

Rechnerarchitektur I (RAI)

Informationsdarstellung

Prof. Dr. Akash Kumar
Chair for Processor Design

Gliederung

2

- 1 Grundbegriffe
- 2 Zielstellung
- 3 Zeichen, Zeichenketten, Alphabete
- 4 Information, Informationskette
- 5 Signale
- 6 Informations-/Datenverarbeitung
- 7 Schichtenmodell des Computers
- 8 Zusammenfassung





Grundbegriffe

3

→ siehe auch Informations- und Kodierungstheorie

Information: ***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: ***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.



Grundbegriffe

4

- Nachricht:** Nachricht ist eine Information, die vom Menschen erzeugt, mit beliebigen Mitteln übertragen und vom Menschen wieder aufgenommen wird. Die abstrakte Information (subjektiv für jeden Einzelnen) wird durch die konkrete Nachricht mitgeteilt.
- Zeichen:** Zeichen als Bestandteile eines Alphabetes entstammen einer Informationsquelle. Das Auftreten einzelner Zeichen eines Alphabetes ist im allgemeinen nicht gleichwahrscheinlich. Zeichen können zu Zeichenketten zusammengefasst (verkettet) werden.
- Signal:** Ein Signal ist die Darstellung von Informationen durch physikalische Größen (Signalträger). Signale sind Träger der Information. Sie sind gezielt veränderbar und vom Empfänger wahrnehmbar.



Zielstellung

5

- 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



Zeichen, Zeichenketten, Alphabete

6

Die Hardware von Computern verarbeitet nur (physikalische) Signale, die als Zeichen z_v eines Alphabetes $Z = \{z_1, z_2, \dots, z_N\}$ interpretiert werden.

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

Zeichen z_v eines Alphabetes $Z = \{z_1, z_2, \dots, z_N\}$ können zu Zeichenketten K beliebiger Länge zusammengefasst werden.

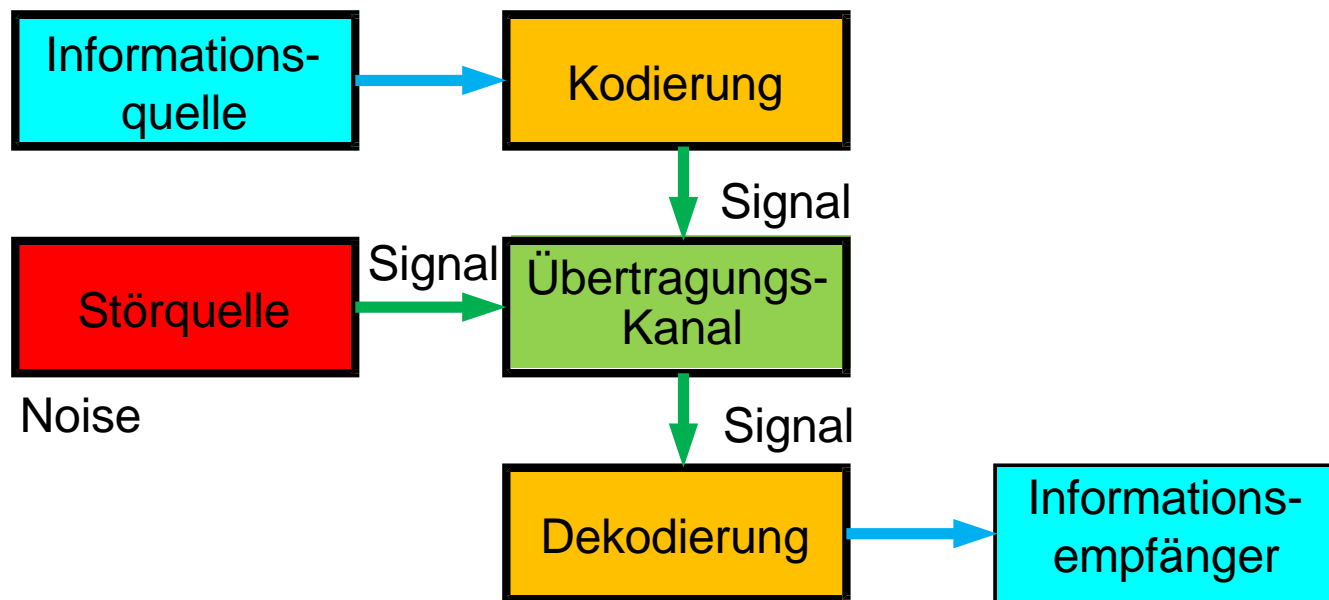
\rightarrow So entstehen Folgen von nicht notwendigerweise verschiedenen Zeichen z_v , die mit dem Verkettungsoperator \circ verkettet wurden.

