
A/D and D/A conversion

- Analog-to-digital converter: combines sample-and-hold function, quantizer, and coder
- Digital-to-analog converter: combines decoder and hold function
- Oversampling and noise shaping

Lecture 8

Finite Word-Length Effects

Gernot KUBIN

Signal Processing and Speech
Communication Laboratory
www.inw.tugraz.at/spsc



A member of the Institute for Communications and Wave Propagation

Contents/Lecture 8

- 8.1. Coefficient sensitivity and quantization
- 8.2. Accuracy vs. stability
- 8.3. Round-off errors: statistical models
- 8.4. Round-off errors: deterministic models
- 8.5. State-space structures: direct error control

Coefficient sensitivity and quantization

- Parallels the tolerance problem for concentrated circuit elements in analog circuits
- Tellegen's theorem
- Pole/zero locations under quantization of difference equation coefficients
- Structures with minimal coefficient sensitivity

↳ quasi pseudo verlustfrei

Accuracy vs. stability

- Round-off errors due to internal computations: map finite alphabet back to same number of bits after multiplications and additions

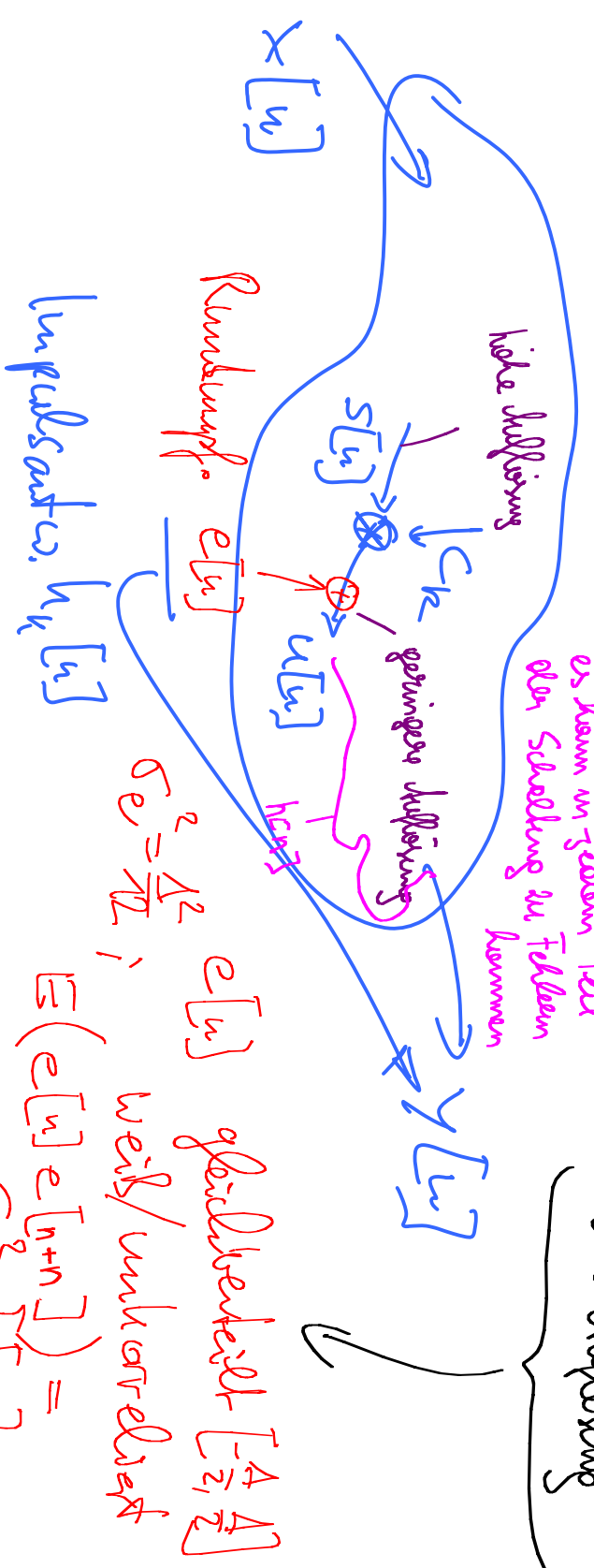
$$\left[\frac{8}{8} \right] \times \left[\frac{8}{8} \right] = \left[\frac{16}{6} \right] \quad \left\{ \right. = \left[\frac{17}{1} \right]$$
- Error generation, propagation, and accumulation most critical in feedback loops
- Total accuracy defined by accumulated errors
- In certain cases, round-off errors drive system into instability, i.e., overflow or underflow; much more serious than loss of accuracy

