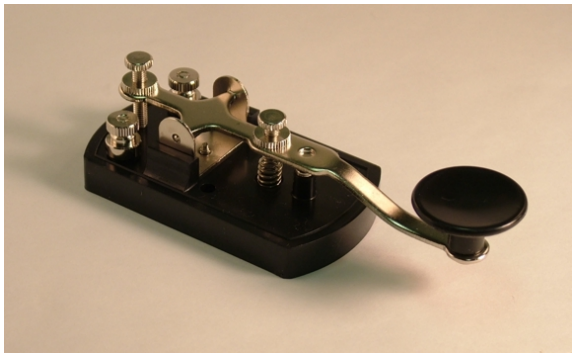
Maths is Coded in Your Genes



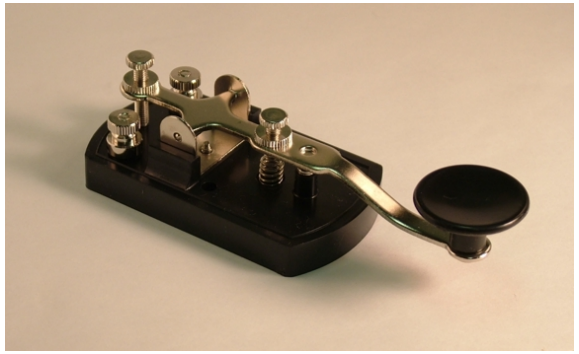
Chris Budd

GRESHAM COLLEGE

IMI

UNIVERSITY OF BATH

We live in a world full of information

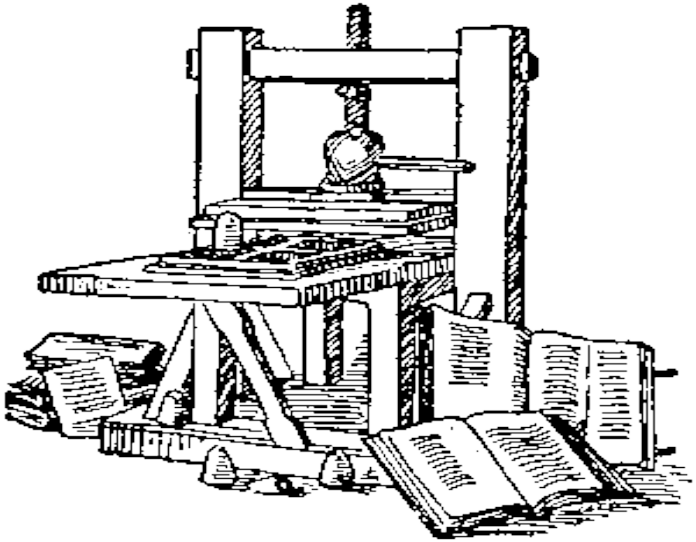


It is essential that this information is sent accurately

Humans: Language. Can be prone to error

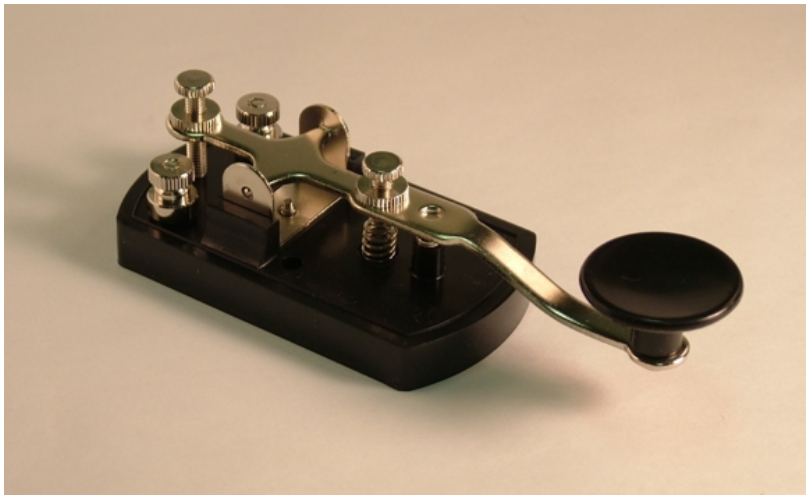
Invention of writing and printing

Cuneiform tablet



Printing press 15th Century

Modern communication is mostly done using electronic means



Signals and information

Information is conveyed by a signal from a transmitter to a receiver through a channel

The channel typically distorts the signal

Can we still recover the information?

Signal may be conveyed over many paths, all of which introduce noise and distortion

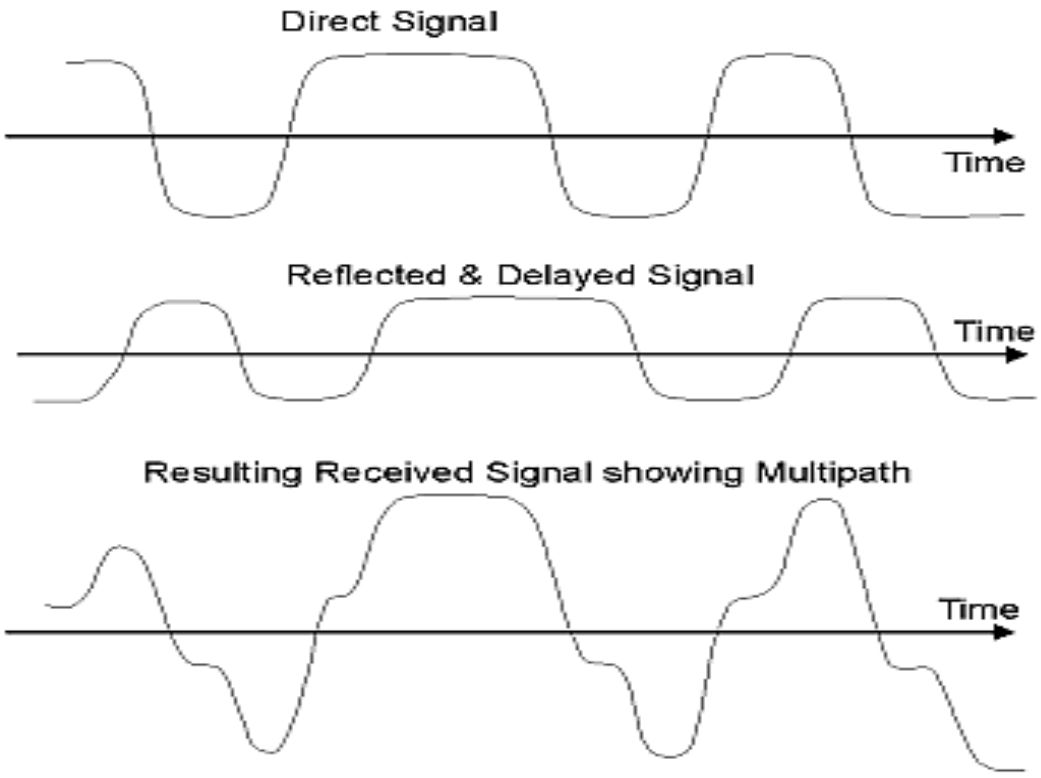
u(t)

Alice

v(t)

Bob

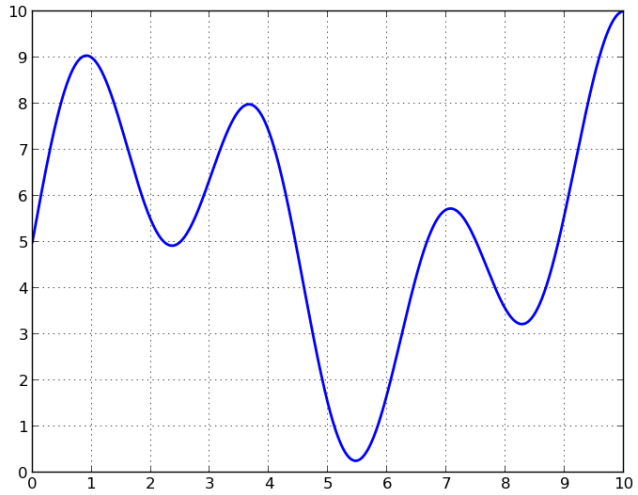
$$v(t) = \sum_{i=1}^{N} f_i(t) * u(t) + e(t)$$



