Codes for Symbol-Pair Read Channels

Yuval Cassuto, *Member, IEEE,* and Mario Blaum, *Fellow, IEEE*

***Abstract*— A new coding framework is established for channels whose outputs are overlapping pairs of symbols. Such channels are motivated by storage applications in which the spatial resolution of the reader may be insufficient to isolate adjacent symbols. Reading symbols as pairs changes the coding-theoretic error model from the standard bounded number of symbol errors to a bounded number of pair errors. Starting from the most basic coding-theoretic questions, the paper studies codes that protect against pair-errors. It provides answers on pair-error correctability conditions, code construction and decoding, and lower and upper bounds on code sizes. Asymptotic analysis of pair-error correction shows that there exist pair-error codes with rates that are strictly higher than the best known codes in the Hamming metric.**

I. Introduction

The central theme of information theory is to manipulate and reason about information when it is containerized in quanta called symbols. The basic information unit is usually fixed at the outset, and the behavior of these units is examined through channels, processing units, and other liabilities, which are discretized correspondingly. In particular, the theory of error-control codes aims at recovering the original information units when some bound is given on their corruption. These corruption bounds can be defined at the code-block level, like a certain number of errors in Hamming-metric codes, or at the individual-symbol level, like symbol-transition restrictions in asymmetric or uni-directional error-correcting codes. The alphabet on which the information unit is defined may change throughout the coding problem, like in soft-decoding, but still it is typically the same unit that is tracked and analyzed.

There are cases where it is incumbent upon the code designer to depart from the coupling of information units with channel uses. In such cases, physical constraints may enforce representing information as one unit, while the channel or impairment may operate on a different unit. This paper's subject is one such instance that is motivated by the application of high-density data storage technologies. In essence, the model treated here is of information units defined over some discrete symbol alphabet, but whose reading from the channel is performed as (potentially corrupted) overlapping *pairs* of symbols. So while the codes are defined as usual over some q-ary symbol alphabet, their design objective is to protect against a certain number of *pair* errors, rather than a certain number of symbol errors. A pair-error is defined as a pair-read in which one or more of the symbols is read in error. The main motivation for the pair-error model is to address scenarios where, due to physical limitations, individual symbols cannot be read off the channel, and therefore, each channel read contains contributions from two adjacent symbols. Such a scenario can be manifested, for example, when information is written to the surface of storage media by a high-resolution write process (e.g. lithography), and later read by a lower-resolution read process (e.g. a magnetic read head). Such a case is described pictorially in Figure 1. Note that even though

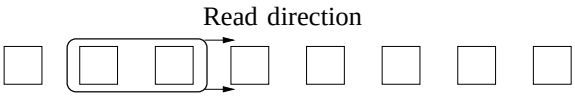

Figure 1. Read process performed in pairs due to resolution deficiency of reader.

the reader is not able to spatially isolate two adjacent symbols, it can still provide hypotheses on the two individual symbols (and not a mixed function thereof) in a single read. The way to resolve the two symbols from one read signal is via modulation at the read head, an example of which is illustrated in Figure 2. In this example, a read head with response that spans two symbols has different gains for the left and right symbols, thereby exploiting amplitude modulation to resolve the two symbols.

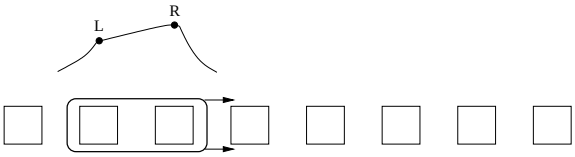

Figure 2. Head response for a symbol pair with different gains. The stronger right (R) symbol can be detected first, then subtracted from the combined signal to detect the weaker left (L) symbol.

Given that in such scenarios two hypotheses are obtained in a single channel use, a single read error event affects either one or both of the symbols in the pair. The overlap between adjacent pair reads suggested by the model is advantageous over the standard method of partitioning the symbols to disjoint pairs, since it provides two observations of each symbol, for the same traversal of the reader over the stored symbols. Even though the main motivation comes from data storage, the pair-channel model may apply to other setups hence we maintain an application-agnostic discussion using standard coding theory terminology. An important question that stood out from the outset is to compare the "strength" of the pair-error model to that of the usual symbol-error model. That is, to compare the achievability of t pair-error correction to that of t symbol-error correction. On one hand, in the pair-error model each symbol is read at two different pairs. On the other hand, errors in two symbols in a pair are counted as a single pair-error. Therefore, our a-priori perception of the pair-error model was as "slightly milder" than the symbol-error model. However, since the minimum pair-distance of a code can be as large as twice its minimum Hamming distance (as explained in

Yuval Cassuto was with Hitachi Global Storage Technologies, San Jose Research Center, California USA, and is now with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Station 14, CH-1015 Lausanne (VD), Switzerland (e-mail: yuval.cassuto@epfl.ch).

Mario Blaum is with the Facultad de Informática, Despacho 431, Universidad Complutense de Madrid (UCM), C/ Profesor José García Santesmases s/n, 28040 Madrid, Spain (e-mail: mario.blaum@fdi.ucm.es).

section II), the gap between the models can potentially be more significant. A partial answer to the model comparison is given in section IV by showing that codes for pair-errors exist with strictly higher asymptotic rates than codes for symbol-errors. Hence, at least in theory, using pair channels may improve the information efficiency of the underlying system.

The proposed model is related to the model of multiple-burst errors, since it treats a pair-error as the error unit, whether one or both of the symbols are received in error. Nevertheless, the model of burst errors still has the property that the transmitted and received units are identical, unlike the pair-error model. While of no direct applicability to the pair-error model, some relevant results from the theory of multiple-burst error correction can be found in [13] and [1]. Another area related to the pair-error model is inter-symbol interference (ISI) mitigation. One can regard pair-error correction as a coding-theoretic framework to handle extreme ISI. Normally ISI is handled at the detector level, providing the error-correcting code decoder with optimal (hard or soft) symbol hypotheses given the ISI. This work goes one step beyond and designs the code itself to mitigate an ISI scenario where each symbol is read combined with either its left or right neighbor. Clearly the pair-error correcting code can too be decoded using soft information from the reader and detector, but exploring this possibility is deferred to a future work.

The initial treatment of pair-errors in this paper pursues some questions that are well known and/or well studied for traditional error-correcting codes. At times the results for the pair-metric are similar to known ones for the Hamming metric, but at other times the behavior of the new metric turns out more surprising and counter-intuitive. In section II, a relevant distance metric is defined, and is used to provide necessary and sufficient conditions for pair-error correctability. In section III, the more "constructive" issues of code construction and decoding are discussed. Some reductions of the construction and decoding problems to the corresponding Hamming-metric problems are examined and concluded to be strictly sub-optimal to direct pair-error correcting code constructions. Accordingly, the theory of cyclic codes is further developed to handle pair-errors. Section IV provides upper and lower bounds on code sizes by combinatorially enumerating spheres in the pair-metric. Asymptotic bounds are obtained by careful manipulation of the combinatorial expressions. The resulting asymptotic expressions prove that codes for pair-errors exist with rates strictly higher than the Gilbert-Varshamov bound, for the same number of corrected errors. This positive result settles (to the affirmative) our initial uncertainty as to whether the slightly milder pair-error model can achieve codes with asymptotically better rates.

