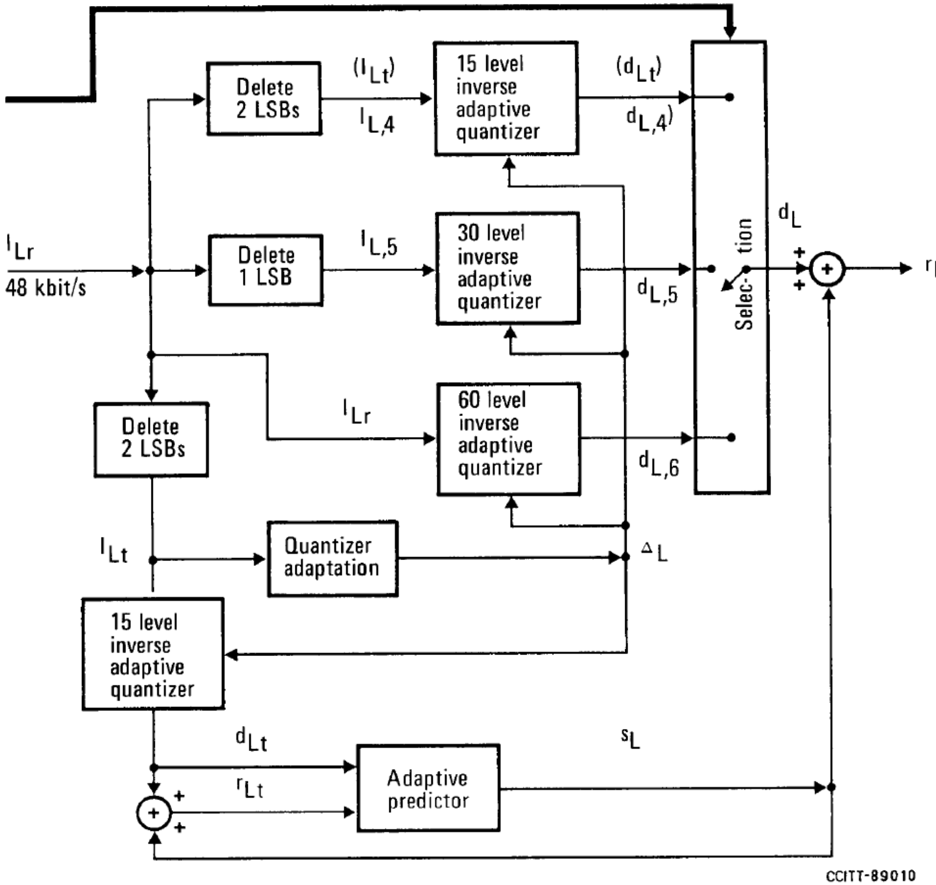
Pole-zero prediction with: $N = 6;\ M = 2$

The standardized ITU-T G.722 codec: Decoder (Receiver)

Lower sub-band

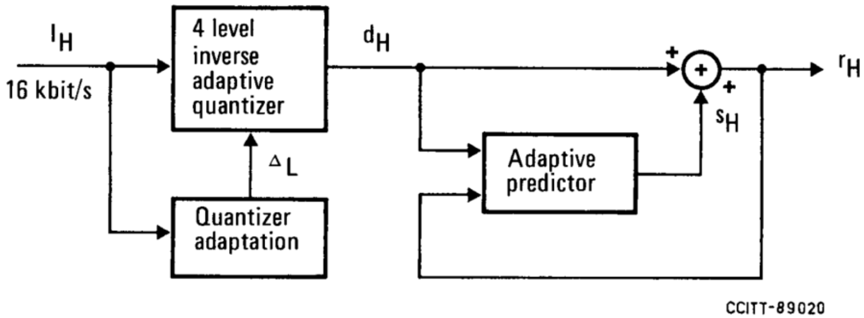
- The modified ADPCM structure:
  Internal predictor running at 4 Bit / sample
  Residual signal up to 6 Bit / sample

Mode indication: 64, 56, or 48 kBit/s



Higher sub-band

- The classic ADPCM decoder structure:
  => two Bit quantization of the
  residual signal



