

Einführung in die Wirtschaftsinformatik

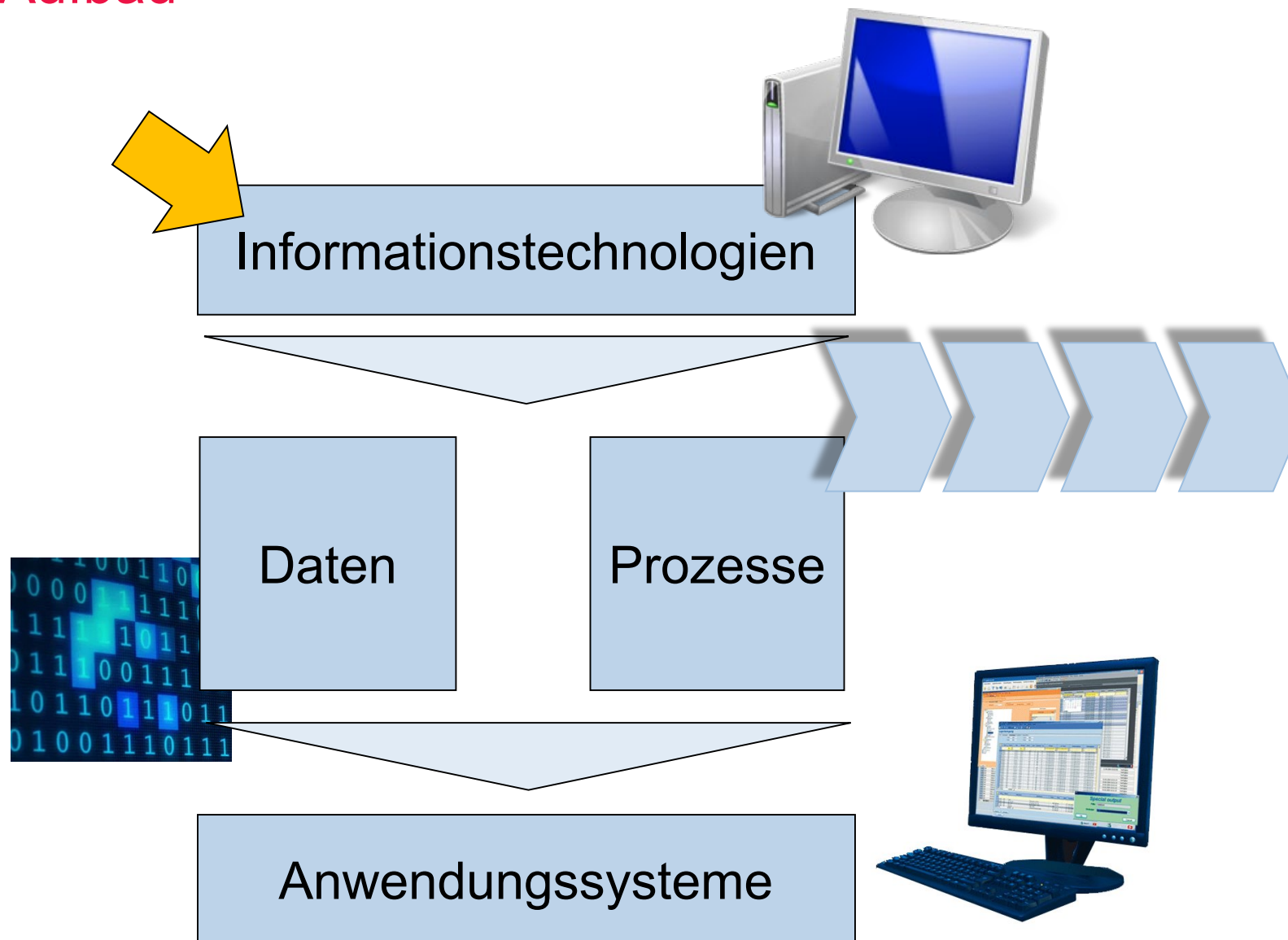
Grundlagen:

Rechner, Befehle und Daten

Prof. Dr. Thomas Myrach
Universität Bern
Institut für Wirtschaftsinformatik
Abteilung Informationsmanagement

Bern, 26. Februar 2020

Logischer Aufbau



- Sie wissen, welche Hauptkomponenten bei der Verarbeitung in einem Rechner eine Rolle spielen.
- Sie kennen Prozessoren und Speicherchips als relevante integrierte Schaltkreise.
- Sie können das Moore'sche Gesetz und dessen Bedeutung beschreiben.
- Sie wissen, dass zwischen der Menschenwelt und der binären Maschinenwelt Übersetzungen und Umcodierungen stattfinden müssen.
- Sie kennen den Unterschied zwischen Maschinencode, Assembler und höheren Programmiersprachen.
- Sie wissen was ein Datentyp ist und können Codierungen für verschiedene Datentypen erklären.
- Sie können Probleme nennen, die bei der Verwendung verschiedener Datentypen auftreten können.



