Connections Between the Error Probability and the r-wise Hamming Distances

(Invited Paper)

Hsuan-Yin Lin*, Stefan M. Moser†‡, Po-Ning Chen‡
*Simula@UiB, N–5020 Bergen, Norway
†Signal and Information Processing Lab, ETH Zurich, Switzerland
‡Institute of Communications Engineering, National Chiao Tung University, Hsinchu, Taiwan

***Abstract*—An extension from the pairwise Hamming distance to the r-*wise Hamming distance* is presented. It can be used to fully characterize the maximum-likelihood decoding (MLD) error of an arbitrary code over the binary erasure channel (BEC). By noting that good codes always have large minimum r-wise Hamming distances for all r, a new design criterion for a code is introduced: the *minimum* r-*wise Hamming distance*. We then prove an upper bound for the minimum r-wise Hamming distance of an arbitrary code, called the *generalized Plotkin bound*, and provide a class of (nonlinear) codes that achieve the bound for every r.**

I. Introduction

It has been an important goal in coding and information theory to design good codes working close to *channel capacity*, i.e., the largest possible reliable transmission rate over a channel. This terminology was first introduced by Shannon in his groundbreaking paper in 1948 [1]. In principle, under the assumption that the best decoder (i.e., a *maximum likelihood decoder*) is employed, Shannon proved that, when transmitting at a rate below capacity, a good code can achieve an arbitrarily small *error probability* as long as the blocklength is large enough. Shannon derived this result without actually computing the actual error probability of any code. Indeed, even for the restricted class of binary memoryless channels, it is still difficult to evaluate the code's exact *maximum likelihood decoding (MLD) error probability* for a given fixed blocklength. This difficulty even remains even if we restrict ourselves to *linear codes* (a family of codes that exhibits a certain linear structure), although linear codes do retain the ability of achieving capacity. Hence, it is still unknown how to find the best code that achieves the smallest MLD error probability among all codes of equal size and blocklength.

As the computation of the MLD error probability is so difficult, it is quite common to elude to the alternative criterion of the *minimum pairwise Hamming distance*, which is linked to the pairwise error probability. One then tries to find a good code that maximizes the minimum pairwise Hamming distance instead of minimizing the MLD error probability. Note that it is possible to find a linear code that attains the largest minimum pairwise Hamming distance among all (linear or nonlinear) codes of equal size and blocklength. Thus, from the minimum pairwise Hamming distance point of view, it is no loss to restrict oneself to linear codes. Unfortunately, a code that achieves the largest minimum pairwise Hamming distance does not necessarily achieve the smallest MLD error probability over symmetric channels.

In this work, since any permutation of the columns of a code is equivalent in the sense of its MLD error probability over memoryless channels, we describe an arbitrary code (linear or nonlinear) in terms of the types of columns in the code matrix. This approach transforms the problem of finding the code that achieves the smallest MLD error probability into a discrete multivariate constrained optimization problem.

Using simple *code parameters* to represent a code of size M and blocklength n, we define the family of (nonlinear) *weak flip codes* by elaborately extracting a subset of all possible code parameters, and introduce its subfamily of *fair weak flip codes* that exist only for certain blocklengths. The fair weak flip codes have some beautiful properties. For example, for every $r \geq 2$, they are guaranteed to achieve the largest *minimum* r-*wise Hamming distance* among all (linear or nonlinear) codes of equal size and blocklength (the r-wise Hamming distance is a generalization of the pairwise Hamming distance, see Definition 10 in Section III-A). This is in contrast to the linear codes that do not necessarily achieve the largest minimum r-wise Hamming distance for some $r \geq 2$.

