Improved Upper Bound for the Redundancy of Fix-Free Codes

Sergey Yekhanin

***Abstract*— A variable-length code is a fix-free code if no codeword is a prefix or a suffix of any other codeword. In a fix-free code any finite sequence of codewords can be decoded in both directions, which can improve the robustness to channel noise and speed up the decoding process. In this paper we prove a new sufficient condition of the existence of fix-free codes and improve the upper bound on the redundancy of optimal fix-free codes.**

***Index Terms*— Fix-free code, redundancy.**

I. Introduction

LET $p = \{p_1, \ldots, p_m\}$ be the probability distribution of a source, and let C be a code for the source. The redundancy R of a code C is defined as the difference between the average codeword length $L(C)$ of this code and the entropy $H(p)$ of the source. We denote the redundancy of an optimal fix-free code by R_f.

Ahlswede *et al.* [1] have proved that $0 \leq R_f < 2$. They have also shown that the lower bound 0 on R_f cannot be improved. Later Ye and Yeung [6], [7] derived several upper bounds on R_f in terms of partial information about the source distribution. The goal of this paper is to improve the upper bound on R_f from 2 to $4 - \log_2 5$, which is approximately 1.678.

Let $\mathbf{v}_n = (k_1, \ldots, k_n)$ be a vector, where k_i are nonnegative integers. By $C(\mathbf{v}_n)$ denote a binary variable-length code containing k_i codewords of length i, for each $i = \overline{1,n}$. The Kraft sum of the vector $\mathbf{v}_n$ is the quantity

$$S(\mathbf{v}_n) = \sum_{i=1}^{n} \frac{k_i}{2^i}. \tag{1}$$

Ahlswede *et al.* [1] conjectured that $S(\mathbf{v}_n) \leq \frac{3}{4}$ is a sufficient condition for the existence of a binary fix-free code $C(\mathbf{v}_n)$. They proved that the conjecture is true in the weaker case when the Kraft sum is at most $\frac{1}{2}$. If the conjecture is true the upper bound on R_f can be improved to $3 - \log_2 3$, which is approximately 1.415 [7]. Since the conjecture was made, several special cases of it were proven [4], [5], [7], [8], although the general conjecture still remains an open problem.

In this paper we prove a new special case of the conjecture. We show that $S(\mathbf{v}_n) \leq \frac{5}{8}$ implies the existence of a fix-free code $C(\mathbf{v}_n)$ (Theorem 1). This result yields an improved upper bound for the redundancy of optimal fix-free codes (Theorem 2).

The author is a Ph.D. student at the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (email: yekhanin@mit.edu)

II. New Sufficient Condition of the Existence of Fix-Free Codes

Let $\mathbf{w}$ be an arbitrary binary vector of length n. The binary vector composed of the first $\{\text{last}\}$ p symbols of $\mathbf{w}$ is called *p-prefix* $\{$*p-suffix*$\}$ of $\mathbf{w}$ and denoted by ${}^{p}\mathbf{w}$ $\{\mathbf{w}^{p}\}$. We say that vector $\mathbf{w}$ has the form $\alpha \star \beta$, where $\alpha, \beta \in \{0,1\}$ if ${}^{1}\mathbf{w} = \alpha$ and $\mathbf{w}^{1} = \beta$.

Consider a binary variable-length fix-free code $C(\mathbf{v}_n)$, where $\mathbf{v}_n = (k_1, \ldots, k_n)$.

A vector $\mathbf{w} \in \{0,1\}^n$ is called *prefix free* $\{$*suffix free*$\}$ over code $C(\mathbf{v}_n)$ if $C(\mathbf{v}_n)$ does not contain any prefix $\{$suffix$\}$ of $\mathbf{w}$.

