

Rechnerarchitektur I

Informationsdarstellung

Technische Universität Dresden
Institut für Technische Informatik

R.G. Spallek

`rgs@ite.inf.tu-dresden.de`

Dresden, 2004

Inhalt

Punkt	Inhalt	Folie
2	Informationsdarstellung	2
2.1	Grundbegriffe	2
2.2	Zeichen, Zeichenketten, Alphabete	4
2.3	Information, Informationskette	5
2.4	Signale	10
2.5	Informations-/Datenverarbeitung	17
2.6	Schichtenmodell des Computers	20
2.7	Zusammenfassung Informationsdarstellung	22

2 Informationsdarstellung

2.1 Grundbegriffe

⇒ Informationstheorie

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

Entropie: 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 $1Sh = 1bit$ (nach Shannon) bezeichnet.

- 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.
- 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.
- 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.

2.2 Zeichen, Zeichenketten, Alphabete

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

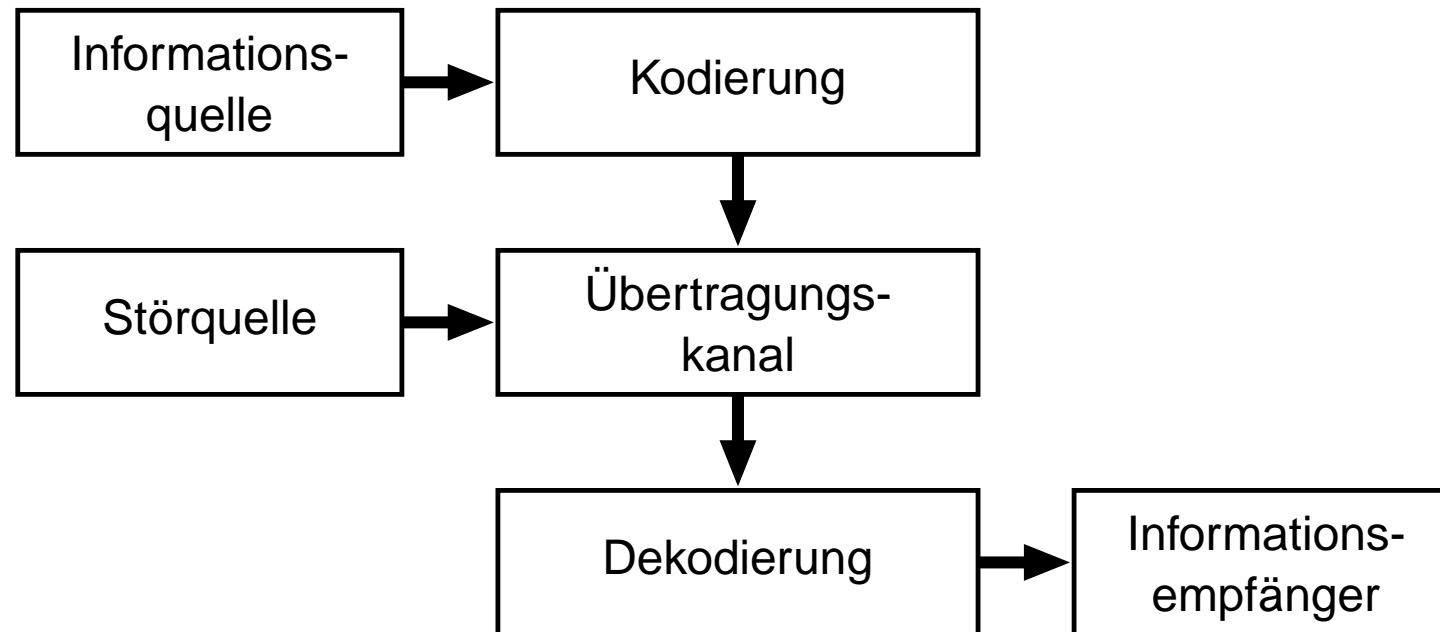
Unter einem Alphabet versteht man eine endliche Menge von Zeichen. Der Zeichenvorrat ist die Menge aller voneinander unterscheidbarer Zeichen.

