

Rechnerstrukturen und -organisation (RSO)

Informations-
darstellung

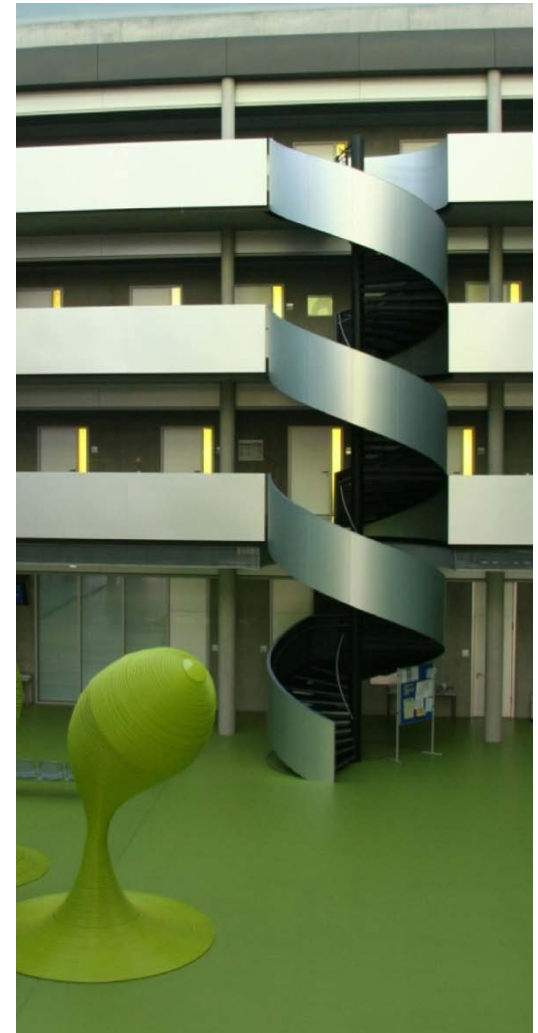
Rainer G. Spallek

TU Dresden, 29.10.2020



Gliederung

- 1 Zielstellung
- 2 Grundbegriffe
- 3 Signale, Diskretisierung
- 4 Informations-/Datenverarbeitung
- 5 Schichtenmodell des Computers
- 6 Zusammenfassung



1 Zielstellung

- Erlangung eines Grundverständnisses für die Informationstheorie
- Auseinandersetzung mit den Begriffen Information, Entropie, Nachricht, Signal, Zeichen, Alphabet, Zeichenkette
- Erwerb von Grundlagen für die Bestimmung des Informationsgehaltes einer Informationsquelle und der Redundanz
- Verständnis im Umgang mit physischen Signalen und digitalen Daten, wie auch mit der Diskretisierung
- Einführung in die Informations- und Datenverarbeitung, Datenformate
- Aufbau und Erläuterung eines einfachen Schichtenmodells des Computers

2 Grundbegriffe

Information, Entropie, Zeichen

Information ist die Beseitigung von Ungewissheit.

Information (im engeren Sinne) ist der Oberbegriff für Mitteilungen, Befehle, Ergebnisse, Daten usw. Im weiteren Sinne ist Information neues Wissen über ein Ereignis.

Entropie ist die Unbestimmtheit der Information.

Der Informationsinhalt ist um so größer, je unbestimmter das Ereignis vor Eintreffen der Information war. Das Maß für die Information ist die Zunahme an Wahrscheinlichkeit. Die kleinste Einheit der Information ist eine Aussage auf eine Ja/Nein-Frage (Antwort) und wird mit $1 \text{ Sh} = 1 \text{ bit}$ (nach *Claude Elwood Shannon*, Informationstheoretiker) bezeichnet.

Zeichen sind Bestandteile eines Alphabetes.

Zeichen entstammen einer Informationsquelle. Das Auftreten einzelner Zeichen eines Alphabetes ist im allgemeinen nicht gleichwahrscheinlich. Zeichen können zu Zeichenketten zusammengefasst (verkettet) werden.

Informationsquelle, Zeichen

- endliche, feste Anzahl von unterscheidbaren Zeichen
- feste Wahrscheinlichkeit für das Erscheinen jedes Zeichen
- das Erscheinen eines Zeichens ist ein unabhängiges Ereignis
- Zeichen sind Elemente eines Zeichensatzes, Alphabetes

N – Anzahl verschiedener Zeichen des Alphabetes

p_i – Wahrscheinlichkeit des Auftretens des i -ten Zeichens

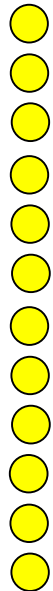
$\sum_{i=1}^N p_i = 1$ – Normierungsbedingung für $0 \leq p_i \leq 1$ mit $1 \leq i \leq N$

$H_i = \log_2 \frac{1}{p_i}$ – Informationsgehalt des i -ten Zeichens

- Je seltener ein Zeichen auftritt, desto mehr Information übermittelt sein Auftreten.

