Nonlinear Codes Outperform the Best Linear Codes on the Binary Erasure Channel

Po-Ning Chen, Hsuan-Yin Lin
Department of Electrical and Computer Engineering
National Chiao-Tung University (NCTU)
Hsinchu, Taiwan
poning@faculty.nctu.edu.tw, lin.hsuanyin@ieee.org

Stefan M. Moser
ETH Zürich, Switzerland, and Department of
Electrical and Computer Engineering, National
Chiao-Tung University (NCTU), Hsinchu, Taiwan
stefan.moser@ieee.org

***Abstract*—The exact value of the average error probability of an arbitrary code (linear or nonlinear) using maximum likelihood decoding is studied on binary erasure channels (BECs) with arbitrary erasure probability $0 < \delta < 1$. The family of the *fair linear codes*, which are equivalent to a concatenation of several Hadamard linear codes, is proven to perform better (in the sense of average error probability with respect to maximum-likelihood decoding) than all other *linear* codes for many values of the blocklength n and for a dimension $k = 3$. It is then noted that the family of fair linear codes and the family of *fair nonlinear weak flip codes* both maximize the minimum Hamming distance under certain blocklengths. However, the fair nonlinear weak flip codes actually outperform the fair linear codes, i.e., linearity and global optimality cannot be simultaneously achieved for the number of codewords being $\mathsf{M} = 2^3$.**

***Index Terms*—Binary erasure channel, generalized Plotkin bound, optimal nonlinear channel coding, r-wise Hamming distance, weak flip codes.**

I. Introduction

In 1948 Claude E. Shannon published his brilliant landmark paper in channel coding theory, entitled "A mathematical theory of communication" [1]. In this ingenious work, Shannon proved that for every communication channel, it is possible to find an information transmission scheme that transmits data with arbitrarily small error probability as long as the rate is below the so-called *channel capacity*. Ever since the publication of Shannon's groundbreaking work, a main goal in conventional coding theory has been to find good codes with a rate close to channel capacity and with acceptable performance with respect to error probability. A large part of this effort has gone into the study of *linear codes* that allow an easier analysis due to their implicit algebraic structure. Although linear codes have been proven to provide good transmission rates with acceptably low error probability over some communications systems, they can still be improved because they are not designed based on the goal of minimizing the decoding error, but rather based on the goal of maximizing some other partially related quantities like, e.g., the minimum Hamming distance. Such improvements will be crucial for future advanced coding systems with a demand for even higher link quality.

At this background, we attempt to break away from traditional information theory principles of finding new theoretical results on bounds of rates or on bounds of average error probability, and instead we focus on an *optimal* (in the sense of *minimizing the exact average error probability* with respect to maximum-likelihood decoding) design of codes for a finite blocklength [2], [3]. We are seeking the basic principles in the concrete design of optimal codes with arbitrary but finite blocklengths. Specifically, for a certain given channel, we fix both the number of codewords M and the blocklength n, and try to find the structure of a code that minimizes the exact average error probability P_{e} among all codes, assuming that an optimal decoder based on maximal-likelihood (ML) criterion is adopted. As a basic principle and following our definitions in [3, Sec. 2], a code $\mathscr{C}^{(\mathsf{M},n)}$ is called *optimal* and denoted by $\mathscr{C}^{(\mathsf{M},n)*}$ if

$$P_{\mathrm{e}}\big(\mathscr{C}^{(\mathsf{M},n)*}\big) \le P_{\mathrm{e}}\big(\mathscr{C}^{(\mathsf{M},n)}\big) \tag{1}$$

for any (linear or nonlinear) code $\mathscr{C}^{(\mathsf{M},n)}$. On the other hand, a linear code is claimed to be the best linear code if we restrict ourselves to consider only the family of linear codes $\mathscr{C}_{\mathrm{lin}}^{(\mathsf{M},n)}$.

