**Fiftieth Annual Allerton Conference**
**Allerton House, UIUC, Illinois, USA**
**October 1 - 5, 2012**

# Weak Flip Codes and Applications to Optimal Code Design on the Binary Erasure Channel

Po-Ning Chen, Hsuan-Yin Lin, and Stefan M. Moser
Department of Electrical and Computer Engineering
National Chiao Tung University (NCTU)
Hsinchu, Taiwan
Email: qponing@mail.nctu.edu.tw, {lin.hsuanyin, stefan.moser}@ieee.org

***Abstract*—A new family of *nonlinear* codes, called *weak flip codes*, is presented and is shown to contain many beautiful properties. In particular, the subfamily of *fair weak flip codes* can be seen as a generalization of linear codes. Different from linear codes that only exist for a number of codewords $\mathsf{M}$ being an integer-power of 2, the fair weak flip code can be defined for an arbitrary $\mathsf{M}$. It is then noted that the fair weak flip codes are related to *binary nonlinear Hadamard codes*: both code families maximize the minimum Hamming distance and meet the Plotkin bound. However, while the binary nonlinear Hadamard codes have only been shown to possess good Hamming-distance properties, the fair weak flip codes are proven to be globally optimal (in the sense of minimizing the error probability) among all linear or nonlinear codes for the binary erasure channel (BEC) for many values of the blocklength $n$ and for $\mathsf{M} \leq 6$. For $\mathsf{M} > 6$, similar optimality results are conjectured.**

**The results in this paper are founded on a new powerful tool for the analysis and generation of block codes: the *column-wise* approach to the codebook matrix.**

## I. Introduction

In traditional coding theory, it is the goal to find good codes that operate close to the ultimate limit of the *channel capacity* as introduced by Shannon [1]. Implicitly, by the definition of capacity, such codes have large blocklength. Moreover, due to the potential simplifications and because for large blocklength such codes do behave very well, conventional coding theory often restricts itself to *linear codes*. It is also quite common to use the *minimum Hamming distance* and the *weight enumerating function (WEF)* as a design and quality criterion [2]. This is motivated by the equivalence of Hamming weight and Hamming distance for linear codes, and by the union bound that converts the global error probability into pairwise error probabilities.

In this work we would like to break away from these traditional simplifications and instead focus on an optimal[1] design of codes for finite blocklength. Since for very short blocklength, it is not realistic to transmit large quantities of information, we start by looking at codes with only a few codewords, so called *ultra-small block-codes*. Such codes have many practical applications, e.g., in the situation of establishing an initial connection in a wireless link. There the amount of information that needs to be transmitted during the setup of the link is limited to usually only a couple of bits, however, these bits need to be transmitted in very short time (e.g., blocklength in the range of $n = 20$ to $n = 30$) with the highest possible reliability [3].

While conventional coding theory in the sense of Shannon often focuses on stating important fundamental insights and properties like, e.g., what rates are possible to achieve and what rates are not achievable, we specifically turn our attention to the concrete *code design*, i.e., we are interested in actually finding a globally optimum code for a certain given channel and a given fixed blocklength.

In this paper, we introduce a new class of codes, called *fair weak flip codes*, that have many beautiful properties similar to those of linear codes. However, while linear codes are very much limited since they only can exist if the number of codewords $\mathsf{M}$ happens to be an integer-power of 2, our class of codes exists for arbitrary $\mathsf{M}$. We will investigate these "quasi-linear" codes and show that they satisfy the Plotkin bound.

Fair weak flip codes are related to a class of binary nonlinear codes that are constructed with the help of Hadamard matrices and Levenshtein's theorem [4, Ch. 2]. These *binary nonlinear Hadamard codes* also meet the Plotkin bound. As a matter of fact, if for the parameters $(\mathsf{M}, n)$ of a given fair weak flip code there exists a Hadamard code, then these two codes are equivalent.[2] In this sense we can consider the fair weak flip codes to be a subclass of Hadamard codes. However, note that there is no guarantee that for every choice of parameters $(\mathsf{M}, n)$ for which fair weak flip codes exist, there also exists a corresponding Hadamard code.

Moreover, also note that while Levenshtein's method is only concerned with an optimal Hamming distance structure, we will show that fair weak flip codes are globally optimal (i.e., they are the best with respect to error probability and not only pairwise Hamming distance, and they are best among *all* codes, linear or nonlinear!) for the *binary erasure channel (BEC)*. We prove this optimality in the case of the number of codewords $\mathsf{M} \leq 6$ and conjecture it for $\mathsf{M} > 6$.

We also define a class of codes called *weak flip codes* that

[1] With *optimal* we always mean *minimum error probability*.

[2] For a precise definition of *equivalent* see Remark 4.

contains as special cases the class of fair weak flip codes, the class of binary nonlinear Hadamard codes, and the class of linear codes. We then specify some weak flip codes that are optimal for the BEC for $\mathsf{M} \le 4$ and for *any* finite blocklength $n$, or for $\mathsf{M} = 5$ and for blocklength $n$ satisfying $n \bmod 10 \in \{0, 3, 5, 7, 9\}$, or for $\mathsf{M} = 6$ and for even blocklength $n$.

This work is a continuation of our previous work [5], [6], where we have studied ultra-small block-codes for the situation of general binary-input binary-output channels and where we have derived the optimal code design for the two special cases of the *Z-channel (ZC)* and the *binary symmetric channel (BSC)*. We will also briefly compare our findings here with these previous results.

The foundations of our insights lie in a new very powerful way of creating and analyzing both linear and nonlinear block-codes. As is quite common, we use the *codebook matrix* containing the codewords in its rows to describe our codes. However, for our code construction and performance analysis, we look at this codebook matrix not row-wise, but *column-wise*. All our proofs and also our definition of the new "quasi-linear" codes are fully based on this new approach to a code. (This is another fundamental difference between our results and the binary nonlinear Hadamard codes that are constructed based on Hadamard matrices and Levenshtein's theorem [4].)

