Introduction to Informatics

Course 05

Non-numeric characters

- extended ASCII is more widely spread in practice (American Standard for Information Interchange)
  - English abc small and capital letters
  - digits
  - punctuation characters
  - special control characters
- 1 byte= 1 character
- 128
  - standard, 7 bit
- +128
  - extended
  - specials, code tables
    - Hungarian: 852, Hungarian Windows: 1250
    - problems: communication between the machines and programs

ASCII standard

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

Source: www.asciitable.com

ASCII standard, extended (Latin-1)

ASCII	0000 0000	0000 0001	0000 0010	0000 0011	0000 0100	0000 0101	0000 0110	0000 0111	0000 1000	0000 1001	0000 1010	0000 1011	0000 1100	0000 1101	0000 1110	0000 1111
0000	$^{N}_{U}$	$^{S}_{H}$	$^{S}_{X}$	$^{E}_{X}$	$^{E}_{T}$	$^{E}_{Q}$	$^{A}_{K}$	$^{B}_{L}$	$^{B}_{S}$	$^{H}_{T}$	$^{L}_{F}$	$^{Y}_{T}$	$^{F}_{F}$	$^{C}_{R}$	$^{S}_{O}$	$^{S}_{I}$
0001	$^{D}_{L}$	$^{D}_{1}$	$^{D}_{2}$	$^{D}_{3}$	$^{D}_{4}$	$^{N}_{K}$	$^{S}_{Y}$	$^{E}_{\Sigma}$	$^{C}_{N}$	$^{E}_{M}$	$^{S}_{B}$	$^{E}_{C}$	$^{F}_{S}$	$^{G}_{S}$	$^{R}_{S}$	$^{U}_{S}$
0010		!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
0011	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
0100	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0101	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
0110	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
0111	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	$^{D}_{T}$
1000	$^{8}_{0}$	$^{8}_{1}$	$^{8}_{2}$	$^{8}_{3}$	$^{I}_{N}$	$^{N}_{L}$	$^{S}_{S}$	$^{E}_{S}$	$^{H}_{S}$	$^{H}_{J}$	$^{Y}_{S}$	$^{P}_{D}$	$^{P}_{V}$	$^{R}_{I}$	$^{S}_{2}$	$^{S}_{3}$
1001	$^{D}_{C}$	$^{P}_{1}$	$^{P}_{Z}$	$^{S}_{E}$	$^{C}_{C}$	$^{M}_{M}$	$^{S}_{P}$	$^{E}_{P}$	$^{O}_{S}$	$^{O}_{O}$	$^{O}_{A}$	$^{C}_{S}$	$^{S}_{T}$	$^{O}_{S}$	$^{P}_{M}$	$^{A}_{P}$
1010	$^{A}_{O}$	¡	¢	£		¥	¦	§	¨	©	ª	«	¬	-	®	¯
1011	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
1100	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
1101	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
1110	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
1111	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Unicode

- Universal Character Set
- the most recent version is Unicode 6.3.
- The Unicode Standard, the latest version of Unicode consists of a repertoire of more than 110,000
- characters covering:
  - 100 scripts
  - a set of code charts
  - an encoding methodology
  - set of standard character encodings
  - an enumeration of character properties
  - a set of reference data computer files
  - a number of related items
  - rules for normalization
  - decomposition
  - collation
  - rendering
  - bidirectional display order

Unicode

- imaginative/virtual limit 2^{31}
  - code space: 0–10FFFF
  - the whole character representation

