Spectral-Null Codes and Null Spaces of Hadamard Submatrices*

Ron M. Roth
Computer Science Department
Technion — Israel Institute of Technology
Haifa 32000, Israel.
e-mail: ronny@cs.technion.ac.il

Abstract

Codes $\mathcal{C}(m,r)$ of length 2^m over $\{1,-1\}$ are defined as null spaces of certain submatrices of Hadamard matrices. It is shown that the codewords of $\mathcal{C}(m,r)$ all have an rth order spectral null at zero frequency. Establishing the connection between $\mathcal{C}(m,r)$ and the parity-check matrix of Reed-Muller codes, the minimum distance of $\mathcal{C}(m,r)$ is obtained along with upper bounds on the redundancy of $\mathcal{C}(m,r)$. An efficient algorithm is presented for encoding unconstrained binary sequences into $\mathcal{C}(m,2)$.

Keywords: Hadamard matrices; Reed-Muller codes; Spectral-null codes.

*This work was presented in part at the French-Israeli Workshop on Algebraic Coding, Paris, July 1993, and was supported in part by the United-States — Israel Binational Science Foundation.

1 Introduction

Let Φ denote the alphabet $\{1, -1\}$. A word $\mathbf{x} = [x_0\, x_1 \ldots x_{n-1}]$ over Φ is said to have an rth order spectral null at zero frequency if $\sum_{j=0}^{n-1} j^i\, x_j = 0$ for $i = 0, 1, \ldots, r-1$, where operations are taken over the integers. Codes consisting of words with prescribed spectral-null properties have appeared in the literature in several applications, e.g., in reducing the notch width of the spectrum of DC-free words at zero frequency [18],[19] or in enhancing the error-correction capability of codes used in partial-response channels [6],[12].

Let $\mathcal{C}$ be a nonempty subset of Φ^n. We refer to n as the length of $\mathcal{C}$ and to $n - \log_2 |\mathcal{C}|$ as the redundancy of $\mathcal{C}$, denoted red($\mathcal{C}$). The minimum distance dist($\mathcal{C}$) of $\mathcal{C}$ is the minimum Hamming distance between any two distinct words in $\mathcal{C}$.

Let $H(n, r)$ denote the integer matrix

$$H(n,r) \;=\; \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 0 & 1 & 2 & \ldots & n-1 \\ 0^2 & 1^2 & 2^2 & \ldots & (n-1)^2 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0^{r-1} & 1^{r-1} & 2^{r-1} & \ldots & (n-1)^{r-1} \end{bmatrix}$$

and let $\mathcal{S}_{\mathbb{R}}(n, r)$ denote the $(n-r)$-dimensional space over the real field $\mathbb{R}$ consisting of all words $\mathbf{x} \in \mathbb{R}^n$ satisfying $H(n, r)\, \mathbf{x}' = \mathbf{0}$. The set $\mathcal{S}(n, r) = \mathcal{S}_{\Phi}(n, r)$ is defined by $\mathcal{S}_{\mathbb{R}}(n, r) \cap \Phi^n$. That is, $\mathcal{S}(n, r)$ is the set of all words in Φ^n with an rth order spectral null at zero frequency.

The case $r = 0$ corresponds to unconstrained words and, therefore, $\mathcal{S}(n, 0) = \Phi^n$ with redundancy zero and minimum distance 1. The set $\mathcal{S}(n, 1)$ consists of all balanced (or DC-free) words, with redundancy $\frac{1}{2} \log_2 n + O(1)$ and minimum distance 2 [13],[18]. Using enumerative encoding [18], it is fairly easy to encode an arbitrary binary sequence of length $k = n - \frac{1}{2} \log_2 n - O(1)$, regarded as a k-bit representation of an integer ℓ, into a word of $\mathcal{S}(n, 1)$ indexed by ℓ according to the standard lexicographic order on $\mathcal{S}(n, 1)$. In [13], Knuth described a simpler encoding method into a subset of $\mathcal{S}(n, 1)$ with redundancy which is approximately twice the redundancy of $\mathcal{S}(n, 1)$. See also [1],[2],[10]. Subsets of $\mathcal{S}(n, 1)$ with certain error-correcting capability have been described in [3],[4],[5],[7],[8],[20], among others.

The problem of analyzing and synthesizing codes $\mathcal{C} \subseteq \mathcal{S}(n,r)$ for values of r greater than 1 has also been dealt with in a number of papers recently, e.g., [12],[15],[17],[19]. Some of the constructions, as in [12] and [15], are based on approaching the set $\mathcal{S}(n,r)$, when n goes to infinity, by a sequence of sets of words that are generated by finite labeled directed graphs. It also follows from [12] and [19] that the minimum distance of $\mathcal{S}(n,r)$ is at least $2r$. Efficient encoding algorithms of unconstrained binary sequences into $\mathcal{S}(n,r)$ are known for small values of r [13],[17]. For general values of r, however, there is still no practical encoding algorithm into $\mathcal{S}(n,r)$ with fairly small redundancy. The case of combining such constructions with prescribed error-correcting capability, and the design of efficient encoders and decoders for such codes, seem to be even more difficult problems that are still unsolved.