By definition, put

$$\begin{aligned}
{}^{0}\overrightarrow{F}(C) &= \{\mathbf{w} | \mathbf{w} \text{ is prefix-free over } C \text{ and } {}^{1}\mathbf{w} = 0\},\\
{}^{1}\overrightarrow{F}(C) &= \{\mathbf{w} | \mathbf{w} \text{ is prefix-free over } C \text{ and } {}^{1}\mathbf{w} = 1\},\\
\overrightarrow{F}(C) &= {}^{0}\overrightarrow{F}(C) \cup {}^{1}\overrightarrow{F}(C),\\
\overleftarrow{F}^{0}(C) &= \{\mathbf{w} | \mathbf{w} \text{ is suffix-free over } C \text{ and } \mathbf{w}^{1} = 0\},\\
\overleftarrow{F}^{1}(C) &= \{\mathbf{w} | \mathbf{w} \text{ is suffix-free over } C \text{ and } \mathbf{w}^{1} = 1\},\\
\overleftarrow{F}(C) &= \overleftarrow{F}^{0}(C) \cup \overleftarrow{F}^{1}(C).
\end{aligned}$$

Let M be an arbitrary subset of $\{0,1\}^n$.

The set M is called *right regular* if all $(n-1)$-suffixes of words from M are pairwise distinct, i.e., $\forall c_1, c_2 \in M,\ c_1 \neq c_2$ implies $c_1^{n-1} \neq c_2^{n-1}$.

Similarly, the set M is called *left regular* if all $(n-1)$-prefixes of words from M are pairwise distinct, i.e., $\forall c_1, c_2 \in M,\ c_1 \neq c_2$ implies ${}^{n-1}c_1 \neq {}^{n-1}c_2$.

Clearly, ${}^{0}\overrightarrow{F}(C)$ and ${}^{1}\overrightarrow{F}(C)$ are right regular sets. Likewise, $\overleftarrow{F}^{0}(C)$ and $\overleftarrow{F}^{1}(C)$ are left regular sets.

Let M_1 and M_2 be arbitrary subsets of $\{0,1\}^n$. By definition, put

$$M_1 \otimes M_2 = \{\mathbf{w} \in \{0,1\}^{n+1} \mid {}^{n}\mathbf{w} \in M_1 \text{ and } \mathbf{w}^{n} \in M_2\}$$

The following lemma is obvious.

Lemma 1: Suppose $C(\mathbf{v}_n)$ is an arbitrary fix-free code; then $\overrightarrow{F}(C) \otimes \overleftarrow{F}(C)$ is the set of all words of length $n+1$ that can be added to $C(\mathbf{v}_n)$ without violation of the fix-free property of the code. Moreover, ${}^{\alpha}\overrightarrow{F}(C) \otimes \overleftarrow{F}^{\beta}(C)$ is the set of all words of the form $\alpha \star \beta$ and length $n+1$ that can be added to $C(\mathbf{v}_n)$ without violation of the fix-free property of the code.

Lemma 2: Suppose M_1 is a right regular subset of $\{0,1\}^n$ and M_2 is a left regular subset of $\{0,1\}^n$; then

$$|M_1 \otimes M_2| \geq |M_1| + |M_2| - 2^{n-1}. \tag{2}$$

Proof: By $M_1^{(n-1)}$ denote the set of $(n-1)$-suffixes of words from M_1. In the same way, by ${}^{(n-1)}M_2$ denote the set of $(n-1)$-prefixes of words from M_2. Since M_1 is right regular, it follows that $|M_1^{(n-1)}| = |M_1|$. Similarly, $|{}^{(n-1)}M_2| = |M_2|$. Since $M_1^{(n-1)}$ and ${}^{(n-1)}M_2$ are subsets of $\{0,1\}^{n-1}$, it follows that $|M_1^{(n-1)} \cup^{(n-1)} M_2| \leq 2^{n-1}$. Therefore, $|M_1^{(n-1)} \cap^{(n-1)} M_2| \geq |M_1| + |M_2| - 2^{n-1}$. Let $\mathbf{b}$ denote an arbitrary element of $M_1^{(n-1)} \cap^{(n-1)} M_2$. It now follows that there exist $a, c \in \{0,1\}$ such that $a\mathbf{b} \in M_1$ and $\mathbf{b}c \in M_2$. Hence, $a\mathbf{b}c \in M_1 \otimes M_2$. Thus, $|M_1 \otimes M_2| \geq |M_1| + |M_2| - 2^{n-1}$. This completes the proof.