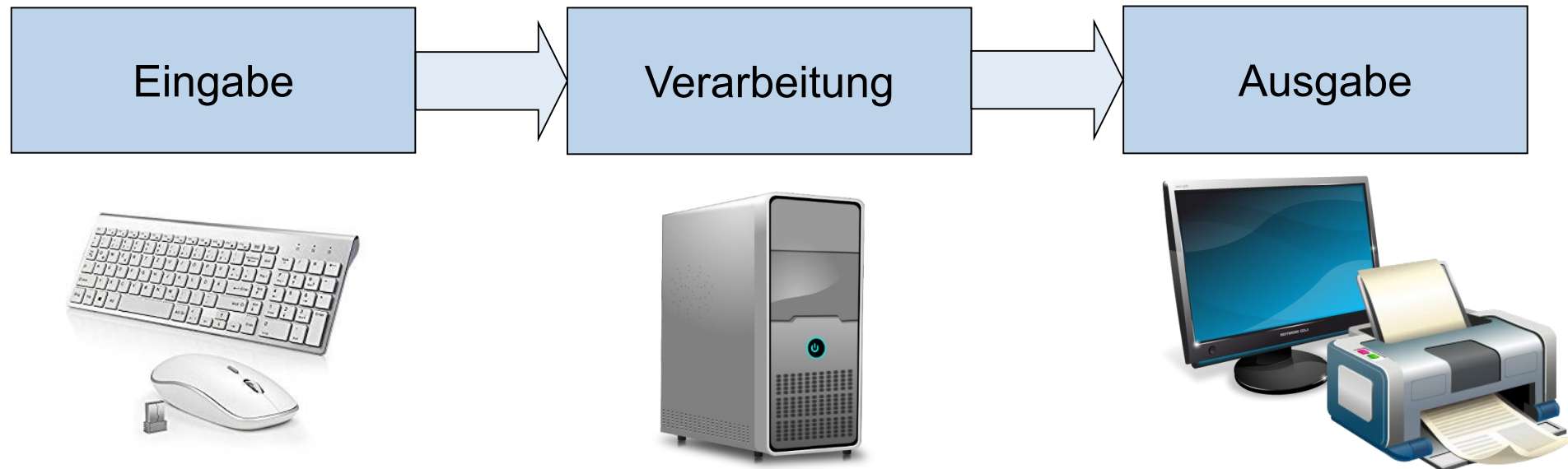
Hauptkomponenten eines Rechners

Codierung von Befehlen und Daten

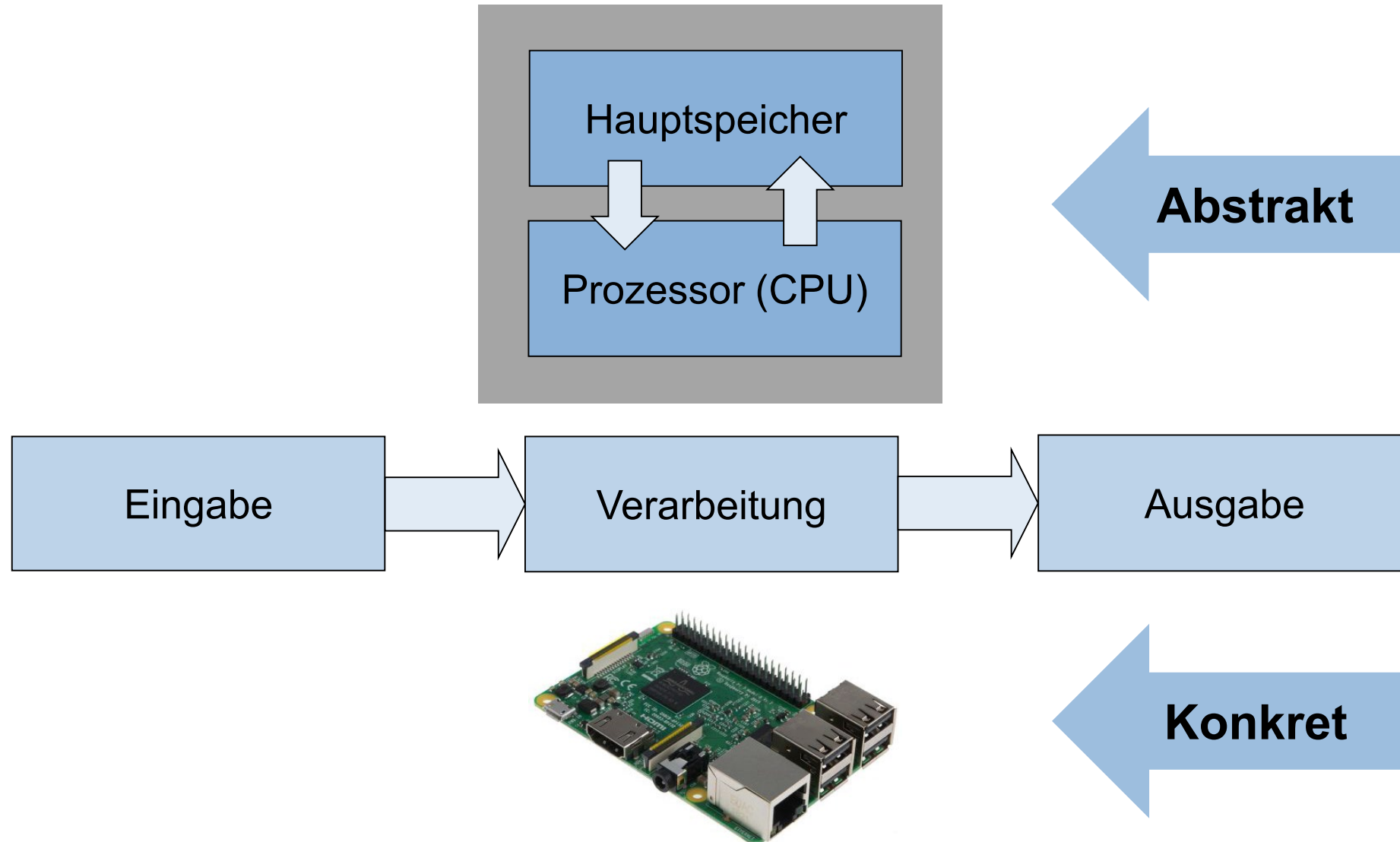
Peripheriegeräte

Rechnerbasierte Informationssysteme

- Rechner (Computer) sind die zentralen Bausteine rechnerbasierter Informationssysteme.
- Sie ermöglichen eine automatisierte Verarbeitung von Daten.



Rechnerbasierte Informationssysteme ("nackt")



- Besteht aus einzelnen adressierbaren Speicherzellen, die bei einem Zugriff gelesen oder beschrieben werden können.
- Eine Speicherzelle besteht aus n Bits ($n > 1$).
- Die ersten Personal Computer (PC) hatten 8 Bit (1 Byte) breite Speicherzellen.
- Heutige PC arbeiten mit einer Speicherzellen von 32 oder 64 Bit.
- Speicherzellen sind physisch in Speicherbausteinen enthalten
- Ein Speicherbaustein (Speicherchip) enthält einen integrierten Schaltkreis (IC).
- Speicherbausteine sind in verschiedenen Formen verfügbar.



Hardware

Prozessor als Marke

- Prozessoren sind die wichtigsten Bausteine eines Rechners.
- Bestimmte Befehlssätze sind vorgegeben.
- Programme müssen letztlich auf den verwendeten Prozessor ausgerichtet sein.
- Sie bestimmen, welche Software auf einem Rechner ablaufen kann.
- Hersteller und Prozessortypen sind von zentraler Bedeutung.
- Dies wird durch spezielle Markennamen angezeigt.