The remainder of this paper is structured as follows. After some comments about our notation, we will introduce the channel model and review some common definitions in Sections II and III. In Section IV we introduce the new family of *weak flip codes*, that also contains the subfamily of *fair weak flip codes*. The main results are then summarized and discussed in Section V.

As it is common in coding theory, vectors (denoted by boldface Roman letters, e.g., $\mathbf{x}$) are row-vectors. However, for simplicity of notation and to avoid a large number of transpose-signs, we slightly misuse this notational convention for one special case: any vector $\mathbf{c}$ is a column-vector. It should be always clear from the context because these vectors are used to build codebook matrices and are therefore also conceptually quite different from the transmitted codeword $\mathbf{x}$ or the received sequence $\mathbf{y}$. Moreover, we use a bar $\bar{\mathbf{x}}$ to denote the flipped version of $\mathbf{x}$, i.e., $\bar{\mathbf{x}} \triangleq \mathbf{x} \oplus \mathbf{1}$ (where $\oplus$ denotes the componentwise XOR operation).

## II. Channel Model and Coding Schemes

We consider the binary erasure channel (BEC) given in Figure 1. The BEC is a discrete memoryless channel (DMC) with binary input $X$ and ternary output $Y$ and with a conditional channel probability

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

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

We next quickly review a few common definitions.

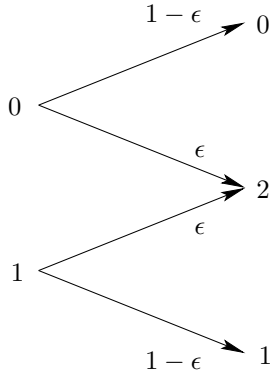


Figure 1. The binary erasure channel (BEC) with erasure probability $\epsilon$. The channel output 2 corresponds to an erasure.

*Definition 1:* An $(\mathsf{M}, n)$ *coding scheme* for a DMC (such as the BEC) consists of a codebook $\mathscr{C}^{(\mathsf{M},n)}$ with $\mathsf{M}$ codewords of length $n$, an encoder that maps every message $m$ into its corresponding codeword $\mathbf{x}_m$, and a decoder that makes a decoding decision $g(\mathbf{y}) \in \{1, \ldots, \mathsf{M}\}$ for every received binary $n$-vector $\mathbf{y}$.

We will always assume that the $\mathsf{M}$ possible messages are equally likely and that the decoder is a *maximum likelihood (ML) decoder:*[3]

$$g(\mathbf{y}) \triangleq \operatorname*{argmax}_{1 \le m \le \mathsf{M}} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}_m). \tag{2}$$

Hence, we are going to be lazy and directly concentrate on the set of codewords $\mathscr{C}^{(\mathsf{M},n)}$, called $(\mathsf{M}, n)$ *codebook* or usually simply $(\mathsf{M}, n)$ *code*. Sometimes we follow the custom of traditional coding theory and use three parameters: $(\mathsf{M}, n, d)$ *code*, where the third parameter $d$ denotes the *minimum Hamming distance*, i.e., the minimum number of components in which any two codewords differ.

*Definition 2:* Given that message $m$ has been sent, let $\lambda_m^{(n)}$ be the *probability of a decoding error* of an $(\mathsf{M}, n)$ code with blocklength $n$:

$$\lambda_m^{(n)} \triangleq \Pr[\, g(\mathbf{Y}) \neq m \mid \mathbf{X} = \mathbf{x}_m \,]. \tag{3}$$

The *average error probability* $P_{\mathrm{e}}^{(n)}$ of an $(\mathsf{M}, n)$ code is defined as

$$P_{\mathrm{e}}^{(n)} = P_{\mathrm{e}}^{(n)}\big(\mathscr{C}^{(\mathsf{M},n)}\big) \triangleq \frac{1}{\mathsf{M}} \sum_{m=1}^{\mathsf{M}} \lambda_m^{(n)}. \tag{4}$$

Sometimes it will be more convenient to focus on the probability of not making any error, denoted *success probability* $\psi_m^{(n)}$:

$$\psi_m^{(n)} \triangleq \Pr[\, g(\mathbf{Y}) = m \mid \mathbf{X} = \mathbf{x}_m \,]. \tag{5}$$

The definition of the *average success probability*[4] $P_{\mathrm{c}}^{(n)}$ follows accordingly.

[3]Note that the ML decoder is optimal in the sense that for a given code and DMC and under the assumption of equally likely messages, it minimizes the average error probability as defined in (4).

[4]The subscript "c" stands for "correct."

*Definition 3:* For a given code $\mathscr{C}^{(\mathsf{M},n)}$, we define the *decoding region* $\mathcal{D}_m$ corresponding to the $m$th codeword $\mathbf{x}_m$ as follows:

$$\mathcal{D}_m \triangleq \{\mathbf{y}\colon g(\mathbf{y}) = m\}. \tag{6}$$

Usually, the codebook $\mathscr{C}^{(\mathsf{M},n)}$ is written as an $\mathsf{M} \times n$ *codebook matrix* with the $\mathsf{M}$ rows corresponding to the $\mathsf{M}$ codewords:

$$\mathscr{C}^{(\mathsf{M},n)} = \begin{pmatrix} — \ \mathbf{x}_1 \ — \\ \vdots \\ — \ \mathbf{x}_{\mathsf{M}} \ — \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ | & | & & | \end{pmatrix}. \tag{7}$$

However, it turns out to be much more convenient to consider the codebook *column-wise* rather than row-wise! We denote the column-vectors of the codebook by $\mathbf{c}$.