Analogue vs. Digital Communication

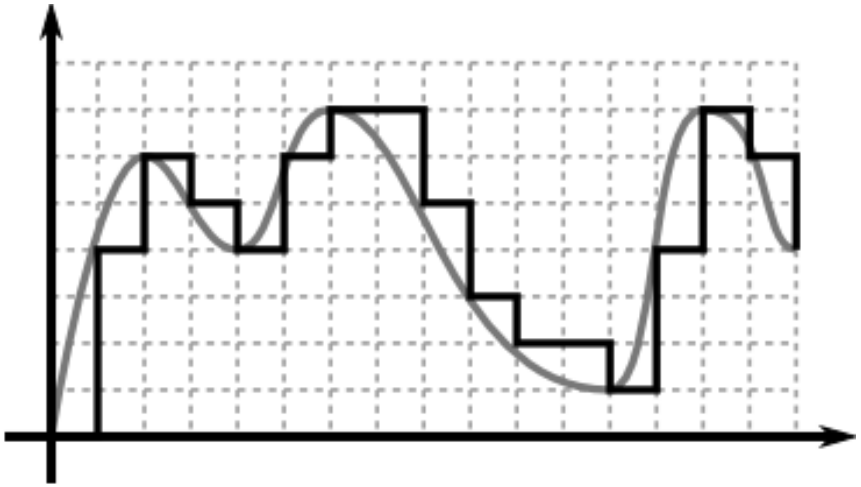
Analogue: Used in speech, records and early TV/Radio.

Information is transmitted as a continuous range of values



Digital Communication

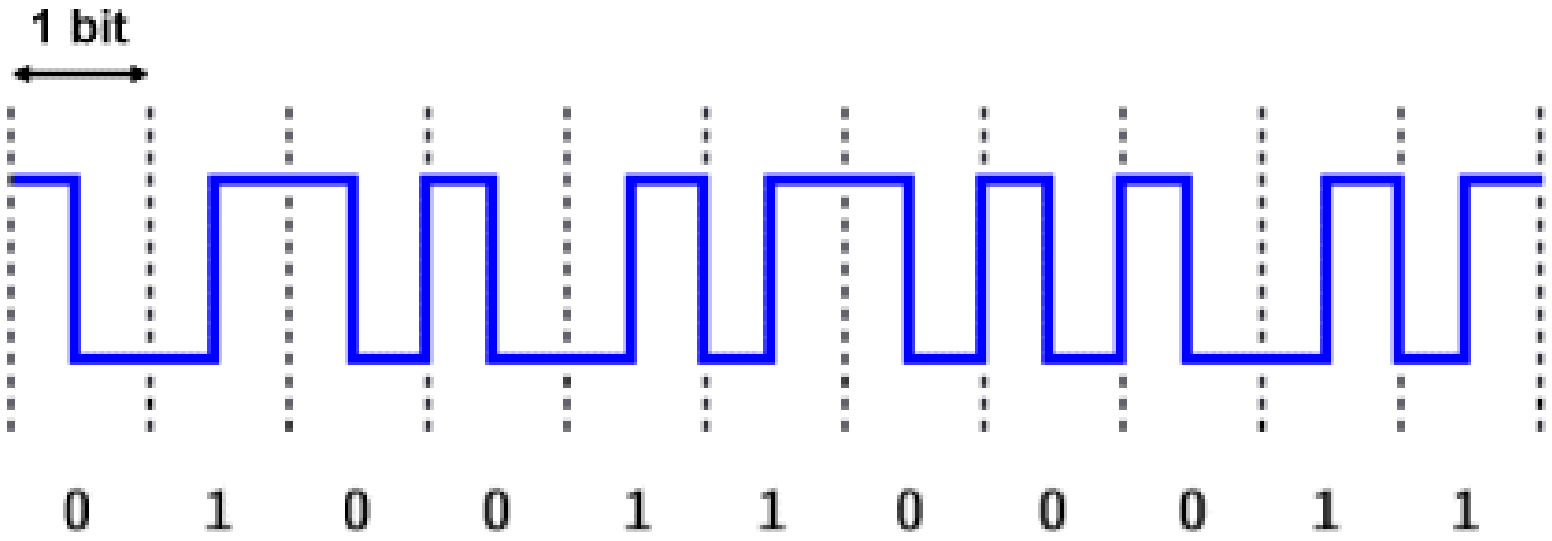
Digital: Used in computers, TV, Radio, Mobile phones, shirt sizes, genetics



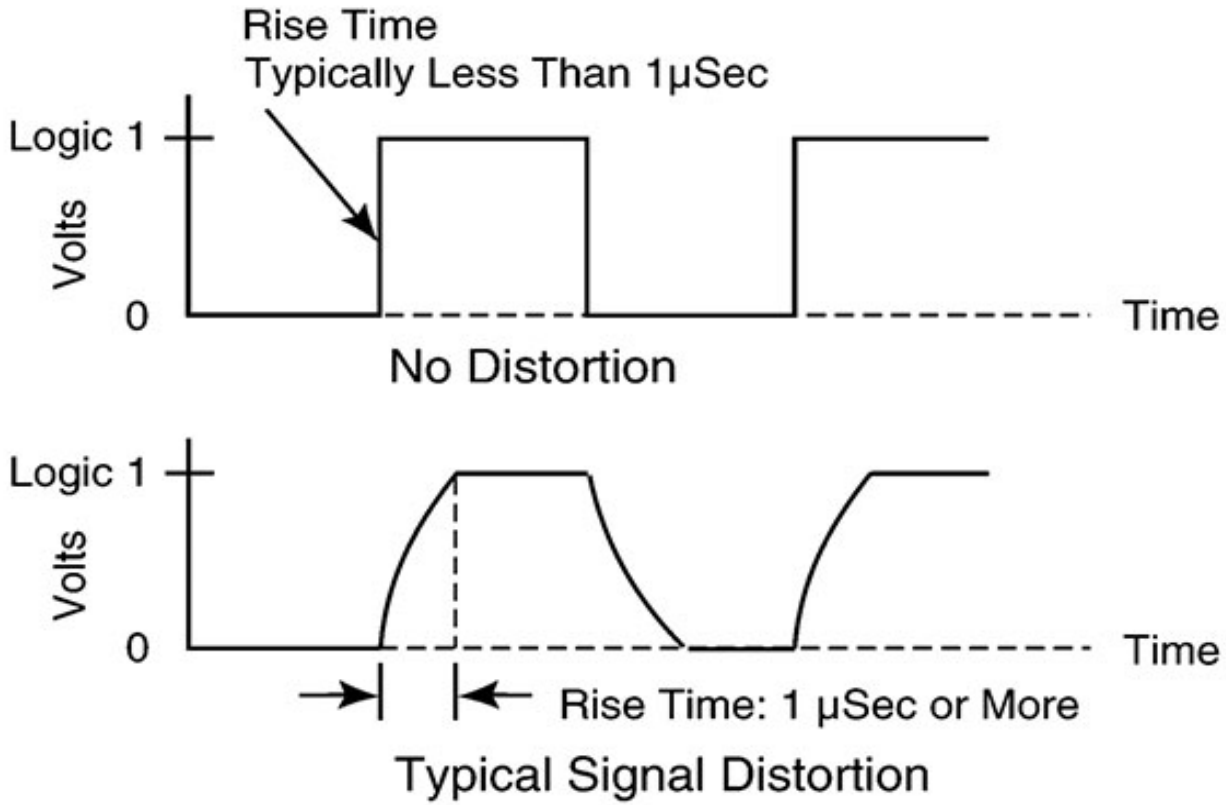
Information is transmitted in discrete amounts

