Representation of information

course "Essentials of computing systems"

Feb.2022

contents

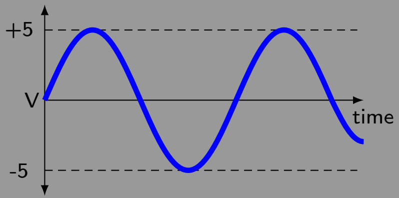
1. Digital abstraction
2. Bits, bytes and words
3. Textual information
4. Machine-level instructions
5. Images
6. Audio

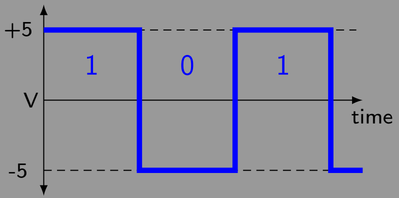
contents

1. Digital abstraction
2. Bits, bytes and words
3. Textual information
4. Machine-level instructions
5. Images
6. Audio

analog and digital signals

- Analog (or continuous) systems process time-varying signals that can take on any value across a continuous range of voltage, current or other metric.
- The same happens with digital systems, but the difference is that one pretends that they do not!
- Digital systems have signals that are represented by discrete (i.e., non-continuous) values.
- A digital signal is modelled as taking on only one of two possible values (‘0’ and ‘1’).





binary system

- The digital approach is not limited to only two values.
- The essential aspect is that the set of possible values is finite.
- The simplest form of digital systems is binary, where there are two possible values for the signals.
- The more values that must be distinguished, the less separation between adjacent values, and the less reliable is the mechanism.
- The binary numeral system is the most reliable method for encoding digital information, since it is easier to distinguish two possible values with physical entities than, say, five or ten.

digital abstraction

- The fundamental advantage of digital systems w.r.t. analog ones is their ability to deal with degraded electrical signals.
- Due to the discrete nature of the output signals, a small variation in an input value is still interpreted correctly.
- In analog circuits, this behaviour does not occur as a slight error at an input generates an error at the output.
- Digital circuits deal with analog voltages and currents.
- The digital abstraction allows analog behaviour to be ignored, so circuits can be modelled as if they really process 0s and 1s.

contents

1 Digital abstraction

2 Bits, bytes and words

3 Textual information

4 Machine-level instructions

5 Images

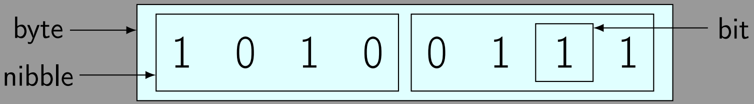
6 Audio

bits

- A logic value, 0 or 1, is often called a bit.
- A single bit is not very useful, since it just permits to represent two possible values.
- If more values must be represented, more bits can be considered.
- When bits are grouped and coded, the elements of any finite set can be represented.
- This can be achieved by assigning some interpretation that gives meaning to the different possible bit patterns.
- With n bits, 2^n different entities can be represented.
- By using a standard character code, the letters and symbols in a document can also be encoded with a set of bits.
- All information in computers is represented by patterns of bits.

byte and nibble

- A block of 8 bits, designated as a byte, is the smallest addressable unit of memory in most computers.
- Each half of a byte is called a nibble, which can be represented by a 4-bit pattern or a hexadecimal digit.



multiple-byte units

decimal			binary		
value		metric	value		metric
10^3	1000	kB kilobyte	2^{10}	1024	KiB kibibyte
10^6	1000^2	MB megabyte	2^{20}	1024^2	MiB mebibyte
10^9	1000^3	GB gigabyte	2^{30}	1024^3	GiB gibibyte
10^{12}	1000^4	TB terabyte	2^{40}	1024^4	TiB tebibyte
10^{15}	1000^5	PB petabyte	2^{50}	1024^5	PiB pebibyte
10^{18}	1000^6	EB exabyte	2^{60}	1024^6	EiB exbibyte
10^{21}	1000^7	ZB zettabyte	2^{70}	1024^7	ZiB zebibyte

main memory

- A machine-level program views main memory as a large array of bytes.
- Every byte in the memory is identified by a unique number, its address.
- A word is the basic unit of data handled by a given family of computers.
- The sizes of words historically range from four bits to 60 bits.
- The word size is a relevant characteristic of any specific processor.