II. Symbol-Pair Read Channels

The standard regime treated by coding theory is where decoder inputs are corrupted versions of symbols output by the encoder. So a typical coding problem defines some alphabet Ξ and a block size n; then specifies the code $\mathcal{C} \subset \Xi^n$ and the decoder $dec : \Xi^n \to \mathcal{C}$. This is obviously a very useful framework to combat noisy memoryless channels that arise in a broad variety of communication and storage applications. In some specific applications, however, this coding-problem definition does not adequately describe the constraints of the physical system, in which case a refinement is needed to get better optimized coding solutions. In the current paper, the subject of study are codes over an alphabet Ξ, with block size n, whose decoder inputs are n (potentially corrupted) *pairs* of adjacent code symbols from Ξ. More precisely, the code is defined, as usual, as a subset $\mathcal{C} \subset \Xi^n$, but the decoder is now a function $pair\_dec : (\Xi, \Xi)^n \to \mathcal{C}$. So the inputs to the decoder are n *pairs* of symbols, where each symbol is independently read in two adjacent pairs. To distinguish pair vectors from standard, symbol vectors, we will over-mark them with the symbol $\leftrightarrow$.

$$\overleftrightarrow{u} = [(\triangleleft u_0, \triangleright u_0), \ldots, (\triangleleft u_{n-1}, \triangleright u_{n-1})]$$

In each pair, $\triangleleft$ and $\triangleright$ designate the left and right symbols, respectively. Each symbol vector in Ξ^n can be represented as a symbol-pair vector as defined below.

Definition 1. *(Symbol-Pair Read Vector)*
Let $\boldsymbol{x} = [x_0, \ldots, x_{n-1}]$ be a vector in Ξ^n. The symbol-pair read vector of $\boldsymbol{x}$ is defined as

$$\pi(\boldsymbol{x}) = [(x_0, x_1), (x_1, x_2), \ldots, (x_{n-2}, x_{n-1}), (x_{n-1}, x_0)].$$

Every vector $\boldsymbol{x} \in \Xi^n$ has a pair representation $\pi(\boldsymbol{x}) \in (\Xi, \Xi)^n$. However, not all pair vectors in $(\Xi, \Xi)^n$ have a corresponding vector in Ξ^n, because they may have two different readings of the same symbol in two adjacent pairs. For example, the pair vector

$$[(0,0)\ ,\ \underset{\nwarrow x_1 \nearrow}{(1,0)}, (0,1), (1,1), (1,0)]$$

does not have a corresponding symbol vector since $x_1 = 0$ at the right of the first pair, but $x_1 = 1$ at the left of the second pair. Pair-vectors that have corresponding symbol vectors will be called *consistent*.
We will hereby be interested in cases where some of the read pairs are corrupted versions of the true symbol pairs. The main error model considered for symbol pairs is when the number of pair-errors is bounded by an integer t, defined as *t-pair error* below.

Definition 2. *(t-Pair Error)*
Let $\boldsymbol{x} = [x_0, \ldots, x_{n-1}]$ be a vector in Ξ^n. A pair vector $\overleftrightarrow{u} = [(\triangleleft u_0, \triangleright u_0), \ldots, (\triangleleft u_{n-1}, \triangleright u_{n-1})]$ is the result of a t-pair error from $\boldsymbol{x}$ if $|\{i : (\triangleleft u_i, \triangleright u_i) \neq (x_i, x_{i+1})\}| \leqslant t$. Indices are taken modulo n and $(a, b) = (c, d)$ if both $a = c$ and $b = d$.

The definition of t-pair error above implies that a pair is counted as error whether one or both of its symbols are received in error. Strengthening the error model to allow two errors in a pair to be counted as a single pair-error captures the reader's property of extracting both symbol hypotheses in a pair from the same physical signal.

A. Conditions for symbol-pair error correctability

After defining the symbol-pair error model, the next natural step is to prove necessary and sufficient conditions on the code for achieving correctability of symbol-pair errors. A central element in the characterization of correctability is the *pair distance*, $D_p(\cdot, \cdot)$ defined below.

Definition 3. *(Pair Distance) Let $\overleftrightarrow{u}, \overleftrightarrow{v}$ be two pair-vectors in $(\Xi, \Xi)^n$. The pair distance between $\overleftrightarrow{u}$ and $\overleftrightarrow{v}$ is defined as*

$$D_p\left(\overleftrightarrow{u}, \overleftrightarrow{v}\right) = |\{i : (\triangleleft u_i, \triangleright u_i) \neq (\triangleleft v_i, \triangleright v_i)\}|.$$

So the pair-distance between two pair-vectors counts how many of the n symbol-pairs differ between them, or in other words, the pair-distance is the Hamming distance over the alphabet (Ξ, Ξ). For notational aesthetics, when a consistent pair-vector is used as an argument to the pair distance, its symbol-vector notation may appear instead of its pair-vector one, i.e.

$$D_p(\boldsymbol{x}, \boldsymbol{y}) \triangleq D_p(\pi(\boldsymbol{x}), \pi(\boldsymbol{y})).$$

The pair-distance is related to the Hamming distance in the following manner.

Proposition 1. *For $\boldsymbol{x}, \boldsymbol{y}$ in Ξ^n, let $0 < D_H(\boldsymbol{x}, \boldsymbol{y}) < n$ be the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$. Then*

$$D_H(\boldsymbol{x}, \boldsymbol{y}) + 1 \leqslant D_p(\boldsymbol{x}, \boldsymbol{y}) \leqslant 2D_H(\boldsymbol{x}, \boldsymbol{y}).$$

In the extreme cases in which $D_H(\boldsymbol{x}, \boldsymbol{y})$ equals 0 or n, clearly $D_p(\boldsymbol{x}, \boldsymbol{y}) = D_H(\boldsymbol{x}, \boldsymbol{y})$.

Proof: Define the set $S_H = \{j : x_j \neq y_j\}$. If $D_H(\boldsymbol{x}, \boldsymbol{y}) = d$, then $|S_H| = d$. In addition, define the set $S_p = \{i : (y_i, y_{i+1}) \neq (x_i, x_{i+1})\}$. Each index $j \in S_H$ appears in exactly two pairs of S_p, which gives the upper bound. Since each pair has exactly two indices, it may seem that the tightest lower bound is d and not $d+1$. However, if $d < n$ there is at least one pair with only one of its indices $i, i+1$ in S_H. ■

Note that for the trivial code $\mathcal{C} = \Xi^n$, the minimum pair distance between distinct codewords is 2, hence it can detect a single pair-error.

Proposition 1 can be regarded as a corollary to the following theorem.

Theorem 2. *For two words $\boldsymbol{x}, \boldsymbol{y}$ in Ξ^n with $0 < D_H(\boldsymbol{x}, \boldsymbol{y}) < n$, define the set $S_H = \{j : x_j \neq y_j\}$. Let $S_H = \cup_{l=1}^{L} B_l$ be a minimal partition of the set S_H to subsets of consecutive*[1] *indices (Each subset $B_l = [s_l, e_l]$ is the sequence of all indices between s_l and e_l, inclusive, and L is the smallest integer that achieves such partition). Then*

$$D_p(\boldsymbol{x}, \boldsymbol{y}) = D_H(\boldsymbol{x}, \boldsymbol{y}) + L \tag{1}$$

Proof: The requirement that the partition be minimal guarantees that there are no two adjacent indices $i, i+1$ that belong to different subsets of S_H (otherwise the two subsets can be merged resulting in a smaller partition). Therefore, the pair distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ can be calculated as the sum of the sizes of the pair subsets $\{(s_l - 1, s_l), (s_l, s_l + 1), \ldots, (e_l, e_l + 1)\}$. The number of pairs in each pair subset l equals $|B_l| + 1$, hence the sum equals $\sum_{l=1}^{L} |B_l| + L = D_H(\boldsymbol{x}, \boldsymbol{y}) + L$. ■

Proposition 1 can be obtained from Theorem 2 by noting that $1 \leqslant L \leqslant D_H(\boldsymbol{x}, \boldsymbol{y})$.

