Nearly Perfect Covering Codes

Avital Boruchovsky and **Tuvi Etzion**

Computer Science Department, Technion, Israel Institute of Technology, Haifa 3200003, Israel

Abstract

Nearly perfect packing codes are those codes that meet the Johnson upper bound on the size of error-correcting codes. This bound is an improvement to the sphere-packing bound. A related bound for covering codes is known as the van Wee bound. Codes that meet this bound will be called nearly perfect covering codes. In this paper, such codes with covering radius one will be considered. It will be proved that these codes can be partitioned into three families depending on the smallest distance between neighboring codewords. Some of the codes contained in these families will be completely characterized. Construction for codes for each such family will be presented, the weight distribution of codes from these families will be examined, and some codes with special properties will be discussed.

I. Introduction

Perfect codes form the family of codes with one of the most fascinating structures in coding theory. These codes meet the well-known sphere-packing bound, but their percentage among all codes is tiny. Therefore, there were many attempts to find almost perfect codes. The nearly perfect codes are probably the codes which are the most close relatives of perfect codes. These codes are the topic of this paper.

A ***code*** of length n contains a set of words of length n. In all the codes which will be considered the words are binary. A ***translate*** of a code $\mathcal{C}$ of length n is the code

$$x + \mathcal{C} \triangleq \{ x + c \ : \ c \in \mathcal{C} \} ~,$$

where x is a word of length n and the sum with c is performed digit by digit. Usually, we assume that the all-zero word, $\mathbf{0}$, is a codeword in $\mathcal{C}$, while it does not belong to the translate of $\mathcal{C}$, i.e., x is not a codeword.

The ***Hamming distance*** between two words of length n is the number of coordinates in which they differ. The Hamming distance between two words x and y will be denoted by $d(x,y)$. Two words x and y of length n will be called ***adjacent*** if $d(x,y)=1$.

An ***error-correcting code*** $\mathcal{C}$ of length n and ***minimum distance*** d, contains codewords of length n, such that for each two distinct codewords c_1 and c_2, the Hamming distance $d(c_1,c_2) \geqslant d$.

A ***covering code*** $\mathcal{C}$ of length n and covering radius R, contains codewords of length n, such that each word of length n is within distance R from some codeword of $\mathcal{C}$. Such a code is also called an (n,R)***-covering code***. For a word $x \in \mathbb{F}_2^n$ we say that x is at distance d from $\mathcal{C}$, $d(x,\mathcal{C})$, if there exists a codeword $c \in \mathcal{C}$ such that $d(x,c)=d$ and there is no codeword $c_1 \in \mathcal{C}$ such that $d(x,c_1) < d$.

For a code $\mathcal{C}$ of length n and minimum distance $2R+1$, the size of $\mathcal{C}$ satisfies

$$|\mathcal{C}| \sum_{i=0}^{R} \binom{n}{i} \leqslant 2^n \tag{1}$$

and this bound is known as the ***sphere-packing bound***.

For a code $\mathcal{C}$ of length n and covering radius R, the size of $\mathcal{C}$ satisfies

$$|\mathcal{C}| \sum_{i=0}^{R} \binom{n}{i} \geqslant 2^n \tag{2}$$

and this bound is known as the ***sphere-covering bound***. Perfect codes are codes that meet the sphere-packing bound and also the sphere-covering bound.

The sphere-packing bound for a code $\mathcal{C}$ of length n and minimum distance $2R+1$ was improved by Johnson [17] as follows

$$|\mathcal{C}| \left(\sum_{i=0}^{R} \binom{n}{i} + \frac{\binom{n}{R}}{\lfloor \frac{n}{R+1} \rfloor} \left(\frac{n-R}{R+1} - \left\lfloor \frac{n-R}{R+1} \right\rfloor \right) \right) \leqslant 2^n \tag{3}$$

and a code that meets this bound is called a ***nearly perfect code***. When $R+1$ divides $n-R$, this bound coincides with the sphere-packing bound. Codes that meet this bound were considered in [15, 19]. There are two families of nontrivial codes that are nearly perfect and are not perfect. One family is the ***shortened Hamming codes***. A second fascinating family is called the ***punctured Preparata codes***. They have length 2^n-1, n even greater than 3, 2^{2^n-2n} codewords, and minimum distance 5. These codes were first found by Preparata [20] and later others found many inequivalent codes with the same parameters [3,18]. Moreover, these codes are very important in constructing other codes, e.g., mixed perfect codes [8] and quasi-perfect codes [10]. A comprehensive work on perfect codes and related ones can be found in [6]. For covering codes, an improvement for the sphere-covering bound akin to the Johnson bound was presented by van Wee [22]. A simplified version of the van Wee lower bound on the size of an (n,R)-covering code was presented by Struik [21]. If $\mathcal{C}$ is an (n,R)-covering code, then

$$|\mathcal{C}| \left(\sum_{i=0}^{R} \binom{n}{i} - \frac{\binom{n}{R}}{\lceil \frac{n-R}{R+1} \rceil} \left(\left\lceil \frac{n+1}{R+1} \right\rceil - \frac{n+1}{R+1} \right) \right) \geqslant 2^n \tag{4}$$

When $R+1$ divides $n+1$, the bound in (4) coincides with the sphere-covering bound. A code that meets this bound will be called a ***nearly perfect covering code***. One can easily see the similarity and the difference between the bounds in (3) and in (4).

Except for perfect codes and some trivial codes, there is only one known set of parameters for codes that meet the bound in (4). These codes have length 2^r, covering radius 1, and 2^{2^r-r} codewords, i.e., these are $(2^r,1)$-nearly perfect (covering) codes (the word covering will be omitted when the other parameters are given). In this paper, we consider the structure of these codes. In Section II we prove that the codewords in such a code can be partitioned into pairs such that for each pair $\{x,y\}$ we have either $d(x,y)=1$ or $d(x,y)=2$ and for each other codeword z we have $d(x,z) \geqslant 3$ and $d(y,z) \geqslant 3$. Based on this property, in Section III we partition these nearly perfect covering codes into three families of codes. We characterize these families and especially the family in which the distance between the codewords of each such pair is one. We also present constructions for codes in each one of these three families. In Section IV we consider the weight distribution of nearly perfect covering codes and in particular we prove that the weight distribution of all the codes in one family is the same for all codes and the same is true for all the codes in another family. In Section V, we concentrate on a class of codes, in the family where the distance between the codewords in a pair is one. which are balanced related to the codewords which differ by exactly one coordinate. The conclusion and a few problems for future research are presented in Section VI.

II. The Structure of Nearly Perfect Covering Codes

In this section, we examine the structure of $(2^r,1)$-nearly perfect codes. Let $\mathcal{C}$ be such a code with the all-zero codeword throughout this section. Such a code has 2^{2^r-r} codewords. Let $\mathcal{B}_t(x)$ be the ball of radius t around the word x, i.e.,

$$\mathcal{B}_t(x) \triangleq \{y \ : \ d(x,y) \leqslant t\}.$$

The ***multiplicity*** of a word $x \in \mathbb{F}_2^{2^r}$ is defined as the number of codewords which cover x, i.e., $|\{c \ : \ c \in \mathcal{C},\ d(x,c) \leqslant 1\}|$. The ***N-redundancies*** of a set $\mathcal{S}$ of a code $\mathcal{C}$ refers to the overhead of words that are covered, i.e., the number of N-redundancies of $\mathcal{S}$ is

$$\sum_{x \in \mathbb{F}_2^{2^r}} \left(|\{c \ : \ c \in \mathcal{S},\ d(x,c) \leqslant 1\}| - 1\right) .$$

The analysis of codes that meet the bound in (4) which was done by Struik [21] implies the following consequence.

Corollary 1. *If x is a word of length n, such that $x \notin \mathcal{C}$, then $\mathcal{B}_1(x)$ contains exactly one word that is covered by two codewords of $\mathcal{C}$ and no word that is covered by more than two codewords of $\mathcal{C}$.*

Proof. Let $x \notin \mathcal{C}$ and consider the translate $\mathcal{C}' = x + \mathcal{C}$. The all-zero word is not a codeword in $\mathcal{C}'$ and hence to cover the all-zero word, $\mathcal{C}'$ contains at least one codeword of weight one. We distinguish between two cases.

Case 1: $\mathcal{C}'$ contains exactly one codeword z of weight one. This codeword of weight one covers itself. To cover the other words of weight one it follows that there are at least 2^{r-1} codewords of weight two in $\mathcal{C}'$. If a word of weight two covers the word z, then $z \in \mathcal{B}_1(\mathbf{0})$ is covered by at least two codewords. Otherwise, there is a word $y \in \mathcal{B}_1(\mathbf{0})$ of weight one which is covered by two codewords of $\mathcal{C}'$. This implies that there is a word in $\mathcal{B}_1(x)$ which is covered by two codewords of $\mathcal{C}$.

Case 2: If $\mathcal{C}'$ contains at least two codewords z_1 z_2 of weight one, then $\mathbf{0} \in \mathcal{B}_1(0)$ is covered by two codewords z_1 and z_2 of $\mathcal{C}'$. This implies that $x + z_1$ and $x + z_2$ are two codeword in $\mathcal{C}$ which cover x and hence x is a word in $\mathcal{B}_1(x)$ which is covered by two codewords of $\mathcal{C}$.

