

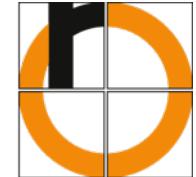


# Prof. Dr. Florian Künzner

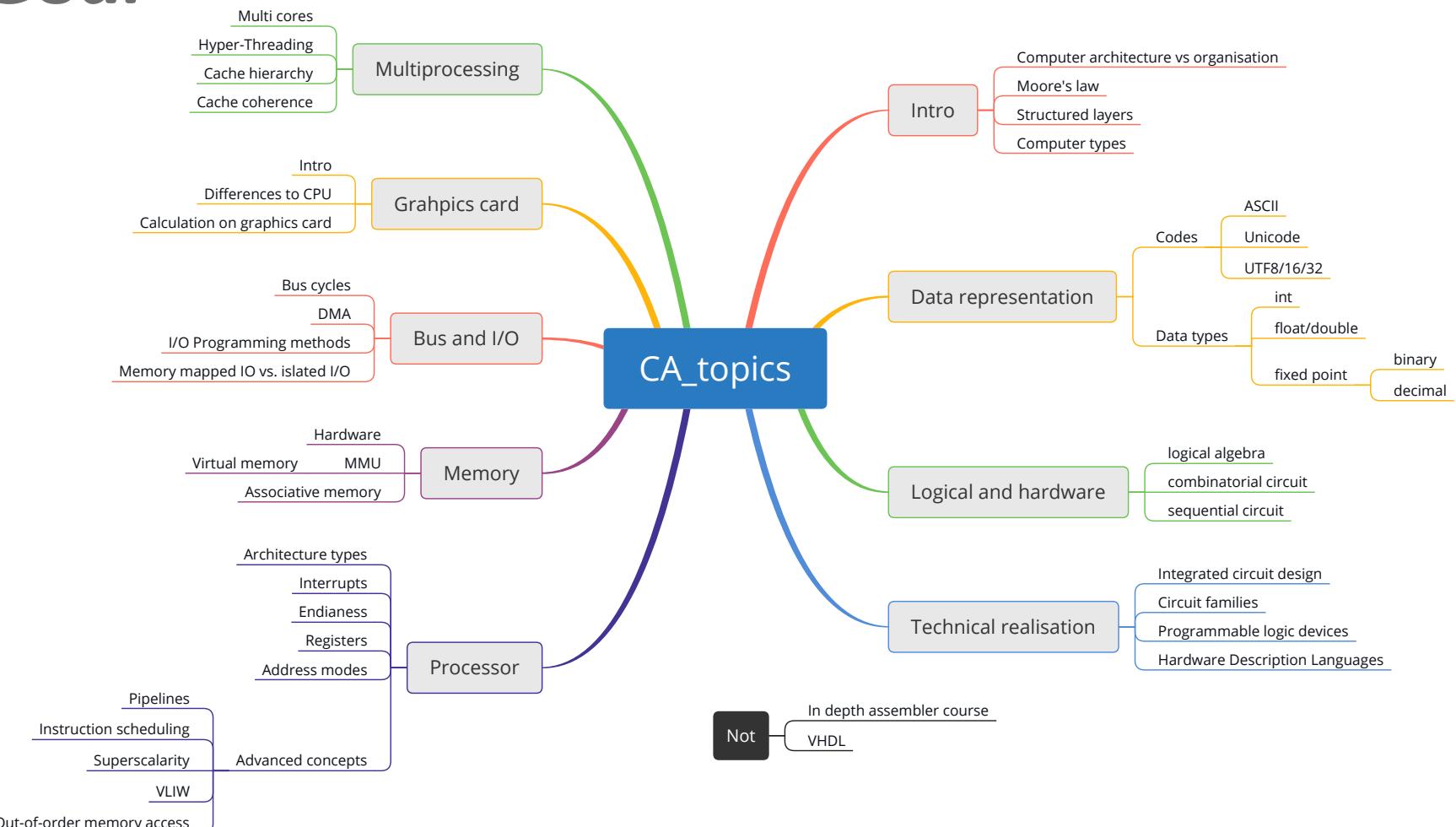
Technical University of Applied Sciences Rosenheim, Computer Science

## CA 2 – Data representation

The lecture is based on the work and the documents of Prof. Dr. Theodor Tempelmeier



# Goal





# Goal

## CA::Data representation

- Important basics
- ASCII
- Unicode and UTF
- Data types: Numbers



# Important basics - numeral systems

How much do you still **know\***  
about **numeral systems?**

low

current knowledge

high →

\*Use a **stamp** for your estimate.

# Important basics

Which **numeral systems** do you know?



# Important basics

## Numeral systems

- DEC: 0, 1, ..., 9; e.g.: 291
- BIN: 0, 1; e.g.: 100100011
- HEX: 0, 1, ..., 9, A, B, ..., F; e.g.: 0x123

## Conversion between:

- HEX  $\leftrightarrow$  DEC
- BIN  $\leftrightarrow$  HEX
- DEC  $\leftrightarrow$  BIN

# Important basics

## Numeral systems

- DEC: 0, 1, ..., 9; e.g.: 291
- BIN: 0, 1; e.g.: 100100011
- HEX: 0, 1, ..., 9, A, B, ..., F; e.g.: 0x123

## Conversion between:

- HEX  $\leftrightarrow$  DEC
- BIN  $\leftrightarrow$  HEX
- DEC  $\leftrightarrow$  BIN

# Important basics

## Numeral systems

- DEC: 0, 1, ..., 9; e.g.: 291
- BIN: 0, 1; e.g.: 100100011
- HEX: 0, 1, ..., 9, A, B, ..., F; e.g.: 0x123

## Conversion between:

- HEX  $\leftrightarrow$  DEC
- BIN  $\leftrightarrow$  HEX
- DEC  $\leftrightarrow$  BIN

# Important basics

## Numeral systems

- DEC: 0, 1, ..., 9; e.g.: 291
- BIN: 0, 1; e.g.: 100100011
- HEX: 0, 1, ..., 9, A, B, ..., F; e.g.: 0x123

## Conversion between:

- HEX  $\leftrightarrow$  DEC
- BIN  $\leftrightarrow$  HEX
- DEC  $\leftrightarrow$  BIN

# Important basics

## Numeral systems

- DEC: 0, 1, ..., 9; e.g.: 291
- BIN: 0, 1; e.g.: 100100011
- HEX: 0, 1, ..., 9, A, B, ..., F; e.g.: 0x123

## Conversion between:

- HEX  $\leftrightarrow$  DEC
- BIN  $\leftrightarrow$  HEX
- DEC  $\leftrightarrow$  BIN

# Important basics

## Numeral systems

- DEC: 0, 1, ..., 9; e.g.: 291
- BIN: 0, 1; e.g.: 100100011
- HEX: 0, 1, ..., 9, A, B, ..., F; e.g.: 0x123

## Conversion between:

- HEX  $\leftrightarrow$  DEC
- BIN  $\leftrightarrow$  HEX
- DEC  $\leftrightarrow$  BIN

# Important basics

## Numeral systems

- DEC: 0, 1, ..., 9; e.g.: 291
- BIN: 0, 1; e.g.: 100100011
- HEX: 0, 1, ..., 9, A, B, ..., F; e.g.: 0x123

## Conversion between:

- HEX  $\leftrightarrow$  DEC
- BIN  $\leftrightarrow$  HEX
- DEC  $\leftrightarrow$  BIN

# Important basics

## Numeral systems

- DEC: 0, 1, ..., 9; e.g.: 291
- BIN: 0, 1; e.g.: 100100011
- HEX: 0, 1, ..., 9, A, B, ..., F; e.g.: 0x123

## Conversion between:

- HEX  $\leftrightarrow$  DEC
- BIN  $\leftrightarrow$  HEX
- DEC  $\leftrightarrow$  BIN

# Important basics - hints

# Important basics - short exercise 1/2

**Convert** HEX:0xCOFE to BIN.

# Important basics - short exercise 2/2

**Convert** BIN:1100000011011110 to HEX.



# Questions?

All right?  $\Rightarrow$



Question?  $\Rightarrow$



and use **chat**

or

*speak after I  
ask you to*

# Binary system

Why is the binary (dual) system used in computer science?