All modern technology uses digitally stored information

Digital information is transmitted and stored as a series of pulses or bits

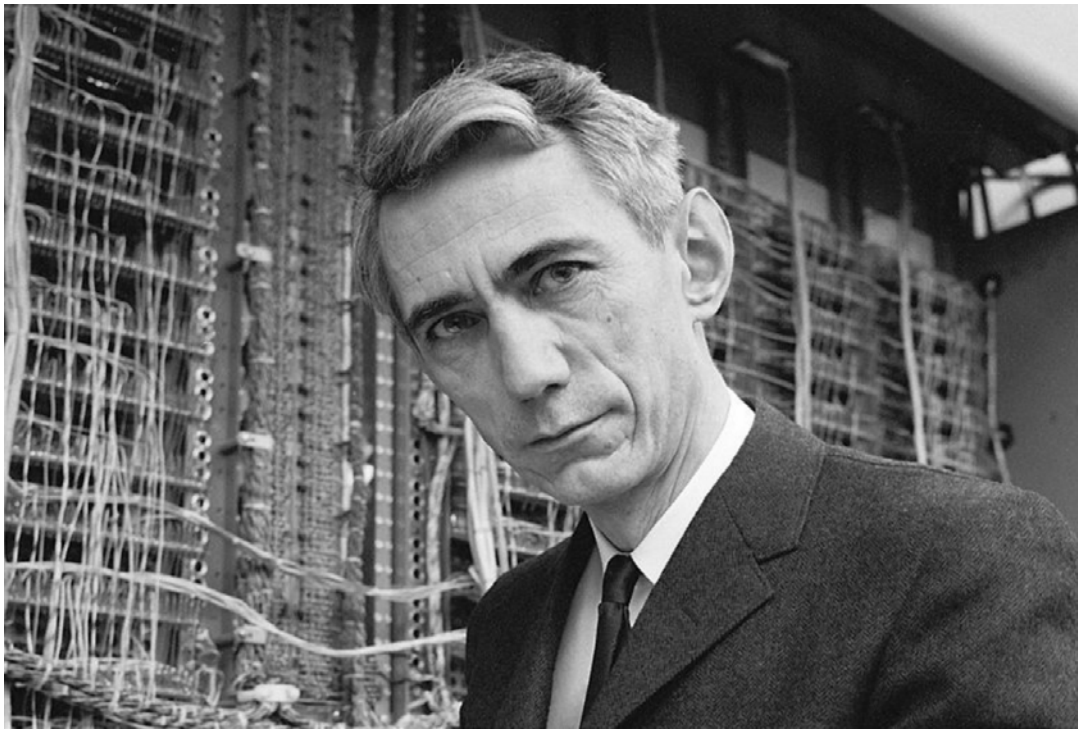


Much less susceptible to distortion



Still recognisable as a single pulse

But errors can still occur! A bit can change from a 1 to a 0 with a probability p



Claude Shannon: Father of information theory

C.E. Shannon, A mathematical theory of communication, Bell System Tech J., 27, (1948), 379-423.

“The fundamental problem of communication is that of reproducing at one point, either exactly or approximately, a message selected at another point."

Q. How much information can be conveyed over a noisy channel?

Shannon’s theorem expressed the maximum information rate at which reliable error free communication is possible using a carefully designed code over a noisy channel.

Its impact has been crucial to the success of the Voyager missions to deep space, the invention of the compact disc, the feasibility of mobile phones, the development of the Internet, the study of linguistics and of human perception, the understanding of black holes, and numerous other fields. Wikipedia

How is digital communication done in practice?

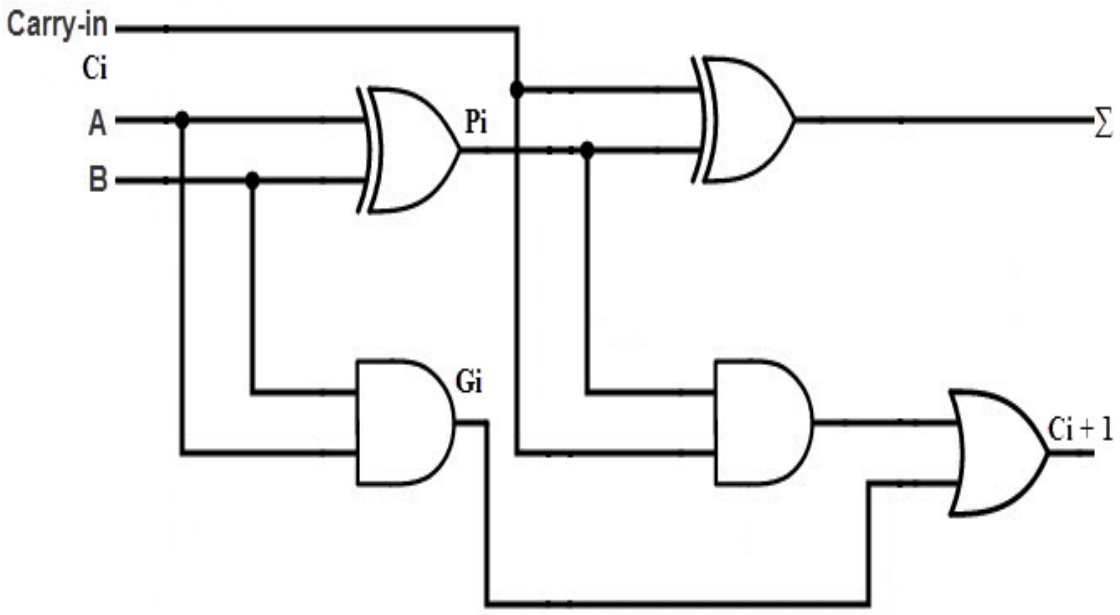
Binary numbers: Leibnitz 1703

0 0 0 0	0	1 0 0 0	8
0 0 0 1	1	1 0 0 1	9
0 0 1 0	2	1 0 1 0	10
0 0 1 1	3	1 0 1 1	11
0 1 0 0	4	1 1 0 0	12
0 1 0 1	5	1 1 0 1	13
0 1 1 0	6	1 1 1 0	14
0 1 1 1	7	1 1 1 1	15

Now used widely for computations and to convey information

George Boole 1854: Boolean algebra

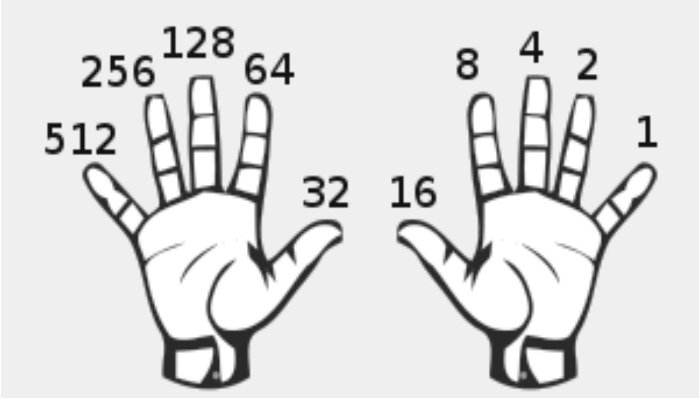
Shannon 1937: Digital communication using relays