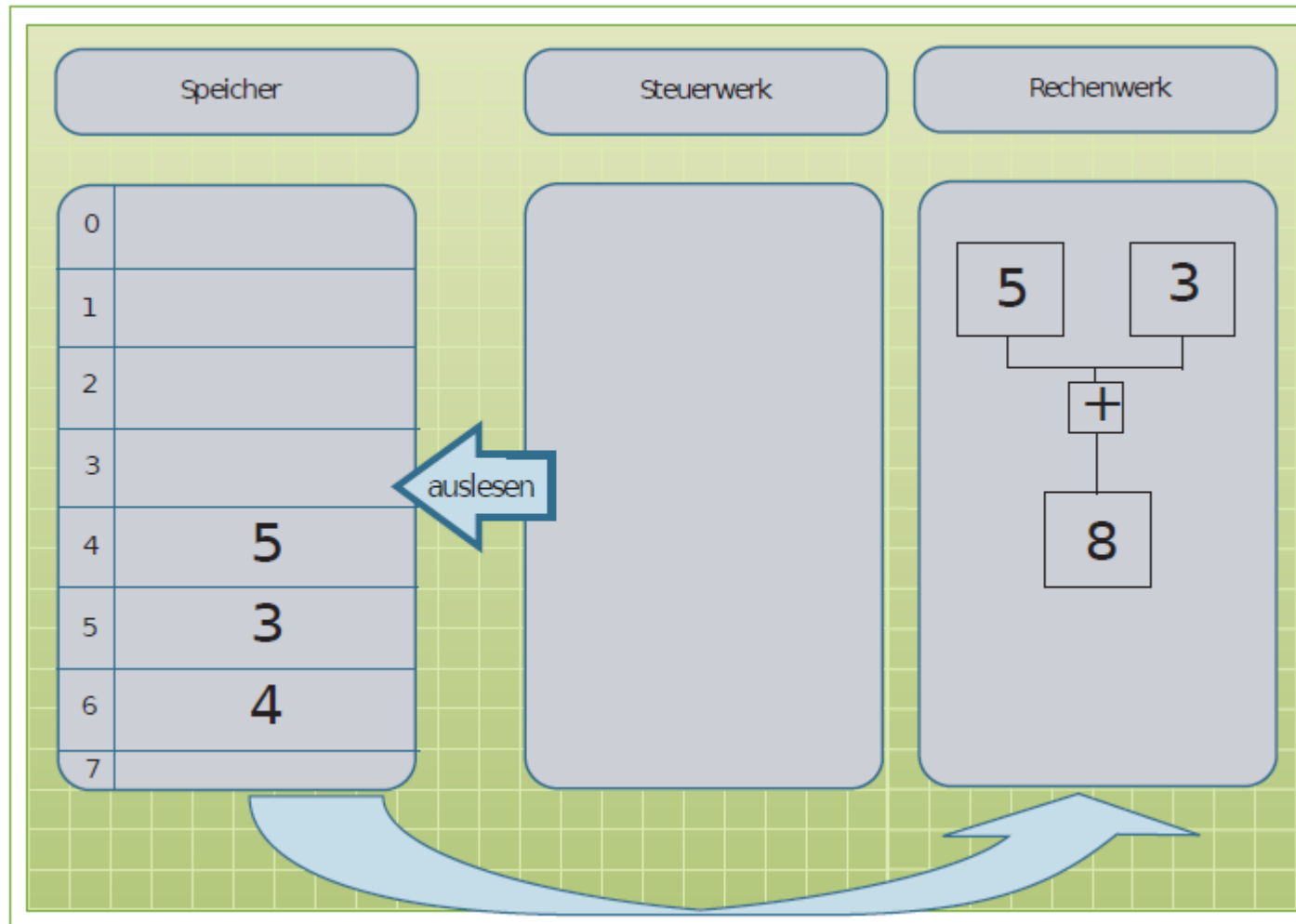


- Prozessoren verarbeiten Daten.
- Prozessoren unterstützen eine Reihe von elementaren Befehlen.
- Zwischen dem Prozessor und dem Hauptspeicher finden Interaktionen statt.
- Daten werden aus dem Speicher ausgelesen bzw. in den Speicher geschrieben.
- Durch die ausgeführte Operation ist festgelegt, welche Operanden benötigt werden.
- Diese werden vom Prozessor ausgelesen.

- Besteht aus Steuerwerk und Rechenwerk.
- Das **Steuerwerk** kontrolliert die Ausführung der Anweisungen.
 - Maschinenbefehle werden ins Befehlsregister geladen.
 - Maschinenbefehl im Befehlsregister werden vom Befehlsdecoder dekodiert und dann vom Rechenwerk ausgeführt.
- Das **Rechenwerk** führt die Elementaroperationen eines Prozessors durch.
 - Es kann sowohl arithmetische (etwa die Addition zweier Zahlen) als auch logische (etwa AND oder OR) Operationen ausführen.
 - Operationen können meist nur mit den Werten direkt ausgeführt werden, die in Arbeitsregistern enthalten sind.
- Bei einem Mikroprozessor sind alle Bausteine des Prozessors auf einem integrierter Schaltkreis (IC) vereinigt.

- Das Steuerwerk liest einen Code für einen Befehl aus.
- Gehört zum Befehl auch das Auslesen einer Speicherzelle, so wird diese vom Steuerwerk adressiert.
- Eingelesene Daten werden in das zuständige Prozessorregister kopiert.
- Das Rechenwerk führt die Operation aus.
- Das Rechenwerk erledigt die eigentliche Verarbeitung innerhalb des Prozessors, beispielsweise die Addition zweier Registerinhalte.
- Das Ergebnis wird wieder in einem der Prozessorregister abgelegt.
- Das Steuerwerk liest den Code für den nächsten Befehl aus.

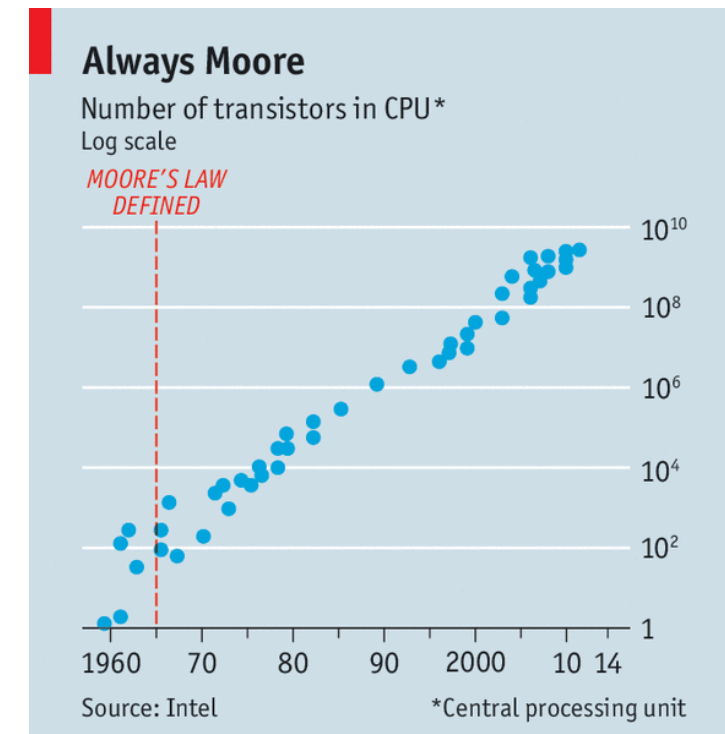
Operationen des Prozessors



Hardware

Moore's Law

- Die Komplexität integrierter Schaltkreise verdoppelt sich in regelmässigen Abständen.
- Komplexität wird ausgedrückt durch die Anzahl der Schaltkreiskomponenten auf einem integrierten Schaltkreis.
- Dadurch werden Rechner bei ungefähr gleichbleibenden Kosten immer leistungsfähiger.
- Diese technische Entwicklung ist Grundlage der digitalen Revolution.