The pair-distance shares the following simple properties with the well-studied Hamming distance.

- $D_p(\boldsymbol{x}, \boldsymbol{y}) \geqslant 0$; with $D_p(\boldsymbol{x}, \boldsymbol{y}) = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{y}$
- $D_p(\boldsymbol{x}, \boldsymbol{y}) = D_p(\boldsymbol{y}, \boldsymbol{x})$ (symmetry).
- $D_p(\boldsymbol{x}, \boldsymbol{y}) \leqslant D_p(\boldsymbol{x}, \overleftrightarrow{u}) + D_p(\overleftrightarrow{u}, \boldsymbol{y})$ (triangle inequality).

Hence the set Ξ^n with the pair-distance is a *metric space*. The first two properties are obvious. To prove the triangle inequality, observe that if for some i we have $(x_i, x_{i+1}) \neq (y_i, y_{i+1})$, then at least one of $(x_i, x_{i+1}) \neq (\triangleleft u_i, \triangleright u_i)$ and $(\triangleleft u_i, \triangleright u_i) \neq (y_i, y_{i+1})$ has to be satisfied.

These properties of the pair-distance enable its use in the statement of necessary and sufficient conditions for pair-error correctability. Define a code $\mathcal{C} \subset \Xi^n$, and let

$$d_p = \min_{\boldsymbol{x} \in \mathcal{C}, \boldsymbol{y} \in \mathcal{C}, \boldsymbol{x} \neq \boldsymbol{y}} D_p(\boldsymbol{x}, \boldsymbol{y})$$

be the minimum pair-distance of $\mathcal{C}$. A necessary and sufficient condition for correctability of t pair-errors is provided in the following proposition.

Proposition 3. *A code $\mathcal{C}$ can correct t pair-errors if and only if $d_p \geqslant 2t + 1$.*

The proof of this proposition is essentially the same as in the Hamming case.

As it turns out, the necessary and sufficient conditions no longer match when the received pair-vectors are guaranteed to be consistent. In the case where decoder inputs are always consistent, there is a gap of one between the necessary and sufficient conditions. We first prove a (weaker) necessary condition for that case, and later show that it is tight.

Theorem 4. *When decoder inputs are consistent pair-vectors, a code $\mathcal{C}$ can correct all t-pair errors only if $d_p \geqslant 2t$.*

A proof of the theorem readily follows from the following lemma.

Lemma 5. *If $D_p(\boldsymbol{x}, \boldsymbol{y}) = 2t - 1$, then there exists a word $\boldsymbol{z} \in \Xi^n$ such that $D_p(\boldsymbol{x}, \boldsymbol{z}) = t$ and $D_p(\boldsymbol{z}, \boldsymbol{y}) \leqslant t$.*

Proof: Define the set $S_H = \{j : x_j \neq y_j\}$, and let $S_H = \cup_{l=1}^{L} B_l$ be its minimal partition to subsets with consecutive indices. For the selection of the word $\boldsymbol{z}$, we now specify how to construct a set T_H from S_H. Define a counter κ and initialize it to $\kappa = t$. If there is a subset $B_l = [s_l, e_l]$ with size $\kappa - 1$ or less, add it to T_H, subtract its size plus one from κ, and repeat the step with the new κ value. When there is no subset smaller than κ, pick any subset $B_l = [s_l, e_l]$, add the subset $[s_l, s_l + \kappa - 2]$ (of size $\kappa - 1$) to T_H, and terminate. Now define the word $\boldsymbol{z}$ as follows:

$$z_j = \begin{cases} y_j & \text{if } j \in T_H \\ x_j & \text{otherwise} \end{cases}$$

Each subset added to T_H contributes a pair-distance increase between $\boldsymbol{x}$ and $\boldsymbol{z}$ amounting to its size plus one. Hence $D_p(\boldsymbol{x}, \boldsymbol{z}) = t$, as a result of the initialization $\kappa = t$. Denote the number of subsets in the minimal partition of T_H by L_1. Denote the number of subsets in the minimal partition of $S_H \setminus T_H$ by L_2. Since at most one subset out of the L subsets of S_H is split between T_H and $S_H \setminus T_H$, then $L_1 + L_2 \leqslant L + 1$. That fact gives the last inequality in the following

$$D_p(\boldsymbol{z}, \boldsymbol{y}) =$$

$$|S_H| - |T_H| + L_2 = (2t - 1 - L) - (t - L_1) + L_2 \leqslant t$$

[1] indices may wrap around modulo n

The first equality is from Theorem 2, with $D_H(\boldsymbol{z},\boldsymbol{y}) = |S_H| - |T_H|$; the second equality is also from Theorem 2, with $D_p(\boldsymbol{x},\boldsymbol{y}) = 2t-1$ and $D_p(\boldsymbol{x},\boldsymbol{z}) = t$. ■

The weaker necessary condition is tight since there exist codes with $d_p = 2t$ that *can* correct all t-pair errors resulting in consistent pair-vectors, such as the code $\{00000, 01110\}$ with $d_p = 4$, which can correct all 2-pair errors. (For example, the received word 01000 in pair-distance 2 from the all-zero codeword is in pair-distance 3 from the codeword 01110, so can be decoded correctly.).

III. Code Construction and Decoding

A. Constructions from Hamming-metric codes

The treatment of adjacent symbols as pairs is reminiscent of the well studied problem of correcting error bursts. Therefore, it is not surprising that *interleaving*, a standard method for error-burst correction, is found useful for the symbol-pair problem as well. In Proposition 1, a potential factor-two gap between the Hamming distance and the pair distance is shown. As a consequence, using codes in the Hamming metric for pair-error correction is sub-optimal. To close the factor-two gap, while still using codes in the Hamming metric, the method of interleaving can be invoked.

Theorem 6. *Let $\mathcal{C}^{(1)}$ be an (n, M_1, d_H) code, and $\mathcal{C}^{(2)}$ be an (n, M_2, d_H) code, using the standard notation of (N, M, D) to denote a length N code with M codewords and minimum Hamming distance D. Then the code obtained by interleaving codewords of $\mathcal{C}^{(1)}$ and $\mathcal{C}^{(2)}$ as in equation* (2) *is a $(2n, M_1M_2, d_H)$ code with $d_p = 2d_H$.*

$\mathcal{C}_0^{(1)}$	$\mathcal{C}_0^{(2)}$	$\mathcal{C}_1^{(1)}$	$\mathcal{C}_1^{(2)}$	$\cdots$	$\mathcal{C}_{n-1}^{(1)}$	$\mathcal{C}_{n-1}^{(2)}$

(2)

Proof: For any two codewords of the interleaved code, the Hamming distance on $\mathcal{C}^{(1)}$'s or $\mathcal{C}^{(2)}$'s coordinates (or both) is at least d_H. Since none of the differing coordinates of the same $\mathcal{C}^{(i)}$ are in consecutive locations, the number of consecutive subsets is at least d_H, and by Theorem 2, $d_p \geqslant d_H + d_H = 2d_H$. The reverse inequality $d_p \leqslant 2d_H$ is proved by fixing a codeword of $\mathcal{C}^{(1)}$ and taking two codewords of $\mathcal{C}^{(2)}$ at distance d_H as two codewords of the interleaved code. ■

B. Direct cyclic-code construction

While codes constructed via interleaving obtain optimal pair-distance for their Hamming distance (factor 2), they are in general inferior to directly constructed codes, even if the constituent Hamming-metric codes are themselves optimal. This fact is proved in the following example.