Digital binary adder

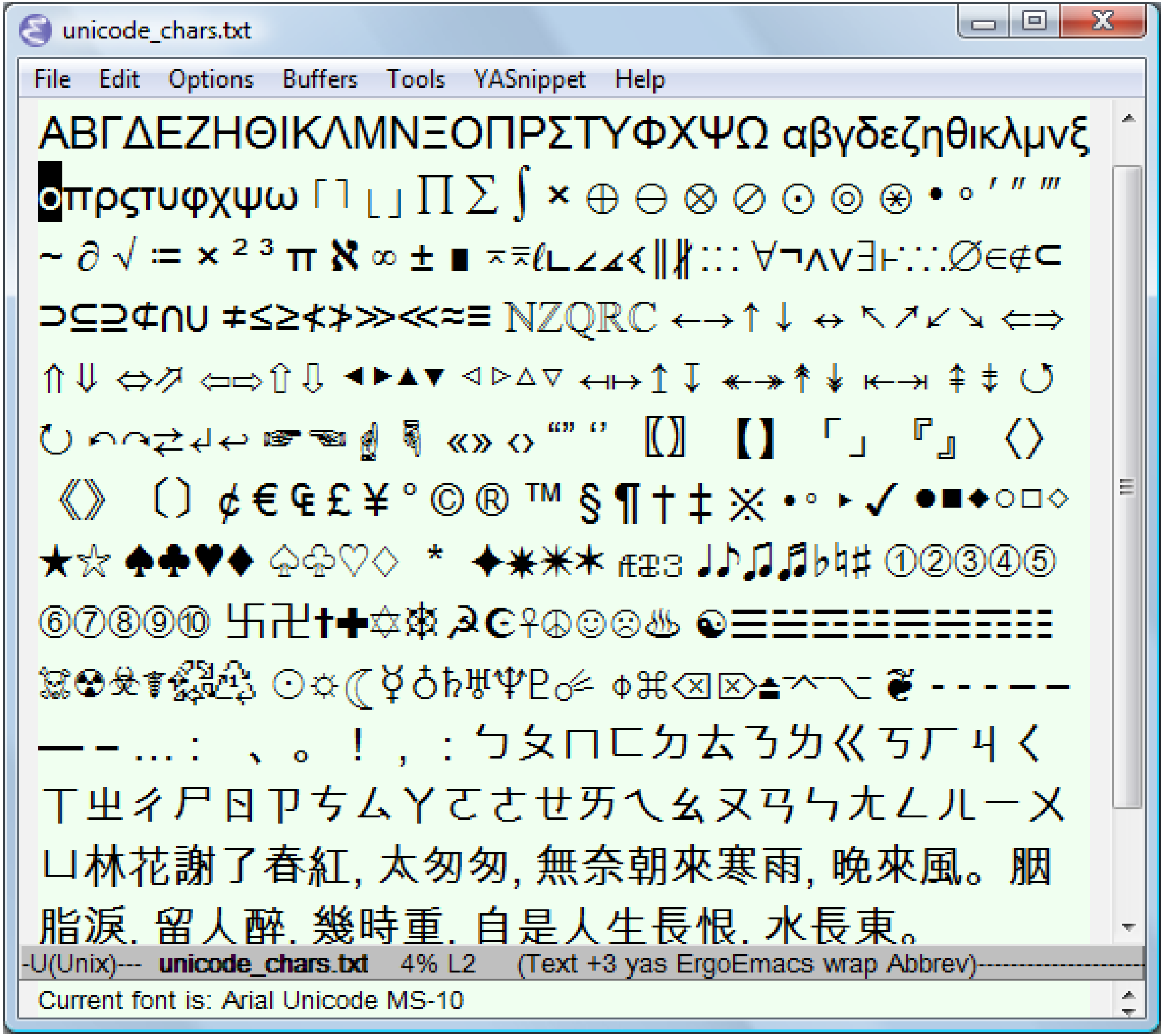
How does a monster count to 30?

On their fingers!



ASCII Code: Character to Binary

0	0011 0000	O	0100 1111	m	0110 1101
1	0011 0001	P	0101 0000	n	0110 1110
2	0011 0010	Q	0101 0001	o	0110 1111
3	0011 0011	R	0101 0010	p	0111 0000
4	0011 0100	S	0101 0011	q	0111 0001
5	0011 0101	T	0101 0100	r	0111 0010
6	0011 0110	U	0101 0101	s	0111 0011
7	0011 0111	V	0101 0110	t	0111 0100
8	0011 1000	W	0101 0111	u	0111 0101
9	0011 1001	X	0101 1000	v	0111 0110
A	0100 0001	Y	0101 1001	w	0111 0111
B	0100 0010	Z	0101 1010	x	0111 1000
C	0100 0011	a	0110 0001	y	0111 1001
D	0100 0100	b	0110 0010	z	0111 1010
E	0100 0101	c	0110 0011	.	0010 1110
F	0100 0110	d	0110 0100	,	0010 0111
G	0100 0111	e	0110 0101	:	0011 1010
H	0100 1000	f	0110 0110	;	0011 1011
I	0100 1001	g	0110 0111	?	0011 1111
J	0100 1010	h	0110 1000	!	0010 0001
K	0100 1011	I	0110 1001	'	0010 1100
L	0100 1100	j	0110 1010	"	0010 0010
M	0100 1101	k	0110 1011	(	0010 1000
N	0100 1110	l	0110 1100	)	0010 1001
				space	0010 0000


unicode_chars.txt
File Edit Options Buffers Tools YASnippet Help
ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΤΥΦΧΨΩ αβγδεζηθικλμνξ
οπρςτυφχψω
ㄅㄆㄇㄈㄉㄊㄋㄌㄍㄎㄏㄐㄑ
ㄒㄓㄔㄕㄖㄗㄘㄙㄚㄛㄜㄝㄞㄟㄠㄡㄢㄣㄤㄥㄦㄧㄨ
ㄩ林花謝了春紅，太匆匆，無奈朝來寒雨，晚來風。胭
脂淚，留人醉，幾時重，自是人生長恨，水長東。
-U(Unix)--- unicode_chars.txt 4% L2 (Text +3 yas ErgoEmacs wrap Abbrev)
Current font is: Arial Unicode MS-10

Need to check messages to see if errors have occurred

Various checking methods were used by early scribes

0	0	0	0	0
1	0	0	1	1
2	0	1	0	1
3	0	1	1	0
4	1	0	0	1
5	1	0	1	0
6	1	1	0	0
7	1	1	1	1

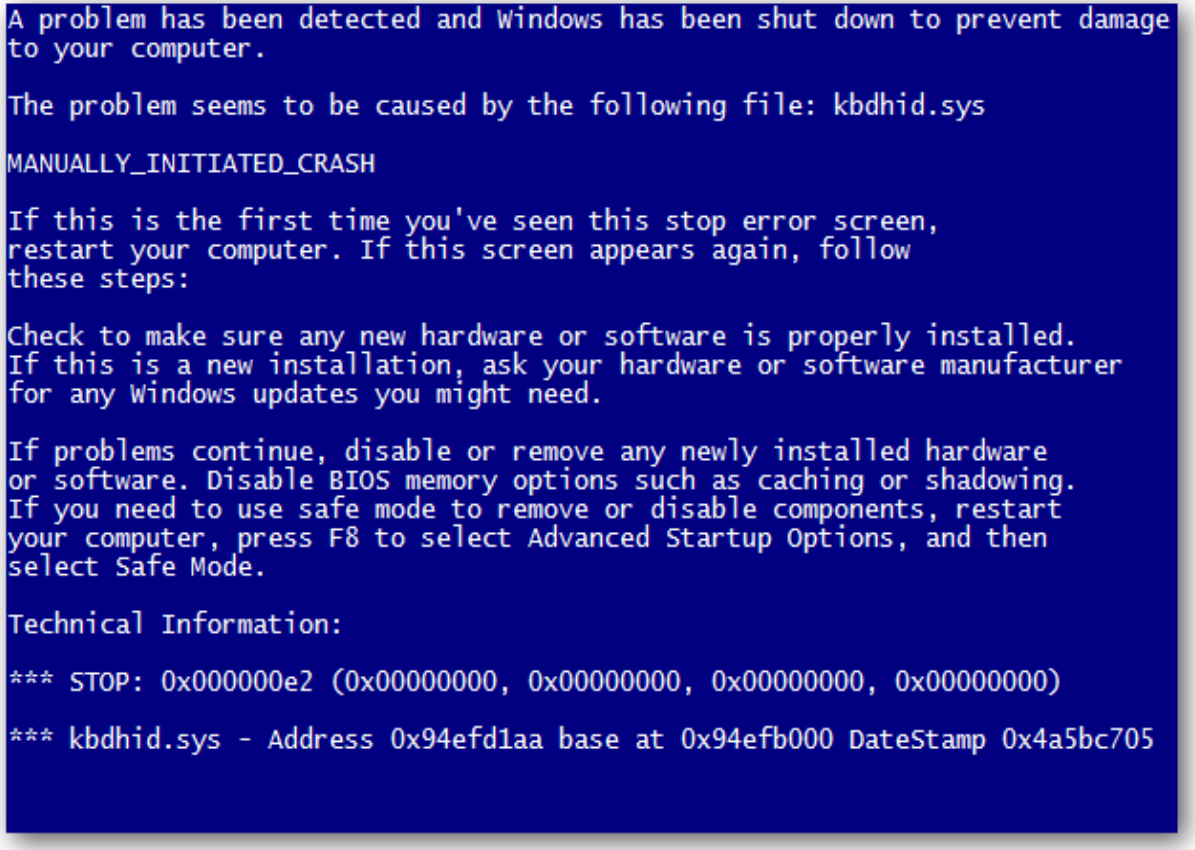
If all correct there are an: even number of 1s

If one mistake there is an: odd number of 1s

Luhn algorithm : used on credit cards and bar codes

Starting from the right of the credit card number, the digits are alternatively doubled. All of the digits are then added up, and the credit card number passes if the resulting sum is divisible by 10. This process checks both the digits and also the order of the digits in the credit card.