We then address codes that minimize the exact MLD error probability over n uses of the *binary erasure channel (BEC)*. It is proved that these r-wise Hamming distances can be exploited to fully characterize the exact MLD error probability of an arbitrary code over the BEC. Using this characterization we then succeed to show that for $\mathsf{M} = 8$ the fair weak flip code outperforms the best linear code over the BEC.[1] Also for blocklengths $n \leq 35$, where no fair weak flip codes exist, a random search indicates that always a weak flip code can be found that beats the best linear code both in having a larger minimum 4-wise Hamming distance and in achieving a better MLD error probability. Thus, maximizing the minimum r-wise Hamming distances can serve as an effective design criterion

This work was supported partially by the Ministry of Science and Technology, Taiwan, under Grant MOST 104-2221-E-009-077-MY2.

[1] We use "best" here in the sense of achieving the smallest MLD error probability among a restricted class of codes of equal size and blocklength.

for nonlinear codes.

We use the following notational conventions. Vectors are usually row-vectors and are represented by boldface italic Roman letters, e.g., $\boldsymbol{x}$. However, we will slightly abuse this convention in one special case: any vector $\boldsymbol{c}$ is a column vector. Random quantities are denoted as upper case letters, e.g., X, and their deterministic counterparts are denoted as lower case letters, e.g., x. We use Greek letters, small Romans, or a special font, e.g., M, to denote constants. Sets are depicted by calligraphic upper case letters, e.g., $\mathcal{I}$, and the cardinality of a set $\mathcal{I}$ is denoted by $|\mathcal{I}|$. A codebook consisting of M codewords of length n is called an (M, n) *code* and depicted by $\mathscr{C}^{(\mathsf{M},n)}$. If they are unambiguous from the context, we might drop the superscripts and simply write $\mathscr{C}$.

Due to page limitations, all proofs are omitted and can be found in [2].

II. Setup and Definitions

A. Binary Erasure Channel (BEC)

In this work we focus on the *binary erasure channel (BEC)*, a discrete memoryless channel (DMC) with a binary input alphabet $\mathcal{X} = \{0, 1\}$ and a ternary output alphabet $\mathcal{Y} = \{0, 1, 2\}$, and with the conditional channel law

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

Here $0 \leq \epsilon < 1$ is called the *erasure probability*.

B. Column-Wise Description of General Binary Codes

An (M, n) code $\mathscr{C}^{(\mathsf{M},n)}$ can be written as an $\mathsf{M} \times n$ matrix with the rows corresponding to the M codewords:

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

In our approach, we prefer to consider the codebook matrix *column-wise* rather than row-wise [3]. We denote the length-M column-vectors of the codebook by $\boldsymbol{c}_j$, $j \in \{1, \ldots, n\}$.

We define a convenient numbering system for all possible columns that can occur in such a codebook matrix.

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

Due to the symmetry of the BEC and [2, Lem. 9], we discard any column starting with a one, i.e., we require $b_1 = 0$. Moreover, as it never helps improving the performance, we exclude the all-zero column. Hence, the set of all possible *candidate columns* of general binary codes is

$$\mathcal{C}^{(\mathsf{M})} \triangleq \bigl\{\boldsymbol{c}_1^{(\mathsf{M})}, \boldsymbol{c}_2^{(\mathsf{M})}, \ldots, \boldsymbol{c}_{\mathsf{J}}^{(\mathsf{M})}\bigr\}, \tag{3}$$

where $\mathsf{J} \triangleq 2^{\mathsf{M}-1} - 1$. For a given codebook and for any $j \in \mathcal{J} \triangleq \{1, \ldots, \mathsf{J}\}$, let t_j be the number of occurrences of the corresponding candidate columns $\boldsymbol{c}_j^{(\mathsf{M})}$ in the codebook matrix of $\mathscr{C}^{(\mathsf{M},n)}$. Since the ordering of the candidate columns is irrelevant with respect to the MLD performance of the code on a DMC, any binary code with blocklength $n = \sum_{j=1}^{\mathsf{J}} t_j$ can therefore be fully described by the parameter vector

$$\boldsymbol{t} \triangleq \bigl[t_1, t_2, \ldots, t_{\mathsf{J}}\bigr]. \tag{4}$$

We say that such a code has a *type vector* (or simply *type*) $\boldsymbol{t}$, and write $\mathscr{C}_{t_1,\ldots,t_{\mathsf{J}}}^{(\mathsf{M},n)}$ or $\mathscr{C}_{\boldsymbol{t}}^{(\mathsf{M},n)}$.