## Binary system for digits and characters

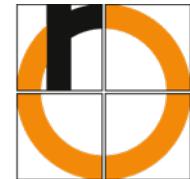
- Technically easy to realise (0/1)
- Well understood theoretical basis
  - Boolean algebra
  - Formal logic

# Binary system

Why is the binary (dual) system used in computer science?

## Binary system for digits and characters

- Technically easy to realise (0/1)
- Well understood theoretical basis
  - Boolean algebra
  - Formal logic



# Subtraction is reduced to addition

**Idea: Complementation and addition of the complement**

**Example:  $11 - 6$  in binary system**

1 11:  $\rightarrow$  01011

2 6:  $\rightarrow$  00110

3 complement of 6: 11001

+ 1

-----

11010

7 addition of  $11 + (-6)$ :

8 11: 01011

9 -6: 11010

-----

11 X00101  $\Rightarrow$  5

# Subtraction is reduced to addition

**Idea: Complementation and addition of the complement**

**Example:  $11 - 6$  in binary system**

1 11:  $\rightarrow$  01011

2 6:  $\rightarrow$  00110

3 complement of 6: 11001

+ 1

-----

11010

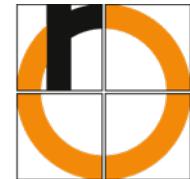
7 addition of  $11 + (-6)$ :

8 11: 01011

9 -6: 11010

-----

11 X00101  $\Rightarrow$  5



# Subtraction is reduced to addition

**Idea: Complementation and addition of the complement**

**Example:  $11 - 6$  in binary system**

1 11:  $\rightarrow$  01011

2 6:  $\rightarrow$  00110

3 complement of 6: 11001  
+ 1  
-----  
11010

4 addition of 11 + (-6):  
11: 01011  
-6: 11010  
-----  
X00101  $\Rightarrow$  5

# Subtraction is reduced to addition

**Idea: Complementation and addition of the complement**

**Example:  $11 - 6$  in binary system**

1 11:  $\rightarrow$  01011

2 6:  $\rightarrow$  00110

3 complement of 6: 11001

+ 1

-----

11010

7 addition of  $11 + (-6)$ :

8 11: 01011

9 -6: 11010

-----

11 X00101  $\Rightarrow$  5

# Subtraction is reduced to addition

**Idea: Complementation and addition of the complement**

**Example:  $11 - 6$  in binary system**

1 11:  $\rightarrow$  01011

2 6:  $\rightarrow$  00110

3 complement of 6: 11001

+ 1

-----

11010

7 addition of 11 + (-6):

8 11: 01011

9 -6: 11010

-----

11 X00101  $\Rightarrow$  5

# Codes

Which codes for characters do you know?



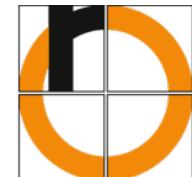
# ASCII (American Standard Code for Information Interchange)

Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0 000	000	<b>NUL</b> (null)	32	20 040	&#32;	<b>Space</b>		64	40 100	&#64;	<b>Ø</b>		96	60 140	&#96;	<b>'</b>	
1	1 001	001	<b>SOH</b> (start of heading)	33	21 041	&#33;	<b>!</b>		65	41 101	&#65;	<b>A</b>		97	61 141	&#97;	<b>a</b>	
2	2 002	002	<b>STX</b> (start of text)	34	22 042	&#34;	<b>"</b>		66	42 102	&#66;	<b>B</b>		98	62 142	&#98;	<b>b</b>	
3	3 003	003	<b>ETX</b> (end of text)	35	23 043	&#35;	<b>#</b>		67	43 103	&#67;	<b>C</b>		99	63 143	&#99;	<b>c</b>	
4	4 004	004	<b>EOT</b> (end of transmission)	36	24 044	&#36;	<b>\$</b>		68	44 104	&#68;	<b>D</b>		100	64 144	&#100;	<b>d</b>	
5	5 005	005	<b>ENQ</b> (enquiry)	37	25 045	&#37;	<b>%</b>		69	45 105	&#69;	<b>E</b>		101	65 145	&#101;	<b>e</b>	
6	6 006	006	<b>ACK</b> (acknowledge)	38	26 046	&#38;	<b>&amp;</b>		70	46 106	&#70;	<b>F</b>		102	66 146	&#102;	<b>f</b>	
7	7 007	007	<b>BEL</b> (bell)	39	27 047	&#39;	<b>'</b>		71	47 107	&#71;	<b>G</b>		103	67 147	&#103;	<b>g</b>	
8	8 010	010	<b>BS</b> (backspace)	40	28 050	&#40;	<b>(</b>		72	48 110	&#72;	<b>H</b>		104	68 150	&#104;	<b>h</b>	
9	9 011	011	<b>TAB</b> (horizontal tab)	41	29 051	&#41;	<b>)</b>		73	49 111	&#73;	<b>I</b>		105	69 151	&#105;	<b>i</b>	
10	A 012	012	<b>LF</b> (NL line feed, new line)	42	2A 052	&#42;	<b>*</b>		74	4A 112	&#74;	<b>J</b>		106	6A 152	&#106;	<b>j</b>	
11	B 013	013	<b>VT</b> (vertical tab)	43	2B 053	&#43;	<b>+</b>		75	4B 113	&#75;	<b>K</b>		107	6B 153	&#107;	<b>k</b>	
12	C 014	014	<b>FF</b> (NP form feed, new page)	44	2C 054	&#44;	<b>,</b>		76	4C 114	&#76;	<b>L</b>		108	6C 154	&#108;	<b>l</b>	
13	D 015	015	<b>CR</b> (carriage return)	45	2D 055	&#45;	<b>-</b>		77	4D 115	&#77;	<b>M</b>		109	6D 155	&#109;	<b>m</b>	
14	E 016	016	<b>SO</b> (shift out)	46	2E 056	&#46;	<b>.</b>		78	4E 116	&#78;	<b>N</b>		110	6E 156	&#110;	<b>n</b>	
15	F 017	017	<b>SI</b> (shift in)	47	2F 057	&#47;	<b>/</b>		79	4F 117	&#79;	<b>O</b>		111	6F 157	&#111;	<b>o</b>	
16	10 020	020	<b>DLE</b> (data link escape)	48	30 060	&#48;	<b>0</b>		80	50 120	&#80;	<b>P</b>		112	70 160	&#112;	<b>p</b>	
17	11 021	021	<b>DC1</b> (device control 1)	49	31 061	&#49;	<b>1</b>		81	51 121	&#81;	<b>Q</b>		113	71 161	&#113;	<b>q</b>	
18	12 022	022	<b>DC2</b> (device control 2)	50	32 062	&#50;	<b>2</b>		82	52 122	&#82;	<b>R</b>		114	72 162	&#114;	<b>r</b>	
19	13 023	023	<b>DC3</b> (device control 3)	51	33 063	&#51;	<b>3</b>		83	53 123	&#83;	<b>S</b>		115	73 163	&#115;	<b>s</b>	
20	14 024	024	<b>DC4</b> (device control 4)	52	34 064	&#52;	<b>4</b>		84	54 124	&#84;	<b>T</b>		116	74 164	&#116;	<b>t</b>	
21	15 025	025	<b>NAK</b> (negative acknowledge)	53	35 065	&#53;	<b>5</b>		85	55 125	&#85;	<b>U</b>		117	75 165	&#117;	<b>u</b>	
22	16 026	026	<b>SYN</b> (synchronous idle)	54	36 066	&#54;	<b>6</b>		86	56 126	&#86;	<b>V</b>		118	76 166	&#118;	<b>v</b>	
23	17 027	027	<b>ETB</b> (end of trans. block)	55	37 067	&#55;	<b>7</b>		87	57 127	&#87;	<b>W</b>		119	77 167	&#119;	<b>w</b>	
24	18 030	030	<b>CAN</b> (cancel)	56	38 070	&#56;	<b>8</b>		88	58 130	&#88;	<b>X</b>		120	78 170	&#120;	<b>x</b>	
25	19 031	031	<b>EM</b> (end of medium)	57	39 071	&#57;	<b>9</b>		89	59 131	&#89;	<b>Y</b>		121	79 171	&#121;	<b>y</b>	
26	1A 032	032	<b>SUB</b> (substitute)	58	3A 072	&#58;	<b>:</b>		90	5A 132	&#90;	<b>Z</b>		122	7A 172	&#122;	<b>z</b>	
27	1B 033	033	<b>ESC</b> (escape)	59	3B 073	&#59;	<b>:</b>		91	5B 133	&#91;	<b>[</b>		123	7B 173	&#123;	<b>{</b>	
28	1C 034	034	<b>FS</b> (file separator)	60	3C 074	&#60;	<b>&lt;</b>		92	5C 134	&#92;	<b>\</b>		124	7C 174	&#124;	<b> </b>	
29	1D 035	035	<b>GS</b> (group separator)	61	3D 075	&#61;	<b>=</b>		93	5D 135	&#93;	<b>]</b>		125	7D 175	&#125;	<b>}</b>	
30	1E 036	036	<b>RS</b> (record separator)	62	3E 076	&#62;	<b>&gt;</b>		94	5E 136	&#94;	<b>^</b>		126	7E 176	&#126;	<b>~</b>	
31	1F 037	037	<b>US</b> (unit separator)	63	3F 077	&#63;	<b>?</b>		95	5F 137	&#95;	<b>_</b>		127	7F 177	&#127;	<b>DEL</b> ☠	

