



# Informatik der digitalen Medien

Ergänzungs-Studienangebot der Mediendidaktik für  
Lehramtstudenten  
Dr. rer. nat. Harald Sack  
Institut für Informatik  
FSU Jena

Sommersemester 2004

## Informatik der digitalen Medien

- 1
- 21.04.2004 – Vorlesung Nr. 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14

### 2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## 2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Warum Kodierung?
- Multimediale Daten im Computer
  - Zeichen, Text, Bild, Audio und Video
- Zahlen
  - Binärcodierung von Dezimalzahlen
  - Negative Zahlen im 2er-Komplement
  - Binärcodierung von Gleitkommazahlen

## 2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Zeichencodes
  - Samuel Morse's Telegrafiecode
  - Fernschreiber und der Baudot-Code
  - ASCII-Code und nationale Erweiterungen
  - Unicode – ein Code für alle Sprachen und Schriften

# Informatik der digitalen Medien

## 2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- **Warum Kodierung?**
- Multimediale Daten im Computer
- Zahlen
- Zeichencodes

## Grundlagen der Digitalisierung – Datenrepräsentation im Computer

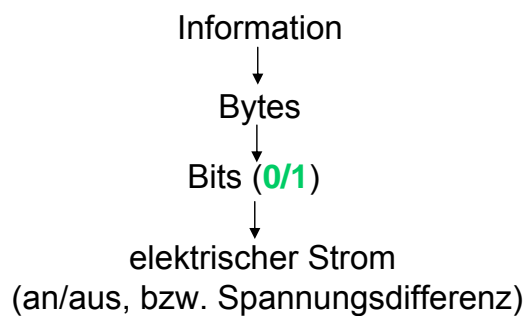
- **Warum Kodierung?**
  - „natürliche Information“ ist **analog**
  - z.B. **Musik**
    - an bestimmter Stelle liegen Töne mit bestimmter
      - Tonhöhe
      - Lautstärke
      - Charakteristik
      - etc...



# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Warum Kodierung?

- „natürliche Information“ ist analog
- aber
- Darstellung im Computer ist **digital**

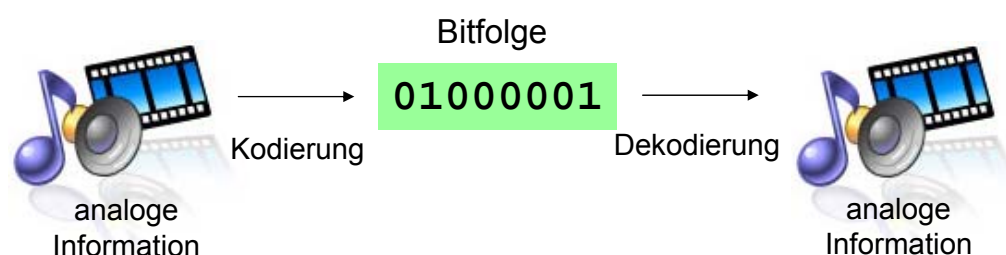


# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Warum Kodierung?

### ○ Was ist Kodierung?

- Vorgang der **Umwandlung** einer Information
  - aus einer bestimmten Darstellung (Ausprägung)
  - in eine **andere Darstellung**
- ohne (bzw. mit zu tolerierendem) Informationsverlust
- Rückwandlung wird als **Dekodierung** bezeichnet



# Informatik der digitalen Medien

## 2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Warum Kodierung?
- **Multimediale Daten im Computer**
- Zahlen
- Zeichencodes

## Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- **Multimediale Daten im Computer**
- **Multimedia?**
  - Information, die mit Hilfe **mehrerer, verschiedenartiger Medien** zur Darstellung kommt
- z.B.
  - Text und Bild
  - Text und Bild und Audio/Video
  - etc.





# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

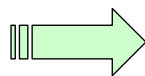
## ○ Multimediale Daten im Computer

### ○ Medientypen

#### ○ Audio

- Kodierung von **akustischer Information**
- Sprache / Musik / Geräusche
- Abhängig von menschlicher Wahrnehmung

- Wichtiger Unterschied zu Text/Grafik:



Wiedergabe von Audio-Daten **muss in Echtzeit erfolgen!**



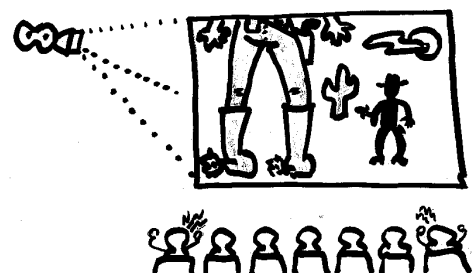
# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Medientypen

#### ○ Video / Animation

- Kodierung von Bild- und Tonfolgen
- Wiedergabe muss ebenfalls in **Echtzeit** erfolgen
- Abhängig von menschlicher Wahrnehmung



# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ **Multimediale Daten im Computer**

### ○ **Medientypen**

#### ○ **Zeitunabhängige Medien**

- Zeitkomponente bei Aufzeichnung/Wiedergabe ohne Bedeutung
- z.B. Text und Grafik
- „diskrete“ Medien

#### ○ **Zeitabhängige Medien**

- Information verändert sich mit der Zeit
- Gehalt der Einzelinformation nicht signifikant
- Gesamtinformation erschließt sich erst aus zeitlichem Ablauf
- Zeitkomponente kritisch bei Aufzeichnung/Wiedergabe
- z.B. Audio und Video

# Informatik der digitalen Medien

## 2. **Grundlagen der Digitalisierung – Datenrepräsentation im Computer**

- Warum Kodierung?
- Multimediale Daten im Computer
- **Zahlen**
- Zeichencodes



# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Zahlen

#### ○ Dezimalsystem

○ Basis **10**

○ Ziffern 0,1,2,3,4,5,6,7,8,9

$$\begin{aligned} 1428 &= 1 \cdot 1.000 + 4 \cdot 100 + 2 \cdot 10 + 8 \cdot 1 \\ &= 1 \cdot 10^3 + 4 \cdot 10^2 + 2 \cdot 10^1 + 8 \cdot 10^0 \end{aligned}$$

#### ○ Binär-/Dualsystem

○ Basis **2**

○ Ziffern 0,1

$$\begin{aligned} 1001 &= 1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 \\ &= 1 \cdot 8 + 0 \cdot 4 + 0 \cdot 2 + 1 \cdot 1 \\ &= 9 \text{ (dezimal)} \end{aligned}$$

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Zahlen

#### ○ Verallgemeinerung: **p**-adische Zahlen

○ Basis **p**,  $p \in \mathbb{N}$

○ Ziffern **0,1,...,p-1**

○ Darstellung einer Zahl  $k \in \mathbb{N}$  als p-adische Zahl

$$k = \sum_{i=0}^n c_i \cdot p^i, \quad c_i \in \{0,1,\dots,p-1\}$$

○ Bsp:  $p=10$ , Ziffern: 0,1,2,3,4,5,6,7,8,9

$$1234 = 1 \cdot 10^3 + 2 \cdot 10^2 + 3 \cdot 10^1 + 4 \cdot 10^0$$

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Zahlen

#### ○ Hexadezimalsystem

- in der Informatik weit verbreitet
- Basis: **p=16**
- Ziffern: **1,2,3,4,5,6,7,8,9,A,B,C,D,E,F**
- Bsp.:

$$\begin{aligned} 2F9 &= 2 \cdot 16^2 + 15 \cdot 16^1 + 9 \cdot 16^0 \\ &= 2 \cdot 256 + 15 \cdot 16 + 9 \cdot 1 \\ &= 512 + 240 + 9 \\ &= 761 \end{aligned}$$

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Zahlen

#### ○ Umrechnung **Dualsystem - Dezimalsystem**

- $101001 = 1 \cdot 2^5 + 1 \cdot 2^3 + 1 \cdot 2^0$   
 $= 1 \cdot 32 + 1 \cdot 8 + 1 \cdot 1$   
 $= 41$  (dezimal)

- $53 = 1 \cdot 32 + 1 \cdot 16 + 0 \cdot 8 + 1 \cdot 4 + 0 \cdot 2 + 1 \cdot 1$   
 $= 1 \cdot 2^5 + 1 \cdot 2^4 + \quad + 1 \cdot 2^2 + 1 \cdot 2^0$   
 $= \mathbf{110101}$