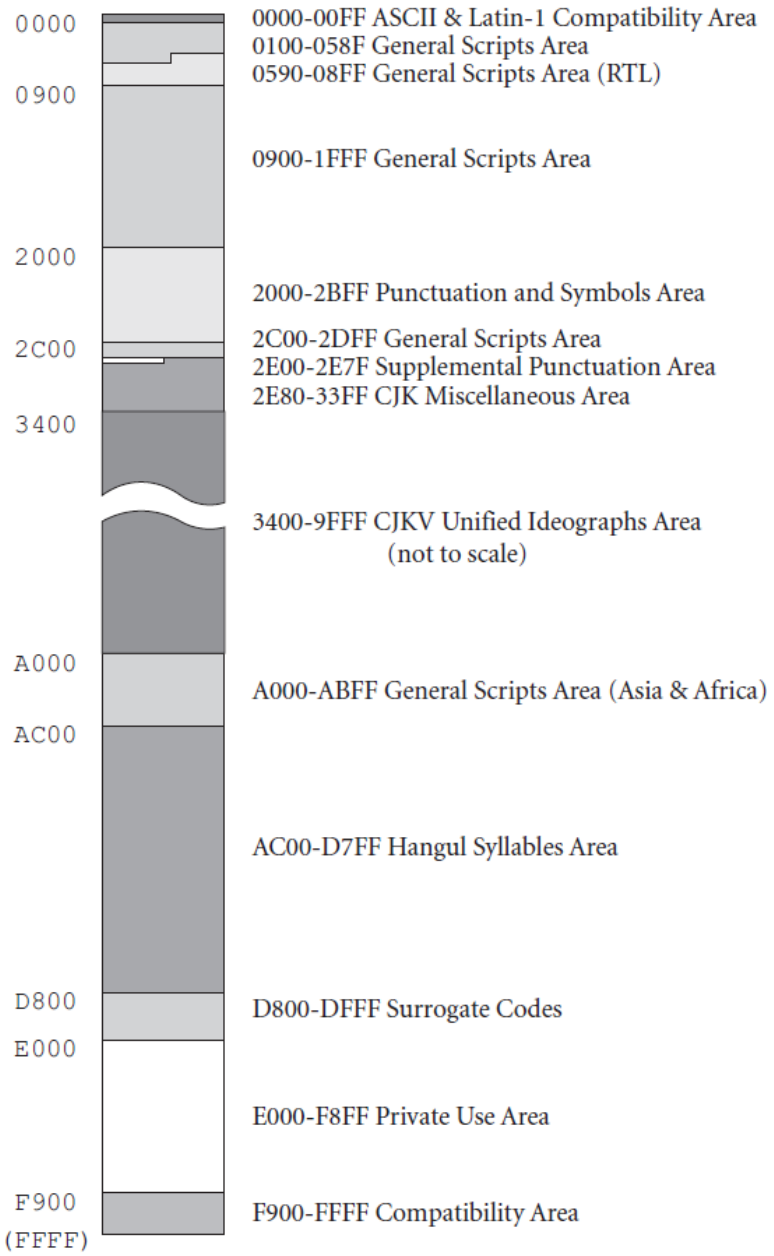
- UTF-8
- UTF-16
- UTF-32

- 16 bits planes
  - the last four hexadecimal digits is the position in the plane
  - leader digits refer to the plane

Unicode planes

- Plane 0, Unicode low 16 bits, Basic Multilingual Plane (BMP)
  - low 128 value: ASCII
  - low 256 value: Latin-1
  - the most frequently used characters of the modern world, rare or historical characters
- Plane 1, Supplementary Multilingual Plane (SMP)
  - rarely used characters: gothic letters, musical notes, domino characters
- Plane 2, Supplementary Ideographic Plane (SIP)
  - very rare CJK characters
- Plane 14, Supplementary Special-purpose Plane (SSP)
  - excluded format characters
- Planes 15 and 16, Private Use Planes

Unicode Plane



Unicode Transformation Format (UTF)

- UTF-32 (32-bit Unicode Transformation Format)
  - complete
  - fixed length codes: 4 byte per character
  - one-one correspondence
- UTF-16 (16-bit Unicode Transformation Format)
  - U+0000…U+FFFF intervals (BMP) 16 bits
  - U+10000…10FFFF intervals (supplementary planes) pair of 16 bits
  - UTF-16 fixed length of the BMP
- UTF-8 (8-bit Unicode Transformation Format)
  - More compact
    - Varying length codes
    - The longest 6 bytes
    - The codes stored in one byte equal ASCII

