The r-wise Hamming Distance and its Operational Interpretation for Block Codes

(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*—We present an extension of the pairwise Hamming distance, the r*-wise Hamming distance*, and show that it can be used to fully characterize the maximum-likelihood decoding (MLD) error of an arbitrary code used over the binary erasure channel (BEC). Based on these insights, we present a new design criterion for a code: the *minimum* r*-wise Hamming distance*. We prove that, for every $r \geq 2$, the class of *fair weak flip codes* achieves the largest minimum r-wise Hamming distance among all codes of equal size M and blocklength n. Thus, it is conjectured that the fair weak flip code is optimal in the sense of achieving the smallest MLD error over the BEC. We confirm this conjecture for $\mathsf{M} \leq 4$ and all $n \geq 1$. For a code size $\mathsf{M} = 8$, we find that the best (in the sense of smallest MLD error) *linear* code cannot achieve the largest minimum 4-wise Hamming distance and is thus strictly outperformed by the fair weak flip code over the BEC.**

***Index Terms*—Binary erasure channel, maximum likelihood decoding, r-wise Hamming distance, weak flip codes.**

I. Introduction

Ever since Shannon introduced the notion of *channel capacity* in his groundbreaking paper in 1948 [1], it has been the goal to design codes that can achieve capacity. Since channel capacity is defined under the assumption that a code can have an arbitrarily large blocklength and arbitrarily many codewords, the criterion of a code being capacity achieving is rather theoretical. Much more relevant in a practical setup is the *error probability* of the code when it is transmitted over a certain channel and when the best possible decoder, the *maximum likelihood decoder*, is employed. The main issue with this practical approach, however, is that it is in general very difficult to evaluate the exact *maximum likelihood decoding (MLD) error probability*. Thus, as an approximation, the alternative criterion of the *minimum pairwise Hamming distance* is commonly adopted. It is important to remember that a code that achieves the largest minimum pairwise Hamming distance does not necessarily also achieve the smallest MLD error probability.

To simplify the code design, encoder, and decoder, it is also common to rely on *linear codes*. It is not difficult to prove that even with the additional constraint of linearity it is possible to find codes that achieve capacity, and it is also possible to find linear codes that maximize the minimum pairwise Hamming distance among all codes of equal size and blocklength. The question is, though, whether this also holds for the MLD error probability.

In this work, we break away from these common simplifications and focus on general codes and on the exact error probability. We aim for codes that minimize the (exact) MLD error probability over n-uses of the *binary erasure channel (BEC)*.

We start by proposing a novel analytical approach that describes an arbitrary code (linear or nonlinear) in terms of the types of columns in the code matrix. This allows us to use simple *code parameters* to represent any code of a certain code size M and a finite blocklength n. Based on this representation, we define the family of (nonlinear) *weak flip codes* by elaborately extracting a subset of all possible code parameters. In particular, the subfamily of *fair weak flip codes* (which exist only for certain blocklengths) is shown to have beautiful code properties and to outperform all linear codes of equal size and blocklength over the BEC.

In a next step we propose an r-wise generalization to the pairwise Hamming distance. We prove 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 prove that for $\mathsf{M} \leq 4$ and $n \geq 1$ the weak flip code is globally optimal[1] over the BEC.

Moreover, in the particular case of $\mathsf{M} = 8$, we find that the best linear code does not maximize the minimum r-wise Hamming distance for all $r \geq 2$. This is in contrast to the fair weak flip code that meets the largest minimum r-wise Hamming distances for every $r \geq 2$. And as expected, the fair weak flip code is found to outperform the best linear code over the BEC. 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.

This work was supported partially by the Ministry of Science and Technology, Taiwan, under Grant MOST 104-2221-E-009-077-MY2, and partially by the Research Council of Norway (grant 240985/F20).

[1] We use "optimal" always in the sense of achieving the smallest MLD error probability among all codes of equal size and blocklength.

978-1-5386-0579-0/18/$31.00 ©2018 European Union

Thus, we conclude that maximizing the minimum r-wise Hamming distances can serve as an effective design and quality criterion for nonlinear codes.

We use the following notation. 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 denotes 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 [2].

II. Setup and Definitions

A. Binary Erasure Channel (BEC)

In this work we focus on the *binary erasure channel (BEC)*. The BEC is 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

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

Here $0 \le \delta < 1$ 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 $b_m \in \{0, 1\}$, $m \in \mathcal{M}$, 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 will never help to improve the performance, we exclude the all-zero column. Hence, the set of all possible *candidate columns* of general binary codes is

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

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

$$\boldsymbol{t} \triangleq \left[t_1, t_2, \ldots, t_{2^{\mathsf{M}-1}-1}\right]. \tag{4}$$

We say that such a code has a *type vector* (or simply *type*) $\boldsymbol{t}$, and write $\mathscr{C}^{(\mathsf{M},n)}_{t_1,\ldots,t_{2^{\mathsf{M}-1}-1}}$ or $\mathscr{C}^{(\mathsf{M},n)}_{\boldsymbol{t}}$.