Zeichen z_v eines Alphabetes $Z = \{z_1, z_2, \dots, z_n\}$ können zu Zeichenketten K beliebiger Länge zusammengefasst werden \rightarrow Folgen nicht notwendig verschiedener Zeichen z_v mit dem Verkettungsoperator \circ .

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

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.

2.3 Information, Informationskette (Shannon)



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

Kennzeichen, Definitionen einer Informationsquelle

- Vorhandensein einer endlichen Anzahl verschiedener Zeichen.
- Feste Wahrscheinlichkeit für das Erscheinen eines Zeichens im Zeichenstrom.
- 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 $p_i > 0$ mit $1 \leq i \leq n$

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

- Je seltener ein Zeichen ist, desto mehr Informationen enthält es.

Entropie einer Informationsquelle (Informationsgehalt)

$$H_Z(p_1, \dots, p_n) = \sum_{i=1}^n p_i * H_i$$

$$H_Z(p_1, \dots, p_n) = \sum_{i=1}^n p_i * \log_2 \frac{1}{p_i}$$

- $H(p_1, \dots, p_n)$ durchschnittlicher Informationsgehalt pro Zeichen
- Bedeutung für die Beurteilung von Kodierungen → effiziente Kodierungen
- $Z(p_1, \dots, p_n)$ stationäre gedächtnislose Zeichenquelle mit der Entropie H_Z
- Die Entropie einer Zeichenquelle wird in *bit / Zeichen* angegeben.

Ein Maximum an Information bei n verschiedenen Zeichen erhält man, wenn alle Zeichen gleichwahrscheinlich sind $p_i = \frac{1}{n} \rightarrow H_Z = \log_2 n$.

Redundanz

Die Redundanz R_C (Weitschweifigkeit, Überfluß) einer Kodierung gibt an, um wieviel der mittlere Kodierungsaufwand H_C vom mittleren Informationsgehalt der Zeichenquelle (Entropie H_Q) abweicht.

R_C	$=$	$H_C - H_Q$	Redundanz absolut
r_C	$=$	$\frac{H_C - H_Q}{H_C} = 1 - \frac{H_Q}{H_C}$	Redundanz relativ in %
η_C	$=$	$\frac{H_Q}{H_C} = 1 - r_C$	Informationsgehalt relativ in %
1	$=$	$r_C + \eta_C$	Summe

Redundanz ist für Fehlererkennung und -korrektur, zur Erhöhung der Übertragungssicherheit von Informationen erforderlich.

Beispiel: Informationsgehalt deutsches Alphabet

27 Zeichen (26 Buchstaben + Leerzeichen)

Zeichenwahrscheinlichkeiten unterschiedlich ($n = 27$):

$$p(e) = 14,5\%, p(n) = 8,7\%, p(s) = 6,5\%, p(q) = 0,02\%, p(y) = 0.01\%, \dots$$

$$H_u(p_1, \dots, p_n) = \sum_{i=1}^{27} p_i * \log_2 \frac{1}{p_i} = 4,04 \text{ bit / Zeichen}$$

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

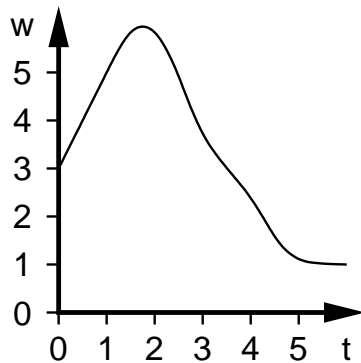
$$p(a) = \frac{1}{27}, p(b) = \frac{1}{27}, \dots, p(z) = \frac{1}{27}$$

$$H_g(p_1, \dots, p_n) = \sum_{i=1}^{27} \frac{1}{27} * \log_2 27 = 4,75 \text{ bit / Zeichen} > H_u = 4,04 \text{ bit / Zeichen}$$