Economist.com



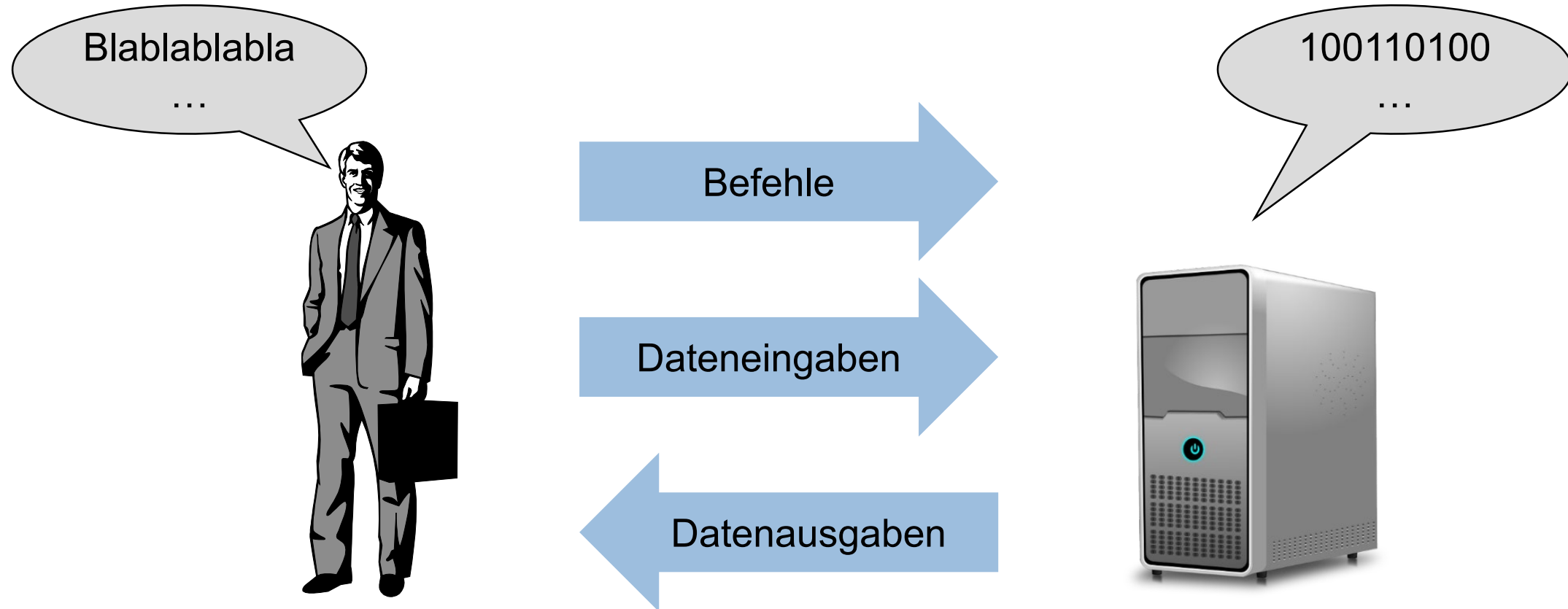
Hauptkomponenten eines Rechners

Codierung von Befehlen und Daten

Peripheriegeräte

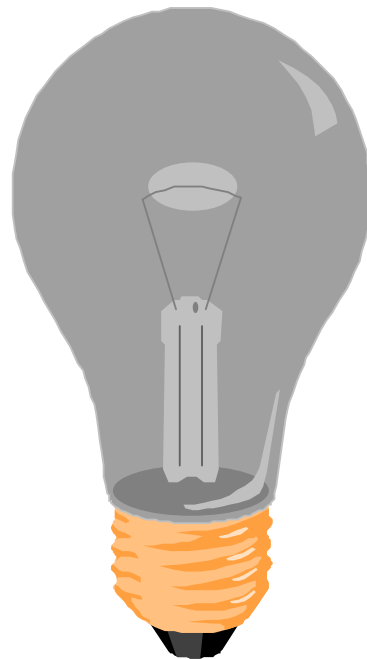
Abstraktionen

Interaktion Mensch-Maschine

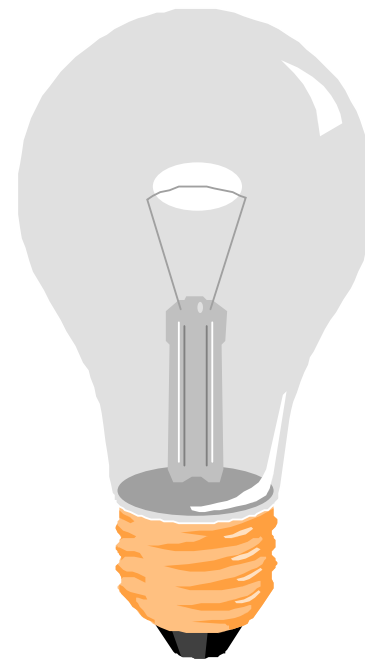


Grundlage der digitalen Welt: das Bit

AUS = 0



EIN = 1



Umrechnung Dezimal und Binär

1 Bit ($2^1 = 2$)

Dezimal	Binär
0	0
1	1

2 Bit ($2^2 = 4$)

Dezimal	Binär
0	00
1	01
2	10
3	11

3 Bit ($2^3 = 8$)

Dezimal	Binär
0	000
1	001
2	010
3	011
4	100
5	101
6	110
7	111

Je nach Anzahl der zusammengehörigen Bits
können unterschiedlich viele Zustände
dargestellt werden.

Speicherbreite

Anzahl der Werte

- Ursprünglich hatten Personal Computer eine Speicherbreite von 8 bit (ein Byte).
- Mit einem Byte können bis zu $2^8=255$ Werte dargestellt werden.
- Häutige Personal Computer haben eine Speicherbreite von 32 bit oder 64 bit.

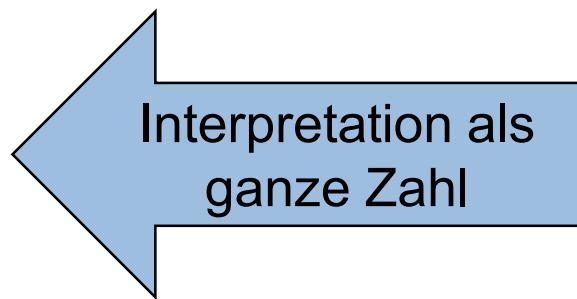
Speicherbreite	Anzahl der Werte
2^8	255
2^{16}	65'536
2^{32}	4'294'967'296
2^{64}	18'446'744'073'709'551'616

Bedeutung der Binärwerte

Befehle und Daten

- Die in einem Computer abgelegten Binärwerte können für unterschiedliche Konzepte stehen.
 - Befehle
 - Daten
- Die Interpretation der Binärwerte durch den Prozessor ergibt sich aus dem Kontext.

#	Dezimal
4	5
5	3
6	4

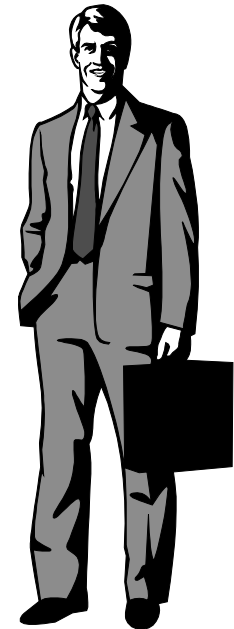


#	Binär
4	00000101
5	00000011
6	00000100

- Der Befehlssatz eines Prozessors besteht aus sehr elementaren Befehlen.
- Diese werden als binäre Maschinenbefehle im Speicher abgelegt und vom Steuerwerk sukzessive aufgerufen.
- Der Befehlsdecoder übersetzt die binären Maschinenbefehle mit Hilfe der Befehlstabelle in entsprechende Anweisungen.
- Diese Anweisungen werden vom Rechenwerk ausgeführt.
- Die Realisierung von Algorithmen erfordert typischerweise etliche Befehle.



- Maschinensprache
 - Direkte Anweisungen an den Prozessor in Maschinencode.
- Assemblersprache
 - Anweisungen an den Prozessor über Befehlskürzel (Mnemonics).
- Höhere Programmiersprachen
 - Anweisungen über mächtigere Befehle, die sich an der menschlichen Logik orientieren.
 - Losgelöst vom konkreten Prozessor.