Examples of Unicode Encoding Forms

Code Point	Encoding Form	Code Unit Sequence
U+004D	UTF-32	0000004D
	UTF-16	004D
	UTF-8	4D
U+0430	UTF-32	00000430
	UTF-16	0430
	UTF-8	D0 B0
U+4E8C	UTF-32	00004E8C
	UTF-16	4E8C
	UTF-8	E4 BA 8C
U+10302	UTF-32	00010302
	UTF-16	D800 DF02
	UTF-8	F0 90 8C 82

Unicode UTF-8

0xxxxxxx					
110xxxxx	**10**xxxxxx				
1110xxxx	**10**xxxxxx	**10**xxxxxx			
11110xxx	**10**xxxxxx	**10**xxxxxx	**10**xxxxxx		
111110xx	**10**xxxxxx	**10**xxxxxx	**10**xxxxxx	**10**xxxxxx	
1111110x	**10**xxxxxx	**10**xxxxxx	**10**xxxxxx	**10**xxxxxx	**10**xxxxxx

Unicode UTF-8 Exercise

- Give the Unicode value of © and the representation of UTF-8 in the hexadecimal form.
- Unicode value: U+00A9

$$1010\ 1001_{(2} = A9_{(16}$$

110xxxxx 10xxxxxx

110xxx10 10101001

11000010 10101001

C2 A9

Unicode UTF-8 Exercise

Character		Binary code point	Binary UTF-8	Hexadecimal UTF-8
$	**U+0024**			
¢	**U+00A2**			
€	**U+20AC**			
𤭢	**U+24B62**			

Unicode UTF-8 Solution

Character		Binary code point	Binary UTF-8	Hexadecimal UTF-8
$	U+0024	0100100	00100100	24
¢	U+00A2	00010100010	11000010 10100010	C2 A2
€	U+20AC	0010000010101100	11100010 10000010 10101100	E2 82 AC
𤭢	U+24B62	000100100101101100010	11110000 10100100 10101101 10100010	F0 A4 AD A2

Exercise

1. Give the Unicode value of the BMP plane's given character and the encoding of UTF-8 in the hexadecimal form.

	0x00	0x01	0x02	0x03	0x04	0x05	0x06	0x07	0x08	0x09	0x0A	0x0B	0x0C	0x0D	0x0E	0x0F
0x3040		ぁ	あ	ぃ	い	ぅ	う	ぇ	え	ぉ	お	か	が	き	ぎ	く
0x3050	ぐ	け	げ	こ	ご	さ	ざ	し	じ	す	ず	せ	ぜ	そ	ぞ	た
0x3060	だ	ち	ぢ	っ	つ	づ	て	で	と	ど	な	に	ぬ	ね	の	は
0x3070	ば	ぱ	ひ	び	ぴ	ふ	ぶ	ぷ	へ	べ	ぺ	ほ	ぼ	ぽ	ま	み
0x3080	む	め	も	ゃ	や	ゅ	ゆ	ょ	よ	ら	り	る	れ	ろ	ゎ	わ
0x3090	ゐ	ゑ	を	ん	ゔ					゛	゜	ゝ	ゞ			

Solution

- U+3086
- 11 0000 1000 0110
- 1110xxxx 10xxxxxx 10xxxxxx
- 11100011 10000010 10000110
- E38286

Exercise

2. Give the Unicode value of the BMP plane's given character and the encoding of UTF-8 in the hexadecimal form.

	0x00	0x01	0x02	0x03	0x04	0x05	0x06	0x07	0x08	0x09	0x0A	0x0B	0x0C	0x0D	0x0E	0x0F
0x3040		ぁ	あ	ぃ	い	ぅ	う	ぇ	え	ぉ	お	か	が	き	ぎ	く
0x3050	ぐ	け	げ	こ	ご	さ	ざ	し	じ	す	ず	せ	ぜ	そ	ぞ	た
0x3060	だ	ち	ぢ	っ	つ	づ	て	で	と	ど	な	に	ぬ	ね	の	は
0x3070	ば	ぱ	ひ	び	ぴ	ふ	ぶ	ぷ	へ	べ	ぺ	ほ	ぼ	ぽ	ま	み
0x3080	む	め	も	ゃ	や	ゅ	ゆ	ょ	よ	ら	り	る	れ	ろ	ゎ	わ
0x3090	ゐ	ゑ	を	ん	ゔ					゛	゜	ゝ	ゞ			