In this work, we aim at finding relatively large subsets of $\mathcal{S}(n,r)$ that have minimum distance which is much larger than $2r$. To this end, we define null spaces $\mathcal{C}(m,r)$ of certain submatrices of Hadamard matrices (Section 2). In Section 3, we show that $\mathcal{C}(m,r)$ is a subset of $\mathcal{S}(2^m,r)$. Using the connection between $\mathcal{C}(m,r)$ and the parity-check matrix of Reed-Muller codes, we then show that the minimum distance of $\mathcal{C}(m,r)$ is 2^r (Section 4). Then, in Section 5, we obtain upper bounds on the redundancy of $\mathcal{C}(m,r)$. Finally, in Section 6, we describe an efficient encoding procedure of unconstrained binary sequences into $\mathcal{C}(m,2)$. Obtaining an encoding procedure into $\mathcal{C}(m,r)$ for general values of r remains still an open problem.

2 Null spaces of Hadamard submatrices

Let F be the alphabet $\{0,1\}$, regarded as a subset of the integers. Denote by $\mathcal{H} = \mathcal{H}(m)$ the $2^m \times 2^m$ Sylvester-type Hadamard matrix whose rows and columns are indexed by elements of F^m and whose entries are given by

$$(\mathcal{H})_{\mathbf{u},\mathbf{v}} \;=\; (-1)^{\mathbf{u}\cdot\mathbf{v}}\;, \quad \mathbf{u},\mathbf{v} \in F^m\;.$$

For example, letting "+" stand for 1 and "−" stand for −1, we have, $\mathcal{H}(0) = [\,+\,]$,

$$\mathcal{H}(1) \;=\; \begin{bmatrix} + & + \\ + & - \end{bmatrix}, \quad \text{and} \quad \mathcal{H}(2) \;=\; \begin{bmatrix} + & + & + & + \\ + & - & + & - \\ + & + & - & - \\ + & - & - & + \end{bmatrix}.$$

For a vector $\mathbf{u} \in F^m$, we denote by $\mathrm{wt}(\mathbf{u})$ the Hamming weight of $\mathbf{u}$. The set of all vectors $\mathbf{u} \in F^m$ with $\mathrm{wt}(\mathbf{u}) < r$ will be denoted by $\mathcal{B}(m,r)$. Clearly, $|\mathcal{B}(m,r)| = V(m,r) = \sum_{i=0}^{r-1} \binom{m}{i}$.

Let $\mathcal{H}(m,r)$ be the $V(m,r) \times 2^m$ submatrix of $\mathcal{H}(m)$ consisting of all rows of $\mathcal{H}(m)$ indexed by $\mathbf{u} \in \mathcal{B}(m,r)$. For example,

$$\mathcal{H}(3,3) \;=\; \begin{bmatrix} + & + & + & + & + & + & + & + \\ + & - & + & - & + & - & + & - \\ + & + & - & - & + & + & - & - \\ + & + & + & + & - & - & - & - \\ + & - & - & + & + & - & - & + \\ + & - & + & - & - & + & - & + \\ + & + & - & - & - & - & + & + \end{bmatrix},$$

where the rows here are indexed in $\mathcal{H}(m)$ by $\mathbf{u} = 000, 001, 010, 100, 011, 101, 110$. The matrix $\mathcal{H}(3,2)$ is obtained by taking the first four rows in $\mathcal{H}(3,3)$.

We now define the set $\mathcal{C}_{\mathbb{R}}(m,r)$ as the right null space of $\mathcal{H}(m,r)$ in $\mathbb{R}^{2^m}$ i.e.,

$$\mathcal{C}_{\mathbb{R}}(m,r) \;=\; \left\{ \mathbf{x} \in \mathbb{R}^{2^m} \;\mid\; \mathcal{H}(m,r)\,\mathbf{x}' = \mathbf{0} \right\}, \quad 0 \le r \le m\,.$$

Hereafter we interchange integer indices of entries of vectors (such as $\mathbf{x}$) with their binary representations.

The code $\mathcal{C}(m,r) = \mathcal{C}_{\Phi}(m,r)$ is defined as the restriction of $\mathcal{C}_{\mathbb{R}}(m,r)$ to Φ^{2^m}, namely,

$$\mathcal{C}(m,r) \;=\; \mathcal{C}_{\mathbb{R}}(m,r) \;\cap\; \Phi^{2^m}\,.$$

Example 1. The code $\mathcal{C}(m,0)$ corresponds to an 'empty' matrix $\mathcal{H}(m,0)$ and, hence, $\mathcal{C}(m,0) = \Phi^{2^m} = \mathcal{S}(2^m,0)$. Since $\mathcal{H}(m,1)$ is the all-one vector $\mathbf{1}$, we have $\mathcal{C}(m,1) = \mathcal{S}(2^m,1)$, namely, $\mathcal{C}(m,1)$ is the set of all balanced words in Φ^{2^m}. •

Let $\mathcal{G}(m,r)$ denote the $V(m,m-r+1) \times 2^m$ submatrix of $\mathcal{H}(m)$ consisting of all rows of $\mathcal{H}(m)$ indexed by $\mathbf{u} \in F^m - \mathcal{B}(m,r)$. Since the rows of $\mathcal{H}(m)$ are orthogonal, the rows of $\mathcal{G}(m,r)$ form a basis of $\mathcal{C}_{\mathbb{R}}(m,r)$. Therefore,

$$\mathcal{C}(m,r) \;=\; \left\{ \mathbf{x} \in \Phi^{2^m} \;\mid\; \mathbf{x} \;=\; \mathbf{y}\,\mathcal{G}(m,r) \;\text{ for some } \mathbf{y} \in \mathbb{R}^{V(m,m-r+1)} \right\}.$$

Note that, in particular, the rows of $\mathcal{G}(m,r)$ are words of $\mathcal{C}(m,r)$.