Suppose that you detect an error?



1. Stop Blue Screen of Death!!

2. Repeat the message

Automatic Repeat Request (ARQ)

Widely used on the Internet

And in Bar Codes:

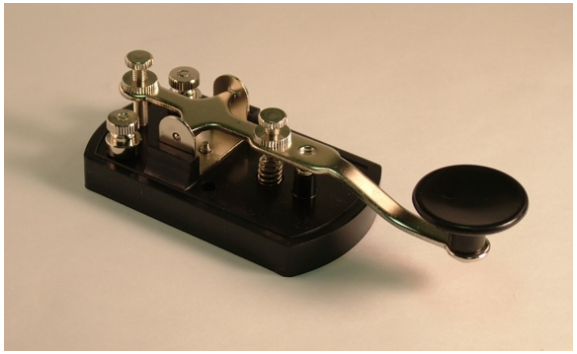
3. Correct the error

Error Correction: The science of sending a message in such a way that it can be decoded even if there is a lot of noise

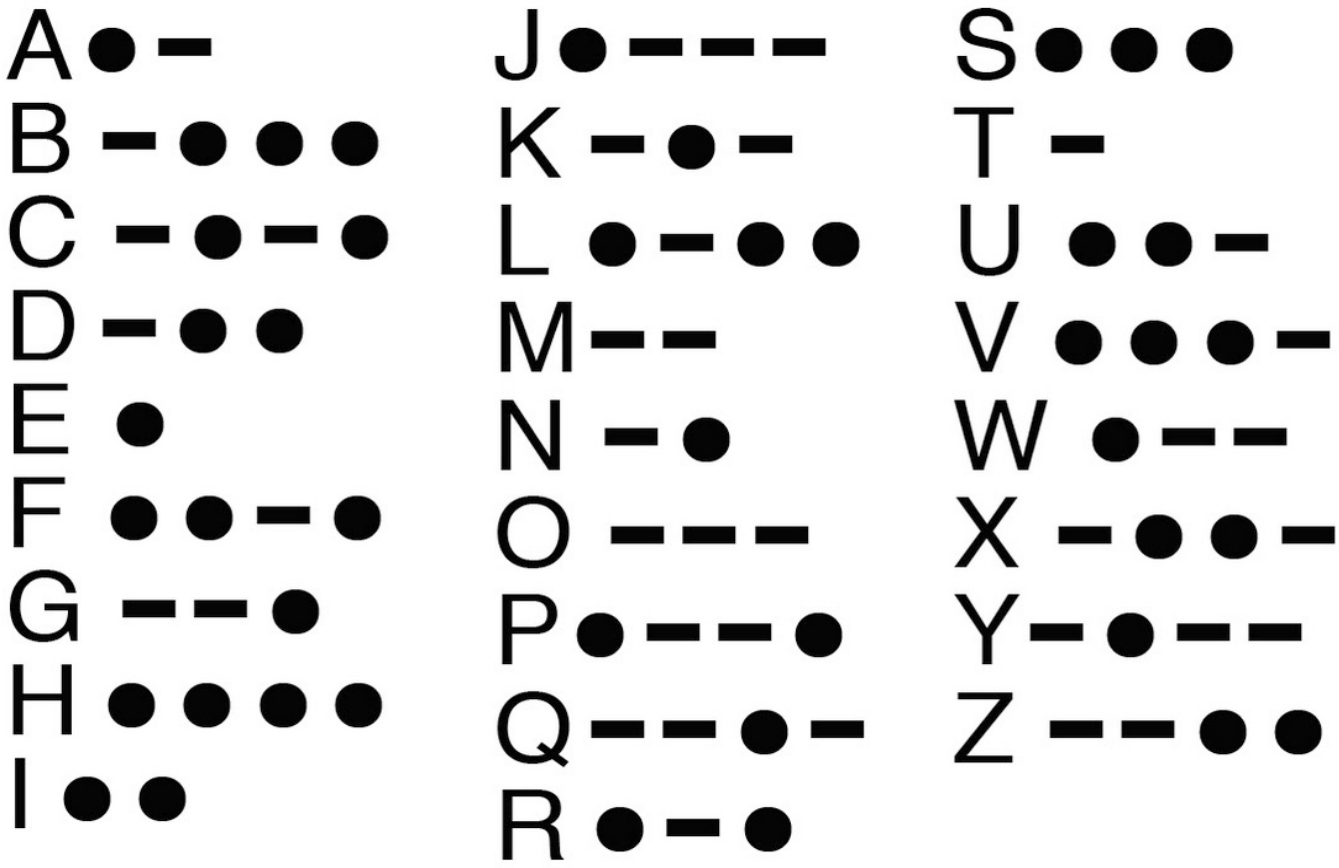
Example Telegraphy is more reliable than Telephony!

Idea:

Convert letters into Morse code



Make the codes as different from each other as possible so that the letters can be distinguished

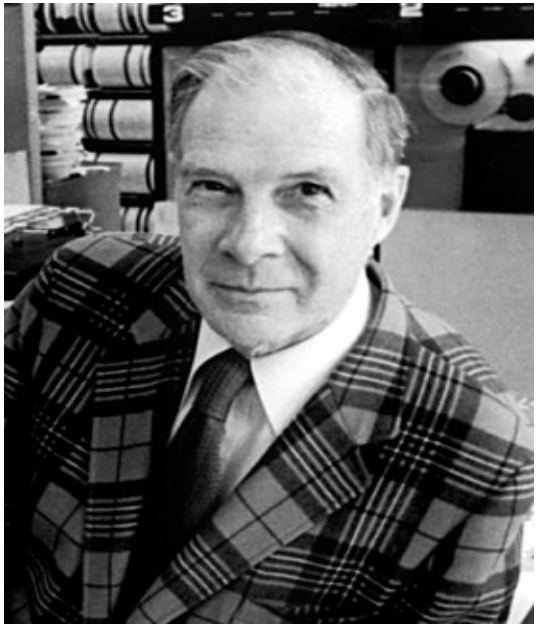

A ● −
B − ● ● ●
C − ● − ●
D − ● ●
E ●
F ● ● − ●
G − − ●
H ● ● ● ●
I ● ●
J ● − − −
K − ● −
L ● − ● ●
M − −
N − ●
O − − −
P ● − − ●
Q − − ● −
R ● − ●
S ● ● ●
T −
U ● ● −
V ● ● ● −
W ● − −
X − ● ● −
Y − ● − −
Z − − ● ●

IDEA In a general message make the code symbols for the characters as different as possible so that they can be distinguished even if they are corrupted by noise.

Error correction does automatically what we do when we read Morse

Hamming Distance measures how different two binary code symbols are

Hamming distance: number of differences of 1s and 0s



Hamming 1947

1 1 0 1 1 0	Original
1 1 1 1 1 0	Hamming distance of 1
1 0 1 1 1 0	Hamming distance of 2

Simple code example

0	000 000
1	001 110
2	010 011
3	011 101
4	100 101
5	101 011
6	110 110
7	111 000

Original Binary

Check digits

0	000 000
1	001 110
2	010 011
3	011 101
4	100 101
5	101 011
6	110 110
7	111 000

All are a Hamming distance of 3 or more apart

Receive 0 0 1 1 0 1

Know there is a one bit error.