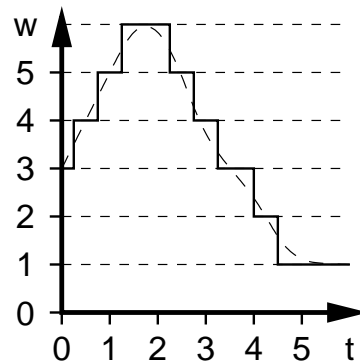
Bei gleichwahrscheinlichen Zeichen höherer Informationsgehalt pro Zeichen!

2.4 Signale

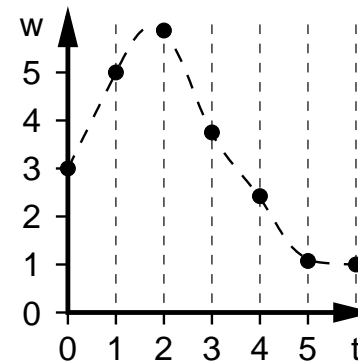
Zeit: kontinuierlich
Wert: analog



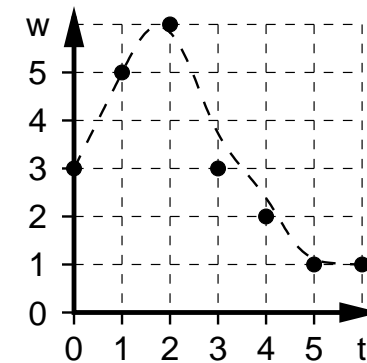
kontinuierlich
diskret



diskret
analog



diskret
diskret



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

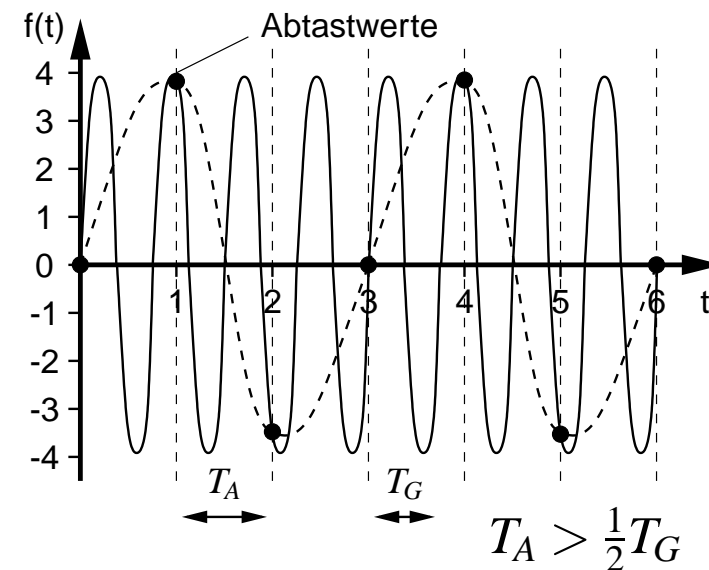
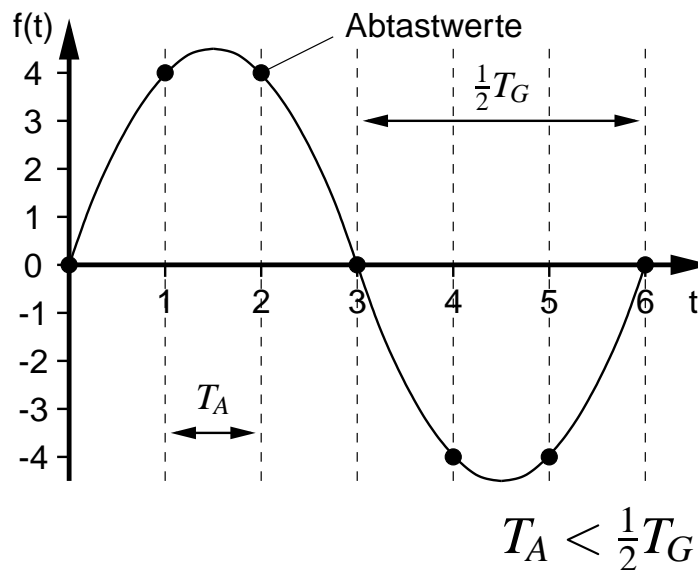
Digitale Signale sind zeit- und wertdiskret und allgemein binär kodiert (Zahlen).