More details about the column-wise description of binary codes can be found in [2, Sec. 2.3].

C. Weak Flip Codes

Definition 2: Given an integer $\mathsf{M} \ge 2$, 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 $\left\lfloor \frac{\mathsf{M}}{2} \right\rfloor$ 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 $\mathcal{C}_{\text{weak}}^{(\mathsf{M})}$.

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

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

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:*

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

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 weak flip codes called *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 goes to infinity, but here we consider finite n. For more details and properties we refer to [2, Sec. 4.2].

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. In summary, we have the following relations among linear, weak flip, and arbitrary (M, n) codes:

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

Next, we are going to investigate linear codes from a column-wise perspective. The goal here is to define *fair linear codes*.

Being a vector 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.[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 $\mathsf{U}_k^{\mathsf{T}}$ 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 $\mathsf{U}_k^{\mathsf{T}}$ 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 $\bigl(\begin{smallmatrix}\mathbf{0}\\ \mathsf{U}_k\end{smallmatrix}\bigr)\mathsf{U}_k^{\mathsf{T}}$, where $\mathbf{0}$ denotes an all-zero row vector of length k.

Thus, the codebook matrix of any linear code can be represented by $\bigl(\begin{smallmatrix}\mathbf{0}\\ \mathsf{U}_k\end{smallmatrix}\bigr)\mathsf{G}_{k\times n}$, 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:

$$\boldsymbol{t}_{\text{lin}} \triangleq \bigl[t_{j_1}, t_{j_2}, \ldots, t_{j_{\mathsf{K}}}\bigr], \tag{8}$$

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 $n = \mathsf{K} = 7$:

$$\mathscr{C}_{\text{lin,fair}}^{(8,7)} = \begin{pmatrix}\mathbf{0}\\ \mathsf{U}_3\end{pmatrix}\mathsf{U}_3^{\mathsf{T}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & 1 & 0 & 1 & 0 & 1\\ 0 & 1 & 1 & 0 & 0 & 1 & 1\\ 1 & 1 & 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 1 & 1 & 1 & 1\\ 1 & 0 & 1 & 1 & 0 & 1 & 0\\ 0 & 1 & 1 & 1 & 1 & 0 & 0\\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \tag{9}$$

with the corresponding reduced type vector

$$\boldsymbol{t}_{\text{lin}} = [t_{85}, t_{51}, t_{102}, t_{15}, t_{90}, t_{60}, t_{105}] = [1, 1, 1, 1, 1, 1, 1]. \tag{10}$$

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]. ♢

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].

While the pairwise Hamming distance vector already contains more information about a particular code than simply the minimum Hamming distance, it is still not sufficient to describe the exact performance of a code. 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

$$d_{i_1\, i_2 \cdots i_r}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr) \triangleq n - a_{i_1\, i_r \cdots i_r}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr), \tag{11}$$

where

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

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:

$$d_{i_1\, i_2 \cdots i_r}\bigl(\mathscr{C}_{\boldsymbol{t}}^{(\mathsf{M},n)}\bigr) = n - \sum_{\substack{j \in \mathcal{J} \text{ s.t.}\\ c_{j,i_1} = c_{j,i_2} = \cdots = c_{j,i_r}}} t_j, \tag{13}$$

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 $\mathsf{J} = 2^{\mathsf{M}-1} - 1$, and

[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.

c_{j,i_ℓ} is the i_ℓth component of the jth candidate column $\boldsymbol{c}_j^{(\mathsf{M})}$ as given in Definition 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 (11) as $d^{(\mathsf{M},n)}_{i_1\, i_2 \cdots i_r}$ or, even shorter, as $d^{(\mathsf{M},n)}_{\mathcal{I}}$ 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 $d_{\text{min};r}$ 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:

$$d_{\text{min};r}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr) \triangleq \min_{\mathcal{I}\subseteq\{1,\ldots,\mathsf{M}\}:\ |\mathcal{I}|=r} d_{\mathcal{I}}\bigl(\mathscr{C}^{(\mathsf{M},n)}\bigr), \quad (14)$$

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 $r = 2, \ldots, \bar{\ell}$:

$$\boldsymbol{d}_{\text{min}} \triangleq \bigl(d_{\text{min};2},\ d_{\text{min};3},\ \ldots,\ d_{\text{min};\bar{\ell}}\bigr). \quad (15)$$