Programmiersprachen

Vergleich Maschinencode, Assemblercode, C-Code

Maschinencode (hexadezimal)	zugehöriger Assemblercode	zugehöriger C-Code	Erläuterung
55 48 89 E5	<code>push rbp</code> <code>mov rbp, rsp</code>	<code>int main() {</code>	Sichere Register RBP auf dem Stack und setze RBP auf den Wert von Register RSP, dem Stackpointer (gehört nicht zur eigentlichen Berechnung). Diese Vorbereitung ist notwendig, um die Werte der Variablen <i>a</i> , <i>b</i> und <i>c</i> auf dem Stack speichern zu können.
C7 45 FC 02	<code>mov DWORD PTR [rbp-4], 2</code>	<code>int a = 2;</code>	Setze Variable <i>a</i> , die durch Register RBP adressiert wird, auf den Wert 2.
C7 45 F8 03	<code>mov DWORD PTR [rbp-8], 3</code>	<code>int b = 3;</code>	Setze Variable <i>b</i> , die durch Register RBP adressiert wird, auf den Wert 3.
8B 45 F8 8B 55 FC 01 D0 89 45 F4	<code>mov eax, DWORD PTR [rbp-8]</code> <code>mov edx, DWORD PTR [rbp-4]</code> <code>add eax, edx</code> <code>mov DWORD PTR [rbp-12], eax</code>	<code>int c = a + b;</code>	Setze Register EAX auf den Wert von Variable <i>b</i> . Setze Register EDX auf den Wert von Variable <i>a</i> . Addiere den Wert von EDX zum Wert von EAX. Setze Variable <i>c</i> , die durch RBP adressiert wird, auf den Wert von EAX.
8B 45 F4	<code>mov eax, DWORD PTR [rbp-12]</code>	<code>return c;</code>	Setze Register EAX auf den Wert von Variable <i>c</i> . Weil Register EAX diesen Wert bereits enthält, könnte diese Anweisung in einem optimierten Programm entfallen.
5D C3	<code>pop rbp</code> <code>ret</code>	<code>}</code>	Setze RBP wieder auf seinen ursprünglichen Wert. Springe zurück an die Stelle des Aufrufs von <i>main</i> . Register EAX enthält den Rückgabewert.

- Höhere Programmiersprache sehen Befehle vor, die nicht direkt vom Prozessor unterstützt werden.
- Die Befehle müssen in Maschinensprache übersetzt werden, damit sie der Prozessor interpretieren kann.
- Interpreter
 - Der Programmcode wird direkt während des Ablaufs in Maschinencode übersetzt.
 - Dazu wird der Programmcode aufgerufen.
- Compiler
 - Der Programmcode wird im Zuge eines Übersetzungsvorgangs in Maschinencode übertragen.
 - Dieser ausführbare Maschinencode wird in eine Datei geschrieben, die anschliessend ausgeführt werden kann.

- Daten werden binär im Speicher abgelegt und vom Steuerwerk im Zuge der Befehlsverarbeitung aufgerufen.
- Daten können einen unterschiedlichen Typ haben:
 - Ganze Zahlen
 - Dezimalzahlen
 - (Text-)Zeichen
 - Datumsangaben
- Je nach Datentyp werden die Daten unterschiedlich codiert.
- Für die korrekte Verarbeitung der Daten ist es also wichtig, den Datentyp zu kennen.

- Die interne Repräsentation von Daten wird festgelegt durch Datentypen.
- Durch einen Datentyp wird bestimmt
 - wie Daten im Computer abgebildet werden
 - welche Werte zulässig sind
 - wie Daten dem Benutzer dargestellt werden
 - welche Operationen mit den Daten zulässig sind
- Üblicherweise steht eine Reihe von vordefinierten Datentypen zur Verfügung, die gewählt werden können.
- In einigen Umgebungen werden Datentypen implizit vom System zugewiesen.

Datentypen

Definition von Daten

- Damit Daten im Rahmen von Programmen verarbeitet werden können, müssen sie entsprechend definiert werden.
- Dies geschieht in höheren Programmiersprachen durch die Definition eines Bezeichners, dem explizit oder implizit ein Datentyp zugeordnet wird.
- Dadurch wird intern ein entsprechender Speicherplatz reserviert.
- Über den Bezeichner wird der Speicherplatz adressiert.

integer a	Ganze Zahl, die durch den Bezeichner a adressiert wird.
real b	Reale Zahl, die durch den Bezeichner b adressiert wird.
char c	Zeichen, das durch den Bezeichner c adressiert wird.

Datentypen

Ganze Zahlen (Integer)

- Der Wertebereich ist endlich.
- Verschiedene Varianten bezüglich der reservierten Speicherbreite.
- Positive und negative Zahlen werden üblicherweise durch Zweierkomplementzahlen ausgedrückt.
- Bei Überschreiten des zulässigen Wertebereichs tritt ein Überlauf auf.

Dezimalzahl	Binärzahl (Zweierkomplement)
Maximum: 255	0 1111 1111
1	0 0000 0001
0	0 0000 0000
-1	1 1111 1111
Minimum: -256	1 0000 0000

Datentypen

Überlaufproblem



- Psy hat mit seinem Lied "Gangnam Style" den Zugriffszähler des Videoportals YouTube überlaufen lassen.
- Für den Zugriffszähler wurde ein 32-Bit-Integer-Wert verwendet.
- Damit konnten Zahlen bis zu 2.147.483.647 abgebildet werden.
- Als "Gangnam Style" diesen Wert überschritt, wechselte der Code aus diesem Grund das Vorzeichen.
- YouTube hat nun den Code auf 64 Bit umgestellt.
- Damit können bis zu neun Trillionen Zugriffe abgebildet werden.

Datentypen

Zahlen mit Nachkommastellen

- Festkommazahlen
 - Besteht aus einer festen Anzahl von Ziffern.
 - Die Position des Kommas ist dabei fest vorgegeben.
 - Die Darstellung einer Zahl ist im gesamten Wertebereich exakt.
- Gleitkommazahlen
 - Eine angenäherte Darstellung einer reellen Zahl.
 - Erfolgt typischerweise in Exponentialschreibweise.
 - Die Darstellung im Computer besteht aus einem Vorzeichen-Bit, einigen Exponenten-Bits und einigen Mantissen-Bits.
 - Wertebereich ist nicht begrenzt, dafür sind die Werte nicht immer exakt.

Codierung von Zeichen

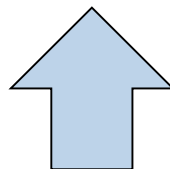
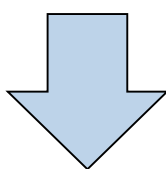
Beispiel: ISO 7-bit Code

0	NUL	16	DLE	32	SPC	48	0	64	@	80	P	96	`	112	p
1	SOH	17	DC1	33	!	49	1	65	A	81	Q	97	a	113	q
2	STX	18	DC2	34	"	50	2	66	B	82	R	98	b	114	r
3	ETX	19	DC3	35	#	51	3	67	C	83	S	99	c	115	s
4	EOT	20	DC4	36	\$	52	4	68	D	84	T	100	d	116	t
5	ENQ	21	NAK	37	%	53	5	69	E	85	U	101	e	117	u
6	ACK	22	SYN	38	&	54	6	70	F	86	V	102	f	118	v
7	BEL	23	ETB	39	'	55	7	71	G	87	W	103	g	119	w
8	BS	24	CAN	40	(56	8	72	H	88	X	104	h	120	x
9	HT	25	EM	41)	57	9	73	I	89	Y	105	i	121	y
10	LF	26	SUB	42	*	58	:	74	J	90	Z	106	j	122	z
11	VT	27	ESC	43	+	59	;	75	K	91	[107	k	123	{
12	FF	28	FS	44	,	60	<	76	L	92	\	108	l	124	
13	CR	29	GS	45	-	61	=	77	M	93]	109	m	125	}
14	SO	30	RS	46	.	62	>	78	N	94	^	110	n	126	~
15	SI	31	US	47	/	63	?	79	O	95	_	111	o	127	DEL

Codierung von Zeichen

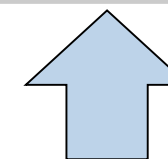
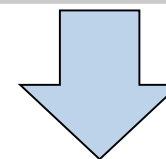
Umrechnung