For designing such optimal code structures that achieve the smallest average ML error probability, the analysis of the exact error performance with ML decoding is vitally important. Unfortunately, the common tools in conventional coding theory such as the *minimum Hamming distance* or the *weight enumerating function (WEF)* do not provide enough information on the analytical evaluation of the exact error probability even for the cases of simple binary-input *discrete memoryless channels (DMCs)*. Note that so far no complete theoretical study in this regard has been presented, except partially in [2], [3]. In [4], the authors study numerical techniques to reduce the search complexity of best linear codes on the binary symmetric channel, and recent results [5], [6] have shown that the linear simplex codes and the possibly nonlinear equidistant codes that maximize the minimum pairwise Hamming distance are strictly suboptimal on the binary symmetric channel.

In this paper we focus on the *binary erasure channel (BEC)*. From our previous work [3], we know that for $\mathsf{M} = 2^1$ or 2^2, optimal codes can be found that are linear. In this paper we treat the case of $\mathsf{M} = 2^3 = 8$ and find that the optimal codes cannot be linear! Note that for an exact performance analysis, an extension of the pairwise Hamming distance, which we name *r-wise Hamming distance*, is needed and developed, and

it is found to be a key to the understanding of the codes' exact performance.

The remainder of this paper is structured as follows. After some comments about our notation, we will introduce a general column-wise description of binary codes in Section II. We review the family of *weak flip codes* including its subfamily of *fair weak flip codes* and re-define the classical *linear codes* from the column-wise point-of-view in Section II-A. Section II-B provides the definition of the r-wise Hamming distance and its corresponding notation. The main results are then summarized and discussed in Sections III: Section III-A generalizes the Plotkin bound for the r-wise Hamming distance, and in Section III-B the $(k = 3)$-dimensional best linear codes of size $\mathsf{M} = 2^3$ on the BEC are presented.

As a convention in coding theory, vectors (denoted by bold face Roman letters, e.g., $\mathbf{x}$) are row-vectors. However, for simplicity of notation and to avoid a large number of transpose-signs '$^\intercal$', we slightly misuse this notational convention for one special case: any vector $\mathbf{c}$ is a column-vector. We use capital letters for random quantities, e.g., X, and small letters for their deterministic realizations, e.g., x; constants are depicted by Greek letters, small Romans, or a special font, e.g., M; sets are denoted by calligraphic letters, e.g., $\mathcal{M}$; and $|\mathcal{M}|$ denotes the cardinality of the set $\mathcal{M}$.

II. Column-Wise Description of Binary Codes

Following our approach in [3], a *codebook matrix* of a general code $\mathscr{C}^{(\mathsf{M},n)}$ with M codewords and with blocklength n can be read either row-wise, where the M rows $\mathbf{x}_m$ correspond to the M codewords of length n, or column-wise with n column vectors $\mathbf{c}_j$ of length M:

$$\mathscr{C}^{(\mathsf{M},n)} = \begin{pmatrix} — \ \mathbf{x}_1 \ — \\ \vdots \\ — \ \mathbf{x}_{\mathsf{M}} \ — \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ | & | & & | \end{pmatrix}. \quad (2)$$

We use a convenient numbering system for the possible columns of the codebook matrix of binary codes as described in the following definition.

Definition 1: For fixed M and $b_m \in \{0,1\}$ with $m \in \mathcal{M} \triangleq \{1,\ldots,\mathsf{M}\}$, we describe the column vector $(b_1\ b_2\ \cdots\ b_{\mathsf{M}})^\intercal$ by its reverse binary representation of nonnegative integers $j = \sum_{m=1}^{\mathsf{M}} b_m\, 2^{\mathsf{M}-m}$, and write $\mathbf{c}_j^{(\mathsf{M})} \triangleq (b_1\ b_2\ \cdots\ b_{\mathsf{M}})^\intercal$.

Note that due to symmetry of the BEC, flipping all zeros to ones and vice-versa will result in a code of identical performance. Thus, from the aspect of finding simply one optimal code, we can neglect all candidate column vectors starting with a one, i.e., we require $b_1 = 0$. By excluding the futile all-zero column, the set $\mathcal{C}^{(\mathsf{M})}$ of all possible length-M *candidate columns* of general binary codes can then be restricted to

$$\mathcal{C}^{(\mathsf{M})} \triangleq \left\{\mathbf{c}_1^{(\mathsf{M})}, \mathbf{c}_2^{(\mathsf{M})}, \ldots, \mathbf{c}_{2^{\mathsf{M}-1}-1}^{(\mathsf{M})}\right\}. \quad (3)$$