Je höher Prozentsatz hat eine erlaubte Unschärfe am Bild; bei $\frac{\infty}{2}$ muss Ungenauigkeit abgefragt werden \Rightarrow Fehler werden geringer!

Round-off errors: statistical models

- Error sources, multipliers, accumulators, guard bits
- Error propagation networks
- Error accumulation and total noise power

Annahme: \cdot Breitbandige Signale
 \cdot volle Multiplikation
 \cdot hohe Multiplikation



$e[n]$ gleichverteilt $[-\frac{A}{2}, \frac{A}{2}]$
 weiß/unkorreliert
 $\sigma_e^2 = \frac{A^2}{12}$

$$E(e[n]e[n+n]) = \sigma_e^2 \delta[n]$$

Quantisierung z.B.: 256 Werte auf 64 abtasten $129 \Rightarrow$ produziert Rundungsrauschen $e[n]$

$$\Delta y [n] = \sum_k (h_k [n] * e_k [n]) \quad , \quad E[(\Delta y)^2]$$

↳ Erwartungswert

$e_k \cdot e_l$ unkorreliert für $k \neq l$

$$E[(\Delta y)^2] = \sum_k E[(h_k * e_k)^2]$$

$$E[(h_k * e_k)^2] = E[(\sum_m h_m [n] e_{n-m})^2]$$

gilt für alle k

$$= E\left(\sum_m \sum_l h_m [n] e_{n-m} h_l [n] e_{n-l}\right)$$

$$= \sum_m \sum_l h_m [n] h_l [n] \cdot E(e_{n-m} e_{n-l})$$

↳ h ist fix, kein stochastisches Verhalten
 h wird einmal gemessen & gespeichert!

$$= \sum_m \sum_e h c m \gamma h c e \gamma \cdot \underbrace{E(e c_{n-m} \gamma e c_{n-l} \gamma)}$$

für $n-m = n-l$ ist dieser
 korreliertes Wert zu 0 verschieden;
 und immer gleich σ_e^2

$$= \sum_m h c m \gamma h c m \gamma \cdot \sigma_e^2 = \sigma_e^2 \cdot \sum_m h c m \gamma^2$$

$$= \sigma_e^2 \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 d\omega$$

$$\sum_m \sum_e h c m \gamma h c e \gamma \cdot E(e c_{n-m} \gamma e c_{n-l} \gamma) = r_{ee} [n] = \delta c m \gamma \cdot \sigma_e^2$$

für korreliertes Rauschen, d.h. $r_{ee} [n] \neq \delta c n \gamma$

Index: $n-m, n-l \Rightarrow m-l = j, l = m-j$
 $n-l = n-m + j$

$$\sum_m \sum_j h[m] h[m-j] \cdot E(e^{cn-m} \cdot e^{cn+m-j})$$

$$r_{ee}[j] = E(e^{cn} \cdot e^{cn+j})$$

↪ hängt nur von j (= Zeitverschiebung) ab!