# Extended ASCII codes

Source : [www.LookupTables.com](http://www.LookupTables.com)

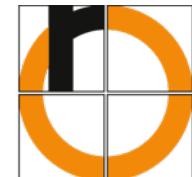
[source: asciitable.com]



# ASCII

**ASCII - American Standard Code for Information Interchange**

Any problems with ASCII?



# Unicode

- International standard (ISO 10646)
- For every character one code
- In the long term: A digital code is defined for each meaningful character or text element of all known cultures, countries/languages, and character systems.
- Is constantly extended
- <http://www.unicode.org>

# Unicode

- International standard (ISO 10646)
- For every character one code
- In the long term: A digital code is defined for each meaningful character or text element of all known cultures, countries/languages, and character systems.
- Is constantly extended
- <http://www.unicode.org>

# Unicode

- International standard (ISO 10646)
- For every character one code
- In the long term: A digital code is defined for each meaningful character or text element of all known cultures, countries/languages, and character systems.
- Is constantly extended
- <http://www.unicode.org>

# Unicode

- International standard (ISO 10646)
- For every character one code
- In the long term: A digital code is defined for each meaningful character or text element of all known cultures, countries/languages, and character systems.
- Is constantly extended
- <http://www.unicode.org>

# Unicode

- International standard (ISO 10646)
- For every character one code
- In the long term: A digital code is defined for each meaningful character or text element of all known cultures, countries/languages, and character systems.
- Is constantly extended
- <http://www.unicode.org>

# Unicode

- International standard (ISO 10646)
- For every character one code
- In the long term: A digital code is defined for each meaningful character or text element of all known cultures, countries/languages, and character systems.
- Is constantly extended
- <http://www.unicode.org>



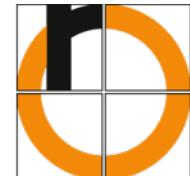
# Unicode

## Character range:

first code U+00 0000  
last code U+10 FFFF

## Character sets

Name	Unit	Calculation	#chars	first	last
------	------	-------------	--------	-------	------



# Unicode

## Character range:

first code U+00 0000  
last code U+10 FFFF

## Character sets

Name	Unit	Calculation	#chars	first	last
------	------	-------------	--------	-------	------

# Unicode

## Character range:

first code U+00 0000

last code U+10 FFFF

## Character sets

Name	Unit	Calculation	#chars	first	last
------	------	-------------	--------	-------	------

# Unicode

## Character range:

first code U+00 0000

last code U+10 FFFF

## Character sets

Name	Unit	Calculation	#chars	first	last
UCS-2	16 Bit	$2^{16}$	65536	U+0000	U+FFFF

# Unicode

## Character range:

first code U+00 0000

last code U+10 FFFF

## Character sets

Name	Unit	Calculation	#chars	first	last
UCS-2	16 Bit	$2^{16}$	65536	U+0000	U+FFFF
UCS-4	17 Planes	$17 * 2^{16}$	1114112	U+00 0000	U+10 FFFF

# Unicode

## Character range:

first code U+00 0000

last code U+10 FFFF

## Character sets

Name	Unit	Calculation	#chars	first	last
UCS-2	16 Bit	$2^{16}$	65536	U+0000	U+FFFF
UCS-4	17 Planes	$17 * 2^{16}$	1114112	U+00 0000	U+10 FFFF

## Examples:

Unicode Full number Character

# Unicode

## Character range:

first code U+00 0000

last code U+10 FFFF

## Character sets

Name	Unit	Calculation	#chars	first	last
UCS-2	16 Bit	$2^{16}$	65536	U+0000	U+FFFF
UCS-4	17 Planes	$17 * 2^{16}$	1114112	U+00 0000	U+10 FFFF

## Examples:

### Unicode Full number Character

U+0041 00 0041 A

# Unicode

## Character range:

first code U+00 0000

last code U+10 FFFF

## Character sets

Name	Unit	Calculation	#chars	first	last
UCS-2	16 Bit	$2^{16}$	65536	U+0000	U+FFFF
UCS-4	17 Planes	$17 * 2^{16}$	1114112	U+00 0000	U+10 FFFF

## Examples:

### Unicode Full number Character

U+0041 00 0041

A

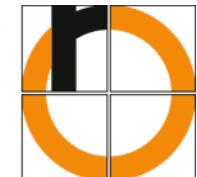
U+1F600 01 F600





# Unicode 10.0 - Planes

Plane 0 00 0000-00 FFFF <b>BMP</b> Basic Multilingual Plane	Plane 1 01 0000-01 FFFF <b>SMP</b> Supplementary Multilingual Plane	Plane 2 02 0000-02 FFFF <b>SIP</b> Supplementary Ideographic Plane	Plane 3 03 0000-03 FFFF unassigned	Plane 4 04 0000-04 FFFF unassigned
Plane 5 05 0000-05 FFFF unassigned	Plane 6 06 0000-06 FFFF unassigned	Plane 7 07 0000-07 FFFF unassigned	Plane 8 08 0000-08 FFFF unassigned	Plane 9 09 0000-09 FFFF unassigned
Plane 10 0A 0000-0A FFFF unassigned	Plane 11 0B 0000-0B FFFF unassigned	Plane 12 0C 0000-0C FFFF unassigned	Plane 13 0D 0000-0D FFFF unassigned	Plane 14 0E 0000-0E FFFF <b>SSP</b> Supplementary Special-purpose Plane
Plane 15 0F 0000-0F FFFF <b>SPUA-A</b> Supplementary Private Use Area planes	Plane 16 10 0000-10 FFFF <b>SPUA-A</b> Supplementary Private Use Area planes			



# Unicode

## Enter unicode characters

OS      Program

Keyboard shortcut

More shortcuts: [wikipedia.org](https://wikipedia.org)

\*must be enabled as input source

# Unicode

## Enter unicode characters

OS	<b>Program</b>
Linux	Terminal, xed, LibreOffice

<b>Keyboard shortcut</b>
CTRL+SHIFT+U + HEX Number

More shortcuts: [wikipedia.org](https://wikipedia.org)