Since the ordering of columns appearing in a codebook matrix is irrelevant for the performance of the code because the BEC is memoryless and stationary, we only need to record the count of different candidate columns. Hence, for a given codebook and for any $j \in \mathcal{J} \triangleq \{1,\ldots,2^{\mathsf{M}-1}-1\}$, we use t_j to denote the number of the corresponding candidate columns $\mathbf{c}_j^{(\mathsf{M})}$ appearing in the codebook matrix of $\mathscr{C}^{(\mathsf{M},n)}$, and describe the code by the type vector

$$\mathbf{t} \triangleq [t_1, t_2, \ldots, t_{2^{\mathsf{M}-1}-1}], \quad (4)$$

where $n = \sum_{j=1}^{2^{\mathsf{M}-1}-1} t_j$. We then say that the code is of *type* $\mathbf{t}$ and write simply $\mathscr{C}_{\mathbf{t}}^{(\mathsf{M},n)}$.

A. Weak Flip Codes and Linear Codes

We recall some special families of binary codes from [2].

Definition 2: Given an integer $\mathsf{M} \geq 2$, a length-M candidate column is called a *weak flip column* if its first component is 0 and its Hamming weight equals $\left\lfloor \frac{\mathsf{M}}{2} \right\rfloor$ or $\left\lceil \frac{\mathsf{M}}{2} \right\rceil$. The collection of all possible weak flip columns is called *weak flip candidate columns set* and is denoted by $\mathcal{C}_{\text{weak}}^{(\mathsf{M})}$.

By its definition, a weak flip column contains an almost equal number of zeros and ones. In the remainder of this paper, we use the following shorthands:

$$\mathsf{J} \triangleq 2^{\mathsf{M}-1} - 1,\ \bar{\ell} \triangleq \left\lceil \frac{\mathsf{M}}{2} \right\rceil,\ \underline{\ell} \triangleq \left\lfloor \frac{\mathsf{M}}{2} \right\rfloor,\ \mathsf{L} \triangleq \binom{2\bar{\ell}-1}{\bar{\ell}}. \quad (5)$$

Note that $\left|\mathcal{C}_{\text{weak}}^{(\mathsf{M})}\right| = \mathsf{L}$ (see [3]).

Definition 3: A *weak flip code* $\mathscr{C}_{\text{weak}}^{(\mathsf{M},n)}$ is constructed only by weak flip columns. Since in its type (see (4)) all positions corresponding to nonweak flip columns are zero, we use a reduced type vector for weak flip codes:

$$\mathbf{t}_{\text{weak}} \triangleq \left[t_{j_1}, t_{j_2}, \ldots, t_{j_{\mathsf{L}}}\right], \quad (6)$$

where $\sum_{w=1}^{\mathsf{L}} t_{j_w} = n$ with j_w, $w = 1,\ldots,\mathsf{L}$, representing the numbers of the corresponding weak flip candidate columns.

We have also defined a special subclass of weak flip codes that possess particular *quasi-linear* properties [3].

Definition 4: A weak flip code is called *fair* if it is constructed by an equal number of all possible weak flip candidate columns in $\mathcal{C}_{\text{weak}}^{(\mathsf{M})}$. Note that by definition the blocklength of a fair weak flip code $\mathscr{C}_{\text{fair}}^{(\mathsf{M},n)}$ is always a multiple of L.

In conventional coding theory, *linear codes* form an important class of error correcting codes that have been shown to possess powerful algebraic properties. We recall here only the common definition of linear codes. For more details we refer to the vast existing literature (e.g., see [7], [8]).

Definition 5: Let $\mathsf{M} = 2^k$, where $k \in \mathbb{N} \triangleq \{1,2,3,\ldots\}$. The binary code $\mathscr{C}_{\text{lin}}^{(\mathsf{M},n)}$ is *linear* if its codewords span a k-dimensional subspace of the n-dimensional vector space over the channel input alphabet.

One of the important properties of a linear code concerns their column weights.

Proposition 6: If an (M, n) binary code is linear, then each column of its codebook matrix has Hamming weight $\frac{\mathsf{M}}{2}$, i.e., the code is a weak flip code.