Now, let ϵ be the average number of N-redundancies in the balls around all the words that are not contained in $\mathcal{C}$. Struik [21] proved that

$$\epsilon \geqslant 2\left(\left\lceil \frac{2^r+1}{2} \right\rceil - \frac{2^r+1}{2}\right) \tag{5}$$

and a $(2^r,1)$-nearly perfect code (meets the bound in (4)) must meet the bound in (5). This implies that $\epsilon = 1$ for a $(2^r,1)$-nearly perfect code. Since we proved that in each ball with a radius one around the words that are not contained in the code, there is at least one N-redundancy (a word that is covered at least twice), it follows that to have $\epsilon = 1$ there is exactly one N-redundancy for the words in each such ball, i.e., there is exactly one word which is covered twice and no word is covered more in each such ball. ∎

Corollary 2. *If x is a word of length n, such that $x \notin \mathcal{C}$, then $|\mathcal{B}_1(x) \cap \mathcal{C}| \leqslant 2$.*

An important result in our exposition is associated with the covering of pairs by triples. A ***covering triple system*** of order n is a pair (Q,B), where Q is an n-set and B is a collection of 3-subsets of Q, called ***blocks***, such that each 2-subset of Q is contained in at least one block of B. Fort and Hedlund [14] proved the following theorem.

Theorem 3. *If (Q,B) is a covering triple system of order n, then*

$$|B| \geqslant \left\lceil \frac{n}{3} \left\lceil \frac{n-1}{2} \right\rceil \right\rceil .$$

In their paper Fort and Hedlund [14] also present codes that meet the bound and find the exact structure of pairs that are covered more than once in these codes. Theorem 3 is used in the proof of Lemma 6, but using the structure of the pairs that are covered more than once, it can be used to prove other claims which we prove differently.

Lemma 4. *There are exactly 2^{2^r-r} N-redundancies of the code $\mathcal{C}$ in $\mathbb{F}_2^{2^r}$.*

Proof. Each codeword of $\mathcal{C}$ covers 2^r+1 words of $\mathbb{F}_2^{2^r}$ and since there are 2^{2^r-r} codewords, it follows that in total $2^{2^r-r}(2^r+1) = 2^{2^r} + 2^{2^r-r}$ words of $\mathbb{F}_2^{2^r}$ are covered (including N-redundancies). Since $\mathbb{F}_2^{2^r}$ has 2^{2^r} words, it follows that there are 2^{2^r-r} N-redundancies and the claim follows. ∎

Corollary 5. *On average, for each codeword of $\mathcal{C}$, there is exactly one N-redundancy associated with the code $\mathcal{C}$.*

Lemma 6. *If $c \in \mathcal{C}$, then there exists at least one codeword $c_1 \in \mathcal{C}$ such that $d(c,c_1) < 3$.*

Proof. W.l.o.g. (without loss of generality) assume that c is the all-zero word and assume to the contrary that $\mathcal{B}_2(c) \cap \mathcal{C} = \{c\}$. The code $\mathcal{C}$ contains a codeword of weight 3 since otherwise there will not be any codeword that covers words of weight two. Since there are no codewords of weight one or two, it follows that the words of weight two are covered only by codewords of weight three. By Theorem 3 the number of codewords of weight three is at least $\frac{2^{2r-1}+1}{3}$. Each such triple (codeword of weight 3) covers three pairs and hence, the number of words with weight two that are covered (including N-redundancies) is at least $2^{2r-1}+1$. The total number of words with weight two in $\mathbb{F}_2^{2^r}$ is $\binom{2^r}{2} = 2^{r-1}(2^r-1) = 2^{2r-1} - 2^{r-1}$. Since by Corollary 1 no word of weight two can be covered more than twice, it follows that there are at least $2^{r-1}+1$ words of weight two that are covered twice. This implies that there exists at least one coordinate, say ℓ, such that two words, say z_1 and z_2, of weight two with a *one* in coordinate ℓ are covered twice by $\mathcal{C}$. Therefore, for the word x of weight one with a *one* in coordinate ℓ we have at least two words in $\mathcal{B}_1(x)$ are covered twice by $\mathcal{C}$, contradicting Corollary 1.

Thus, there exists at least one codeword $c_1 \in \mathcal{C}$ such that $d(c,c_1) < 3$. ∎

Corollary 7. *For each codeword $c \in \mathcal{C}$, $|\mathcal{B}_2(c) \cap \mathcal{C}| = 2$, i.e., there exists exactly one codeword $c_1 \in \mathcal{C}$ such that either $d(c,c_1) = 1$ or $d(c,c_1) = 2$.*

Proof. Again, w.l.o.g. assume that c is the all-zero word. By Lemma 6 $\mathcal{C}$ has a codeword of weight one or a codeword of weight two. If there exists one codeword c_1 of weight one, then c covers c and c_1, and also c_1 covers c and c_1. If there exists one codeword c_1 of weight two, then there are two words x_2 and x_3 of weight one such that x_2 is covered by c and c_1 and also x_3 is covered by c and c_1. These two cases satisfy the condition of Corollary 5.

If there exist at least two codewords of weights one or two, then the average multiple covering is damaged.

For example, if there exist two codewords c_1 and c_2 of weight one, then c is covered three times, while c_1 and c_2 are covered twice each. Hence, we have four N-redundancies on three codewords. Note, that each additional codeword to this set of codewords adds at least one more multiplicity keeping more N-redundancies than codewords, contradicting Corollary 5.

Another example, if there exist two codewords c_1 and c_2 of weight two, which intersect in one coordinate, then the word of weight one with a *one* in this coordinate is covered three times, and another two words of weight one are covered twice. This makes it four N-redundancies on three codewords. Again, this case contradicts Corollary 5.

Similar contradictions arise in the other cases. ∎

Corollary 8. *If x is a word of length n, then $|\mathcal{B}_1(x) \cap \mathcal{C}| \leqslant 2$.*

Corollary 9. *There are exactly 2^{2^r-r} words in $\mathbb{F}_2^{2^r}$ which are covered twice in $\mathcal{C}$ and no word is covered more than twice.*

Corollary 10. *The codewords of $\mathcal{C}$ can be partitioned uniquely into 2^{2^r-r-1} pairs such that if $\{x,y\}$ is a pair in this partition, then either $d(x,y)=1$ or $d(x,y)=2$.*

Corollary 10 yields two types of "spheres" for a pair of codewords. The first type, called type A, has in each sphere two centers x and y such that $d(x,y)=1$. The second type, called type B, has in each sphere two centers x and y such that $d(x,y)=2$.

Corollary 11. *Each partition of $\mathbb{F}_2^{2^r}$ into spheres of type A and type B yields a $(2^r,1)$-nearly perfect code.*

Corollary 12. *The words of $\mathbb{F}_2^{2^r}$ which are covered twice in a $(2^r,1)$-nearly perfect code are the codewords of the pairs $\{x,y\}$ such that $d(x,y)=1$ and the pairs of words $\{u,v\}$ derived from a pair of codewords $\{x,y\}$ such that $d(x,y)=2$, $d(x,u)=d(x,v)=d(y,u)=d(y,v)=1$.*

For any codeword $c \in \mathcal{C}$ the codeword $c' \in \mathcal{C}$ such that $d(c,c')=1$ or $d(c,c')=2$ will be called the ***partner*** of c. These two codewords will be referred to as a ***pair of codewords***.

III. Constructions of Nearly Perfect Covering Codes

All the constructions of nearly perfect covering codes which will be presented in this section are based on perfect codes and their properties. Hence, we start this section by presenting some basics of perfect codes. A perfect code is a code that meets the bounds of (1) and (2). We will consider only codes for which $R=1$ in these equations. Such a code has length 2^r-1 and 2^{2^r-r-1} codewords [6, 12]. For each length 2^r-1, there is one linear perfect code known as the Hamming code, but by abuse of notation, we will say that a ***Hamming code*** is any perfect code with the all-zero codeword. Any other perfect code $\mathcal{C}$ without the all-zero word is also a translate of a Hamming code since for any $c \in \mathcal{C}$ we have that $c+\mathcal{C} \triangleq \{c+x \ : \ x \in \mathcal{C}\}$ is a Hamming code. The number of nonequivalent perfect codes is very large and it was considered throughout the years [5,6]. For example, it was proved in [12] that the number of nonequivalent perfect codes of length n, for sufficiently large n and a constant $c=0.5-\epsilon$, is $2^{2^{cn}}$. Analysis of various constructions of such codes can be found in [5, pp. 296–310].

The ***extended Hamming code*** is obtained from a Hamming code (i.e., a perfect code) by adding an even parity in a new coordinate. Clearly, such an extended Hamming code contains the all-zero codeword. There are two type of translates for an extended Hamming code, an odd translate and an even translate. An ***odd translate*** of an extended Hamming code contains only words with odd weight including exactly one word of weight one. An ***even translate*** of an extended Hamming code of length 2^r contains only words of even weight including 2^{r-1} words of weight two. Since the Hamming code is a perfect code with covering radius one, the following lemmas are followed.