Informationsgehalt der Zeichen einer Quelle

Informationsquelle
Zeichen ● ◆



$$H_{\bullet} \rightarrow 0$$

Informationsquelle
Zeichen ● ◆



$$H_{\blacklozenge} \gg H_{\bullet}$$

Informationsquelle
Zeichen ● ◆



$$H_{\blacklozenge} = H_{\bullet}$$

Entropie einer Informationsquelle

$$H_Z(p_1, p_2, \dots, p_N) = \sum_{i=1}^N (p_i \cdot H_i) \quad \text{Informationsgehalt der Quelle Z}$$

$$H_Z = \sum_{i=1}^N \left(p_i \cdot \text{ld} \frac{1}{p_i} \right) = -\sum_{i=1}^N (p_i \cdot \text{ld} p_i)$$

$Z(p_1, p_2, \dots, p_N)$ stationäre gedächtnislose Zeichenquelle mit Entropie H_Z

$H_Z(p_1, p_2, \dots, p_N)$ durchschnittlicher Informationsgehalt der Zeichen 1...N

- Die Entropie einer Zeichenquelle wird in $\frac{\text{bit}}{\text{Zeichen}}$ angegeben.
- Bedeutung für die Beurteilung von Codierungen → effiziente Codierung
- Ein Maximum an Information bei N verschiedenen Zeichen erhält man, wenn alle Zeichen gleichwahrscheinlich sind
→ $p_i = \frac{1}{N} \rightarrow H_Z = H_i = \text{ld} N$

Zeichen, Zeichenfolge, Alphabete

Unter einem Alphabet versteht man nach DIN 44300 eine total geordnete endliche Menge von unterscheidbaren Zeichen (\rightarrow Zeichenvorrat).

Zeichen z_μ eines Alphabetes $Z = \{z_1, z_2, \dots, z_N\}$ mit N verschiedenen Zeichen können zu Zeichenfolgen K beliebiger Länge zusammengefasst werden.

Darstellung der so entstehen Folgen K von nicht notwendigerweise verschiedenen Zeichen z_μ , die mit dem Verkettungsoperator \circ verkettet wurden:

$$K = z_1 \circ z_2 \circ z_3 \circ \dots \circ z_\mu \quad \text{oder kurz} \quad K = z_1 z_2 z_3 \dots z_\mu \quad \text{mit } z_i \in Z$$

Zeichenfolgen – Zeichenketten - Wörter

Ihre Bedeutung wird durch die Wahl der Zeichen und ihre Stellung innerhalb der Kette bestimmt. Bei eindeutigem Zusammenhang kann der Verkettungsoperator entfallen, er wird dann implizit angenommen.

Redundanz

Die Redundanz R_C (Weitschweifigkeit, Überfluss) einer Codierung gibt an, um wie viel der mittlere Codierungsaufwand H_C vom mittleren Informationsgehalt der Zeichenquelle (Entropie H_Q) abweicht.

$$R_C = H_C - H_Q \quad \text{Redundanz absolut}$$

$$r_C = \frac{H_C - H_Q}{H_C} = 1 - \frac{H_Q}{H_C} \quad \text{Redundanz relativ}$$

$$\eta_C = \frac{H_Q}{H_C} = 1 - r_C \quad \text{Informationsgehalt relativ}$$

$$1 = r_C + \eta_C \quad \text{Summe}$$

Redundanz ist für Fehlererkennung und -korrektur notwendig (z.B. ECC).
Erhöhung der Speicher- und Übertragungssicherheit von Informationen.