Solution

- U+307F
- 11 0000 0111 1111
- 1110xxxx 10xxxxxx 10xxxxxx
- 11100011 10000001 10111111
- E381BF

Unicode table and converter

- Unicode table:
  - http://www.tamasoft.co.jp/en/general-info/unicode.html
- Unicode converter
  - http://rishida.net/tools/conversion/

Logical Operations

- the lowest level in the computer's hardver structure – the digital logic level – consist of the gate circuits
  - analogue components
  - with their operation they serve as the base of the digital (binary) system
- in digital circuits we distinguished between two sign levels
  - low (L) level (between 0 and 1 Volt voltage)
    - false
    - 0
  - highest (H) level (between 2 and 5 Volt voltage)
    - truth
    - 1

Logical Operations

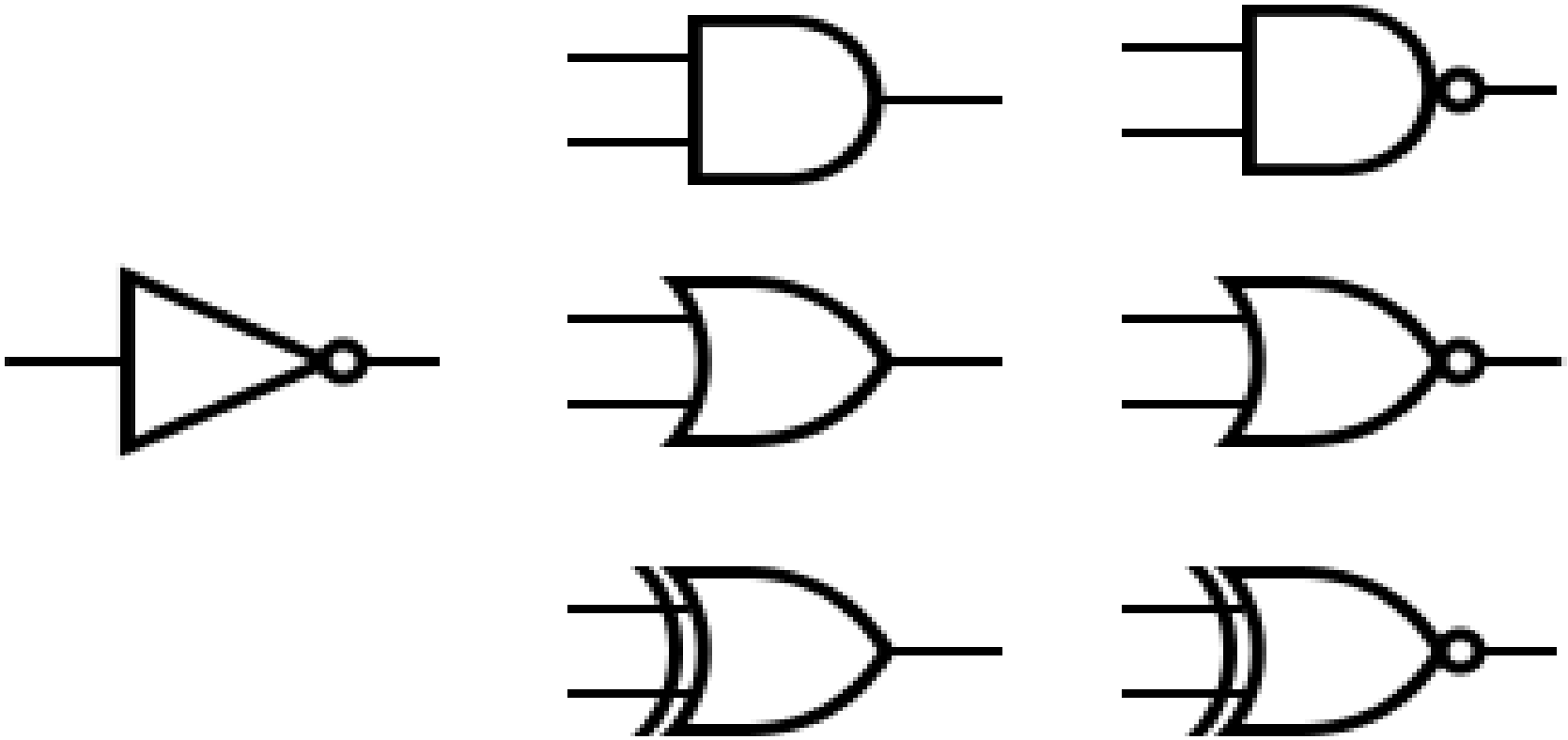
- Basic operations
  - NOT
  - AND
  - OR
- for the description of the circuits built from the combination of the gates
  - the variables and functions can be 0 and 1 value
  - Boole-algebra
    - Gottfried Wilhelm Leibniz (1646-1716)
    - George Boole (1815-1864)

