On Nearly Perfect Covering Codes

Avital Boruchovsky, *Student Member, IEEE*, Tuvi Etzion, *Life Fellow, IEEE*, and Ron M. Roth, *Life Fellow, IEEE*

***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. This work studies such codes with covering radius 1. It is shown that the set of these codes can be partitioned into three families, depending on the distribution of the Hamming distances between neighboring codewords. General properties of these code families are presented, including a characterization of their weight and distance distributions. Constructions of codes for each of the families are presented. Finally, extended perfect covering codes are considered. Their punctured codes yield a variety of nearly perfect covering codes.**

***Index Terms*—Nearly perfect codes, Perfect codes, Sphere-covering bound, Weight distribution.**

I. Introduction

PERFECT codes are among the most fascinating structures in coding theory. They meet the well-known sphere-packing bound, yet they are very rare. Therefore, there have been many attempts to find either packing or covering codes that are "almost perfect." One class of such covering codes is the topic of this work.

All the codes herein are over the binary field $\mathbb{F}_2$, and by an (n, M) *code* (of length n and size M) we mean a subset $\mathcal{C} \subseteq \mathbb{F}_2^n$ of size $|\mathcal{C}| = M$. For integers $\ell \leqslant m$, we use the notation $[\ell : m]$ for the integer interval $\{\ell, \ell{+}1, \ldots, m\}$, with $[m]$ standing for $[1 : m]$.

A *translate* of an (n, M) code $\mathcal{C}$ is the set

$$\boldsymbol{e} + \mathcal{C} \triangleq \{\boldsymbol{e} + \boldsymbol{c} \,:\, \boldsymbol{c} \in \mathcal{C}\},$$

where $\boldsymbol{e} \in \mathbb{F}_2^n$ (and addition is over $\mathbb{F}_2$). When the all-zero word, $\mathbf{0}$, is a codeword in $\mathcal{C}$ we say that the code is *zeroed*. A translate $\boldsymbol{e} + \mathcal{C}$ with $\boldsymbol{e} \in \mathcal{C}$ is a zeroed code. When $\boldsymbol{e} \notin \mathcal{C}$ we say that $\boldsymbol{e} + \mathcal{C}$ is a non-zeroed translate.

The *(Hamming) distance* between two words $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}_2^n$ will be denoted by $\mathsf{d}(\boldsymbol{x}, \boldsymbol{y})$, and $\mathsf{w}(\boldsymbol{x})$ will denote the *weight* of $\boldsymbol{x}$, i.e., the size of the support, $\mathsf{Supp}(\boldsymbol{x})$, of $\boldsymbol{x}$ (the notation extends to integer vectors as well). The *radius-t ball* centered at a word $\boldsymbol{x} \in \mathbb{F}_2^n$ is denoted by

$$\mathfrak{B}_t(\boldsymbol{x}) \triangleq \{\boldsymbol{y} \in \mathbb{F}_2^n \,:\, \mathsf{d}(\boldsymbol{x}, \boldsymbol{y}) \leqslant t\}$$

(where, for simplicity of notation, we make the dependence on n implicit). We also define the boundary (sphere)

$$\partial\mathfrak{B}_t(\boldsymbol{x}) \triangleq \{\boldsymbol{y} \in \mathbb{F}_2^n \,:\, \mathsf{d}(\boldsymbol{x}, \boldsymbol{y}) = t\}$$

and

$$\mathcal{S}_t \triangleq \partial\mathfrak{B}_t(\mathbf{0}) = \{\boldsymbol{y} \in \mathbb{F}_2^n \,:\, \mathsf{w}(\boldsymbol{y}) = t\}.$$

The words in $\partial\mathfrak{B}_t(\boldsymbol{x})$ will be called the *t-neighbors* of $\boldsymbol{x}$. The *minimum distance* of an (n, M) code $\mathcal{C}$ is the smallest distance between any two distinct codewords in $\mathcal{C}$, and the distance of a word $\boldsymbol{x} \in \mathbb{F}_2^n$ from $\mathcal{C}$ is defined by $\mathsf{d}(\boldsymbol{x}, \mathcal{C}) = \min_{\boldsymbol{c} \in \mathcal{C}} \mathsf{d}(\boldsymbol{x}, \boldsymbol{c})$. The *covering radius* of $\mathcal{C}$ is defined by

$$R = \max_{\boldsymbol{x} \in \mathbb{F}_2^n} \mathsf{d}(\boldsymbol{x}, \mathcal{C}),$$

and we say that a code $\mathcal{C}$ is *R-covering* if its covering radius is at most R. When $\mathcal{C}$ is a linear code over $\mathbb{F}_2$ (of dimension $\log_2 M$) and H is any full-rank $r \times n$ parity-check matrix of $\mathcal{C}$ over $\mathbb{F}_2$ (where $r = n - \log_2 M$), the covering radius of $\mathcal{C}$ equals the smallest R such that every vector in $\mathbb{F}_2^r$ can be expressed as a linear combination (over $\mathbb{F}_2$) of R columns of H.

For an (n, M) code $\mathcal{C}$ with minimum distance $2R + 1$ we have the *sphere-packing bound*

$$M \cdot \sum_{i=0}^{R} \binom{n}{i} \leqslant 2^n, \tag{1}$$

and if $\mathcal{C}$ has covering radius R we have the *sphere-covering bound*

$$M \cdot \sum_{i=0}^{R} \binom{n}{i} \geqslant 2^n. \tag{2}$$

Perfect codes meet both bounds.

The sphere-packing bound for an (n, M) code with minimum distance $2R + 1$ was improved by Johnson [18] to

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

(see [22, p. 533]), and a code that meets this bound is called a *nearly perfect (packing) code*. When $R + 1$ divides $n + 1$, this bound coincides with (1). Codes that meet the bound (3) were considered in [15], [21]. There are two families of nontrivial codes that are nearly perfect yet not perfect. One family is the set of *shortened Hamming codes*. A second family consists of the *punctured Preparata codes*. These codes were first found by Preparata [25] and later others found many inequivalent codes with the same parameters [3], [19]. Moreover, these codes are very important in constructing other codes, e.g., see [9]. A comprehensive work on perfect codes and related codes can be found in [7]. For R-covering codes,

This work was supported in part by Grants Nos. 222/19, 1713/20, 2369/24, and 2462/24 from the Israel Science Foundation.

Avital Boruchovsky is with the Computer Science Department, Technion, Haifa 3200003, Israel (email: avital.bor@cs.technion.ac.il).

Tuvi Etzion is with the Computer Science Department, Technion, Haifa 3200003, Israel (email: etzion@cs.technion.ac.il).

Ron M. Roth is with the Computer Science Department, Technion, Haifa 3200003, Israel (email: ronny@cs.technion.ac.il).

an improvement on the sphere-covering bound was obtained by van Wee [27]. A simplified version of his bound was presented by Struik [26] and takes the form

$$M \cdot \left(\sum_{i=0}^{R} \binom{n}{i} - \frac{\binom{n}{R}}{\lfloor \frac{n}{R+1} \rfloor} \left(\left\lceil \frac{n+1}{R+1} \right\rceil - \frac{n+1}{R+1} \right) \right) \geqslant 2^n \quad (4)$$

(notice the similarity to (3)). When $R+1$ divides $n+1$, this bound coincides with (2). A code that meets the bound (4) will be called a *nearly perfect covering code*. For even n and $R = 1$, this bound becomes

$$M \geqslant \frac{2^n}{n}. \quad (5)$$

Except for perfect codes and some trivial codes, $R = 1$ is the only radius for which we currently know of codes that meet the bound (4); from (5), these codes have length $n = 2^r$ and size $M = 2^{2^r - r}$, for some positive integer r. A code with these parameters will be called a *a nearly perfect* 1*-covering code* (in short, *NP1CC*). In the case of linear codes, there is a very simple characterization of NP1CCs, as we show in the next example.

Example 1. Let $\mathcal{C}$ be an $(n{=}2^r, M{=}2^{n-r})$ linear code over $\mathbb{F}_2$ and let H be any full-rank $r \times n$ parity-check matrix of $\mathcal{C}$ over $\mathbb{F}_2$. Then $\mathcal{C}$ is 1-covering (and, hence, is an NP1CC), if and only if each nonzero vector in $\mathbb{F}_2^r$ appears as a column in H. Thus, there are two possible cases. The first is when the columns of H range over all the vectors of $\mathbb{F}_2^r$ (including the all-zero vector); the second case is similar, except that the all-zero column is replaced by some nonzero vector of $\mathbb{F}_2^r$. □

In this work, we consider the structure of general (not necessarily linear) NP1CCs. In Section II, we show that in any NP1CC $\mathcal{C}$, each codeword $\boldsymbol{c} \in \mathcal{C}$ has a unique other codeword $\boldsymbol{c}'$ in $\mathfrak{B}_2(\boldsymbol{c})$; this, in turn, induces a partition of the code $\mathcal{C}$ into pairs $\{\boldsymbol{c}, \boldsymbol{c}'\}$. Based on this property, we classify NP1CCs into three types:

- Type $\mathbb{A}$ codes, in which the codewords in every pair $\{\boldsymbol{c}, \boldsymbol{c}'\}$ are 1-neighbors (the first case in Example 1 belongs to this type),
- Type $\mathbb{B}$ codes, in which the codewords in every such pair are 2-neighbors (the second case in the example is of this type), and—
- Type $\mathbb{C}$ codes (which are all the remaining NP1CCs).

In Section III, we present constructions of codes for each type and then provide a complete characterization of NP1CCs of Type $\mathbb{A}$. In Section IV, we consider the weight and distance distributions of NP1CCs; in particular, we prove that there are exactly two possible weight distributions for all zeroed NP1CCs and two other weight distributions for all their non-zeroed translates. Moreover, we show that Type $\mathbb{A}$ and Type $\mathbb{B}$ codes are distance invariant. In Section V, we concentrate on a class of Type $\mathbb{A}$ codes in which the number of codeword pairs $\{\boldsymbol{c}, \boldsymbol{c}'\}$ that differ only on any given coordinate is the same for all coordinates. Extended NP1CCs are discussed in Section VI and are shown to provide—through puncturing—a method for obtaining new NP1CCs from others. A conclusion and a few problems for future research are presented in Section VII.

II. Structure of NP1CCs

In this section, we examine the structure of NP1CCs.

Let $\mathcal{C}$ be an (n, M) code. Given a word $\boldsymbol{x} \in \mathbb{F}_2^n$, we say that a codeword $\boldsymbol{c} \in \mathcal{C}$ *covers* $\boldsymbol{x}$ if $\boldsymbol{c} \in \mathfrak{B}_1(\boldsymbol{x})$. Clearly, if $\mathcal{C}$ is 1-covering then every word $\boldsymbol{x} \in \mathbb{F}_2^n$ is covered by at least one codeword of $\mathcal{C}$. The *over-covering* of a subset $\mathcal{Y} \subseteq \mathbb{F}_2^n$ (with respect to a 1-covering code $\mathcal{C}$) is defined by

$$\sum_{\boldsymbol{y} \in \mathcal{Y}} (|\mathfrak{B}_1(\boldsymbol{y}) \cap \mathcal{C}| - 1) = \Big(\sum_{\boldsymbol{y} \in \mathcal{Y}} |\mathfrak{B}_1(\boldsymbol{y}) \cap \mathcal{C}| \Big) - |\mathcal{Y}|.$$

Thus, while each word in $\mathcal{Y}$ is covered by at least one codeword of $\mathcal{C}$, the over-covering of $\mathcal{Y}$ measures how many *additional* codewords cover each one of the words in $\mathcal{Y}$. The following lemma follows from the analysis of Struik in [26] (although, as stated, it does not appear explicitly there).