*Remark 4:* Since we assume equally likely messages, any permutation of rows only changes the assignment of codewords to messages and has therefore no impact on the performance. We thus consider two codes with permuted rows as being *equal* (this agrees with the thinking of a code being a *set* of codewords, where the ordering of the codewords is irrelevant). Furthermore, since we only consider memoryless channels, any permutation of the columns of $\mathscr{C}^{(\mathsf{M},n)}$ will lead to another code that will result in the same error probability. We say that such two codes are *equivalent*. We would like to emphasize that two codes being equivalent is not the same as two codes being equal. However, as we are mainly interested in the performance of a code, we usually treat two equivalent codes as being the same.

Due to the symmetry of the BEC, we have an additional equivalence in the codebook design.

*Lemma 5:* Consider an arbitrary code $\mathscr{C}^{(\mathsf{M},n)}$ to be used on the BEC and consider an arbitrary $\mathsf{M}$-vector $\mathbf{c}$. Now construct a new length-$(n+1)$ code $\mathscr{C}^{(\mathsf{M},n+1)}$ by appending $\mathbf{c}$ to the codebook matrix of $\mathscr{C}^{(\mathsf{M},n)}$ and a new length-$(n+1)$ code $\overline{\mathscr{C}}^{(\mathsf{M},n+1)}$ by appending the flipped vector $\bar{\mathbf{c}} = \mathbf{c} \oplus \mathbf{1}$ to the codebook matrix of $\mathscr{C}^{(\mathsf{M},n)}$. Then the performance of these two new codes is identical:

$$P_{\mathrm{e}}^{(n+1)}\big(\mathscr{C}^{(\mathsf{M},n+1)}\big) = P_{\mathrm{e}}^{(n+1)}\Big(\overline{\mathscr{C}}^{(\mathsf{M},n+1)}\Big). \tag{8}$$

We remind the reader that our ultimate goal is to find the structure of an optimal code $\mathscr{C}^{(\mathsf{M},n)*}$ that satisfies

$$P_{\mathrm{e}}^{(n)}\big(\mathscr{C}^{(\mathsf{M},n)*}\big) \le P_{\mathrm{e}}^{(n)}\big(\mathscr{C}^{(\mathsf{M},n)}\big) \tag{9}$$

for any code $\mathscr{C}^{(\mathsf{M},n)}$.

## III. Pairwise Hamming Distance

The minimum Hamming distance is a well-known and often used quality criterion of a code. Unfortunately, a design based on the minimum Hamming distance can be strictly suboptimal even for a very symmetric channel like the BSC and even for linear codes, although the error probability performance of a BSC is completely specified by the Hamming distances between codewords and received vectors [6].

We therefore define a slightly more general and more concise description of a code: the *pairwise Hamming distance vector*.

*Definition 6:* Given a code $\mathscr{C}^{(\mathsf{M},n)}$ with codewords $\mathbf{x}_m$, we define the *pairwise Hamming distance vector* $\mathbf{d}^{(\mathsf{M},n)}$ of length $\frac{(\mathsf{M}-1)\mathsf{M}}{2}$ as

$$\mathbf{d}^{(\mathsf{M},n)} \triangleq \Big(d_{12}^{(n)},\ d_{13}^{(n)}, d_{23}^{(n)},\ d_{14}^{(n)}, d_{24}^{(n)}, d_{34}^{(n)}, \ldots, d_{1\mathsf{M}}^{(n)}, d_{2\mathsf{M}}^{(n)}, \ldots, d_{(\mathsf{M}-1)\mathsf{M}}^{(n)}\Big) \tag{10}$$

with $d_{mm'}^{(n)} \triangleq d_{\mathrm{H}}(\mathbf{x}_m, \mathbf{x}_{m'})$, $1 \le m < m' \le \mathsf{M}$, where $d_{\mathrm{H}}(\cdot,\cdot)$ is the well known Hamming distance function. The *minimum Hamming distance* $d_{\min}$ is defined as the minimum component of the pairwise Hamming distance vector $\mathbf{d}^{(\mathsf{M},n)}$.

## IV. Weak Flip Codes and Hadamard Codes

We next introduce some special families of binary codes. We start with a family of codes with two codewords.

*Definition 7:* The *flip code of type* $t$ for $t \in \big\{0, 1, \ldots, \lfloor \frac{n}{2} \rfloor\big\}$ is a code with $\mathsf{M} = 2$ codewords defined by the following codebook matrix $\mathscr{C}_t^{(2,n)}$:

$$\mathscr{C}_t^{(2,n)} \triangleq \begin{pmatrix} \mathbf{x} \\ \bar{\mathbf{x}} \end{pmatrix} = \begin{pmatrix} 0 & \cdots & 0 & \overbrace{1 & \cdots & 1}^{t \text{ columns}} \\ 1 & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}. \tag{11}$$

Defining the column vectors

$$\left\{\mathbf{c}_1^{(2)} \triangleq \begin{pmatrix} 0 \\ 1 \end{pmatrix},\ \mathbf{c}_2^{(2)} \triangleq \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}, \tag{12}$$

we see that a flip code of type $t$ is given by a codebook matrix that consists of $n - t$ columns $\mathbf{c}_1^{(2)}$ and $t$ columns $\mathbf{c}_2^{(2)}$.

Note that while the flip code of type 0 corresponds to a repetition code, the general flip code of type $t$ with $t > 0$ neither is a repetition code nor is it even linear.

We have shown in [6] that for any blocklength $n$ and for a correct choice[5] of $t$, the flip codes are optimal on *any* binary-input binary-output channel for arbitrary channel parameters. In particular, they are optimal for the BSC and the ZC [6].

The columns given by the set (12) are called *candidate columns*. We see that they are flipped versions of each other, therefore also the name of the code.

To be able to generalize the definition of flip codes to $\mathsf{M} > 2$, we give the following definition.