Lemma 13. *If $\mathcal{C}$ is an extended perfect code, then deleting any one of its coordinates yields a perfect code.*

Lemma 14. *For each word $x \in \mathbb{F}_2^{2^r}$ of odd weight there exists exactly one codeword c in the extended Hamming code such that $d(c,x) = 1$.*

Lemma 15. *For each word $x \in \mathbb{F}_2^{2^r}$ of even weight there exists exactly one codeword c in an odd translate of the extended Hamming code such that $d(c,x) = 1$.*

Lemma 16. *For each word $x \in \mathbb{F}_2^{2^r}$ of odd weight there exists exactly one codeword c in an even translate of the extended Hamming code such that $d(c,x) = 1$.*

Recall that a $(2^r,1)$-nearly perfect code contains 2^{2^r-r} codewords and its codewords can be partitioned into 2^{2^r-r-1} pairs such that if $\{x,y\}$ is a pair in the partition, then either $d(x,y) = 1$ or $d(x,y) = 2$. Moreover, there is no other codeword z such that $d(x,z) \leqslant 2$ or $d(y,z) \leqslant 2$. We consider three types of nearly perfect covering codes. In the first type A, the distance between the two codewords in each pair of codewords is one. In the second type B, the distance between the two codewords in each pair of codewords is two. In the third type C, there are some pairs of codewords for which the distance is one and some pairs of codewords for which the distance is two. We will try to characterize some of these types, find some more properties that they have, and construct codes for each type. The following simple construction can be used to obtain nearly perfect covering codes for all three types.

Theorem 17. *If $\mathcal{C}_1$ and $\mathcal{C}_2$ are perfect codes of length 2^r-1, where $\mathbf{0} \in \mathcal{C}_1$, then the code*

$$\mathcal{C} \triangleq \{(c,0)\ :\ c \in \mathcal{C}_1\} \cup \{(c,1)\ :\ c \in \mathcal{C}_2\}$$

is a $(2^r,1)$-nearly perfect code.

Proof. To prove the claim of the theorem we only have to show that each word of length 2^r is within distance one from at least one codeword of $\mathcal{C}$.

If $x \in \mathbb{F}_2^{2^r-1}$, then since $\mathcal{C}_1$ and $\mathcal{C}_2$ are perfect codes, it follows that $d(x,\mathcal{C}_1) \leqslant 1$ and $d(x,\mathcal{C}_2) \leqslant 1$. Hence, there exists a codeword $c_1 \in \mathcal{C}_1$ such that $d((x,0),(c_1,0)) \leqslant 1$ and there exists a codeword $c_2 \in \mathcal{C}_2$ such that $d((x,1),(c_2,1)) \leqslant 1$. Moreover, since $\mathcal{C}_1$ contains the all-zero codeword, it follows that $\mathcal{C}$ contains the all-zero codeword.

Thus, $\mathcal{C}$ is a $(2^r,1)$-nearly perfect code. ∎

Corollary 18. *If $\mathcal{C}_1 = \mathcal{C}_2$ in Theorem 17, then the code $\mathcal{C}$ is a $(2^r,1)$-nearly perfect code of type A.*

Corollary 19. *If $\mathcal{C}_1$ in Theorem 17 is a Hamming code and $\mathcal{C}_2$ is a translate of $\mathcal{C}_1$, then the code $\mathcal{C}$ is a $(2^r,1)$-nearly perfect code of type B.*

Proof. Since $\mathcal{C}_2$ is a translate of $\mathcal{C}_1$, it follows by our choice of translates that $\mathcal{C}_1 \cap \mathcal{C}_2 = \varnothing$ and for each codeword $c_1 \in \mathcal{C}_1$ there exists a codeword $c_2 \in \mathcal{C}_2$ such that $d(c_1, c_2) = 1$ and therefore $d((c_1, 0), (c_2, 1)) = 2$. This implies that $\mathcal{C}$ is a $(2^r, 1)$-nearly perfect code of type B. ∎

Corollary 20. *If $\mathcal{C}_1$ in Theorem 17 is a Hamming code and $\mathcal{C}_2$ is a perfect code, such that $\mathcal{C}_1 \neq \mathcal{C}_2$ and $\mathcal{C}_1 \cap \mathcal{C}_2 \neq \varnothing$, then the code $\mathcal{C}$ is a $(2^r, 1)$-nearly perfect code of type C.*

Proof. If $\mathcal{S} = \mathcal{C}_1 \cap \mathcal{C}_2$, then for each $x \in \mathcal{S}$ we have that $(x, 0), (x, 1) \in \mathcal{C}$ and therefore $d((x, 0), (x, 1)) = 1$. Since $\mathcal{C}_2$ is a perfect code, it follows that for each $c_1 \in \mathcal{C}_1 \setminus \mathcal{S}$, there exists a codeword $c_2 \in \mathcal{C}_2 \setminus \mathcal{S}$ such that $d(c_1, c_2) = 1$ and hence $d((c_1, 0), (c_2, 1)) = 2$. This implies that $\mathcal{C}$ is a nearly perfect code of type C. ∎

Corollary 21. *If $\mathcal{C}_1$ and $\mathcal{C}_2$ in Theorem 17 are Hamming codes, such that and $|\mathcal{C}_1 \cap \mathcal{C}_2| = k$, then the code $\mathcal{C}$ is a $(2^r, 1)$-nearly perfect code of type C with exactly k pairs $\{c_1, c_2\}$ such that $d(c_1, c_2) = 1$.*

Corollaries 20 and 21 raise an interesting question associated with $(2^r, 1)$-nearly perfect codes of type C. For which integer k, $1 \leqslant k \leqslant 2^{2^r - r - 1} - 1$, there exists a $(2^r, 1)$-nearly perfect code $\mathcal{C}$ with exactly k pairs $\{c_1, c_2\}$, $c_1, c_2 \in \mathcal{C}$, such that $d(c_1, c_2) = 1$, and $2^{2^r - r - 1} - k$ pairs $\{c_1, c_2\}$, $c_1, c_2 \in \mathcal{C}$, such that $d(c_1, c_2) = 2$. Corollary 21 implies that such codes can be constructed from two Hamming codes whose intersection is k. It was proved by Avgustinovich, Heden, and Solov'eva [2] that for each even integer k such that $0 \leqslant k \leqslant 2^{n+1-2\log_2(n+1)}$ there exist two Hamming codes whose intersection is k. The minimum possible nonzero intersection of two Hamming codes is 2 and two such codes were found in [13]. This intersection problem was initiated in [12] and further investigated by Avgustinovich, Heden, and Solov'eva [1]. A summary of the results with complete analysis were given by Heden, Solov'eva, and Mogilnykh [16].

Our first observation is that the number k of pairs of codewords at distance one must be an even integer.

Lemma 22. *If $\mathcal{C}$ is a $(2^r, 1)$-nearly perfect code of type C, then the number of pairs $\{c_1, c_2\}$, $c_1, c_2 \in \mathcal{C}$, such that $d(c_1, c_2) = 1$, is an even integer.*

Proof. Each sphere of type A has in its center two words x, y such that $d(x, y) = 1$, and hence it contains 2^r words of even weight and 2^r words of odd weight. Each sphere of type B has in its center two words x, y such that $d(x, y) = 2$. Hence, it contains either 2 words of even weight (at the center) and $2^{r+1} - 2$ words of odd weight or 2 words of odd weight (at the center) and $2^{r+1} - 2$ words of even weight.

Since in $\mathbb{F}_2^{2^r}$ there are 2^{r-1} words of even weight and 2^{r-1} word of odd weight and in a sphere of type A the number of even words equals the number of odd words, it follows that the number of spheres of type B is even. This implies that also the number of spheres of type A is even. ∎

The proof of Lemma 22 implies also the following result.

Lemma 23. *In a $(2^r, 1)$-nearly perfect code the number of codewords with even weight is equal to the number of codewords with odd weight. Moreover, the number of pairs $\{x, y\}$, such that $x, y \in \mathcal{C}$, $d(x, y) = 2$, and the weights of x and y is even equal to the number of pairs $\{x, y\}$, such that $x, y \in \mathcal{C}$, $d(x, y) = 2$, and the weights of x and y is odd.*

The structure of a $(2^r, 1)$-nearly perfect code can be further revealed with the following result.

Lemma 24. *In a $(2^r, 1)$-nearly perfect code the number of* zeros *in each coordinate is $2^{2^r - r - 1}$ and it equals the number of* ones *in each coordinate.*

Proof. Assume for the contrary, w.l.o.g., that the number of *zeros* in the last coordinate is larger by $2k$ than the number of *ones*, i.e., there are $2^{2^r-r-1}-k$ codewords which end with a *one* and $2^{2^r-r-1}+k$ codewords which end with a *zero*.

We will count the number of *zeros* in the the last coordinate of the words covered by the codewords of $\mathcal{C}$. Since $\mathcal{C}$ is a $(2^r,1)$-nearly perfect code, it follows that each word of $\mathbb{F}_2^{2^r}$ is covered at least once. By Corollary 12 we can observe that in the words which are covered more than once, the number of *zeros* in the last coordinate equals the number of *zeros* in the codewords of $\mathcal{C}$ in the last coordinate. Therefore, the number of covered words that end with a *zero* is $2^{2^r-1}+2^{2^r-r-1}+k$.