Lemma 1 ([26])**.** *Let $\mathcal{C}$ be an (n, M) NP1CC and let $\boldsymbol{x} \in \mathbb{F}_2^n \setminus \mathcal{C}$ be a non-codeword. Then $\mathfrak{B}_1(\boldsymbol{x})$ 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. We provide the steps of the proof through pointers to [26]. It follows from Eq. (6) therein that for each non-codeword $\boldsymbol{x} \in \mathbb{F}_2^n \setminus \mathcal{C}$, the over-covering of the ball $\mathfrak{B}_1(\boldsymbol{x})$ is at least 1. Denoting by ϵ the average of these over-coverings, we then get that $\epsilon \geqslant 1$ (see Eq. (7) in [26]). Then, equality in the Van Wee bound (Eq. (9) in [26]) forces the equality $\epsilon = 1$, which means that the over-covering of each ball $\mathfrak{B}_1(\boldsymbol{x})$ must be exactly 1. □

Corollary 2. *Let $\mathcal{C}$ be an (n, M) NP1CC. For every non-codeword $\boldsymbol{x} \in \mathbb{F}_2^n \setminus \mathcal{C}$,*

$$|\mathfrak{B}_1(\boldsymbol{x}) \cap \mathcal{C}| \leqslant 2.$$

A non-codeword $\boldsymbol{x} \in \mathbb{F}_2^n \setminus \mathcal{C}$ for which $|\mathfrak{B}_1(\boldsymbol{x}) \cap \mathcal{C}| = 2$ will be called a *midword*.

While midwords differ from the remaining non-codewords in the size of the intersection $\mathfrak{B}_1(\boldsymbol{x}) \cap \mathcal{C}$, those sizes become the same if we look at balls of radius 2. This property, which we prove in the next theorem, will be instrumental in Section IV for deriving the weight and distance distributions of NP1CCs.

Theorem 3. *Let $\mathcal{C}$ be an (n, M) NP1CC. For every non-codeword $\boldsymbol{x} \in \mathbb{F}_2^n \setminus \mathcal{C}$,*

$$|\mathfrak{B}_2(\boldsymbol{x}) \cap \mathcal{C}| = \frac{n}{2} + 1.$$

Proof. Consider first the case where $\boldsymbol{x}$ is a midword. By Lemma 1, no other word in $\mathfrak{B}_1(\boldsymbol{x})$ is covered by two codewords; namely, the set of 1-neighbors of $\boldsymbol{x}$ consists of two codewords $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ (none of which is a 1-neighbor of a codeword) and $n-2$ non-codewords $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n-2}$ (none of which is a midword). Each $\boldsymbol{y}_i$, in turn, is covered by a unique codeword (which belongs to $\partial\mathfrak{B}_2(\boldsymbol{x})$). Conversely, each codeword in $\partial\mathfrak{B}_2(\boldsymbol{x})$ covers exactly two words among the $\boldsymbol{y}_i$'s (and none of the codewords $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$). We conclude that $\mathfrak{B}_2(\boldsymbol{x})$ contains exactly $n/2{+}1$ codewords: the codewords $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$, and $n/2-1$ codewords that cover the $\boldsymbol{y}_i$'s.

We next turn to the case where $\boldsymbol{x}$ is not a midword. One (and only one) of the 1-neighbors of $\boldsymbol{x}$ is a codeword, $\boldsymbol{c}$, and, by Lemma 1, there is a unique 1-neighbor $\boldsymbol{y}_0$ of $\boldsymbol{x}$ that is covered by two codewords. We distinguish between two cases.

Case 1: $\boldsymbol{y}_0 = \boldsymbol{c}$. The $n-1$ remaining 1-neighbors of $\boldsymbol{x}$ are non-codewords $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n-1}$, and each $\boldsymbol{y}_i$ (including $\boldsymbol{y}_0$) is covered by a unique codeword in $\partial\mathfrak{B}_2(\boldsymbol{x})$. Conversely, each codeword in $\partial\mathfrak{B}_2(\boldsymbol{x})$ covers exactly two words among the $\boldsymbol{y}_i$'s. These $n/2$ codewords, along with $\boldsymbol{c}$, are (all) the $n/2+1$ codewords in $\mathfrak{B}_2(\boldsymbol{x})$.

Case 2: $\boldsymbol{y}_0 \neq \boldsymbol{c}$, namely, $\boldsymbol{y}_0$ is a midword, which is covered by two codewords $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \partial\mathfrak{B}_2(\boldsymbol{x})$. The 1-neighbors of $\boldsymbol{x}$ other than $\boldsymbol{c}$ and $\boldsymbol{y}_0$ are non-codewords $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{n-2}$, each covered by a unique codeword (in $\partial\mathfrak{B}_2(\boldsymbol{x})$). Conversely, each codeword in $\partial\mathfrak{B}_2(\boldsymbol{x})$ covers exactly two words among the $\boldsymbol{y}_i$'s (where $\boldsymbol{y}_0$ is covered by two codewords). We conclude that $\mathfrak{B}_2(\boldsymbol{x})$ contains exactly $n/2+1$ codewords: (i) the codeword $\boldsymbol{c}$, (ii) the codewords $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$, which cover both $\boldsymbol{y}_0$ and two other $\boldsymbol{y}_i$'s, and (iii) $n/2-2$ codewords that cover the $n-4$ remaining $\boldsymbol{y}_i$'s. □

The following bound is due to Fort and Hedlund [14] and will be used in our next proof.

Lemma 4 ([14]). *Let $\mathcal{X}$ be an (n, M) code whose codewords are all in $\mathcal{S}_3$ and, in addition, every word in $\mathcal{S}_2$ is covered by at least one codeword in $\mathcal{X}$. Then*

$$|\mathcal{X}| \geqslant \left\lceil \frac{n}{3} \left\lfloor \frac{n}{2} \right\rfloor \right\rceil .$$

Lemma 5. *Let $\mathcal{C}$ be an (n, M) NP1CC. For every codeword $\boldsymbol{c} \in \mathcal{C}$,*

$$|\mathfrak{B}_2(\boldsymbol{c}) \cap \mathcal{C}| \geqslant 2.$$

Proof. The result is immediate when $n = 2$, so we assume hereafter in the proof that $n = 2^r \geqslant 4$ and (by possibly translating the code) that $\boldsymbol{c} = \mathbf{0}$. Suppose to the contrary that $\mathfrak{B}_2(\mathbf{0}) \cap \mathcal{C} = \{\mathbf{0}\}$. Then $\mathcal{S}_1 \cap \mathcal{C} = \mathcal{S}_2 \cap \mathcal{C} = \emptyset$ and, so, all the words in $\mathcal{S}_2$ are covered (only) by codewords in $\mathcal{S}_3 \cap \mathcal{C}$. By Lemma 4 we then get that $|\mathcal{S}_3 \cap \mathcal{C}| \geqslant (n^2/2+1)/3$. Now, each codeword in $\mathcal{S}_3 \cap \mathcal{C}$ covers three words in $\mathcal{S}_2$ and, hence, the over-covering of $\mathcal{S}_2$ (with respect to $\mathcal{C}$) satisfies

$$\begin{aligned} &\sum_{\boldsymbol{y} \in \mathcal{S}_2} \left(|\mathfrak{B}_1(\boldsymbol{y}) \cap \mathcal{C}| - 1 \right) \\ &\quad = \; 3\,|\mathcal{S}_3 \cap \mathcal{C}| - |\mathcal{S}_2| \geqslant \frac{n^2}{2} + 1 - \binom{n}{2} = \frac{n}{2} + 1. \end{aligned}$$

On the other hand, by Corollary 2, $|\mathfrak{B}_1(\boldsymbol{y}) \cap \mathcal{C}| \in \{1, 2\}$ for every $\boldsymbol{y} \in \mathcal{S}_2$. Hence, there are at least $n/2+1$ words $\boldsymbol{y} \in \mathcal{S}_2$ for which $|\mathfrak{B}_1(\boldsymbol{y}) \cap \mathcal{C}| = 2$, which means that at least two of these words, say $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, must have a '1' at the same position. Let $\boldsymbol{x}$ be the word in $\mathcal{S}_1$ that has its (only) '1' at that position. Then $\mathfrak{B}_1(\boldsymbol{x})$ contains two words, $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, each covered by two codewords, thereby contradicting Lemma 1. We thus conclude that $|\mathfrak{B}_2(\mathbf{0}) \cap \mathcal{C}| \geqslant 2$. □

The next theorem presents the counterpart of Theorem 3 for radius-2 balls that are centered at codewords of an NP1CC.

Theorem 6. *Let $\mathcal{C}$ be an (n, M) NP1CC. For every codeword $\boldsymbol{c} \in \mathcal{C}$,*

$$|\mathfrak{B}_2(\boldsymbol{c}) \cap \mathcal{C}| = 2.$$

Proof. We consider the sum

$$\rho = \sum_{\boldsymbol{x} \in \mathbb{F}_2^n} |\mathfrak{B}_2(\boldsymbol{x}) \cap \mathcal{C}| .$$

Every codeword $\boldsymbol{c} \in \mathcal{C}$ is counted in this sum exactly $|\mathfrak{B}_2(\boldsymbol{c})| = |\mathfrak{B}_2(\mathbf{0})|$ times; so,

$$\begin{aligned} \rho = M \cdot |\mathfrak{B}_2(\mathbf{0})| &= M \cdot \left(\binom{n}{2} + n + 1 \right) \\ &= M \cdot \left(\frac{n^2}{2} + \frac{n}{2} + 1 \right). \end{aligned}$$

Next, we write $\rho = \sigma + \tau$, where

$$\sigma = \sum_{\boldsymbol{x} \in \mathbb{F}_2^n \setminus \mathcal{C}} |\mathfrak{B}_2(\boldsymbol{x}) \cap \mathcal{C}|$$

and

$$\tau = \sum_{\boldsymbol{c} \in \mathcal{C}} |\mathfrak{B}_2(\boldsymbol{c}) \cap \mathcal{C}| . \tag{6}$$

By Theorem 3 it follows that

$$\begin{aligned} \sigma = (2^n - M) \left(\frac{n}{2} + 1 \right) &= M \cdot (n-1) \left(\frac{n}{2} + 1 \right) \\ &= M \cdot \left(\frac{n^2}{2} + \frac{n}{2} - 1 \right) \end{aligned}$$

and, so,

$$\tau = \rho - \sigma = 2M.$$

Now, by Lemma 5, each of the M summands in (6) is at least 2; hence, each of them must in fact be equal to 2. □

For any codeword $\boldsymbol{c}$ in an NP1CC $\mathcal{C}$, the unique other codeword $\boldsymbol{c}'$ in $\mathfrak{B}_2(\boldsymbol{c})$ will be called the *partner* of $\boldsymbol{c}$. A pair of partners $\{\boldsymbol{c}, \boldsymbol{c}'\}$ in which $\boldsymbol{c}$ and $\boldsymbol{c}'$ are at distance 1 (respectively, 2) apart will be called a *Type* I (respectively, *Type* II) *pair*.