The above proposition concludes that linear codes are weak flip codes. However, the converse of Proposition 6 does not

necessarily hold, i.e., even if $\mathsf{M} = 2^k$ for some $k \in \mathbb{N}$, a weak flip code $\mathscr{C}^{(\mathsf{M},n)}$ is not necessarily linear. In summary, we have the following relations among linear, weak flip, and arbitrary (M, n) codes:

$$\Big\{\mathscr{C}_{\text{lin}}^{(\mathsf{M},n)}\Big\} \subset \Big\{\mathscr{C}_{\text{weak}}^{(\mathsf{M},n)}\Big\} \subset \Big\{\mathscr{C}^{(\mathsf{M},n)}\Big\}. \tag{7}$$

Next, we will derive the set $\mathcal{C}_{\text{lin}}^{(\mathsf{M})}$ of all possible length-M candidate columns for the codebook matrices of binary linear codes with $\mathsf{M} = 2^k$ codewords. Being a subspace, linear codes are usually represented by a generator matrix $\mathsf{G}_{k\times n}$. We now apply our column-wise point-of-view to the construction of generator matrices.[1] The generator matrix $\mathsf{G}_{k\times n}$ consists of n column vectors $\mathbf{c}_j$ of length k similar to (2). Note that since the generator matrix is a basis of the code subspace, only a column of all zeros is useless, i.e., there are totally $\mathsf{K} \triangleq 2^k - 1 = \mathsf{M} - 1$ possible candidate columns for $\mathsf{G}_{k\times n}$: $\mathbf{c}_j^{(k)} \triangleq (b_1\ b_2\ \cdots\ b_k)^{\mathsf{T}}$, where $j = \sum_{i=1}^{k} b_i\, 2^{k-i}$. Here b_1 is not necessarily equal to zero. Let U_k be an auxiliary K-by-k matrix consisting of all possible K candidate columns for the generator matrix: $\mathsf{U}_k^{\mathsf{T}} = \big(\mathbf{c}_1^{(k)}\ \cdots\ \mathbf{c}_{\mathsf{K}}^{(k)}\big)$. This matrix U_k then allows us to create the set of all possible candidate columns of length $\mathsf{M} = 2^k$ for the codebook matrix of a linear code.

Lemma 7: Given a dimension k, the *candidate columns set* $\mathcal{C}_{\text{lin}}^{(\mathsf{M})}$ for linear codes is given by the columns of the matrix

$$\begin{pmatrix}\mathbf{0}\\ \mathsf{U}_k\end{pmatrix}\mathsf{U}_k^{\mathsf{T}}, \tag{8}$$

where $\mathbf{0}$ denotes an all-zero row vector of length k.

Thus, the codebook matrix of any linear code can be represented by

$$\mathscr{C}_{\text{lin}}^{(\mathsf{M},n)} = \begin{pmatrix}\mathbf{0}\\ \mathsf{U}_k\end{pmatrix}\mathsf{G}_{k\times n}, \tag{9}$$

which consists of columns taken only from $\mathcal{C}_{\text{lin}}^{(\mathsf{M})}$. Similarly to (6), since in its type all positions corresponding to candidate columns not in $\mathcal{C}_{\text{lin}}^{(\mathsf{M})}$ are zero, we can also use a reduced type vector to describe a k-dimensional linear code:

$$\mathbf{t}_{\text{lin}} \triangleq \big[t_{j_1}, t_{j_2}, \ldots, t_{j_{\mathsf{K}}}\big], \tag{10}$$

where $\sum_{\ell=1}^{\mathsf{K}} t_{j_\ell} = n$ with j_ℓ, $\ell = 1, \ldots, \mathsf{K}$, representing the numbers of the corresponding candidate columns in $\mathcal{C}_{\text{lin}}^{(\mathsf{M})}$.

Definition 8: A linear code is called *fair* if its codebook matrix is constructed by an equal number of all possible candidate columns in $\mathcal{C}_{\text{lin}}^{(\mathsf{M})}$. Hence the blocklength of a fair linear code $\mathscr{C}_{\text{lin, fair}}^{(\mathsf{M},n)}$ is always a multiple of K.

[1]The authors in [4] have also used this approach to examine exhaustively all possible linear codes.