Note that we restrict ourselves to $r \leq \bar{\ell}$ because for weak flip codes the minimum r-wise Hamming distance is only relevant for $2 \leq r \leq \bar{\ell}$; 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 (9) 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 (9) 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. $\Diamond$

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

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

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 (15) 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 Hamming distance (or equivalently, the largest minimum 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 $\mathscr{C}_{\text{equidist}}^{(\mathsf{M},n)}$ also achieves the generalized Plotkin bound (and hence achieves the largest minimum r-wise Hamming distance), i.e., it satisfies

$$d_{\text{min};r}\Bigl(\mathscr{C}_{\text{equidist}}^{(\mathsf{M},n)}\Bigr) = n \cdot \left(1 - \frac{\binom{\bar{\ell}-1}{r-1}}{\binom{2\bar{\ell}-1}{r-1}}\right), \quad (17)$$

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-E.

IV. Performance Analysis of 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 turns 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 $0 \le \delta < 1$ and an arbitrary code $\mathscr{C}_{\boldsymbol{t}}^{(\mathsf{M},n)}$ with $\mathsf{M} \ge 2$. Its MLD error probability can be expressed using the type vector $\boldsymbol{t}$ as follows:

$$P_{\mathrm{e}}\left(\mathscr{C}_{\boldsymbol{t}}^{(\mathsf{M},n)}\right) = \frac{1}{\mathsf{M}} \sum_{r=2}^{\mathsf{M}} (-1)^r \sum_{\substack{\mathcal{I} \subseteq \{1,\ldots,\mathsf{M}\}: \\ |\mathcal{I}|=r}} \delta^{d_{\mathcal{I}}^{(\mathsf{M},n)}}, \quad (18)$$

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

B. Codes Achieving Optimal MLD Error Probability with Three or Four Codewords ($\mathsf{M} = 3, 4$)

We start to investigate the optimal codes for $\mathsf{M} \ge 3$, since an optimal code for $\mathsf{M} = 2$ on the BEC is quite trivially the repetition code.

Even though we know the exact MLD error probability for a code with an arbitrary number of codewords M on the BEC, the optimal code structure is not obvious.

Theorem 17: For the BEC and for any $n \ge 2$, an optimal code with $\mathsf{M} = 3$ or $\mathsf{M} = 4$ codewords is the weak flip code of type $\boldsymbol{t}_{\text{weak}}^{*}$, where for $\mathsf{M} = 3$

$$t_1^* = \left\lfloor \frac{n+2}{3} \right\rfloor, \quad t_2^* = \left\lfloor \frac{n+1}{3} \right\rfloor, \quad t_3^* = \left\lfloor \frac{n}{3} \right\rfloor \quad (19)$$

and for $\mathsf{M} = 4$

$$t_3^* = \left\lfloor \frac{n+2}{3} \right\rfloor, \quad t_5^* = \left\lfloor \frac{n+1}{3} \right\rfloor, \quad t_6^* = \left\lfloor \frac{n}{3} \right\rfloor. \quad (20)$$

Using the shorthand $k \triangleq \left\lfloor \frac{n}{3} \right\rfloor$, the code parameters of these optimal codes for $\mathsf{M} = 3$ and $\mathsf{M} = 4$ can be summarized as

$$\boldsymbol{t}_{\text{weak}}^{*} = \begin{cases} [k, k, k] & \text{if } n \bmod 3 = 0, \\ [k+1, k, k] & \text{if } n \bmod 3 = 1, \\ [k+1, k+1, k] & \text{if } n \bmod 3 = 2. \end{cases} \quad (21)$$

We would like to emphasize here that for $\mathsf{M} = 3, 4$ and n mod $3 = 0$, the corresponding optimal codes are fair weak flip codes.

C. Bounds on MLD Error Probabilities

Since we now know the optimal code structure, we can compare its performance to the best known bounds in the literature.

Figure 1 compares the exact optimal performance for $\mathsf{M} = 4$ with the following bounds: the SGB upper and lower bounds based on the optimal code as used by Shannon, Gallager, and Berlekamp [4] (thereby confirming that this lower bound is valid generally), the Gallager upper bound [9], and also the PPV upper and lower bounds based on the *random coding* and *meta-converse* methodology, respectively, proposed by Polyanskiy, Poor, and Verdú [10].