G.722 Characteristics

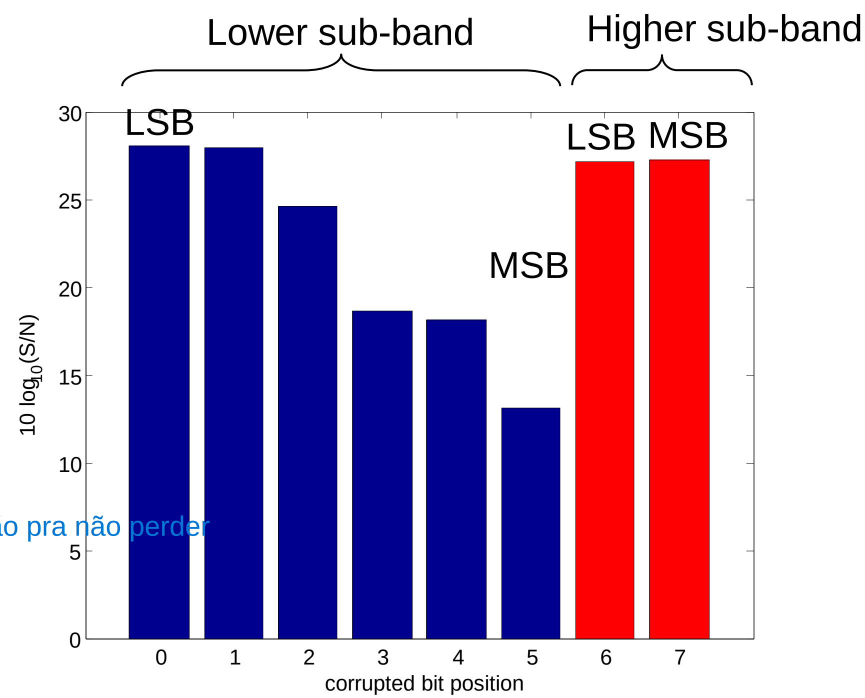
- **Sensitivity to bit errors:**

  Test by corruption of a specific bit position with a BER of 1%:

- 56 kBit Mode: Bit 0 is cancelled
  48 kBit Mode: Bit 0 + 1 are cancelled

- SNR is most sensitive to a corruption of Bit 4 + 5
  => have to be protected by channel coding

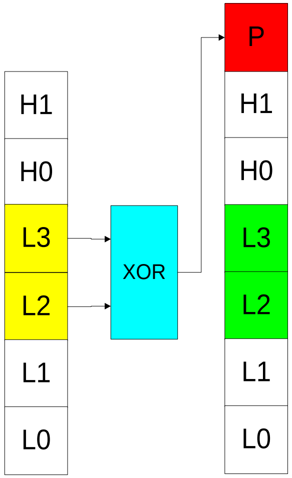
Talvez vc precise usar bits de amostragem como bits de verificação pra não perder



Channel Coding Approaches (48kBit mode)

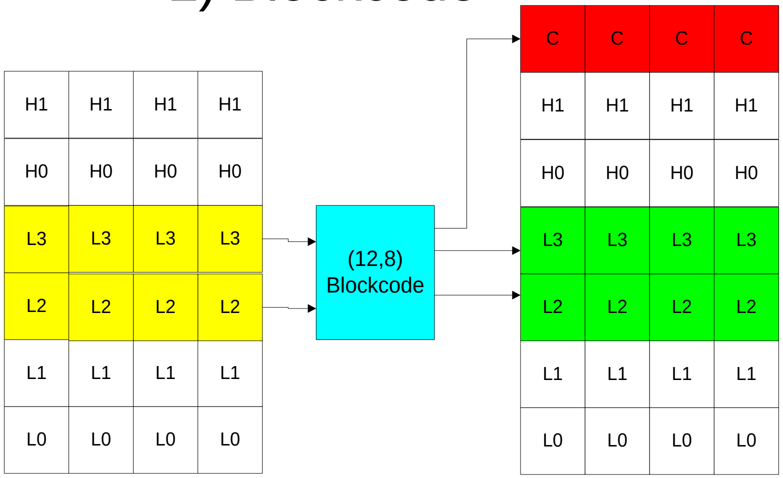
- **Three different possibilities for channel coding:**