Theorem 1: If $S(\mathbf{v}_n) \leq \frac{5}{8}$, then there exists a fix-free code $C(\mathbf{v}_n)$.

Proof: Clearly, it suffices to prove that $S(\mathbf{v}_n) = \frac{5}{8}$, implies the existence of a fix-free code $C(\mathbf{v}_n)$. Let us consider three cases.

1) $k_1 = 1$
2) $k_1 = 0,\ k_2 = 2$
3) $k_1 = 0,\ k_2 \leq 1$

In every case we construct the code $C(\mathbf{v}_n)$ in n steps. On step t we add k_t words of length t to the code. The input of step t is a code $C(\mathbf{v}_{t-1})$, the output is a code $C(\mathbf{v}_t)$. Thus, on step n we construct $C(\mathbf{v}_n)$.

Proof of case 1: We shall now prove that $S(\mathbf{v}_n) \leq \frac{3}{4}$ and $k_1 = 1$ imply the existence of a fix-free code $C(\mathbf{v}_n)$. This claim is stronger than the assertion of the theorem. Put $C(\mathbf{v}_1) = \{0\}$. Suppose that a fix-free code $C = C(\mathbf{v}_{t-1})$ is constructed; we shall prove that on step t we can add k_t words of length t to the code without violation of the fix-free property. By lemma 1, it is sufficient to prove that $|\overrightarrow{F}(C) \otimes \overleftarrow{F}(C)| \geq k_t$. Put $\delta = S(\mathbf{v}_{t-1})$. Using (1), we get $\delta + \frac{k_t}{2^t} \leq \frac{3}{4}$. Hence,

$$k_t \leq 3 * 2^{t-2} - \delta * 2^t. \tag{3}$$

Now note that since $0 \in C(\mathbf{v}_{t-1})$, it follows that ${}^0\overrightarrow{F}(C) = \overleftarrow{F}^0(C) = \emptyset$. Therefore, $\overrightarrow{F}(C)$ is right regular and $\overleftarrow{F}(C)$ is left regular. It can be easily checked that $|\overrightarrow{F}(C)| = |\overleftarrow{F}(C)| = 2^{t-1}(1-\delta)$. The application of lemma 2 yields

$$|\overrightarrow{F}(C) \otimes \overleftarrow{F}(C)| \geq 3 * 2^{t-2} - \delta * 2^t. \tag{4}$$

Combining (3) and (4), we obtain $|\overrightarrow{F}(C) \otimes \overleftarrow{F}(C)| \geq k_t$. This completes the proof of the first case of Theorem 1.

Proof of case 2: We shall now prove that $S(\mathbf{v}_n) \leq \frac{3}{4}$, $k_1 = 0$ and $k_2 = 2$ imply the existence of a fix-free code $C(\mathbf{v}_n)$. Again, our claim is stronger than the assertion of the theorem. Put $C(\mathbf{v}_2) = \{00, 11\}$. Suppose that a fix-free code $C = C(\mathbf{v}_{t-1})$ is constructed; we shall prove that $|\overrightarrow{F}(C) \otimes \overleftarrow{F}(C)| \geq k_t$. It is sufficient to prove that both inequalities (3) and (4) are fulfilled. The proof of inequality (3) is exactly the same as above, so we proceed to inequality (4).

Let us show that $\overrightarrow{F}(C)$ is right regular. Assume the converse. Then there exists a vector $\mathbf{b} \in \{0,1\}^{t-2}$ such that both words $0\mathbf{b}$ and $1\mathbf{b}$ are prefix free over $C(\mathbf{v}_{t-1})$. Let us consider the two cases ${}^1\mathbf{b} = 0$ and ${}^1\mathbf{b} = 1$ separately. In the first case $0\mathbf{b}$ is prefixed by the codeword 00. In the second case $1\mathbf{b}$ is prefixed by the codeword 11. Thus, we have come to a contradiction. By the same argument, $\overleftarrow{F}(C)$ is left regular. As above, $|\overrightarrow{F}(C)| = |\overleftarrow{F}(C)| = 2^{t-1}(1-\delta)$. The application of lemma 2 yields (4). This completes the proof of the second case of Theorem 1.