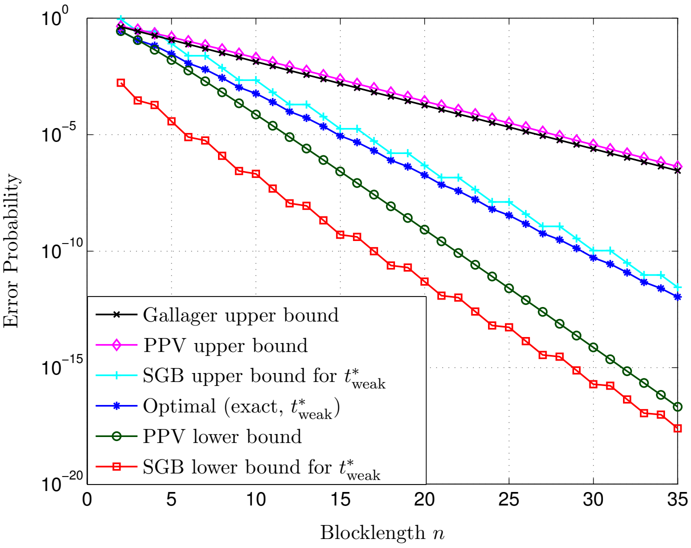

Figure 1. Exact value of, and bounds on, the MLD performance of an optimal code with $\mathsf{M} = 4$ codewords on the BEC with $\delta = 0.3$.

We can see that the SGB upper bound is closer to the exact optimal performance (and hence tighter) than the PPV upper bound and the Gallager upper bound. Note that the SGB upper bound does exhibit the correct error exponent. It is shown in [10] that when n goes to infinity under fixed M, the PPV upper bound only tends to the suboptimal Gallager exponent [9]; this fact is also confirmed by the figure.

Regarding the lower bounds we see that the PPV lower bound is much better for finite n than the SGB lower bound. However, the exponential growth rate of the PPV lower bound only approaches that of the sphere-packing bound [11], and does not equal the optimal exponent either [4].

D. r-Wise Hamming Distances for Codes with Arbitrary M

We have already pointed out that a code having a large (or even the largest) pairwise Hamming distance is not necessarily an optimal MLD-error code. It is crucial to look at all r-wise Hamming distances for $2 \le r \le \bar{\ell}$. In the following theorem we will once again confirm this intuition.

Theorem 18: Let the number of codewords be $\mathsf{M} = 2\bar{\ell}$ or $2\bar{\ell} - 1$ where $\bar{\ell}$ is an arbitrary positive even integer, and let the blocklength n be such that n mod $\mathsf{L} = 0$. Then an $(\bar{\ell}-1)$-wise equidistant weak flip code that achieves the largest minimum $(\bar{\ell}-1)$-wise Hamming distance[4] but not the largest minimum $\bar{\ell}$-wise Hamming distance has a strictly worse performance on the BEC than the fair weak flip code.

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

In this work, we are not really interested in linear codes as our focus lies on optimality in the sense of achieving the smallest MLD error probability. Nevertheless it is important to show the superiority of our proposed weak flip codes. To that goal we will next compare linear codes with nonlinear weak

[4] By Theorem 14, such a weak flip code is also s-wise equidistant and maximizes the s-wise Hamming distances for all $2 \le s \le \bar{\ell} - 1$.

Table I
THE MINIMUM r-WISE HAMMING DISTANCES OF THE (NUMERICALLY) BEST 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

flip codes for $\mathsf{M} = 8$. We will see that the best linear codes (with the smallest MLD error probability) are always strictly suboptimal in all cases that we numerically examine.

The following theorem shows that the fair linear code with $\mathsf{M} = 8$ codewords, which only achieves the 2-wise Plotkin bound, is strictly suboptimal on the BEC.

Theorem 19: 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. This is in agreement with Theorem 18 and explains why the fair linear code is outperformed on the BEC.

Based on these insights, we actually believe that the fair weak flip code is globally optimal and 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 are next going to elaborate on this.

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 [12] 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. 5.7.1].

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, $8 \leq n \leq 35$. 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 found best weak flip code always has a larger minimum 4-wise Hamming distance and thus strictly outperforms the best linear code over the BEC. This is consistent with Theorem 19.

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

V. CONCLUSION

In this paper, we employ a column-wise perspective to define and analyze the family of *weak flip codes* (including the special cases of the *linear codes* and the *fair weak flip codes*). We introduce the *r-wise Hamming distances* as an extension to the pairwise Hamming distance and use them to determine the exact MLD error probability of a code over the BEC. We prove that the fair weak flip codes achieve the largest minimum r-wise Hamming distance among codes of size M and certain blocklengths n for every $r \geq 2$, and that they are optimal in the sense of achieving the smallest MLD error probability over the BEC for $\mathsf{M} = 3, 4$. A numerical study for $\mathsf{M} = 8$ indicates that the best linear codes have a smaller minimum 4-wise Hamming distance than the optimal (nonlinear) weak flip codes, and are thus strictly suboptimal in 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," Jun. 2016, subm. to *IEEE Trans. Inf. Theory*. Available: https://arxiv.org/abs/1711.03310

[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. Jessy 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] Robert G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[10] Yury Polyanskiy, H. Vincent Poor, and Sergio Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[11] Yury Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.

[12] 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.