Corollary 7. *Let $\mathcal{C}$ be an $(n{=}2^r, M{=}2^{n-r})$ NP1CC. Then $\mathcal{C}$ can be partitioned uniquely into $M/2 = 2^{n-r-1}$ (unordered) pairs $\{\boldsymbol{c}, \boldsymbol{c}'\}$, where $\boldsymbol{c}$ and $\boldsymbol{c}'$ are partners.*

For a pair of partners $\{\boldsymbol{c}, \boldsymbol{c}'\}$, consider the "capsule" $\mathfrak{B}_1(\boldsymbol{c}) \cup \mathfrak{B}_1(\boldsymbol{c}')$. We can distinguish between two types of capsules, depending on whether the pair $\{\boldsymbol{c}, \boldsymbol{c}'\}$ is of Type I or of Type II. Interestingly, the two types of capsules have the same size, $2n$. The midwords are precisely the words that belong to the intersections $\mathfrak{B}_1(\boldsymbol{c}) \cap \mathfrak{B}_1(\boldsymbol{c}')$ when the pair is of Type II.

Theorem 8. *Let $\mathcal{C}$ be an $(n{=}2^r, M{=}2^{n-r})$ NP1CC. There are exactly $M = 2^{n-r}$ words in $\mathbb{F}_2^n$ that are covered by two codewords of $\mathcal{C}$ and no word is covered by more than two codewords.*

Proof. Each codeword of $\mathcal{C}$ covers $n+1$ words of $\mathbb{F}_2^n$ and, so,

$$\sum_{\boldsymbol{x} \in \mathbb{F}_2^n} |\mathfrak{B}_1(\boldsymbol{x}) \cap \mathcal{C}| = M(n+1) = |\mathbb{F}_2^n| + M.$$

The result follows from Corollary 2 and Theorem 6, which imply that $|\mathfrak{B}_1(\boldsymbol{x}) \cap \mathcal{C}| \in \{1, 2\}$ for every $\boldsymbol{x} \in \mathbb{F}_2^n$. □

The words in Theorem 8 that are covered by two codewords are (i) the midwords and (ii) the partners in Type I pairs.

We end this section by presenting sufficient conditions for a code to be an NP1CC.

Corollary 9. *Let $\mathcal{C}$ an (n, M) code where M is even, and suppose that $\mathcal{C}$ can be partitioned into $M/2$ unordered pairs $\{\boldsymbol{c}, \boldsymbol{c}'\}$ where $\mathsf{d}(\boldsymbol{c}, \boldsymbol{c}') \leqslant 2$. Suppose in addition that the respective $M/2$ capsules form a partition of $\mathbb{F}_2^n$. Then $\mathcal{C}$ is an NP1CC.*

Proof. The code $\mathcal{C}$ is 1-covering since every word in $\mathbb{F}_2^n$ is contained in some capsule. And since the size of each capsule is $2n$ we get equality in (5). ☐

Corollary 10. *Let $\mathcal{C}$ an $(n{=}2^r, M{=}2^{n-r})$ code where r is a positive integer. Then $\mathcal{C}$ is an NP1CC, if and only if $|\mathfrak{B}_2(\boldsymbol{c}) \cap \mathcal{C}| = 2$ for every codeword $\boldsymbol{c} \in \mathcal{C}$.*

Proof. Theorem 6 establishes the “only if” part, so we prove sufficiency. Let $\mathcal{C}$ an $(n{=}2^r, M{=}2^{n-r})$ code such that $|\mathfrak{B}_2(\boldsymbol{c}) \cap \mathcal{C}| = 2$ for every codeword $\boldsymbol{c} \in \mathcal{C}$. We can then partition $\mathcal{C}$ (uniquely) into $M/2$ unordered pairs $\{\boldsymbol{c}, \boldsymbol{c}'\}$ where $\mathsf{d}(\boldsymbol{c}, \boldsymbol{c}') \leqslant 2$. We show that the capsules that correspond to distinct pairs are disjoint.

Indeed, suppose that the capsules that correspond to the pairs $\{\boldsymbol{c}_1, \boldsymbol{c}_2\}$ and $\{\boldsymbol{c}_3, \boldsymbol{c}_4\}$ intersect, i.e., there exists a word $\boldsymbol{x} \in \mathbb{F}_2^n$ in the intersection

$$(\mathfrak{B}_1(\boldsymbol{c}_1) \cup \mathfrak{B}_1(\boldsymbol{c}_2)) \cap (\mathfrak{B}_1(\boldsymbol{c}_3) \cup \mathfrak{B}_1(\boldsymbol{c}_4)) .$$

This means that $\boldsymbol{x} \in \mathfrak{B}_1(\boldsymbol{c}_i) \cap \mathfrak{B}_1(\boldsymbol{c}_j)$ for some $i \in \{1, 2\}$ and $j \in \{3, 4\}$. By the triangle inequality,

$$\mathsf{d}(\boldsymbol{c}_i, \boldsymbol{c}_j) \leqslant \mathsf{d}(\boldsymbol{c}_i, \boldsymbol{x}) + \mathsf{d}(\boldsymbol{c}_j, \boldsymbol{x}) \leqslant 2,$$

which means that $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ are in the same capsule. Yet this is possible only if $\{\boldsymbol{c}_1, \boldsymbol{c}_2\} = \{\boldsymbol{c}_3, \boldsymbol{c}_4\}$.

Since the $M/2$ capsules are disjoint, the size of their union is $(M/2)(2n) = 2^n$. Hence, they form a partition of $\mathbb{F}_2^n$, and the result follows from Corollary 9. ☐

III. Elementary Constructions of NP1CCs

All the constructions of NP1CCs 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. Recall that a perfect code is a code that meets the bounds (1) and (2). We will consider only codes for which $R = 1$ in these equations; such codes have length $n = 2^r - 1$ and size $M = 2^{n-r}$. For each length n, there is an essentially unique linear perfect code—the Hamming code. A perfect code can be a zeroed perfect code or its non-zeroed translate. It is known that the number of nonequivalent (nonlinear) perfect codes grows (at least) doubly-exponentially with n [6, pp. 296–310], [12], [24], [28].

An *extended zeroed perfect code* is obtained from a zeroed perfect code $\mathcal{C}$ by adding an even parity to each codeword of $\mathcal{C}$. We distinguish between two types of non-zeroed translates of an extended zeroed perfect code of length 2^r: an *odd translate* of such a code contains only words with odd weight including exactly one word of weight 1, and an *even translate* contains only words of even weight including 2^{r-1} words of weight 2.

The following two lemmas are straightforward.

Lemma 11. *Puncturing an extended zeroed perfect code on any one of its coordinates yields a zeroed perfect code.*

Lemma 12. *Let $\mathcal{C}$ be an extended zeroed perfect code of length $n = 2^r$.*

(1) *Each word in $\mathbb{F}_2^n$ of odd weight has a unique 1-neighbor in $\mathcal{C}$.*
(2) *Each word in $\mathbb{F}_2^n$ of even weight has a unique 1-neighbor in any odd translate of $\mathcal{C}$.*
(3) *Each word in $\mathbb{F}_2^n$ of odd weight has a unique 1-neighbor in any even translate of $\mathcal{C}$.*

The simple construction in the next theorem yields NP1CCs for all the three types that were listed in Section I.

Theorem 13. *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be $(n{-}1, M/2)$ perfect codes, where $n = 2^r$ and $M = 2^{n-r}$. Then the code*

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

is an (n, M) NP1CC.

Proof. Since $|\mathcal{C}| = M = 2^{n-r}$, it suffices to show that $\mathsf{d}(\boldsymbol{x}, \mathcal{C}) \leqslant 1$ for every word $\boldsymbol{x} \in \mathbb{F}_2^n$. Write $\boldsymbol{x} = (\boldsymbol{y}, b)$ where $b \in \mathbb{F}_2$. Since $\mathcal{C}_1$ and $\mathcal{C}_2$ are perfect codes we have $\mathsf{d}(\boldsymbol{y}, \mathcal{C}_1) \leqslant 1$ and $\mathsf{d}(\boldsymbol{y}, \mathcal{C}_2) \leqslant 1$; hence, $\mathsf{d}(\boldsymbol{x}, \mathcal{C}) \leqslant 1$ regardless of b. ☐

Corollary 14. *Using the notation of Theorem 13, the code $\mathcal{C}$ therein is*

(1) *of Type $\mathbb{A}$, if $\mathcal{C}_1 = \mathcal{C}_2$,*
(2) *of Type $\mathbb{B}$, if $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$, and—*
(3) *of Type $\mathbb{C}$, if $\mathcal{C}_1$ and $\mathcal{C}_2$ are distinct and, in addition, $|\mathcal{C}_1 \cap \mathcal{C}_2| = k > 0$, where k is the number of Type* I *pairs in $\mathcal{C}$.*

Proof. The pairs of partners in $\mathcal{C}$ range over all pairs $\{(\boldsymbol{c}_1, 0), (\boldsymbol{c}_2, 1)\}$ where $\boldsymbol{c}_1 \in \mathcal{C}_1$, $\boldsymbol{c}_2 \in \mathcal{C}_2$, and $\mathsf{d}(\boldsymbol{c}_1, \boldsymbol{c}_2) \leqslant 1$, with Type I pairs corresponding to the case where $\boldsymbol{c}_1 = \boldsymbol{c}_2$ ($\in \mathcal{C}_1 \cap \mathcal{C}_2$). ☐

Corollary 14(3) raises an interesting question regarding (n, M) NP1CCs of Type $\mathbb{C}$: for which integers $k \in [M/2 - 1]$ do there exist NP1CCs with exactly k pairs of Type I (and $M/2 - k$ pairs of Type II)? The corollary implies that such NP1CCs can be constructed from two perfect codes that intersect on exactly k codewords. It was proved in [2] that when $r \geqslant 4$, for any even integer k in $\left[0 : 2^{2^r - 2r}\right]$ there exist two perfect codes of length $2^r - 1$ whose intersection has size k. See also [1], [12], [13], [17], Corollary 22 in Section IV, and the paragraph before Proposition 31 in Section VI.

Corollary 15. *There exist NP1CCs of Type $\mathbb{A}$, of Type $\mathbb{B}$, and of Type $\mathbb{C}$.*

The next theorem and corollary provide a complete characterization of NP1CCs of Type $\mathbb{A}$.