*Definition 8:* Given an $\mathsf{M} > 2$, a length-$\mathsf{M}$ candidate column $\mathbf{c}$ is called a *weak flip column* if its first component is 0 and its Hamming weight equals to $\lfloor \frac{\mathsf{M}}{2} \rfloor$ 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 $\mathcal{C}^{(\mathsf{M})}$.

We see that a weak flip column contains an almost equal number of zeros and ones. The restriction of the first component to be zero is based on the insight of Lemma 5. For

[5] We would like to emphasize that the optimal choice of $t$ for many binary channels is not 0, i.e., the linear repetition code is not optimal!

the remainder of this paper, we introduce the shorthand

$$\ell \triangleq \left\lceil \frac{\mathsf{M}}{2} \right\rceil. \tag{13}$$

*Lemma 9:* The cardinality of a weak flip candidate columns set is

$$\left|\mathcal{C}^{(\mathsf{M})}\right| = \binom{2\ell-1}{\ell}. \tag{14}$$

We are now ready to generalize Definition 7.

*Definition 10:* A *weak flip code* is a codebook that is constructed only by weak flip columns.

Concretely, for $\mathsf{M}=3$ or $\mathsf{M}=4$, we have the following.

*Definition 11:* The *weak flip code of type* $(t_2,t_3)$ for $\mathsf{M}=3$ or $\mathsf{M}=4$ codewords is defined by a codebook matrix $\mathscr{C}_{t_2,t_3}^{(\mathsf{M},n)}$ that consists of $t_1 \triangleq n-t_2-t_3$ columns $\mathbf{c}_1^{(\mathsf{M})}$, $t_2$ columns $\mathbf{c}_2^{(\mathsf{M})}$, and $t_3$ columns $\mathbf{c}_3^{(\mathsf{M})}$, where

$$\left\{ \mathbf{c}_1^{(3)} \triangleq \begin{pmatrix}0\\0\\1\end{pmatrix},\ \mathbf{c}_2^{(3)} \triangleq \begin{pmatrix}0\\1\\0\end{pmatrix},\ \mathbf{c}_3^{(3)} \triangleq \begin{pmatrix}0\\1\\1\end{pmatrix} \right\} \tag{15}$$

or

$$\left\{ \mathbf{c}_1^{(4)} \triangleq \begin{pmatrix}0\\0\\1\\1\end{pmatrix},\ \mathbf{c}_2^{(4)} \triangleq \begin{pmatrix}0\\1\\0\\1\end{pmatrix},\ \mathbf{c}_3^{(4)} \triangleq \begin{pmatrix}0\\1\\1\\0\end{pmatrix} \right\}, \tag{16}$$

respectively. We often describe a weak flip code of type $(t_2,t_3)$ by the *code parameters* $[t_1,t_2,t_3]$.

*Lemma 12:* The pairwise Hamming distance vector of a weak flip code of type $(t_2,t_3)$ can be computed as follows:

$$\mathbf{d}^{(3,n)} = (t_2+t_3, t_1+t_3, t_1+t_2),$$
$$\mathbf{d}^{(4,n)} = (t_2+t_3, t_1+t_3, t_1+t_2, t_1+t_2, t_1+t_3, t_2+t_3).$$

A similar definition can be given also for larger $\mathsf{M}$, however, one needs to be aware that the number of weak flip candidate columns is increasing fast. For $\mathsf{M}=5$ or $\mathsf{M}=6$ we have ten weak flip candidate columns:

$$\left\{ \mathbf{c}_1^{(5)} \triangleq \begin{pmatrix}0\\0\\0\\1\\1\end{pmatrix},\ \mathbf{c}_2^{(5)} \triangleq \begin{pmatrix}0\\0\\1\\0\\1\end{pmatrix},\ \mathbf{c}_3^{(5)} \triangleq \begin{pmatrix}0\\0\\1\\1\\0\end{pmatrix},\right.$$
$$\mathbf{c}_4^{(5)} \triangleq \begin{pmatrix}0\\0\\1\\1\\1\end{pmatrix}, \mathbf{c}_5^{(5)} \triangleq \begin{pmatrix}0\\1\\0\\0\\1\end{pmatrix},\ \mathbf{c}_6^{(5)} \triangleq \begin{pmatrix}0\\1\\0\\1\\0\end{pmatrix},\ \mathbf{c}_7^{(5)} \triangleq \begin{pmatrix}0\\1\\0\\1\\1\end{pmatrix},$$
$$\left. \mathbf{c}_8^{(5)} \triangleq \begin{pmatrix}0\\1\\1\\0\\0\end{pmatrix},\ \mathbf{c}_9^{(5)} \triangleq \begin{pmatrix}0\\1\\1\\0\\1\end{pmatrix},\ \mathbf{c}_{10}^{(5)} \triangleq \begin{pmatrix}0\\1\\1\\1\\0\end{pmatrix} \right\}, \tag{17}$$

and

$$\left\{ \mathbf{c}_1^{(6)} \triangleq \begin{pmatrix}0\\0\\0\\1\\1\\1\end{pmatrix},\ \mathbf{c}_2^{(6)} \triangleq \begin{pmatrix}0\\0\\1\\0\\1\\1\end{pmatrix},\ \mathbf{c}_3^{(6)} \triangleq \begin{pmatrix}0\\0\\1\\1\\0\\1\end{pmatrix},\right.$$
$$\mathbf{c}_4^{(6)} \triangleq \begin{pmatrix}0\\0\\1\\1\\1\\0\end{pmatrix}, \mathbf{c}_5^{(6)} \triangleq \begin{pmatrix}0\\1\\0\\0\\1\\1\end{pmatrix},\ \mathbf{c}_6^{(6)} \triangleq \begin{pmatrix}0\\1\\0\\1\\0\\1\end{pmatrix},\ \mathbf{c}_7^{(6)} \triangleq \begin{pmatrix}0\\1\\0\\1\\1\\0\end{pmatrix},$$
$$\left. \mathbf{c}_8^{(6)} \triangleq \begin{pmatrix}0\\1\\1\\0\\0\\1\end{pmatrix},\ \mathbf{c}_9^{(6)} \triangleq \begin{pmatrix}0\\1\\1\\0\\1\\0\end{pmatrix},\ \mathbf{c}_{10}^{(6)} \triangleq \begin{pmatrix}0\\1\\1\\1\\0\\0\end{pmatrix} \right\}, \tag{18}$$

respectively.