Diskretisierung der Zeit: Abtasten der Zeitfunktion an diskreten Zeitpunkten

Diskretisierung der Werte: Quantisierung der Werte durch diskrete Stufen

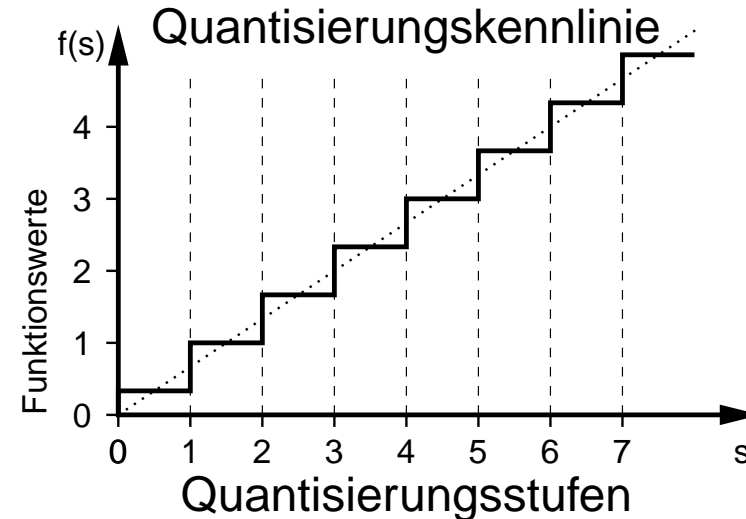
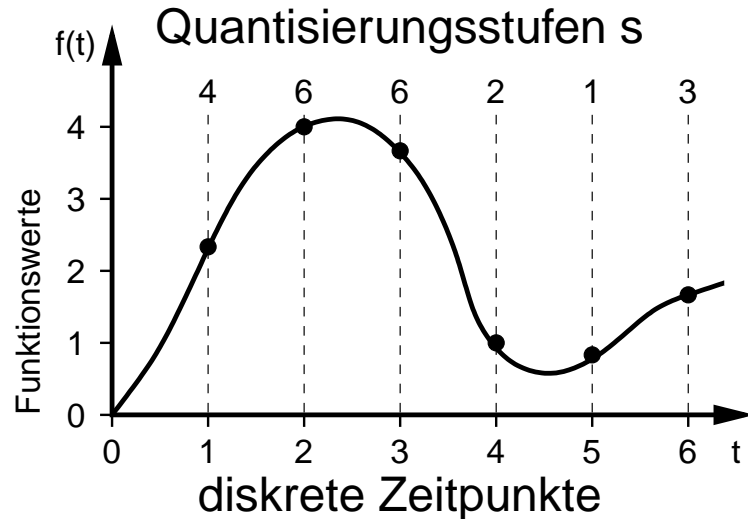
Zeitquantisierung – Abtasttheorem von Shannon

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).



Wertquantisierung – Analog-Digital-Wandlung

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 entsprechend einer Quantisierungskennlinie. Die Anzahl der dem Analogwert entsprechenden Quantisierungsstufen s kann binär kodiert dargestellt werden.



Quantisierungsstufen – Quantisierungsintervall

Die Amplitudenquantisierung beschreibt das Signal durch s diskrete (unterschiedliche) Quantisierungsstufen.