\*must be enabled as input source

# Unicode

## Enter unicode characters

**OS      Program**

Linux    Terminal, xed, LibreOffice

Windows Microsoft Word, Excel, WordPad

**Keyboard shortcut**

CTRL+SHIFT+U + HEX Number

HEX Number + ALT+C

More shortcuts: [wikipedia.org](https://wikipedia.org)

\*must be enabled as input source

# Unicode

## Enter unicode characters

### OS      Program

Linux      Terminal, xed, LibreOffice

Windows      Microsoft Word, Excel, WordPad

macOS\*      Console, Text

### Keyboard shortcut

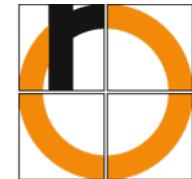
CTRL+SHIFT+U + HEX Number

HEX Number + ALT+C

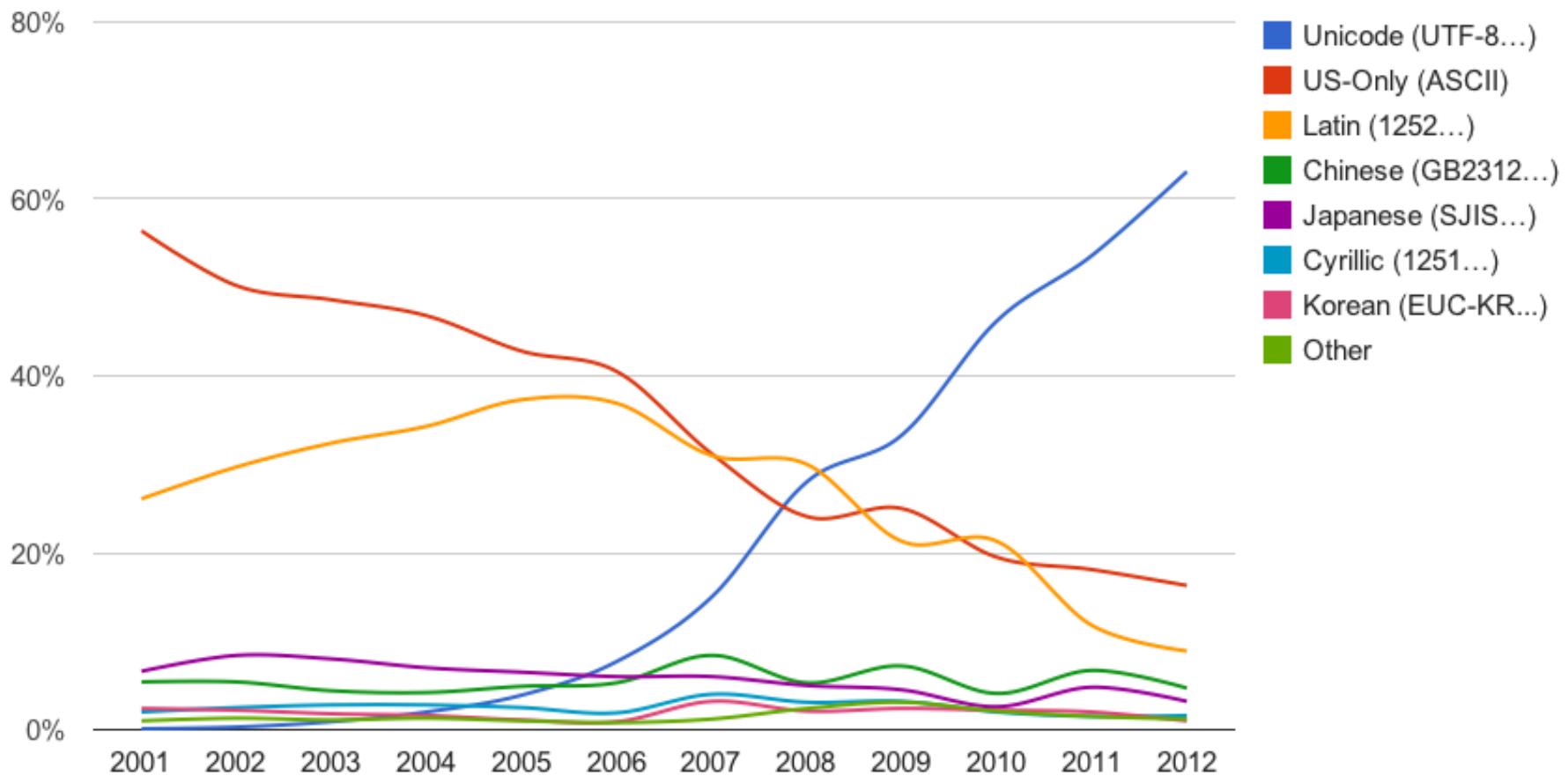
ALT + HEX Number

More shortcuts: [wikipedia.org](https://wikipedia.org)

\*must be enabled as input source



# Unicode usage



[source: [googleblog.blogspot.com](http://googleblog.blogspot.com)], Link to current statistics: [w3techs.com](http://w3techs.com)



# Questions?

All right?  $\Rightarrow$



Question?  $\Rightarrow$



and use **chat**

or

*speak after I  
ask you to*

# Unicode

Character set vs. character encoding?

Unicode vs UTF

# Unicode

Character set vs. character encoding?

**Unicode vs UTF**

# UTF - Unicode Transformation Format

UTF maps all unicode code points to a unique sequence of bytes.

## Used for

- Store information into files, databases, ...
- Transfer data (websites, e-mail, ...)

## Choice depends on

- Storage space
- Source code compatibility
- Interoperability with other systems
- Runtime for encoding/decoding

# UTF - Unicode Transformation Format

UTF maps all unicode code points to a unique sequence of bytes.

## Used for

- Store information into files, databases, ...
- Transfer data (websites, e-mail, ...)

## Choice depends on

- Storage space
- Source code compatibility
- Interoperability with other systems
- Runtime for encoding/decoding

# UTF - Unicode Transformation Format

UTF maps all unicode code points to a unique sequence of bytes.

## Used for

- Store information into files, databases, ...
- Transfer data (websites, e-mail, ...)

## Choice depends on

- Storage space
- Source code compatibility
- Interoperability with other systems
- Runtime for encoding/decoding

# UTF - Unicode Transformation Format

## Overview of UTF encodings

Encoding	Bits	Length	Common use
UTF-8	8-bit	Variable length: 1 to 4 bytes	Internet, Linux
UTF-16	16-bit	Variable length: 2 or 4 bytes	Qt, Java, Tcl
UTF-32	32-bit	Fixed length: 4 bytes	

# UTF-8

## UTF-8 length

Number of bytes	Bits for code point
-----------------	---------------------

Number of bytes	Bits for code point	Unicode range	Comment
1	7	0 - 00 007F	Compatible with ASCII
2	11	80 - 00 07FF	
3	16	800 - 00 FFFF	
4	21	1 0000 - 10 FFFF	

## UTF-8 encoding details

Unicode range	Byte 1	Byte 2	Byte 3	Byte 4
0 - 00 007F	0xxxxxxx			
80 - 00 07FF	110xxxxx	10xxxxxx		
800 - 00 FFFF	1110xxxx	10xxxxxx	10xxxxxx	
1 0000 - 10 FFFF	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx

# UTF-8

## UTF-8 length

Number Bits for

of bytes code point

Unicode range

Comment

1	7	0 - 00 007F	Compatible with ASCII
2	11	80 - 00 07FF	
3	16	800 - 00 FFFF	
4	21	1 0000 - 10 FFFF	

## UTF-8 encoding details

Unicode range      Byte 1      Byte 2      Byte 3      Byte 4

0 - 00 007F      0xxxxxxx

80 - 00 07FF      110xxxxx      10xxxxxx

800 - 00 FFFF      1110xxxx      10xxxxxx      10xxxxxx

1 0000 - 10 FFFF      11110xxx      10xxxxxx      10xxxxxx      10xxxxxx



# UTF-8 - example

## Encode the „ü“ into UTF-8!

[ü: [https://en.wikipedia.org/wiki/Latin-1\\_Supplement\\_\(Unicode\\_block\)](https://en.wikipedia.org/wiki/Latin-1_Supplement_(Unicode_block))]