Theorem 16. *A code $\mathcal{C}$ is a zeroed $(n{=}2^r, M{=}2^{n-r})$ NP1CC of Type $\mathbb{A}$, if and only if it is the union of an extended zeroed perfect code of length $n = 2^r$ with an odd translate of an extended zeroed perfect code of the same length.*

Proof. Suppose that $\mathcal{C}$ is a zeroed $(n{=}2^r, M{=}2^{n-r})$ NP1CC of Type $\mathbb{A}$. Since $\mathcal{C}$ can be partitioned into Type I pairs, exactly half of the codewords have even weight. Moreover, since there are no two codewords in $\mathcal{C}$ at distance 2 apart, it follows that the sub-code that consists of the even-weight (respectively, odd-weight) codewords has minimum distance (at least) 4. Therefore, the even-weight codewords in $\mathcal{C}$ form an extended zeroed perfect code, and the odd-weight codewords form an odd translate of an extended zeroed perfect code.

Conversely, suppose that $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$, where $\mathcal{C}_1$ is an extended zeroed perfect code of length $n = 2^r$ and $\mathcal{C}_2$ is an odd translate of an extended zeroed perfect code of the same length. By Lemma 12(1), every word $\boldsymbol{x} \in \mathbb{F}_2^n$ of odd weight has a 1-neighbor in $\mathcal{C}_1$ (and that applies also when $\boldsymbol{x} \in \mathcal{C}_2$); similarly, by Lemma 12(2), every word $\boldsymbol{x} \in \mathbb{F}_2^n$ of even weight has a 1-neighbor in $\mathcal{C}_2$ (and that includes the case where $\boldsymbol{x} \in \mathcal{C}_1$). Moreover, $|\mathcal{C}| = 2^{n-r}$ and, hence, $\mathcal{C}$ is a zeroed NP1CC of Type $\mathbb{A}$. □

Corollary 17. *A code is a non-zeroed translate of an* $(n{=}2^r, M{=}2^{n-r})$ *NP1CC of Type* $\mathbb{A}$*, if and only if it is the union of an even translate of an extended zeroed perfect code of length* 2^r *with an odd translate of an extended zeroed perfect code of the same length.*

Other constructions in which an NP1CC of one type is obtained from an NP1CC of another type will be given in Section VI.

IV. Weight Distribution of NP1CCs

In this section, we characterize the weight distribution of NP1CCs. In particular, we show that zeroed NP1CCs can have one out of two weight distributions: one distribution is unique to NP1CCs of Type $\mathbb{A}$, and the other is unique to NP1CCs of Type $\mathbb{B}$ (zeroed NP1CCs of Type $\mathbb{C}$ can have any of these two distributions).

Our analysis will make use of some known properties of weight distributions, all of which can be found in Chapters 5 and 6 in [22]. For the ease of reference, we have summarized them in Section IV-A.

A. Definitions and background

Given an (n, M) code $\mathcal{C}$, the *weight distribution* of $\mathcal{C}$ is the integer vector $\boldsymbol{A} = \boldsymbol{A}_{\mathcal{C}} = (A_i)_{i\in[0:n]}$ with entries

$$A_i = |\mathcal{C} \cap \mathcal{S}_i| .$$

The respective *weight enumerator* is the bivariate homogeneous polynomial

$$\mathsf{A}(x, y) = \sum_{i\in[0:n]} A_i x^{n-i} y^i ,$$

or the univariate polynomial $\mathsf{A}(y) \triangleq \mathsf{A}(1, y)$. The *distance distribution* of an (n, M) code $\mathcal{C}$ is the rational vector $\boldsymbol{B} = \boldsymbol{B}_{\mathcal{C}} = (B_i)_{i\in[0:n]}$ whose entries are

$$B_i = \frac{1}{M} \left| \left\{ (\boldsymbol{c}, \boldsymbol{c}') \in \mathcal{C} \times \mathcal{C} \,:\, \mathsf{d}(\boldsymbol{c}, \boldsymbol{c}') = i \right\} \right| .$$

Thus,

$$\boldsymbol{B} = \frac{1}{M} \sum_{\boldsymbol{e}\in\mathcal{C}} \boldsymbol{A}_{\boldsymbol{e}+\mathcal{C}} . \tag{7}$$

The respective *distance enumerator* is the bivariate homogeneous polynomial

$$\mathsf{B}(x, y) = \sum_{i\in[0:n]} B_i x^{n-i} y^i ,$$

or the univariate polynomial $\mathsf{B}(y) \triangleq \mathsf{B}(1, y)$.

A zeroed code $\mathcal{C}$ is called *distance invariant* if $\boldsymbol{A}_{\boldsymbol{e}+\mathcal{C}} = \boldsymbol{A}_{\mathcal{C}}$ for every codeword $\boldsymbol{e} \in \mathcal{C}$. For such codes we have $\boldsymbol{B} = \boldsymbol{A}$. All linear codes are distance invariant.

Let $\boldsymbol{z} = (z_j)_{j\in[n]}$ be a vector of real indeterminates and define the ring

$$\mathfrak{R}_n = \mathbb{R}[\boldsymbol{z}] / \langle z_1^2 - 1, z_2^2 - 1, \ldots, z_n^2 - 1 \rangle .$$

Namely, the elements and arithmetic in $\mathfrak{R}_n$ are obtained from those in $\mathbb{R}[\boldsymbol{z}]$ by reducing modulo 2 the exponents of powers of the indeterminates (and so those powers can be seen as the elements 0 and 1 of $\mathbb{F}_2$). For $\boldsymbol{v} = (v_j)_{j\in[n]} \in \mathbb{F}_2^n$, we introduce the shorthand notation

$$\boldsymbol{z}^{\boldsymbol{v}} \triangleq \prod_{j\in[n]} z_j^{v_j} .$$

For each $\boldsymbol{u} = (u_j)_{j\in[n]} \in \mathbb{F}_2^n$, we define the *character* $\chi_{\boldsymbol{u}} : \mathfrak{R}_n \to \mathbb{R}$ which maps any

$$\mathsf{G} = \mathsf{G}(\boldsymbol{z}) = \sum_{\boldsymbol{v}\in\mathbb{F}_2^n} g_{\boldsymbol{v}} \boldsymbol{z}^{\boldsymbol{v}} \in \mathfrak{R}_n$$

to its value at $\boldsymbol{z} = ((-1)^{u_j})_{j\in[n]}$:

$$\chi_{\boldsymbol{u}}(\mathsf{G}(\boldsymbol{z})) = \sum_{\boldsymbol{v}\in\mathbb{F}_2^n} g_{\boldsymbol{v}} \cdot (-1)^{\langle \boldsymbol{u}, \boldsymbol{v} \rangle} ,$$

where $\langle \cdot, \cdot \rangle$ denotes dot product. Clearly, $\chi_{\boldsymbol{u}}$ is linear over $\mathbb{R}$ and multiplicative.

With each (n, M) code $\mathcal{C}$ we associate its *generating function* in $\mathfrak{R}_n$:

$$\mathsf{C}(\boldsymbol{z}) = \sum_{\boldsymbol{v}\in\mathcal{C}} \boldsymbol{z}^{\boldsymbol{v}} .$$

Given an (n, M) code $\mathcal{C}$, the *transform* of the weight distribution $\boldsymbol{A}_{\mathcal{C}}$ is the rational vector $\boldsymbol{A}' = \boldsymbol{A}'_{\mathcal{C}} = (A'_i)_{i\in[0:n]}$ with the entries

$$A'_i = \frac{1}{M} \sum_{\boldsymbol{u}\in\mathcal{S}_i} \chi_{\boldsymbol{u}}(\mathsf{C}(\boldsymbol{z})) . \tag{8}$$

In particular, $A'_0 \equiv 1$. The respective enumerator polynomial,

$$\mathsf{A}'(x, y) = \sum_{i\in[0:n]} A'_i x^{n-i} y^i ,$$

is related to $\mathsf{A}(x, y)$ by *MacWilliams' identities*:

$$\mathsf{A}'(x, y) = \frac{1}{M} \cdot \mathsf{A}(x + y, x - y) \tag{9}$$

and

$$\mathsf{A}(x, y) = \frac{M}{2^n} \cdot \mathsf{A}'(x + y, x - y) . \tag{10}$$

When $\mathcal{C}$ is linear, the transform $\boldsymbol{A}'$ is the weight distribution of the dual code, $\mathcal{C}^{\perp}$, of $\mathcal{C}$.

Example 2. Let $\mathcal{C}_1$ be the Type $\mathbb{A}$ linear NP1CC in Example 1. The dual code $\mathcal{C}_1^\perp$ is the simplex code padded with an extra zero coordinate; hence,

$$\mathsf{A}'_{\mathcal{C}_1}(x,y) = x^n + (n-1)\,x^{n/2}y^{n/2}.$$

The weight enumerator of $\mathcal{C}_1$ is therefore $\mathsf{A}_{\mathcal{C}_1}(y) = \mathsf{A}_1(y)$, where

$$\begin{aligned}\mathsf{A}_1(y) &\triangleq \frac{1}{n}(1+y)^n + \left(1-\frac{1}{n}\right)(1+y)^{n/2}(1-y)^{n/2} \\ &= \frac{1}{n}(1+y)^n + \left(1-\frac{1}{n}\right)(1-y^2)^{n/2}. \qquad (11)\end{aligned}$$

Let $\mathcal{C}_2$ be the Type $\mathbb{B}$ linear NP1CC in that example. The dual code $\mathcal{C}_2^\perp$ is the simplex code padded with a replica of one of the coordinates. Here

$$\mathsf{A}'_{\mathcal{C}_2}(x,y) = x^n + \left(\frac{n}{2}-1\right)x^{n/2}y^{n/2} + \frac{n}{2}\,x^{n/2-1}y^{n/2+1}$$

and, so, $\mathsf{A}_{\mathcal{C}_2}(y) = \mathsf{A}_2(y)$, where

$$\begin{aligned}\mathsf{A}_2(y) &\triangleq \frac{1}{n}(1+y)^n + \left(\frac{1}{2}-\frac{1}{n}\right)(1+y)^{n/2}(1-y)^{n/2} \\ &\quad + \frac{1}{2}(1+y)^{n/2-1}(1-y)^{n/2+1}. \qquad (12)\end{aligned}$$

□

The transform of the distance distribution $\boldsymbol{B}$ is the rational vector $\boldsymbol{B}' = (B'_i)_{i\in[0:n]}$ with the entries

$$B'_i = \frac{1}{M^2}\sum_{\boldsymbol{u}\in\mathcal{S}_i}\big(\chi_{\boldsymbol{u}}(\mathsf{C}(\boldsymbol{z}))\big)^2. \qquad (13)$$

The respective enumerator polynomial,

$$\mathsf{B}'(x,y) = \sum_{i\in[0:n]} B'_i x^{n-i}y^i,$$

is related to $\mathsf{B}(x,y)$ by MacWilliams' identities (9)–(10), with $\mathsf{A}(x,y)$ and $\mathsf{A}'(x,y)$ therein replaced by $\mathsf{B}(x,y)$ and $\mathsf{B}'(x,y)$. When a zeroed code $\mathcal{C}$ is distance invariant we have $\boldsymbol{B}' = \boldsymbol{A}'$.