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Zahlen

#### ○ Umrechnung Dualsystem - Dezimalsystem

$$\begin{aligned} \text{○ } 53 &= 1 \cdot 32 + 1 \cdot 16 + 0 \cdot 8 + 1 \cdot 4 + 0 \cdot 2 + 1 \cdot 1 \\ &= 1 \cdot 2^5 + 1 \cdot 2^4 + \quad + 1 \cdot 2^2 + 1 \cdot 2^0 \\ &= \mathbf{110101} \end{aligned}$$

#### ○ Einfacher Algorithmus:

53:2 =	26	Rest	1	↑ Leserichtung
26:2 =	13	Rest	0	
13:2 =	6	Rest	1	
6:2 =	3	Rest	0	
3:2 =	1	Rest	1	
1:2 =	0	Rest	1	

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Zahlen

#### ○ Darstellung im Computer

○ Bytes (1 Byte = 8 Bit)

○ Bit (Wert: 0 oder 1)

○ Bsp.:

○ 53 als **8-Bit Dualzahl** = 00110101

Stelle	7	6	5	4	3	2	1	0
Bitwert	<b>0</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>1</b>
Dezimalwert	$2^7$	$2^6$	$2^5$	$2^4$	$2^3$	$2^2$	$2^1$	$2^0$

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Zahlen

#### ○ Einfache Arithmetik

##### ○ Addition

$$\begin{array}{r} 19 \quad 00010011 \\ +9 \quad 00001001 \\ \hline \quad \quad \quad 11 \\ 28 \quad 00011100 \end{array}$$

##### ○ Subtraktion

$$\begin{array}{r} 19 \quad 00010011 \\ -9 \quad 00001001 \\ \hline \quad \quad \quad 1 \\ 10 \quad 00001010 \end{array}$$

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Negative Dualzahlen

#### ○ 1-Komplement:

- für negative Zahlen werden alle Bits **invertiert** (umgedreht) (1 wird 0 und 0 wird 1)
- und ein zusätzliches Bit (**Vorzeichenbit**) =1 gesetzt

#### ○ Bsp.:

- 53 als **8-Bit Dualzahl** = 00110101
- -53 als **8-Bit Dualzahl im 1-Komplement** = 11001010

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## Multimediale Daten im Computer

### Negative Zahlen

Bsp.:

53 als **8-Bit Dualzahl** = 00110101

-53 als **8-Bit Dualzahl im 1-Komplement** = 11001010

V	6	5	4	3	2	1	0
<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>
	$2^6$	$2^5$	$2^4$	$2^3$	$2^2$	$2^1$	$2^0$

Stelle (pointing to the top row)

Bitwert (pointing to the middle row)

Dezimalwert (pointing to the bottom row)

Vorzeichenbit (pointing to the 'V' column)

Informatik der digitalen Medien  
Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## Multimediale Daten im Computer

### Negative Zahlen

Problem bei 1-Komplement:

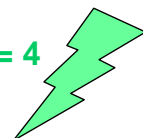
Es gibt 2 verschiedene Kodierungen für 0:

00000000 = 11111111

Addition und Subtraktion mit negativen Zahlen ergibt falsches Ergebnis

z.B.  $17 + (-12)$

$$\begin{array}{r}
 17 \quad 00010001 \\
 -12 \quad 11110011 \\
 \hline
 = \quad 00000100 = 4
 \end{array}$$



Informatik der digitalen Medien  
Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Negative Zahlen

#### ○ 2-Komplement:

1. für negative Zahlen werden alle Bits **invertiert** (umgedreht) (1 wird 0 und 0 wird 1)
2. zur invertierten Zahl wird **1 hinzuaddiert**
3. ein zusätzliches Bit (**Vorzeichenbit**) =1 wird gesetzt

#### ○ Bsp.:

- 53 als **8-Bit Dualzahl** = 00110101
- -53 als **8-Bit Dualzahl im 1-Komplement** = 11001010
- -53 als **8-Bit Dualzahl im 2-Komplement** = **11001011**

**Bemerkung:** anstelle (2) kann auch vor der Invertierung von der zu invertierenden Zahl 1 subtrahiert werden.

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Negative Zahlen

#### ○ 2-Komplement:

- es gibt nur noch eine Null: 00000000
- Addition und Subtraktion funktionieren korrekt, auch für negative Zahlen im 2-Komplement

z.B. 17 + (-12)

$$\begin{array}{r} 17 \quad 00010001 \\ -12 \quad 11110100 \\ \hline = \quad 00000101 \end{array} = 5$$