$$K = z_1 \circ z_2 \circ z_3 \circ \dots \circ z_v; \quad K = z_1 z_2 z_3 \dots z_v \text{ mit } z_i \in Z$$

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

Information, Informationskette (Shannon)

7



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



Informationsquelle - Definition, Kennzeichen

8

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

N – Anzahl verschiedener Zeichen

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 enthält es.



Entropie einer Informationsquelle

9

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

$$H_m = -\sum_{i=1}^N p_i \cdot \text{ld } p_i = \sum_{i=1}^N p_i \cdot \text{ld } \frac{1}{p_i}$$

- $Z(p_1, p_2, \dots, p_N)$ stationäre gedächtnislose Zeichenquelle mit Entropie H_Z
- $H_Z(p_1, p_2, \dots, p_N)$ durchschnittlicher Informationsgehalt der Zeichen $1 \dots N$
- Die Entropie einer Zeichenquelle wird in $\frac{\text{bit}}{\text{Zeichen}}$ angegeben.
- Bedeutung für die Beurteilung von Kodierungen → effiziente Kodierung

Ein Maximum an Information bei N verschiedenen Zeichen erhält man, wenn

alle Zeichen gleichwahrscheinlich sind $p_i = \frac{1}{N} \rightarrow H_i = H_Z = \text{ld } N$



Redundanz

10

Die Redundanz R_C (Weitschweifigkeit, Überfluss) einer Kodierung gibt an, um wie viel der mittlere Kodierungsaufwand 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 Übertragungssicherheit von Informationen.

Beispiel: Informationsgehalt dt. Alphabet

11

27 Zeichen (26 Buchstaben und Leerzeichen) dt. Schriftsprache

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

$$p_e = 17,9\%, p_n = 10,2\%, p_s = 6,1\%, p_q = 0,04\% p_y = 0,03\%, \dots$$

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

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

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

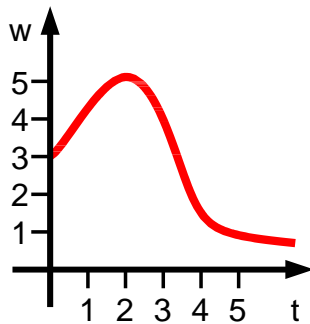
$$H_m = H_g(p_1, p_2, \dots, p_{27}) = \sum_{i=1}^{27} \frac{1}{27} \cdot \lg 27 = \lg 27 = 4,75 \frac{\text{bit}}{\text{Zeichen}}$$

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

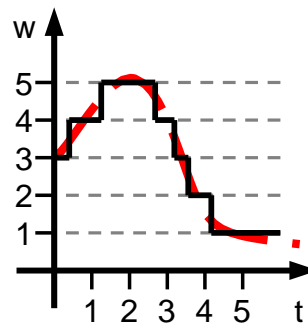
Signale

12

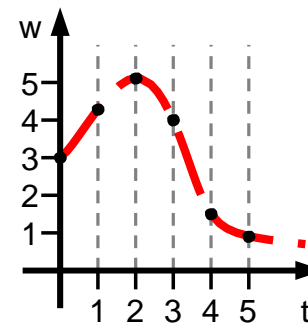
Zeit: kontinuierlich
Wert: analog



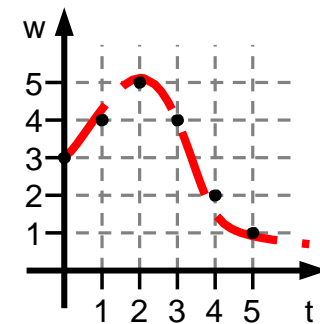
kontinuierlich
diskret



diskret
analog



diskret
diskret



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

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

Diskretisierung der Zeit: Abtasten der Zeitfunktion zu diskreten Zeitpunkten.