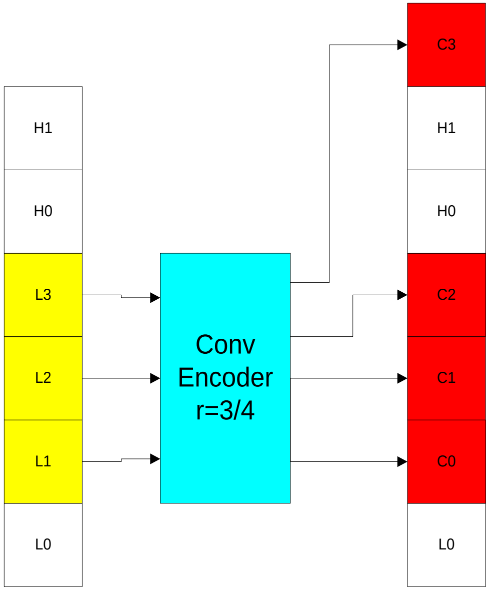
Não parece bom -> Não da pra corrigir



1) Parity

2) Blockcode





3) Convolutional Encoding

Comparison of channel coding strategies

$10 \cdot \log_{10}(E_S/N_0)$	BER_{uncod}	No protection	Parity	Block	Conv.
7.3	$\sim 10^{-2}$				
8.5	$\sim 4 \cdot 10^{-3}$				
9.8	$\sim 10^{-3}$				

Subjective and objective quality evaluation methods

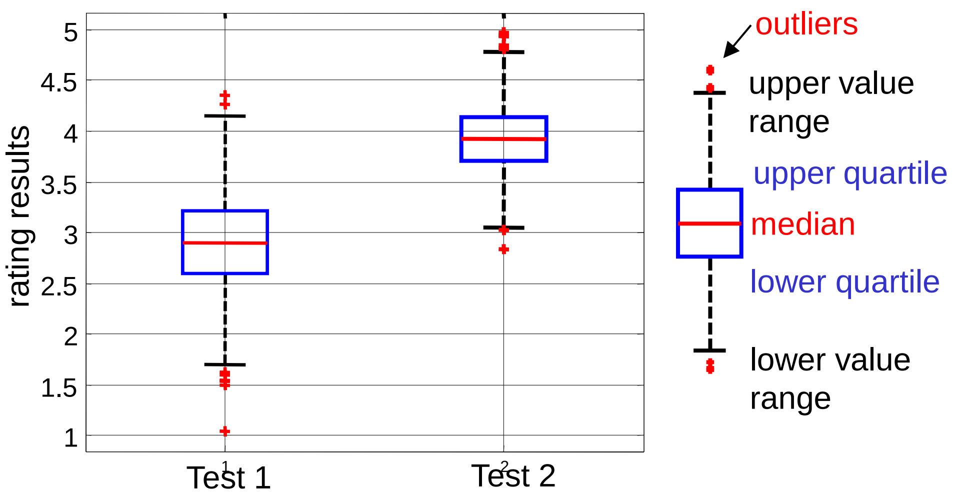
- **Long history in subjective quality ratings:**
  - Assessment of telephone band systems, e.g., speech codecs.

- First type of subjective tests: **Subjective absolute rating**

  - 1993: **Absolute category rating (ACR) test method (ITU-T P.800)**
  - User is regarded to have a reference of a telephone signal "in the mind"
  - Rating according to an absolute scale:

Impairment	Grade
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

Resulting in a **MOS (mean opinion score)** rating when asking a significant number of test subjects



Subjective and objective quality evaluation methods

- Second type of subjective tests: **Subjective relative rating**
  - 1994/97: Procedure for quality rating of wide-band coded audio signals
  - Comparison with reference signal (anchor)
  - **ITU-R BS.1116**:
    "Methods for the Subjective Assessment of small Impairments in Audio Systems including Multichannel Sound Systems"
  - Double-blind triple-stimulus with hidden reference => reference is hidden within the test signals as one test signal.