Example 1. *As a directly constructed pair-error code we take the $[30, 22]$ shortened cyclic code generated by the polynomial $1 + x^2 + x^3 + x^8$, whose minimum pair-distance is $d_p = 7$. By Proposition 3, this code can correct 3 pair-errors, one more than the $[30, 22]$ code obtained from interleaving the (perfect) $[15, 11]$ Hamming code with itself, whose minimum pair-distance is only $d_p = 6$.*

The problem with the interleaving approach is that it optimizes the pair-distance given the Hamming distance, with no attempt to optimize the Hamming distance itself (interleaved codes are known to have poor Hamming distance for their length). Therefore, in the remainder of the section we take a more balanced approach of (more modestly) lower-bounding the pair-distance given the Hamming distance, but using codes that enjoy better Hamming distances to begin with: cyclic linear codes. New algebraic methods in the realm of the theory of cyclic codes will be sought as a framework for analysis and synthesis of pair-error correcting codes. For codes in the Hamming metric, the most powerful, flexible and practical codes in use are cyclic codes. So the purpose of the forthcoming discussion is to explore how the structure of cyclic codes can be exploited for pair-error correction as well. We start with common definitions and notations for cyclic codes [9]. Let $g(x)$ be a polynomial of degree r over $\mathbb{F}_q$. If $g(x)|(x^n-1)$, then $g(x)$ defines a cyclic linear code $\mathcal{C}$ of length n with dimension $k = n-r$. The codewords of $\mathcal{C}$ are all polynomials that can be written as $c(x) = g(x)f(x)$, for some polynomial $f(x)$ over $\mathbb{F}_q$, where polynomial multiplication is carried out over $\mathbb{F}_q[x]/(x^n-1)$, the ring of q-ary polynomials modulo x^n-1. Let $\mathbb{F}_{q^t}$ be the splitting field of $\mathbb{F}_q$, i.e. the smallest field in which x^n-1 can be factored into linear factors. So $\mathbb{F}_{q^t}$ contains a primitive n^{th} root of unity α, such that $x^n - 1 = \prod_{i=0}^{n-1}(x-\alpha^i)$. For a polynomial $c(x) \in \mathbb{F}_q[x]/(x^n-1)$, we define the j^{th} DFT (Discrete Fourier Transform) coefficient as $\hat{C}_j = c(\alpha^j)$. The BCH bound [2],[5], the most fundamental result in the theory of cyclic codes, can be formulated in terms of the DFT coefficients of codeword polynomials. We now include Proposition 7 as presented in [8]. Unless noted otherwise, polynomial and DFT indices are taken modulo n.

Proposition 7. *Let $c(x)$ be a polynomial in $\mathbb{F}_q[x]/(x^n-1)$ whose DFT satisfies $\hat{C}_{b+1} = \hat{C}_{b+2} = \cdots = \hat{C}_{b+\delta} = 0$, for some integer b. Then the number of non-zero coefficients in $c(x)$ is at least $\delta + 1$.*

To prove lower bounds on the minimum pair-distance, a dual version of Proposition 7 will become useful.

Proposition 8. *Let $c(x)$ be a polynomial in $\mathbb{F}_q[x]/(x^n-1)$ whose coefficients satisfy $c_{b+1} = c_{b+2} = \cdots = c_{b+\delta} = 0$, for some integer b. Then the number of non-zero elements in the DFT sequence $\{\hat{C}_j\}_{j=0}^{n-1}$ is at least $\delta + 1$.*

The next (simple) step is to transpose[2] Proposition 8 and get the following lemma.

Lemma 9. *Let $\{\hat{C}_j\}_{j=0}^{n-1}$ be a DFT sequence with at least d zero elements. Then $c(x)$ does <u>not</u> have a set of $n-d$ consecutive coefficients such that $c_{b+1} = c_{b+2} = \cdots = c_{b+n-d} = 0$.*

Finally, an algebraic lower bound on the minimum pair-distance of a cyclic code is provided in the following theorem. We remark that the linearity of the codes allows considering the pair-weight, i.e. the pair-distance to the all-zero codeword, when proving results on the minimum pair-distance of the codes.

Theorem 10. *Let $g(x)$ be a generator polynomial of a cyclic code $\mathcal{C}$ with minimum Hamming distance d_H. If $g(x)$ has at*

[2]Logical transposition means change from $(a \Rightarrow b)$ to $(\text{not } b \Rightarrow \text{not } a)$.

least d_H roots in $\mathbb{F}_{q^t}$, then the minimum pair-distance of $\mathcal{C}$ is at least $d_H + 2$.

Proof: If $g(x)$ has at least d_H roots, then any codeword $c(x) = g(x)f(x)$ has a DFT sequence with at least d_H zeros. By Lemma 9, $c(x)$ does not have a set of $n - d_H$ consecutive zero coefficients. We distinguish two cases. If $c(x)$ has Hamming weight exactly d_H, then its non-zeros cannot fall into a single set of consecutive locations. This implies a pair-weight of at least $d_H + 2$. Alternatively, if $c(x)$ has Hamming weight strictly larger than d_H, this also implies a pair-weight of at least $d_H + 2$. ■

The importance of Theorem 10 is that it provides an algebraic $d_H + 2$ lower bound on the pair-distance of a code that is strictly better than the combinatorial $d_H + 1$ lower bound of section II. This algebraic lower bound applies to all linear cyclic codes that are not MDS (Maximum Distance Separable). The next example shows how this improved lower bound can prove that *cyclic* Hamming codes are "perfect" in the pair-metric as well.

Example 2. *Let α be a primitive element in $\mathbb{F}_{2^t}$, for some $t > 2$. With $n = 2^t - 1$ we define $g(x) \in \mathbb{F}_2[x]/(x^n - 1)$ to be the lowest degree polynomial that satisfies $g(\alpha) = 0$. It is well known that the length n cyclic code generated by $g(x)$ has roots $\alpha^1, \alpha^2, \alpha^{2^2}, \ldots, \alpha^{2^{t-1}}$, and by the BCH bound has minimum Hamming distance $d_H = 3$. This code is the length $2^t - 1$ binary Hamming code, constructed as a primitive narrow-sense BCH code with designed minimum Hamming distance 3 (all binary Hamming codes can be constructed in that form, but not all q-ary ones). We now turn to analyze the pair-error correction on the Hamming code generated by $g(x)$. For any $t > 2$, $g(x)$ has at least $d_H = 3$ roots. Hence by Theorem 10, it has minimum pair-distance $d_p \geqslant 5$. So cyclic Hamming codes of length $n \geqslant 7$ can correct 2 pair-errors. Later in the paper (Theorem 19), a pair-metric sphere-packing bound will be used to prove that these cyclic Hamming codes are perfect in the pair-metric as well. Hence it follows that cyclic Hamming codes have exactly $d_p = 5$.*

We note that the $d_p \geqslant 5$ bound obtained with Theorem 10 in Example 2 applies exclusively to Hamming codes that are cyclic. For example, there are equivalent ways to construct Hamming codes that yield (non-cyclic) codes with $d_p = 4$ (see "textbook" Hamming code in Figure 3). Clearly the code represented by the parity-check matrix in Figure 3 is equivalent to the code generated by $g(x)$ in Example 2 (identical up to reordering of the code coordinates). But the fact that the two codes have different minimum pair-distances demonstrates the sensitivity of pair-error correctability to such coordinate reordering.

$$H_{[7,4]} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Figure 3. Parity check matrix of a $[7, 4]$ Hamming code. The word 0001110 is a codeword, thus the minimum pair-distance is only 4.