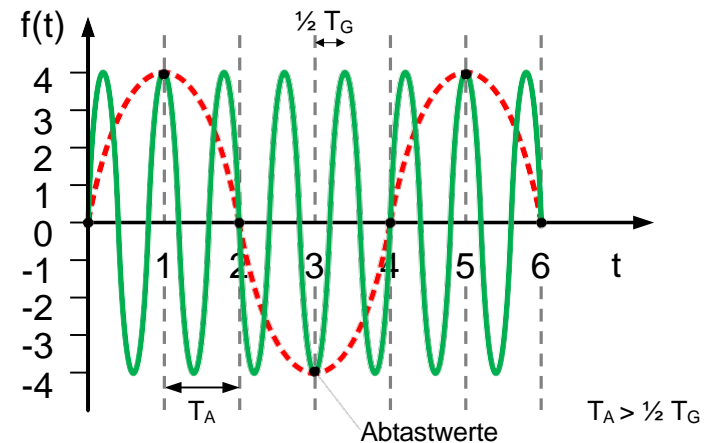
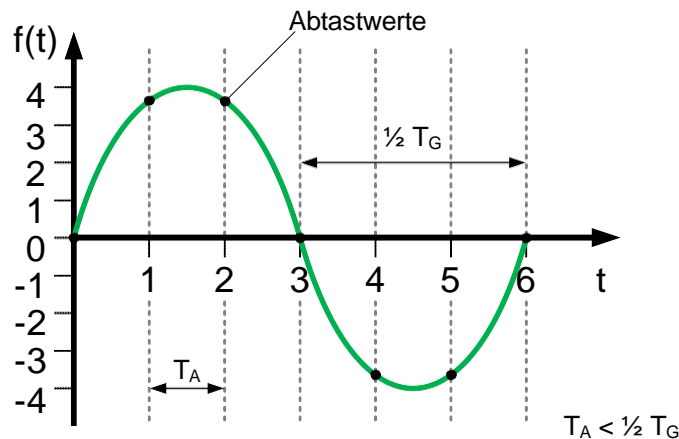
Diskretisierung der Werte: Quantisierung der Werte durch diskrete Stufen.



Zeitquantisierung – Abtasttheorem (Shannon)

13

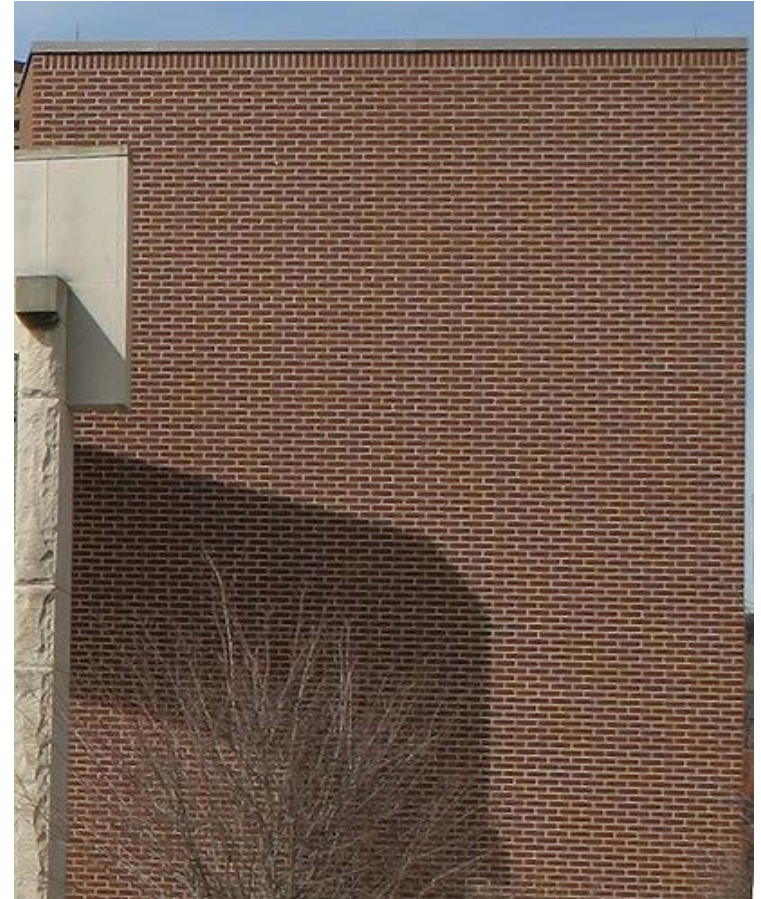
Kontinuierliche Zeitfunktionen müssen vor der Verarbeitung in Digitalrechnern quantisiert (diskretisiert) werden. Jede kontinuierliche Zeitfunktion $f(t)$ mit der oberen Grenzfrequenz $f_G = \frac{1}{T_G}$ wird eindeutig durch diskrete Werte $x(k \cdot T_A)$ mit $k = 0, \pm 1, \pm 2, \dots$ beschrieben, wenn $T_A < \frac{1}{2} T_G$ ($f_A = \frac{1}{T_A}$). Anderenfalls kann es zur Überlagerung von Signal- und Abtastfrequenz kommen, Stroboskopeffekt.