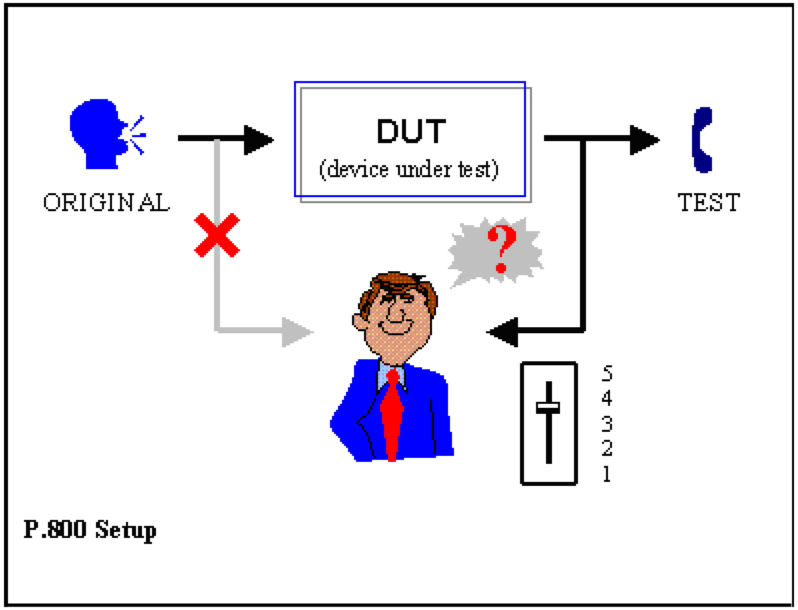
Impairment	**Grade**	**SDG**
Imperceptible	5	0
Perceptible, not annoying	4	-1
Slightly annoying	3	-2
Annoying	2	-3
Very annoying	1	-4