Which was the original?

	Code	Received	Hamming Distance
0	000 000	001 101	3
1	001 110	001 101	2
2	010 011	001 101	3
3	011 101	001 101	1
4	100 101	001 101	2
5	101 011	001 101	3
6	110 110	001 101	5
7	111 000	001 101	4

Closest symbol is 3. So correct received symbol to 3
With error in the second digit.

Things to consider when constructing an error correcting code

1. Must get through the noise

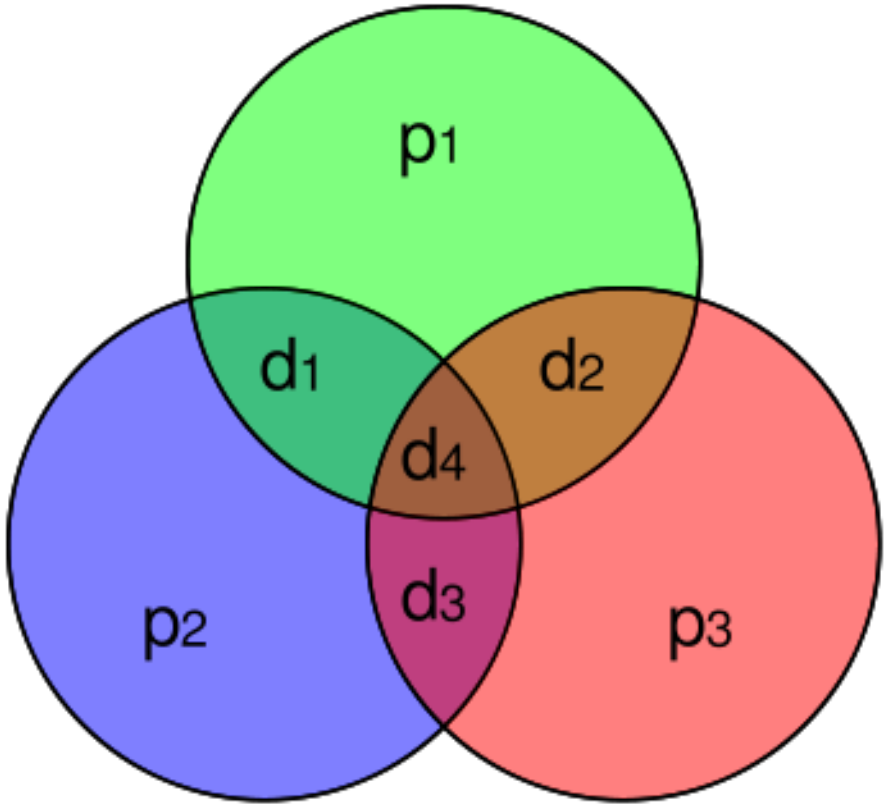
2. Must allow rapid transmission

3. Must be easy to find errors and decode

Evariste Galois

Examples of Error correcting codes.

- Reed-Solomon code 1960 Polynomial over finite fields
- Hamming (7,4) code 1950: Uses three parity digits



Original message x = (d1,d2,d3,d4)

Transmitted code y = (p1,p2,d1,p3,d2,d3,d4)

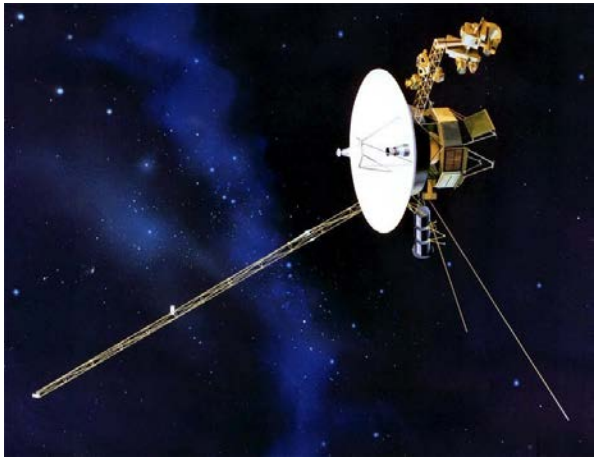
Linear code y = G x

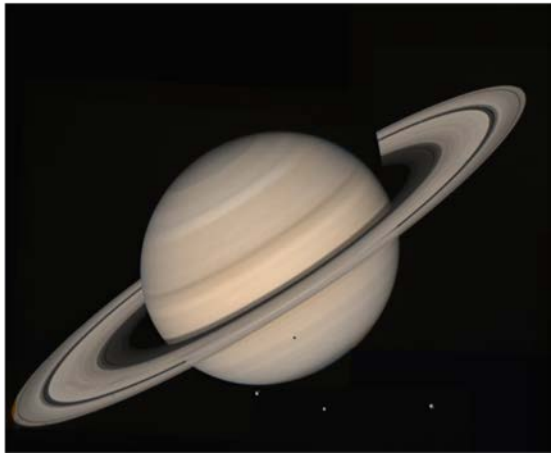
Received message z

Error digit given by d = H z

$$\mathbf{G} := \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{H} := \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Examples of use of the Reed-Solomon Code





Data Compression

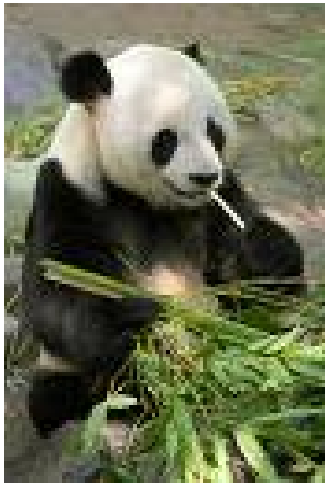
For example

Instead of sending this message which has lots of vowels in it which we don't really need

W cn snd ths mssg nstd whch ds nt hv ny vwls t ll

R vn ths n

Camera takes picture made up of PIXELS



8 BITS per pixel ….. 256 range of intensity

3 000 000 Pixels per Picture

Total 3M Byte per picture

One bite

JPEG (Joint Photographic Experts Group) compresses the picture by ignoring high frequencies beyond the visual range.

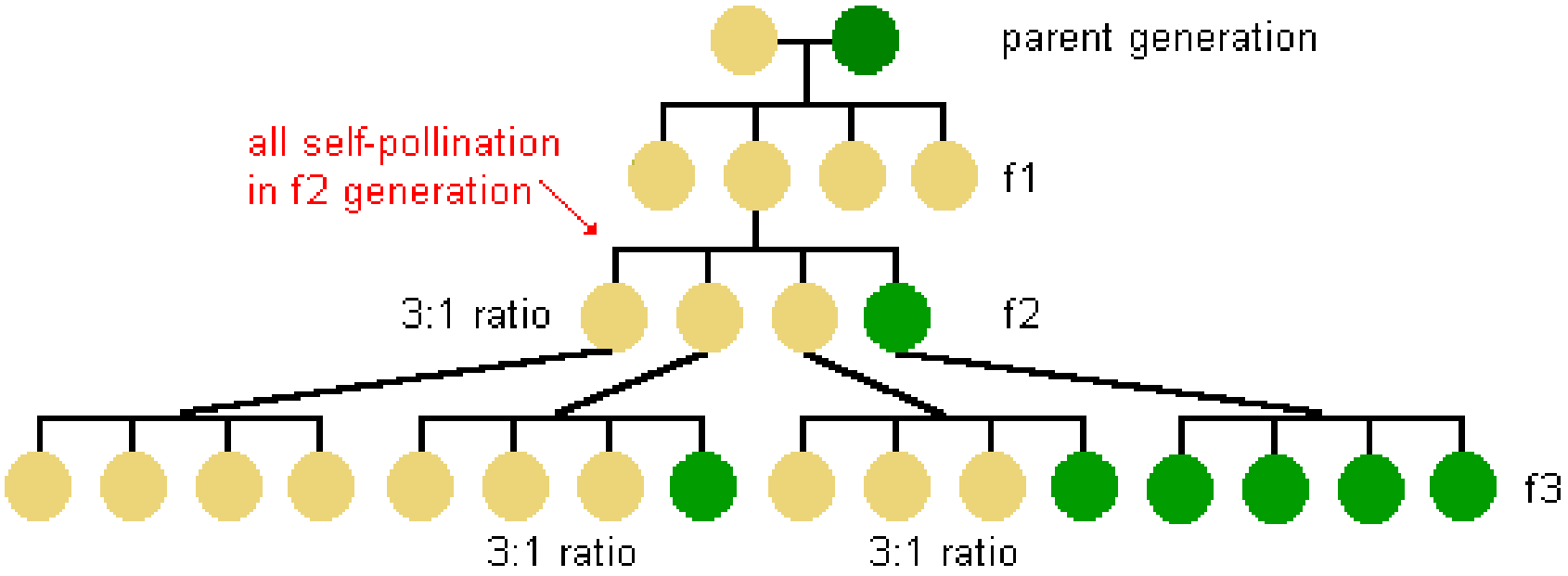
Achieved by using the Discrete Cosine Transform