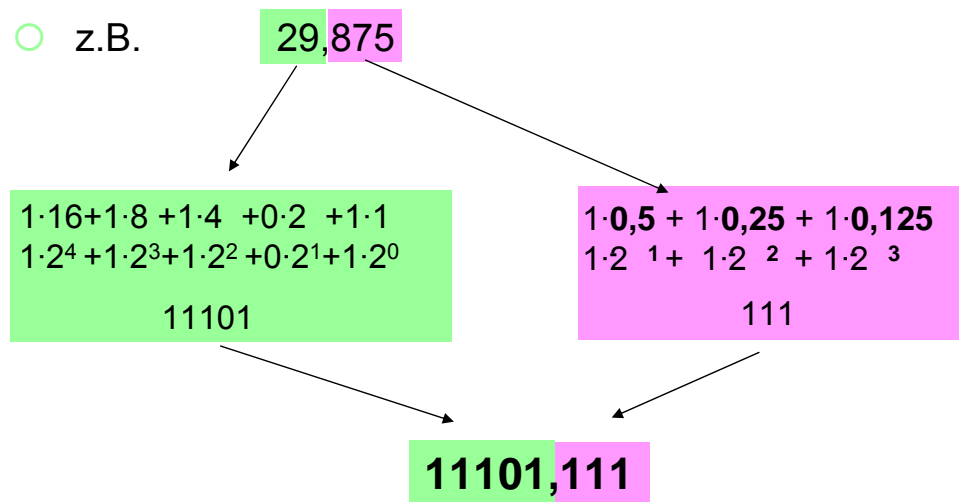


# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Gleitkommazahlen

○ z.B.



Informatik der digitalen Medien  
Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Multimediale Daten im Computer

### ○ Gleitkommazahlen

○ das Komma steht immer an einer anderen Stelle.....

○ kann man das Komma fixieren?

$$\begin{array}{rclcl} \text{○ } 29,875 & = & 2,9875 & \cdot 10^1 \\ 3,14159 & = & 3,14159 & \cdot 10^0 \\ 12342,7 & = & 1,23427 & \cdot 10^4 \end{array}$$

○ Fixiere Komma an erster Stelle (**Normalisierung**)

$$1,23427 \cdot 10^4$$

Informatik der digitalen Medien  
Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## Multimediale Daten im Computer

### Gleitkommazahlen

- Fixiere Komma an erster Stelle (**Normalisierung**)



- Trennung in **Mantisse** und **Exponent**

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## Multimediale Daten im Computer

### Binäre Gleitkommazahlen

○ 29,875 = 1101,111

Normalisierung

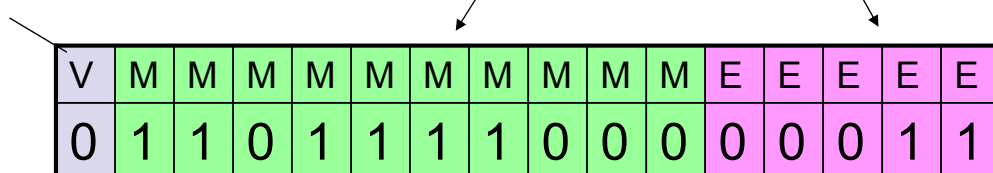
**1,101111**  
Mantisse

**2<sup>3</sup>**  
Exponent

1,101111

11

Vorzeichenbit



**16-Bit Gleitkommazahl:**

**Vorzeichenbit:**

1 Bit

**Mantisse:**

10 Bit

**Exponent:**

5 Bit



# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ **Multimediale Daten im Computer**

### ○ **Binäre Gleitkommazahlen**

- üblich ist heute 32/64-Bit Gleitkommadarstellung
  - IEEE 754 Floating Point Representation
    - 1 Bit Vorzeichen
    - 8/11 Bit Exponent
    - 23/52 Bit Mantisse
  - [Beispielanwendung](#)

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ **Multimediale Daten im Computer**

### ○ **Zahlendarstellung - Zusammenfassung**

- Zahlen werden im Computer stets als **Binärzahl** repräsentiert
- Unterscheide:
  - Einfache, positive Binärdarstellung
  - Negative Binärdarstellung
    - 1-Komplement
    - 2-Komplement
  - normalisierte Gleitkommadarstellung