Nyquist-Shannon Sampling theorem

14

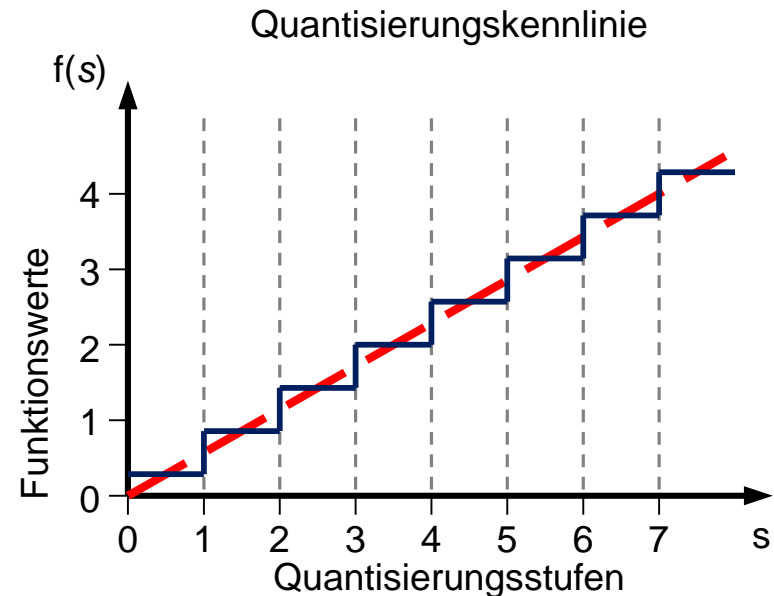
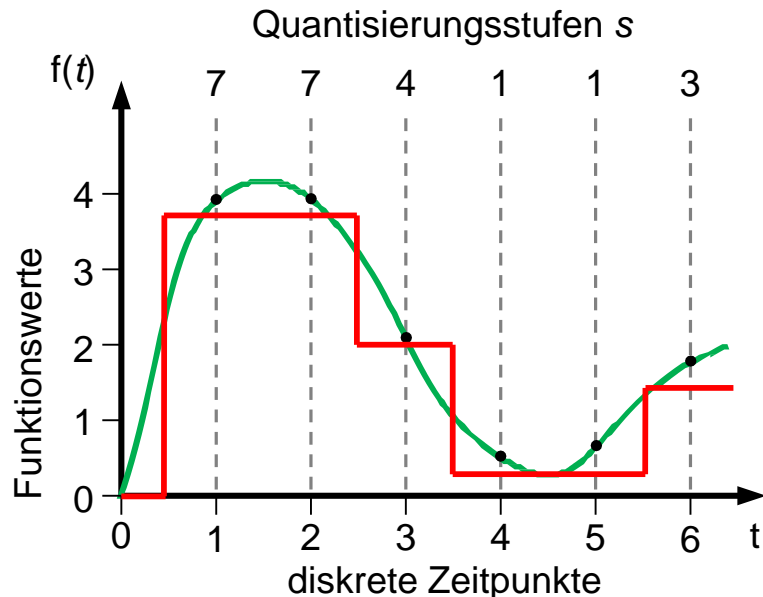


Source: Wikipedia

Wertquantisierung – Analog-Digital-Wandlung

15

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





Quantisierungsstufen – Quantisierungsintervall

16

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.