The algebraic structure of cyclic codes serves the theory of pair-error correction beyond the result of Theorem 10. Most conveniently, we can harness deeper (than the BCH bound) results on cyclic codes to obtain stronger bounds on the minimum pair-distance. This possibility is proved with the following theorem.

Theorem 11. *Let $g(x)$ be a generator polynomial of a cyclic code $\mathcal{C}$ with prime length n and minimum Hamming distance d_H. If $g(x)$ has at least m roots in $\mathbb{F}_{q^t}$, and $d_H \leqslant \min(2m - n + 2, m - 1)$, then the minimum pair-distance of $\mathcal{C}$ is at least $d_H + 3$.*

Proof: The main tool to prove this theorem is the use of the dual of the Hartmann-Tzeng bound [4], which generalizes the BCH bound to multiple sets of consecutive zeros (see also [6, Ch.4] for examples). We include a special case of the Hartmann-Tzeng bound in the following lemma.

Lemma 12. *(Hartmann-Tzeng Bound) Let $c(x)$ be a polynomial in $\mathbb{F}_q[x]/(x^n - 1)$ whose DFT satisfies $\hat{C}_{b+1} = \hat{C}_{b+2} = \cdots = \hat{C}_{b+\delta-1} = 0$, and $\hat{C}_{a+b+1} = \hat{C}_{a+b+2} = \cdots = \hat{C}_{a+b+\delta-1} = 0$ for some integers b, a, with $\gcd(a, n) < \delta$. Then the number of non-zero coefficients in $c(x)$ is at least $\delta + 1$.*

As we did with the BCH bound, a transposed version of the dual of Lemma 12 is stated in the following lemma (we skip the dual and move directly to the transposed dual).

Lemma 13. *Let $\{\hat{C}_j\}_{j=0}^{n-1}$ be a DFT sequence with at least m zero elements. Then $c(x)$ does <u>not</u> have two sets of $n - m - 1$ consecutive coefficients such that $c_{b+1} = c_{b+2} = \cdots = c_{b+n-m-1} = 0$ and $c_{a+b+1} = c_{a+b+2} = \cdots = c_{a+b+n-m-1} = 0$, for any b and any a with $\gcd(a, n) < n - m$.*

Now Lemma 13 implies that for a prime n, there are no two sets of $n-m-1$ consecutive zero coefficients, for any spacing a between them (for n prime $\gcd(a, n) = 1$ for any $a < n$). Denote by $D_H(c)$ the number of non-zero coefficients in the polynomial $c(x)$. $D_p(c)$ will denote the pair-weight of the vector of coefficients of $c(x)$. If $c(x)$ does not have two sets of $n - m - 1$ consecutive zero coefficients, then at least one of the following is true:

1) The non-zero coefficients of $c(x)$ fall into 3 or more consecutive subsets.
2) $D_H(c) > n - 2(n - m - 1) = 2m - n + 2$.

In other words, if the non-zero coefficients of $c(x)$ fall into 2 consecutive subsets, then their number must be greater than $2m - n + 2$ to avoid two consecutive zero subsets of the forbidden size. Note that $d_H \leqslant m - 1$ excludes the possibility of a single consecutive set of non-zeros in $c(x)$, as proved in Theorem 10.

If option 1 is true, then the theorem trivially follows from (1). If option 2 is true, then $D_H(c) \geqslant d_H + 1$ from the condition $d_H \leqslant 2m - n + 2$. We distinguish two cases. If $D_H(c) = d_H + 1$, the condition $d_H \leqslant m - 1$ guarantees that $D_H(c) \leqslant m$, and by Theorem 10 $D_p(c) \geqslant D_H(c) + 2 \geqslant d_H + 3$. If $D_H(c) \geqslant d_H + 2$ then the bound $D_p(c) \geqslant d_H + 3$ follows trivially from the combinatorial relation $D_p \geqslant D_H + 1$. ■

More insight and guarantees on pair-error correction of cyclic codes may be gained by carefully analyzing subsequent improvements of the BCH bound that have appeared in the literature. Examples for these include the Roos lower bound [10] and the van Lint-Wilson bounding technique [11].

C. Decoding

So far, for the correctability proofs of sub-section II-A, a decoding function in the pair-metric was assumed, without regard to the algorithmic aspects of achieving such a decoder (nearest-codeword pair-decoding by enumerating the pair-distances from the received pair-vector to all codewords and finding the closest one is always possible, though not practical). Similarly to decoding in the Hamming metric, there is a need to devise efficient decoding algorithms that will allow the implementation of coding schemes in the pair-metric. Clearly the decoder design will be governed by the specific code of choice, but in the following we attempt a more generic treatment of symbol-pair decoding. In Algorithm 1 below, we propose a reduction of the pair-decoding problem to error+erasure decoding in the Hamming metric.

Algorithm 1. *Let* $\overleftrightarrow{u} = [(\triangleleft u_0, \triangleright u_0), \ldots, (\triangleleft u_{n-1}, \triangleright u_{n-1})]$ *be the received pair-vector. Define the* n *symbols of the vector* $\boldsymbol{z}$ *as*

$$z_i = \begin{cases} \triangleleft u_i & \text{if } \triangleleft u_i = \triangleright u_{i-1} \\ * & \text{otherwise} \end{cases}$$

The symbol $*$ *represents symbol erasure and is used when symbol hypotheses from the two pairs are in conflict. The vector* $\boldsymbol{z}$ *is now input to an error/erasure decoder in the Hamming metric.*

While any code (either interleaved or directly constructed) for the pair-error model can be decoded using Algorithm 1, the question to ask at this moment is whether this decoder provides the decoding guarantees of Proposition 3, i.e. whether it is guaranteed to find the unique codeword (if exists) within a pair-ball of radius $\lfloor (d_p - 1)/2 \rfloor$ around the received pair-vector. The answer turns out to be *no* in general, and *yes* for interleaved codes. To prove that the algorithm is inferior, in general, to a bounded-distance pair-decoder we show an example. Suppose a single pair-error correcting code with the two codewords $\{00000, 01100\}$ (minimum pair-distance 3) is used, and the pair-vector $\overleftrightarrow{u} = [(0,0),(1,1),(0,0),(0,0),(0,0)]$ is received. Then Algorithm 1 will transform $\overleftrightarrow{u}$ into $\boldsymbol{z} = [0, *, *, 0, 0]$, and a Hamming-metric decoder will fail to decode (both codewords are equally likely given the decoder input $\boldsymbol{z}$). On the other hand, a pair-decoder will discern that $\overleftrightarrow{u}$ is at pair-distance 1 to 00000 and at pair-distance 2 to 01100, hence successfully choosing the vector 00000 from within the radius-1 pair-ball.
The equivalence of Algorithm 1 to optimal bounded-distance pair-decoding for interleaved codes is proved in the following theorem.