# Informatik der digitalen Medien

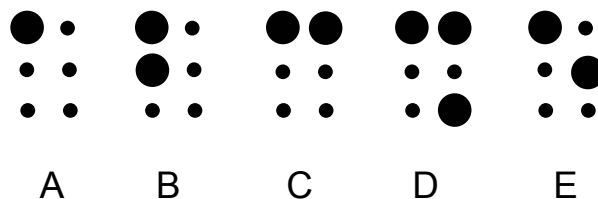
## 2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Warum Kodierung?
- Multimediale Daten im Computer
- Zahlen
- **Zeichencodes**

Informatik der digitalen Medien  
Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

## Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- **Textkodierung**
- Alphabet: a,b,c,d,e,...,A,B,C,D,E,...,0,1,2,...
- Z.B.: Brailleschrift



**Louis Braille (1809-1852)**  
- Zeichen werden in 3x2-BinärMatrix kodiert

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ Morsecode

•	E	— •	N	...	— — • •	Z
—	T	— —	M	...	— • — —	Y
		• •	I	...	— — • —	Q
		• —	A	...	• — — —	J



Samuel Morse  
1791 to 1872

### Samuel Morse (1791-1872)

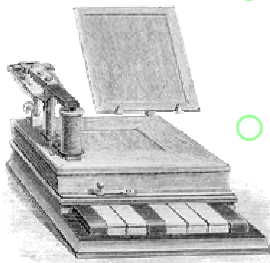
- **binärer** Zeichencode (1840) (Strom an/aus)
  - Buchstaben (1 – 4 Signale)
  - Zahlen (5 Signale)
  - Interpunktion (6 Signale)
- Zeichen/Wortgrenzen werden durch Pausen markiert
- Zeichenlänge abhängig von mittlerer Häufigkeit des Zeichens

Informatik der digitalen Medien  
Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ Baudot-Code



- Problem:
  - Morsecode wegen unterschiedlicher Zeichenlänge zu kompliziert für **Fernschreiber**
- daher konstante Zeichenlänge: **5 Bit** (Emile Baudot's Fernschreibmaschine, 1874)
  - **Problem:** reicht aber nur für 32 verschiedene Zeichen



- Idee: Doppelbelegungen
  - spezielles **Steuerzeichen** schaltet zwischen Ziffern und Buchstaben um

Informatik der digitalen Medien  
Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ ASCII-Code

- (American Standard Code for Information Interchange)
  - in 50er Jahre gab es keinen Kodierungsstandard für Computer
  - 1961 (Robert Bemer, IBM) entwickelt 7-Bit Kodierung
    - basiert auf 7-Bit FIELDATA Code
    - 99 Zeichen (Ziffern, Großbuchstaben, Steuersymbole)
    - ECMA belegt Rest mit Kleinbuchstaben
  - 1963 erstmals von ANSI standardisiert
  - 1974 ISOI-646 Standard
  - von IBM aber erst 1981 eingesetzt