1 ü → 252 → 0xFC

2

3 ü in Unicode:

4 U+00 00FC (8 bits → 2 bytes required)

5 F C

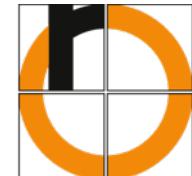
6 1111 1100

7

8 ü in UTF-8:

9 11000011 10111100

10 C 3 B C → 0xC3BC



# UTF-8 - example

## Encode the „ü“ into UTF-8!

[ü: [https://en.wikipedia.org/wiki/Latin-1\\_Supplement\\_\(Unicode\\_block\)](https://en.wikipedia.org/wiki/Latin-1_Supplement_(Unicode_block))]

1 ü → 252 → 0xFC

2  
3 ü in Unicode:  
4 U+00 00FC (8 bits → 2 bytes required)

5 F C  
6 1111 1100

7  
8 ü in UTF-8:  
9 11000011 10111100  
10 C 3 B C → 0xC3BC

# UTF-8 - example

## Encode the „ü“ into UTF-8!

[ü: [https://en.wikipedia.org/wiki/Latin-1\\_Supplement\\_\(Unicode\\_block\)](https://en.wikipedia.org/wiki/Latin-1_Supplement_(Unicode_block))]

1 ü → 252 → 0xFC

2

3 ü in Unicode:

4 U+00 00FC (8 bits → 2 bytes required)

5 F C

6 1111 1100

7

8 ü in UTF-8:

9 11000011 10111100

10 C 3 B C → 0xC3BC

# UTF-8 - example

## Encode the „ü“ into UTF-8!

[ü: [https://en.wikipedia.org/wiki/Latin-1\\_Supplement\\_\(Unicode\\_block\)](https://en.wikipedia.org/wiki/Latin-1_Supplement_(Unicode_block))]

1 ü → 252 → 0xFC

2

3 ü in Unicode:

4 U+00 00FC (8 bits → 2 bytes required)

5 F C

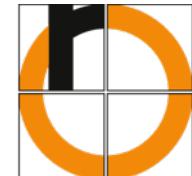
6 1111 1100

7

8 ü in UTF-8:

9 11000011 10111100

10 C 3 B C → 0xC3BC



# Questions?

All right?  $\Rightarrow$



Question?  $\Rightarrow$



and use **chat**

or

*speak after I  
ask you to*

# UTF-16

## UTF-16 length

Number of bytes	Bits for code point	Unicode range	Comment
2	16	0 - 00 FFFF	
4	20	01 0000 - 10 FFFF	subtraction required: U+XXXXXX - 0x10000

## UTF-16 encoding details

Unicode range	Byte 1	Byte 2	Byte 3	Byte 4
0 - 00 FFFF	xxxxxxxx	xxxxxxxx		
		High surrogate		Low surrogate
01 0000 - 10 FFFF	110110xx	xxxxxxxx	110111xx	xxxxxxxx

# UTF-16

## UTF-16 length

Number of bytes	Bits for code point	Unicode range	Comment
2	16	0 - 00 FFFF	
4	20	01 0000 - 10 FFFF	subtraction required: U+XXXXXX - 0x10000

## UTF-16 encoding details

Unicode range	Byte 1	Byte 2	Byte 3	Byte 4
0 - 00 FFFF	xxxxxxxx	xxxxxxxx		
		High surrogate		Low surrogate
01 0000 - 10 FFFF	110110xx	xxxxxxxx	110111xx	xxxxxxxx

# UTF-16 - example

**Encode the „“ (U+1F600) into UTF-16!**

1 4 byte variant and therefore correction required:

2  $0x1F600 - 0x10000 = 0xF600$

3

4 F 6 0 0

5 1111 0110 0000 0000

6

7 In UTF-16:

8 High surrogate	Low surrogate
9 11011000 00111101	11011110 00000000
10 D 8 3 D	D E 0 0

→ 0xD83DDE00

# UTF-16 - example

**Encode the „“ (U+1F600) into UTF-16!**

- 1 4 byte variant and therefore correction required:
- 2  $0x1F600 - 0x10000 = 0xF600$

3  
4 F 6 0 0  
5 1111 0110 0000 0000

6  
7 In UTF-16:  
8      High surrogate      Low surrogate  
9      11011000 00111101      11011110 00000000  
10     D    8    3    D      D    E      0    0      -> 0xD83DDE00

# UTF-16 - example

**Encode the „“ (U+1F600) into UTF-16!**

- 1 4 byte variant and therefore correction required:
- 2  $0x1F600 - 0x10000 = 0xF600$

3  
4 F 6 0 0  
5 1111 0110 0000 0000

6  
7 In UTF-16:  
8      High surrogate      Low surrogate  
9      11011000 00111101      11011110 00000000  
10     D    8    3    D      D    E      0    0      -> 0xD83DDE00

# UTF-16 - example

**Encode the „“ (U+1F600) into UTF-16!**

1 4 byte variant and therefore correction required:

2  $0x1F600 - 0x10000 = 0xF600$

3

4 F 6 0 0

5 1111 0110 0000 0000

6

7 In UTF-16:

8 High surrogate

9 11011000 00111101

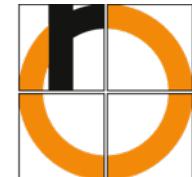
10 D 8 3 D

Low surrogate

11011110 00000000

D E 0 0

-> 0xD83DDE00



# Questions?

All right?  $\Rightarrow$



Question?  $\Rightarrow$



and use **chat**

or

*speak after I  
ask you to*

# UTF-32

## UTF-32 length

Number of bytes	Bits for code point	Unicode range	Comment
4	21	00 0000 - 10 FFFF	directly representable

## UTF-32 encoding details

Unicode range	Byte 1	Byte 2	Byte 3	Byte 4
0 - 10 FFFF	00000000	000xxxxx	xxxxxxx	xxxxxxx

# UTF-32

## UTF-32 length

Number of bytes	Bits for code point	Unicode range	Comment
4	21	00 0000 - 10 FFFF	directly representable

## UTF-32 encoding details

Unicode range	Byte 1	Byte 2	Byte 3	Byte 4
0 - 10 FFFF	00000000	000xxxxx	xxxxxxxx	xxxxxxxx

# UTF-32 - example

Encode the „“ (U+1F600) into UTF-32!

1 Only the 4 byte variant exists

2 0x1F600

3

4 1 F 6 0 0  
5 0001 1111 0110 0000 0000

6

7 In UTF-32:

8 00000000 00000001 11110110 00000000  
9 0 0 0 1 F 6 0 0      -> 0x0001F600

# UTF-32 - example

Encode the „“ (U+1F600) into UTF-32!

- 1 Only the 4 byte variant exists
- 2 0x1F600

3  
4 1 F 6 0 0  
5 0001 1111 0110 0000 0000

6  
7 In UTF-32:  
8 00000000 00000001 11110110 00000000  
9 0 0 0 1 F 6 0 0      -> 0x0001F600

# UTF-32 - example

Encode the „“ (U+1F600) into UTF-32!

1 Only the 4 byte variant exists

2 0x1F600

3

4 1 F 6 0 0  
5 0001 1111 0110 0000 0000

6

7 In UTF-32:

8 00000000 00000001 11110110 00000000  
9 0 0 0 1 F 6 0 0      -> 0x0001F600

# UTF-32 - example

Encode the „“ (U+1F600) into UTF-32!

