ASCII AND EBCDIC CODES

How Computers Represent Data

ON

🔴 **OR** 🟢 **= 1 bit**

0   1

🔴 🔴 🟢 🟢 🔴 🟢 🔴 🔴 **= 1 Byte**

0   0   1   1   0   1   0   0

OR

🔴 🟢 🔴 🔴 🔴 🔴 🔴 🟢 **= 1 Byte**

0   1   0   0   0   0   0   1
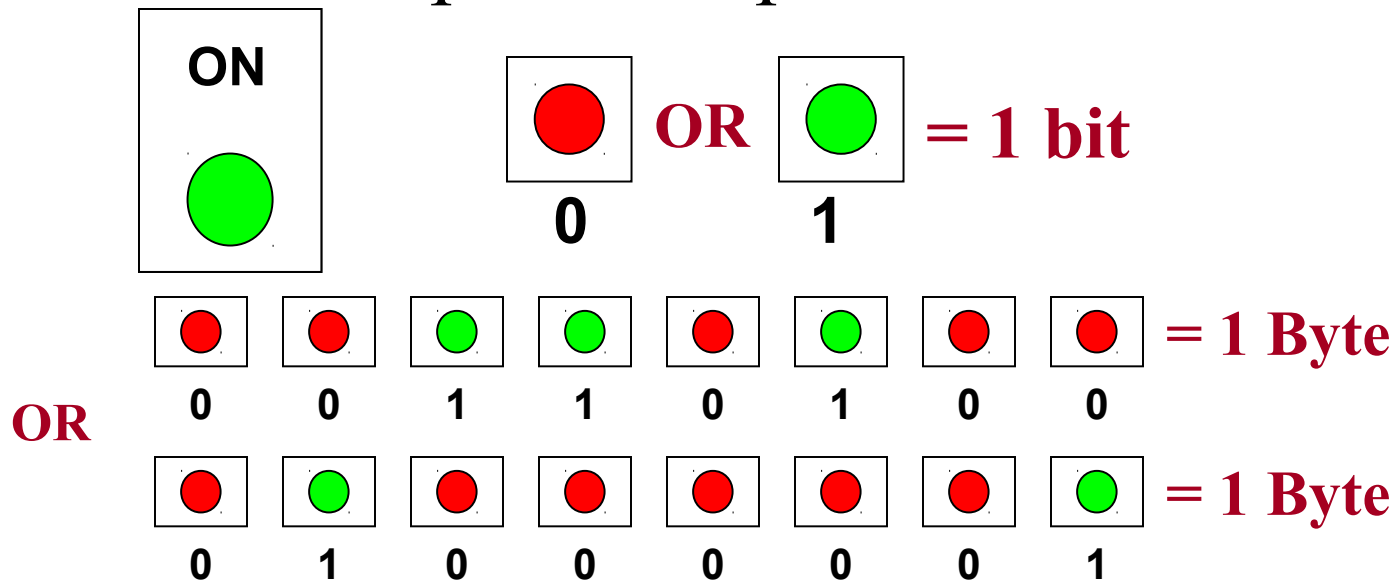
Bit (Binary digit) – On or off state of electric current; considered the basic unit of information; represented by 1s and 0s (binary numbers)

Byte – Eight bits grouped together to represent a character (an alphabetical letter, a number, or a punctuation symbol); 256 different combinations

Bits

> 1000 bits = 1 kilobit (kb)
>
> 1,000,000 bits = 1 megabit (mb)
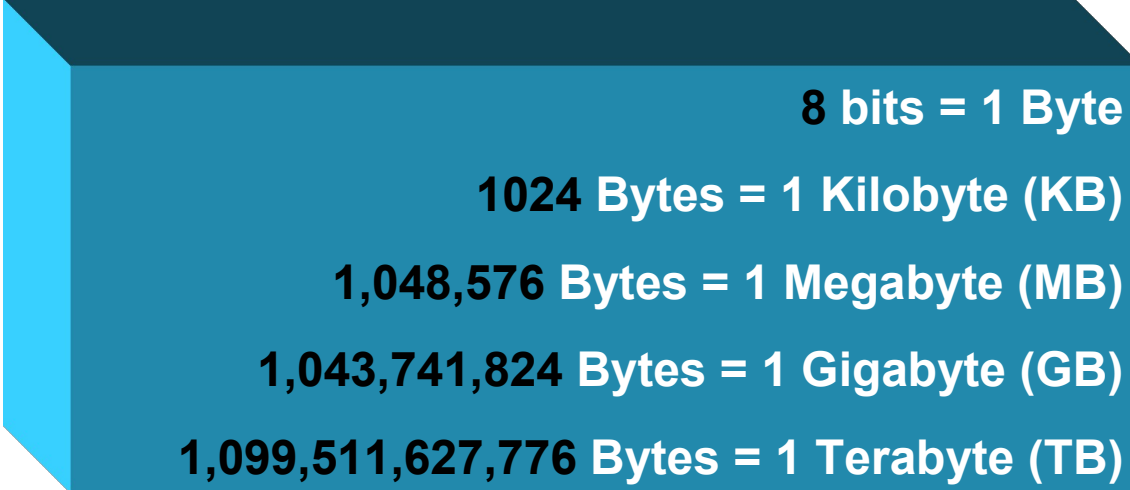>
> 1,000,000,000 bits = 1 gigabit (gb)

Kilobits per second (Kbps), **megabits per second** (Mbps), and **gigabits per second** (Gbps) are terms that describe units of data used in measuring data transfer rates

☐ Example: 56 Kbps modem

Bytes

8 bits = 1 Byte

1024 Bytes = 1 Kilobyte (KB)

1,048,576 Bytes = 1 Megabyte (MB)