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

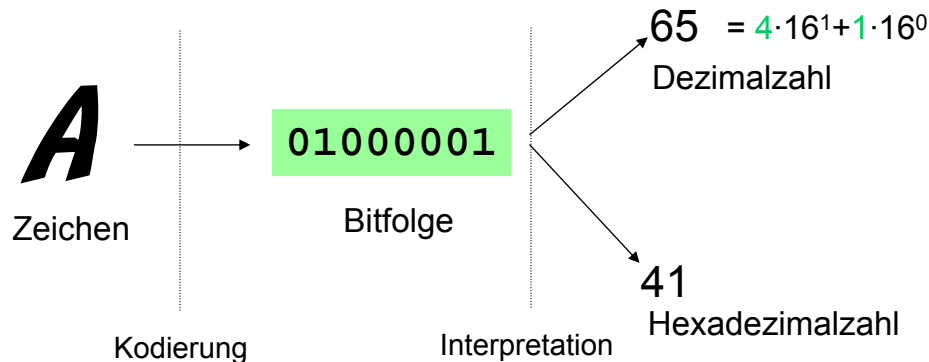
### ○ ASCII-Code

ASCII value	Character	Control character	ASCII value	Character	ASCII value	Character	ASCII value	Character
000	(null)	NUL	032	(space)	064	@	096	
001	☺	SOH	033	!	065	A	097	a
002	☹	STX	034	"	066	B	098	b
003	☛	ETX	035	#	067	C	099	c
004	☜	EOT	036	\$	068	D	100	d
005	☞	ENQ	037	%	069	E	101	e
006	☟	ACK	038	&	070	F	102	f
007	(beep)	BEL	039	'	071	G	103	g
008	☐	BS	040	(	072	H	104	h
009	(tab)	HT	041	)	073	I	105	i
010	(line feed)	LF	042	*	074	J	106	j
011	(home)	VT	043	+	075	K	107	k
012	(form feed)	FF	044	,	076	L	108	l
013	(carriage return)	CR	045	-	077	M	109	m
014	☪	SO	046	.	078	N	110	n
015	☫	SI	047	/	079	O	111	o
016	☛	DLE	048	0	080	P	112	p
017	☜	DC1	049	1	081	Q	113	q
018	☞	DC2	050	2	082	R	114	r
019	☟	DC3	051	3	083	S	115	s
020	☛	DC4	052	4	084	T	116	t
021	☜	NAK	053	5	085	U	117	u
022	☞	SYN	054	6	086	V	118	v
023	☟	ETB	055	7	087	W	119	w
024	☛	CAN	056	8	088	X	120	x
025	☜	EM	057	9	089	Y	121	y
026	☞	SUB	058	:	090	Z	122	z
027	←	ESC	059	;	091	[	123	{
028	(cursor right)	FS	060	<	092	\	124	}
029	(cursor left)	GS	061	=	093	]	125	~
030	(cursor up)	RS	062	>	094	^	126	
031	(cursor down)	US	063	?	095	_	127	␣

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ ASCII-Code



# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ ASCII-Code

#### ○ Problem:

- 7 Bit ausreichend für 128 Zeichen
- International existieren aber viele Umlaute und Sonderzeichen

#### ○ Lösung:

- **ISO 8859-x** Standard,
  - 8-Bit ASCII-Kodierung mit nationalen Erweiterungen (Umlaute)
  - 0-127 identisch mit Standard-ASCII
  - 128-159 seltene Steuerzeichen
  - **160-255 nationale Erweiterungen**

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ ASCII-Code

#### ○ nationalen Erweiterungen

- ISO-8859-1 Westeuropa, Amerika, Australien, Afrika (ISO-8859-15)
- ISO-8859-2 Osteuropa (ISO-8859-16)
- ISO-8859-3 Esperanto und Maltesisch
- ISO-8859-4 Baltisch, Grönland, Lappland
- ISO-8859-5 Bulgarien, Mazedonien, Russisch, Serbien, Ukraine
- ISO-8859-6 Arabisch (ohne Persisch/Urdu)
- ISO-8869-7 Griechenland
- ISO-8859-8 Hebräisch
- ISO-8859-9 Island, Türkei
- ISO-8859-10 Grönland, Lappland
- ISO-8859-11 Thai
- ISO-8859-12 Indien
- ISO-8859-13 Baltikum
- ISO-8859-14 Gälisch, Walisisch

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ ASCII-Code

#### ○ Problem:

- 8 Bit ausreichend für 256 Zeichen
- Chinesische, japanische, koreanische oder indische Schriftzeichen lassen sich damit nur schwer repräsentieren
- Bsp.: chinesische Schriftzeichen in Japan
  - Gakashu Kanji:  
1006 Zeichen (Grundschule)
  - Joyo Kanji  
1945 Zeichen (offizielle Dokumente, Zeitung)
  - Jinmei-yo Kanji  
285 Zeichen (Namen)

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ ASCII-Code

#### ○ Problem:

- 8 Bit ausreichend für 256 Zeichen
- Chinesische, japanische, koreanische oder indische Schriftzeichen lassen sich damit nur schwer repräsentieren
- Unterschiedliche Laufrichtungen
- Bsp.:       deutsch       **hallo**  
                  hebräisch    מזל טוב

#### ○ Multilinguale Dokumente

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ ASCII-Code

#### ○ Problem:

- 8 Bit ausreichend für 256 Zeichen
- Chinesische, japanische, koreanische oder indische Schriftzeichen lassen sich damit nur schwer repräsentieren
- Multilinguale Dokumente