Beispiel: Informationsgehalt der deutschen Schriftsprache

27 Zeichen (26 Buchstaben und Leerzeichen) dt. Schriftsprache

Auftrittswahrscheinlichkeit der einzelnen Zeichen unterschiedlich ($N = 27$):

z.B.: $p_e = 17,9 \%$, $p_n = 10,2 \%$, $p_s = 6,1 \%$, $p_q = 0,04 \%$, $p_y = 0,03 \%$, ...

$$H_u(p_1, p_2, \dots, p_{27}) = - \sum_{i=1}^{27} (p_i \cdot \lg p_i) = 4,06 \frac{\text{bit}}{\text{Zeichen}}$$

Annahme: Alle Zeichen gleichwahrscheinlich ($N = 27$, $p_i = \frac{1}{27}$)

$$p_a = \frac{1}{27}, p_b = \frac{1}{27}, \dots, p_z = \frac{1}{27}, p_{\text{ }} = \frac{1}{27}$$

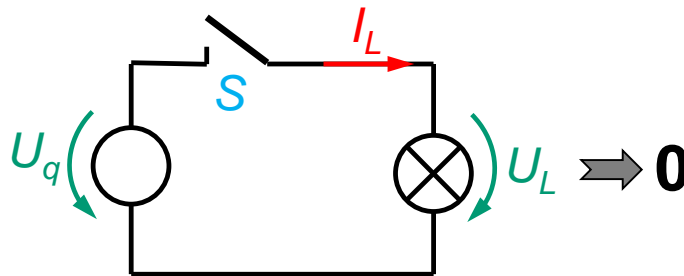
$$H_g(p_1, p_2, \dots, p_{27}) = \sum_{i=1}^{27} \left(\frac{1}{27} \cdot \lg 27 \right) = \lg 27 = 4,75 \frac{\text{bit}}{\text{Zeichen}}$$

Bei gleichwahrscheinlichen Zeichen ergibt sich eine höhere Entropie $H_g > H_u$

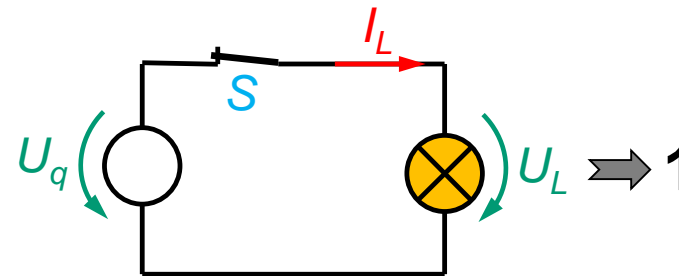
Signal

Ein Signal ist die Darstellung von Informationen durch physische Größen (Signalträger). Signale sind Träger der Information. Sie sind gezielt veränderbar und vom Empfänger wahrnehmbar.