Genetics: Nature's digital code

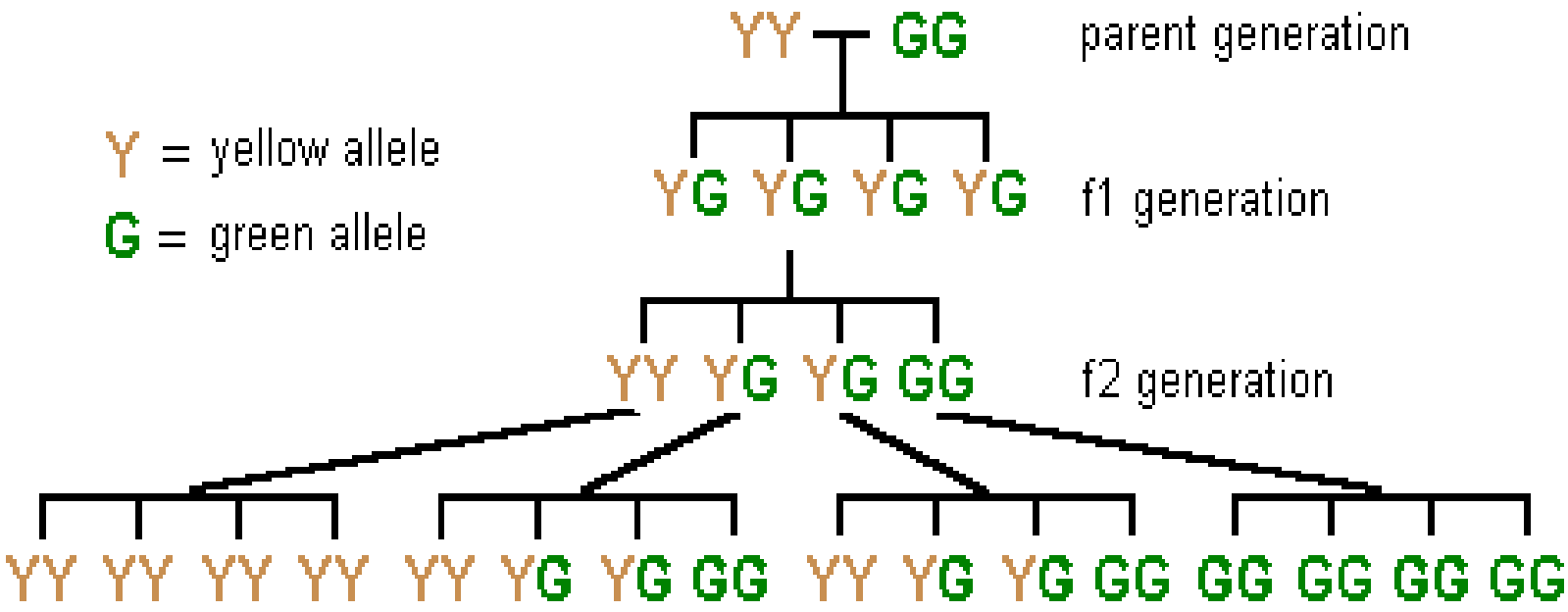
A huge amount of information has to be passed from one generation to the next for an organism to function

Mendel Studied the inheritance of 8000 pea plants

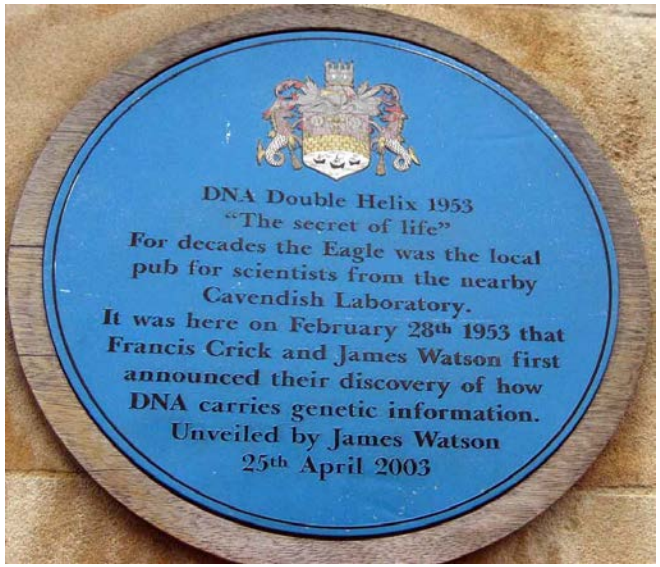
Observed a discrete pattern of inheritance with a strong mathematical structure

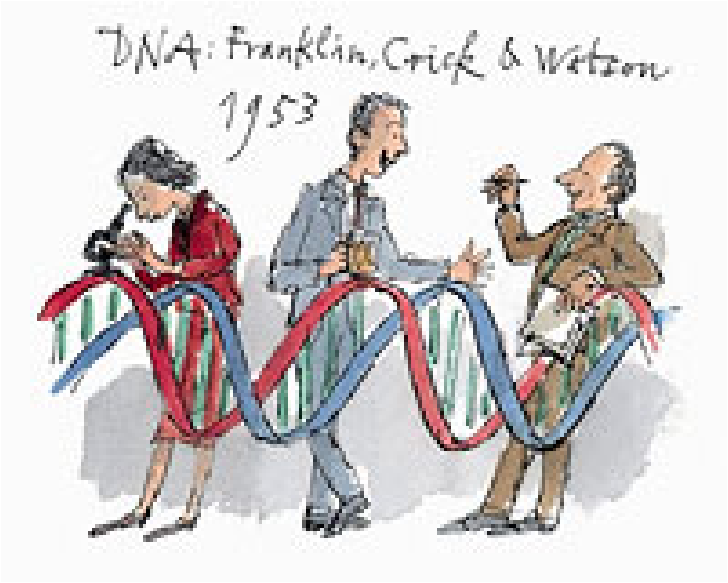


Deduced that this was due to dominant/recessive alleles defining inherited characteristics



Franklin, Watson and Crick: Identified the mechanism via the structure of the DNA molecule





