code

Canonical name	Code
Date of creation	2013-03-22 14:21:21
Last modified on	2013-03-22 14:21:21
Owner	mathcam (2727)
Last modified by	mathcam (2727)
Numerical id	9
Author	mathcam (2727)
Entry type	Definition
Classification	msc 68P05
Classification	msc 68P30
Defines	code
Defines	block length
Defines	minimum distance

Let A be an alphabet. A *code over A* is any subset C of the set of words A^* on the alphabet A such that C has “uniquue factorization into letters,” i.e., such that for whenever $a_1 \dots a_n = b_1 \dots b_m$, with all $a_i, b_j \in C$, then we have $n = m$ and $a_i = b_i$ for all i. In other words, every “word over A” generated by C (considered as an alphabet) can be uniquely factored into “letters” in C.

An example of a subset of A^* which is *not* a code is given by $C = \{ab, c, a, bc\}$. Here the word abc can be written either as $(ab)c$ or as $a(bc)$ in terms of elements of C. Since $ab \neq a$ nor $c \neq bc$, C is not a code.

If we fix a length n for the words, i.e. we require that $C \subset A^n$, then we call C a *block code*, and call n the *block length* of the code. An important property of a code is the code's *minimum distance*, given by the minimum Hamming distance between any pair of words in C.

This notion of code is obviously very general. In practice (i.e., in coding theory) one typically takes codes with a little more structure. See, in particular, linear codes.