Signale: elektrisch, mechanisch, optisch, magnetisch, pneumatisch, hydraulisch



S - offen
 $U_L=0, I_L=0$
Lampe aus

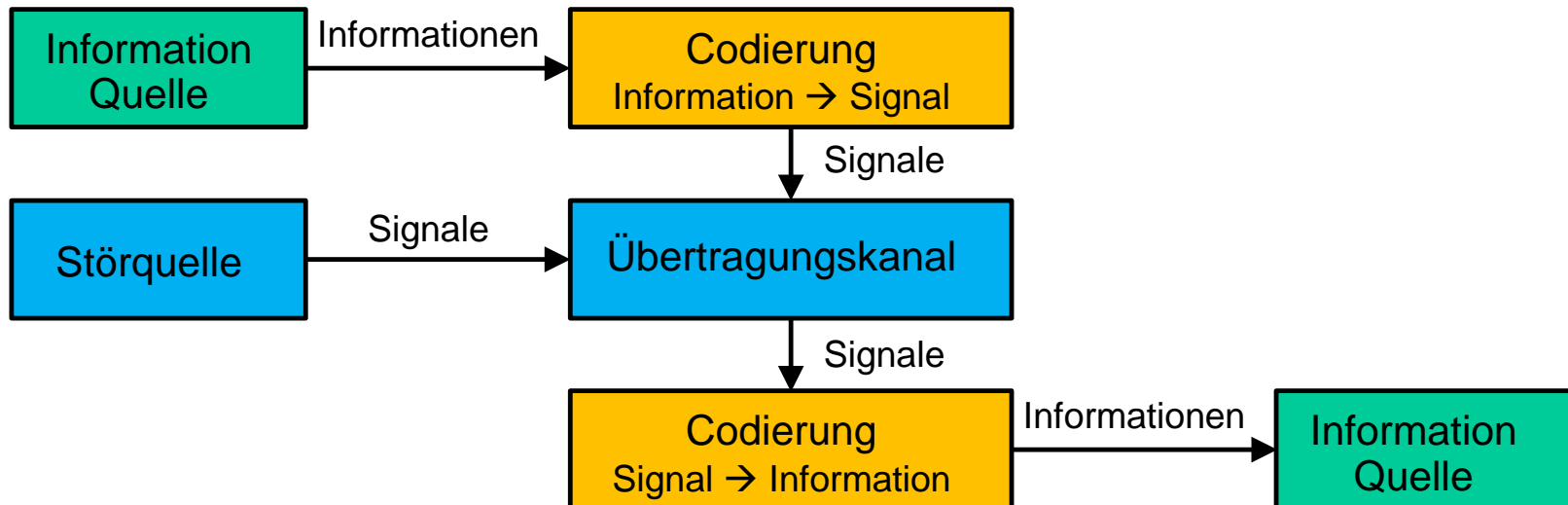


S - geschlossen
 $U_L=U_q, I_L>0$
Lampe ein

Nachricht

- Die Nachricht ist eine Information die aus Zeichen oder Zeichenfolgen besteht und mit beliebigen Mitteln von einem Sender zu einem Empfänger übertragen wird.
- Die abstrakte Information der Nachricht wird durch die physische Nachricht in Form von Signalen übertragen.
- Die Transformation von abstrakter Information zu physischen Signalen erfolgt durch Codierung.
- Nachrichten sind das Grundelement jeder Kommunikation.

Information, Informationskette (Shannon)



Die Informationsübertragung vom Sender zum Empfänger erfolgt durch codierte Signale über einen Übertragungskanal, der je nach Bedingung gestört sein kann und so die codierte Information verfälscht (Störquelle).

3 Signale, Diskretisierung

Zeit: kontinuierlich

kontinuierlich

diskret

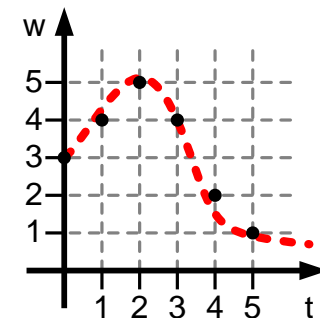
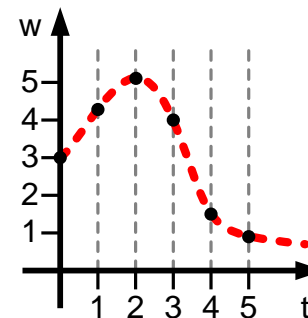
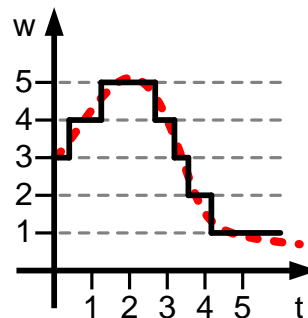
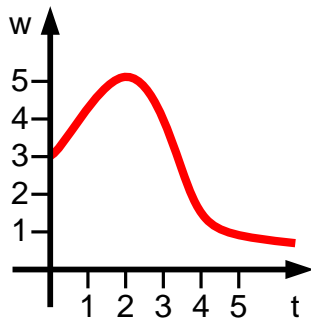
diskret

Wert: analog

diskret

analog

diskret



In der Natur ist die Zeit kontinuierlich und die Werte sind analog.

Digitale Signale sind zeit- und wertdiskret und im allgemeinen binär codiert.

Diskretisierung der Zeit: Abtasten der Zeitfunktion zu diskreten Zeitpunkten.

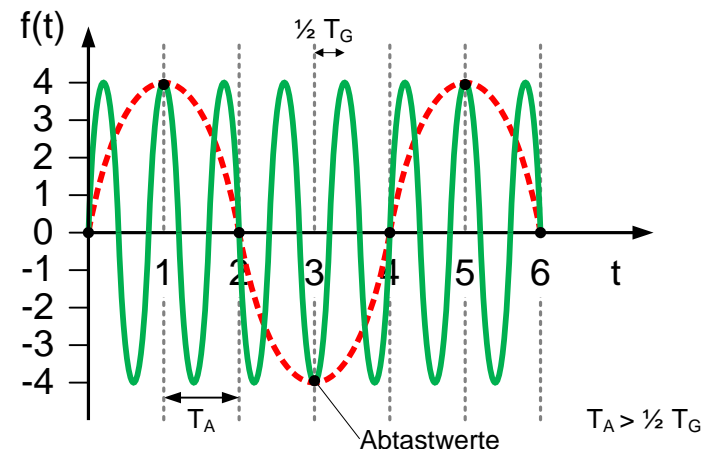
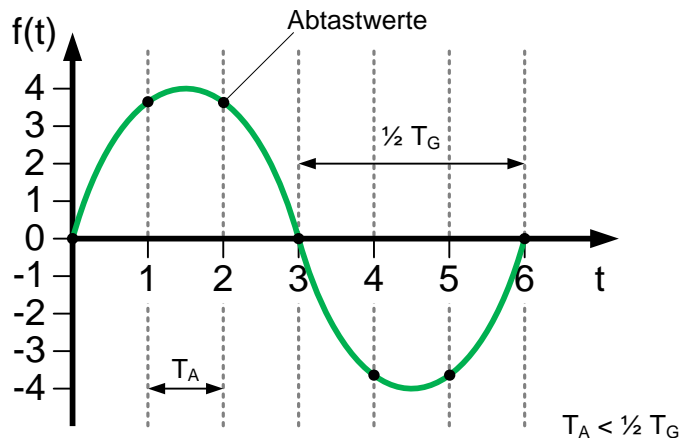
Diskretisierung der Werte: Quantisierung der Werte durch diskrete Stufen.