#### ○ Lösung:

- mehr Bits pro Zeichen, um größere Zeichenvorräte kodieren zu können



# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ Unicode

- ab 1984, **ISO 10646** Standard seit 1992
- ursprünglich 16-Bit, dann **21** (32)-Bit Kodierung
- ermöglicht **multilinguale Textverarbeitung**
- genutzt werden 17 Ebenen (planes) mit je 65.536 Zeichen
- potenziell aber 2.147.483.648 Zeichen möglich
- **B**asic **M**ultilingual **P**lane (BMP, Unicode 3.2.0, 2002)
  - 49194 ausgewiesene Zeichen
  - unterstützt fast alle gebräuchlichen modernen Schreibsysteme

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ Unicode

- **B**asic **M**ultilingual **P**lane (BMP)
  - kann in 16-Bit UTF-16 kodiert werden
  - Schreibweise: **U+xxxx<sub>16</sub>**
  - BMP umfasst
    - 10236 Buchstaben
    - 27786 **CJK**-Unihan-Zeichen (vereinheitlichte **c**hinesische, **j**apanische und **k**oreanische Schrift)
    - 11172 koreanische Hangul-Zeichen
    - 8515 Kontrollsymbole
- Supplementary Multilingual Plane (SMP)
  - archaische und nicht mehr genutzte Schriftsysteme, Musik, Mathe



# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ Unicode

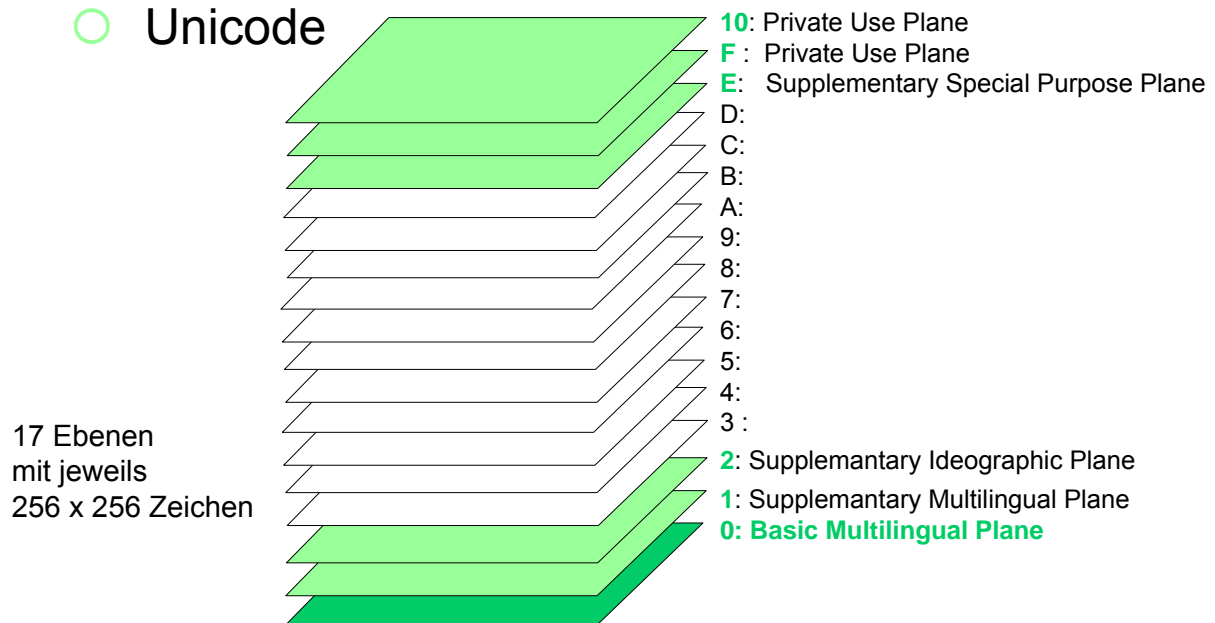
- Supplementary Ideographic Plane (SIP)
  - CJK-Erweiterungen
  - Seltene und unübliche chinesische Schriftzeichen
- Supplementary Special Purpose (SSP)
  - Kontrollsymbole
  - Protokollanweisungen
- Private Use Planes (PUP)

Informatik der digitalen Medien  
Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ Unicode