1,043,741,824 Bytes = 1 Gigabyte (GB)

1,099,511,627,776 Bytes = 1 Terabyte (TB)

Kilobyte, **megabyte**, **gigabyte**, and **terabyte** are terms that describe large units of data used in measuring data storage

☐Example: 20 GB hard disk

Representing Characters: Character Codes

Character codes translate numerical data into characters readable by humans

- **American Standard Code for Information Interchange** (ASCII) – Eight bits equals one character; used by minicomputers and personal computers

- **Extended Binary Coded Decimal Interchange Code** (EBCDIC) – Eight bits equals one character; used by mainframe computers

- **Unicode –** Sixteen bits equals one character; over 65,000 combinations; used for foreign language symbols

ASCII 0 0 1 1 0 1 0 0 = 4

EBCDIC 1 1 1 1 0 1 0 0 = 4
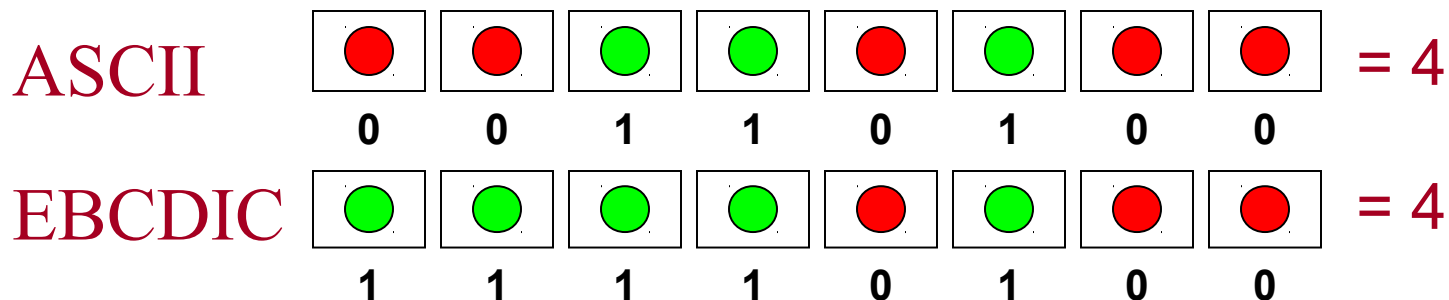
Codes and Characters

- The problem:
  - Representing text strings, such as "Hello, world", in a computer
- Each character is coded as a byte (= 8 bits)
- Most common coding system is ASCII
- ASCII = <u>A</u>merican National <u>S</u>tandard <u>C</u>ode for <u>I</u>nformation <u>I</u>nterchange

ASCII

- The **American Standard Code for Information Interchange** is a standard seven-bit code that was proposed by ANSI (American National Standards Institute) in 1963, and finalized in 1968.

- ASCII, pronounced "ask-key", is the common code for microcomputer . The standard ASCII character set consists of 128 decimal numbers ranging from zero through 127 assigned to letters, numbers, punctuation marks, and the most common special characters (see ASCII Table).

ASCII

- Computers can only understand numbers, so an ASCII code is the numerical representation of a character such as 'a' or '@' or an action of some sort.

ASCII Features

- 7-bit code
- 8th bit is unused (or used for a parity bit)
- $2^7 = 128$ codes
- Two general types of codes:
  - 95 are "Graphic" codes (displayable on a console)
  - 33 are "Control" codes (control features of the console or communications channel)

	000	001	010	011	100	101	110	111
0000	NULL	DLE		0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EDT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l	|
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

	000	001	010	011	100	101	110	111
0000	NULL	DLE		0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EDT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l		
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

$$\text{i.e. 'a'} = 1100001_2 = 97_{10} = 61_{16}$$

	000	001	010	011	100	101	110	111
0000	NULL	DLE		0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EDT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l	|
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

95 Graphic codes

	000	001	010	011	100	101	110	111
0000	NULL	DLE		0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EDT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l	|
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

33 Control codes

	000	001	010	011	100	101	110	111
0000	NULL	DLE		0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EDT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l		
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

Alphabetic codes

"Hello, world" Example

		Binary		Hexadecimal		Decimal
H	=	01001000	=	48	=	72
e	=	01100101	=	65	=	101
l	=	01101100	=	6C	=	108
l	=	01101100	=	6C	=	108
o	=	01101111	=	6F	=	111
,	=	00101100	=	2C	=	44
	=	00100000	=	20	=	32
w	=	01110111	=	77	=	119
o	=	01100111	=	67	=	103
r	=	01110010	=	72	=	114
l	=	01101100	=	6C	=	108
d	=	01100100	=	64	=	100

Note: 12 characters – requires 12 bytes
Each character requires 1 byte

	000	001	010	011	100	101	110	111
0000	NULL	DLE		0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EDT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l	|
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

Numeric codes

"4+15" Example

		Binary		Hexadecimal		Decimal
4	=	00110100	=	34	=	52
+	=	00101011	=	2B	=	43
1	=	00110001	=	31	=	49
5	=	00110101	=	35	=	53

"4+15" is "00110100 00101011 00110001 00110101"

or "$34_{16}2B_{16}31_{16}35_{16}$"

	000	001	010	011	100	101	110	111
0000	NULL	DLE		0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EDT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l		
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

Punctuation, etc.

EBCDIC

- <u>E</u>xtended <u>BCD</u> <u>I</u>nterchange <u>C</u>ode (pronounced *ebb'-se-dick*)
- 8-bit code
- Developed by IBM
- Rarely used today
- IBM mainframes only