We will next introduce a special subclass of weak flip codes that, as we will see in Section V, possess particularly beautiful properties.

*Definition 13:* A weak flip code is called *fair* if it is constructed by an equal number of all possible weak flip candidate columns in $\mathcal{C}^{(\mathsf{M})}$. Note that by definition the blocklength of a fair weak flip code is always a multiple of $\binom{2\ell-1}{\ell}$, $\ell \geq 2$.

Fair weak flip codes have been used by Shannon *et al.* [7] for the derivation of error exponents, although the codes were not named at that time. Note that in their derivation, the error exponents are defined when blocklength $n$ goes to infinity, but in this work we consider finite $n$.

Related to the weak flip codes and the fair weak flip codes are the families of *Hadamard codes* [4, Ch. 2].

*Definition 14:* For an even integer $n$, a (*normalized*) *Hadamard matrix* $\mathsf{H}_n$ of order $n$ is an $n \times n$ matrix with entries $+1$ and $-1$ and with the first row and column being all $+1$, such that

$$\mathsf{H}_n \mathsf{H}_n^{\mathsf{T}} = n\mathsf{I}_n, \tag{19}$$

if such a matrix exists. Here $\mathsf{I}_n$ is the identity matrix of size $n$. If the entries $+1$ are replaced by 0 and the entries $-1$ by 1, $\mathsf{H}_n$ is changed into the *binary Hadamard matrix* $\mathsf{A}_n$.

Note that a necessary (but not sufficient) condition for the existence of $\mathsf{H}_n$ (and the corresponding $\mathsf{A}_n$) is that $n$ is 1, 2, or a multiple of 4 [4, Ch. 2].

*Definition 15:* The binary Hadamard matrix $\mathsf{A}_n$ gives rise to three families of Hadamard codes:

1) The $\left(n, n-1, \frac{n}{2}\right)$ *Hadamard code* $\mathscr{H}_{1,n}$ consists of the rows of $\mathsf{A}_n$ with the first column deleted. The codewords in $\mathscr{H}_{1,n}$ that begin with 0 form the $\left(\frac{n}{2}, n-2, \frac{n}{2}\right)$ *Hadamard code* $\mathscr{H}'_{1,n}$ if the initial zero is deleted.

2) The $\left(2n, n-1, \frac{n}{2}-1\right)$ *Hadamard code* $\mathscr{H}_{2,n}$ consists of $\mathscr{H}_{1,n}$ together with the complements of all its codewords.
3) The $\left(2n, n, \frac{n}{2}\right)$ *Hadamard code* $\mathscr{H}_{3,n}$ consists of the rows of $\mathsf{A}_n$ and their complements.

Further Hadamard codes can be created by an arbitrary combinations of the codebook matrices of different Hadamard codes.

*Example 16:* Consider an $(8, 7, 4)$ $\mathscr{H}_{1,8}$ code:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{20}$$

From this code, an $(8, 35, 20)$ Hadamard code can be constructed by simply concatenating $\mathscr{H}_{1,8}$ five times. $\Diamond$

Note that since the rows of $\mathsf{H}_n$ are orthogonal, any two rows of $\mathsf{A}_n$ agree in $\frac{1}{2}n$ places and differ in $\frac{1}{2}n$ places, i.e., they have a Hamming distance $\frac{n}{2}$. Moreover, by definition the first row of a binary Hadamard matrix is the all-zero row. Hence, we see that all Hadamard codes are weak flip codes, i.e., the family of weak flip codes is a superset of Hadamard codes.

On the other hand, fair weak flip codes can be seen as a "subset" of Hadamard codes because for all parameters $(\mathsf{M}, n)$, for which fair weak flip codes and also Hadamard codes exist, a Hadamard code can be constructed that is also a fair weak flip code. The problem with this statement lies in the fact that the Hadamard codes rely on the existence of Hadamard matrices, which in general is not guaranteed. So it is very difficult to predict whether for a given pair $(\mathsf{M}, n)$, a Hadamard code will exist or not. This is in stark contrast to weak flip codes (which exist for all $\mathsf{M}$ and $n$) and fair weak flip codes (which exist for all $\mathsf{M}$ and all $n$ being a multiple of $\binom{2\ell-1}{\ell}$).

We also remark that a Hadamard code of parameters $(\mathsf{M}, n)$, for which fair weak flip codes exist, is not necessarily equivalent to a fair weak flip code.

*Example 17:* We continue with Example 16 and note that the $(8, 35, 20)$ Hadamard code that is constructed by five repetitions of the matrix given in (20) is actually not a fair weak flip code since we have not used all possible weak flip candidate columns. However, it is possible to find five $(8, 7, 4)$ Hadamard codes that combine to a $(8, 35, 20)$ fair weak flip code. $\Diamond$

Note that two Hadamard matrices can be *equivalent* if one can be obtained from the other by permuting rows and columns and multiplying rows and columns by $-1$. In other words, Hadamard codes can actually be constructed from weak candidate columns. This also follows directly from the already mentioned fact that Hadamard codes are weak flip codes.

## V. Main Results

### A. Characteristics of Weak Flip Codes

In conventional coding theory, most results are restricted to so called *linear codes* that possess very powerful algebraic properties. For the following definitions see, e.g., [2], [4].