Zeichen	Dezimal
A	65



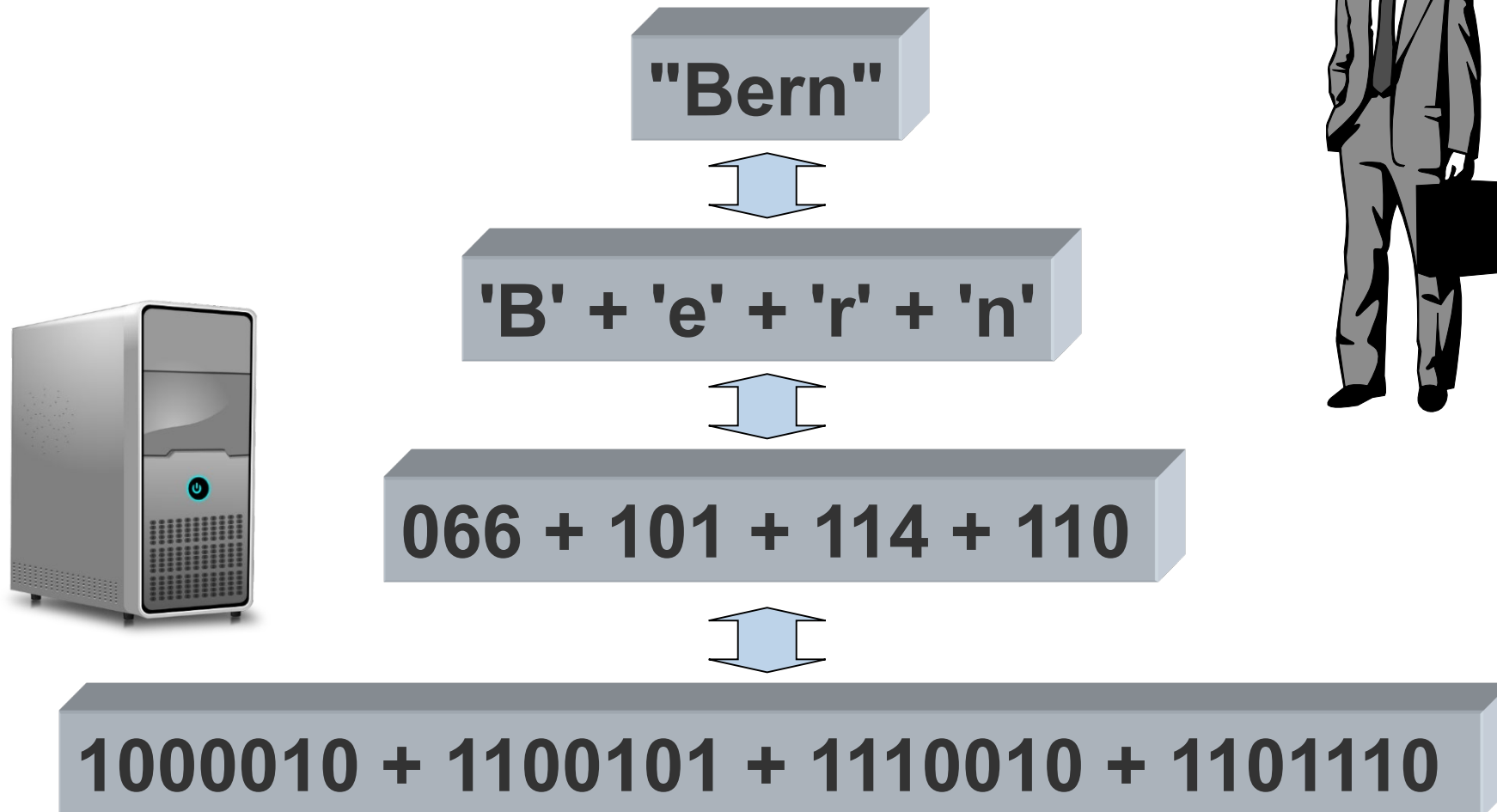
0	NUL	16	DLE	32	SPC	48	0	64	@	80	P	96	`	112	p
1	SOH	17	DC1	33	!	49	1	65	A	81	Q	97	a	113	q
2	STX	18	DC2	34	"	50	2	66	B	82	R	98	b	114	r
3	ETX	19	DC3	35	#	51	3	67	C	83	S	99	c	115	s
4	EOT	20	DC4	36	\$	52	4	68	D	84	T	100	d	116	t
5	ENQ	21	NAK	37	%	53	5	69	E	85	U	101	e	117	u
6	ACK	22	SYN	38	&	54	6	70	F	86	V	102	f	118	v
7	BEL	23	ETB	39	'	55	7	71	G	87	W	103	g	119	w
8	BS	24	CAN	40	(56	8	72	H	88	X	104	h	120	x
9	HT	25	EM	41)	57	9	73	I	89	Y	105	i	121	y
10	LF	26	SUB	42	*	58	:	74	J	90	Z	106	j	122	z
11	VT	27	ESC	43	+	59	;	75	K	91	[107	k	123	{
12	FF	28	FS	44	,	60	<	76	L	92	\	108	l	124	
13	CR	29	GS	45	-	61	=	77	M	93]	109	m	125	}
14	SO	30	RS	46	.	62	>	78	N	94	^	110	n	126	~
15	SI	31	US	47	/	63	?	79	O	95	_	111	o	127	DEL

Dezimal	Binär
65	1000001



Zahl	Quotient	Rest
65:2	32	1
32:2	16	0
16:2	8	0
8:4	4	0
4:2	2	0
2:2	1	0
1:2	0	1

Codierung von Zeichenketten



Codierung von Zeichenketten

Problem Sonderzeichen

- Die genormte Zeichentabelle beruht ursprünglich auf dem Standard der US-Amerikanischen Normungsbehörde (ASCII).
- Sonderzeichen sind dabei nicht vorgesehen.
- Immer wieder kommt es bei diesen zu Darstellungsproblemen.

Bildnummer	1412178
Bildtitel *	Geschäftsfrau hält Daumen hoch
Hintergrund- informationen zum Bild	Zufriedene Geschäftsfrau hält ihren Daumen nach oben
Keywords * trennen mit ,	Frau; Daumen; hoch; Erfolg; Gratulation; Geschäftsfrau; glücklich; Gewinner; Business; Kommunikation; Leistung; Sieger; Zustimmung; Manager; gewinnen; lachen; Freude; Studentin; Azubi; Auszubildende; Hand; lächelnd; applaudieren; freudig; Sieg; Gewinn;
Aufnahmeort	Köln, Nordrhein-Westfalen, Deutschland

Weit hinten, hinter den Wortbergen, fern der Länder
Vokalien und Konsonantien leben die Blindtexte.
Abgeschieden wohnen Sie in Buchstabenhäusern an der
Küste des Semantik, eines großen Sprachozeans. Ein
kleines Böschlein namens Duden fließt durch ihren Ort
und versorgt sie mit den nötigen Regelialien. Es ist ein
paradiesmatisches Land, in dem einem gebratene Satzteile
in den Mund fliegen. Nicht einmal von der allmächtigen
Interpunktion werden die Blindtexte beherrscht ? ein
geradezu unorthographisches Leben. Eines Tages aber
beschloß eine kleine Zeile Blindtext, ihr Name war Lorem
Ipsum, hinaus zu gehen in die weite Grammatik. Der
große Oxymer riet ihr davon ab, da es dort wimmelte von

Codierung von Zeichenketten

Weitreichende Konsequenzen



- Die Bundesverwaltung nutzt zur Erfassung von Personendaten verschiedene Informationssysteme.
- Den wichtigsten und grössten Datensatz verwaltet das Bundesamt für Justiz im Personenstandsregister Infostar.
- Infostar erfasst die komplette Schweizer Wohnbevölkerung und sämtliche Auslandschweizer.
- Ausgerechnet das System Infostar verwendet als einziges einen reduzierten Zeichensatz ohne das ć.
- Die Tabelle beruht auf der weltweit angewendeten ISO-Norm 8859-15 für westeuropäischen Sprachen.
- Hätte man sich für eine andere Richtlinie der Normgruppe entschieden, könnte man den Namen von Robert Matešić heute fehlerfrei in seinen Schweizer Pass schreiben.