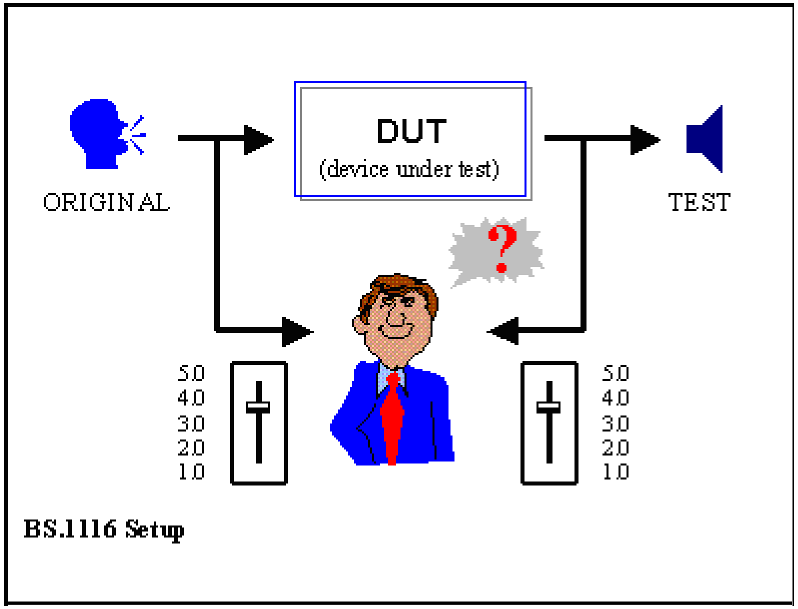
SDG: subjective difference grade

SDG = Grade of test signal –
Grade of hidden reference

Subjective and objective quality evaluation methods

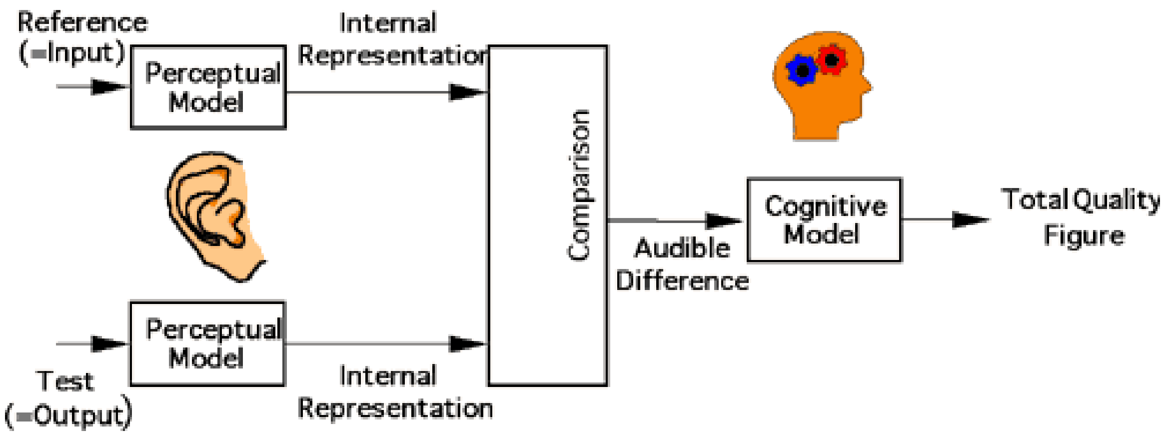
- **Different setups of subjective tests in an overview:**





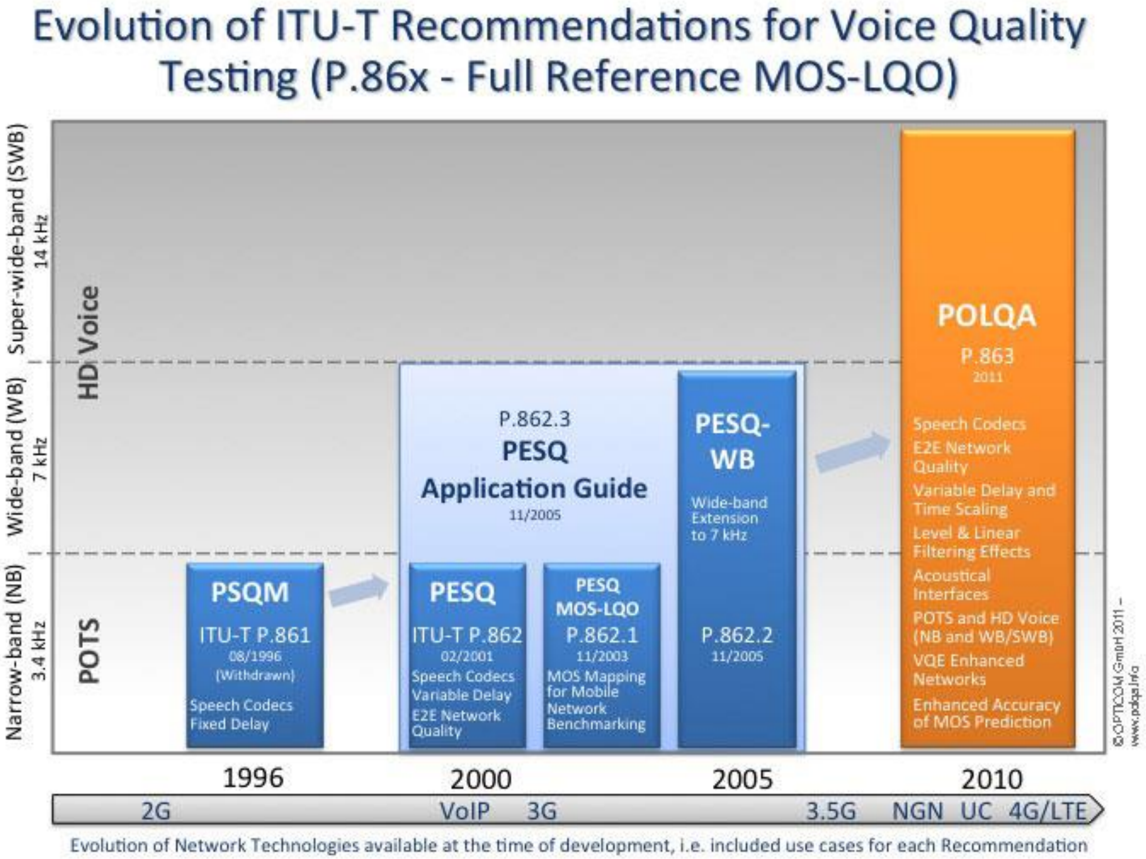
Subjective and objective quality evaluation methods