Zeitquantisierung – Abtasttheorem (Shannon)

Kontinuierliche Zeitfunktionen müssen vor der Verarbeitung quantisiert (diskretisiert) werden. Jede kontinuierliche Zeitfunktion $f(t)$ mit der oberen Grenzfrequenz $f_G = \frac{1}{T_G}$ wird eindeutig durch diskrete Werte beschrieben:

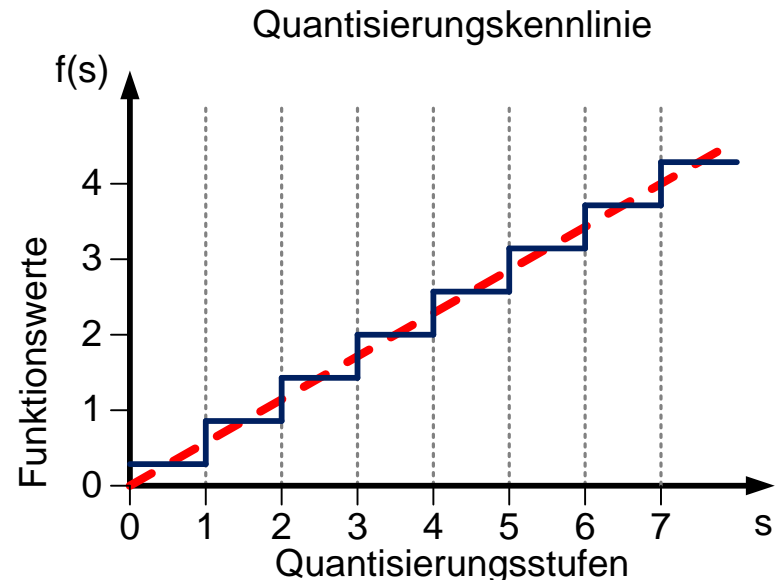
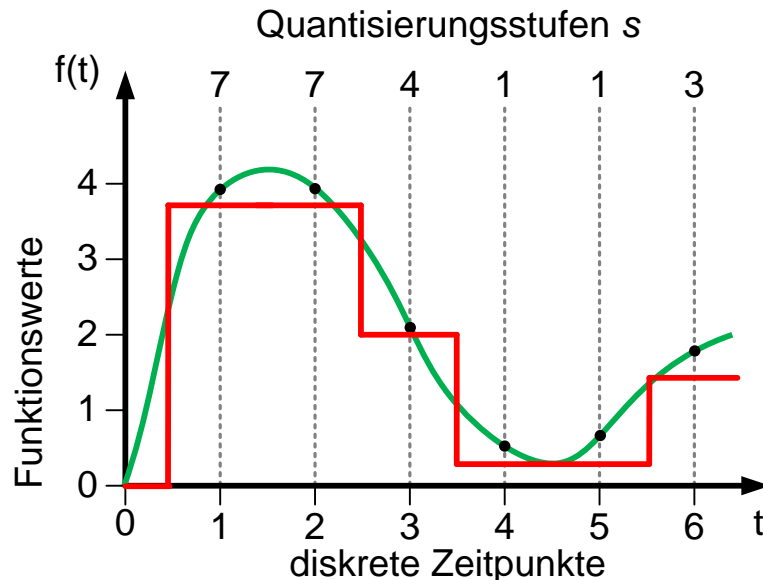
$$t = k \cdot T_A \text{ mit } k = 0, \pm 1, \pm 2, \dots, \text{ mit } T_A = \frac{1}{f_A} < \frac{1}{2} T_G \text{ (konstante Schrittweite).}$$

Anderenfalls kann es zur Überlagerung von Signal- und Abtastfrequenz kommen, Stroboskopeffekt.