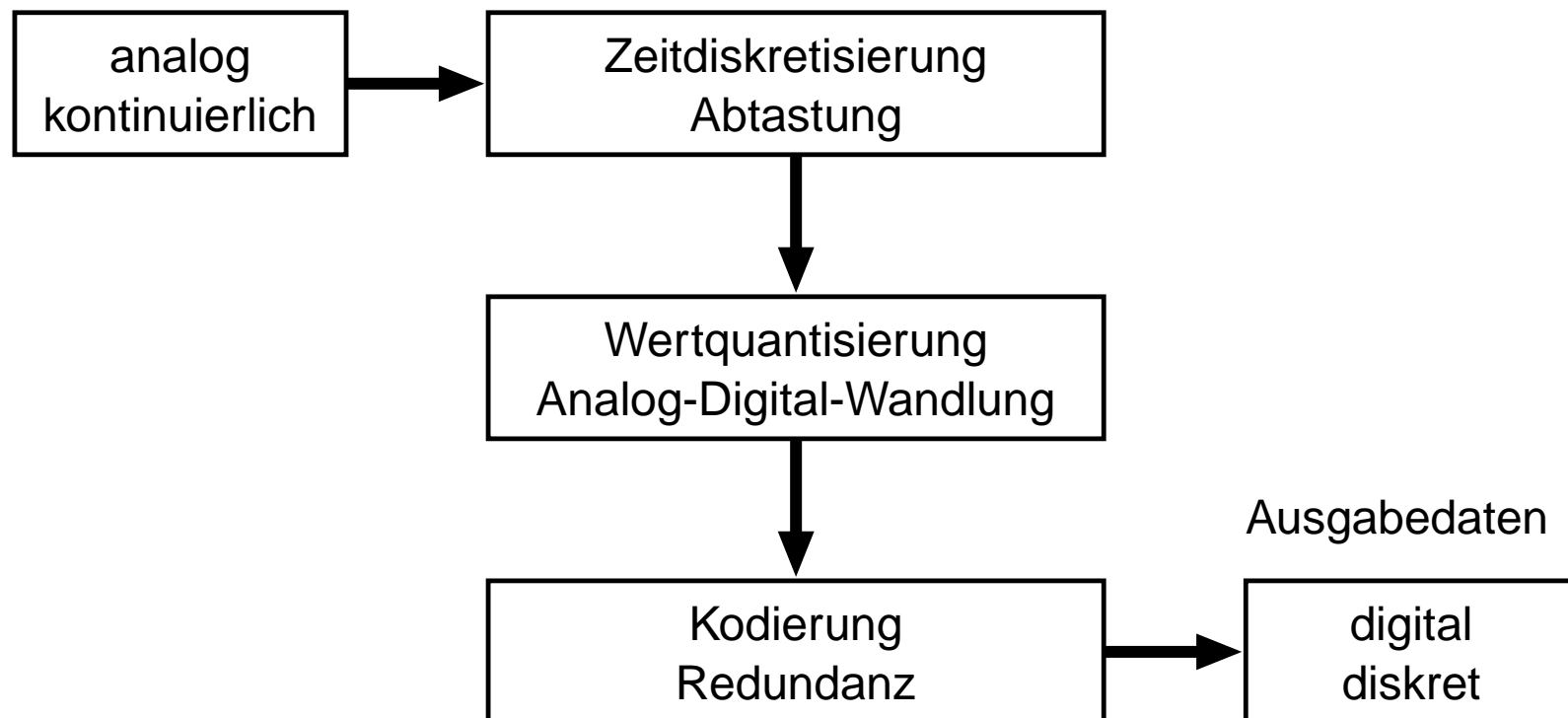
Die Anzahl der erforderlichen Quantisierungsstufen ist vom Wertebereich des Signals und der Größe der einzelnen Quantisierungsintervall abhängig.