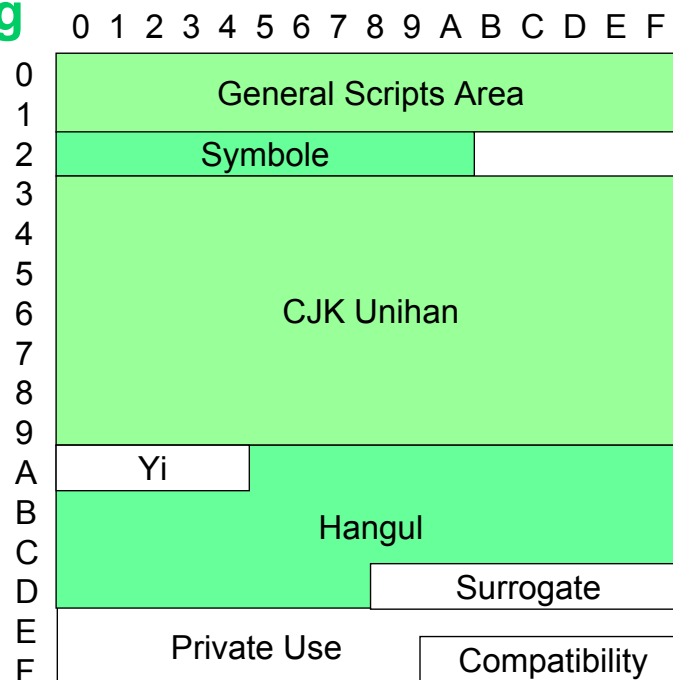
Informatik der digitalen Medien  
Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ Unicode

#### ● BMP



Informatik der digitalen Medien  
Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ Unicode

#### ● BMP – General Script Area

- 0-127 entspricht ASCII-Kodierung (Kompatibilität) (U+0000 – U+007F)
- 0-255 entspricht ISO-8859-1 (U+0000 – U+00FF)
- umfasst Lateinschrift, Griechisch, Kyrillisch, Hebräisch, Arabisch, Devanagari, Bengali, Äthiopisch, Cherokee, Kmehr, Runen, Thai, Laotisch, Malayalam, Mongolisch, Tibetisch, Telegu, Georgisch, Ogham, Gumurkhi, Oriya, Tamil, Sinhala, Thaana, etc.

vgl. [www.unicode.org](http://www.unicode.org)

Informatik der digitalen Medien  
Dr.rer.nat. Harald Sack, Institut für Informatik, FSU Jena, Ernst-Abbe-Platz 2-4, D-0744 Jena, E-Mail: sack@minet.uni-jena.de

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ Textkodierung

### ○ Unicode – **U**niversal **T**ransformation **F**ormats

- UTF-8
  - 8-Bit Unicodevariante, identisch mit ASCII
- UTF-16
  - 16-Bit Kodierung der BMP-Unicode-Zeichen
- UTF-32
  - 32-Bit Unicode Kodierung nach ISO 10646
  - UCS-4 (Universal Multiple Octet Coded Character Set)

# Informatik der digitalen Medien

## 2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

- Warum Kodierung?
- Multimediale Daten im Computer
- Zahlen
- Zeichencodes

# Grundlagen der Digitalisierung – Datenrepräsentation im Computer

## ○ **Multimediale Daten im Computer**

### ○ **Zeichenkodierung - Zusammenfassung**

- Die Anzahl der jeweils darstellbaren Zeichen hängt von der Länge der verwendeten Zeichenkodierung ab
  - z.B. Baudot: 5 Bit, ASCII: 7 Bit
- Entwicklung der Zeichenkodierung
  - Morse Code
  - Baudot Code
  - ASCII
  - Unicode

## Informatik der digitalen Medien

### 2. Grundlagen der Digitalisierung – Datenrepräsentation im Computer

#### ○ **Literatur**

- Ch. Meinel, H. Sack:  
**WWW– Kommunikation, Internetworking, Web-Technologien**,  
Springer, 2004.
- P.A. Henning:  
**Taschenbuch Multimedia**,  
3. Aufl., Fachbuchverlag Leipzig, 2003.
- R. Gilliam:  
**Unicode Demystified –  
A Practical Programmer’s Guide to the Encoding Standard**  
Addison Wesley, 2003.
- G. Brookshear:  
**„1.5 The Binary System“,  
in Computer Science – an Overview**,  
Addison Wesley, NY, 2003, pp. 43–56.
- The Unicode Consortium:  
<http://www.unicode.org>