Each codeword that ends in a *one* covers one word that ends with a *zero*. Each codeword that ends in a *zero* covers 2^r words that end with a *zero*. Therefore, the total number of covered words which end with a *zero* is

$$2^{2^r-r-1}-k+(2^{2^r-r-1}+k)\cdot 2^r = 2^{2^r-1}+2^{2^r-r-1}-k+k\cdot 2^r \ ,$$

a contradiction if $r>1$ and $k>0$. ∎

Theorem 25. *Any $(2^r,1)$-nearly perfect code $\mathcal{C}$ of Type A is constructed as a union of an extended Hamming code of length 2^r with an odd translate of an extended Hamming code of length 2^r.*

Proof. Since the codewords can be partitioned into pairs with distance one, it follows that half of the codewords are of even weight and half are of odd-weight. Moreover, since the distance between the codewords of the pairs is one, it follows by Corollary 7 that there are no pairs of codewords with a distance of 2. Therefore, the codewords of even weight in $\mathcal{C}$ form an extended Hamming code, and the codewords of odd weight form an odd translate of an extended Hamming code. ∎

Theorem 26. *A union of an extended Hamming code of length 2^r with an odd translate of an extended Hamming code of length 2^r is a $(2^r,1)$-nearly perfect code $\mathcal{C}$ of type A.*

Proof. Let $\mathcal{C}_1$ be an extended Hamming code of even weight and $\mathcal{C}_2$ be an odd translate of an extended Hamming code. If $x\in\mathbb{F}_2^{2^r}$ is of even weight, then by Lemma 15 there exists a codeword $c\in\mathcal{C}_2$ such that $d(x,c)=1$. If $x\in\mathbb{F}_2^{2^r}$ is of odd weight, then by Lemma 14 there exists a codeword $c\in\mathcal{C}_1$ such that $d(x,c)=1$. Moreover, the number of codewords in $\mathcal{C}_1\cup\mathcal{C}_2$ is 2^{2^r-r} and hence $\mathcal{C}$ is a $(2^r,1)$-nearly perfect code of type A. ∎

Corollary 27. *A translate of a $(2^r,1)$-nearly perfect code $\mathcal{C}$ of type A is constructed as a union of an even translate of an extended Hamming code of length 2^r with an odd translate of an extended Hamming code of length 2^r.*

Corollary 28. *A union of an even translate of an extended Hamming code of length 2^r with an odd translate of an extended Hamming code of length 2^r is a translate of a $(2^r,1)$-nearly perfect code $\mathcal{C}$ of type A.*

Corollary 29. *There is a one-to-one correspondence between the pairs of an extended Hamming code and length 2^r with an odd translate of an extended Hamming code of length 2^r, and the $(2^r,1)$-nearly perfect codes of type A.*

Corollary 30. *There is a one-to-one correspondence between the pairs of an even translate of an extended Hamming code and length 2^r with an odd translate of an extended Hamming code of length 2^r, and the translates of $(2^r,1)$-nearly perfect codes of type A.*

After characterizing the nearly perfect covering codes of type A we would like to characterize the set of nearly perfect codes of type B. A partial result is the following theorem.

Theorem 31. *A nearly perfect covering code of type A, in which there are no two codewords that differ only in one of the coordinates, yields a nearly perfect covering codes of type B.*

Proof. Let $\mathcal{C}'$ be a $(2^r,1)$-nearly perfect code of Type A, where w.l.o.g. there are no two codeword that differ only in the last coordinates. By Theorem 25 the code $\mathcal{C}'$ is a union of an extended Hamming code $\mathcal{C}_1$ and an odd translate of an extended Hamming code $\mathcal{C}_2$. We construct the following code

$$\begin{aligned}\mathcal{C} \triangleq \{(x,0) \ : \ x \in \mathbb{F}_2^{2^r-1}, \ (x,b) \in \mathcal{C}', \ \text{wt}\,(x,b) \equiv 0 \pmod 2\} \\ \cup\{(x,1) \ : \ x \in \mathbb{F}_2^{2^r-1}, \ (x,b) \in \mathcal{C}', \ \text{wt}\,(x,b) \equiv 1 \pmod 2\} \ ,\end{aligned}$$

where $\text{wt}\,(z)$ is the weight of z.

By Lemma 13 the code $\{x \ : \ x \in \mathbb{F}_2^{2^r-1}, \ (x,b) \in \mathcal{C}', \ \text{wt}\,(x,b) \equiv 0 \pmod 2\}$ is a Hamming code and $\{x \ : \ x \in \mathbb{F}_2^{2^r-1}, \ (x,b) \in \mathcal{C}', \ \text{wt}\,(x,b) \equiv 1 \pmod 2\}$ is a translate of a Hamming code. Hence, by Theorem 26, $\mathcal{C}$ is a $(2^r,1)$-nearly perfect code. Now, by Corollary 19 $\mathcal{C}$ is a nearly perfect covering code of type B. ∎

If the same construction in the proof of Theorem 31 is applied on a code in which some pairs (but not all of them) differ in the last coordinate and the associated codewords remain unchanged, then the constructed code $\mathcal{C}$ is a nearly perfect covering code of type C. In the constructed code $\mathcal{C}$, the number of pairs of codewords $\{x,y\}$ for which $d(x,y)=1$ is the same number of such pairs which differ only in the last coordinate of $\mathcal{C}'$.

IV. Weight Distribution of Nearly Perfect Covering Codes

The weight distribution of a code is one of its important parameters. For a code $\mathcal{C}$ let $\mathcal{A}_i(\mathcal{C})$ be the number of codewords in $\mathcal{C}$ whose weight is i. The weight distribution of a perfect code and a translate of a perfect code was found in [12]. If $\mathcal{C}$ is a perfect code of length $n = 2^r - 1 = 2\nu + 1$, then

$$\mathcal{A}_i(\mathcal{C}) = \frac{\binom{n}{i} + n\Delta_i}{n+1} \triangleq \alpha_i, \tag{6}$$

where

$$\Delta_i = \begin{cases} \binom{\nu}{\lfloor i/2 \rfloor} & i \equiv 0,3 \pmod 4 \\ -\binom{\nu}{\lfloor i/2 \rfloor} & i \equiv 1,2 \pmod 4 \end{cases} . \tag{7}$$

If $\mathcal{C}$ is a translate of a perfect code of length $n = 2^r - 1$, then

$$\mathcal{A}_i(\mathcal{C}) = \frac{\binom{n}{i} - \Delta_i}{n+1} \triangleq \beta_i, \tag{8}$$

The weight distribution of a perfect code given in (6) and the value of Δ_i in (7) implies the weight distribution of an extended Hamming code and also of translates of the extended Hamming code as follows.

Lemma 32. *If $\mathcal{C}$ is an extended Hamming code of length $n+1 = 2^r = 2\nu+2$, then $\mathcal{A}_i(\mathcal{C}) = 0$ for odd i and its weight distribution for even weights is given by*

$$\mathcal{A}_i(\mathcal{C}) = \alpha_i + \alpha_{i-1} = \frac{\binom{n+1}{i} + n(\Delta_i + \Delta_{i-1})}{n+1} = \begin{cases} \frac{\binom{n+1}{i}+n\binom{\nu+1}{i/2}}{n+1} & i \equiv 0 \pmod 4 \\ \frac{\binom{n+1}{i}-n\binom{\nu+1}{i/2}}{n+1} & i \equiv 2 \pmod 4 \end{cases}. \tag{9}$$

Lemma 33. *If $\mathcal{C}$ is an odd translate of an extended Hamming code of length $n+1 = 2^r = 2\nu+2$, then $\mathcal{A}_i(\mathcal{C}) = 0$ for even i and its weight distribution for odd weights is given by*

$$\mathcal{A}_i(\mathcal{C}) = \alpha_i + \alpha_{i-1} = \frac{1}{n+1}\binom{n+1}{i}, \quad i \equiv 1,3 \pmod 4. \tag{10}$$

Lemma 34. *If $\mathcal{C}$ is an even translate of an extended Hamming code of length $n+1 = 2^r = 2\nu+2$, then $\mathcal{A}_i(\mathcal{C}) = 0$ for odd i and its weight distribution for even weights is given by*

$$\mathcal{A}_i(\mathcal{C}) = \beta_i + \beta_{i-1} = \frac{\binom{n+1}{i} - (\Delta_i + \Delta_{i-1})}{n+1} = \begin{cases} \frac{\binom{n+1}{i}-\binom{\nu+1}{i/2}}{n+1} & i \equiv 0 \pmod 4 \\ \frac{\binom{n+1}{i}+\binom{\nu+1}{i/2}}{n+1} & i \equiv 2 \pmod 4 \end{cases}. \tag{11}$$

By Corollaries 29 and 30 and since there is a unique weight distribution for an extended Hamming codes and its translates of even and odd weights we have that $(2^r, 1)$-nearly perfect codes of type A and their translates have a unique weight distribution as follows.

Theorem 35. *If $\mathcal{C}$ is a $(2^r, 1)$-nearly perfect code of type A, where $n = 2^r - 1$, then*

$$\mathcal{A}_0(\mathcal{C}) = \mathcal{A}_{n+1}(\mathcal{C}) = 1$$

and for $1 \leqslant i \leqslant n$,

$$\mathcal{A}_i(\mathcal{C}) = \alpha_i + \alpha_{i-1} = \frac{\binom{n+1}{i} + n(\Delta_{i-1} + \Delta_i)}{n+1}.$$

Proof. By Theorem 25 any nearly perfect covering code of type A is constructed as a union of an extended Hamming code and an odd translate of an extended Hamming code. Hence, the weight distribution of such code is determined by Eq. (9) and (10). ∎

Similarly, we have the following theorem.