More details about the column-wise description of binary codes can be found in [2, Sec. II.C].

C. Weak Flip Codes

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

We see that a weak flip column contains an almost equal or equal number of zeros and ones. We introduce the following shorthands:

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

Definition 3: A *weak flip code* $\mathscr{C}_{\text{weak}}^{(\mathsf{M},n)}$ contains only weak flip columns in its codebook matrix. Since all positions corresponding to nonweak flip columns are zero, the type vector (4) can be reduced to a *reduced type vector:*

$$\boldsymbol{t}_{\text{weak}} \triangleq \bigl[t_{j_1}, t_{j_2}, \ldots, t_{j_{\mathsf{L}}}\bigr], \tag{6}$$

where $\sum_{w=1}^{\mathsf{L}} t_{j_w} = n$ with j_w being the reverse binary representation of the corresponding weak flip column.

As an example, for $\mathsf{M} = 4$ we have $\boldsymbol{t}_{\text{weak}} = [t_3, t_5, t_6]$. Note that the number of weak flip columns is increasing exponentially fast; e.g., for $\mathsf{M} = 5$, we already have ten weak flip columns.

Next, we introduce the subclass of *fair weak flip codes*.

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

Fair weak flip codes have been used by Shannon *et al.* [4] for the derivation of error exponents, although the codes were not named at that time. Note that in [4] the error exponents are defined when blocklength n tends to infinity, but here we consider finite n. For more details and properties we refer to [2, Sec. IV.B].

D. Linear Codes

In conventional coding theory, *linear codes* constitute an important and well-known class of error correcting codes that have been shown to possess powerful algebraic properties (e.g., see [5], [6]). We focus briefly on certain properties of linear codes that are important in the context of this work.

We start by categorizing linear codes as a special case of weak flip codes.

Proposition 5: Every linear code is a weak flip code.

Note that linear codes only exist if $\mathsf{M} = 2^k$, while weak flip codes are defined for any M. Also note that the converse of Proposition 5 does not necessarily hold, i.e., even if $\mathsf{M} = 2^k$ for some $k \in \mathbb{N} \triangleq \{1, 2, 3, \ldots\}$, a weak flip code $\mathscr{C}^{(\mathsf{M},n)}$ is not necessarily linear.

We next investigate linear codes from a column-wise perspective. The goal here is to define *fair linear codes*. As a vector subspace, linear codes are usually represented by a generator matrix $\mathsf{G}_{k\times n}$. We can apply our column-wise point-of-view to the construction of generator matrices.[2] The generator matrix $\mathsf{G}_{k\times n}$ consists of n column vectors $\boldsymbol{c}_j$ of length k similar to (2). Note that in the generator matrix the all-zero column is useless and is therefore excluded. Thus there are totally $\mathsf{K} \triangleq 2^k - 1 = \mathsf{M} - 1$ possible candidate columns for $\mathsf{G}_{k\times n}$: $\boldsymbol{c}_j^{(k)} \triangleq (b_1\ b_2\ \cdots\ b_k)^\mathsf{T}$, where $j = \sum_{i=1}^{k} b_i\, 2^{k-i}$ and where b_1 is not necessarily equal to zero. Let U_k^T be an auxiliary $k \times \mathsf{K}$ matrix consisting of all possible K candidate columns for the generator matrix: $\mathsf{U}_k^\mathsf{T} = \bigl(\boldsymbol{c}_1^{(k)}\ \cdots\ \boldsymbol{c}_\mathsf{K}^{(k)}\bigr)$. This matrix U_k^T then allows us to create the set $\mathcal{C}_{\text{lin}}^{(\mathsf{M})}$ of all possible length-M candidate columns of length $\mathsf{M} = 2^k$ for the codebook matrix of a binary linear code with $\mathsf{M} = 2^k$ codewords.