*Definition 18:* Let $\mathsf{M} = 2^k$, where $k \in \mathbb{N}$. The binary code $\mathscr{C}_{\text{lin}}^{(\mathsf{M},n)}$ is *linear* if its codewords span a $k$-dimensional subspace of $\{0,1\}^n$.

One of the most important property of a linear code is as follows.

*Proposition 19:* Let $\mathscr{C}_{\text{lin}}$ be linear and let $\mathbf{x}_m \in \mathscr{C}_{\text{lin}}$ be given. Then the code that we obtain by adding $\mathbf{x}_m$ to each codeword of $\mathscr{C}_{\text{lin}}$ is equal to $\mathscr{C}_{\text{lin}}$.

Another property concerns the column weights.

*Proposition 20:* If an $(\mathsf{M}, n)$ 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.

Hence, linear codes are weak flip codes. Note, however, that linear codes only exist if $\mathsf{M} = 2^k$, where $k \in \mathbb{N}$, while weak flip codes are defined for any $\mathsf{M}$. Also note that the converse of Proposition 20 does not 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. It is not even the case that a fair weak flip code for $\mathsf{M} = 2^k$ is necessarily linear!

Now the question arises as to which of the many powerful algebraic properties of linear codes are retained in weak flip codes.

*Theorem 21:* Consider a weak flip code $\mathscr{C}^{(\mathsf{M},n)}$ and fix some codeword $\mathbf{x}_m \in \mathscr{C}^{(\mathsf{M},n)}$. If we add this codeword to all codewords in $\mathscr{C}^{(\mathsf{M},n)}$, then the resulting code $\tilde{\mathscr{C}}^{(\mathsf{M},n)} \triangleq \left\{\mathbf{x}_m \oplus \mathbf{x} \,\middle|\, \forall\, \mathbf{x} \in \mathscr{C}^{(\mathsf{M},n)}\right\}$ is still a weak flip code, however, it is not necessarily the same one.

Theorem 21 is a beautiful property of weak flip codes; however, it still represents a considerable weakening of the powerful property of linear codes given in Proposition 19. This can be fixed by considering the subfamily of fair weak flip codes.

*Theorem 22 (Quasi-Linear Codes):* Let $\mathscr{C}$ be a *fair* weak flip code and let $\mathbf{x}_m \in \mathscr{C}$ be given. Then the code $\tilde{\mathscr{C}} = \left\{\mathbf{x}_m \oplus \mathbf{x} \,\middle|\, \forall\, \mathbf{x} \in \mathscr{C}^{(\mathsf{M},n)}\right\}$ is equivalent to $\mathscr{C}$.

If we recall Proposition 20 and the discussion after it, we realize that the definition of the quasi-linear fair weak flip code is a considerable enlargement of the set of codes having the property given in Theorem 22.

The following corollary is a direct consequence of Theorem 22.

*Corollary 23:* The Hamming weights of each codeword of a fair weak flip code are all identical except the all-zero

codeword $\mathbf{x}_1$. In other words, if we let $w_{\mathrm{H}}(\cdot)$ be the Hamming weight function, then

$$w_{\mathrm{H}}(\mathbf{x}_2) = w_{\mathrm{H}}(\mathbf{x}_3) = \cdots = w_{\mathrm{H}}(\mathbf{x}_{\mathsf{M}}). \tag{21}$$

Before we next investigate the minimum Hamming distance for the quasi-linear fair weak flip codes, we quickly recall an important bound that holds for any $\big(\mathsf{M}, n, d\big)$ code.

*Lemma 24 (Plotkin Bound [4]):* The minimum distance of an $(\mathsf{M}, n)$ binary code $\mathscr{C}^{(\mathsf{M},n)}$ always satisfies

$$d_{\min}\big(\mathscr{C}^{(\mathsf{M},n)}\big) \leq \begin{cases} \frac{n\cdot\frac{\mathsf{M}}{2}}{\mathsf{M}-1} & \mathsf{M} \text{ even}, \\ \frac{n\cdot\frac{\mathsf{M}+1}{2}}{\mathsf{M}} & \mathsf{M} \text{ odd}. \end{cases} \tag{22}$$

It can be seen that a necessary condition for a codebook to meet the Plotkin-bound is that the codebook is composed by weak flip candidate columns. Furthermore, Levenshtein [4, Ch. 2] proved that the Plotkin bound can be achieved, provided that Hadamard matrices exist.

*Theorem 25:* Fix some $\mathsf{M}$ and a blocklength $n$ with $n \bmod \binom{2\ell-1}{\ell} = 0$. Then a fair weak flip code $\mathscr{C}^{(\mathsf{M},n)}$ achieves the largest minimum Hamming distance among all codes of given blocklength and satisfies

$$d_{\min}\big(\mathscr{C}^{(\mathsf{M},n)}\big) = \frac{n\cdot\ell}{2\ell-1}. \tag{23}$$

### B. Optimal Codes on BEC

The definitions of the flip, the weak flip, and the fair weak flip codes are interesting not only due to their generalization of the concept of linear codes, but also because we can show that they are optimal for the BEC for many values of the blocklength $n$.

*Theorem 26:* For a BEC and for any $n \geq 1$, an optimal codebook with $\mathsf{M} = 2$ codewords is the flip code of type $t$ for any $t \in \big\{0, 1, \ldots, \big\lfloor \frac{n}{2} \big\rfloor\big\}$.