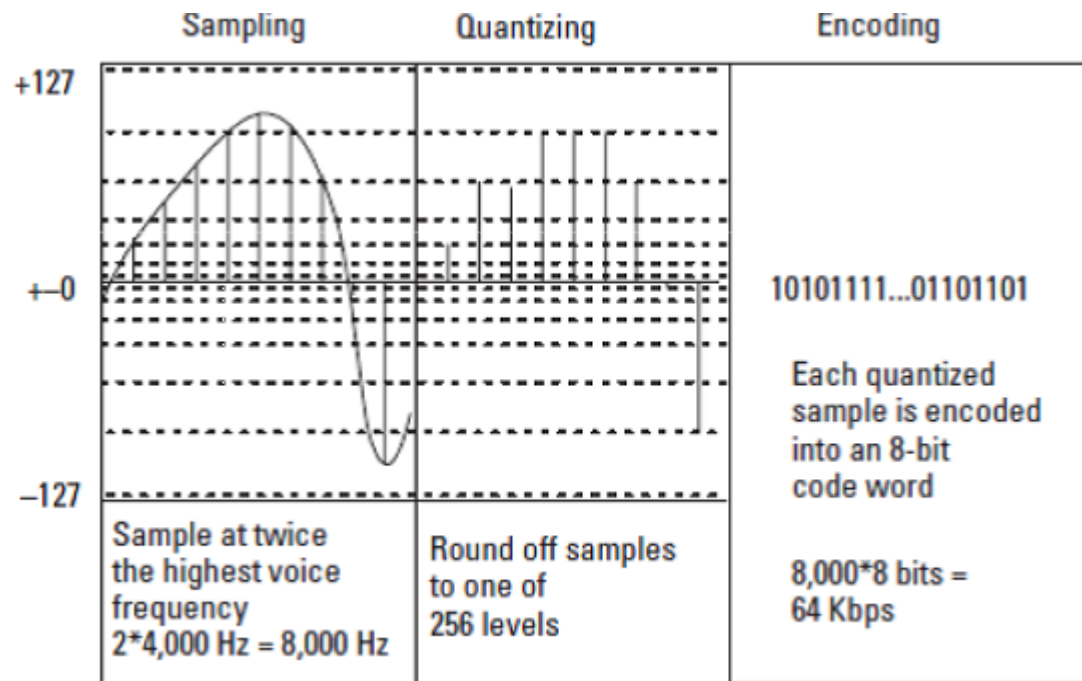
Die Größe der Quantisierungsintervalle ergibt sich aus der geforderten Abtastgenauigkeit für das Signal.

Sowohl durch die Zeit- wie auch 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.