Example 2. For $r = m$, the matrix $\mathcal{G}(m,r)$ contains one row, namely, the last row of $\mathcal{H}(m)$. This row vector is also known as a *Morse sequence* [9] and will be denoted hereafter by $\mu(m)$. The $\mathbf{v}$th entry in $\mu(m)$ is given by $= (-1)^{\mathrm{wt}(\mathbf{v})}$ for every $\mathbf{v} \in F^m$. Therefore, the code $\mathcal{C}(m,m)$ contains two words, $\mu(m)$ and $-\mu(m)$, and, as such, it has redundancy $2^m - 1$. The minimum distance of $\mathcal{C}(m,m)$ is 2^m. •

An equivalent definition for $\mathcal{C}_{\mathbb{R}}(m,r)$ and $\mathcal{C}(m,r)$ can be obtained by using the parity-check matrix of Reed-Muller (RM) codes, as we now show. This alternative definition will turn out to be useful while analyzing the spectral properties and the minimum distance of $\mathcal{C}(m,r)$.

Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in F^m. We write $\mathbf{u} \leq \mathbf{v}$ if the inequality holds component-by-component. Similarly, we write $\mathbf{u} < \mathbf{v}$ if $\mathbf{u} \leq \mathbf{v}$ and $\mathbf{u} \neq \mathbf{v}$. Define the $2^m \times 2^m$ matrix $H_{\mathrm{RM}} = H_{\mathrm{RM}}(m)$ by

$$(H_{\mathrm{RM}})_{\mathbf{u},\mathbf{v}} \;=\; \begin{cases} 1 & \text{if } \mathbf{u} \leq \mathbf{v} \\ 0 & \text{otherwise} \end{cases}, \quad \mathbf{u}, \mathbf{v} \in F^m \,.$$

For example, $H_{\mathrm{RM}}(0) = [\,1\,]$,

$$H_{\mathrm{RM}}(1) \;=\; \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad H_{\mathrm{RM}}(2) \;=\; \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The parity-check matrix of the binary $(m-r)$th order Reed-Muller (RM) code of length 2^m is defined as the $V(m,r) \times 2^m$ submatrix of $H_{\mathrm{RM}}(m)$ consisting of all rows of $H_{\mathrm{RM}}(m)$ indexed

by $\mathbf{u} \in \mathcal{B}(m,r)$. For example,

$$H_{\mathrm{RM}}(3,3) \;=\; \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} ,$$

where the rows here are indexed in $H_{\mathrm{RM}}(m)$ by $\mathbf{u} = 000$, 001, 010, 100, 011, 101, 110. The matrix $H_{\mathrm{RM}}(3,2)$ is obtained by taking the first four rows of $H_{\mathrm{RM}}(3,3)$. Note that we regard the matrices $H_{\mathrm{RM}}(m,r)$ as *integer* matrices with entries in F, rather than their common definition as parity-check matrices over $GF(2)$ [14, Ch. 13].

For integer vectors $\mathbf{x} = [x_0\, x_1\, \ldots\, x_{\ell-1}]$ and $\mathbf{y} = [y_0\, y_1\, \ldots\, y_{\ell-1}]$, we denote by $\mathbf{x} \oplus \mathbf{y}$ the (unique) vector in F^ℓ whose sth entry is congruent to $x_s + y_s$ modulo 2 for every $s = 0, 1, \ldots, \ell-1$. Also, we denote by $\mathbf{x} * \mathbf{y}$ the integer vector whose sth entry is given by $x_s y_s$.

It is easy to verify that the rows of $\mathcal{H}$ that are indexed by $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} \oplus \mathbf{v}$ are related by

$$(\mathcal{H})_{\mathbf{u}\oplus\mathbf{v}} \;=\; (\mathcal{H})_{\mathbf{u}} \;*\; (\mathcal{H})_{\mathbf{v}} \,.$$

If, in addition, we have $\mathbf{u} * \mathbf{v} = \mathbf{0}$, then a similar relation holds also for the rows of H_{RM}:

$$(H_{\mathrm{RM}})_{\mathbf{u}\oplus\mathbf{v}} \;=\; (H_{\mathrm{RM}})_{\mathbf{u}} \;*\; (H_{\mathrm{RM}})_{\mathbf{v}} \,.$$

The following lemma establishes a connection between $\mathcal{H}$ and H_{RM}.