$$= \sum_m \sum_j h[m] \cdot h[m-j] \cdot r_{ee}[j]$$

$r_{hh} \dots$ überminimale auto-korrelationsfunktion oder
Impulsantwort $h[n]$

$$r_{hh} = \sum_m h[m] h[m+k]$$

$$= \sum_j r_{hh} [^{-j}] r_{ee}[j] = \sum_j r_{hh} [h^{-j}] r_{ee}[j] \Big|_{h=0}$$

$$y[n] = (h_k * (u + e)) \cdot [n]$$

Foller am Ausgang: $\Delta y[n] = (h_k * e)[n]$

$$E(\Delta y^2[n]) = E\left\{\left(\sum_m h_k[m] e[u-m]\right)^2\right\} =$$

$$= E\left\{\sum_m \sum_l h_k[m] h_l[l] e[u-m] e[u-l]\right\}$$

$$= \sum_m \sum_l h_k[m] h_l[l] E\{e[u-m] e[u-l]\}$$

$$m-l = j \quad ; \quad l = m-j$$

$$m-l = n-m+j$$

$$= \sum_m \sum_j h_k[m] h_l[m-j] E\{e[u-m] e[u-m+j]\}$$

$$\underbrace{\quad}_{\sigma_e^2 \delta[j]}$$

$$E(A_y^2) = \sum_j \underbrace{\sum_m h_k[m] h_k[m-j]}_{\text{def. Autokorr. Impulsantwort}} \text{real}[j] = \sum_j r_{hh}[-j] \text{real}[j] = \text{real}[j]$$

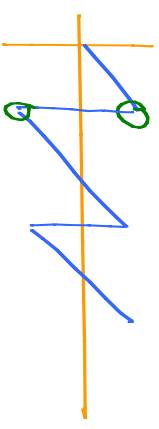
(Korr am Ausgang) = (Korr am Eingang) * (Korr d. Impulsantwort)

Da $\text{real}[j] = \text{se}^2 \cdot \delta[j]$

$$E(A_y^2) = \sum_m h_k^2[m] h_k^2[m] \text{se}^2 = \sum_m h_k^2[m] \cdot \underbrace{\frac{A^2}{12}}_{\substack{\text{1} \\ = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\theta})|^2 d\theta}}$$

Round-off errors: deterministic models

- **Overflow** and underflow characteristics
- Scaling Werte können nicht beliebig klein oder groß werden
- Overflow and granular limit cycles, chaos



des Komplexions
Sprung von max auf min

Granulare Grenzyklen (\Leftrightarrow Unterlauf); "Zero-Lupet"

$$y[n] = a \cdot y[n-1] + x[n]$$

Angabe

$$x[n] = 8[n] \cdot \left(\frac{7}{8}\right) \text{ durch Angabe}$$

↳ Betrag-Runden nach oben (auf ganze $\frac{1}{8}$)
Speicherzellen
y[-1] = 0 ... leer über alle

Schleife über Felder

$$y[n] = \begin{cases} \frac{7}{8} & \frac{7}{8} \left(\frac{1}{2}\right) = -\frac{7}{16} \Rightarrow \frac{8}{16} = \frac{1}{2} \\ \frac{1}{8} & \left(-\frac{1}{2}\right) = +\frac{1}{4}, \left(+\frac{1}{4}\right) = -\frac{1}{8}, \dots \end{cases}$$

relativ bei nicht-stetigem Input kann auch die Grenzyklen ein Schwingen aufweisen \Rightarrow Pfeilen am Ausgang selbst bei Input = 0

State-space: direct error control

- State-space representation of causal LTI filters
- Pseudo-energy concepts for digital filters
- State-transition matrix design

Zustandsraumfilter

$$\vec{z}[n] = \mathbb{R} \{A \cdot \vec{z}[n+1]\} + \vec{b} \cdot x[n]$$

↳ Rundungsoperationen \Rightarrow soll oben System Energie enthalten \sim verhindert Schwingung (angelehnt an Dämpfung aus Physik)

\Rightarrow führt auf Strukturen die prinzipiell frei von Grenzyklen sind (kannet zusätzlichen Hardware aufwenden).