- **Objective quality rating:**
  - Target: Replace time-consuming and expensive subjective test by a computational method
  - Not easy to achieve: Subjective quality rating has to be modeled.
  - General: Two types of signals to be rated:
    - Voice for telephone / coding quality rating and
    - Music for high quality audio coder (e.g., MP3) rating



Subjective and objective quality evaluation methods

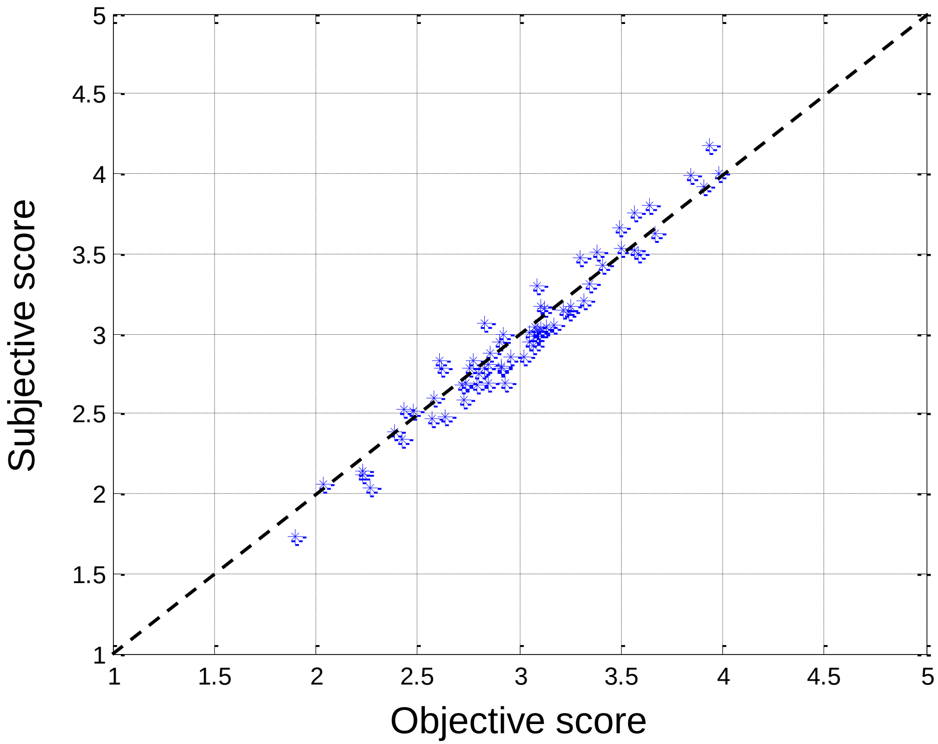
- **Evolution of different methods of quality rating of speech signals** (MOS-LQO: the corresponding objective measure compared to MOS, LQO: listening quality objective).



Reference for pictures: http://www.opticom.de

Subjective and objective quality evaluation methods

- Comparing the subjective and objective MOS scores:
  The closer the values are located with respect to the dashed line, the better do objective and subjective results fit.



Correlation coefficient:

$$\delta = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2 \sum_i (y_i - \overline{y})^2}}$$

Subjective and objective quality evaluation methods

- **History of speech quality rating methods:**

- 1996: **PSQM** (perceptual speech quality measure), ITU-T P.861:
  - Objective analysis of speech codecs for narrow-band telephone speech.

- 2000/1: **PESQ** (perceptual evaluation of speech quality), ITU-T.862:
  - Allows assessing additional degrees of degradation such as varying delays, packet loss artefacts by VoIP transmission.