Wertquantisierung – Analog-Digital-Wandlung

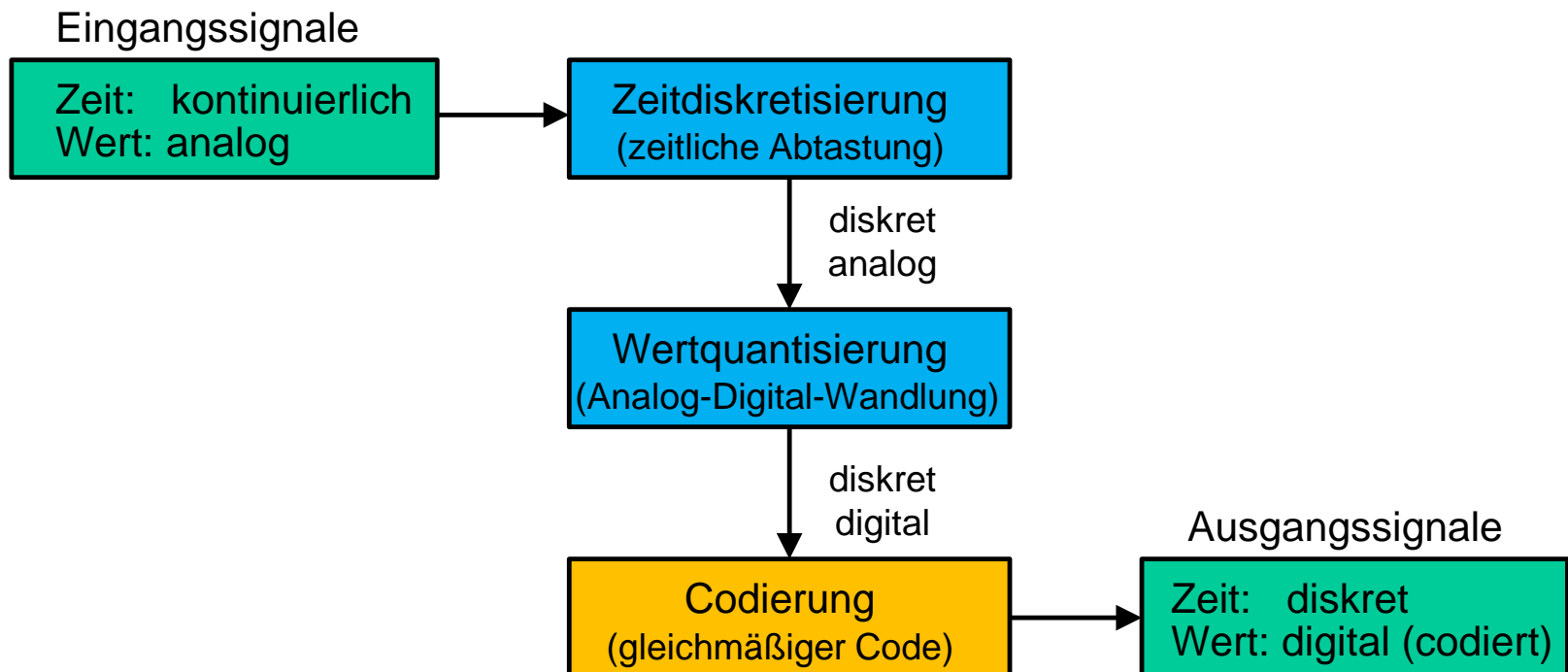
Analoge Signale müssen vor der Verarbeitung quantisiert (diskreditiert) werden. Die Zuordnung der Analogwerte zu den diskreten Werten, Quantisierungsstufen erfolgt durch einen Analog-Digital-Wandler (ADC) entsprechend einer Quantisierungskennlinie. Die Anzahl der dem Analogwert entsprechenden Quantisierungsstufen s kann codiert dargestellt werden.



Quantisierungsstufen – Quantisierungsintervall

- Die Wertquantisierung beschreibt das analoge Signal durch s diskrete (unterschiedliche) Quantisierungsstufen.
- Die Anzahl der erforderlichen Quantisierungsstufen ist vom Wertebereich des Signals und der Größe der einzelnen Quantisierungsintervalle abhängig.
- Die Quantisierungskennlinie kann linear oder auch nichtlinear sein.
- Die Größe der Quantisierungsintervalle ergibt sich aus der geforderten Abtastgenauigkeit für das Signal.
- Sowohl durch die Zeit- als auch durch die Wertquantisierung wird das Signal nur näherungsweise beschrieben. Die dabei auftretenden Ungenauigkeiten finden sich im Quantisierungsrauschen wieder.
- Das Quantisierungsrauschen ist indirekt ein Maß für die Güte der Quantisierung