Example 9: Consider the fair linear code with dimension $k = 3$ and blocklength $n = \mathsf{K} = 7$:

$$\mathscr{C}_{\text{lin, fair}}^{(8,7)} = \begin{pmatrix}\mathbf{0}\\ \mathsf{U}_3\end{pmatrix}\mathsf{U}_3^{\mathsf{T}} = \begin{pmatrix} 0&0&0&0&0&0&0\\ 1&0&1&0&1&0&1\\ 0&1&1&0&0&1&1\\ 1&1&0&0&1&1&0\\ 0&0&0&1&1&1&1\\ 1&0&1&1&0&1&0\\ 0&1&1&1&1&0&0\\ 1&1&0&1&0&0&1 \end{pmatrix} \tag{11}$$

with the corresponding code type vector $\mathbf{t}_{\text{lin}} = [t_{85}, t_{51}, t_{102}, t_{15}, t_{90}, t_{60}, t_{105}] = [1,1,1,1,1,1,1]$. Note that the fair linear code with $k = 3$ and $n = 7$ is an $(8, 7)$ Hadamard linear code with all pairwise Hamming distances equal to 4.

B. r-wise Hamming Distance and r-wise Hamming Match

We once again emphasize that the pairwise Hamming distance is not sufficient for the description of the exact performance of a code. We therefore define the *r-wise Hamming distance* and show that in combination with the code type vector $\mathbf{t}$ it allows a precise formulation of the exact error probability of codes over the BEC.

Definition 10: For a given general codebook $\mathscr{C}^{(\mathsf{M},n)}$ and an arbitrary integer $2 \le r \le \mathsf{M}$, we fix some integers $1 \le i_1 < i_2 < \cdots < i_r \le \mathsf{M}$ and define the *r-wise Hamming match* $a_{i_1\, i_2\, \cdots\, i_r}\big(\mathscr{C}^{(\mathsf{M},n)}\big)$ to be the cardinality of the index set

$$a_{i_1\, i_2\, \cdots\, i_r}\big(\mathscr{C}^{(\mathsf{M},n)}\big) \triangleq \big|\big\{j \in \{1,\ldots,n\}\colon c_{j,i_1} = c_{j,i_2} = \cdots = c_{j,i_r}\big\}\big|. \tag{12}$$

The *r-wise Hamming distance* $d_{i_1\, i_2\, \cdots\, i_r}\big(\mathscr{C}^{(\mathsf{M},n)}\big)$ is accordingly defined as

$$d_{i_1\, i_2\, \cdots\, i_r}\big(\mathscr{C}^{(\mathsf{M},n)}\big) \triangleq n - a_{i_1\, i_r\, \cdots\, i_r}\big(\mathscr{C}^{(\mathsf{M},n)}\big). \tag{13}$$

Note that the r-wise Hamming distance can be written elegantly with the help of the type vector describing the corresponding code:

$$d_{i_1\, i_2\, \cdots\, i_r}\Big(\mathscr{C}_{\mathbf{t}}^{(\mathsf{M},n)}\Big) = n - \sum_{\substack{j\in\mathcal{J}\ \text{s.t.}\\ c_{j,i_1}=c_{j,i_2}=\cdots=c_{j,i_r}}} t_j, \tag{14}$$

Here t_j denotes the jth component of the code type vector $\mathbf{t}$ of length J, and c_{j,i_ℓ} is the i_ℓth component of the jth candidate column $\mathbf{c}_j^{(\mathsf{M})}$ as given in Definition 1. Usually we will omit the specification of the code and abbreviate the r-wise notation in (12) and (13) as $a_{i_1\, i_2\, \cdots\, i_r}^{(\mathsf{M},n)}$ and $d_{i_1\, i_2\, \cdots\, i_r}^{(\mathsf{M},n)}$ or, even shorter, $a_{\mathcal{I}}$ and $d_{\mathcal{I}}$ for some given $\mathcal{I} = \{i_1, i_2, \ldots, i_r\}$, respectively.

Definition 11: The *minimum r-wise Hamming distance* $d_{\text{min};r}$ is the minimum of all possible r-wise Hamming distances $d_{\mathcal{I}}$ for a given (M, n) code. Correspondingly, we are also interested in the *maximum r-wise Hamming match* $a_{\text{max};r}$, which is the maximum of all possible r-wise Hamming matches $a_{\mathcal{I}}$ and is given by $a_{\text{max};r} = n - d_{\text{min};r}$.