By (13) it follows that

$$B'_i = 0 \iff \chi_{\boldsymbol{u}}(\mathsf{C}(\boldsymbol{z})) = 0 \text{ for all } \boldsymbol{u}\in\mathcal{S}_i. \qquad (14)$$

Hence, by (8),

$$\mathsf{Supp}(\boldsymbol{A}') \subseteq \mathsf{Supp}(\boldsymbol{B}'). \qquad (15)$$

The *external distance* of $\mathcal{C}$ is defined by

$$s' = |\mathsf{Supp}(\boldsymbol{B}')\setminus\{0\}| = \mathsf{w}(\boldsymbol{B}') - 1.$$

Theorem 18 ([22, Ch. 6, Thm. 20])**.** *Let $\mathcal{C}$ be an (n,M) code with external distance s'. Then for any $\boldsymbol{e}\in\mathbb{F}_2^n$, the entries of $\boldsymbol{A}_{\boldsymbol{e}+\mathcal{C}}$ are uniquely determined by n, M, $\mathsf{Supp}(\boldsymbol{B}')$, and the first s' entries of $\boldsymbol{A}_{\boldsymbol{e}+\mathcal{C}}$.*

It follows from (the proof of) this theorem that a code is distance invariant whenever its external distance does not exceed its minimum distance. Moreover, the external distance bounds from above the covering radius of the code.

B. Characterization of the weight distribution of NP1CCs

Our next theorem will be the main tool for characterizing the weight distribution of NP1CCs. Our proof will use the following notation. For $i\in[0:n]$, we let $\mathsf{Y}_i(\boldsymbol{z})$ be the ith *elementary symmetric function* in the entries of $\boldsymbol{z}$:

$$\mathsf{Y}_i(\boldsymbol{z}) = \sum_{\boldsymbol{v}\in\mathcal{S}_i}\boldsymbol{z}^{\boldsymbol{v}}.$$

It is known (see [22, p. 135]) that for any $\boldsymbol{u}\in\mathcal{S}_w$,

$$\chi_{\boldsymbol{u}}(\mathsf{Y}_i(\boldsymbol{z})) = P_i(w), \qquad (16)$$

where $P_i(\cdot)$ is the ith *Krawtchouk polynomial*:

$$P_i(w) \triangleq \sum_{j\in[0:i]}(-1)^j\binom{w}{j}\binom{n-w}{i-j}.$$

Theorem 19. *Let $\mathcal{C}$ be an (n,M) NP1CC and let $\boldsymbol{B}'$ be the transform of its distance distribution. Then*

$$\mathsf{Supp}(\boldsymbol{B}') \subseteq \{0, n/2, n/2+1\},$$

i.e., $s' \leqslant 2$.

Proof. Let $\mathsf{C}(\boldsymbol{z})$ be the generating function of $\mathcal{C}$ and consider the following multinomial (in $\mathfrak{R}_n$):

$$\mathsf{C}(\boldsymbol{z})\cdot\sum_{\boldsymbol{v}\in\mathcal{S}_1\cup\mathcal{S}_2}\boldsymbol{z}^{\boldsymbol{v}} = \mathsf{C}(\boldsymbol{z})\left(\mathsf{Y}_1(\boldsymbol{z})+\mathsf{Y}_2(\boldsymbol{z})\right).$$

For any word $\boldsymbol{x}\in\mathbb{F}_2^n$, the coefficient of $\boldsymbol{z}^{\boldsymbol{x}}$ in this multinomial equals the number of codewords at distance 1 or 2 from $\boldsymbol{x}$. By Theorems 3 and 6, this number is

$$\begin{cases} \dfrac{n}{2}+1 & \text{if } \boldsymbol{x} \text{ is a non-codeword,} \\ 1 & \text{if } \boldsymbol{x} \text{ is a codeword.}\end{cases}$$

Hence,

$$\begin{aligned}\mathsf{C}(\boldsymbol{z})\left(\frac{n}{2}+\mathsf{Y}_1(\boldsymbol{z})+\mathsf{Y}_2(\boldsymbol{z})\right) &= \left(\frac{n}{2}+1\right)\sum_{\boldsymbol{v}\in\mathbb{F}_2^n}\boldsymbol{z}^{\boldsymbol{v}} \\ &= \left(\frac{n}{2}+1\right)\prod_{j\in[n]}(1+z_j)^n\end{aligned}$$

and, so, for every $\boldsymbol{u}\in\mathbb{F}_2^n\setminus\{\boldsymbol{0}\}$,

$$\chi_{\boldsymbol{u}}\left(\mathsf{C}(\boldsymbol{z})\left(\frac{n}{2}+\mathsf{Y}_1(\boldsymbol{z})+\mathsf{Y}_2(\boldsymbol{z})\right)\right) = 0.$$

By (16) and the additivity and multiplicativity of $\chi_{\boldsymbol{u}}(\cdot)$ we get

$$\chi_{\boldsymbol{u}}(\mathsf{C}(\boldsymbol{z}))\cdot\beta(\mathsf{w}(\boldsymbol{u})) = 0, \qquad (17)$$

where $\beta(\cdot)$ is the following polynomial:

$$\begin{aligned}\beta(w) &= \frac{n}{2}+P_1(w)+P_2(w) \\ &= \frac{n}{2}+(n-2w)+\left(\binom{n}{2}-2nw+2w^2\right) \\ &= 2\left(w-\frac{n}{2}\right)\left(w-\frac{n}{2}-1\right).\end{aligned}$$

Let w be a nonzero element in $\mathsf{Supp}(\boldsymbol{B}')$, namely, $B'_w\neq 0$. By (14), there exists at least one word $\boldsymbol{u}\in\mathcal{S}_w$ such that $\chi_{\boldsymbol{u}}(\mathsf{C}(\boldsymbol{z}))\neq 0$. Hence, by (17), we conclude that

$$\beta(w) = 0$$

TABLE I
PARAMETERS OF THE FOUR POSSIBLE WEIGHT DISTRIBUTIONS OF NP1CCS.

Case	A_0	A_1	$A'_{n/2}$	$A'_{n/2+1}$	Types
$\boldsymbol{e} \in \mathcal{C}$ and $\lvert\mathfrak{B}_1(\boldsymbol{e}) \cap \mathcal{C}\rvert = 2$	1	1	$n-1$	0	$\mathbb{A}, \mathbb{C}$
$\boldsymbol{e} \in \mathcal{C}$ and $\lvert\mathfrak{B}_1(\boldsymbol{e}) \cap \mathcal{C}\rvert = 1$	1	0	$n/2-1$	$n/2$	$\mathbb{B}, \mathbb{C}$
$\boldsymbol{e} \notin \mathcal{C}$ and $\lvert\mathfrak{B}_1(\boldsymbol{e}) \cap \mathcal{C}\rvert = 2$	0	2	$n/2-1$	$-n/2$	$\mathbb{B}, \mathbb{C}$
$\boldsymbol{e} \notin \mathcal{C}$ and $\lvert\mathfrak{B}_1(\boldsymbol{e}) \cap \mathcal{C}\rvert = 1$	0	1	-1	0	$\mathbb{A}, \mathbb{B}, \mathbb{C}$

(see Lemma 19 in [22, Ch. 6]), i.e., $w \in \{n/2, n/2+1\}$. $\square$

Let $\mathcal{C}$ be an (n, M) NP1CC which, without any loss of generality, we assume to be zeroed, and let $\boldsymbol{e} + \mathcal{C}$ be any of its translates. By Theorem 19 we have $s' \leqslant 2$ and, so, by Theorem 18, the weight distribution, $\boldsymbol{A} = (A_i)_{i\in[0:n]}$, of $\boldsymbol{e}+\mathcal{C}$ is uniquely determined by its first two entries, namely, by the pair $(A_0\ A_1)$. And by Corollary 2 and Theorem 6, this pair can take (only) four values, as shown in the first three columns in Table I. In what follows, we compute the explicit dependence of the weight enumerator $\mathsf{A}(y)$ (and, hence, of the weight distribution $\boldsymbol{A}$) on $(A_0\ A_1)$. We do this by first determining the transform $\mathsf{A}'(x, y)$ using the first set of MacWilliams' identities (9); then, we use the second set (10) to obtain the complete weight enumerator $\mathsf{A}(x, y)$.

Substituting $(x, y) = (1, 1)$ in both sides of (9) and recalling that $A'_0 \equiv 1$ and (from (15) and Theorem 19) that $\mathsf{Supp}(\boldsymbol{A}') \subseteq \mathsf{Supp}(\boldsymbol{B}') \subseteq \{0, n/2, n/2+1\}$, we get

$$1 + A'_{n/2} + A'_{n/2+1} = n.$$

Next, differentiating both sides of (9) with respect to y and doing the same substitution yields

$$\frac{n}{2}\,A'_{n/2} + \left(\frac{n}{2}+1\right)A'_{n/2+1} = \frac{n}{2}(nA_0 - A_1).$$

Solving the last two equations for $A'_{n/2}$ and $A'_{n/2+1}$ in terms of $(A_0\ A_1)$ results in:

$$\begin{aligned} A'_{n/2} &= nA_0 - \frac{n}{2}(1-A_1) - 1 \\ A'_{n/2+1} &= \frac{n}{2}(1-A_1). \end{aligned}$$

The fourth and fifth columns in Table I present the solutions for $A'_{n/2}$ and $A'_{n/2+1}$ (and, thus, the complete characterization of the transform $\mathsf{A}'(x, y)$) for each of the four cases in the table. Knowing now all the nonzero coefficients in $\mathsf{A}'(x, y)$, we get from (10) the complete weight enumerator $\mathsf{A}(y)$, in terms of $(A_0\ A_1)$:

$$\begin{aligned} \mathsf{A}(y) &= \frac{1}{n}(1+y)^n \\ &\quad + \left(A_0 - \frac{1-A_1}{2} - \frac{1}{n}\right)(1+y)^{n/2}(1-y)^{n/2} \\ &\quad + \frac{1-A_1}{2}\cdot(1+y)^{n/2-1}(1-y)^{n/2+1}. \end{aligned}$$

Rearranging terms leads to the following result.

Theorem 20. *Let $\mathcal{C}$ be a zeroed (n, M) NP1CC and let $\boldsymbol{e}$ be a word in $\mathbb{F}_2^n$. Then the weight enumerator of $\boldsymbol{e}+\mathcal{C}$ is given by*

$$\begin{aligned} \mathsf{A}(y) &= \frac{1}{n}(1+y)^n \\ &+ \left(A_0 - \frac{1}{n} + \left(A_0 + A_1 - 1 - \frac{1}{n}\right)y\right)(1-y)(1-y^2)^{n/2-1}, \end{aligned} \tag{18}$$

where $(A_0\ \ A_1)$ is determined from $\mathcal{C}$ and $\boldsymbol{e}$ according to Table I.