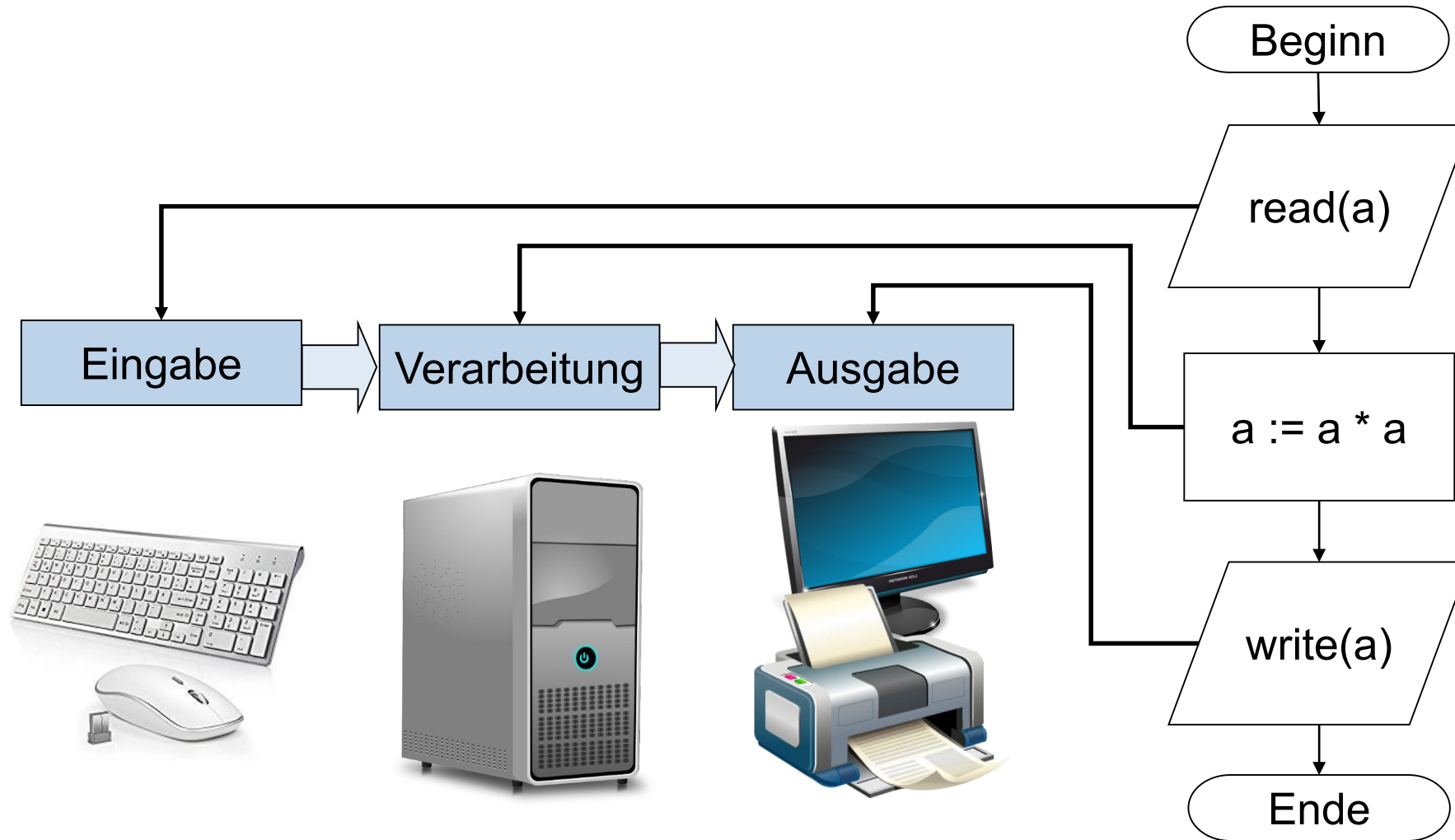
Hauptkomponenten eines Rechners

Codierung von Befehlen und Daten

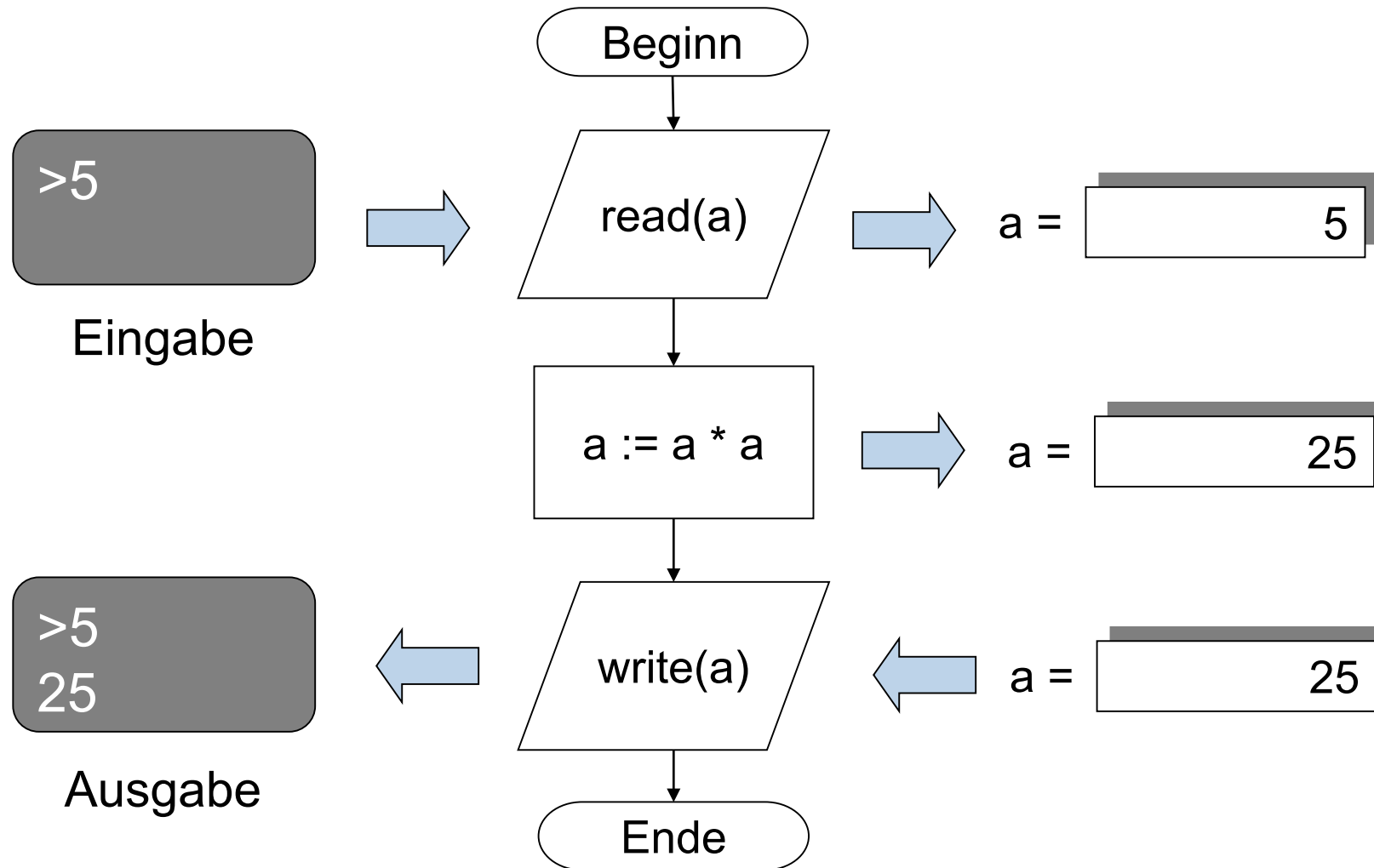
Peripheriegeräte

- Im Rahmen von Programmen muss es möglich sein, mit unterschiedlichen Daten zu arbeiten.
- Die in einem Programm verwendeten Daten müssen eingelesen und ausgegeben werden können.
- Dazu stehen in Programmiersprachen entsprechende Programmkonstrukte zur Verfügung.
- Als Standard-Eingabemechanismus ist die Tastatur anzusehen.
- Als Standard-Ausgabemechanismus gilt der Monitor.
- Die Eingabe oder Ausgabe von Daten kann auch ein Speichermedium betreffen.

Input-Process-Output



Flussdiagramm



Externer Speicher

- Im Rahmen eines Rechnersystems stehen typischerweise externe Speicher zur Verfügung.
- Dies hat verschiedene Vorteile:
 - Die beschränkte Kapazität des Hauptspeichers wird ausgeweitet.
 - Die Daten auf externen Speichern sind persistent.
- Extern heisst, dass der Prozessor nicht direkt auf den Speicher zugreifen kann.



- Primärspeicher
 - Üblicherweise der Arbeits- oder Hauptspeicher eines Rechners.
 - Direkter Zugriff durch den Prozessor.
 - Daten sind zum grossen Teil nicht persistent.
- Sekundarspeicher
 - Schneller externer Festspeicher.
 - Üblicherweise schneller Zugriff.
 - Wird für die dauerhafte (persistente) Aufbewahrung von Daten verwendet.
- Tertiärspeicher
 - Externer Festspeicher
 - Üblicherweise langsamer und nicht notwendigerweise permanenter Zugriff.
 - Wird als Archivierungsmedium gebraucht.

- Menge zusammengehöriger gleichartiger Daten auf einem externen Speichermedium.
- Ablage und Zugriff auf Dateien erfolgen über ein Betriebssystem.