Logical Operations

- logical functions
  - one or more input variables
  - fraction value depends only on the value of the logical variables
  - it gives the correlation between the input and output variable values in the logical operation

- manifestation forms
  - gate circuits
  - truth tables
  - set theory correspondence

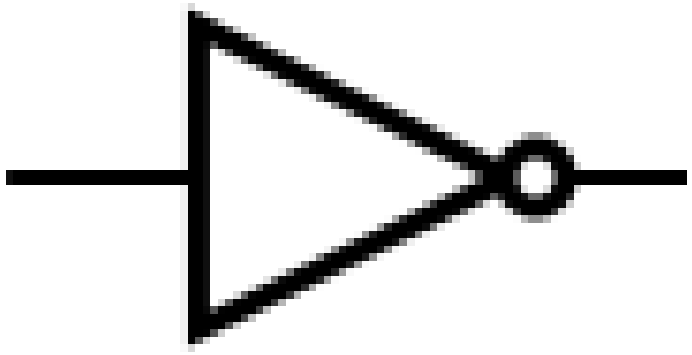
IEEE Standard Graphic Symbols for Logic Functions – Logic Gates



Digital logic gates

- logic gates
- circuit diagram
- Boolean operation
- conjunction (AND-gates)
- disjunction (OR-gates
- complement (inverters)
- input wires or ports
- output port