Proof of case 3: Since $k_1 = 0$ and $k_2 \leq 1$, it follows that the vector $\mathbf{v}_n$ can be uniquely represented as a sum of four vectors $\mathbf{v}_n^1, \mathbf{v}_n^2, \mathbf{v}_n^3, \mathbf{v}_n^4$ such that

$$\begin{cases} \mathbf{v}_n^i = \{k_1^i, \ldots, k_n^i\}, \quad \text{i=1,2,3,4,} \\ S(\mathbf{v}_n^1) = \frac{1}{4}, \\ S(\mathbf{v}_n^2) = S(\mathbf{v}_n^3) = S(\mathbf{v}_n^4) = \frac{1}{8}, \\ \text{If } k_t^i \neq 0, \text{ then } \forall i' > i, t' < t \quad k_{t'}^{i'} = 0. \end{cases} \tag{5}$$

Consider the following example of such representation.

$$\begin{array}{ll} \mathbf{v}_n = \{0,0,2,1,2,6,20\} & S(\mathbf{v}_n) = \frac{5}{8}, \\ \mathbf{v}_n^1 = \{0,0,2,0,0,0,0\} & S(\mathbf{v}_n^1) = \frac{1}{4}, \\ \mathbf{v}_n^2 = \{0,0,0,1,2,0,0\} & S(\mathbf{v}_n^2) = \frac{1}{8}, \\ \mathbf{v}_n^3 = \{0,0,0,0,0,6,4\} & S(\mathbf{v}_n^3) = \frac{1}{8}, \\ \mathbf{v}_n^4 = \{0,0,0,0,0,0,16\} & S(\mathbf{v}_n^4) = \frac{1}{8}. \end{array}$$

We shall construct a code $C(\mathbf{v}_n)$ that is a union of four codes $C(\mathbf{v}_n) = C^{00}(\mathbf{v}_n^1) \cup C^{01}(\mathbf{v}_n^2) \cup C^{10}(\mathbf{v}_n^3) \cup C^{11}(\mathbf{v}_n^4)$, where each code $C^{\alpha\beta}(\mathbf{v}_n^i)$ contains only codewords of the form $\alpha \star \beta$.

Thus, for each $t = \overline{1,n}$ the set of codewords of length t is composed of k_t^1 codewords of the form $0 \star 0$, k_t^2 codewords of the form $0 \star 1$, k_t^3 codewords of the form $1 \star 0$ and k_t^4 codewords of the form $1 \star 1$.

We start with an empty code $C(\mathbf{v}_1) = \emptyset$. Suppose that a fix-free code $C = C(\mathbf{v}_{t-1})$ is constructed; we shall prove that on step t the code can be extended with k_t^1 $0\star0$ codewords, k_t^2 $0\star1$ codewords, k_t^3 $1\star0$ codewords and k_t^4 $1\star1$ codewords of length t without violation of the fix-free property. By lemma 1, it is sufficient to prove that

$$\begin{array}{l} |{}^0\overrightarrow{F}(C) \otimes \overleftarrow{F}^0(C)| \geq k_t^1, \\ |{}^0\overrightarrow{F}(C) \otimes \overleftarrow{F}^1(C)| \geq k_t^2, \\ |{}^1\overrightarrow{F}(C) \otimes \overleftarrow{F}^0(C)| \geq k_t^3, \\ |{}^1\overrightarrow{F}(C) \otimes \overleftarrow{F}^1(C)| \geq k_t^4. \end{array} \tag{6}$$

Put $\delta_i = S(\mathbf{v}_{t-1}^i)$. Note that, by construction, $\delta_i = 0$ and $\delta_i < S(\mathbf{v}_n^i)$ both imply $\delta_{i+1} = 0$. We shall consider four possible cases:

1) $\delta_1 < \frac{1}{4}, \delta_2 = \delta_3 = \delta_4 = 0$
2) $\delta_1 = \frac{1}{4}, \delta_2 < \frac{1}{8}, \delta_3 = \delta_4 = 0$
3) $\delta_1 = \frac{1}{4}, \delta_2 = \frac{1}{8}, \delta_3 < \frac{1}{8}, \delta_4 = 0$
4) $\delta_1 = \frac{1}{4}, \delta_2 = \frac{1}{8}, \delta_3 = \frac{1}{8}, \delta_4 < \frac{1}{8}$

In all the cases we use the fact that

$$k_t^i \leq 2^t(S(\mathbf{v}_n^i) - \delta_i). \tag{7}$$

Case 3.1: $\delta_1 < \frac{1}{4}, \delta_2 = \delta_3 = \delta_4 = 0$. Using (7), we get

$$\begin{array}{cc} k_t^1 \leq 2^{t-2} - \delta_1 * 2^t, & k_t^2 \leq 2^{t-3}, \\ k_t^3 \leq 2^{t-3}, & k_t^4 \leq 2^{t-3}. \end{array}$$

It can be easily checked that

$$\begin{array}{c} |{}^0\overrightarrow{F}(C)| = |\overleftarrow{F}^0(C)| = 2^{t-2} - \delta_1 * 2^{t-1}, \\ |{}^1\overrightarrow{F}(C)| = |\overleftarrow{F}^1(C)| = 2^{t-2}. \end{array}$$

The application of lemma 2 yields

$$|^0\overrightarrow{F}(C)\otimes\overleftarrow{F}^0(C)|\geq 2^{t-2}-\delta_1*2^t\geq k_t^1,$$
$$|^0\overrightarrow{F}(C)\otimes\overleftarrow{F}^1(C)|\geq 2^{t-2}-\delta_1*2^{t-1}>2^{t-3}\geq k_t^2,$$
$$|^1\overrightarrow{F}(C)\otimes\overleftarrow{F}^0(C)|\geq 2^{t-2}-\delta_1*2^{t-1}>2^{t-3}\geq k_t^3,$$
$$|^1\overrightarrow{F}(C)\otimes\overleftarrow{F}^1(C)|\geq 2^{t-2}>k_t^4.$$

This completes the proof of case 3.1.

Case 3.2: $\delta_1=\frac{1}{4},\delta_2<\frac{1}{8},\delta_3=\delta_4=0$. By the same argument as above

$$k_t^1=0,\quad k_t^2\leq 2^{t-3}-\delta_2*2^t,$$
$$k_t^3\leq 2^{t-3},\quad k_t^4\leq 2^{t-3}.$$

We see that

$$|^0\overrightarrow{F}(C)|=2^{t-2}-(\tfrac{1}{4}+\delta_2)*2^{t-1},$$
$$|\overleftarrow{F}^0(C)|=2^{t-3},$$
$$|^1\overrightarrow{F}(C)|=2^{t-2}$$
$$|\overleftarrow{F}^1(C)|=2^{t-2}-\delta_2*2^{t-1}.$$

By lemma 2, we have

$$|^0\overrightarrow{F}(C)\otimes\overleftarrow{F}^1(C)|\geq 2^{t-3}-\delta_2*2^t\geq k_t^2,$$
$$|^1\overrightarrow{F}(C)\otimes\overleftarrow{F}^0(C)|\geq 2^{t-3}\geq k_t^3,$$
$$|^1\overrightarrow{F}(C)\otimes\overleftarrow{F}^1(C)|\geq 2^{t-2}-\delta_2*2^{t-1}>2^{t-3}\geq k_t^4.$$

This completes the proof of case 3.2.