1 Only the 4 byte variant exists

2 0x1F600

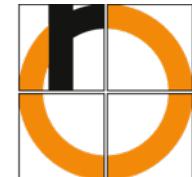
3

4 1 F 6 0 0  
5 0001 1111 0110 0000 0000

6

7 In UTF-32:

8 00000000 00000001 11110110 00000000  
9 0 0 0 1 F 6 0 0      -> 0x0001F600



# Questions?

All right?  $\Rightarrow$



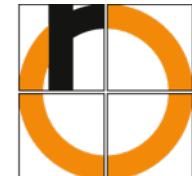
Question?  $\Rightarrow$



and use **chat**

or

*speak after I  
ask you to*



# Numbers

## Type

Integer

## Common data type

unsigned int, int, ...

## Realisation

Hardware: ALU

Floating point – binary

float, double, ...

Hardware: FPU

Floating point – decimal

decimal32, decimal64, ...

Mostly in software

Fixed point – binary

Often not well integrated

Mostly in software

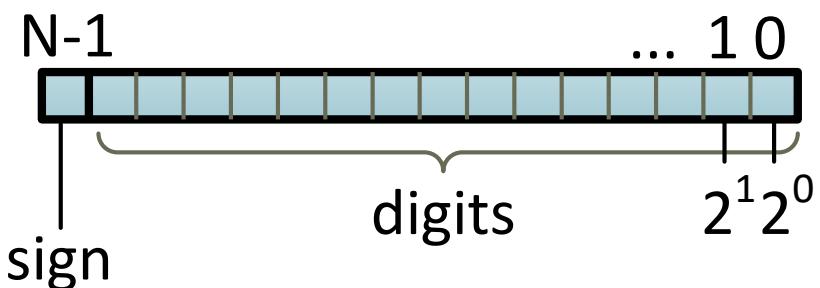
Fixed point – decimal

Often not well integrated

Mostly in software

# Integer (signed)

**Example:** short int



Positive number:

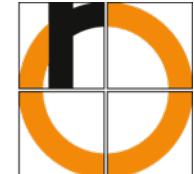
The weight for position  $i$  is  $2^i$

Negative number:

The sign is interpreted as  $-2^N$

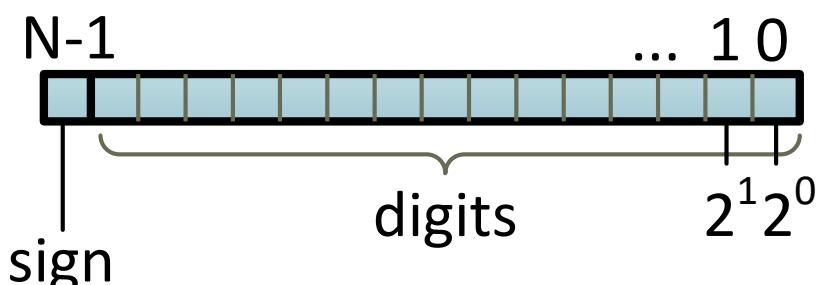
Example short int: Minimum: -32768; Maximum: 32767

limits: <http://www.cplusplus.com/reference/climits>



# Integer (signed)

Example: short int



Positive number:

The weight for position  $i$  is  $2^i$

Negative number:

The sign is interpreted as  $-2^N$

Example short int: Minimum: -32768; Maximum: 32767

limits: <http://www.cplusplus.com/reference/climits>

# Floating point – binary

**Usually scientific numbers with mantissa and exponent.**  
Requires hardware support (FPU - floating point unit).

Format:  $x = m \cdot B^e$  (m = mantissa, B = basis, and e = exponent)

## Examples:

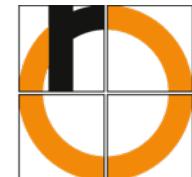
- C: float x;
- Ada: x: float

# Floating point – binary

Floating point binary formats are defined in the **IEEE Standard for Floating-Point Arithmetic (IEEE 754)**.

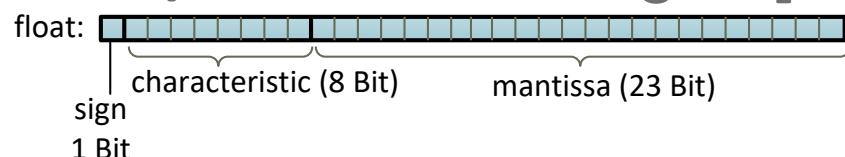
Name	Common name	of bits	Number Characteristic	Mantissa
binary16	Half precision	16	5 bits; $c = e + 15$	10 bits
float	Single precision	32	8 bits; $c = e + 127$	23 bits
double	Double precision	64	11 bits; $c = e + 1023$	52 bits
long double	Quadruple precision	128	15 bits; $c = e + 16383$	112 bits
(long long double?)	Octuple precision	256	19 bits; $c = e + 262143$	236 bits

IEEE 754 on Wikipedia: [https://en.wikipedia.org/wiki/IEEE\\_754](https://en.wikipedia.org/wiki/IEEE_754)



# Floating point – binary

## Example: float (single precision)



Exponent  $-126, \dots, +127$  Exponent is represented via the characteristic

Characteristic  $c = e + 127$

Mantissa  $1 \leq m < B$

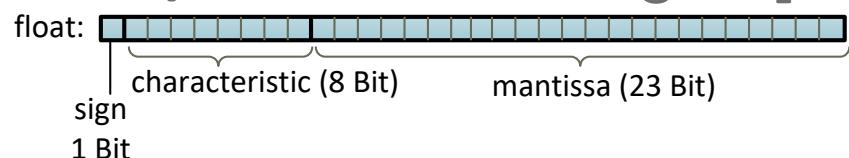
Is normalised in the binary system:

$1.MMM\dots M$

Advantage: 1 doesn't have to be saved!

# Floating point – binary

## Example: float (single precision)



Exponent  $-126, \dots, +127$  Exponent is represented via the characteristic

Characteristic  $c = e + 127$

Mantissa  $1 \leq m < B$

Is normalised in the binary system:

$1.MMM\dots M$

Advantage: 1 doesn't have to be saved!

# Floating point – binary

**Convert the decimal number 1.75 into the binary32 (float) representation.**

1 1.75 → binary:

2 01.11000...0 → it has already the required form  
of 1.MMM...M ( $\Rightarrow e=0$ )

3

$$5 \quad c = e + 127 = 0 + 127 = 127$$

6

7 S|C |M

8 0|01111111|11000000000000000000000000000000|

9

10 Hex representation:

11 0x3fe00000

# Floating point – binary

**Convert the decimal number 1.75 into the binary32 (float) representation.**

- 1 1.75 -> binary:
- 2 01.11000...0 -> it has already the required form of 1.MMM...M ( $\Rightarrow e=0$ )
- 3
- 4

$$5 c = e + 127 = 0 + 127 = 127$$

6  
7 S|C |M  
8 0|01111111|11000000000000000000000000000000 |

- 9
- 10 Hex representation:
- 11 0x3fe00000

# Floating point – binary

**Convert the decimal number 1.75 into the binary32 (float) representation.**

1 1.75 -> binary:

2 01.11000...0 -> it has already the required form  
3 of 1.MMM...M ( $\Rightarrow e=0$ )

4

5  $c = e + 127 = 0 + 127 = 127$

6

7 S|C |M  
8 0|01111111|11000000000000000000000000000000 |

9

10 Hex representation:

11 0x3fe00000

# Floating point – binary

**Convert the decimal number 1.75 into the binary32 (float) representation.**

1 1.75 -> binary:

2 01.11000...0 -> it has already the required form  
3 of 1.MMM...M ( $\Rightarrow e=0$ )

4

5  $c = e + 127 = 0 + 127 = 127$

6

7 S|C |M

8 0|01111111|11000000000000000000000000000000|

9

10 Hex representation:

11 0x3fe00000

# Floating point – binary

**Convert the decimal number 1.75 into the binary32 (float) representation.**

1 1.75 → binary:

2 01.11000...0 → it has already the required form  
3 of 1.MMM...M ( $\Rightarrow e=0$ )

4

5  $c = e + 127 = 0 + 127 = 127$

6

7 S | C | M

8 0 | 011 | 1111 | 110 | 00000000000000000000 |

9

10 Hex representation:

11 **0x3fe00000**

# Floating point – binary

Let's do some (binary) floating point number crunching.