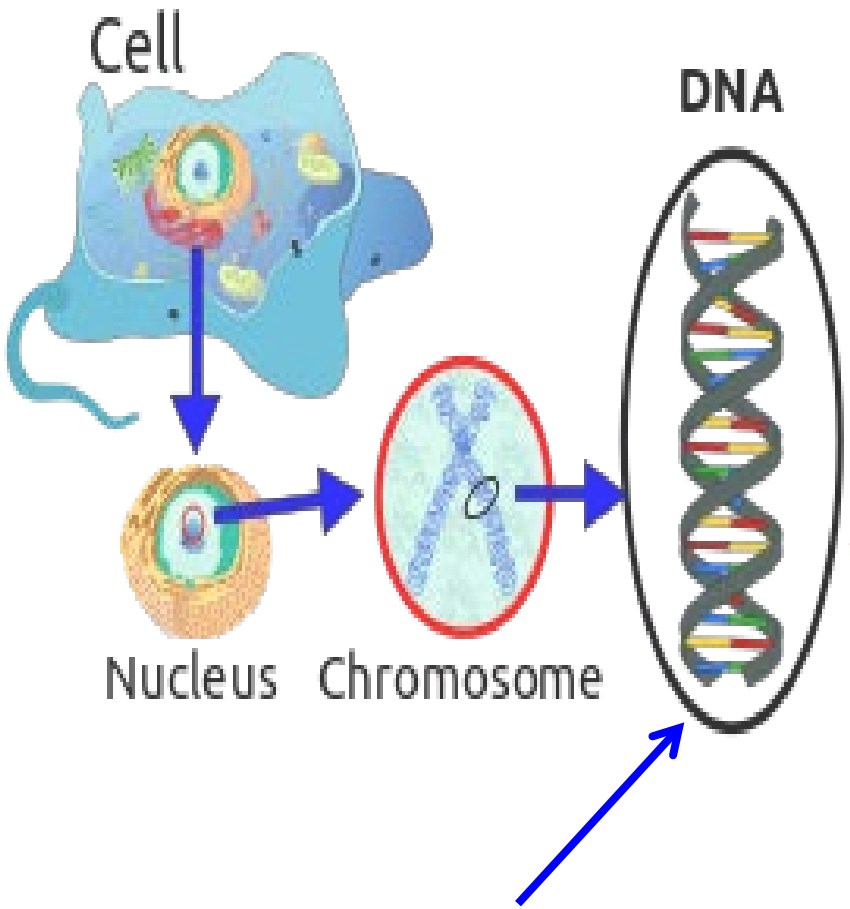
A gene itself is a sequence of the DNA or RNA molecules which gives the code for a molecule such as a protein

A chromosome is a long strand of DNA containing many genes.

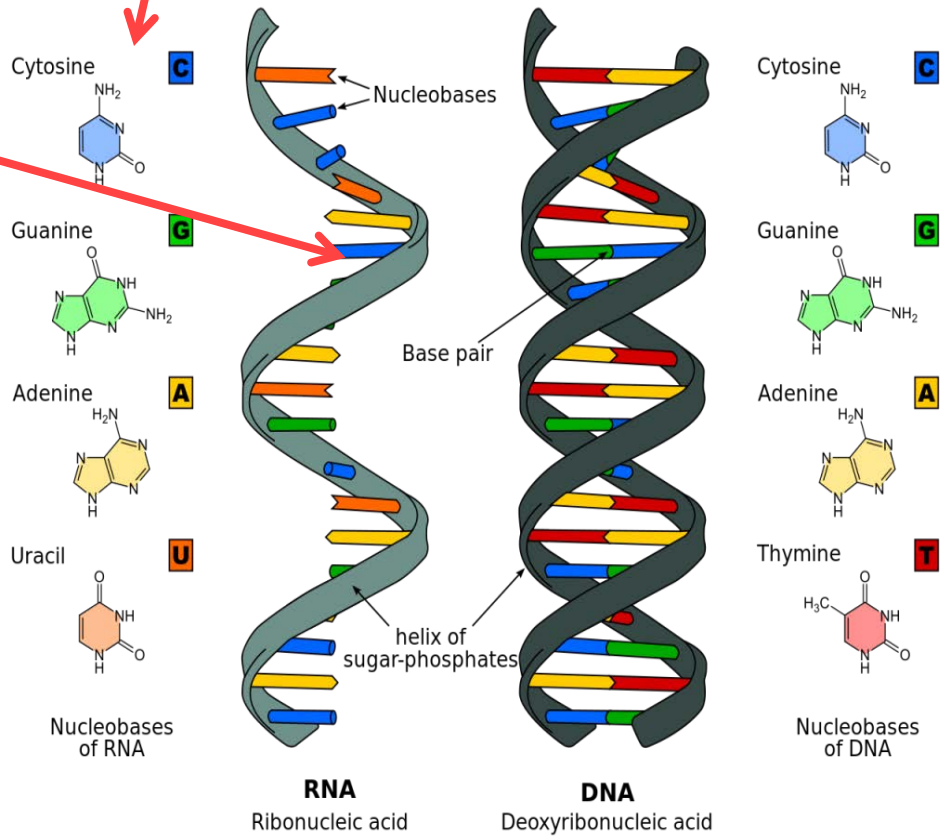
A human chromosome can have up to 500 million base pairs of DNA with thousands of genes.

The sequence of genes is called the geneotype of the individual

The transmission of genes to that individuals offspring is the basis of the inheritance of its phenotypical traits



C G A U elements in RNA

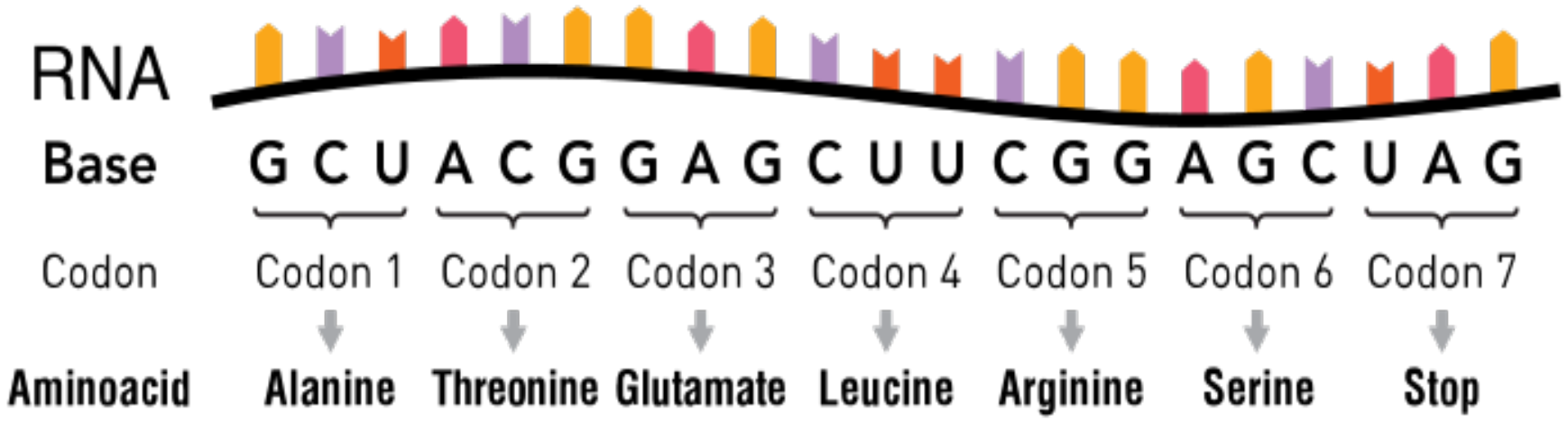


Huge mathematical regularity!

A gene codes information for amino acids through a sequence of codons

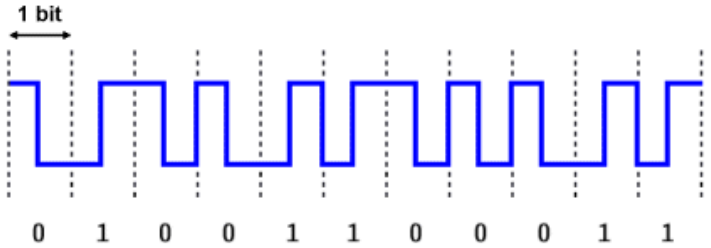
Codons are made up of combinations of 3 units of A C G U

These codons then code 20 Amino acids



Inheritance comes from one generation passing on its genes to the next

The codons play a very similar role in this to the digital codes we considered earlier



Eg. A C G or A C U are discrete codes

4*4*4 = 64 Different codes 20 Amino acids

So have a degree of error correction

Very effective process

1/100 000 chance of error

Most errors are bad or neutral

But some are good,
and lead to evolution

So it’s good that nature’s
error correction is not quite perfect