HOW DO WE PROTECT INFORMATION?

claude shannon
1916-2016

many thanks to
Erdal Arikan

design
alesiom.ch

«Redundancy» protects information against errors. Natural languages have built-in redundancy. Communication systems use «error correcting codes» to insert redundancy.

REDUNDANCY

Redundancy in Language

Suppose the sentence «I know him» is mistyped as «I knoe him». This error can be detected because «knoe» is not a valid word in English. Most people will have no difficulty in not only detecting but correcting the error since «know» is the most likely word in the present context to be mistyped as «knoe».
If «know» is mistyped in the same sentence as «snow», it is still likely that this error will be detected and corrected from the context, although «snow» is a valid word. Detection and correction may be harder or impossible if «know» is mistyped as «knew». As these examples show, redundancy is the key to recovery from errors.

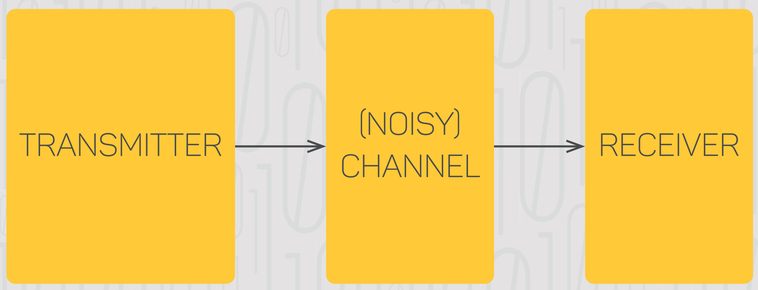
Repetition Coding

Redundancy through repetition coding was used since ancient times, mostly for strategic purposes. It is the simplest method of inserting redundancy into a message: **repeat it**. For example, in a triple repetition code for sending binary data, 0 and 1 are «encoded» as 000 and 111 respectively. A data block such as 010 is encoded into the code block 000 111 000. If the code block is corrupted by noise and received as 001 100 1111, we can detect that there are errors in the first two symbols. The first error can be corrected by «majority voting», the second word is «misdecoded» as 0, and the error in the third word goes undetected.

Hamming Code

DATA D1, D2, D3, D4	**REDUNDANCY** P1, P2, P3
0000	000
0001	011
0010	010
0011	101
0100	111
0101	100
0110	**001**
0111	010
1000	101
1001	110
1010	011
1011	000
1100	010
1101	001
1110	100
1111	111

A triple-repetition code requires a large overhead for the ability to correct single errors. A (7,4) Hamming code that encodes 4 bits of data into codewords of 7 bits, as shown in the table on the left, can correct single errors more efficiently. This code encodes, for instance, the data block 0110 into the codeword 0110001. All other codewords differ from 0110001 in **at least three positions**, which ensures that any single error can be corrected.



SHANNON CAPACITY

What is the largest rate (or minimum amount of redundancy, maximum amount of information) that enables to reliably communicate over a given channel?

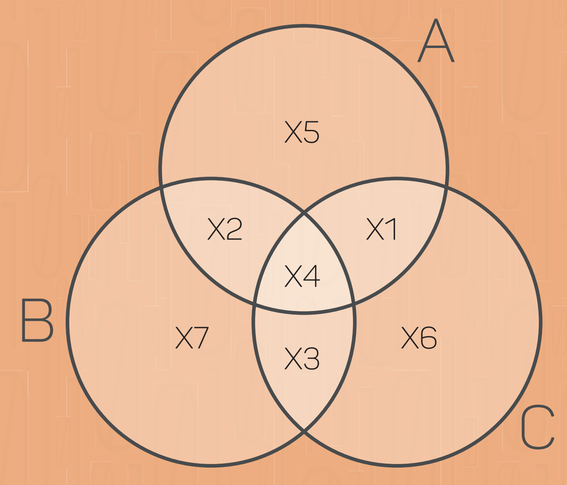
Shannon's capacity theorem answers this question for a broad class of channels and sets ultimate limits on efficiency and performance.
Communication engineers have developed a variety of codes, such as LDPC and turbo codes, that achieve the efficiency limits set by Shannon theory. However, the field of error-correcting codes is still a vibrant research area, and is likely to remain so for many decades spurred by emerging applications.

PARITY-CHECK CODES

In a parity-check code, each redundant bit is computed as a modulo-2 sum of some subset of data bits. The (7,4) Hamming code is an example of a parity-check code. Most codes used in today's communication systems are parity-check codes.

Example

Let us decode a message with the Hamming Code and a visualization introduced by McEliece. It is actually possible to retrieve a lost bit by using Venn diagram and a simple rule. Take the following figure:



The circle A encompasses x1, x2, x4 and x5.
The circle B encompasses x2, x3, x4 and x7.
Finally, the circle C encompasses x1, x3, x4 and x6.
Consider that we get a new codeword of seven digits as explained above : 0,1,0,1,0,1,0. Transpose them into (x1, x2, x3, x4, x5, x6, x7). One observation that you can do is that the sum of all X's in A, B and C are pair. This is actually a sufficient constraint that can help us to decode codewords with missing digits.

i) x1 + x2 + x4 + x5 must be pair
ii) x2 + x3 + x4 + x7 must be pair
iii) x1 + x3 + x4 + x6 must be pair

If you get a codeword of the form (0, ?, ?, 1, 0, ?, 0), you can recover it by first using rule i) to discover the second digit, than ii) to discover the third digit and finally iii) to discover the 6th digit.

Notice that recovering 3 digits will not always work (e.g. ?,?,0,?,0,1,0), but recovering one digit is always possible with this method.

STATE-OF-THE-ART IN CODING

Today, error-correcting codes are used everywhere from wireless communications to mass storage devices to fiber-optic links. Some major classes of error-correcting codes and their application areas are given in the table below.

FUTURE CHALLENGES

- How to know the capacity of even small networks
- Need to design codes for new applications
- How to jointly design how to process and encode information over networks

TYPE OF CODE	REED-SOLOMON	LOW-DENSITY PARITY-CHECK (LDPC)	TURBO
APPLICATIONS	 DATA STORAGE (CD/DVD)	 WIFI, BROADCASTING	 CELLULAR (3G, 4G), SATELLITE COMMUNICATIONS