Die Größe des Quantisierungsintervalles 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.

Digitalisierung analoger zeitkontinuierlicher Signale

Eingangssignale



Beispiel digitale Aufzeichnung Musik-CD

- obere Grenzfrequenz: $f_G \text{ ca. } 20000\text{Hz}$ (menschliches Hörvermögen)
- Abtastfrequenz: $f_A = 44100\text{Hz}$
- Digitalisierung der Audiowerte: 16-Bit (65536 mögliche diskrete Abtastwerte)
- Datenvolumen: $2 \cdot 16 \text{ Bit}$ pro Abtastzeitpunkt (2-Kanal-Stereo)
- Ausgabedatenstrom: $\rightarrow 1,4 \text{ MBit/s}$ Rohdaten
- Redundanz: Zusatzdaten im Ausgabestrom zur Korrektur von Fehlern bei der Wiedergabe (Staubteilchen, Kratzer, ...)

Beispiel ISDN-Telekommunikation

Übertragungskanäle für Sprachkommunikation: A-Kanal, B-Kanal

- Übertragungsrate : 64 kBit/s Datenstrom
- Zeitdiskretisierung : 8 kHz Abtastfrequenz
- Wertquantisierung : 8 Bit Auflösung = 256 Quantisierungsstufen

2.5 Informations-/Datenverarbeitung

Computer sind hochautomatisierte 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 (Instruktionen) auf Operanden (Rechengrößen). Die Darstellung von Steuerung, Befehlen und Operanden werden oft gleichermaßen als Daten bezeichnet und behandelt.

Computer – Informationsverarbeitung

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

- | | |
|---|--|
| Verarbeitungsdaten:
(Operanden) | umfassen die zu verarbeitenden Informationen und die Resultate der Verarbeitung (Ein-/Ausgabedaten, ...). |
| Programmdaten:
(Befehle, Instruktionen) | bestimmen Art und Reihenfolge der Verarbeitungsschritte (Operationen, Befehle, ...). |
| Steuerdaten: | steuern und registrieren Arbeitsmoden und unterstützen den geordneten Programmablauf (Programmstatus, Interrupt, ...). |

Darstellungsarten von Verarbeitungsdaten:

numerisch, alphanumerisch, Bild, Ton, Sensor, Aktuator ...