Pulse Code Modulation

17

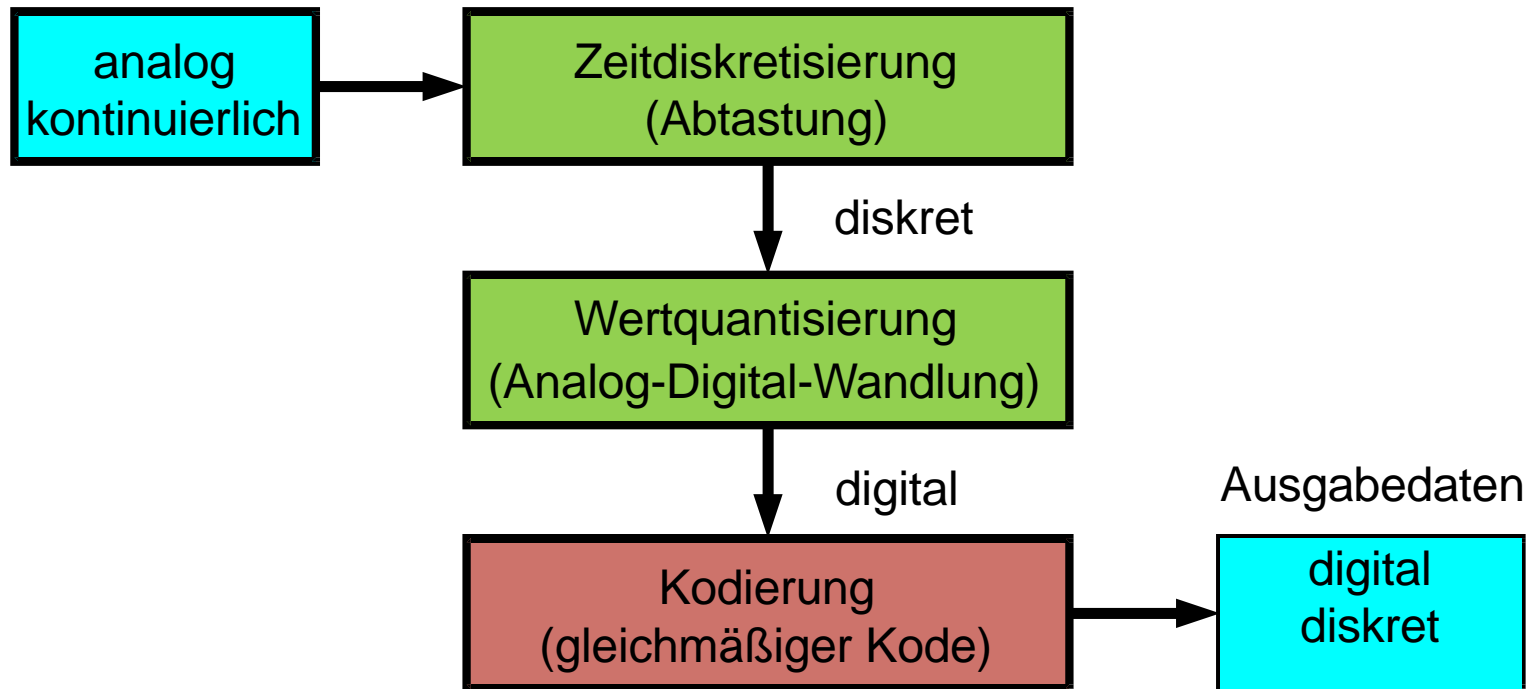


Source: <http://ecedunia.blogspot.com/2016/03/pulse-code-modulation.html>

Digitalisierung von analogen zeitkontinuierlichen Signalen

18

Eingangssignale



Beispiel digitale Aufzeichnung einer Audio-CD

19

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: 1,4 MBit/s Rohdaten
Redundanz: Zusatzdaten im Ausgabestrom zur Erkennung und Korrektur von Fehlern bei der Wiedergabe (Staubteilchen, Kratzer, ...).

Beispiel I SDN-Telekommunikation

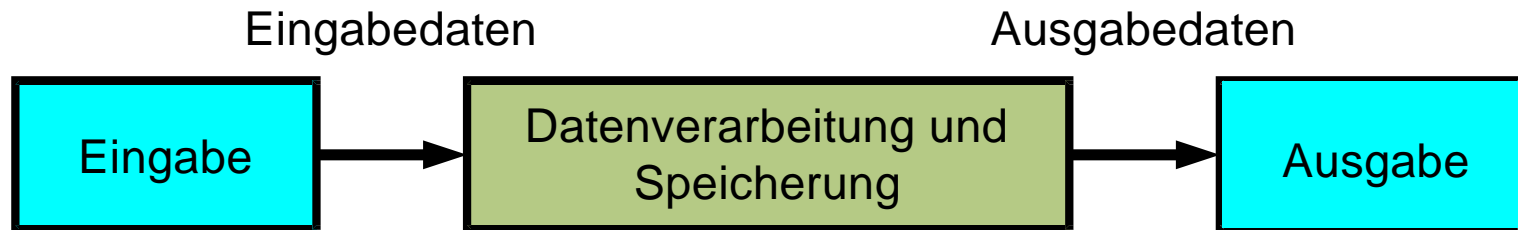
Übertragungsrate für Sprachkommunikation: A-Kanal, B-Kanal
Übertragungsrate: 64 kBit/s Datenstrom pro Kanal
Zeitdiskretisierung: 8 kHz Abtastfrequenz
Wertquantisierung: 8 Bit Auflösung = 256 Quantisierungsstufen

Informations-/Datenverarbeitung

20

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 (Steuerung) Ausführung von Befehlen auf Operanden. Die Darstellung von Steuerung, Befehlen und Operanden werden oft gleichermaßen als Daten bezeichnet und behandelt.



Computer – Informationsverarbeitung

21

Die Hardware eines Computers ist nur 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 (Operanden) Resultate der Verarbeitung (Ein-/Ausgabedaten, ...).

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

Steuerdaten: steuern und registrieren Arbeitsmodi und unterstützen den geordneten Programmablauf (Programmstatus, Interrupt, ...).

Darstellungsarten von Verarbeitungsdaten können sein:
numerische und alphanumerische Daten, Bild- und Tondaten, Sensordaten, Aktordaten, ...