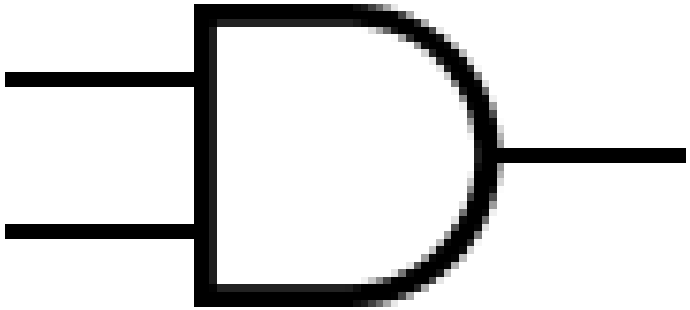
Logical NOT operation
NOT gate



A	Q
0	1
1	0

$$\text{NOT A} = \overline{A}$$

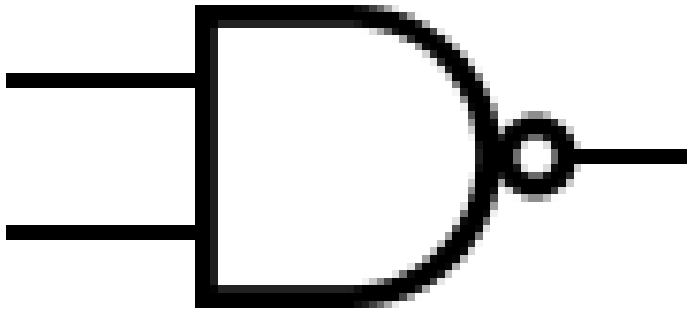
Logical AND operation
AND gate



A AND B $= A \cdot B$

A	B	Q
0	0	0
0	1	0
1	0	0
1	1	1

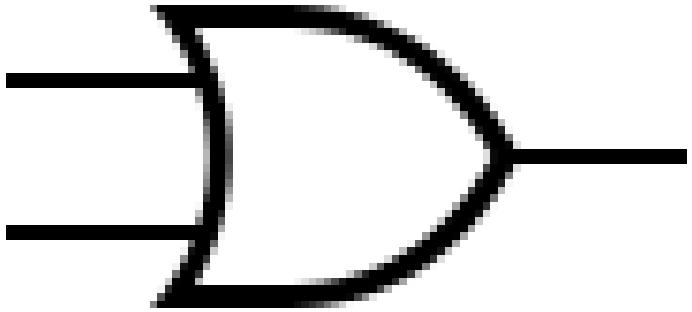
Logical NAND operation
NAND gate



$$A \text{ NAND } B = \overline{A \cdot B}$$

A	B	Q
0	0	1
0	1	1
1	0	1
1	1	0

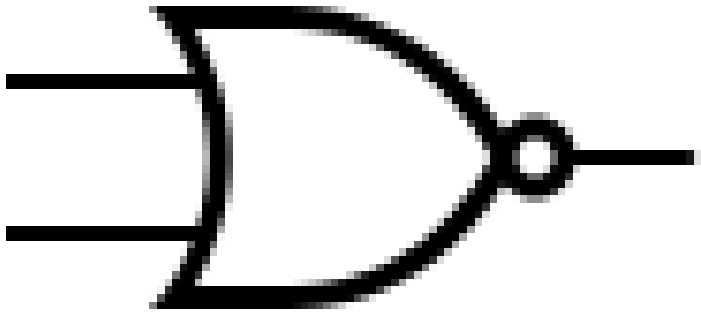
Logical OR operation
OR gate



A OR B $= A + B$

A	B	Q
0	0	0
0	1	1
1	0	1
1	1	1

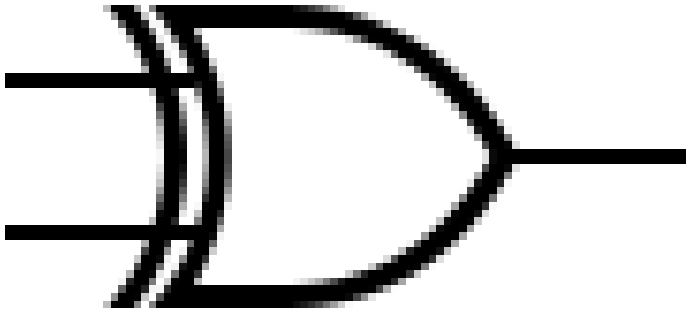
Logical NOR operation
NOR gate



$A \text{ NOR } B = \overline{A + B}$

A	B	Q
0	0	1
0	1	0
1	0	0
1	1	0

Logical XOR operation
XOR gate

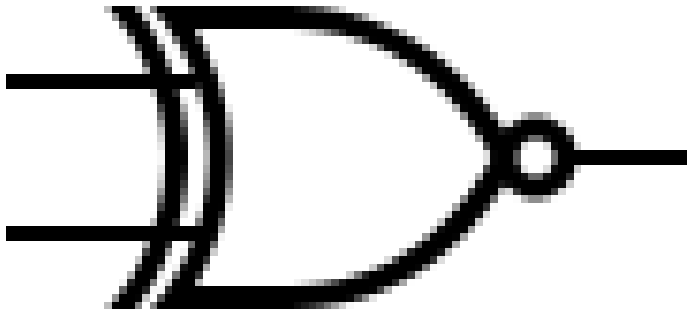


A XOR B = $A \oplus B$

$$A \cdot \overline{B} + \overline{A} \cdot B$$

A	B	Q
0	0	0
0	1	1
1	0	1
1	1	0

Logical XNOR operation



$$\text{A XNOR B} = \overline{A \oplus B}$$

$$A \cdot B + \overline{A} \cdot \overline{B}$$

A	B	Q
0	0	1
0	1	0
1	0	0
1	1	1

Truth Table