Theorem 36. *If $\mathcal{C}$ is a translate of a $(2^r, 1)$-nearly perfect code of type A, where $n = 2^r - 1$, then*

$$\mathcal{A}_0(\mathcal{C}) = \mathcal{A}_{n+1}(\mathcal{C}) = 0$$

and for $1 \leqslant i \leqslant n$,

$$\mathcal{A}_i(\mathcal{C}) = \beta_i + \beta_{i-1} = \frac{\binom{n+1}{i} - (\Delta_{i-1} + \Delta_i)}{n+1}.$$

Proof. A translate of any nearly perfect covering code of type A is constructed as a union of an odd translate of an extended Hamming code and an even translate of an extended Hamming code. This can be done by adding a coordinate to a translate of an Hamming code and to each codeword in the translate having both a *zero* and a *one* in this coordinate. ∎

We proved that there is a unique weight distribution for a nearly perfect covering code of type A and also a unique weight distribution for a translate of such code. A natural question is about the weight distribution of a $(2^r,1)$-nearly perfect code of type B and also about its translates. Do these codes have a unique weight distribution? To analyze the weight distribution of these codes we will define a few sets of variables which characterize for each weight the covered words and especially those words which are covered more than once in the code. For the rest of this section until Theorem 43 let $\mathcal{C}$ be a $(2^r,1)$-nearly perfect codes of type B. The first three sets of variables are as follows.

1) $\mathcal{L}_i$ - the set of codewords of weight i, for which the partner of each codeword has weight $i-2$, $2 \leqslant i \leqslant 2^r$.
2) $\mathcal{M}_i$ - the set of codewords of weight i, for which the partner of each codeword has weight i, $0 \leqslant i \leqslant 2^r$.
3) $\mathcal{R}_i$ - the set of codewords of weight i, for which partner of each codeword has weight $i+2$, $0 \leqslant i \leqslant 2^r-2$.

Note, that the partner of a codeword in $\mathcal{R}_i$ is a codeword in $\mathcal{L}_{i+2}$ and vice versa and the partner of a codeword in $\mathcal{M}_i$ is also a codeword in $\mathcal{M}_i$.

The following two lemmas can be observed from our previous results and they will be used to prove the main theorem of this section

Lemma 37. *If $d(x,y)=1$ for $x,y \in \mathbb{F}_2^{2^r}$, then the N-redundancies of $\{x,y\}$ is at most one.*

Proof. If $x \in \mathcal{C}$, then x is covered once by $\mathcal{C}$ and hence w.l.o.g. we assume that $x \notin \mathcal{C}$. By Corollary 1 we have that $\mathcal{B}_1(x)$ has exactly one word which is covered twice by $\mathcal{C}$. Since $y \in \mathcal{B}_1(x)$ it follows that the N-redundancies of $\{x,y\}$ is at most one. ∎

Lemma 38.

- *If $x \in \mathcal{L}_i$, then $\mathcal{B}_1(x)$ contains exactly two words that are covered twice by $\mathcal{C}$, and these two words have weight $i-1$.*
- *If $x \in \mathcal{M}_i$, then $\mathcal{B}_1(x)$ contains exactly two words that are covered twice by $\mathcal{C}$, one word of weight $i-1$ and one word of weight $i+1$.*
- *If $x \in \mathcal{R}_i$, then $\mathcal{B}_1(x)$ contains exactly two words that are covered twice by $\mathcal{C}$, and these two words have weight $i+1$.*

Proof. By the definition of the sets $\mathcal{L}_i$, $\mathcal{M}_i$, and $\mathcal{R}_i$, and considering their partners we have that each such pair covers two words which are covered twice. By Corollary 9 this implies that these are exactly all the words which are covered twice. ∎

The other sets of variables that are going to be used are as follows.

1) D_i - the number of codewords of weight i is $\mathcal{C}$.
2) T_i - the set of words of weight i that are covered twice by $\mathcal{C}$.
3) $\mathcal{U}_i^2$ - the set of words of weight i that are covered once by $\mathcal{C}$ and have an adjacent word of weight $i+1$ that is covered twice by $\mathcal{C}$.
4) $\mathcal{U}_i^1$ - the set of words of weight i that are covered once by $\mathcal{C}$ and have no adjacent word of weight $i+1$ that is covered twice by $\mathcal{C}$.
5) $\mathcal{V}_i^2$ - the set of words of weight i that are covered once by $\mathcal{C}$ and have an adjacent word of weight $i-1$ that is covered twice by $\mathcal{C}$.

There are many equalities that tie together the variables which were defined. Some of these equalities are given in the following lemma.

Lemma 39.

$$|\mathcal{L}_{i+2}| = |\mathcal{R}_i|\,, \tag{12}$$

$$D_i = |\mathcal{L}_i| + |\mathcal{M}_i| + |\mathcal{R}_i|\,, \tag{13}$$

$$\binom{n}{i} = (n+i-1)D_{i-1} - \frac{1}{2}|\mathcal{M}_{i-1}| - |\mathcal{R}_{i-1}| + D_i + (i+1)D_{i+1} - \frac{1}{2}|\mathcal{M}_{i+1}| - |\mathcal{L}_{i+1}|\,, \tag{14}$$

$$|T_i| = |\mathcal{R}_{i-1}| + \frac{1}{2}|\mathcal{M}_{i-1}| + |\mathcal{L}_{i+1}| + \frac{1}{2}|\mathcal{M}_{i+1}| = 2\,|\mathcal{R}_{i-1}| + \frac{1}{2}|\mathcal{M}_{i-1}| + \frac{1}{2}|\mathcal{M}_{i+1}|\,, \tag{15}$$

$$\binom{n}{i} = |T_i| + \left|\mathcal{U}_i^2\right| + \left|\mathcal{U}_i^1\right|\,, \tag{16}$$

$$\left|\mathcal{V}_i^2\right| = \left|\mathcal{U}_i^1\right| + |\mathcal{M}_i|\,. \tag{17}$$

Proof. The partner of each word in $\mathcal{L}_{i+2}$ is in $\mathcal{R}_i$ and hence $|\mathcal{L}_{i+2}| = |\mathcal{R}_i|$.

By definition D_i is the number of codewords of weight i and each such codeword is either in $\mathcal{L}_i$ or in $\mathcal{M}_i$ or $\mathcal{R}_i$. Hence, $D_i = |\mathcal{L}_i| + |\mathcal{M}_i| + |\mathcal{R}_i|$.

The $\binom{n}{i}$ words of weight i are covered by codewords of weight $i-1$, i, or $i+1$. Each codeword of weight $i-1$ covers $n-i+1$ words of weight i, each codeword of weight i covers only one word of weight i, and each codeword of weight $i+1$ covers $i+1$ words of weight i. In this enumeration, some words that are covered twice by a pair of codewords. Each pair of codewords, one from $\mathcal{R}_{i-1}$ and one from $\mathcal{L}_{i+1}$ covers two words of weight i twice and hence we have to remove either $2\,|\mathcal{R}_{i-1}|$ or $2\,|\mathcal{L}_{i+1}|$ or $|\mathcal{R}_{i-1}| + |\mathcal{L}_{i+1}|$ from this enumeration. Each pair of codewords from $\mathcal{M}_{i-1}$ covers one word of weight i twice and hence we have to remove $\frac{1}{2}|\mathcal{M}_{i-1}|$ from the computation. Similarly, each pair of codewords from $\mathcal{M}_{i+1}$ covers one word of weight i twice and hence we have to remove $\frac{1}{2}|\mathcal{M}_{i+1}|$ from the computation. this analysis implies Eq. (14).

Words which removed from the computation in Eq. (14), i.e.,

$$|\mathcal{R}_{i-1}| + |\mathcal{L}_{i+1}| + \frac{1}{2}|\mathcal{M}_{i-1}| + \frac{1}{2}|\mathcal{M}_{i+1}|$$

are covered twice by $\mathcal{C}$. This implies Eq. (15).

Each word of weight i is either covered twice by $\mathcal{C}$ or covered only once. If it is covered only once then either it has one adjacent word of weight $i+1$ which is covered twice (no more than one by Corollary 1 since T_i does not contain codewords) or all its adjacent words of weight $i+1$ are covered once. This implies Eq. (16).

For Eq. (17) we consider words of weight i which are covered once by $\mathcal{C}$ and do not have adjacent words of weight $i-1$ which are covered twice by $\mathcal{C}$. For each word x that is covered once, $\mathcal{B}_1(x)$ has at least one word that is covered twice (if x is a codeword, then there are exactly two such words, and if x is not a codeword, then there is exactly one such word.). Hence, $\mathcal{U}_i^1 \subseteq \mathcal{V}_i^2$. The other words of weight i are in T_i and they are not contained in $\mathcal{V}_i^2$ and not contained in $\mathcal{U}_i^2$ (i.e., they are not contained in $\mathcal{V}_i^2 \cup \mathcal{U}_i^2$). The only words in $\mathcal{U}_i^2$, that have an adjacent word with weight $i-1$ that is covered twice, are the words of $\mathcal{M}_i$ and hence Eq. (17) is implied. ∎

We assume now that the values of $|\mathcal{L}_i|$, $|\mathcal{M}_i|$, and $|\mathcal{R}_i|$, are known for each $0 \leqslant i \leqslant k$. We also assume that the value of $|T_i|$ is known for each $0 \leqslant i \leqslant k-1$.