*Theorem 27:* For a BEC and for any $n \geq 2$, an optimal codebook with $\mathsf{M} = 3$ or $\mathsf{M} = 4$ codewords is the weak flip code of type $(t_2^*, t_3^*)$, where

$$t_2^* \triangleq \left\lfloor \frac{n}{3} \right\rfloor, \quad t_3^* \triangleq \left\lfloor \frac{n+1}{3} \right\rfloor. \tag{24}$$

This optimal codebook can be constructed recursively in the blocklength $n$. We start with an optimal codebook for $n = 2$:

$$\mathcal{C}_{\mathrm{BEC}}^{(\mathsf{M},2)*} = \Big(\mathbf{c}_1^{(\mathsf{M})}, \mathbf{c}_3^{(\mathsf{M})}\Big). \tag{25}$$

Then, from the optimal code $\mathcal{C}_{\mathrm{BEC}}^{(\mathsf{M},n-1)*}$ of blocklength $n-1$, we can recursively construct the optimal codebook of blocklength $n$ by appending

$$\begin{cases} \mathbf{c}_2^{(\mathsf{M})} & \text{if } n \bmod 3 = 0, \\ \mathbf{c}_1^{(\mathsf{M})} & \text{if } n \bmod 3 = 1, \\ \mathbf{c}_3^{(\mathsf{M})} & \text{if } n \bmod 3 = 2. \end{cases} \tag{26}$$

This theorem suggests that for a given fixed code size $\mathsf{M}$, a sequence of good codes can be generated by appending proper columns to a code of smaller blocklength. The proof is based on this recursive generation and follows similar ideas as in [6, App. C], i.e., it is based on a column-wise analysis of the codebook matrix and on a mathematical induction on $n$. For a given DMC and a code of blocklength $n$, we ask the question what is the optimal improvement (i.e., the maximum reduction of error probability) when increasing the blocklength $n$ to $n+\gamma$, where $\gamma = 1$ when $\mathsf{M} = 3$ or $4$ (and may be larger than $1$ when $\mathsf{M} \geq 5$). The answer to this question then leads to the recursive construction of (26).

Note that the idea of designing an optimal code recursively promises to be a very powerful approach. Unfortunately, for larger values of $\mathsf{M}$, we might need a recursion from $n$ to $n+\gamma$ with a step-size $\gamma > 1$ that might be a function of blocklength $n$. However, based on our definition of fair weak flip codes and on Theorem 29 below, we conjecture that the necessary step-size satisfies $\gamma \leq \binom{2\ell-1}{\ell}$.

We have successfully applied this recursive approach also to the cases of $\mathsf{M} = 5$ and $\mathsf{M} = 6$.

*Theorem 28:* For a BEC and for any $n \geq 3$, an optimal codebook with $\mathsf{M} = 5$ codewords can be constructed recursively in the blocklength $n$. We start with an optimal codebook for $n = 3$:

$$\mathscr{C}_{\mathrm{BEC}}^{(\mathsf{M},3)*} = \Big(\mathbf{c}_1^{(\mathsf{M})}, \mathbf{c}_2^{(\mathsf{M})}, \mathbf{c}_5^{(\mathsf{M})}\Big) \tag{27}$$

and recursively construct the optimal codebook for $n \geq 5$ by using $\mathscr{C}_{\mathrm{BEC}}^{(\mathsf{M},n-\gamma)*}$, $\gamma \in \{1,2,3\}$, and appending

$$\begin{cases} \Big(\mathbf{c}_1^{(\mathsf{M})}, \mathbf{c}_2^{(\mathsf{M})}, \mathbf{c}_5^{(\mathsf{M})}\Big) & \text{if } n \bmod 10 = 3, \\ \Big(\mathbf{c}_3^{(\mathsf{M})}, \mathbf{c}_6^{(\mathsf{M})}\Big) & \text{if } n \bmod 10 = 5, \\ \Big(\mathbf{c}_9^{(\mathsf{M})}, \mathbf{c}_{10}^{(\mathsf{M})}\Big) & \text{if } n \bmod 10 = 7, \\ \Big(\mathbf{c}_4^{(\mathsf{M})}, \mathbf{c}_7^{(\mathsf{M})}\Big) & \text{if } n \bmod 10 = 9, \\ \mathbf{c}_8^{(\mathsf{M})} & \text{if } n \bmod 10 = 0. \end{cases} \tag{28}$$

For $\mathsf{M} = 6$ codewords, an optimal codebook can be constructed recursively in the blocklength $n$ by starting with an optimal codebook for $n = 4$:

$$\mathscr{C}_{\mathrm{BEC}}^{(\mathsf{M},3)*} = \Big(\mathbf{c}_1^{(\mathsf{M})}, \mathbf{c}_2^{(\mathsf{M})}, \mathbf{c}_6^{(\mathsf{M})}, \mathbf{c}_8^{(\mathsf{M})}\Big). \tag{29}$$

Then we recursively construct the optimal codebook for $n \geq 6$ by using $\mathscr{C}_{\text{BEC}}^{(\mathsf{M},n-2)*}$ and appending

$$\begin{cases} \left(\mathbf{c}_1^{(\mathsf{M})}, \mathbf{c}_2^{(\mathsf{M})}\right) & \text{if } n \bmod 10 = 2, \\ \left(\mathbf{c}_6^{(\mathsf{M})}, \mathbf{c}_8^{(\mathsf{M})}\right) & \text{if } n \bmod 10 = 4, \\ \left((\mathbf{c}_3^{(\mathsf{M})}, \mathbf{c}_5^{(\mathsf{M})}\right) & \text{if } n \bmod 10 = 6, \\ \left(\mathbf{c}_4^{(\mathsf{M})}, \mathbf{c}_7^{(\mathsf{M})}\right) & \text{if } n \bmod 10 = 8, \\ \left(\mathbf{c}_9^{(\mathsf{M})}, \mathbf{c}_{10}^{(\mathsf{M})}\right) & \text{if } n \bmod 10 = 0. \end{cases} \tag{30}$$

An interesting special case of Theorem 28 is as follows.

*Theorem 29:* For a BEC and for any $n$ being a multiple of 10, an optimal codebook with $\mathsf{M} = 5$ or $\mathsf{M} = 6$ codewords is the corresponding fair weak flip code.