Note that in traditional coding theory it is customary to specify a code with three parameters $(\mathsf{M}, n, d_{\text{H,min}})$, where the third

parameter specifies the minimum pairwise Hamming distance (which corresponds to the 2-wise Hamming distance according to Definition 10). We follow this tradition but replace the minimum pairwise Hamming distance by a vector containing all minimum r-wise Hamming distances for $r = 2, \ldots, \bar{\ell}$:

$$\mathbf{d} \triangleq \big(d_{\min;2},\ d_{\min;3},\ \ldots,\ d_{\min;\bar{\ell}}\big). \tag{15}$$

The reason why we restrict $r \leq \bar{\ell}$ lies in the fact that for weak flip codes the minimum r-wise Hamming distance is only relevant for $2 \leq r \leq \bar{\ell}$; see the remark after Theorem 13.

Example 12: We continue with Example 9. One can show that the fair linear code with $k = 3$ and $n = 7$ is a $(8, 7, \mathbf{d})$ Hadamard linear code with $\mathbf{d} = (d_{\min;2}, d_{\min;3}, d_{\min;4}) = (4, 6, 6)$. Similarly, the fair linear code with $k = 3$ and $n = 35$ is a $(8, 35, \mathbf{d})$ Hadamard linear code with $\mathbf{d} = (d_{\min;2}, d_{\min;3}, d_{\min;4}) = (20, 30, 30)$. They are obviously not fair weak flip codes for $\mathsf{M} = 8$. Later in Theorem 15 we will show that the fair weak flip code with $\mathsf{M} = 8$ codewords is actually a $\big(8, 35, (20, 30, 34)\big)$ code.

III. Main Results

A. Generalized Plotkin Bound for r-wise Hamming Distance

The r-wise Hamming distance (together with the code type vector $\mathbf{t}$) plays an important role in the closed-form expression of the average ML error probability for an arbitrary code $\mathscr{C}_{\mathbf{t}}^{(\mathsf{M},n)}$ over a BEC. It is therefore interesting to find some bounds on the r-wise Hamming distance. We start with a generalization of the Plotkin bound for r-wise Hamming distance.

Theorem 13 (Plotkin Bound for r-wise Hamming Distance): The minimum r-wise Hamming distance with $2 \leq r \leq \mathsf{M}$ of an (M, n) binary code always satisfies

$$d_{\min;r} \leq \begin{cases} n\left(1 - \frac{\binom{\bar{\ell}-1}{r-1}}{\binom{2\bar{\ell}-1}{r-1}}\right) & \text{if } 2 \leq r \leq \bar{\ell}, \\ n & \text{if } \bar{\ell} < r \leq \mathsf{M}. \end{cases} \tag{16}$$

The above theorem only provides absorbing bounds to the r-wise Hamming distance for $2 \leq r \leq \bar{\ell}$. For larger values of the parameter r it only renders the trivial bound $d_{\min;r} \leq n$. Moreover, for $r > \bar{\ell}$, the minimum r-wise Hamming distance of a weak flip code is always equal to this trivial upper bound n and is therefore irrelevant to the code's exact error performance. Thus we only consider the minimum r-wise Hamming distances for $2 \leq r \leq \bar{\ell}$ in the code type vector in (15).

It is well-known that the largest minimum pairwise Hamming distance (or equivalently, the largest minimum 2-wise Hamming distance) can be achieved by Hadamard codes, provided that the corresponding Hadamard matrix exists [8, Ch. 2]. Moreover, we have shown in [3] that the fair weak flip codes maximize the minimum pairwise Hamming distance, and we have conjectured that it is globally optimal in the sense of minimizing the average ML error probability on the BEC. The question therefore arises whether the fair weak flip code achieves the generalized Plotkin Bound (16). We can answer this question by referring to s-designs [9] from combinatorial design theory.