Lemma 40. *The values of* $\left|\mathcal{U}_{k-1}^1\right|$ *and* $\left|\mathcal{U}_{k-1}^2\right|$ *can be determined.*

Proof. In an $(n,1)$-nearly perfect code of type B there are no adjacent codewords, and hence each codeword is covered exactly once. Hence, if $x \in T_{k-2}$, then x is not a codeword and by Corollary 1 exactly one word in $\mathcal{B}_1(x)$, which is x, is covered twice and the $n-k+2$ words of weight $k-1$ in $\mathcal{B}_1(x)$ are covered once. By the definition of $\mathcal{V}_{k-1}^2$ all these $n-k+2$ words are in $\mathcal{V}_{k-1}^2$ and $\mathcal{V}_{k-1}^2$ contains only these words. By lemma 38, the only words of $\mathcal{V}_{k-1}^2$ that have more than one (exactly two) adjacent words in T_{k-2} are all the words of $\mathcal{L}_{k-1}$ and hence we have (using also Eq. (17)) that

$$|T_{k-2}| = \frac{|\mathcal{L}_{k-1}| + \left|\mathcal{V}_{k-1}^2\right|}{n-k+2} = \frac{|\mathcal{L}_{k-1}| + \left|\mathcal{U}_{k-1}^1\right| + |\mathcal{M}_{k-1}|}{n-k+2}. \tag{18}$$

Since the values of $|\mathcal{L}_{k-1}|$, $|\mathcal{M}_{k-1}|$, and $|T_{k-2}|$ are known, it follows by Eq. (18) that we can determine the value of $\left|\mathcal{U}_{k-1}^1\right|$. Since the values of $|T_{k-1}|$ and $\left|\mathcal{U}_{k-1}^1\right|$ are known, it follows from Eq. (16) that we can determine the value of $\left|\mathcal{U}_{k-1}^2\right|$. ∎

Lemma 41. *The value of* $|T_k|$ *can be determined.*

Proof. In an $(n,1)$-nearly perfect code of type B there are no adjacent codewords, and hence each codeword is covered exactly once. Hence, if $x \in T_{k-1}$, then x is not a codeword and by Corollary 1 exactly one word in $\mathcal{B}_1(x)$, which is x, is covered twice and the $n-k+1$ words of weight k in $\mathcal{B}_1(x)$ are covered once. By the definition of $\mathcal{U}_{k-1}^1$ also for each word in $\mathcal{U}_{k-1}^1$ all $n-k+1$ the adjacent words of weight k are covered once. Hence, if we take a word $x \in T_k$ and convert one of its *ones* to a *zero* to obtain a word x', then x' will be a word in $\mathcal{U}_{k-1}^2$.

Now, for each $x \in \mathcal{U}_{k-1}^2$ we count the number of adjacent words which are contained in T_k. If x is not a codeword or $x \in \mathcal{M}_{k-1}$, i.e., $x \in \mathcal{U}_{k-1}^2 \setminus \mathcal{R}_{k-1}$, then by Corollary 1, $\mathcal{B}_1(x)$ contains exactly one word of weight k which is covered twice. If x is a codeword, i.e., $x \in \mathcal{R}_{k-1}$, then by Lemma 38, $\mathcal{B}_1(x)$ contains exactly two words which are covered twice. Thus, the number of words of weight k which are covered twice is

$$|T_k| = \frac{\left|\mathcal{U}_{k-1}^2\right| + |\mathcal{R}_{k-1}|}{k}, \tag{19}$$

where the division by k comes from the fact each word is counted once for each *one* which is converted to a *zero*.

By Lemma 40 we have that $\left|\mathcal{U}_{k-1}^2\right|$ can be determined and since $|\mathcal{R}_{k-1}|$ is known, it follows by Eq. (19) that $|T_k|$ can be determined. ∎

Lemma 42. *The values of* $|\mathcal{L}_{k+1}|$, $|\mathcal{M}_{k+1}|$, *and* $|\mathcal{R}_{k+1}|$, *can be determined.*

Proof. By Eq. (12) we have that $|\mathcal{L}_{k+1}| = |\mathcal{R}_{k-1}|$. By Lemma 41 the value of T_k is determined. Hence, by Eq. (15) the value of $|\mathcal{M}_{k+1}|$ can be determined. This implies by Eq. (14) and the fact that D_{k-1} and D_k were determined, that also D_{k+1} can be determined. Hence, by Eq. (13) also $|\mathcal{R}_{k+1}|$ can be determined. ∎

Theorem 43. *All* $(2^r,1)$*-nearly perfect codes of type B have the same weight distribution.*

Proof. The proof is by induction on k to compute D_k, the number of codewords with weight k, where we require that the values of $|\mathcal{L}_i|$, $|\mathcal{M}_i|$, and $|\mathcal{R}_i|$, are known for each $0 \leqslant i \leqslant k$ and also that the value of $|T_i|$ is known for each $0 \leqslant i \leqslant k-1$. The initial conditions are for $k=2$:

$$|\mathcal{L}_0| = |\mathcal{M}_0| = 0, \ \ |\mathcal{R}_0| = 1, \ \ |\mathcal{L}_1| = |\mathcal{M}_1| = |\mathcal{R}_1| = 0, \ \ |\mathcal{L}_2| = 1, \ \ |\mathcal{M}_2| = |\mathcal{R}_2| = 0,$$

$$|T_0| = 0, \ \ |T_1| = 2.$$

In the induction step $|T_k|$ is determined by Lemma 41 and $|\mathcal{L}_{k+1}|$, $|\mathcal{M}_{k+1}|$, $|\mathcal{R}_{k+1}|$ are determined in Lemma 42. Thus, by Eq. (13) also D_{k+1} is determined. ∎

Corollary 44. *The values of* $|\mathcal{L}_i|$, $|\mathcal{M}_i|$, *and* $|\mathcal{R}_i|$, *of all the* $(2^r,1)$*-nearly perfect codes of type B are the same for all* $0 \leqslant i \leqslant n$.

Theorem 45. *If* $\mathcal{C}$ *is a* $(2^r,1)$*-nearly perfect code of type B, where* $n = 2^r - 1$, *then*

$$\mathcal{A}_0(\mathcal{C}) = \mathcal{A}_n(\mathcal{C}) = 1, \ \mathcal{A}_{n+1}(\mathcal{C}) = 0$$

and for $1 \leqslant i \leqslant n-1$,

$$\mathcal{A}_i(\mathcal{C}) = \alpha_i + \beta_{i-1} = \frac{\binom{n}{i} + n\Delta_i}{n+1} + \frac{\binom{n}{i-1} - \Delta_{i-1}}{n+1} = \frac{\binom{n+1}{i} + n\Delta_i - \Delta_{i-1}}{n+1}$$

Proof. By Theorem 43 all the $(2^r,1)$-nearly perfect codes have the same weight distributions. This weight distribution can be computed by the codes constructed in Theorem 17 and Corollary 19 and from Eqs. (6) and (8). ∎

Theorem 46. *There are two possible weight distributions for a translate of a* $(2^r,1)$*-nearly perfect codes of type B.*

Proof. In a translate of nearly perfect code of type B, there is no codeword of weight zero. To cover the all-zero word, by Corollary 1 the code must have either one or two codewords of weight one. It is easy to verify that these two scenarios are possible. The values of In both cases, we can apply the same analysis as done in this section and the weight distribution is unique for each case if we have initial conditions as were given in Theorem 43. $|\mathcal{M}_2|$ and $|\mathcal{R}_2|$ are also forced to cover each word of weight one and to satisfy Corollary 1. If there is a unique codeword of weight one then the initial conditions are

$$|\mathcal{L}_0| = |\mathcal{M}_0| = |\mathcal{R}_0| = 0, \ \ |\mathcal{L}_1| = |\mathcal{M}_1| = 0, \ \ |\mathcal{R}_1| = 1, \ \ |\mathcal{L}_2| = 0, \ \ |\mathcal{M}_2| = 2, \ \ |\mathcal{R}_2| = \frac{2^r - 4}{2},$$

$$|T_0| = 0, \ \ |T_1| = 1.$$

If there are exactly two codewords of weight one then the initial conditions are

$$|\mathcal{L}_0| = |\mathcal{M}_0| = |\mathcal{R}_0| = 0, \ \ |\mathcal{L}_1| = 0, \ \ |\mathcal{M}_1| = 2, \ \ |\mathcal{R}_1| = 0, \ \ |\mathcal{L}_2| = 0, \ \ |\mathcal{M}_2| = 0, \ \ |\mathcal{R}_2| = \frac{2^r - 2}{2},$$

$$|T_0| = 1, \ \ |T_1| = 0.$$

∎

The computation of the weight distribution for the two possible translates of $(2^r,1)$-nearly perfect codes is performed similar to previous computations. When there is exactly one codeword of weight one we have $|T_0| = 0$, $|T_1| = 1$, and we form the code

$$\mathcal{C} \triangleq \{(c,0) \ : \ c \in \mathcal{C}_1\} \cup \{(c,1) \ : \ c \in \mathcal{C}_2\},$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are two disjoint translates of a Hamming code. We have $\mathcal{A}_0(\mathcal{C}) = \mathcal{A}_{n+1}(\mathcal{C}) = 0$ and $\mathcal{A}_i(\mathcal{C}) = \beta_i + \beta_{i-1}$ for $1 \leqslant i \leqslant n$.

When there are exactly two codewords of weight one we have $|T_0| = 1$, $|T_1| = 0$, and we form the code

$$\mathcal{C} \triangleq \{(c,0) \;:\; c \in \mathcal{C}_1\} \cup \{(c,1) \;:\; c \in \mathcal{C}_2\},$$

where $\mathcal{C}_2$ is a Hamming code and $\mathcal{C}_1$ is a translate of $\mathcal{C}_2$. We have $\mathcal{A}_0(\mathcal{C}) = \mathcal{A}_{n+1}(\mathcal{C}) = 0$ and $\mathcal{A}_i(\mathcal{C}) = \beta_i + \alpha_{i-1}$ for $1 \leqslant i \leqslant n$.