Lemma 1. *$\mathcal{H} = {H_{\mathrm{RM}}}'\, D\, H_{\mathrm{RM}}$, where D is the $2^m \times 2^m$ diagonal matrix $(D)_{\mathbf{u},\mathbf{u}} = (-2)^{\mathrm{wt}(\mathbf{u})}$, $\mathbf{u} \in F^m$.*

Proof. We show by induction on $\mathrm{wt}(\mathbf{u})$ that

$$(\mathcal{H})_{\mathbf{u}} \;=\; \sum_{\mathbf{v}\le\mathbf{u}} (-2)^{\mathrm{wt}(\mathbf{v})}\, (H_{\mathrm{RM}})_{\mathbf{v}} \,.$$

For $\mathbf{u} = \mathbf{0}$ we have $(\mathcal{H})_{\mathbf{0}} = (H_{\mathrm{RM}})_{\mathbf{0}}$ and for every unit vector $\mathbf{e} \in F^m$ we have $(\mathcal{H})_{\mathbf{e}} = (H_{\mathrm{RM}})_{\mathbf{0}} - 2\,(H_{\mathrm{RM}})_{\mathbf{e}}$.

Assume now that $\mathrm{wt}(\mathbf{u}) > 1$ and write $\mathbf{u} = \mathbf{e} \oplus \mathbf{z}$ for some unit vector $\mathbf{e} \leq \mathbf{u}$. Since $\mathrm{wt}(\mathbf{z}) = \mathrm{wt}(\mathbf{u}) - 1$, we can apply the induction hypothesis to $(\mathcal{H})_{\mathbf{e}}$ and $(\mathcal{H})_{\mathbf{z}}$ to obtain

$$\begin{aligned}
(\mathcal{H})_{\mathbf{u}} \;=\; (\mathcal{H})_{\mathbf{e}\oplus\mathbf{z}} \;&=\; (\mathcal{H})_{\mathbf{e}} * (\mathcal{H})_{\mathbf{z}} \\
&=\; \Big((H_{\mathrm{RM}})_{\mathbf{0}} - 2\,(H_{\mathrm{RM}})_{\mathbf{e}}\Big) * \Big(\sum_{\mathbf{v}\leq\mathbf{z}} (-2)^{\mathrm{wt}(\mathbf{v})}\,(H_{\mathrm{RM}})_{\mathbf{v}}\Big) \\
&=\; \sum_{\mathbf{v}\leq\mathbf{z}} (-2)^{\mathrm{wt}(\mathbf{v})}\,(H_{\mathrm{RM}})_{\mathbf{v}} \;-\; 2\sum_{\mathbf{v}\leq\mathbf{z}} (-2)^{\mathrm{wt}(\mathbf{v})}\,\Big((H_{\mathrm{RM}})_{\mathbf{e}} * (H_{\mathrm{RM}})_{\mathbf{v}}\Big) \\
&=\; \sum_{\mathbf{v}\leq\mathbf{z}} (-2)^{\mathrm{wt}(\mathbf{v})}\,(H_{\mathrm{RM}})_{\mathbf{v}} \;+\; \sum_{\mathbf{v}\leq\mathbf{z}} (-2)^{\mathrm{wt}(\mathbf{e}\oplus\mathbf{v})}\,(H_{\mathrm{RM}})_{\mathbf{e}\oplus\mathbf{v}} \\
&=\; \sum_{\mathbf{v}\leq\mathbf{u}} (-2)^{\mathrm{wt}(\mathbf{v})}\,(H_{\mathrm{RM}})_{\mathbf{v}}\;,
\end{aligned}$$

as claimed. □

Proposition 1. $\mathcal{C}_{\mathbb{R}}(m,r) = \Big\{\, \mathbf{x} \in \mathbb{R}^{2^m} \mid H_{\mathrm{RM}}(m,r)\,\mathbf{x}' = \mathbf{0} \,\Big\}$.

Proof. We show that the rows of $\mathcal{H}(m,r)$ and $H_{\mathrm{RM}}(m,r)$ span the same real vector space. Indeed, since H_{RM} is upper-triangular, then, by Lemma 1, each row $(\mathcal{H})_{\mathbf{u}}$ is a linear combination of rows $(H_{\mathrm{RM}})_{\mathbf{v}}$ for $\mathbf{v} \leq \mathbf{u}$. Conversely, the inverse of H_{RM} is also upper-triangular (in fact, $({H_{\mathrm{RM}}}^{-1})_{\mathbf{u},\mathbf{v}} = (-1)^{\mathrm{wt}(\mathbf{u})+\mathrm{wt}(\mathbf{v})}\,(H_{\mathrm{RM}})_{\mathbf{u},\mathbf{v}}$). Writing $H_{\mathrm{RM}} = D^{-1}({H_{\mathrm{RM}}}')^{-1}\mathcal{H}$, each row $(H_{\mathrm{RM}})_{\mathbf{u}}$ is a linear combination of rows $(\mathcal{H})_{\mathbf{v}}$ for $\mathbf{v} \leq \mathbf{u}$. □