Definition 14 ([9, Ch. 9]): Let v, κ, λ_s, and s be positive integers such that $v > \kappa \geq s$. An s-(v, κ, λ_s) *design* or simply *s-design* is a pair $(\mathcal{X}, \mathscr{B})$, where $\mathcal{X}$ is a set of size v and $\mathscr{B}$ is a collection of subsets of $\mathcal{X}$ (called *blocks*), such that the following properties are satisfied:

1) each block $\mathcal{B} \in \mathscr{B}$ contains exactly κ points, and
2) every set of s distinct points is contained in exactly λ_s blocks.

We now claim that the fair weak flip code for an arbitrary M and for certain blocklengths can be seen as an r-design with $2 \leq r \leq \bar{\ell}$, and that it achieves the Plotkin upper bound (16) for r-wise Hamming distances (again, it is trivial that its $d_{\min;r}$ for $r > \bar{\ell}$ are equal to n).

Theorem 15: Fix some M and a blocklength n with $n \bmod \mathsf{L} = 0$. Then a fair weak flip code $\mathscr{C}_{\text{fair}}^{(\mathsf{M},n)}$ achieves the largest minimum r-wise Hamming distance for all $2 \leq r \leq \bar{\ell}$ among all (M, n) codes and satisfies

$$d_{\min;r}\big(\mathscr{C}_{\text{fair}}^{(\mathsf{M},n)}\big) = n\left(1 - \frac{\binom{\bar{\ell}-1}{r-1}}{\binom{2\bar{\ell}-1}{r-1}}\right), \quad 2 \leq r \leq \bar{\ell}. \tag{17}$$

A final remark to Theorem 15 is that the fair linear code always meets the Plotkin bound for the 2-wise Hamming distance; however, it does not necessarily meet the Plotkin bound for r-wise Hamming distances with $r > 2$ as a fair weak flip code $\mathscr{C}_{\text{fair}}^{(\mathsf{M},n)}$ does. This serves as an indication that a fair linear code may perform strictly worse than the optimal fair weak flip code even if it is the best linear code. Evidence of this will be given in the next section.

B. Best Linear Codes with $\mathsf{M} = 8$ on BEC

In [3], a new approach has been proposed for the derivation of the exact average ML error probability. It is based on the Inclusion–Exclusion principle in probability theory [10]. Combined with the r-wise Hamming distance and the code parameter properties, it allows for a closed-form expression of the exact value of the average ML error probability of an arbitrary code $\mathscr{C}_{\mathbf{t}}^{(\mathsf{M},n)}$ used on a BEC [3, Thm. 46].

Theorem 16 (Exact Average ML Error Probability of Arbitrary $\mathscr{C}^{(\mathsf{M},n)}$ on BEC): Consider a BEC with the conditional channel probability

$$P_{Y|X}(y|x) = \begin{cases} 1-\delta & \text{if } y = x,\ x \in \{0,1\}, \\ \delta & \text{if } y = 2,\ x \in \{0,1\}, \end{cases} \tag{18}$$

where the erasure probability satisfies $0 < \delta < 1$. For a given code $\mathscr{C}_{\mathbf{t}}^{(\mathsf{M},n)}$, the average ML error probability can be expressed using the code parameters $\mathbf{t}$ as follows:

$$P_{\text{e}}\Big(\mathscr{C}_{\mathbf{t}}^{(\mathsf{M},n)}\Big) = \frac{1}{\mathsf{M}} \sum_{r=2}^{\mathsf{M}} (-1)^r \sum_{\substack{\mathcal{I} \subseteq \{1,\ldots,\mathsf{M}\} \\ |\mathcal{I}| = r}} \delta^{d_{\mathcal{I}}}, \tag{19}$$

where $d_{\mathcal{I}}$ denotes the r-wise Hamming distance as given in Definition 10.

In order to find the best linear codes, we consider this closed-form expression of the exact average ML error probability and turn the minimization problem into an optimization problem on the discrete variables $\mathbf{t}_{\text{lin}}$ subject to the condition that $\sum_{j=1}^{\mathsf{K}} t_j = n$. For the blocklength n being a multiple of $\mathsf{K} = 7$, we succeed to find the best linear codes of dimension $k = 3$.

Theorem 17: For a BEC and for any blocklength n being a multiple of $\mathsf{K} = 7$, a best linear code with dimension $k = 3$ is the fair linear code.