Digitalisierung analoger zeitkontinuierlicher Signale



Beispiel: Digitale Audio-CD

obere Grenzfrequenz: f_G ca. 20 kHz (menschliche Hörvermögen)

Abtastfrequenz: $f_A = 44,1$ kHz

Digitalisierung der Audiowerte: 16-bit (65536 mögliche diskrete Abtastwerte)

Datenvolumen: $2 \cdot 16$ bit pro Abtastzeitpunkt (2-Kanal, Stereo)

Ausgabedatenstrom: 1,411 Mbit/s Rohdaten

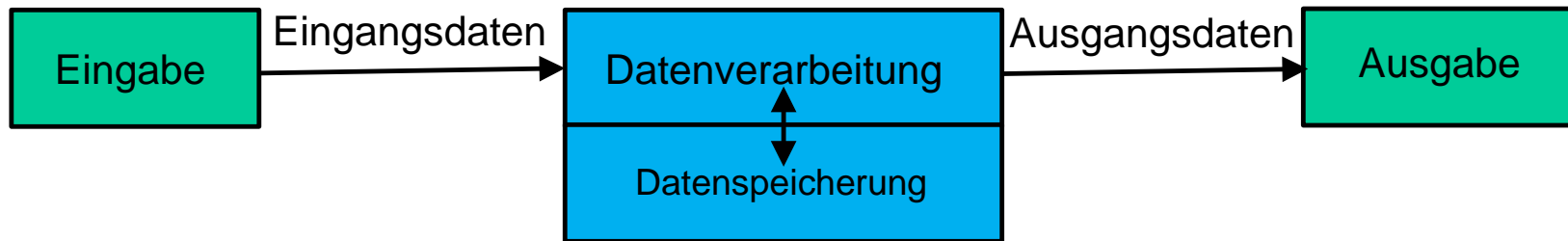
Ausgabedatenstrom
decodiert: 176,4 kbit/s

Redundanz: Zusatzdaten im Ausgabestrom zur Erkennung und Korrektur von Fehlern bei der Wiedergabe (Staubteilchen, Kratzer, ...).

4 Informations-/Datenverarbeitung

Computer sind hochautomatische technische Systeme der Informationsverarbeitung (Datenverarbeitung) mit folgenden Merkmalen:

- Eingabe, Verarbeitung, Speicherung und Ausgabe von Informationen,
- frei programmierbar, Anpassung an unterschiedlichste Aufgabenstellungen,
- arbeiten auf der Basis streng mathematischer Prinzipien.



Die Informationsverarbeitung in Computern erfolgt durch die gesteuerte Ausführung von Befehlen auf Operanden. Die Darstellung von Steuerung, Befehlen und Operanden erfolgt oft gleichermaßen als Daten.

Computer – Informationsverarbeitung

Die Hardware eines Computers ist auf wenige Daten, genau spezifizierte Informationsdarstellungen, fest zugeschnitten. Nur mit diesen Daten sind ausgewählte, spezifische Manipulationen im Computer möglich.

Verarbeitungsdaten: umfassen die zu verarbeitenden Informationen und die Resultate der Verarbeitung (E/A - Daten).
(Operanden)

Programmdaten: bestimmen Art und Reihenfolge der
(Befehle, Instruktionen) Verarbeitungsschritte (Operationen, Befehle, ...).

Steuerdaten: steuern und registrieren Arbeitsmodi und
(Control, Status) unterstützen den geordneten Programmablauf.

Darstellungsarten von Verarbeitungsdaten können sein:

numerische und alphanumerische Daten, Bild- und Ton-Daten, Sensor-Daten, Aktuator-Daten, ...

Informationsdarstellung – Formate

Die Darstellung der verschiedenen Informationen (Daten) im Computer erfolgt binär codiert in festen Formaten.