For nearly perfect covering codes of type C, the computation of the possible weight distributions is more complicated. Open problems on these codes will be discussed in Section VI.

V. Balanced Nearly Perfect Covering Codes

There might be many types of nearly perfect covering codes with special properties. Such an example can be codes of type A in which for each given coordinate the number of pairs of codewords $\{x,y\}$ for which $d(x,y) = 1$ and x and y differ in the given coordinate is 2^{2^r-2r-1}. In other words, this number is the same for all coordinates. Such a code will be called a ***balanced nearly perfect covering code*** and such a code can be constructed recursively as follows.

A ***self-dual*** sequence is a binary cyclic sequence that is equal to its complement. The following two cyclic sequences $\mathcal{S}_1 = [0001101111100100]$ and $\mathcal{S}_2 = [0001101011100101]$ are self-dual sequences of length 16. We consider all the 32 words obtained by any 8 consecutive symbols of $\mathcal{S}_1$ and $\mathcal{S}_2$. In these 32 words, we have 16 even-weight words of length 8 and 16 odd-weight words of length 8. Let $\mathcal{C}$ be the code obtained from these 32 words. Let $\mathcal{C}_e$ be the 16 even-weight words of $\mathcal{C}$ and $\mathcal{C}_o$ be the 16 odd-weight words of $\mathcal{C}$. $\mathcal{C}_e$ is an even translate of an extended Hamming code of length 8 and $\mathcal{C}_o$ is an odd translate of an extended Hamming code of length 8. Therefore, by Corollary 28 their union is a nearly perfect covering code. Finally, for each one of the 8 coordinates, there are exactly two pairs of words from $\mathcal{C}_e$ and $\mathcal{C}_o$ which differ exactly in this coordinate, and hence the code is balanced. To obtain a nearly perfect covering code from this translate we have to translate it by one of its codewords.

Example 1. *Three more pairs of sequences can be used as $\mathcal{S}_1$ and $\mathcal{S}_2$ (each two pairs have disjoint codewords of length 8 and they can be obtained from each other by decimation)*

$$\mathcal{S}_1 = [0100111110110000], \quad \mathcal{S}_2 = [0100111010110001],$$
$$\mathcal{S}_1 = [0111011110001000], \quad \mathcal{S}_2 = [0111011010001001],$$
$$\mathcal{S}_1 = [0010001011011101], \quad \mathcal{S}_2 = [0010001111011100].$$

Generally, we consider 2^{2^r-2r-1} self-dual sequences of length 2^{r+1}. Let $\mathcal{C}$ be the set of 2^{2^r-r} words obtained by each 2^r consecutive symbols in these self-dual words. Assume further that all these 2^{2^r-r} words of length 2^r are different. Let $\mathcal{C}_e$ be the set of even-weight words in $\mathcal{C}$ and $\mathcal{C}_o$ be the set of odd-weight words in $\mathcal{C}$. Assume further that $\mathcal{C}_e$ and $\mathcal{C}_o$ are two translates of extended Hamming codes of length 2^r (one even translate and one odd translate). Assume further that the 2^{2^r-2r-1} self-dual sequences can be ordered in pairs

$$\mathcal{P}_i = ([X \;\bar{X}], [X' \;\bar{X}']), \quad 1 \leqslant i \leqslant 2^{2^r-2r-2},$$

where X and X' are sequence of length 2^r which start with a *zero* and differ only in their last symbol, and $\bar{X}$ is the binary complement of X.

This partition into pairs of self-dual sequences implies that the words of $\mathcal{C}_e$ and $\mathcal{C}_o$ can be partitioned into pairs of words defined by the following set (see also the proof of Lemma 51).

$$\mathcal{Q} \triangleq \{\{x,y\} \ : \ x \in \mathcal{C}_e, \ y \in \mathcal{C}_o, \ d(x,y) = 1\},$$

where $\mathcal{Q}$ contains exactly 2^{2^r-r-1} pairs of words and each word of $\mathcal{C}_e$ and each word of $\mathcal{C}_o$ is contained in exactly one such pair. Such a definition for $\mathcal{Q}$ and the definition of the pairs in $\mathcal{P}_i$, $1 \leqslant i \leqslant 2^{2^r-2r-2}$, imply that for each one of the 2^r coordinates, there are 2^{2^r-2r-1} pairs of words that differ in this coordinate.

For each pair of self-dual sequences $\mathcal{P}_i = ([X \ \bar{X}],[X' \ \bar{X}'])$, $1 \leqslant i \leqslant 2^{2^r-2r-2}$, and any word $V = 0Z$ of length 2^r, where Z is an even-weight word of length $2^r - 1$ we form the following pair

$$\mathcal{P}_{iV} = ([V \ X+V \ \bar{V} \ X+\bar{V}],[V \ X'+V \ \bar{V} \ X'+\bar{V}]).$$

The following lemma is an immediate observation.

Lemma 47. *The two sequences in $\mathcal{P}_{iY}$ are self-dual sequences. They have the form $[X_1 \ X_2 \ \bar{X}_1 \ \bar{X}_2]$ and $[X_1 \ X_2' \ \bar{X}_1 \ \bar{X}_2']$, where X_1 and X_2 are words of length 2^r that start with a* zero.

Let $\mathcal{S}$ be the set of even-weight words of length 2^r that start with a *zero*. Let $\mathbb{C}$ be the code defined by taking the union of all the sequences in these pairs and from each sequence taking 2^{r+2} codewords obtained from the 2^{r+2} consecutive 2^{r+1} bits of the sequences.

The construction of the pair of sequences is very similar to the constructions presented in [7,9,11]. The same code was defined and analyzed for another purpose in [4]. The following observations lead to the main result. The first lemma was proved in [7,9,11].

Lemma 48. *All the words of length 2^{r+1} obtained from all the pairs $\mathcal{P}_{iV}$, $1 \leqslant i \leqslant 2^{2^r-2r-2}$, $V \in \mathcal{S}$ are distinct.*

Corollary 49. *The code $\mathbb{C}$ contains $2^{2^{r+1}-r-1}$ codewords.*

The following lemma was proved in [4].

Lemma 50. *The code $\mathbb{C}$ is a $(2^{r+1},1)$-covering code.*

Proof. The form of the two sequences in a pair implies that we can partition the $2^{2^{r+1}-r-1}$ codewords of $\mathbb{C}$ into two sets, one with words of even weight and one with words of odd weight. we claim that there are no two codewords with distance 2. Assume for the contrary that there are two such distinct codewords, (X_1X_2) and (Y_1Y_2) where X_1, X_2, Y_1, Y_2 are words of length 2^r and $d((X_1X_2),(Y_1Y_2)) = 2$. The associated two self-dual sequences (not necessarily distinct) of length 2^{r+1} are

$$[X_1 \ X_2 \ \bar{X}_1 \ \bar{X}_2] \ \text{ and } \ [Y_1 \ Y_2 \ \bar{Y}_1 \ \bar{Y}_2] \ .$$

We distinguish now between two cases:
Case 1: $d(X_1,Y_1) = 2$ and $X_2 = Y_2$ (the case $d(X_2,Y_2) = 2$ and $X_1 = Y_1$ is equivalent). The code $\mathcal{C}$ contains the codewords $X_1 + X_2$ and $Y_1 + Y_2$, where $d(X_1 + X_2, Y_1 + Y_2) = 2$, a contradiction.
Case 2: $d(X_1,Y_1) = 1$ and $d(X_2,Y_2) = 1$. The code $\mathcal{C}$ contains the codewords $X_1 + X_2$ and $Y_1 + Y_2$, where either $d(X_1 + X_2, Y_1 + Y_2) = 2$ or $d(X_1 + X_2, Y_1 + Y_2) = 0$. $d(X_1 + X_2, Y_1 + Y_2) = 2$ is not possible since the code $\mathcal{C}$ does not contains two codewords with distance 2. $d(X_1 + X_2, Y_1 + Y_2) = 0$

implies that the coordinate in which X_1 and Y_1 differ is the same coordinate as X_2 and Y_2 differ. This implies that the two distinct self-dual sequences

$$[X_1\ X_2\ \bar{X}_1\ \bar{X}_2] \text{ and } [Y_1\ Y_2\ \bar{Y}_1\ \bar{Y}_2]\ . \tag{20}$$

are obtained from the same self-dual sequences $[X_1 + X_2\ \bar{X}_1 + X_2] = [Y_1 + Y_2\ \bar{Y}_1 + Y_2]$. The two sequences in (20) differ in four positions, each two are separated by $2^r - 1$ equal positions. But, by our choice of $V = 0Z$ of length 2^r, where Z has even weight, cannot yield two sequences that differ in exactly one position among 2^r consecutive coordinates, a contradiction.