Lemma 6: Given a dimension k, the *candidate columns set* $\mathcal{C}_{\text{lin}}^{(\mathsf{M})}$ for linear codes is given by the columns of the $\mathsf{M} \times (\mathsf{M}-1)$ matrix $\left(\begin{smallmatrix}\mathbf{0}\\ \mathsf{U}_k\end{smallmatrix}\right)\mathsf{U}_k^\mathsf{T}$, where $\mathbf{0}$ denotes an all-zero row vector of length k.

Thus, the codebook matrix of any linear code can be represented by $\left(\begin{smallmatrix}\mathbf{0}\\ \mathsf{U}_k\end{smallmatrix}\right)\mathsf{G}_{k\times n}$, which consists of columns taken only from $\mathcal{C}_{\text{lin}}^{(\mathsf{M})}$. Since in its type, all positions corresponding to candidate columns not in $\mathcal{C}_{\text{lin}}^{(\mathsf{M})}$ are zero, we can again use a reduced type vector to describe a k-dimensional linear code:

$$\boldsymbol{t}_{\text{lin}} \triangleq \bigl[t_{j_1}, t_{j_2}, \ldots, t_{j_\mathsf{K}}\bigr], \tag{7}$$

where $\sum_{\ell=1}^{\mathsf{K}} t_{j_\ell} = n$ with j_ℓ being the reverse binary representation of the corresponding weak flip column.

Definition 7: 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[3] $\mathscr{C}_{\text{lin,fair}}^{(\mathsf{M},n)}$ is always a multiple of $\mathsf{K} = \mathsf{M} - 1$.

Example 8: Consider the fair linear code with dimension $k = 3$ and blocklength $n = \mathsf{K} = 7$:

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

[2] The authors in [7] have also used this approach to exhaustively examine all possible linear codes.

[3] We point out that a fair linear code actually is a binary simplex code [6, Ch. 1]. However, to remain synchronized with the description of fair weak flip codes, we will stick to the name *fair linear codes* throughout this paper.

with the corresponding reduced type vector

$$\boldsymbol{t}_{\text{lin}} = [t_{85}, t_{51}, t_{102}, t_{15}, t_{90}, t_{60}, t_{105}] = [1, 1, 1, 1, 1, 1, 1]. \tag{9}$$

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 [6, Ch. 2]. $\Diamond$

III. Column-Wise Analysis of Codes

A. r-Wise Hamming Distance

The minimum pairwise Hamming distance is a well-known and widely used quality criterion of a code. Unfortunately, a design solely based on the minimum pairwise Hamming distance can be strictly suboptimal even for a very symmetric channel like the *binary symmetric channel (BSC)* and even for linear codes [3], [8]. We will therefore next provide an extension of the pairwise Hamming distance: the so-called *r-wise Hamming distance of a code*. We will see that this generalization (in combination with the type vector $\boldsymbol{t}$) allows a precise formulation of the exact MLD error probability of a code over the BEC.

Definition 9 (r-Wise Hamming Distance): For a given general code $\mathscr{C}^{(\mathsf{M},n)}$ and an arbitrary integer $r \in \{2, \ldots, \mathsf{M}\}$, we fix some integers $1 \le i_1 < i_2 < \cdots < i_r \le \mathsf{M}$. The *$r$-wise Hamming distance* $d_{i_1\, i_2\, \cdots\, i_r}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr)$ is defined as

$$d_{i_1\, i_2\, \cdots\, i_r}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr) \triangleq n - a_{i_1\, i_r\, \cdots\, i_r}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr), \tag{10}$$

where

$$a_{i_1\, i_2\, \cdots\, i_r}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr) \triangleq \bigl|\{j \in \{1, \ldots, n\} \colon c_{j,i_1} = c_{j,i_2} = \cdots = c_{j,i_r}\}\bigr|, \tag{11}$$

and c_{j,i_ℓ} is the i_ℓth component of the jth candidate column $\boldsymbol{c}_j^{(\mathsf{M})}$ as given in Definition 1.