We next present an explicit expression for the entries of the weight distribution $\boldsymbol{A} = (A_i)_{i\in[0:n]}$. For $i \in [0:n]$, let

$$\Delta_i \triangleq (-1)^{\lceil i/2 \rceil}\binom{n/2-1}{\lfloor i/2 \rfloor}$$

(where the binomial coefficient is assumed to be zero for invalid parameters); it can be verified that

$$(1-y)(1-y^2)^{n/2-1} = \sum_{i\in[0:n]} \Delta_i y^i.$$

By (18) it then follows that for every $i \in [0:n]$,

$$A_i = \frac{1}{n}\binom{n}{i} + \Big(A_0 - \frac{1}{n}\Big)\Delta_i + \Big(A_0 + A_1 - 1 - \frac{1}{n}\Big)\Delta_{i-1}.$$

When $(A_0\ A_1) = (1\ 1)$, Eq. (18) becomes $\mathsf{A}_1(y)$ in (11). Note that this case can occur only when $\mathcal{C}$ is either of Type $\mathbb{A}$ or of Type $\mathbb{C}$ (see the last column in Table I). Moreover, if $\mathcal{C}$ is of Type $\mathbb{A}$, then $\mathsf{A}_1(y)$ is the weight enumerator of $\boldsymbol{e}+\mathcal{C}$ for *every* codeword $\boldsymbol{e} \in \mathcal{C}$. Hence, Type $\mathbb{A}$ codes are distance invariant: in their case $\boldsymbol{B} = \boldsymbol{A}$ and $\boldsymbol{B}' = \boldsymbol{A}'$ and, consequently, their external distance is 1 (which is also their minimum distance).

When $(A_0\ A_1) = (1\ 0)$, Eq. (18) becomes $\mathsf{A}_2(y)$ in (12). This case can occur only when $\mathcal{C}$ is either of Type $\mathbb{B}$ or of Type $\mathbb{C}$. By a similar reasoning as before we conclude that Type $\mathbb{B}$ codes are distance invariant as well and their external distance, as well as their minimum distance, is 2 (except when $n = 2$, where the external distance is 1).

The case $(A_0\ A_1) = (0\ 2)$ also pertains to Type $\mathbb{B}$ and Type $\mathbb{C}$ codes, as it occurs when $\boldsymbol{e}$ is a midword. Eq. (18) is then similar to (12) except that the sign of the last term in (12) is flipped.

Finally, the case $(A_0\ A_1) = (0\ 1)$ corresponds to $\boldsymbol{e}$ being a non-codeword that is not a midword. This case can occur in all types, and the weight enumerator is

$$\frac{1}{n}\left((1+y)^n - (1-y^2)^{n/2}\right).$$

Type $\mathbb{C}$ codes cannot be distance invariant, since a fraction $B_1 \in (0,1)$ of the codewords have 1-neighbors while the other codewords do not. Still, by (7), we get a complete characterization of their distance enumerator:

$$\mathsf{B}(y) = B_1 \cdot \mathsf{A}_1(y) + (1-B_1)\cdot \mathsf{A}_2(y).$$

Corollary 21. *Let $\mathcal{C}$ be an (n,M) N1PCC where $n>2$. Then exactly half of the codewords in $\mathcal{C}$ have even weight.*

Proof. It follows from (18) that

$$\sum_{i \text{ even}} A_i - \sum_{i \text{ odd}} A_i = \mathsf{A}(-1) = 0.$$

□

Corollary 22. *Let $\mathcal{C}$ be an (n,M) N1PCC where $n>2$. Then the number, k, of Type* I *pairs in $\mathcal{C}$ is even (and so is the number, $M/2-k$, of Type* II *pairs). Moreover, exactly half of the Type* II *pairs consist of even-weight partners.*

Proof. Within each Type I pair, one (and only one) of the partners has even weight. Hence, in the subset $\mathcal{C}_{\mathrm{I}}$ of $\mathcal{C}$ formed by the union of all Type I pairs, exactly half of the codewords have even weight. By Corollary 21 it then follows that the same must hold in the subset $\mathcal{C}_{\mathrm{II}} = \mathcal{C} \setminus \mathcal{C}_{\mathrm{I}}$, which is formed by the union of all Type II pairs. Yet in each Type II pair, the parity of the partners must be the same; hence, there are as many Type II pairs with even-weight partners as such pairs with odd-weight partners. We conclude that $|\mathcal{C}_{\mathrm{II}}|$ is even and, therefore, so is $k = |\mathcal{C}_{\mathrm{I}}| = M/2 - |\mathcal{C}_{\mathrm{II}}|$. □

Remark 1. The weight distributions of Type $\mathbb{A}$ and Type $\mathbb{B}$ NP1CCs were shown in [4] using a different technique. Another approach for computing the weight distributions of the three types was suggested by the reviewer and is based on equitable partitions and quotient matrices [20], [23]. This method completely solves the weight and distance distributions for Type $\mathbb{A}$ and Type $\mathbb{B}$. For Type $\mathbb{C}$, we need to consider the same technique for the extended code and analyze its punctured code after the solution of the weight distribution. However this method does not recover any information on the distance distribution. □

V. Balanced Nearly Perfect Covering Codes

Some NP1CCs have additional special properties. One example of such a property pertains to (n,M) codes of Type $\mathbb{A}$ where for any coordinate there is at least one codeword that disagrees with its partner on that coordinate (this property will turn out to be useful in Section VI). In this section, we construct such codes. Moreover, in the codes that we present, the number of codewords that disagree with their respective partners on any given coordinate equals M/n (i.e., it is the same for all coordinates). Such a code will be called a *balanced NP1CC*.

We start by introducing some notation. A *self-dual* sequence is a binary cyclic sequence that is equal to its complement. If there is no periodicity in the sequence (which will the case henceforth), then it can be written as $[X\ \overline{X}]$, where $\overline{X}$ is the binary complement of X. Letting n be the length of X, the *orbit* of $[X\ \overline{X}]$, denoted $\mathrm{Orb}(X)$, is the set of all words in $\mathbb{F}_2^n$ that are obtained by reading any n (cyclically-)consecutive symbols in $[X\ \overline{X}]$. Equivalently, defining the *complemented cyclic shift* operator $\varphi_n : \mathbb{F}_2^n \to \mathbb{F}_2^n$ by

$$\varphi_n(x_1\, x_2\, x_3\, \ldots\, x_n) = x_2\, x_3 \ldots x_n\, \overline{x}_1,$$

the elements of $\mathrm{Orb}(X)$ are all the distinct words of the form $\varphi_n^i(X)$, for some integer $i \geqslant 0$ (where $\varphi_n^0(X) \equiv X$) [16, p. 171–173]. It is easy to see that $|\mathrm{Orb}(X)|$ divides $2n$ (in our case there will be equality) and that $\varphi_n(\cdot)$ is distance preserving: for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}_2^n$,

$$\mathsf{d}(\varphi_n(\boldsymbol{x}), \varphi_n(\boldsymbol{y})) = \mathsf{d}(\boldsymbol{x}, \boldsymbol{y}).$$

For each $r \geqslant 3$, we define the code $\mathcal{C}_r$ of length $n = 2^r$ by

$$\mathcal{C}_r = \bigcup_{[X\ \overline{X}] \in \mathcal{G}_r} \mathrm{Orb}(X), \tag{19}$$

where $\mathcal{G}_r$ is a set of 2^{n-2r-1} self-dual sequences of length $2n$ that will be defined recursively. We show by induction on r that $\mathcal{C}_r$ is an NP1CC of Type $\mathbb{A}$, and the balancing property will then follow from the closure of $\mathcal{C}_r$ under $\varphi_n(\cdot)$.

For $r=3$, we take

$$\mathcal{G}_3 = \{[00011010\ 11100101], [00011011\ 11100100]\}, \tag{20}$$

resulting in an $(8,32)$ code $\mathcal{C}_3$. Since the sequences in $\mathcal{G}_3$ differ only in two positions (the eighth and the 16th), each codeword in $\mathcal{C}_3$ has a 1-neighbor in $\mathcal{C}_3$; moreover, a simple inspection reveals that no two codewords in $\mathcal{C}_3$ are at distance 2 apart. Hence, by Corollary 10, the code $\mathcal{C}_3$ is a (non-zeroed) NP1CC of Type $\mathbb{A}$.

Example 3. Three more pairs of sequences can be used instead of (20) (the pairs can be obtained from one another by decimation):

$$\begin{array}{c}\{[01001110\ 10110001], [01001111\ 10110000]\},\\ \{[01110110\ 10001001], [01110111\ 10001000]\},\\ \{[00100011\ 11011100], [00100010\ 11011101]\}.\end{array}$$

□

Now, given some $r \geqslant 3$, write $n = 2^r$ and suppose (by induction) that $\mathcal{G}_r$ is a set of 2^{n-2r-1} self-dual sequences of length $2n$ for which the construction (19) yields an $(n{=}2^r, M{=}2^{n-r})$ NP1CC of Type $\mathbb{A}$. Also, let $\mathcal{E}_r$ be the set of all 2^{n-2} even-weight words in $\mathbb{F}_2^n$ that start with a '0'. We define $\mathcal{G}_{r+1}$ to be the following set of self-dual sequences of length $4n$:

$$\mathcal{G}_{r+1} \triangleq \left\{[U\ \ U{+}X\ \ \overline{U}\ \ \overline{U}{+}X] \,:\, U \in \mathcal{E}_r \text{ and } [X\ \overline{X}] \in \mathcal{G}_r\right\}. \tag{21}$$

This recursive construction of $\mathcal{G}_{r+1}$ from $\mathcal{G}_r$ is very similar to the constructions presented in [8], [10], [11] and is reminiscent of the well-known $(\boldsymbol{u}, \boldsymbol{u}{+}\boldsymbol{v})$ code construction method [22, p. 76].

Example 4. Applying the recursion, we can compute the set $\mathcal{G}_4$ (of size 128) from the set $\mathcal{G}_3$ in (20). Some of the sequences of $\mathcal{G}_4$ are shown in Table II. □

Proposition 23. *Under the induction hypothesis on $\mathcal{C}_r$ and $|\mathcal{G}_r|$, the code $\mathcal{C}_{r+1}$ that is constructed from $\mathcal{G}_{r+1}$ by (19) is a balanced $(2n, 2^{2n-r-1})$ NP1CC, and $|\mathcal{G}_{r+1}| = 2^{2n-2r-3}$.*

Proof. We first show that every word in $\mathbb{F}_2^{2n}$ has a 1-neighbor in $\mathcal{C}_{r+1}$. Let $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2) \in \mathbb{F}_2^{2n}$ where $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{F}_2^n$. Since (by induction) $\mathcal{C}_r$ is an NP1CC of Type $\mathbb{A}$, there exists $\boldsymbol{c} \in \mathcal{C}_r$ such that $\boldsymbol{y}_1 + \boldsymbol{y}_2 = \boldsymbol{c} + \boldsymbol{e}$, for some $\boldsymbol{e} \in \mathcal{S}_1$ (and that holds also when $\boldsymbol{y}_1 + \boldsymbol{y}_2 \in \mathcal{C}_r$). Let $[X\ \overline{X}]$ be the self-dual sequence in $\mathcal{G}_r$ whose orbit contains $\boldsymbol{c}$; then $X = \varphi_n^i(\boldsymbol{c})$, for some $i \in [0 : 2n{-}1]$. Since $\varphi_{2n}(\cdot)$ is distance preserving and $\mathcal{C}_{r+1}$ is closed under this operator, it suffices to show that the word $\boldsymbol{u} = \varphi_{2n}^i(\boldsymbol{y})$ has a 1-neighbor in $\mathcal{C}_{r+1}$.

