Some properties of Fix – Free Codes

R.Ahlswede
Fakultät für Mathematik,
Universität Bielefeld,
Postfach 100131,
33501 Bielefeld,
Germany

B.Balkenhol*
Fakultät für Mathematik,
Universität Bielefeld,
Postfach 100131,
33501 Bielefeld,
Germany

L.Khachatrian†
Fakultät für Mathematik,
Universität Bielefeld,
Postfach 100131,
33501 Bielefeld,
Germany

Abstract

A (variable length) code is fix – free code if no codeword is a prefix or a suffix of any other. A database constructed by a fix – free code is instantaneously decodeable from both sides. We discuss the existence of fix – free codes, relations to the deBrujin Network and shadow problems. Particulary we draw attention to a remarkable **conjecture**: For numbers $l_1, \ldots, l_N$ satisfying $\sum_{i=1}^{N} 2^{-l_i} \leq \frac{3}{4}$ a fix–free code with lengths $l_1, \ldots l_N$ exists. If true, this bound is best possible.

*email:bernhard@mathematik.uni-bielefeld.de
†email:lk@mathematik.uni-bielefeld.de

1 Basic Definitions

For a finite set $\mathcal{X} = \{0, \dots, a-1\}$, called alphabet, we form $\mathcal{X}^n = \prod_1^n \mathcal{X}$, the words of length n, with letters from $\mathcal{X}$ and $\mathcal{X}^* = \bigcup_{n=0}^{\infty} \mathcal{X}^n$, the set of all finite length words including the empty word e from $\mathcal{X}^0 = \{e\}$, $\mathcal{X}^*$ is equipped with an associative operation, called concatenation, defined by

$$(x_1, \dots, x_n)(y_1, \dots, y_m) = (x_1, \dots, x_n, y_1, \dots, y_m).$$

We skip the brackets whenever this results in no confusion, in particular we write the letter x instead of (x). We also write $\mathcal{X}^+ = \mathcal{X}^* \setminus \{e\}$ for the set of non–empty words.

The **length** $\mid x^n \mid$ of the word $x^n = x_1 \dots x_n$ is the number n of letters in x^n.

A word $w \in \mathcal{X}^*$ is a **factor** of a word $x \in \mathcal{X}^*$ if there exist $u, v \in \mathcal{X}^*$ such that $x = uwv$. A factor w of x is **proper** if $w \neq x$.
For subsets $\mathcal{Y}, \mathcal{Z}$ of $\mathcal{X}^*$ and a word $w \in \mathcal{X}^*$, we define

$$\mathcal{Y}w = \{yw \in \mathcal{X}^* : y \in \mathcal{Y}\},$$

$$\mathcal{Y}\mathcal{Z} = \{yz \in \mathcal{X}^* : y \in \mathcal{Y}, z \in \mathcal{Z}\}$$

and

$$\mathcal{Y}w^{-1} = \{z \in \mathcal{X}^* : zw \in \mathcal{Y}\}.$$

A set of words $\mathcal{C} \subset \mathcal{X}^*$ is called a **code**.

Recall that a code is called prefix–free (resp. suffix–free), if no codeword is beginning (resp. ending) of another one.

Definition 1 *A code, which is simultaneously prefix–free and suffix–free, is called biprefix or fix–free. This can be expressed by the equations*

$$\mathcal{C}\mathcal{X}^+ \cap \mathcal{C} = \phi \text{ and } \mathcal{X}^+\mathcal{C} \cap \mathcal{C} = \phi.$$

Definition 2 *A code $\mathcal{C} = \{c_1, \dots, c_N\}$ over an a–letter alphabet $\mathcal{X}$ is said to be* **complete** *if it satisfies equality in Kraft's inequality, i.e. for*
$\ell_i = \mid c_i \mid$,

$$\sum_{i=1}^{N} a^{-\ell_i} = 1.$$

Definition 3 *A fix–free code $\mathcal{C}$ is called* **saturated**, *if it is not possible to find a fix–free code $\mathcal{C}'$ containing $\mathcal{C}$ properly, that is, $\mid \mathcal{C}' \mid > \mid \mathcal{C} \mid$.*