It is straightforward to verify that the 2-wise Hamming distances are identical to the pairwise Hamming distances.

The r-wise Hamming distances can be written elegantly with the help of the type vector:

$$d_{i_1\, i_2\, \cdots\, i_r}\bigl(\mathscr{C}_{\boldsymbol{t}}^{(\mathsf{M},n)}\bigr) = n - \sum_{\substack{j \in \mathcal{J} \text{ s.t.}\\ c_{j,i_1} = c_{j,i_2} = \cdots = c_{j,i_r}}} t_j, \tag{12}$$

where $1 \le i_1 < i_2 < \cdots < i_r \le \mathsf{M}$. Here t_j denotes the jth component of the type vector $\boldsymbol{t}$ of length $\mathsf{J} = 2^{\mathsf{M}-1} - 1$.

When the considered type-$\boldsymbol{t}$ code is unambiguous from the context, we will usually omit the explicit specification of the code and abbreviate (10) as $d_{i_1\, i_2\, \cdots\, i_r}^{(\mathsf{M},n)}$ or, even shorter, as $d_{\mathcal{I}}^{(\mathsf{M},n)}$ for some given $\mathcal{I} = \{i_1, i_2, \ldots, i_r\}$.

The definition of the r-wise Hamming distances leads to a natural extension of the minimum pairwise Hamming distance.

Definition 10 (Minimum r-Wise Hamming Distance): For a given $r \in \{2, \ldots, \mathsf{M}\}$, the *minimum r-wise Hamming distance* $d_{\min;r}$ *of a code* $\mathscr{C}^{(\mathsf{M},n)}$ is defined as the minimum of all possible r-wise Hamming distances of this (M, n) code:

$$d_{\min;r}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr) \triangleq \min_{\mathcal{I} \subseteq \{1,\ldots,\mathsf{M}\} \colon |\mathcal{I}| = r} d_{\mathcal{I}}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr), \tag{13}$$

where the minimization is taken over all size-r subsets $\mathcal{I} \subseteq \{1, \ldots, \mathsf{M}\}$.

Recall that in traditional coding theory, it is customary to specify a code with three parameters $(\mathsf{M}, n, d_{\text{min}})$, where the third parameter specifies the minimum pairwise Hamming distance. We follow this tradition but replace the minimum pairwise Hamming distance by a vector containing all minimum r-wise Hamming distances for $r = 2, \ldots, \bar{\ell} \triangleq \lceil \mathsf{M}/2 \rceil$:

$$\boldsymbol{d}_{\text{min}} \triangleq \bigl(d_{\text{min};2},\ d_{\text{min};3},\ \ldots,\ d_{\text{min};\bar{\ell}}\bigr). \tag{14}$$

Note that we restrict ourselves to $r \leq \bar{\ell}$ because for weak flip codes the minimum r-wise Hamming distance is equal to n for $\bar{\ell} < r \leq \mathsf{M}$; see the discussion after Theorem 13 below.

Example 11: We continue with Example 8. The fair linear code with $k = 3$ and $n = 7$ given in (8) is an $(8, 7, \boldsymbol{d}_{\text{min}})$ Hadamard linear code with $\boldsymbol{d}_{\text{min}} = (4, 6, 6)$. Similarly, the fair linear code with $k = 3$ and $n = 35$ that is created by concatenating the codebook matrix (8) five times is an $\bigl(8, 35, (20, 30, 30)\bigr)$ Hadamard linear code. Both codes are obviously not fair weak flip codes. Later in Theorem 14 we will show that the fair weak flip code with $\mathsf{M} = 8$ codewords is actually an $\bigl(8, 35, (20, 30, 34)\bigr)$ code. ♢

Following the classical definition of an *equidistant code*, i.e., a code whose pairwise Hamming distance between all codewords is the same, we define *r-wise equidistant codes*.