Informationsdarstellung – Formate

22

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

Verarbeitungsdaten:

- ❑ Datenformat
- ❑ Darstellungsformat der Operanden

Programmdaten:

- ❑ Befehlsformat
- ❑ Darstellungsformat der Befehle (Instruktionen)

Steuerdaten:

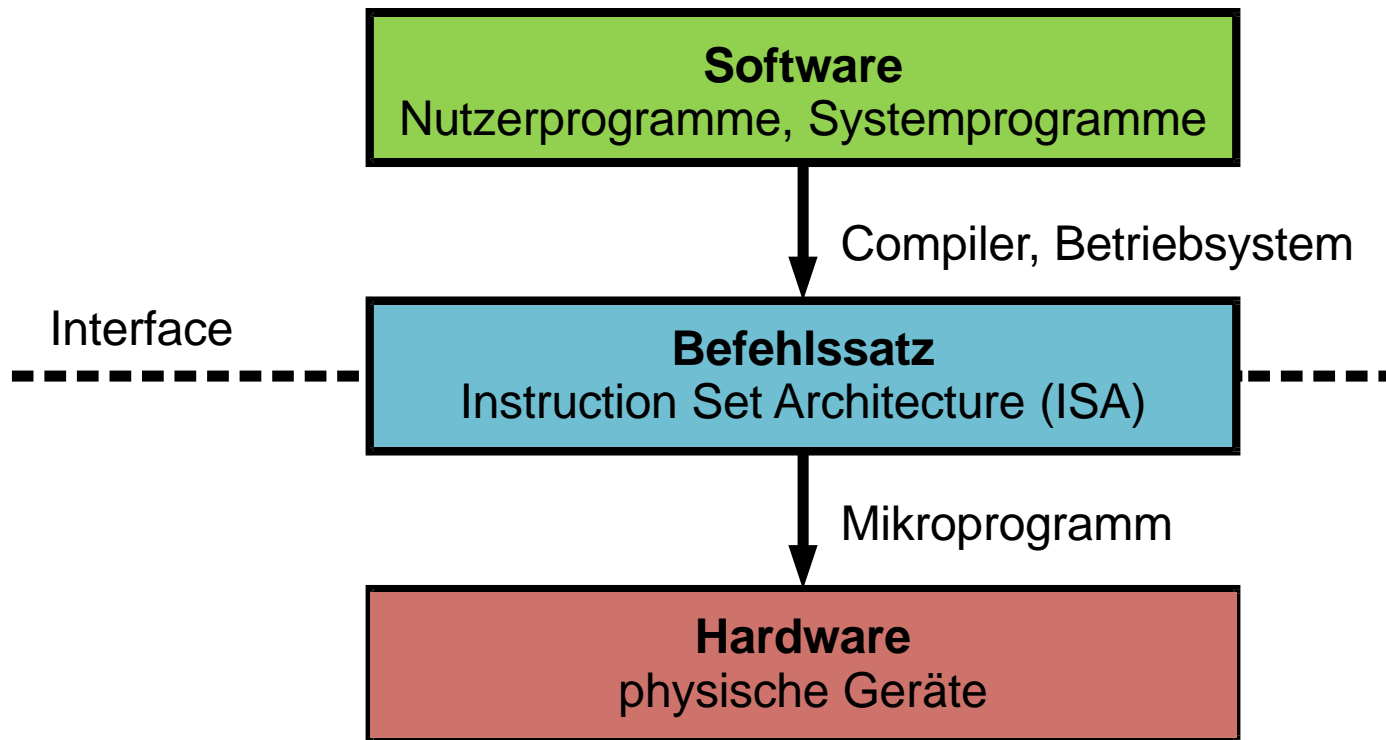
- ❑ Steuerformat
- ❑ Darstellung der Steuerungsinformationen

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 kodiert.

Die Daten eines Computers unterliegen einem strengen Ordnungsprinzip.

Schichtenmodell des Computers (einfach)

23





Hardware - Software

24

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

Zusammenfassung

25

- Information ist messbar, als Entropie einer Informationsquelle.
- Informationen werden durch physikalische Signale repräsentiert.
- Kodierung von Information als Zeichen oder Zeichenketten.
- Die Entropie einer Informationsquelle ist für die Beurteilung einer Kodierung von Bedeutung.
- Redundanz von Kodierung 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 deutlich, Befehlssatzebene.