Note that the restriction on $n$ comes from the restriction that fair weak flip codes are only defined for $n$ with $n \bmod \binom{2\ell-1}{\ell} = n \bmod 10 = 0$. Even though Theorem 29 actually follows as special case from Theorem 28, it can be proven directly and more elegantly using the properties of fair weak flip codes derived in Section V-A.

How about the optimal codes on BEC for higher number of codewords $\mathsf{M}$? We strongly believe that Theorem 29 can be generalized to arbitrary $\mathsf{M}$.

*Conjecture 30:* For a BEC and for an arbitrary $\mathsf{M}$, the optimal code for a blocklength $n$ that satisfies $n \bmod \binom{2\ell-1}{\ell} = 0$ is the corresponding fair weak flip code.

### C. Quick Comparison between BSC and BEC

In [6] it has been shown that optimal codes for $\mathsf{M} = 3$ or $\mathsf{M} = 4$ are weak flip codes with code parameters:

$$[t_1^*, t_2^*, t_3^*] = \begin{cases} [k+1, k-1, k] & \text{if } n \bmod 3 = 0, \\ [k+1, k, k] & \text{if } n \bmod 3 = 1, \\ [k+1, k, k+1] & \text{if } n \bmod 3 = 2, \end{cases} \tag{31}$$

where we use

$$k \triangleq \left\lfloor \frac{n}{3} \right\rfloor. \tag{32}$$

The corresponding pairwise Hamming distance vectors (see Lemma 12) are

$$\begin{cases} (2k-1, 2k, 2k+1) & \text{if } n \bmod 3 = 0, \\ (2k, 2k+1, 2k+1) & \text{if } n \bmod 3 = 1, \\ (2k+1, 2k+2, 2k+1) & \text{if } n \bmod 3 = 2. \end{cases} \tag{33}$$

If we compare this to Theorem 27:

$$[t_1^*, t_2^*, t_3^*] = \begin{cases} [k, k, k] & \text{if } n \bmod 3 = 0, \\ [k+1, k, k] & \text{if } n \bmod 3 = 1, \\ [k+1, k, k+1] & \text{if } n \bmod 3 = 2 \end{cases} \tag{34}$$

with corresponding pairwise Hamming distance vectors

$$\begin{cases} (2k, 2k, 2k) & \text{if } n \bmod 3 = 0, \\ (2k, 2k+1, 2k+1) & \text{if } n \bmod 3 = 1, \\ (2k+1, 2k+2, 2k+1) & \text{if } n \bmod 3 = 2, \end{cases} \tag{35}$$

we can conclude the following.

*Corollary 31:* Apart from $n \bmod 3 = 0$, the optimal codes for a BSC are identical to the optimal codes for a BEC for $\mathsf{M} = 3$ or $\mathsf{M} = 4$ codewords.

It is interesting to note that for $n \bmod 3 = 0$ the optimal codes for the BEC are fair and therefore maximize the minimum Hamming distance, while this is not the case for the (very symmetric!) BSC. However, note that the converse is *not* true: if a code maximizes the minimum Hamming distance, then it is not necessarily an optimal code for the BEC! So, in particular, it is not clear if binary nonlinear Hadamard codes are optimal.

## VI. Conclusion

In this paper, we have introduced the *weak flip codes*, a new class of codes containing both the class of binary nonlinear Hadamard codes and the class of linear codes as special cases. We have shown that weak flip codes have many desirable properties; in particular, we have been able to prove that besides retaining many of the good Hamming distance properties of Hadamard codes, they are actually optimal with respect to the minimum error probability over a binary erasure channel (BEC) for certain numbers of codewords $\mathsf{M}$ and many finite blocklengths $n$.

We have also introduced the subclass of *fair weak flip codes* that can be seen as a generalization of linear codes to arbitrary numbers of codewords $\mathsf{M}$. We have shown that fair weak flip codes are optimal with respect to the error probability for the BEC for $\mathsf{M} \leq 6$ and a blocklength that depends on $\mathsf{M}$, and we have conjectured that this result continues to hold also for $\mathsf{M} > 6$.

Note that while it has been known for quite some time that binary nonlinear Hadamard codes have good Hamming distance properties [4], so far not much has been known about their behavior with respect to error probability for finite $n$. Furthermore, also note that while fair weak flip codes have been used before (although without being named) in the derivation of results related to error probability [7], so far it has only been shown that the optimal error exponents can be achieved by fair weak flip codes, but they have not been proven to be actually optimal in error probability among all possible linear or nonlinear codes for finite blocklength.

In conclusion, we try to build a bridge between coding theory, which usually is concerned with the design of codes with good Hamming distance properties (like, e.g., the binary

nonlinear Hadamard codes), and information theory, which deals with error probability and the existence of codes that have good or optimal error probability behavior.

## 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 System Techn. J.*, vol. 27, pp. 379–423 and 623–656, Jul. and Oct. 1948.

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

[3] C.-L. Wu, P.-N. Chen, Y. S. Han, and Y.-X. Zheng, "On the coding scheme for joint channel estimation and error correction over block fading channels," in *Proc. IEEE Int. Symp. Pers., Indoor and Mob. Radio Commun.*, Tokyo, Japan, Sep. 13–16, 2009, pp. 1272–1276.

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

[5] P.-N. Chen, H.-Y. Lin, and S. M. Moser, "Ultra-small block-codes for binary discrete memoryless channels," in *Proc. IEEE Inf. Theory Workshop*, Paraty, Brazil, Oct. 16–20, 2011, pp. 175–179.

[6] ——, "Optimal ultra-small block-codes for binary discrete memoryless channels," Mar. 2012, subm. to *IEEE Trans. Inf. Theory*. [Online]. Available: http://moser.cm.nctu.edu.tw/publications.html

[7] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels," *Inform. Contr.*, pp. 522–552, May 1967, part II.