Let $G_{\mathrm{RM}}(m,r)$ be the $V(m,m{-}r{+}1) \times 2^m$ matrix whose rows are given by

$$(G_{\mathrm{RM}}(m,r))_{\mathbf{u}} \;=\; (H_{\mathrm{RM}}(m,m{-}r{+}1))_{\mathbf{u}} \;*\; \mu(m)\;, \quad \mathbf{u} \in \mathcal{B}(m,m{-}r{+}1)\;,$$

where $\mu(m)$ is the Morse sequence of length 2^m (see Example 2). It is easy to verify that $H_{\mathrm{RM}}(m,r)\,G_{\mathrm{RM}}(m,r)' = 0$ and, therefore, the rows of $G_{\mathrm{RM}}(m,r)$ form a basis of $\mathcal{C}_{\mathbb{R}}(m,r)$. In other words, the codes $\mathcal{C}(m,r)$ and $\mathcal{C}(m,m{-}r{+}1)$ are dual codes, up to a fixed sign-flipping of components of the words in one of the codes according to a Morse sequence.

3 Spectral properties of $\mathcal{C}(m,r)$

The next theorem, following the lemma, exhibits the spectral properties of $\mathcal{C}_{\mathbb{R}}(m,r)$ (and of $\mathcal{C}(m,r)$).

Lemma 2. *Let $\ell \le m$ be a positive integer. Then, there exist integers $\alpha_{\mathbf{u}}$, $\mathbf{u} \in \mathcal{B}(m,\ell)$, such that for every $\mathbf{v} = [v_0\, v_1\, \ldots\, v_{\ell-1}] \in F^m$,*

$$\Big(\sum_{s=0}^{m-1} v_s\, 2^s\Big)^{\ell-1} \;=\; \sum_{\mathbf{u}\in\mathcal{B}(m,\ell)} \alpha_{\mathbf{u}}\, (H_{\mathrm{RM}})_{\mathbf{u},\mathbf{v}}\ . \tag{1}$$

Proof. Expanding the left-hand side of (1) we have, for every $\mathbf{v} \in F^m$,

$$\Big(\sum_{s=0}^{m-1} v_s\, 2^s\Big)^{\ell-1} \;=\; \sum_{U} \alpha_U \prod_{s\in U} v_s\ ,$$

where U ranges over all subsets of $\{0,1,\ldots,m-1\}$ of size smaller than ℓ and the α_U's are integers that do not depend on $\mathbf{v}$. Note that since $\mathbf{v}$ is restricted to F^m, the degree of each v_s in every term of the expansion can be assumed to be 1 at most. Let $\mathbf{u} = [u_0\, u_1\, \ldots\, u_{m-1}]$ be the characteristic vector of U, that is, $u_s = 1$ if $s \in U$ and $u_s = 0$ otherwise. The lemma now follows by the easily-verified equality $\prod_{s\in U} v_s = (H_{\mathrm{RM}})_{\mathbf{u},\mathbf{v}}$. □

Theorem 1. $\mathcal{C}_{\mathbb{R}}(m,r) \subseteq \mathcal{S}_{\mathbb{R}}(2^m,r)$.

Proof. Let $\mathbf{x} = [x_0\, x_1\, \ldots\, x_{2^m-1}] = [\,x_{\mathbf{v}}\,]_{\mathbf{v}\in F^m}$ be a word in $\mathcal{C}_{\mathbb{R}}(m,r)$. We compute the first r entries of $H(2^m,r)\,\mathbf{x}'$. By Lemma 2 we have, for every $1 \le \ell \le r$,

$$\begin{aligned}
(H(2^m,r)\,\mathbf{x}')_{\ell-1} \;=\; \sum_{j=0}^{2^m-1} j^{\ell-1}\, x_j \;&=\; \sum_{\mathbf{v}=[v_0 v_1 \ldots v_{m-1}]\in F^m} \Big(\sum_{s=0}^{m-1} v_s\, 2^s\Big)^{\ell-1} x_{\mathbf{v}} \\
&=\; \sum_{\mathbf{v}\in F^m} \sum_{\mathbf{u}\in\mathcal{B}(m,\ell)} \alpha_{\mathbf{u}}\, (H_{\mathrm{RM}})_{\mathbf{u},\mathbf{v}}\, x_{\mathbf{v}} \\
&=\; \sum_{\mathbf{u}\in\mathcal{B}(m,\ell)} \alpha_{\mathbf{u}}\, (H_{\mathrm{RM}}\,\mathbf{x}')_{\mathbf{u}}\ .
\end{aligned}$$

However, $(H_{\mathrm{RM}}\,\mathbf{x}')_{\mathbf{u}} = 0$ for every $\mathbf{u} \in \mathcal{B}(m,r)$. Hence, $H(2^m,r)\,\mathbf{x}' = \mathbf{0}$. □

Remark 1. By Theorem 1 we have $\mathcal{C}(m,r) \subseteq \mathcal{S}(2^m,r)$, with equality for $r = 0, 1$. It is also known that $|\mathcal{S}(2^m,m)| = 2$ for $m \leq 5$ [18, p. 243]. Hence, $\mathcal{C}(m,m) = \mathcal{S}(2^m,m) = \{\mu(m), -\mu(m)\}$ for $m \leq 5$. However, for $m = 6$, the set $\mathcal{S}(64,6)$ contains words other than a Morse sequence, e.g., the word $++---++-+--+-++++--+--+--+--++$, concatenated with its reflection. •

4 Distance properties of $\mathcal{C}(m,r)$

Next we obtain the minimum distance of $\mathcal{C}(m,r)$.