Definition 12 (r-Wise Equidistant Codes): For a given integer $r \in \{2, \ldots, \mathsf{M}\}$, an (M, n) code $\mathscr{C}^{(\mathsf{M},n)}$ is called *r-wise equidistant* if all r-wise Hamming distances are equal, i.e., if for all choices of integers $1 \leq i_1 < i_2 < \cdots < i_r \leq \mathsf{M}$, $d_{i_1 \cdots i_r}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr) = \text{constant}$.

B. Generalized Plotkin Bound for r-Wise Hamming Distances

The r-wise Hamming distance (together with the type vector $\boldsymbol{t}$) plays an important role in the closed-form expression of the MLD error probability for an arbitrary code $\mathscr{C}_{\boldsymbol{t}}^{(\mathsf{M},n)}$ over the 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 the minimum pairwise Hamming distance to the situation of the minimum r-wise Hamming distance.

Theorem 13 (Plotkin Bound for Minimum r-wise Hamming Distances): For some $r \in \{2, \ldots, \mathsf{M}\}$, the minimum r-wise Hamming distance of an (M, n) binary code satisfies

$$d_{\text{min};r}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr) \leq \begin{cases} n\left(1 - \frac{\binom{\bar{\ell}-1}{r-1}}{\binom{2\bar{\ell}-1}{r-1}}\right) & \text{if } 2 \leq r \leq \bar{\ell}, \\ n & \text{if } \bar{\ell} < r \leq \mathsf{M}. \end{cases} \tag{15}$$

The above theorem only provides absorbing bounds to the r-wise Hamming distance for $2 \leq r \leq \bar{\ell}$, while further increasing the parameter r only renders trivially $d_{\text{min};r} \leq n$. Since the minimum r-wise Hamming distance of a weak flip code for $r > \bar{\ell}$ is always equal to this trivial bound n and therefore is irrelevant for the exact MLD error performance over the BEC, the vector (14) contains the minimum r-wise Hamming distances for $2 \leq r \leq \bar{\ell}$ only.

It is well-known that Hadamard codes achieve the Plotkin bound with equality, i.e., they achieve the largest minimum pairwise or 2-wise Hamming distance [6, Ch. 2]. Moreover, Hadamard codes are also pairwise equidistant. In the following we will investigate generalizations of these two properties for weak flip codes.

Theorem 14: Fix some M, a blocklength n with $n \bmod \mathsf{L} = 0$, and some $r \in \{2, \ldots, \bar{\ell}\}$. Then if a weak flip code is r-wise equidistant, then it is also s-wise equidistant for all $2 \leq s < r$. Moreover, if this r-wise equidistant weak flip code $\mathscr{C}_{\text{equidist}}^{(\mathsf{M},n)}$ achieves the generalized Plotkin bound, i.e., it satisfies

$$d_{\text{min};r}\Bigl(\mathscr{C}_{\text{equidist}}^{(\mathsf{M},n)}\Bigr) = n \cdot \left(1 - \frac{\binom{\bar{\ell}-1}{r-1}}{\binom{2\bar{\ell}-1}{r-1}}\right), \tag{16}$$

then $\mathscr{C}_{\text{equidist}}^{(\mathsf{M},n)}$ must also achieve the largest minimum s-wise Hamming distances for all $2 \leq s < r$.

The following corollary can be obtained from Theorem 14.

Corollary 15: The 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.

We make the following remark to Corollary 15: The fair *linear* code always meets the Plotkin bound for the 2-wise Hamming distance; however, in contrast to the fair weak flip code $\mathscr{C}_{\text{fair}}^{(\mathsf{M},n)}$, it does not necessarily meet the Plotkin bound for r-wise Hamming distances for $r > 2$. This gives rise to our conjecture that a fair linear code performs strictly worse than the optimal fair weak flip code even if it is the best linear code with the smallest MLD error probability over the BEC. More evidence for this claim will be given in Section IV-B.