Writing $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2)$ where $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{F}_2^n$, we have

$$\boldsymbol{u}_1 + \boldsymbol{u}_2 = \varphi_n^i(\boldsymbol{y}_1 + \boldsymbol{y}_2) = \varphi_n^i(\boldsymbol{c} + \boldsymbol{e}) = X + \boldsymbol{e}', \qquad (22)$$

for some $\boldsymbol{e}' \in \mathcal{S}_1$ (which is a rotation of $\boldsymbol{e}$ by i positions). Now, if $\boldsymbol{u}_1$ has even weight, then

$$(\boldsymbol{u}_1, \boldsymbol{u}_1{+}X) \overset{(22)}{=} (\boldsymbol{u}_1, \boldsymbol{u}_2{+}\boldsymbol{e}')$$

is a codeword of $\mathcal{C}_{r+1}$ at distance 1 from $\boldsymbol{u}$. Otherwise, $\boldsymbol{u}_1' \triangleq \boldsymbol{u}_1 + \boldsymbol{e}'$ has even weight, in which case

$$(\boldsymbol{u}_1', \boldsymbol{u}_1'{+}X) \overset{(22)}{=} (\boldsymbol{u}_1', \boldsymbol{u}_2)$$

is a codeword of $\mathcal{C}_{r+1}$ at distance 1 from $\boldsymbol{u}$. In summary, every word $\boldsymbol{y} \in \mathbb{F}_2^{2n}$ has a 1-neighbor in $\mathcal{C}_{r+1}$ (and that applies also when $\boldsymbol{y} \in \mathcal{C}_{r+1}$).

Turning now to the size of $\mathcal{C}_{r+1}$, we have

$$|\mathcal{C}_{r+1}| \overset{(19)}{\leqslant} 4n \cdot |\mathcal{G}_{r+1}| \overset{(21)}{\leqslant} 4n \cdot |\mathcal{E}_r| \cdot |\mathcal{G}_r| \leqslant 2^{2n-r-1},$$

namely, $\mathcal{C}_{r+1}$ is an NP1CC of Type $\mathbb{A}$ and the inequalities in the last equation all hold with equality; in particular, $|\mathcal{G}_{r+1}| = 2^{2n-2r-3}$. Finally, the balancing property follows from the closure of $\mathcal{C}_{r+1}$ under $\varphi_{2n}(\cdot)$. □

The construction $\mathcal{C}_r$ was defined and analyzed for another purpose in [5], where it was mentioned (without proof) that it is 1-covering. The construction will also work with self-dual sequences on other 1-covering of $\mathbb{F}_2^n$ if every word in $\mathbb{F}_2^n$ (including codewords) has a 1-neighbor.

VI. Extended NP1CCs and their properties

In this section, we study the properties of extended NP1CCs.

Given an $(n{=}2^r, M{=}2^{n-r})$ NP1CC $\mathcal{C}$, the respective extended NP1CC (in short, *ENP1CC*) is the $(n{+}1, M)$ code $\mathcal{C}^*$ that is obtained by adding to each codeword of $\mathcal{C}$ an even parity as an $(n{+}1)$st coordinate. The next property follows from Theorem 6.

Proposition 24. *In an ENP1CC $\mathcal{C}^*$, for each codeword $\boldsymbol{c} \in \mathcal{C}^*$ there is a unique codeword $\boldsymbol{c}' \in \mathcal{C}^*$ such that $\mathsf{d}(\boldsymbol{c}, \boldsymbol{c}') = 2$ (and $\mathsf{d}(\boldsymbol{c}, \boldsymbol{c}'') \geqslant 4$ for any other codeword $\boldsymbol{c}'' \in \mathcal{C}^* \setminus \{\boldsymbol{c}, \boldsymbol{c}'\}$).*

Similarly to NP1CCs, two codewords in an $(n{+}1, M)$ ENP1CC $\mathcal{C}^*$ that are at distance 2 apart will be called *partners*, and the $M/2$ pairs of partners form a partition of $\mathcal{C}^*$.

Corollary 10 and Proposition 24 imply the following consequence (compare with Lemma 11).

Corollary 25. *Puncturing an ENP1CC on any one of its coordinates yields an NP1CC.*

Thus, ENP1CCs induce a partition of the set of NP1CCs into equivalence classes, where two NP1CCs are said to be equivalent if one can be obtained from the other by extension followed by puncturing. Each equivalence class can contain NP1CCs of several types, and we are interested in identifying which of the three types of NP1CCs can belong to the same equivalence class.

A necessary and sufficient condition that a puncturing of an ENP1CC will be of a certain type can be inferred as an immediate observation from the definitions of Type $\mathbb{A}$, Type $\mathbb{B}$, and Type $\mathbb{C}$.

Lemma 26. *Let $\mathcal{C}^*$ is an ENP1CC. The type of an NP1CC obtained by puncturing of $\mathcal{C}^*$ on a given coordinate j is determined as follows:*

1. *it is of Type $\mathbb{A}$, if and only if all codewords in $\mathcal{C}^*$ disagree with their respective partners on the jth coordinate,*
2. *it is of Type $\mathbb{B}$, if and only if all codewords in $\mathcal{C}^*$ agree with their partners on that coordinate, and—*
3. *it is of Type $\mathbb{C}$ otherwise (i.e., there is at least one codeword in $\mathcal{C}^*$ that agrees with its partner on the jth coordinate and at least one codeword that doesn't).*

Proposition 24 and Lemma 26 imply the following negative result.

Proposition 27. *There are no ENP1CCs whose punctured codes are all of Type $\mathbb{A}$, or all of Type $\mathbb{B}$.*

Proposition 28. *Let $\mathcal{C}$ be an (n, M) NP1CC obtained from the union of an extended zeroed perfect code $\mathcal{C}_1$ and an odd translate $\mathcal{C}_2$ of $\mathcal{C}_1$. The punctured codes of $\mathcal{C}^*$ are either of Type $\mathbb{A}$ or of Type $\mathbb{B}$ (with at least one code of each type).*

Proof. Noting that $\mathcal{C}_2 = \boldsymbol{e} + \mathcal{C}_1$ where $\boldsymbol{e} \in \mathcal{S}_1$, let $\mathsf{Supp}(\boldsymbol{e}) = \{j\}$. All codewords in $\mathcal{C}$ disagree with their respective partners on the jth coordinate (and only on that coordinate). Therefore, all codewords in $\mathcal{C}^*$ disagree with their partners on that coordinate and on the $(n{+}1)$st coordinate (and agree on the remaining $n-1$ coordinates). Thus, by Lemma 26(1), puncturing on one of these two coordinates yields an NP1CC of Type $\mathbb{A}$, while, by Lemma 26(2), puncturing on any of the remaining $n-1$ coordinates yields an NP1CC of Type $\mathbb{B}$. □

It is easy to verify by Lemma 26 that Proposition 28 characterizes all the ENP1CCs whose punctured codes are of Type $\mathbb{A}$ or of Type $\mathbb{B}$ (with at least one code of each type).

Proposition 29. *Let $\mathcal{C}$ be an NP1CC of Type $\mathbb{A}$ in which for each coordinate there is at least one codeword that disagrees with its partner on that coordinate. The punctured codes of $\mathcal{C}^*$ are either of Type $\mathbb{A}$ or of Type $\mathbb{C}$ (with at least one code of each type).*

TABLE II
ELEMENTS OF THE SET $\mathcal{G}_4$, SORTED ACCORDING TO THE STANDARD LEXICOGRAPHIC ORDERING ON $\mathcal{E}_3 \times \mathcal{G}_3$.

[00000000 00011010 11111111 11100101],	[00000000 00011011 11111111 11100100],
[00000011 00011001 11111100 11100110],	[00000011 00011000 11111100 11100111],
[00000101 00011111 11111010 11100000],	[00000101 00011110 11111010 11100001],
[00000110 00011100 11111001 11100011],	[00000110 00011101 11111001 11100010],
[00001001 00010011 11110110 11101100],	[00001001 00010010 11110110 11101101],
[00001010 00010000 11110101 11101111],	[00001010 00010001 11110101 11101110],
[00001100 00010110 11110011 11101001],	[00001100 00010111 11110011 11101000],
[00001111 00010101 11110000 11101010],	[00001111 00010100 11110000 11101011],
⋮	
[01111000 01100010 10000111 10011101],	[01111000 01100011 10000111 10011100],
[01111011 01100001 10000100 10011110],	[01111011 01100000 10000100 10011111],
[01111101 01100111 10000010 10011000],	[01111101 01100110 10000010 10011001],
[01111110 01100100 10000001 10011011],	[01111110 01100101 10000001 10011010].

Proof. Clearly, $\mathcal{C}$ is one of the punctured codes (of Type $\mathbb{A}$) and, so, by Lemma 26(1), every codeword in $\mathcal{C}^*$ disagrees with its partner on the $(n+1)$st coordinate. Since partners in each pair in $\mathcal{C}^*$ are at distance 2 apart, there can be at most one additional coordinate on which all codewords in $\mathcal{C}^*$ disagree with their partners (in which case puncturing on that coordinate yields an NP1CC of Type $\mathbb{A}$). For each of the remaining coordinates there exists some codeword that agrees with its partner on that coordinate (and, by the condition of the proposition, also some codeword that doesn't). Hence, by Lemma 26(3), puncturing on any of these coordinates yields an NP1CC of Type $\mathbb{C}$. □

We note that a balanced NP1CC is an NP1CC of Type $\mathbb{A}$ which satisfies the requirements of Proposition 29. It is easy to verify by Lemma 26 that Proposition 29 characterizes all the ENP1CCs whose punctured codes are of Type $\mathbb{A}$ or of Type $\mathbb{C}$.

Proposition 30. *Let $\mathcal{C}^*$ be an ENP1CC whose punctured codes range over all three types (with at least one code of each type). There is exactly one coordinate on which all codewords in $\mathcal{C}^*$ disagree with their respective partners, and at least one coordinate on which all agree with their partners.*

Proof. By Lemma 26(1), there is a least one coordinate on which all codewords in $\mathcal{C}^*$ disagree with their partners. Now, if there were two such coordinates, then all codewords would have to agree with their partners on the remaining coordinates, thereby contradicting Lemma 26(3). Finally, by Lemma 26(2), there is a least one coordinate on which all codewords agree with their partners. □

We next construct an ENP1CC that satisfies the conditions of Proposition 30, based on an idea presented in [12]. By [12], there exist two zeroed perfect codes of length $n-1 = 2^r - 1$ which differ only in $2^{n/2-1}$ codewords and only on one coordinate, say the first coordinate. Let $\mathcal{C}_1$ be the extended code of the first code and $\mathcal{C}_2$ be an odd translate of the extended code of the second (where the extended code and its translate differ only on the last coordinate).