Verarbeitungsdaten: Datenformat

Darstellungsformat der Operanden

Programmdaten: Befehlsformat

Darstellungsformat der Befehle (Instruktionen)

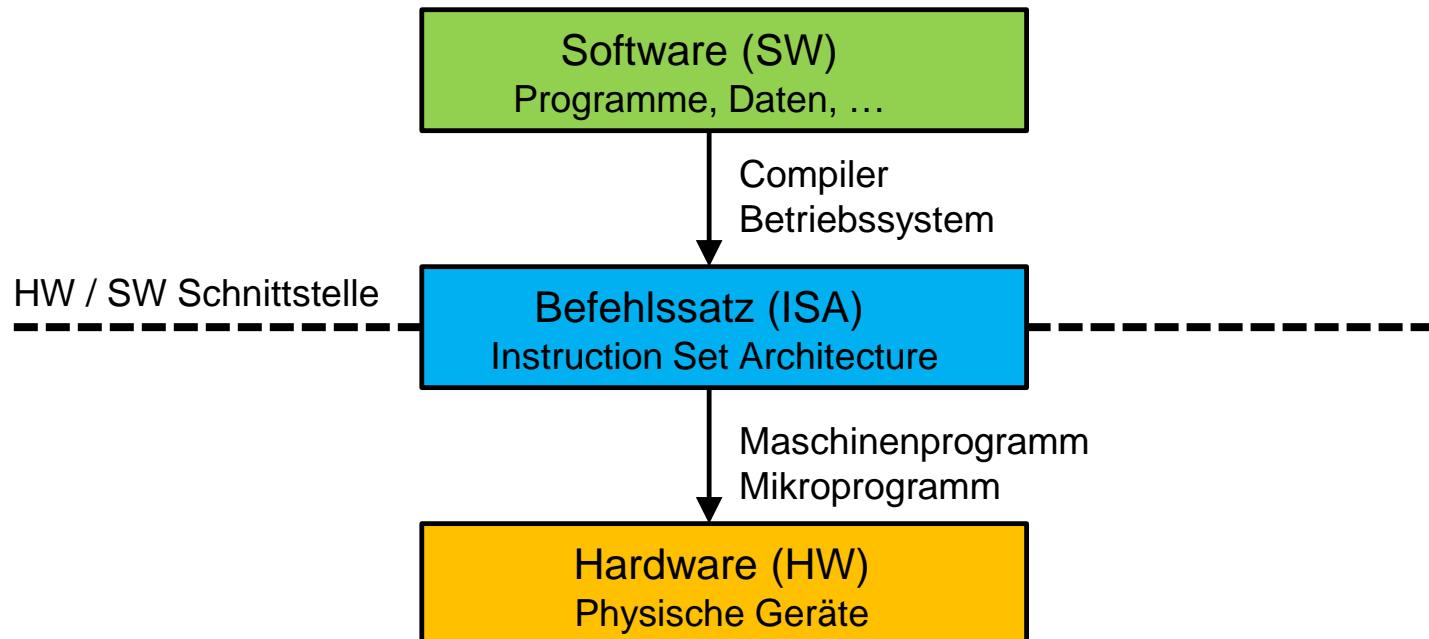
Steuerdaten: Steuerformat

Darstellung der Steuerungsinformationen

Format bedeutet dabei: Einzelne Informationsbestandteile stehen immer an der selben Stelle innerhalb der Daten einer Klasse bzw. die Daten einer Klasse sind identisch aufgebaut und codiert.

Die Daten eines Computers unterliegen einem strengen Ordnungsprinzip.

5 Schichtenmodell des Computers (einfach)



Hardware – Software

Hardware: umfasst alle mechanischen, elektrischen und elektronischen Bauelemente und Baugruppen (integrierte Schaltkreise, Leiterplatten, Netzteil, Gehäuse, ...), physische Elemente.

Software: umfasst alle Programme, Betriebssysteme und Daten (Compiler, Editor, ...).

Firmware: umfasst hardwarenahe Mikroprogramme, Funktionen und Geräteeinstellungen, Übergangsbereich zwischen Hard- und Software (meist Software, die fest in Hardware realisiert ist).

Configware: umfasst Daten für die hardwaremäßige Konfiguration von einzelnen Komponenten eines Computers (meist Software für Hardware-programmierbare Bauelemente).

Der Informationsaustausch zwischen Hardware und Software erfolgt durch Daten. Diese unterliegen ebenfalls festen Darstellungsformaten (Vorschriften).

6 Zusammenfassung

- Information ist messbar, als Entropie einer Informationsquelle.
- Informationen werden durch physische Signale repräsentiert.
- Codierung von Information als Zeichen oder Zeichenketten.
- Die Entropie einer Informationsquelle ist für die Beurteilung einer Codierung von Bedeutung.
- Redundanz von Codierung dient der Erhöhung der Datensicherheit.
- Die Digitalisierung von Signalen erfolgt durch Zeitdiskretisierung und Wertquantisierung.
- Computer sind hochautomatisierte technische Systeme der Informationsverarbeitung.
- Die Verkopplung von Hardware und Software wird im Schichtenmodell des Rechners verdeutlicht, Befehlssatzebene wichtigste Schnittstelle.