Lemma 3. [14, p. 374].

$$H_{\text{RM}}(m,r) \;=\; \begin{bmatrix} H_{\text{RM}}(m{-}1,r) & H_{\text{RM}}(m{-}1,r) \\ 0 & H_{\text{RM}}(m{-}1,r{-}1) \end{bmatrix} .$$

Theorem 2. *The minimum distance of* $\mathcal{C}(m,r)$ *is* 2^r.

Proof. The proof is very similar to that used for RM codes. We show by induction on m that the nonzero words of $\mathcal{C}_{\mathbb{R}}(m,r)$ all have Hamming weight at least 2^r. This will imply a lower bound of 2^r on $\text{dist}(\mathcal{C}(m,r))$.

The claim is trivial for $m = 0$. As for $m > 0$, let $\mathbf{x}$ be a nonzero word in $\mathcal{C}_{\mathbb{R}}(m,r)$ and write $\mathbf{x} = [\,\mathbf{y}\;\;\mathbf{z}\,]$, where $\mathbf{y}$ and $\mathbf{z}$ are words of length 2^{m-1}. By Lemma 3 we have $H_{\text{RM}}(m{-}1,r{-}1)\,\mathbf{z}' = \mathbf{0}$ and $H_{\text{RM}}(m{-}1,r)\,(\mathbf{y}+\mathbf{z})' = \mathbf{0}$. Therefore, $\mathbf{z} \in \mathcal{C}_{\mathbb{R}}(m{-}1,r{-}1)$ and $\mathbf{y}+\mathbf{z} \in \mathcal{C}_{\mathbb{R}}(m{-}1,r)$. Now, if $\mathbf{y} \neq -\mathbf{z}$, then, by the induction hypothesis, the Hamming weight of $\mathbf{y}+\mathbf{z}$, and therefore that of $\mathbf{x}$, must be at least 2^r. On the other hand, if $\mathbf{y} = -\mathbf{z}$, then $\mathbf{z} \neq \mathbf{0}$ and, so, the Hamming weight of $\mathbf{z}$ is at least 2^{r-1}. Hence, the Hamming weight of $\mathbf{x}$, being in this case twice that of $\mathbf{z}$, must be at least 2^r.

To show that the 2^r lower bound is tight, note that if $\mathbf{y}$ and $\mathbf{z}$ are two words in $\mathcal{C}(m{-}1,r{-}1)$ at distance 2^{r-1}, then $[\,-\mathbf{y}\;\;\;\mathbf{y}\,]$ and $[\,-\mathbf{z}\;\;\;\mathbf{z}\,]$ are two words in $\mathcal{C}(m,r)$ at distance 2^r. The tightness now follows by induction on r, starting with $\mathcal{C}(m,0) = \Phi^{2^m}$. □

Example 3. The matrix $\mathcal{G}(m, m-1)$ has $m+1$ rows and, hence, $|\mathcal{C}(m, m-1)| \geq 2m+2$. We now show by induction on m that this bound is tight. This is obviously true for $m = 1$. Assume now that $m > 1$ and let $[\,\mathbf{y}\ \mathbf{z}\,]$ be a word in $\mathcal{C}(m, m-1)$, where $\mathbf{y}$ and $\mathbf{z}$ are words of length 2^{m-1}. By Lemma 3 we have $\mathbf{z} \in \mathcal{C}(m-1, m-2)$ and $\mathbf{y} + \mathbf{z} \in \mathcal{C}_{\mathbb{R}}(m-1, m-1)$. However, since $\mathcal{C}_{\mathbb{R}}(m-1, m-1)$ consists of scalar multiplies of $\mu(m-1)$, all the nonzero words in $\mathcal{C}_{\mathbb{R}}(m-1, m-1)$ do not contain any zero components. Therefore, we must have either $\mathbf{y} = -\mathbf{z}$ or $\mathbf{y} = \mathbf{z}$. Moreover, when $\mathbf{y} = \mathbf{z}$ we have $\mathbf{z} = \pm\mu(m-1)$. Hence,

$$|\mathcal{C}(m, m-1)| \ \leq \ |\mathcal{C}(m-1, m-2)| \ + \ 2 \ \leq \ 2m+2\ .$$

We thus conclude that $\mathrm{red}(\mathcal{C}(m, m-1)) = 2^m - 1 - \log_2(m+1)$. The minimum distance of $\mathcal{C}(m, m-1)$ is 2^{m-1}. •