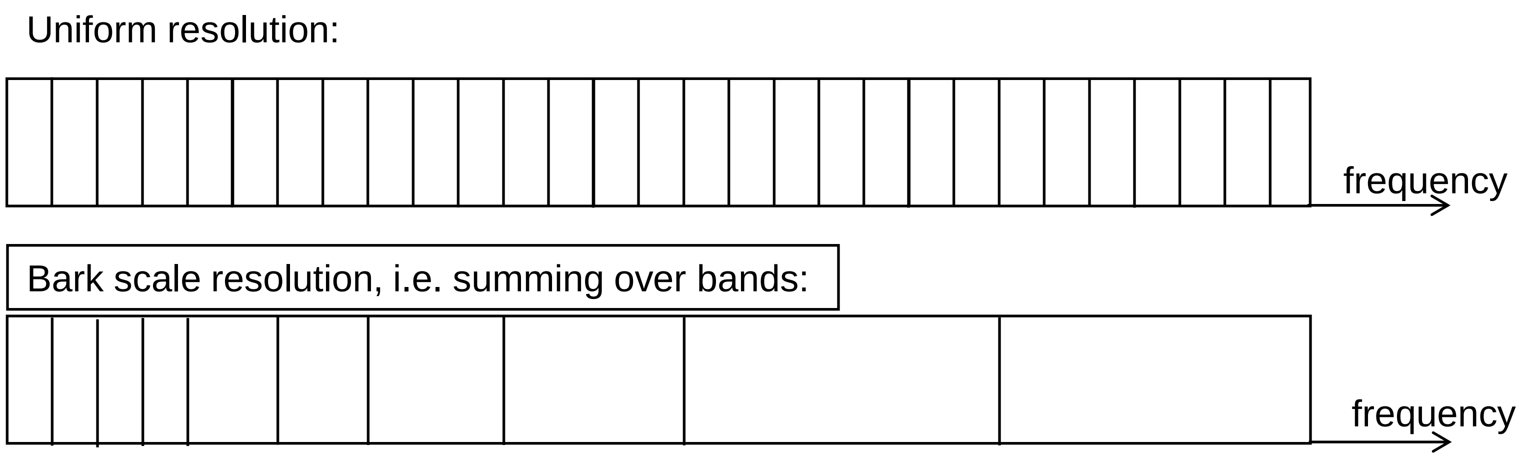
- 2005: **PESQ-WB** generalizing the PESQ concept for wideband speech signals.

- 2010: **POLQA** (perceptual objective listening quality analysis), ITU-T.863:
  - New perceptual models, especially designed for the assessment of ultra-wide band speech.

- **Having a closer look at the PESQ method:**

- Performed computational steps:
  - **STFT** is applied to each of the signals (reference and signal under test)

  - **The power in Bark bands** is calculated.



pulei esse PESQ por motivos de MUITO CHATO

- **Having a closer look at the PESQ method:**

- Performed computational steps:
  - **Time and frequency dependent equalization**:
    - The mean PSD is calculated for each of the signals, where only values 30 dB above the hearing threshold are considered for the average.
    - Spectral differences (time independent !) are equalized up to 20 dB.

    => overall spectral equalization => G(f)

    - Short-term gain compensation is applied. Therefore in each frame the power values of the reference and the signal under test are compared. The gain compensation is limited to [3*10^(-4), 5] and applied to the signal under test. Before the gain application, the gain values are smoothed.

    => short-term gain equalization => g(t)

Subjective and objective quality evaluation methods

- **Having a closer look at the PESQ method:**

- Performed computational steps:
  - **Loudness scale mapping**:
    - The Bark-scale based spectrogram (“pitch power density”) is transformed to a loudness scale [sone] Bark spectrogram.

  - **Disturbance density calculation**:
    - In the loudness scale Bark spectrogram small difference values are set to zero by masking => no perception effect.
    - Two different disturbance densities are estimated based on the difference. one symmetric: $D(f,n)$ and one none-symmetric with higher weight on added distortions: $DA(f,n)$

Subjective and objective quality evaluation methods

- **Having a closer look at the PESQ method:**

- Performed computational steps:
  - The “average disturbance value” and the “average asymmetrical disturbance value” are calculated by a weighted average over frequency:

$$D(n) = M(n) \sqrt[3]{\sum_f \left(W(f)\,|D(f,n)|\right)^3}$$

average disturbance value with emphasis on high values

$$DA(n) = M(n) \sum_f \left(W(f)\,|DA(f,n)|\right)$$

average asymmetrical disturbance value

$W(f)$: Weighing proportional to the width of the modified Bark bands

$M(n)$: Emphasis of the disturbances that occur during silences in the original speech fragment

  - **The final PESQ score** is a linear combination of the average disturbance value and the average asymmetrical disturbance value

Summary & Outlook

- First part on audio coding schemes:
- Target:
  - Remove redundancy of the signal which is coded.
  - Transmit only the relevant information.
- All coding schemes are based on prediction error filtering

- **This lecture**:
  - Signal form coder => no transmission of prediction coefficients.