Theorem 14. *Let* $\mathcal{C}$ *be a* $(2n, M_1 M_2, d_H)$ *code constructed by interleaving two codes* $\mathcal{C}^{(1)}$ *and* $\mathcal{C}^{(2)}$*, each with Hamming distance* d_H *(hence by Theorem* 6 d_p *of* $\mathcal{C}$ *equals* $2d_H$*). Then the decoder of Algorithm* 1 *can correct up to* $\lfloor (d_p - 1)/2 \rfloor$ *pair-errors.*

Proof: We observe that a chain of ℓ consecutive pair-errors induces up to 2 symbol erasures and $\ell - 1$ symbol errors. For odd ℓ, each constituent code $\mathcal{C}^{(i)}$ suffers one erasure and $(\ell - 1)/2$ errors. For even ℓ, one constituent code suffers two erasures and $\ell/2 - 1$ errors and the other suffers zero erasures and $\ell/2$ errors. When taking the weighted sum of $2 \cdot \#\text{errors} + \#\text{erasures}$, each code has a worst case sum of t, where t is the number of pair-errors. Substituting $d_p = 2d_H$ from Theorem 6 into Proposition 3 gives $t \leqslant \lfloor (2d_H - 1)/2 \rfloor = d_H - 1$, and from elementary coding theory, bounded-distance Hamming decoders of the constituent codes will be able to correct an error/erasure weighted sum of $d_H - 1$. ■

IV. Bounds on Code Sizes

A. Combinatorial Bounds

The existence of necessary and sufficient conditions for pair-error correctability allows the derivation of upper and lower bounds, respectively, on the code size. A well known technique, used for both types of bounds, is to count the number of Ξ^n words in distance d from a given word. In the Hamming-distance metric, this counting task is very simple, and is used to derive the sphere-packing (upper) bound and the Gilbert-Varshamov (lower) bound, among many other bounds [7]. Given a word of Ξ^n, the pair-distance metric entails the complication of having part of the pair-error vectors (the consistent ones) result in words of Ξ^n, while others – non-consistent pair-error vectors – result in non-consistent pair-vectors. Thus the challenge is to count only the consistent pair-vectors at pair-distance d from the given word. Theorem 2 guides the solution toward solving the following combinatorial problem.

Problem 15. *Count how many of the subsets of the coordinate set* $[0, n-1]$ *have size* l*, and minimal partition of* $L = d - l$ *(cyclically) consecutive subsets.*

If this problem is solved, then all words that differ from the given word on these size-l subsets are known to be at pair-distance d from that word. Let $D(n, l, L)$ be the number of size l subsets of $[1, n]$ that occupy L cyclically consecutive subsets. The closed form formula for $D(n, l, L)$ is given in the following theorem.

Theorem 16. *For any triple* $n > l \geqslant L$*,*

$$D(n,l,L) = \binom{l-1}{L-1}\left[\binom{n-l-1}{L} + 2\binom{n-l-1}{L-1}\right] \quad (3)$$

$$+ \binom{n-l-1}{L-1}\binom{l-1}{L} \quad (4)$$

Proof: A subset that meets the (n, l, L) specification has one of the layouts depicted in Figure 4. The dark rectangles represent elements in the size-l subset. The white rectangles represent elements *not* in the subset. The three layouts in (a) are ones that have no wrap around from $n-1$ to 0. The layout in (b) has a consecutive subset that wraps around. For any L, dark and white rectangles are placed in alternation (Figure 4 presents an example with $L = 3$). Each rectangle represents a subset with size *strictly* larger than 0. The sizes of the dark rectangles in each layout sum to l. The sizes of the white rectangles sum to $n-l$. From elementary counting, the number of ways to divide m identical elements into k numbered bins that are non-empty is $\binom{m-1}{k-1}$, [12, Ch.13]. The identical elements are the dark/white indices, and the bins are the dark/white rectangles of consecutive indices. We observe that the dark and the white elements can be independently grouped

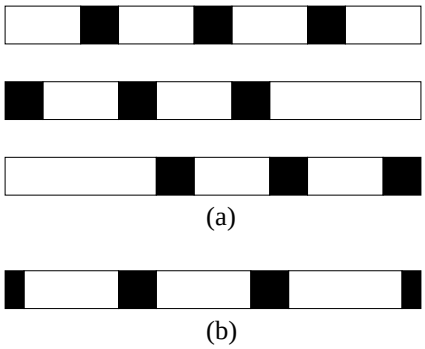

Figure 4. Layouts of (n, l, L) subsets. (a) Non-all-around layouts. (b) All-around layouts.

to rectangles, and each such grouping gives a distinct (n, l, L) subset. Therefore, the number of (n, l, L) subsets from each layout is the product of the number of dark groupings by the number of white groupings. In particular, the 4 layouts in Figure 4 have the following m,k parameters, listed from top to bottom.

1) Darks: $m = l$, $k = L$. Whites: $m = n - l$, $k = L + 1$
2) Darks: $m = l$, $k = L$. Whites: $m = n - l$, $k = L$
3) Darks: $m = l$, $k = L$. Whites: $m = n - l$, $k = L$
4) Darks: $m = l$, $k = L + 1$. Whites: $m = n - l$, $k = L$

Now the closed form expression immediately follows with (3) for the non-all-around layouts and (4) for the all-around layouts. ■

It is worth noting interesting special cases of $D(n, l, L)$ from Theorem 16.

- $D(n, l, 1) = n$ (a single set with arbitrary shift)
- $D(n, l, L) = 0$ if $L > l$ or $L > n - l$.
- For $D(n, l, l)$ the contribution of (4) is zero (singleton sets cannot go all around).

With a closed form formula for $D(n, l, L)$ it is possible to obtain a closed form formula for the number of Ξ^n words at pair-distance h from a given word. For that purpose we now define $\mathcal{S}_h(\boldsymbol{x})$, the radius-$h$ *pair-sphere* around a word $\boldsymbol{x}$.

Definition 4. *For a word $\boldsymbol{x} \in \Xi^n$, define the pair-sphere $\mathcal{S}_h(\boldsymbol{x})$ as the set of all $\boldsymbol{y} \in \Xi^n$ such that $D_p(\boldsymbol{x}, \boldsymbol{y}) = h$.*

The size of h-spheres is given in the following proposition.

Proposition 17. *For any $\boldsymbol{x} \in \Xi^n$, and $0 < h < n$*

$$|\mathcal{S}_h(\boldsymbol{x})| = \sum_{l=\lceil h/2 \rceil}^{h-1} D(n, l, h - l)(q - 1)^l$$

where $q = |\Xi|$ is the size of the alphabet.

Note that $|\mathcal{S}_1(\boldsymbol{x})| = 0$, as needed, and $|\mathcal{S}_2(\boldsymbol{x})| = n(q - 1)$, which coincides with the Hamming sphere of radius 1. The extreme cases $h = 0$ and $h = n$ are $|\mathcal{S}_0(\boldsymbol{x})| = 1$ and $|\mathcal{S}_n(\boldsymbol{x})| = (q-1)^n$, respectively – identically to the Hamming spheres with the same radii. The *pair-ball* $\mathcal{B}_h(\boldsymbol{x})$ consists of all words with pair-distance h or less from $\boldsymbol{x}$, and clearly

$$|\mathcal{B}_h(\boldsymbol{x})| = 1 + \sum_{i=1}^{h} |\mathcal{S}_i(\boldsymbol{x})|. \tag{5}$$

The ability to enumerate pair-balls allows generalizing useful bounds to the pair-metric.

Proposition 18. *(Pair-Sphere Packing Bound) If $\mathcal{C} \subset \Xi^n$ is a code with M codewords that corrects all t-pair errors, then*

$$M|\mathcal{B}_t(\boldsymbol{x})| \leqslant q^n.$$

This sphere-packing bound in the pair-metric can be used to prove that the cyclic Hamming codes analyzed in Example 2 are perfect in the pair-metric as well.