Nr.	Code	different	equal
-----	------	-----------	-------

# Floating point – binary

Let's do some (binary) floating point number crunching.

Nr.	Code	different	equal
1	36.2 != 36.2		

# Floating point – binary

Let's do some (binary) floating point number crunching.

Nr.	Code	different	equal
1	36.2 != 36.2		
2	0.362 * 100.0 != 36.2		

# Floating point – binary

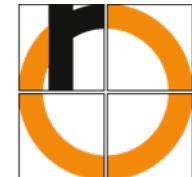
Let's do some (binary) floating point number crunching.

Nr.	Code	different	equal
1	$36.2 \neq 36.2$		
2	$0.362 * 100.0 \neq 36.2$		
3	$0.362 * (100.0 / 100.0) \neq 0.362$		

# Floating point – binary

Let's do some (binary) floating point number crunching.

Nr.	Code	different	equal
1	$36.2 \neq 36.2$		
2	$0.362 * 100.0 \neq 36.2$		
3	$0.362 * (100.0 / 100.0) \neq 0.362$		
4	$(0.362 * 100.0) / 100.0 \neq 0.362$		



# Questions?

All right?  $\Rightarrow$



Question?  $\Rightarrow$



and use **chat**

or

*speak after I  
ask you to*

# Floating point – decimal

Floating point decimal formats are defined in the **IEEE Standard for Floating-Point Arithmetic (IEEE 754)**.

Format:  $x = (-1)^{\text{signbit}} \times 10^{\text{exponentbits}_2 - 101_{10}} \times \text{true significand}_{10}$

Name	Number of decimal digits	Exponent min.	Exponent max.
decimal32	7	-95	+96
decimal64	16	-383	+384
decimal128	34	-6143	+6144

IEEE 754 on Wikipedia: [https://en.wikipedia.org/wiki/IEEE\\_754](https://en.wikipedia.org/wiki/IEEE_754)

- Possible in gnu C with `_Decimal32`, `_Decimal64`, and `_Decimal128`
- Example C: `_Decimal32 x = 0.1df;`
- Possible in gnu C++ with `decimal32`, `decimal64`, and `decimal128`
- Example C++: `std::decimal::decimal32 x(0.1);`

More details on the format (on Wikipedia): [https://en.wikipedia.org/wiki/Decimal32\\_floating-point\\_format](https://en.wikipedia.org/wiki/Decimal32_floating-point_format)

# Floating point – decimal

Floating point decimal formats are defined in the **IEEE Standard for Floating-Point Arithmetic (IEEE 754)**.

Format:  $x = (-1)^{\text{signbit}} \times 10^{\text{exponentbits}_2 - 101_{10}} \times \text{truesignificand}_{10}$

Name	Number of decimal digits	Exponent min.	Exponent max.
decimal32	7	-95	+96
decimal64	16	-383	+384
decimal128	34	-6143	+6144

IEEE 754 on Wikipedia: [https://en.wikipedia.org/wiki/IEEE\\_754](https://en.wikipedia.org/wiki/IEEE_754)

- Possible in gnu C with `_Decimal32`, `_Decimal64`, and `_Decimal128`
- Example C: `_Decimal32 x = 0.1df;`
- Possible in gnu C++ with `decimal32`, `decimal64`, and `decimal128`
- Example C++: `std::decimal::decimal32 x(0.1);`

More details on the format (on Wikipedia): [https://en.wikipedia.org/wiki/Decimal32\\_floating-point\\_format](https://en.wikipedia.org/wiki/Decimal32_floating-point_format)

# Floating point – decimal

Floating point decimal formats are defined in the **IEEE Standard for Floating-Point Arithmetic (IEEE 754)**.

Format:  $x = (-1)^{\text{signbit}} \times 10^{\text{exponentbits}_2 - 101_{10}} \times \text{truesignificand}_{10}$

Name	Number of decimal digits	Exponent min.	Exponent max.
decimal32	7	-95	+96
decimal64	16	-383	+384
decimal128	34	-6143	+6144

IEEE 754 on Wikipedia: [https://en.wikipedia.org/wiki/IEEE\\_754](https://en.wikipedia.org/wiki/IEEE_754)

- Possible in gnu C with `_Decimal32`, `_Decimal64`, and `_Decimal128`
- Example C: `_Decimal32 x = 0.1df;`
- Possible in gnu C++ with `decimal32`, `decimal64`, and `decimal128`
- Example C++: `std::decimal::decimal32 x(0.1);`

More details on the format (on Wikipedia): [https://en.wikipedia.org/wiki/Decimal32\\_floating-point\\_format](https://en.wikipedia.org/wiki/Decimal32_floating-point_format)

# Floating point – decimal

Floating point decimal formats are defined in the **IEEE Standard for Floating-Point Arithmetic (IEEE 754)**.

Format:  $x = (-1)^{\text{signbit}} \times 10^{\text{exponentbits}_2 - 101_10} \times \text{truesignificand}_{10}$

Name	Number of decimal digits	Exponent min.	Exponent max.
decimal32	7	-95	+96
decimal64	16	-383	+384
decimal128	34	-6143	+6144

IEEE 754 on Wikipedia: [https://en.wikipedia.org/wiki/IEEE\\_754](https://en.wikipedia.org/wiki/IEEE_754)

- Possible in gnu C with `_Decimal32`, `_Decimal64`, and `_Decimal128`
- Example C: `_Decimal32 x = 0.1df;`
- Possible in gnu C++ with `decimal32`, `decimal64`, and `decimal128`
- Example C++: `std::decimal::decimal32 x(0.1);`

More details on the format (on Wikipedia): [https://en.wikipedia.org/wiki/Decimal32\\_floating-point\\_format](https://en.wikipedia.org/wiki/Decimal32_floating-point_format)

# Floating point – decimal

Floating point decimal formats are defined in the **IEEE Standard for Floating-Point Arithmetic (IEEE 754)**.

Format:  $x = (-1)^{\text{signbit}} \times 10^{\text{exponentbits}_2 - 101_10} \times \text{truesignificand}_{10}$

Name	Number of decimal digits	Exponent min.	Exponent max.
decimal32	7	-95	+96
decimal64	16	-383	+384
decimal128	34	-6143	+6144

IEEE 754 on Wikipedia: [https://en.wikipedia.org/wiki/IEEE\\_754](https://en.wikipedia.org/wiki/IEEE_754)

/ Java: Big Decimal

- Possible in gnu C with \_Decimal32, \_Decimal64, and \_Decimal128
- Example C: \_Decimal32 x = 0.1df;
- Possible in gnu C++ with decimal32, decimal64, and decimal128
- Example C++: std::decimal::decimal32 x(0.1);

More details on the format (on Wikipedia): [https://en.wikipedia.org/wiki/Decimal32\\_floating-point\\_format](https://en.wikipedia.org/wiki/Decimal32_floating-point_format)

# Floating point – decimal

Let's do some (decimal) floating point number crunching.

Nr.	Code	different	equal
-----	------	-----------	-------

# Floating point – decimal

Let's do some (decimal) floating point number crunching.

Nr.	Code	different	equal
1	36.2 != 36.2		

# Floating point – decimal

Let's do some (decimal) floating point number crunching.

Nr.	Code	different	equal
1	36.2 != 36.2		
2	0.362 * 100.0 != 36.2		

# Floating point – decimal

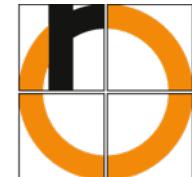
Let's do some (decimal) floating point number crunching.

Nr.	Code	different	equal
1	$36.2 \neq 36.2$		
2	$0.362 * 100.0 \neq 36.2$		
3	$0.362 * (100.0 / 100.0) \neq 0.362$		

# Floating point – decimal

Let's do some (decimal) floating point number crunching.