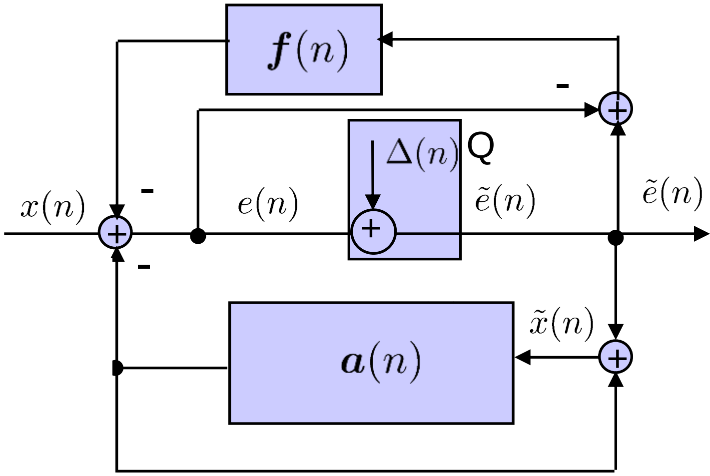
- **Next lecture**:
  - Continuation of signal form coders: Vocoder and Hybrid coder
  - Frequency domain / sub-band coders:

    MP3 and AAC coders of MPEG2 and MPEG4 standards

Appendix (to slide 16)

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Coloring the quantization noise**



$$\tilde{X}(z) = \hat{X}(z) + \tilde{E}(z) = A(z)\,\tilde{X}(z) + \tilde{E}(z) = \frac{\tilde{E}(z)}{1 - A(z)}$$

$$\hat{X}(z) = \tilde{X}(z) - \tilde{E}(z) = \tilde{E}(z)\frac{A(z)}{1 - A(z)}$$

$$E(z) = X(z) - \hat{X}(z) - F(z)\,\Delta(z)$$

$$= X(z) - \underbrace{\tilde{E}(z)}_{E(z)+\Delta(z)} \frac{A(z)}{1 - A(z)} - F(z)\,\Delta(z)$$

$$= X(z) - E(z)\frac{A(z)}{1 - A(z)} + \Delta(z)\left[-F(z) - \frac{A(z)}{1 - A(z)}\right]$$

$$E(z)\left[1 + \frac{A(z)}{1 - A(z)}\right] = X(z) + \Delta(z)\left[-F(z) - \frac{A(z)}{1 - A(z)}\right]$$

$$E(z) = X(z)(1 - A(z)) - \Delta(z)\,\left[F(z)(1 - A(z)) + A(z)\right]$$

with: $\tilde{E}(z) = E(z) + \Delta(z)$

$$\tilde{E}(z) = X(z)(1 - A(z)) - \Delta(z)\,\left[F(z)(1 - A(z)) + A(z) - 1\right]$$

Appendix (to slide 16)

$$\tilde{E}(z) = X(z)(1-A(z)) - \Delta(z)\,[F(z)(1-A(z)) + A(z) - 1]$$

$$\frac{\tilde{E}(z)}{1-A(z)} = \tilde{X}(z) = X(z) - \Delta(z)\,[F(z) - 1]$$

$$\boxed{\tilde{X}(z) = X(z) + \Delta(z)\,[1 - F(z)]}$$

Choose closed-loop structure: $\tilde{X}(z) = X(z) + \Delta(z)$ => $F(z) = 0$

Choose open-loop structure: $\tilde{X}(z) = X(z) + \dfrac{\Delta(z)}{1-A(z)}$ => $F(z) = \dfrac{-A(z)}{1-A(z)}$

Continuous fading with γ:

$$1 - F(z) = \frac{1-A(z/\gamma)}{1-A(z)}$$

$$\boxed{F(z) = \frac{A(z/\gamma) - A(z)}{1-A(z)}}$$

Choose closed-loop structure: $\gamma = 1$

Choose open-loop structure: $\gamma = 0$