IV. Performance Analysis over the BEC

In Section II-B we have shown that any codebook can be described by the type vector $\boldsymbol{t}$. Therefore the minimization of the MLD error probability among all possible codebooks is transformed into an optimization problem on the discrete vector $\boldsymbol{t}$, subject to the condition that $\sum_{j=1}^{\mathsf{J}} t_j = n$. Consequently, the r-wise Hamming distance and the properties of the type vector play an important role in our analysis.

A. Exact MLD Error Probability of a Code over the BEC

In terms of r-wise Hamming distances, we are able to give a closed-form expression for the exact MLD error probability of an arbitrary code $\mathscr{C}_{\boldsymbol{t}}^{(\mathsf{M},n)}$ used on the BEC.

Theorem 16 (MLD Error Probability on the BEC): Consider the BEC with erasure probability $0 \leq \epsilon < 1$ and an arbitrary code $\mathscr{C}_{\boldsymbol{t}}^{(\mathsf{M},n)}$ with $\mathsf{M} \geq 2$. Its MLD error probability can be expressed using the type vector $\boldsymbol{t}$ as follows:

$$P_{\text{e}}\Bigl(\mathscr{C}_{\boldsymbol{t}}^{(\mathsf{M},n)}\Bigr) = \frac{1}{\mathsf{M}} \sum_{r=2}^{\mathsf{M}} (-1)^r \sum_{\substack{\mathcal{I} \subseteq \{1,\ldots,\mathsf{M}\}: \\ |\mathcal{I}| = r}} \epsilon^{d_{\mathcal{I}}^{(\mathsf{M},n)}}, \tag{17}$$

where $d_{\mathcal{I}}^{(\mathsf{M},n)}$ denotes the r-wise Hamming distance as given in Definition 9.

Table I

THE MINIMUM r-WISE HAMMING DISTANCES OF THE NUMERICALLY FOUND WEAK FLIP CODE AND THE BEST LINEAR CODE WITH $\mathsf{M} = 8$.

n	8		10		12		14		16		18		20		21		22		24		26		28		30		32		34		35	
	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$	$\boldsymbol{t}^{\diamond}_{\text{weak}}$	$\boldsymbol{t}^{*}_{\text{lin}}$
$d_{\min;2}$	4	4	5	5	6	6	8	8	8	8	10	10	11	11	12	12	12	12	13	13	14	14	16	16	16	16	18	18	19	19	20	20
$d_{\min;3}$	6	6	8	8	10	10	12	12	13	13	15	15	17	17	18	18	18	18	20	20	22	22	24	24	25	25	27	27	29	29	30	30
$d_{\min;4}$	7	6	9	8	11	10	13	12	15	13	17	15	19	17	20	18	21	18	23	20	25	22	27	24	29	25	31	27	33	29	34	30

B. *Linear vs. Nonlinear Codes: Comparisons for* $\mathsf{M} = 8$

In this section we will compare linear codes with nonlinear weak flip codes for $\mathsf{M} = 8$. We will see that the best linear codes (with the smallest MLD error probability among all linear codes) are strictly suboptimal in all cases that we examine. This further substantiates the superiority of our proposed weak flip codes over the widely used linear codes.

Theorem 17: For $n \bmod 7 = 0$ except for $n = 7$, the fair linear code with $\mathsf{M} = 8$ codewords is strictly suboptimal over the BEC.

It is interesting that for $\mathsf{M} = 8$ and for all blocklengths $n \bmod 35 = 0$, both the fair linear code and the fair weak flip code are 2-wise and 3-wise equidistant and both achieve the 2-wise and the 3-wise Plotkin bounds. However, only the fair weak flip code is also 4-wise equidistant and achieves the 4-wise Plotkin bound.

We conjecture that the fair weak flip code is globally optimal in the sense of minimizing the MLD error probability and actually show that the so-called *generalized fair weak flip codes* (see [2, App. C]) outperform the best linear codes for $\mathsf{M} = 8$. For blocklengths $n \bmod \mathsf{L} \neq 0$, the situation is in general unclear because the optimal discrete solution to the "fair noninteger" distribution among all weak flip columns might even end up with nonweak flip columns (see [2, Conj. 55]). Still, we have numerical evidence that the best found weak flip codes are superior to the best linear codes. We next elaborate on this numerically.