Case 3.3: $\delta_1=\frac{1}{4},\delta_2=\frac{1}{8},\delta_3<\frac{1}{8},\delta_4=0$. As above,

$$k_t^1=0,\quad k_t^2=0,$$
$$k_t^3\leq 2^{t-3}-\delta_3*2^t,\quad k_t^4\leq 2^{t-3}.$$

It is easily shown that

$$|\overleftarrow{F}^0(C)|=2^{t-3}-\delta_3*2^{t-1},$$
$$|^1\overrightarrow{F}(C)|=2^{t-2}-\delta_3*2^{t-1}$$
$$|\overleftarrow{F}^1(C)|=3*2^{t-4}.$$

Applying lemma 2, we obtain

$$|^1\overrightarrow{F}(C)\otimes\overleftarrow{F}^0(C)|\geq 2^{t-3}-\delta_3*2^t\geq k_t^3,$$
$$|^1\overrightarrow{F}(C)\otimes\overleftarrow{F}^1(C)|\geq 3*2^{t-4}-\delta_3*2^{t-1}>2^{t-3}\geq k_t^4.$$

This completes the proof of case 3.3.

Case 3.4: $\delta_1=\frac{1}{4},\delta_2=\frac{1}{8},\delta_3=\frac{1}{8},\delta_4<\frac{1}{8}$. As above,

$$k_t^1=0,\quad k_t^2=0,$$
$$k_t^3=0,\quad k_t^4\leq 2^{t-3}-\delta_4*2^t.$$

One can easily see that

$$|^1\overrightarrow{F}(C)|=|\overleftarrow{F}^1(C)|=2^{t-2}-(\tfrac{1}{8}+\delta_4)*2^{t-1}.$$

By lemma 2, we have

$$|^1\overrightarrow{F}(C)\otimes\overleftarrow{F}^1(C)|\geq 2^{t-3}-\delta_4*2^t\geq k_t^4.$$

This completes the proof of the theorem.

III. Upper Bound for the Redundancy

Theorem 2: For each probability distribution $p=\{p_1,\ldots,p_m\}$ there exists a binary fix-free code C where the average length of the codewords $L(C)$ satisfies

$$L(C)<H(p)+4-\log_2 5.$$

Proof: By $l_1,\ldots,l_m$ denote the codeword lengths. We define

$$l_i=\lceil -\log_2 p_i+3-\log_2 5\rceil.$$

It follows that

$$\sum_{i=1}^m 2^{-l_i}\leq\sum_{i=1}^m 2^{\log_2 p_i-3+\log_2 5}=\frac{5}{8}\sum_{i=1}^m p_i=\frac{5}{8}.$$

By theorem 1 there exists a fix-free code C with the codeword lengths $l_1,\ldots,l_m$. The average length of this code is

$$L(C)=\sum_{i=1}^m p_i*l_i<\sum_{i=1}^m p_i(-\log_2 p_i+4-\log_2 5)=$$
$$H(p)+(4-\log_2 5)\sum_{i=1}^m p_i=H(p)+4-\log_2 5.$$

This completes the proof.

References

[1] R. Ahlswede, B. Balkenhol and L. Khachatrian, "Some properties of fix-free codes," in *Proc. 1-st Int. Sem. on Coding Theory and Combinatorics*, Thahkadzor, Armenia, pp. 20-33, 1996.

[2] D. Gillman and R.L. Rivest,"Complete variable-length fix-free codes," *Des., Codes Cryptogr.*, vol. 5, pp. 109-114, 1995.

[3] A.S. Fraenkel and S.T. Klein,"Bidirectional Huffman coding," *Computer J.*, vol. 33, pp. 296-307, 1990.

[4] K. Harada and K. Kobayashi, "A note on the fix-free code property," in *IEICE Trans. Fund. Electron., Commun. Comput. Sci.*, Vol. E82-A, no. 10, pp. 2121-2128, Oct, 1999.

[5] Z. Kukorelly, K. Zeger, "New binary fix-free codes with Kraft sum 3/4,' *Proc. Int. Symp. on Information Theory*, Lausanne, Switzerland, p. 178, June 2002.

[6] C. Ye, R.W. Yeung, "On fix-free codes," *Proc. Int. Symp. on Information Theory*, Sorrento, Italy, p. 426, June 2000.

[7] C. Ye, R.W. Yeung, "Some basic properties of fix-free codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 1, pp. 72-87, Jan. 2001.

[8] S. Yekhanin, "Sufficient conditions of existence of fix-free codes," *Proc. Int. Symp. on Information Theory*, Washington D.C., USA, p. 284, June 2001.