- Dateien werden durch das Betriebssystem über einen Namen adressiert.
- Die Codierung von Daten in einer Datei wird durch ein Dateiformat festgelegt.
- Das Dateiformat wird vom Betriebssystem angezeigt.
- Dies geschieht üblicherweise durch eine Namenserverweiterung (Extension).

Dateiformate

Textdateien

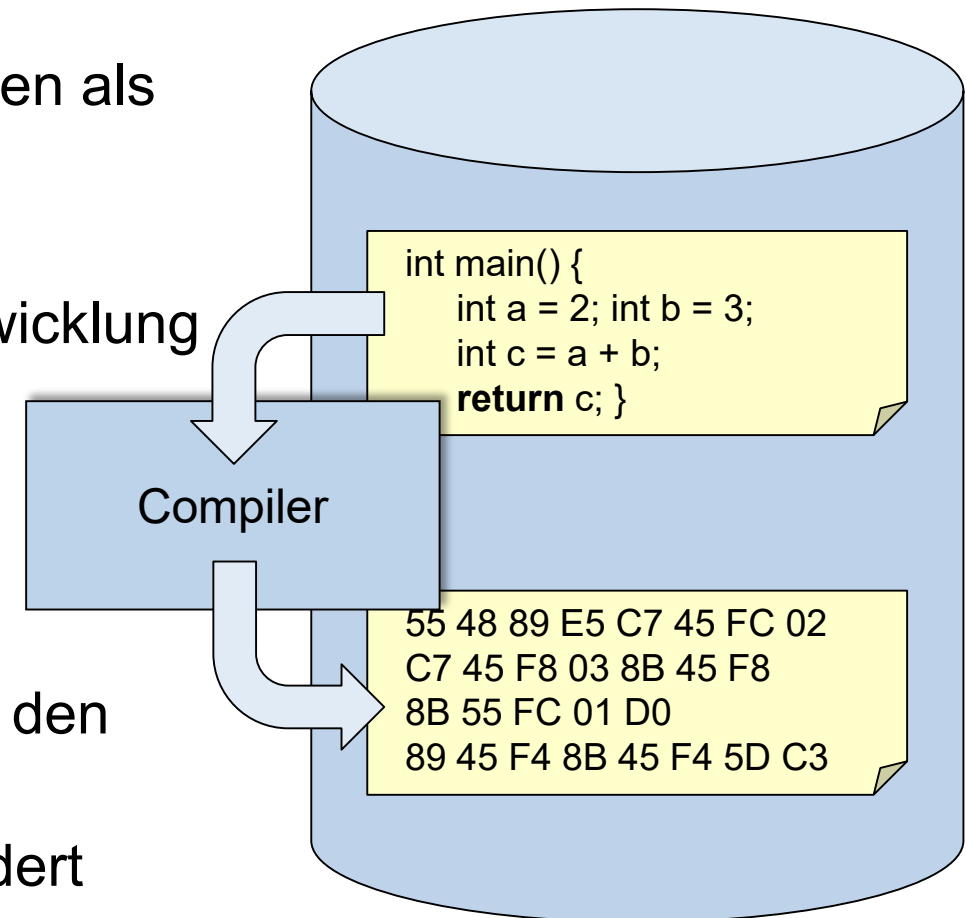
- Es existieren sehr unterschiedliche Dateiformate für verschiedene Zwecke.
- Ein sehr einfaches Dateiformat sind Textdateien.
- Diese sind etwa in Windows mit der Erweiterung **.txt* gekennzeichnet.
- Textdateien sind geordnete Sammlung von Zeichen.
- Die einzelnen Zeichen sind z.B. nach ISO 7-bit codiert.



Dateiformate

Programmdateien

- Programme in höheren Programmiersprachen liegen als Textdateien vor.
- Bezeichnung als Quellcode.
- Quellcode wird für die Abänderung und Weiterentwicklung benötigt.
- Ein Compiler erzeugt aus dem Quellcode einen Maschinencode.
- Der Maschinencode wird für die Ausführung durch den Prozessor benötigt.
- Maschinencode kann nicht ohne weiteres abgeändert werden.



- Computer verarbeiten Befehle und Daten in binärer Form.
- Binäre Zeichenketten werden in den Speicherzellen des Hauptspeichers abgelegt.
- Der Prozessor greift auf die Speicherzellen zu und interpretiert die Zeichenketten je nach Kontext.
- Der Prozessor verfügt über einen beschränkten Satz elementarer Befehle.
- Menschen programmieren Computer kaum mehr auf der Maschinenebene sondern mit höheren Konstrukten.
- Zwischen den verschiedenen Bezugsebenen müssen Übersetzungen und Umcodierungen stattfinden.
- Im Zusammenhang mit der Datenverarbeitung sind auch Ein- und Ausgabegeräte und externe Speicher relevant.