Theorem 19. *(Pair-perfect Hamming Codes) If a code is perfect in the Hamming metric with $d_H = 3$, and in addition has minimum pair-distance $d_p = 5$, then it is perfect in the pair-metric as well.*

Proof: The proof follows from the aforementioned fact that the pair-sphere $\mathcal{S}_2(\boldsymbol{x})$ is identical to the Hamming sphere $\mathcal{S}_1^H(\boldsymbol{x})$ (radius-1 Hamming sphere). Therefore, a code that perfectly packs the volume with Hamming spheres of radius 1, also packs perfectly with pair-spheres of radius 2. ■

Theorem 19 implies that Hamming codes cannot have minimum pair-distance $d_p = 6$, even though $d_p = 6$ satisfies the combinatorial bound $d_p \leqslant 2d_H$. It is important to note that this theorem is specific for the case $(d_H, d_p) = (3, 5)$, and in general perfect codes in the Hamming metric may not be perfect in the pair-metric, and vice-versa.

Next we write a Gilbert-Varshamov lower bound for pair-error correcting codes. The modification from the Hamming case is similarly done by considering pair-balls instead of Hamming balls.

Proposition 20. *(Pair Gilbert-Varshamov Bound) There exists a code $\mathcal{C} \subset \Xi^n$ with M codewords and minimum pair-distance d if*

$$M|\mathcal{B}_{d-1}(\boldsymbol{x})| \leqslant q^n.$$

B. Asymptotic Bounds

The combinatorial bounds of the previous sub-section use an exact enumeration of pair-spheres, and are thus a useful tool to bound the sizes of codes with given parameter sets. However, to get a general insight about the achievability and limits of coding in the pair-error model, an asymptotic analysis is needed. The main task toward an asymptotic analysis is to derive concise bounds on the sizes of pair-balls. Then the resulting simple expressions are used to bound the rates of codes with fractional minimum pair-distance $\delta = d_p/n$ (as the code length n tends to infinity). Our goal is to obtain asymptotic bounds on the size of pair-balls that will be tight enough to show a non-vanishing rate advantage of coding in the pair scheme over coding in the Hamming scheme. We note that it is not a-priori clear that such an advantage exists. Examining the bound $d_H + 1 \leqslant d_p \leqslant 2d_H$: if asymptotically good pair-codes have pair-distance at the low end closer to $d_H + 1$, then they are not likely to give any advantage over Hamming-metric codes; on the other extreme, if asymptotically good pair-codes have pair-distance at the high end closer to $2d_H$, then a significant advantage will emerge in favor of pair-codes. Thus the main purpose of the analysis below is to see whether asymptotically good pair-codes fall at the low end, high end, or somewhere in between

(asymptotic advantage, but less dramatic than doubling the relative distance).
We start by obtaining a simple upper bound on $D(n,l,L)$ by the following inequality

$$\begin{aligned} D(n,l,L) &= \binom{l-1}{L-1}\binom{n-l-1}{L} + 2\binom{l-1}{L-1}\binom{n-l-1}{L-1} \\ &+ \binom{l-1}{L}\binom{n-l-1}{L-1} \\ &< 4\binom{l}{L}\binom{n-l}{L} \end{aligned} \tag{6}$$

the inequality follows from the basic binomial recursion that gives the following inequalities

$$\binom{a-1}{b} = \binom{a}{b} - \binom{a-1}{b-1} < \binom{a}{b}$$

and

$$\binom{a-1}{b-1} = \binom{a}{b} - \binom{a-1}{b} < \binom{a}{b}$$

(substitute $b = L$ and $a = l$ or $a = n-l$ to get (6)). As we proceed, we restrict ourselves to binary codes ($q = 2$), though derivation for general q is not substantially different. The size of a pair-sphere can now be bounded using (6)

$$|\mathcal{S}_h(\boldsymbol{x})| = \sum_{l=\lceil h/2\rceil}^{h-1} D(n,l,h-l) < 4 \sum_{l=\lceil h/2\rceil}^{h-1} \binom{l}{h-l}\binom{n-l}{h-l}$$

Substituting the former into (5), we get

$$|\mathcal{B}_h(\boldsymbol{x})| = 1 + \sum_{i=1}^{h} |\mathcal{S}_i(\boldsymbol{x})| \leqslant 4 \sum_{i=1}^{h} \sum_{l=\lceil i/2\rceil}^{i-1} \binom{l}{i-l}\binom{n-l}{i-l}$$

Since $\binom{n-l}{i-l} = \binom{n-l}{n-i}$, we have

$$|\mathcal{B}_h(\boldsymbol{x})| \leqslant 4 \sum_{i=1}^{h} \sum_{l=\lceil i/2\rceil}^{i-1} \binom{l}{i-l}\binom{n-l}{n-i}$$

From the inner summation $l \geqslant \lceil i/2\rceil$, hence we can move the second multiplicand out of the inner sum:

$$4 \sum_{i=1}^{h} \sum_{l=\lceil i/2\rceil}^{i-1} \binom{l}{i-l}\binom{n-l}{n-i} < 4 \sum_{i=1}^{h} \binom{n-\lceil i/2\rceil}{n-i} \sum_{l=\lceil i/2\rceil}^{i-1} \binom{l}{i-l}$$

Now we can separate and bound each sum individually. We start with the right sum

$$\begin{aligned} \sum_{l=\lceil i/2\rceil}^{i-1} \binom{l}{i-l} &= \sum_{\ell=0}^{\lfloor i/2\rfloor-1} \binom{\lceil i/2\rceil+\ell}{\lfloor i/2\rfloor-\ell} && (7) \\ &= \sum_{\ell=0}^{\lfloor i/2\rfloor-1} \binom{\lceil i/2\rceil+\ell}{2\ell+i \bmod 2} && (8) \\ &< \sum_{\ell=0}^{\lfloor i/2\rfloor-1} \binom{i-1}{2\ell+i \bmod 2} && (9) \\ &< \sum_{\ell=0}^{\lfloor i/2\rfloor-1} \binom{i}{2\ell+1} && (10) \\ &< \sum_{s=0}^{i} \binom{i}{s} = 2^i \leqslant 2^h && (11) \end{aligned}$$

(7) is obtained by taking $\ell = l - \lceil i/2\rceil$. In (8) the binomial coefficient is written in its complementary form. (9) is implied by $\ell \leqslant \lfloor i/2\rfloor - 1$. (10) follows from the basic binomial recursion, and the first inequality in (11) is obtained by taking a super-set of the summation arguments in (10). The last inequality in (11) is simply by the fact that $i \leqslant h$.
Moving to the left sum, we rewrite its argument

$$\binom{n-\lceil i/2\rceil}{n-i} = \binom{n-\lceil i/2\rceil}{\lfloor i/2\rfloor}$$

And continue with the following chain of inequalities for the left sum

$$\begin{aligned} \sum_{i=1}^{h} \binom{n-\lceil i/2\rceil}{\lfloor i/2\rfloor} &< \sum_{i=1}^{h} \binom{n}{\lfloor i/2\rfloor} \\ &< 2 \sum_{j=0}^{\lfloor h/2\rfloor} \binom{n}{j} \\ &< 2^{1+nH\left(\frac{h}{2n}\right)} \end{aligned} \tag{12}$$

and $H(\alpha)$ is the binary entropy function. Combining (12) and (11) we get

$$|\mathcal{B}_h(\boldsymbol{x})| < 4 \cdot 2^{1+nH\left(\frac{h}{2n}\right)+h}$$

Denoting $\chi = h/n$ as the fractional pair-ball radius, gives

$$|\mathcal{B}_h(\boldsymbol{x})| < 2^{3+n\left[H\left(\frac{\chi}{2}\right)+\chi\right]}$$