Nr.	Code	different	equal
1	$36.2 \neq 36.2$		
2	$0.362 * 100.0 \neq 36.2$		
3	$0.362 * (100.0 / 100.0) \neq 0.362$		
4	$(0.362 * 100.0) / 100.0 \neq 0.362$		



# Questions?

All right?  $\Rightarrow$



Question?  $\Rightarrow$



and use **chat**

or

*speak after I  
ask you to*

# Fixed point

Fixed point numbers have a **fixed imaginary point** that is not moved.

## Usage:

- Areas where rounding errors must be avoided (e.g. commercial applications)
- If no floating point hardware (FPU) is available (e.g. in embedded systems)
- Devices use the numbers in this format anyway (e.g. analog/digital converter)

## Two variants:

Type

Binary fixed point

Decimal fixed point

Usage

technical

economical

# Fixed point

Fixed point numbers have a **fixed imaginary point** that is not moved.

## Usage:

- Areas where rounding errors must be avoided (e.g. commercial applications)
- If no floating point hardware (FPU) is available (e.g. in embedded systems)
- Devices use the numbers in this format anyway (e.g. analog/digital converter)

## Two variants:

Type	Usage
Binary fixed point	technical
Decimal fixed point	economical

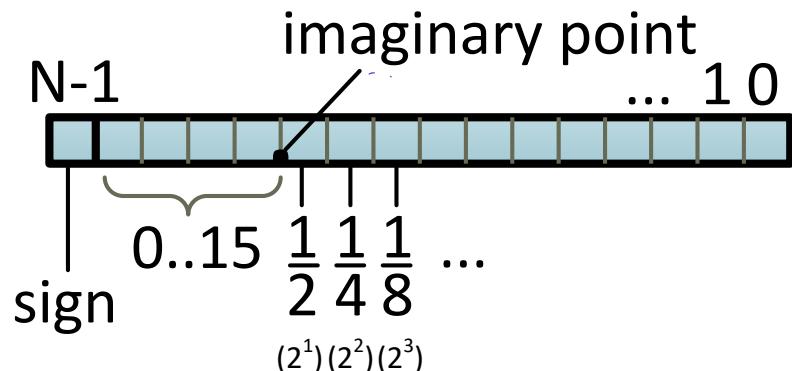


# Fixed point – binary

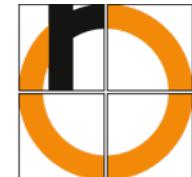
Uses integers with an **imaginary binary point**.

Often without specialised hardware: *Poor man's floating point.*

Ada: `type analog_input is delta 0.125 range -16.0..15.0;`



[C++ library example: Compositional Numeric Library]



# Questions?

All right?  $\Rightarrow$



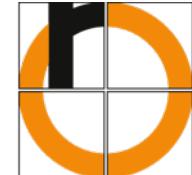
Question?  $\Rightarrow$



and use **chat**

or

*speak after I  
ask you to*



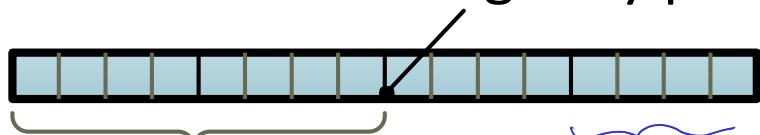
# Fixed point – decimal

Uses the **binary coded decimal (BCD)** system with an imaginary decimal point and BCD arithmetic.

Used in IBM main frame. Sometimes there exists specialised hardware.

**BCD:** Every digit (0 – 9) is represented by 4 bits

Ada: `type money is delta 0.01 digits 8;`  
imaginary point

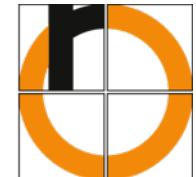


2 digits per byte

$$\rightarrow 9 \Rightarrow 1001$$

$0x85$   
↑  
0000.1000 0101

[C++ library example: Decimal data type for C++ ]



# Questions?

All right?  $\Rightarrow$



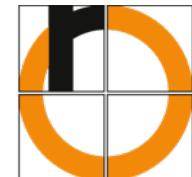
Question?  $\Rightarrow$



and use **chat**

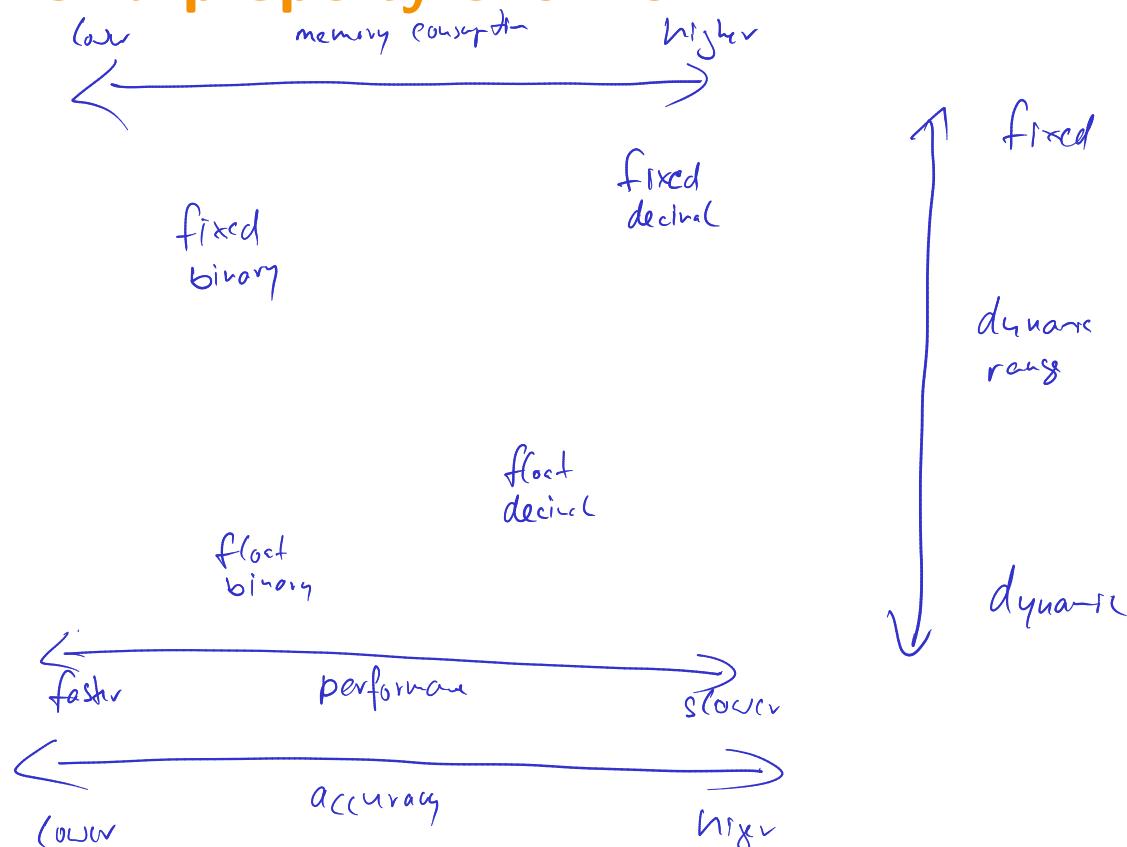
or

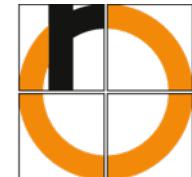
*speak after I  
ask you to*



# When to use what?

## A first try for a property overview





# Questions?

All right?  $\Rightarrow$



Question?  $\Rightarrow$



and use **chat**

or

*speak after I  
ask you to*



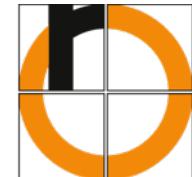
# Summary and outlook

## Summary

- Important basics
- ASCII
- Unicode and UTF
- Data types: Numbers

## Outlook

- Logical hardware



# Summary and outlook

## Summary

- Important basics
- ASCII
- Unicode and UTF
- Data types: Numbers

## Outlook

- Logical hardware