A	B	NOT A	A AND B	A NAND B	A OR B	A NOR B	A XOR B	A XNOR B
0	0	1	0	1	0	1	0	1
0	1	1	0	1	1	0	1	0
1	0	0	0	1	1	0	1	0
1	1	0	1	0	1	0	0	1

Exercise

Calculate the following logical operations.

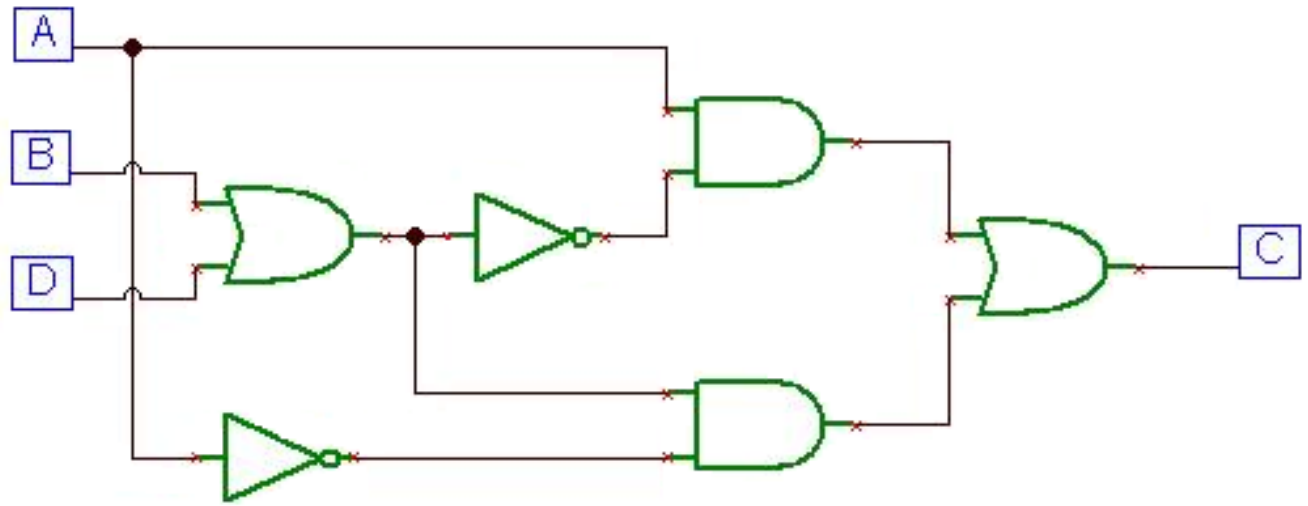
1. **NOT** (10010101)=
2. 01001010 **AND** 10110011=
3. 11001011 **OR** 10111011=
4. 01011010 **XOR** 10000011=
5. NOT(01001010) **AND** (10110011 **OR** 10110111)=
6. 11110101 **OR** (11010011 **NAND** 10111101)=
7. (01001010 **NOR** 10110011) **AND** 11000001=
8. 01001010 **XOR** (10110011 **AND** 11110111)=

Logical circuits

1. Describe the following circuit with logic expression. According to this give the mathematical equivalents. What is the value of expression, if A=0, B=1, D=0?

Exercise

2. Describe the following circuit with logic expression. According to this give the mathematical equivalents. What is the value of expression, if A=0, B=1, D=0?



Solution

(A AND NOT(B OR D)) OR ((B OR D) AND NOT(A))

$$\left[A \cdot \left(\overline{B + D}\right)\right] + \left[(B + D) \cdot \overline{A}\right]$$

1