For $\mathsf{M} = 8$ and for any blocklength $n \leq 35$, the best linear codes of type $\boldsymbol{t}^{*}_{\text{lin}}$ are found by an exhaustive search over all possible linear code parameters $\boldsymbol{t}_{\text{lin}}$. Unfortunately, the same approach does not work for the weak flip codes, because the exhaustive search that varies over 35 weak flip columns results in a numerically unmanageable complexity. Instead, we use a simulated annealing algorithm [9] to determine a "good" weak flip code type $\boldsymbol{t}^{\diamond}_{\text{weak}}$ (which therefore is not guaranteed to be optimal). The simulated annealing algorithm we use is briefly summarized in [2, Sec. V.G.1].

Table I lists the resulting minimum r-wise Hamming distances for $r = 2, 3, 4$ for both $\boldsymbol{t}^{*}_{\text{lin}}$ and $\boldsymbol{t}^{\diamond}_{\text{weak}}$ for all even n and for all n being a multiple of 7, where $8 \leq n \leq 35$. Note that for $n \leq 7$, $\boldsymbol{t}^{\diamond}_{\text{weak}}$ is equivalent to $\boldsymbol{t}^{*}_{\text{lin}}$. We observe that $d_{\min;4}$ increases as n grows and that the best weak flip code always has a larger minimum 4-wise Hamming distance and strictly outperforms the best linear code over the BEC. This is consistent with Theorem 17.

Finally, we remark that the same insights still hold true when we increase the number of codewords to $\mathsf{M} = 16$ (see [2, Table II]), i.e., the numerically found (nonlinear) weak flip codes are superior to the corresponding best linear codes.

V. CONCLUSION

In this paper, a column-wise description of codes is employed to define the family of *weak flip codes* that include the special classes of the *linear codes* and the *fair weak flip codes*. The *r-wise Hamming distance* is proposed as an extension to the pairwise Hamming distance. Moreover, we derive a Plotkin-type bound on the r-wise Hamming distances for binary codes. It is then shown that the r-wise Hamming distances have an important application as a mean to express the exact MLD error probability of a code over the BEC. We prove that in contrast to linear codes, the fair weak flip codes achieve the largest minimum r-wise Hamming distance among all codes of equal size and blocklength for every $r \geq 2$. Finally, a numerical study for $\mathsf{M} = 8$ shows that we can find nonlinear weak flip codes that have a larger minimum 4-wise Hamming distance than the best linear codes, indicating that linear codes are strictly suboptimal with regard to MLD performance over the BEC.

REFERENCES

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

[2] Hsuan-Yin Lin, Stefan M. Moser, and Po-Ning Chen, "Weak flip codes and their optimality on the binary erasure channel," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5191–5218, Jul. 2018.

[3] Po-Ning Chen, Hsuan-Yin Lin, and Stefan M. Moser, "Optimal ultrasmall block-codes for binary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7346–7378, Nov. 2013.

[4] Claude E. Shannon, Robert G. Gallager, and Elwyn R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels," *Inf. Contr.*, pp. 65–103, Feb. 1967, part I.

[5] Shu Lin and Daniel J. Costello, Jr., *Error Control Coding*, 2nd ed. Upper Saddle River, NJ, USA: Pearson Prentice Hall, 2004.

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

[7] 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.

[8] Po-Ning Chen, Hsuan-Yin Lin, and Stefan 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. 7–12, 2013, pp. 3015–3019.

[9] Abbas A. El Gamal, Lane A. Hemachandra, Itzhak Shperling, and Victor K. Wei, "Using simulated annealing to design good codes," *IEEE Trans. Inf. Theory*, vol. 33, no. 1, pp. 116–123, Jan. 1987.