The syndrome-based decoding algorithm for RM codes, as described in [16], can be adapted easily to correct up to $2^{r-1} - 1$ errors (and detect 2^{r-1} errors) in any word of $\mathcal{C}_{\mathbb{R}}(m, r)$ and, hence, in any word of $\mathcal{C}(m, r)$. The outline of the algorithm is as follows. Let $\mathbf{x} = [\,\mathbf{y}\ \mathbf{z}\,] \in \mathcal{C}_{\mathbb{R}}(m, r)$ be the 'transmitted' word, where, by Lemma 3, $\mathbf{z} \in \mathcal{C}_{\mathbb{R}}(m-1, r-1)$ and $\mathbf{y} + \mathbf{z} \in \mathcal{C}_{\mathbb{R}}(m-1, r)$. In fact, since $H_{\mathrm{RM}}(m-1, r-1)$ is a submatrix of $H_{\mathrm{RM}}(m-1, r)$, we also have $\mathbf{y} \in \mathcal{C}_{\mathbb{R}}(m-1, r-1)$. Assume that at most 2^{r-1} errors have occurred, resulting in a 'received' word $\tilde{\mathbf{x}} = [\,\tilde{\mathbf{y}}\ \tilde{\mathbf{z}}\,]$. We first correct the errors in $\tilde{\mathbf{y}} + \tilde{\mathbf{z}}$ to obtain $\mathbf{y} + \mathbf{z}$ by applying recursively the decoding algorithm for $\mathcal{C}_{\mathbb{R}}(m-1, r)$. Next, we attempt to correct each of the words $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{z}}$ by applying the decoding algorithm for $\mathcal{C}_{\mathbb{R}}(m-1, r-1)$. Since one of these words contains no more than 2^{r-1} errors, at least one of the decoding attempts will end up with the correct error locations or (if the number of errors in $\tilde{\mathbf{x}}$ is exactly 2^{r-1}) with the right detection indication. Knowing $\mathbf{y} + \mathbf{z}$, we can then reconstruct the other half of $\mathbf{x}$. The complexity analysis in [16] shows that the running time of such a decoding algorithm is $O(2^r\, V(m, r))$ i.e., it is polynomial in the minimum distance and the number of rows of $H_{\mathrm{RM}}(m, r)$.

5 Bounds on the redundancy of $\mathcal{C}(m, r)$

We now turn to bounding the redundancy of $\mathcal{C}(m, r)$ from above, starting with the following theorem.

Theorem 3. $\mathrm{red}(\mathcal{C}(m,r)) \leq \log_2(2^{m-r}+1) \cdot V(m,r)$.

Proof. It is easy to verify that the theorem holds for $r = m$. For $r < m$ we prove by induction on r, where the claim obviously holds for $r = 0$.

For an integer vector $\mathbf{s} \in \mathcal{Z}^{V(m-1,r)}$, denote by $\mathcal{C}(m{-}1, r; \mathbf{s})$ the set of all words $\mathbf{x} \in \mathcal{C}(m{-}1, r{-}1)$ such that $H_{\mathrm{RM}}(m{-}1, r)\,\mathbf{x}' = \mathbf{s}$. Also, let $A(m{-}1, r)$ denote the set of all vectors $\mathbf{s}$ for which $\mathcal{C}(m{-}1, r; \mathbf{s})$ is nonempty.

Consider the sets

$$\Delta(m, r; \mathbf{s}) \;=\; \Big\{ [\,\mathbf{x}\,\mathbf{y}\,] \in \Phi^{2^m} \mid \mathbf{x} \in \mathcal{C}(m{-}1, r; \mathbf{s}) \text{ and } \mathbf{y} \in \mathcal{C}(m{-}1, r; -\mathbf{s}) \Big\}$$

which are defined for every $\mathbf{s} \in A(m{-}1, r)$. By Lemma 3 we have

$$\mathcal{C}(m, r) \;=\; \bigcup_{\mathbf{s} \in A(m-1,r)} \Delta(m, r; \mathbf{s})\ .$$

Furthermore, by definition, the sets $\Delta(m, r; \mathbf{s})$ are disjoint for distinct vectors $\mathbf{s}$ and $|\Delta(m, r; \mathbf{s})| = |\mathcal{C}(m{-}1, r; \mathbf{s})|^2$. Hence,

$$|\mathcal{C}(m, r)| \;=\; \sum_{\mathbf{s} \in A(m-1,r)} |\mathcal{C}(m{-}1, r; \mathbf{s})|^2\ .$$

Combining this equality with the constraint $\sum_{\mathbf{s} \in A(m-1,r)} |\mathcal{C}(m{-}1, r; \mathbf{s})| = |\mathcal{C}(m{-}1, r{-}1)|$, we obtain

$$|\mathcal{C}(m, r)| \;\geq\; \frac{|\mathcal{C}(m{-}1, r{-}1)|^2}{|A(m{-}1, r)|}\ . \tag{2}$$

Taking logarithms, we finally get the recursion

$$\mathrm{red}(\mathcal{C}(m, r)) \;\leq\; 2\,\mathrm{red}(\mathcal{C}(m{-}1, r{-}1)) \;+\; \log_2 |A(m{-}1, r)|\ . \tag{3}$$