To obtain asymptotic bounds on code rates it is useful to bound the ratio $2^n/|\mathcal{B}_h(\boldsymbol{x})|$

$$\frac{2^n}{|\mathcal{B}_h(\boldsymbol{x})|} > 2^{n-n\left[H\left(\frac{\chi}{2}\right)+\chi\right]-3}$$

Taking the logarithm (base 2) and normalizing by n we get

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{2^n}{|\mathcal{B}_h(\boldsymbol{x})|} \geqslant 1 - \left[H\left(\frac{\chi}{2}\right) + \chi\right] \tag{13}$$

Finally, we can write down the asymptotic Gilbert-Varshamov lower bound for pair-errors, as Corollary 21 to Proposition 20

Corollary 21. *(Asymptotic Pair Gilbert-Varshamov Bound) There exist codes with fractional minimum pair-distance δ and rate*

$$R(\delta) > 1 - H\left(\frac{\delta}{2}\right) - \delta$$

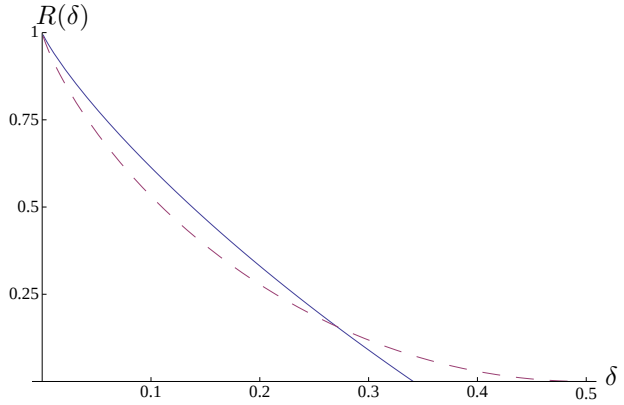

Figure 5. Comparison between the asymptotic Gilbert-Varshamov bounds for pair-distance (solid) and Hamming distance (dashed).

The implication from Figure 5 is that codes for pair-errors are now *provably* known to exist with asymptotic rates that are strictly higher than provably achievable rates of codes for symbol-errors. This statement applies to all fractional distances $\delta < 0.27$ (the x-coordinate of the intersection point between the two curves in Figure 5). Moreover, if, as widely conjectured, the binary Gilbert-Varshamov rate-bound for symbol-errors is tight [3], then Figure 5 proves an asymptotic gap in the correction capability between pair-errors and symbol-errors.

As for an asymptotic sphere-packing upper bound for pair-errors, the inequality in (13) *cannot* be used to exclude existence of codes with higher rate (a reverse inequality is needed for that). However, it is possible to use the asymptotic bound to obtain a lower bound on the sphere-packing upper bound for pair-errors. The usefulness of such a lower bound is expressed in the following proposition

Proposition 22. *(Bound on the Size of Pair-Perfect Codes) If there exist perfect codes in the pair-metric (code families that satisfy the bound from Proposition* 18 *with equality), with fractional minimum pair-distance δ, then their asymptotic rates satisfy*

$$R(\delta) > 1 - H\left(\frac{\delta}{4}\right) - \frac{\delta}{2}$$

V. Conclusions and Open Questions

The main purpose of this paper is the initial installation of the coding-theoretic infrastructure to study pair-error channels. First steps toward constructive handling of pair-errors are made by sample code constructions that are evaluated against upper and lower bounds. However, similarly to symbol-error channels, achieving the end goal of reliable storage or communication over pair-error channels is a task for a comprehensive study by many researchers. To help guide such a research effort that will hopefully follow, we point possible directions. In addition to the algebraic cyclic-code construction techniques that were proposed in this paper, another potential construction tool may come from graph theory. The code alphabet symbols can be regarded as graph vertices, and symbol-pairs as the edges of the graph. Constructing good pair-error codes can then be formulated as a problem of finding many graph walks with small edge overlap. A more formal description of this graph-theoretic formulation is given in the Appendix below. It is possible that graph-theoretic insight in conjunction with known coding-theory results may lead to code constructions that are superior to currently known ones. Another interesting direction is to generalize the pair-error model to triple errors and beyond – toward a general framework of multi-symbol sliding window readers.

VI. Acknowledgment

The authors wish to thank Alexander Vardy for useful discussions and suggestions.

Appendix: Pair-error Code Construction as a Graph Problem

Let $V = \{v_1, \ldots, v_q\}$ be a set of graph vertices. Let $E \subseteq V \times V$ be a set of directed edges, each denoted as an ordered pair of vertices $(v^{\text{out}}, v^{\text{in}})$. A *walk* on a graph is a list of n, not necessarily distinct, edges from E: $[(v_1^{\text{out}}, v_1^{\text{in}}), (v_2^{\text{out}}, v_2^{\text{in}}), \ldots, (v_n^{\text{out}}, v_n^{\text{in}})]$, where $v_i^{\text{in}} = v_{i+1}^{\text{out}}$, for all $1 \leqslant i < n$. A walk is *closed* if in addition $v_n^{\text{in}} = v_1^{\text{out}}$. Now define the graph $\mathcal{G}$ to be the complete directed graph with self loops and q vertices. The problem of constructing codes over Ξ^n with minimum pair-distance d can be formulated as

Problem 23. *Given the graph $\mathcal{G}$, find a set of length n closed walks whose pairwise overlap is at most $n - d$ edges.*

As an example, we draw in Figure 6 the graph that corresponds to the binary alphabet.

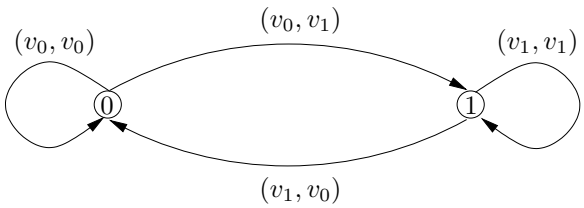

Figure 6. Complete directed graph for pair-error code construction over the binary alphabet.

References

[1] G. Benelli, C. Bianciardi, and V. Capellini, "Redundancy bounds for multiple-burst error-correcting codes," *Electronic Letters*, vol. 13, no. 13, pp. 389–390, June 1977.

[2] R. Bose and D. Ray-Chaudhuri, "On a class of error correcting binary group codes," *Information and Control*, vol. 3, pp. 68–79, 1960.

[3] V. Goppa, "Bounds for codes," *Doklady Akademii Nauk SSSR*, vol. 333, p. 423, 1993.

[4] C. Hartmann and K. Tzeng, "Generalizations of the BCH bound," *Information and Control*, vol. 20, pp. 489–498, 1972.

[5] A. Hocquenghem, "Codes correcteurs d'erreurs," *Chiffres*, vol. 2, pp. 147–156, 1959.

[6] W. Huffman and V. Pless, *Fundamentals of Error-Correcting Codes*. Cambridge, UK: Cambridge university press, 2003.

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

[8] R. J. McEliece, *The Theory of Information and Coding*. Cambridge University Press, 2002.
[9] W. Peterson and E. Weldon, *Error-Correcting Codes*. Cambridge, MA: MIT Press, 1972.
[10] C. Roos, "A new lower bound for the minimum distance of a cyclic code," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 330–332, 1983.
[11] J. van Lint and R. Wilson, "On the minimum distance of cyclic codes," *IEEE Transactions on Information Theory*, vol. 32, no. 1, pp. 23–40, 1986.
[12] ——, *A Course in Combinatorics, second edition*. Cambridge UK: Cambridge University Press, 2001.
[13] S. Wainberg and J. Wolf, "Burst decoding of binary block codes on q-ary output channels," *IEEE Transactions on Information Theory*, vol. 18, no. 5, pp. 684–686, 1972.