Proposition 31. *Given the above notation, the ENP1CC $\mathcal{C}^*$ obtained by extending the code $\mathcal{C} \triangleq \mathcal{C}_1 \cup \mathcal{C}_2$ is an ENP1CC whose punctured codes range over all three types.*

Proof. There is one coordinate—the $(n+1)$st—on which all the codewords in $\mathcal{C}^*$ disagree with their respective partners; two coordinates—the first and the nth—on which some (but not all) the codewords agree with their partners; and $n-2$ coordinates on which there is full agreement. The result follows from Lemma 26. □

Propositions 27–29 and 31 raise the question whether there exists an ENP1CC with no punctured codes of Type $\mathbb{A}$.

We end this section by a characterization of the weight distribution of a zeroed ENP1CC. This distribution turns out to be unique and independent of the type of the NP1CC that was extended (this also implies that ENP1CCs are distance invariant).

Theorem 32. *Let $\mathcal{C}^*$ be a zeroed $(n+1=2^r+1, M=2^{n-r})$ ENP1CC. Its weight enumerator is given by*

$$\begin{aligned}\mathsf{A}^*(y) &= \frac{1}{2n}\big((1+y)^{n+1} + (1-y)^{n+1}\big) \\ &\quad + \left(1 - \frac{1}{n}\right)(1-y^2)^{n/2}.\end{aligned}$$

Proof. Let $\mathcal{C}$ be the zeroed (n, M) NP1CC that was extended and let $\mathsf{A}(y) = \sum_{i\in[0:n]}^{n} A_i y^i$ be its weight enumerator. It is easy to see that the weight distribution of $\mathcal{C}^*$ is given by

$$A_0^* = 1, \quad A_{n+1}^* = 0,$$

and, for $i \in [n]$:

$$A_i^* = \begin{cases} A_i + A_{i-1} & \text{if } i \text{ is even} \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$\begin{aligned}\mathsf{A}^*(y) &= \sum_{i\in[0:n+1]} A_i^* y^i \\ &= \frac{1}{2}\big(\mathsf{A}(y) + \mathsf{A}(-y) + y(\mathsf{A}(y) - \mathsf{A}(-y))\big) \\ &= \frac{1}{2}\big((1+y)\mathsf{A}(y) + (1-y)\mathsf{A}(-y)\big). \qquad (23)\end{aligned}$$

Substituting either (11) or (12) into (23) yields the result. □

VII. Conclusion and Future Work

In this work, we studied the structure of NP1CCs. We showed that NP1CCs can be classified into three types, depending on the distribution of the Hamming distances between neighboring codewords. Properties of these types were presented, including a characterization of their weight and distance distributions. We derived several methods for constructing codes for each type, including a method that is based on puncturing ENP1CCs. The latter method motivated us to study which types of NP1CCs can be obtained from any given ENP1CC.

Our exposition leads to several interesting open problems.

1) Is it true that there exist two $(n-1, M/2)$ perfect codes that intersect on exactly k codewords, if and only if there exists an (n, M) NP1CC with exactly k pairs of Type I? What is the maximum possible number of pairs of Type I in an NP1CC of Type $\mathbb{C}$?
2) Let $\mathcal{X}$ and $\mathcal{Y}$ be two distinct nonempty sets of pairwise disjoint capsules in $\mathbb{F}_2^n$ such that
$$\bigcup_{V \in \mathcal{X}} V = \bigcup_{V \in \mathcal{Y}} V.$$
What is the minimum size of $\mathcal{X}$ and $\mathcal{Y}$?
3) Is there an NP1CC of Type $\mathbb{B}$ such that for each pair of coordinates there exists some codeword that disagrees with its partner on both of these coordinates?
4) We proved that there exist NP1CCs of Type $\mathbb{A}$ that are balanced. Is there a counterpart of this result for codes of Type $\mathbb{B}$? One possible definition for balanced (n, M) NP1CCs of Type $\mathbb{B}$ is that for any two (cyclically-)adjacent coordinates there are exactly M/n codewords that disagree with their respective partners only on those coordinates.
5) Is there an ENP1CC such that none of its punctured codes is of Type $\mathbb{A}$?

Acknowledgment

Tuvi Etzion would like to thank Denis Krotov for the insightful and inspirational discussions on topics that relate to this work.

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," *Des. Codes 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. Inf. Theory,* 29 (1983), 342–345.
[4] A. Boruchovsky and T. Etzion, "Nearly perfect covering codes," https://arxiv.org/abs/2405.00258v1 (2024).
[5] Y. M. Chee, T. Etzion, H. Ta, and V. K. Vu, "On de Bruijn covering sequences and arrays," *Proc. IEEE Int'l Symp. on Information Theory,* Athens, Greece (2024), 1343–1348.
[6] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering Codes*, North-Holland, Amsterdam, 1997.
[7] T. Etzion, *Perfect Codes and Related Structures*, World Scientific, Singapore, 2022.
[8] T. Etzion, *Sequences and the de Bruijn Graph: Properties, Constructions, and Applications,* Elsevier, London, 2024.
[9] T. Etzion and G. Greenberg, "Constructions for perfect mixed codes and other covering codes," *IEEE Trans. Inf. Theory,* 39 (1993), 209–214.
[10] T. Etzion and A. Lempel, "Construction of de Bruijn sequences of minimal complexity," *IEEE Trans. Inf. Theory,* 30 (1984), 705–709.
[11] T. Etzion and K. G. Paterson, "Near optimal single-track Gray codes," *IEEE Trans. Inf. Theory,* 42 (1996), 779–789.
[12] T. Etzion and A. Vardy, "Perfect binary codes: Constructions, properties, and enumeration," *IEEE Trans. Inf. Theory,* 40 (1994), 754–763.
[13] T. Etzion and A. Vardy, "On perfect codes and tilings: Problems and solutions," *SIAM J. 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.,* 3 (1972), 65–88.
[16] S. W. Golomb, *Shift Register Sequences,* Revised Edition, Aegean Park Press, Laguna Hills, California, 1982.
[17] O. Heden, F. I. Solov'eva, and I. Yu. Mogilnykh, "Intersection of perfect binary codes," *Proc. 2010 IEEE Region 8 Int'l Conf. on Computational Technologies in Electrical and Electronics Engineering (SIBIRCON),* Irkutsk, Russia (2010), 52–54.
[18] S. M. Johnson, "A new upper bound for error-correcting codes," *IRE Trans. Inf. Theory,* 8 (1962), 203–207.
[19] W. M. Kantor, "On the inequivalence of generalized Preparata codes," *IEEE Trans. Inf. Theory,* 29 (1983), 345–348.
[20] D. Krotov, "On weight distributions of perfect colorings and completely regular codes," *Des. Codes Cryptogr.,* 61 (2011), 315–329.
[21] K. Lindström, "All nearly perfect codes are known," *Inform. Control,* 35 (1977), 40–47.
[22] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes,* North-Holland, Amsterdam, 1977.
[23] W. J. Martin, *Completely regular subsets,* Ph.D. thesis, University of Waterloo, 1992. http://users.wpi.edu/ martin/RESEARCH/THESIS.
[24] K. T. Phelps, "A general product construction for error-correcting codes," *SIAM J. Alg. Disc. Meth.,* 5 (1984), 224–228.
[25] F. P. Preparata, "A class of optimum nonlinear double-error-correcting codes," *Inform. Control,* 13 (1968), 378–400.
[26] R. Struik, "An improvement of the Van Wee bound for binary linear covering codes," *IEEE Trans. Inf. Theory,* 40 (1994), 1280–1284.
[27] G. J. M. van Wee, "Improved sphere bounds on the covering radius of codes," *IEEE Trans. Inf. Theory,* 34 (1988), 237–245.
[28] J. L. Vasil'ev, "On nongroup close-packed codes," *Probl. Kibernet.,* 8 (1962), 337–339. Also in: *Algebraic Coding Theory: History and Development,* I. F. Blake (Ed.), Dowden, Hutchinson and Ross, 1973, pp. 351–357.

Avital Boruchovsky (Student Member, IEEE) received the B.Sc. degree in mathematics and computer science and the M.Sc. degree in computer science from the Technion–Israel Institute of Technology, Haifa, Israel, in 2021 and 2024, respectively. He is currently pursuing the Ph.D. degree with the Computer Science Department at the Technion. His research interests include algorithms, discrete mathematics, coding theory, and DNA storage.

Tuvi Etzion (Life Fellow, IEEE) was born in Tel Aviv, Israel, in 1956. He received the B.A., M.Sc., and D.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, Israel, in 1980, 1982, and 1984, respectively.

From 1984 he held positions in the Department of Computer Science at the Technion, where currently he is an Emeritus Professor. From 1985 to 1987 he was Visiting Research Fellow with the Department of Electrical Engineering—Systems, the University of Southern California, Los Angeles. During the summers of 1990 and 1991 he was visiting Bellcore, Morristown, New Jersey. From 1994 to 1996, he was a Visiting Research Fellow with the Computer Science Department, Royal Holloway University of London, Egham, England. He also had several visits to the Coordinated Science Laboratory at University of Illinois in Urbana–Champaign from 1995 to 1998, two visits to HP Bristol during the summers of 1996 and 2000, a few visits to the Department of Electrical Engineering, University of California at San Diego from 2000 to 2022, several visits to the Mathematics Department at Royal Holloway University of London, from 2007 to 2022, a few visits to the School of Physical and Mathematical Science (SPMS), Nanyang Technological University and to the department of Industrial Systems Engineering and Management the National University of Singapore, Singapore, from 2016 to 2019, and from 2022 to 2025, where he is currently a Visiting Research Professor. He also had a few visits to Jiaotong University, Beijing, from 2017 to 2019 and in 2024. His research interests include applications of discrete mathematics to problems in computer science and information theory, coding theory, coding for storage, and combinatorial designs.

Dr. Etzion was an Associate Editor for Coding Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2006 till 2009. From 2004 to 2009, he was an Editor for the *Journal of Combinatorial Designs*. From 2011 till 2024 he was an Editor for Designs, Codes, and Cryptography, and an Editor for Advances of Mathematics in Communications from 2013 till 2023. He has been an Editor-in-Chief for Journal of Combinatorial Theory, Series A, since 2021.

Ron M. Roth (Life Fellow, IEEE) received the B.Sc. degree in computer engineering, the M.Sc. in electrical engineering, and the D.Sc. in computer science from Technion—Israel Institute of Technology, Haifa, Israel, in 1980, 1984, and 1988, respectively. Since 1988 he has been with the Computer Science Department at Technion, where he now holds the General Yaakov Dori Chair in Engineering. During the academic years 1989–91 he was a Visiting Scientist at IBM Research Division, Almaden Research Center, San Jose, California. During 1996–97, 2004–05, and 2011–2012 he was on sabbatical leave at Hewlett–Packard Laboratories, Palo Alto, California, and in 2022 at Hewlett Packard Enterprise, Milpitas, California. He is the author of the book *Introduction to Coding Theory*, published by Cambridge University Press in 2006. Dr. Roth was an associate editor for coding theory in IEEE TRANSACTIONS ON INFORMATION THEORY from 1998 till 2001, and he is now serving as an associate editor in *SIAM Journal on Discrete Mathematics*. His research interests include coding theory, information theory, and their application to storage, computation, and the theory of complexity.