Unfortunately, this best linear code is not necessarily a globally optimal code among all possible codes (including nonlinear codes).

Example 18: Consider the fair linear code and the nonlinear fair weak flip code for $\mathsf{M} = 2^3$ and $n = 35$. From Theorem 16 we obtain

$$\begin{aligned} &P_{\text{e}}\Big(\mathscr{C}_{\text{lin, fair}}^{(8,35)}\Big) \\ &= \frac{1}{8}\Bigg(\binom{8}{2}\delta^{n-15} - \binom{8}{3}\delta^{n-5} + 14\delta^{n-5} + \Bigg(\binom{8}{4} - 14\Bigg)\delta^n \\ &\qquad - \binom{8}{5}\delta^n + \binom{8}{6}\delta^n - \binom{8}{7}\delta^n + \binom{8}{8}\delta^n\Bigg), \end{aligned} \tag{20}$$

and from Theorems 15 and also 16, we get

$$\begin{aligned} &P_{\text{e}}\Big(\mathscr{C}_{\text{fair}}^{(8,35)}\Big) \\ &= \frac{1}{8}\Bigg(\binom{8}{2}\delta^{n-15} - \binom{8}{3}\delta^{n-5} + \binom{8}{4}\delta^{n-1} \\ &\qquad - \binom{8}{5}\delta^n + \binom{8}{6}\delta^n - \binom{8}{7}\delta^n + \binom{8}{8}\delta^n\Bigg). \end{aligned} \tag{21}$$

Thus,

$$P_{\text{e}}\Big(\mathscr{C}_{\text{lin, fair}}^{(8,35)}\Big) - P_{\text{e}}\Big(\mathscr{C}_{\text{fair}}^{(8,35)}\Big) = \frac{14}{8}\big(\delta^{n-5} + 4\delta^n - 5\delta^{n-1}\big), \tag{22}$$

which is strictly positive because the arithmetic mean is strictly larger than the geometric mean. Hence, the fair linear code with dimension $k = 3$ and blocklength $n = 35$ is not globally optimal among all possible codes even if it beats any other linear code in performance.

Actually, this example can be generalized to any blocklength being a multiple of 7 except $n = 7$. The derivation is based on elaborately extracting n columns from the codebook matrix of a fair weak flip code with blocklength larger than n to form a new $(8, n)$ nonlinear code that is a concatenation of *nonlinear* Hadamard codes. The technique however fails for $n = 7$ because taking any seven columns from the code matrix of the $(8, 35)$ fair weak flip code always results in a Hadamard *linear* code. We omit the details and only summarize the main statement.

Proposition 19: For $n \bmod 7 = 0$ apart from $n = 7$, the fair linear code with $\mathsf{M} = 8$ codewords given in Theorem 17 is strictly suboptimal over the BEC.

Acknowledgment

This work was supported by the National Science Council under NSC 100–2221–E–009–068–MY3.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Jul. and Oct. 1948.

[2] P.-N. Chen, H.-Y. Lin, and S. M. Moser, "Optimal ultrasmall block-codes for binary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7346–7378, Nov. 2013.

[3] ——, "Weak flip codes and its optimality on the binary erasure channel," Apr. 2014, subm. to *IEEE Trans. Inf. Theory*. [Online]. Available: http://moser-isi.ethz.ch/publications.html

[4] A. B. Fontaine and W. W. Peterson, "Group code equivalence and optimum codes," *IRE Trans. Inf. Theory*, vol. 5, no. 5, pp. 60–70, May 1959.

[5] T. Helleseth, T. Kløve, and V. I. Levenshtein, "The simplex codes and other even-weight binary linear codes for error correction," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2818–2823, Nov. 2004.

[6] P.-N. Chen, H.-Y. Lin, and S. M. Moser, "Equidistant codes meeting the Plotkin bound are not optimal on the binary symmetric channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 7–13, 2013, pp. 3015–3019.

[7] S. Lin and D. J. Costello, Jr., *Error Control Coding*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2004.

[8] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.

[9] D. R. Stinson, *Combinatorial Designs: Constructions and Analysis*. Springer Verlag, 2003.

[10] R. A. Brualdi, *Introductory Combinatorics*, 5th ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2010.