Hence, the minimum distance in each set of codewords is four, which implies that each set of words has the parameters of the extended Hamming code. Thus, $\mathbb{C}$ is a $(2^{r+1}, 1)$-nearly perfect code. ∎

Lemma 51. *The code $\mathbb{C}$ is a balanced $(2^{r+1}, 1)$-nearly perfect code.*

Proof. By Corollary 49 and Lemma 50 we have that $\mathbb{C}$ is a $(2^{r+1}, 1)$-nearly perfect code. Two pairs of sequences differ in positions 2^{r+1} and 2^{r+2}. These two positions are associated with the last coordinate of the codewords that start in the first bit and bit $2^{r+1} + 1$ of these two sequences. Since the codewords are formed from the 2^{r+1} consecutive bits in each pair of such sequences, the codewords which start in the next bits differ in the previous positions and so on. It follows that for each position γ there are exactly two pairs of codewords from these two sequences which differ exactly in position γ. Therefore, $\mathbb{C}$ is a balanced $(2^{r+1}, 1)$-nearly perfect code. ∎

Example 2. *For $r = 3$, there is one pair given by*

$$\mathcal{P} = ([00011011\ 11100100], [00011010\ 11100101])$$

Applying the recursion we obtain the following 64 pairs (the first eight and the last four are given), where the index is their place in the lexicographic order and the first 8 bits are ordered by this lexicographic order

$$\mathcal{P}_1 = ([00000000\ 00011011\ 11111111\ 11100100], [00000000\ 00011010\ 11111111\ 11100101])$$
$$\mathcal{P}_2 = ([00000011\ 00011000\ 11111100\ 11100111], [00000011\ 00011001\ 11111100\ 11100110])$$
$$\mathcal{P}_3 = ([00000101\ 00011110\ 11111010\ 11100001], [00000101\ 00011111\ 11111010\ 11100000])$$
$$\mathcal{P}_4 = ([00000110\ 00011101\ 11111001\ 11100010], [00000110\ 00011100\ 11111001\ 11100011])$$
$$\mathcal{P}_5 = ([00001001\ 00010010\ 11110110\ 11101101], [00001001\ 00010011\ 11110110\ 11101100])$$
$$\mathcal{P}_6 = ([00001010\ 00010001\ 11110101\ 11101110], [00001010\ 00010000\ 11110101\ 11101111])$$
$$\mathcal{P}_7 = ([00001100\ 00010111\ 11110011\ 11101000], [00001100\ 00010110\ 11110011\ 11101001])$$
$$\mathcal{P}_8 = ([00001111\ 00010100\ 11110000\ 11101011], [00001111\ 00010101\ 11110000\ 11101010])$$
$$\vdots$$
$$\mathcal{P}_{61} = ([01110111\ 01101100\ 10001000\ 10010011], [01110111\ 01101101\ 10001000\ 10010010])$$
$$\mathcal{P}_{62} = ([01111011\ 01100000\ 10000100\ 10011111], [01111011\ 01100001\ 10000100\ 10011110])$$
$$\mathcal{P}_{63} = ([01111101\ 01100110\ 10000010\ 10011001], [01111101\ 01100111\ 10000010\ 10011000])$$
$$\mathcal{P}_{64} = ([01111110\ 01100101\ 10000001\ 10011010], [01111110\ 01100100\ 10000001\ 10011011])$$

VI. Conclusion and Future Work

The structure of $(2^r,1)$-nearly perfect (covering) codes was considered. It was proved that there are three types of such codes which depend on the distance between each codeword to its nearest codeword. The structure of these codes and their weight distribution are examined in the paper. Constructions of a large number of codes of each type were given. Our exposition leads to a few interesting open problems.

1) What is the minimum possible number of pairs $\{x,y\}$ such that $x,y \in \mathcal{C}$ and $d(x,y)=2$ in a $(2^r,1)$-nearly perfect code of Type C?
2) What is the minimum number k of disjoint spheres of type A that cover the same area as k disjoint spheres of type B?
3) Is it true that there exist two perfect codes of length 2^r-1 and intersection k if and only if there exists a $(2^r,1)$-nearly perfect code of type C with exactly k pairs $\{x,y\}$ such that $d(x,y)=1$?
4) What is the smallest size of two different sets of disjoint spheres of type A and type B which cover the same area?
5) Is there a characterization of nearly perfect covering codes of type B similar to the one given in Corollary 29 (union of an extended Hamming code and an odd translate of extended Hamming code)? Given a $(2^r,1)$-nearly perfect code of type B, from each pair of codewords one can remove one of the partners from the code. The obtained code has minimum Hamming distance 3. Is there a way to remove such a codeword from each pair of codewords and to delete one of the coordinates such that the obtained code will be a Hamming code?
6) Is there a way in which a nearly perfect covering code of type A can be constructed from a nearly perfect code of type B in a similar way to the construction in the proof of Theorem 31?
7) It is easy to verify that $(2^r,1)$-nearly perfect codes of type C can have various different weight distributions. Is it possible to characterize these weight distributions? Can they be characterized based on the weight distributions of type A or type B?
8) We have proved that there exists a balanced $(2^r,1)$-nearly perfect code of type A. Does there exist a similar code of type B? First, does there exist a $(2^r,1)$-nearly perfect code in which for each two coordinates i and j, there exist two codewords $x,y \in \mathcal{C}$, such that $d(x,y)=2$ and these two codewords x and y differ in coordinates i and j?
9) Does there exist such a related balanced code (a simple enumeration shows that the number of such pairs cannot be the same for each pair of coordinates)? One possible simple definition for balanced $(2^r,1)$-nearly perfect codes of type B is that pair of codewords differ only in coordinates i and $i+1$, $1 \leqslant 2^r-1$ or coordinates 1 and 2^r and the number of pair of codewords for each pair of coordinates is the same. Another possibility is that they differ in coordinates i and j, for each $i<j$, such that $j-i$ is an odd integer and the number of pairs of codewords for each pair of coordinates is the same. Are there balanced nearly perfect codes for each definition?

References

[1] S. V. Avgustinovich, O. Heden, and F. I. Solov'eva, *On intersection of perfect binary codes*, *Bayreuther Mathematische Schriften,* 71 (2005), 8–13.

[2] S. V. Avgustinovich, O. Heden, and F. I. Solov'eva, *On intersection problem for perfect binary codes*, *Designs, Codes and Crypto.,* 39 (2006), 317–322.

[3] R. D. Baker, J. H. van Lint, and R. W. Wilson, *On the Preparata and Goethals codes*, *IEEE Trans. Infor. Theory,* 29 (1983), 342–345.

[4] Y. M. CHEE, T. ETZION, H. TA, AND V. K. VU, *On de Bruijn Covering Sequences and Arrays*, to be presented in the IEEE Symposium on Information Theory, Athens, Greece 2024.
[5] G. COHEN, I. HONKALA, S. LITSYN, AND A. LOBSTEIN, *Covering Codes*, *North-Holland, Amsterdam, 1997.*
[6] T. ETZION, *Perfect Codes and Related Structures*, *World Scientific, 2022.*
[7] T. ETZION, *Sequences and the de Bruijn Graph: Properties, Constructions, and Applications*, *Elsevier, 2024*.
[8] T. ETZION AND G. GREENBERG, *Constructions for perfect mixed codes and other covering codes*, *IEEE Trans. Infor. Theory,* 39 (1993), 209–214.
[9] T. ETZION AND A. LEMPEL, *Construction of de Bruijn sequences of minimal complexity*, *IEEE Trans. Infor. Theory,* 30 (1984), 705–709.
[10] T. ETZION AND B. MOUNITS, *Quasi-perfect codes with small distance*, *IEEE Trans. Infor. Theory,* 51 (2005), 3928–3946.
[11] T. ETZION AND K. G. PATERSON, *Near optimal single-track Gray codes*, *IEEE Trans. Infor. Theory,* 42 (1996), 779–789.
[12] T. ETZION AND A. VARDY, *Perfect binary codes: constructions, properties, and enumeration*, *IEEE Trans. Infor. Theory,* 40 (1994), 754–763.
[13] T. ETZION AND A. VARDY, *On perfect codes and tilings: problems and solutions*, *SIAM J. on Discrete Math.,* 11 (1998), 203–223.
[14] M. K. FORT, JR. AND G. A. HEDLUND, *Minimal coverings of pairs by triples*, *Pacific J. Math.,* 8 (1958), 709–717.
[15] J. M. GOETHALS AND S. L. SNOVER, *Nearly perfect binary codes*, *Disc. Math.,* 1-3 (1972), 65–88.
[16] O. HEDEN, F. I. SOLOV'EVA, AND I. YU. MOGILNYKH, *Intersection of perfect binary codes*, *2010 IEEE Region 8 International Conference on Computational Technologies in Electrical and Electronics Engineering (SIBIRCON),* (2010), 52–54.
[17] S. M. JOHNSON, *A new upper bound for error-correcting codes*, *IRE Trans. Infor. Theory,* 8 (1962), 203–207.
[18] W. M. KANTOR, *On the inequivalence of generalized Preparata codes*, *IEEE Trans. Infor. Theory,* 29 (1983), 345–348.
[19] K. LINDSTRÖM, *All nearly perfect codes are known*, *Infor. and Control,* 35 (1977), 40–47.
[20] F. P. PREPARATA, *A class of optimum nonlinear double-error-correcting codes*, *Infor. Contr.,* 13 (1968), 378–400.
[21] R. STRUIK, *An Improvement of the Van Wee Bound for Binary Linear Covering Codes*, *IEEE Trans. Infor. Theory,* 40 (1994), 1280–1284.
[22] G. J. M. VAN WEE, *Improved sphere bounds on the covering radius of codes*, *IEEE Trans. Infor. Theory,* 34 (1988), 237–245.