0000 0000	0010 1000
0000 0001	1110 1001
0000 0010	0001 0000
0000 0011	0100 0111
0000 0100	1100 0110
0000 0101	0100 1101
...	...
1111 1101	0000 0001
1111 1110	1111 0000
1111 1111	0100 1100

contents

1 Digital abstraction

2 Bits, bytes and words

3 Textual information

4 Machine-level instructions

5 Images

6 Audio

ASCII

- Text is the most common type of nonnumeric data that humans use, so computers need to represent it.
- In a computer, each alphanumeric character is represented by a bit pattern according to an established convention (code).
- The most commonly used character encoding standard is ASCII (American Standard Code for Information Interchange).
- The ASCII code is used in computers, telecommunications equipment, and other devices, and represents each character with a 7-bit string.
- An ASCII symbol is stored in a byte, with the leftmost bit usually set to 0.

ASCII

binary	char	binary	char	binary	char	binary	char
0000000	NUL	0100000	space	1000000	@	1100000	`
0000001	SOH	0100001	!	1000001	A	1100001	a
0000010	STX	0100010	"	1000010	B	1100010	b
0000011	ETX	0100011	#	1000011	C	1100011	c
0000100	EOT	0100100	$	1000100	D	1100100	d
0000101	ENQ	0100101	%	1000101	E	1100101	e
0000110	ACK	0100110	&	1000110	F	1100110	f
0000111	BEL	0100111	'	1000111	G	1100111	g
0001000	BS	0101000	(	1001000	H	1101000	h
0001001	HT	0101001	)	1001001	I	1101001	i
0001010	LF	0101010	*	1001010	J	1101010	j
0001011	VT	0101011	+	1001011	K	1101011	k
0001100	FF	0101100	,	1001100	L	1101100	l
0001101	CR	0101101	-	1001101	M	1101101	m
0001110	SO	0101110	.	1001110	N	1101110	n
0001111	SI	0101111	/	1001111	O	1101111	o
0010000	DLE	0110000	0	1010000	P	1110000	p
0010001	DC1	0110001	1	1010001	Q	1110001	q
0010010	DC2	0110010	2	1010010	R	1110010	r
...	...	...	...	...	...	...	...

ASCII

- A string is encoded in the C programming language by an array of ASCII characters, terminated by the null character

D	i	g	i	t	a	l	'\0'
01000100	01101001	01100111	01101001	01110100	01100001	01101100	00000000

- The set of symbols provided by ASCII is too short.
- Modern computers use Unicode, a standard for text expressed in most of the world's writing systems.
- As of March 2020, Unicode has a total of 143,859 characters.

contents

1 Digital abstraction

2 Bits, bytes and words

3 Textual information

4 Machine-level instructions

5 Images

6 Audio

format of instructions

- A computer program is a sequence of instructions.
- At the machine-level, each instruction is represented by a bit pattern.
- It consists of an opcode and some additional information, such as where operands come from and where to store the results.

opcode

opcode	arg

opcode	arg1	arg2

opcode	arg1	arg2	arg3

format of instructions

- On some machines, all instructions have the same length; on others there may be many different lengths.
- The opcode for each instruction type must be associated with a unique bit pattern, to identify it univocally.
- The instructions of the MIPS processor all have 32 bits.
- They are classified according to five types (R, I, J, FR, FI).