Informationsdarstellung – Formate

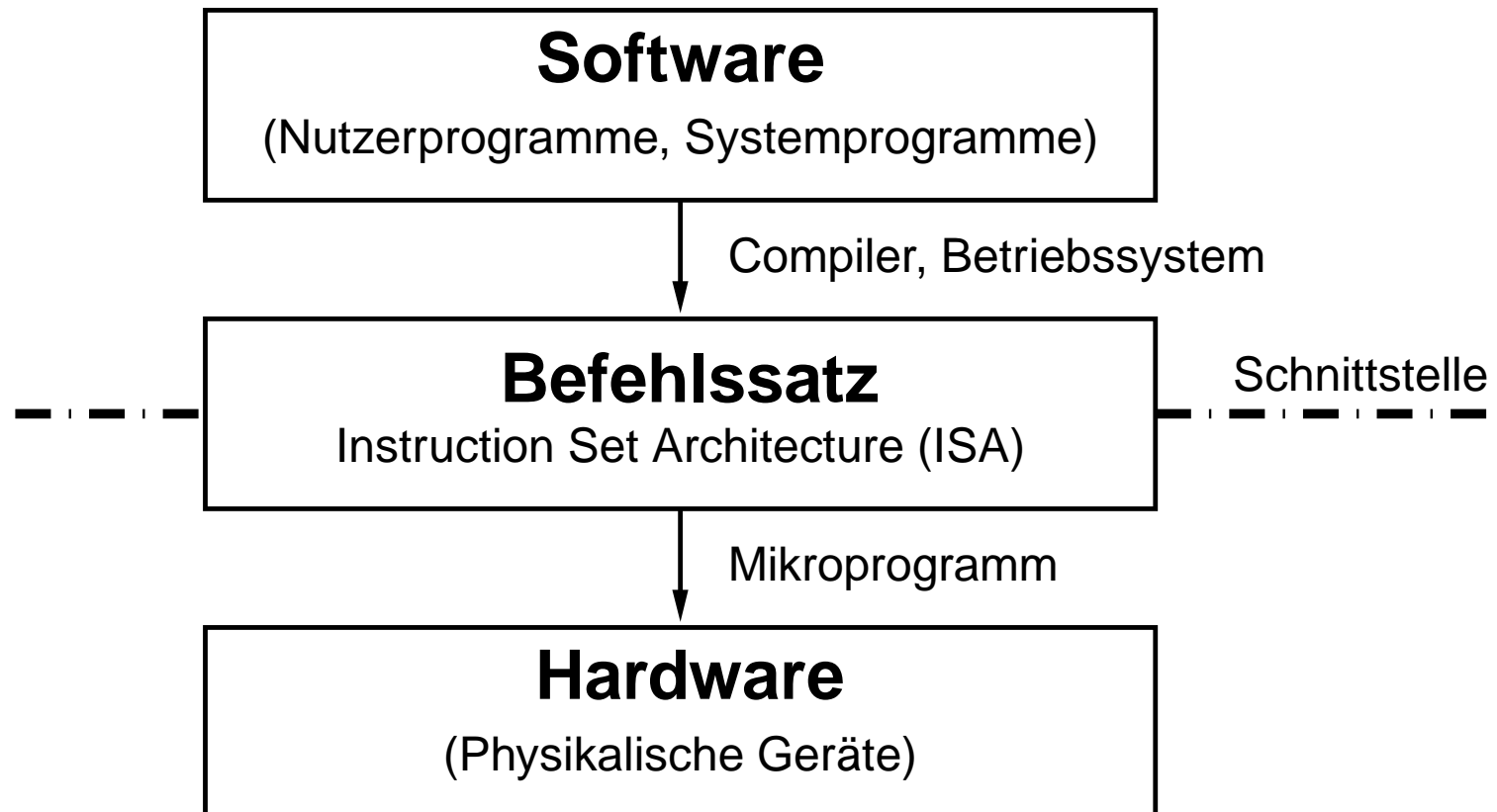
Die Darstellung der verschiedenen Informationen (Daten) im Computer erfolgt binär kodiert in festen Formaten.

Verarbeitungsdaten:	Datenformat Darstellungsformat der Operanden
Programmdaten:	Befehlsformat Darstellungsformat der Befehle (Instruktionen)
Steuerdaten:	Steuerformat Darstellungsformat 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.

2.6 Schichtenmodell des Computers (vereinfacht)



Hardware und Software

- Hardware:** umfasst alle mechanischen, elektrischen und elektronischen Bauelemente und Baugruppen (integrierte Schaltkreise, Leiterplatten, Netzteil, Gehäuse, ...).
- Software:** umfasst alle Programme, Betriebssysteme und Daten (Compiler, Editoren, ...).
- 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 hardwareprogrammierbare Bauelemente).

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

2.7 Zusammenfassung Informationsdarstellung

1. Information ist meßbar, Entropie einer Informationsquelle.
2. Informationen werden durch physikalische Signale repräsentiert.
3. Kodierung von Information als Zeichen, Zeichenketten.
4. Die Entropie einer Informationsquelle ist für die Beurteilung von Kodierungen von Bedeutung.
5. Redundanz von Kodierungen dient der Erhöhung der Datensicherheit.
6. Die Digitalisierung von Signalen erfolgt durch Zeitdiskretisierung und Wertquantisierung.
7. Computer sind hochautomatisierte technische Systeme der Informationsverarbeitung.
8. Die Verkopplung von Hardware und Software wird im Schichtenmodell des Computers deutlich.