format of instructions

- The R instructions have all the data values located in registers.
- The syntax of the R instructions is: `OP rd, rs, rt`.
- Format of an R instruction:

opcode	rs	rt	rd	shamt	funct
6 bits	5 bits	5 bits	5 bits	5 bits	6 bits

- Example: `add $t1, $t2, $t3`

000000	01010	01011	01001	00000	100000

contents

1 Digital abstraction

2 Bits, bytes and words

3 Textual information

4 Machine-level instructions

5 Images

6 Audio

raster images

- A digital image can be represented by a grid of small points.
- This image is a raster image (or bitmap image).
- Each point is called a pixel and is represented by a binary pattern.
- Black and white images represent, for example, black by ‘1’ and white by ‘0’.
- This image is represented by:
  01101001 10011111 10011001.
- The size of the image needs to be stored and is part of the metadata.

0	1	1	0
1	0	0	1
1	0	0	1
1	1	1	1
1	0	0	1
1	0	0	1

colour depth

- Adding colours entails enlarging the number of bits that are used to represent each pixel.
- With two bits, four colours can be represented: 00 white, 01 blue, 10 green, and 11 red.
- This image is represented by: 00000000 00110000 11111100 01010100 01000100 10101010.
- The number of bits used to store each pixel is the colour depth.
- Images with more possible colours require more bits to specify each one, so they are stored in larger files.

00	00	00	00
00	11	00	00
11	11	11	00
01	01	01	00
01	00	01	00
10	10	10	10

image resolution

- The image quality depends on the image resolution, which is related to how close the pixels are.
- It is usually measured in dots per inch (dpi), i.e., the number of dots that can be placed in a line within the span of 1 inch.
- In a low-resolution image, the pixels are larger so fewer are needed to fill the space.

contents

1 Digital abstraction

2 Bits, bytes and words

3 Textual information

4 Machine-level instructions

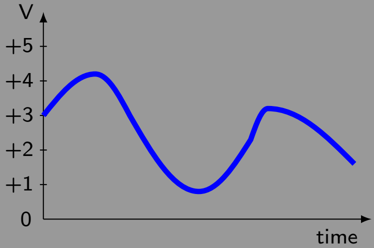
5 Images

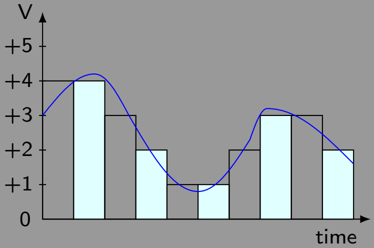
6 Audio

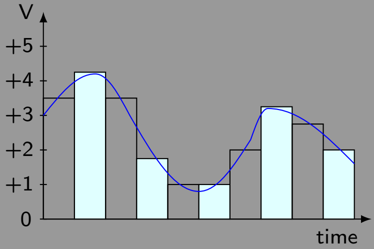
sampling

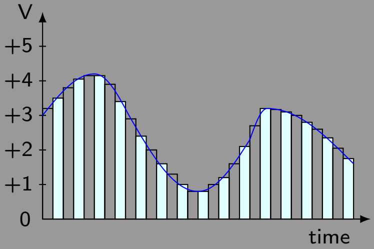
- To process sound (or audio), computers need to convert it into a digital format.
- Sound is recorded using a microphone that translates sound waves into an electrical signal.
- Periodic measurements of the level of that signal are registered.
- The process that reduces a continuous-time signal to a discrete-time signal is called sampling.
- The value at a point in time is designated sample.
- The samples are then simply converted into binary, using a unique binary code.
- Afterwards, the digital sound can be processed by a computer as a sequence of bits.

different samples of the same signal









sampling rate

- The number of samples taken per second, measured in Hertz (Hz), is the sampling rate.
- The higher the sampling rate, the better the quality of the audio digital signal.
- Size s of a sound signal in bits: $s = f \times r \times t$.
  - f: sampling rate (Hertz)
  - r: sample resolution (number of bits)
  - t: duration of the signal (seconds)