Now, for every $\mathbf{s} \in A(m{-}1, r)$, the entries corresponding to rows of $H_{\mathrm{RM}}(m{-}1, r)$ that are indexed by $\mathbf{v} \in \mathcal{B}(m{-}1, r{-}1)$, all equal to zero. The remaining $\binom{m-1}{r-1}$ entries of $\mathbf{s}$ correspond to rows in $H_{\mathrm{RM}}(m{-}1, r)$ whose Hamming weight is 2^{m-r} and, as such, each of these entries in $\mathbf{s}$ is an even integer that may take at most $2^{m-r}{+}1$ values. Hence, we can bound $|A(m{-}1, r)|$ from above by $(2^{m-r}+1)^{\binom{m-1}{r-1}}$, and (3) thus becomes

$$\mathrm{red}(\mathcal{C}(m, r)) \;\leq\; 2\,\mathrm{red}(\mathcal{C}(m{-}1, r{-}1)) \;+\; \tbinom{m-1}{r-1} \log_2(2^{m-r}+1)\ . \tag{4}$$

Applying the induction hypothesis on (4) we finally obtain

$$\begin{aligned} \mathrm{red}(\mathcal{C}(m,r)) &\leq \log_2(2^{m-r}+1)\cdot\left(2\,V(m{-}1,r{-}1)+\tbinom{m-1}{r-1}\right) \\ &= \log_2(2^{m-r}+1)\cdot V(m,r)\ , \end{aligned}$$

as desired. □

The bound of Theorem 3 is tight for the extreme cases $r = 0$ and $r = m$, but not for values of r in between. One way of improving the bound is by taking into account the fact that the last column of $H_{\mathrm{RM}}(m{-}1,r)$ is an all-one vector which, in turn, allows to provide a sharper estimate on the size of $A(m{-}1,r)$. Another improvement can be obtained by observing that the sizes of the sets $\mathcal{C}(m{-}1,r;\mathbf{s})$ vary with $\mathbf{s} \in A(m{-}1,r)$. We elaborate more on this latter improvement in the appendix, proving the following asymptotic result.

Theorem 4. $\mathrm{red}(\mathcal{C}(m,r)) \leq (m/2)\,V(m,r)\left(1+O((r/m)\,\log\,(m{+}1))\right)$.

We can therefore conclude that when $r = o(m/\log\,m)$, the redundancy of $\mathcal{C}(m,r)$ is not greater than approximately $m/2$ times the redundancy of the respective $(m{-}r)$th order RM code of length 2^m. In particular, for $r = 1$, we obtain the well-known fact that the redundancy of $\mathcal{S}(2^m,1) = \mathcal{C}(m,1)$ is approximately $m/2$. Comparing the bound of Theorem 3 with that of Theorem 4, we can see that the latter is about half the value of the former for the range $r = o(m/\log\,m)$.

It is worthwhile comparing the bounds of Theorems 3 and 4 to the known bounds on $\mathrm{red}(\mathcal{S}(n,r))$. For $n = 2^m$ we have [17]

$$\mathrm{red}(\mathcal{S}(2^m,r)) \;=\; O(2^r m)\ , \tag{5}$$

compared to the bound

$$\mathrm{red}(\mathcal{C}(m,r)) \;=\; O(m^r)$$

implied by the theorems of this section. Recall, however, that the minimum distance of $\mathcal{C}(m,r)$ is *exponential* in r, whereas the minimum distance of $\mathcal{S}(2^m,r)$ is guaranteed to be only *linear* in r [12],[19] and is bounded from above by $r\,(r{-}1)+2$ [11, p. 506],[17]. We point out, however, that it is not yet known as to what extent the bound of (5) is tight: the lower bound on $\mathrm{red}(\mathcal{S}(2^m,r))$ shown in [17] is $(r{-}1)\,(m-r+1)$.

$\frac{m}{r}$	1	2	3	4	5	6	7
1	**1**	**1.42**	**1.87**	**2.35**	**2.84**	**3.33**	**3.83**
2		**3**	**5**	8.75	15.13	22.07	29.93
3			**7**	**12.68**	24.08	46.56	86.06
4				**15**	**28.42**	55.45	109.72
5					**31**	**60.19**	118.88
6						**63**	**124**
7							**127**

Table 1: Upper bounds on $\mathrm{red}(\mathcal{C}(m,r))$ for $m \leq 7$.

We turn next to upper bounds on $\mathrm{red}(\mathcal{C}(m,r))$ for specific values of m and $1 \leq r \leq m$. Table 1 presents such bounds for $m \leq 7$. Boldface entries in the table correspond to exact values of $\mathrm{red}(\mathcal{C}(m,r))$. The values for $r = 1$ were computed by the formula

$$\mathrm{red}(\mathcal{C}(m,1)) \;=\; 2^m \;-\; \log_2 \binom{2^m}{2^{m-1}} \,,$$

and the last two entries in each column equal $2^m - 1 - \log_2(m+1)$ and $2^m - 1$ (see Examples 2 and 3). The other entries in Table 1 were computed using the recursion of Lemma 5 in the appendix, along with the following simple result.

Proposition 2.

$$|\mathcal{C}(m,r)| \;\geq\; |\mathcal{C}(m-1,r)|^2 \;+\; |\mathcal{C}(m-1,r-1)| \;-\; |\mathcal{C}(m-1,r)| \;.$$

Proof. By Lemma 3, the concatenation of any two words in $\mathcal{C}(m-1,r)$ produces a word in $\mathcal{C}(m,r)$, and so does the concatenation of any word $\mathbf{y} \in \mathcal{C}(m-1,r-1)$ with its negated copy $-\mathbf{y}$ (see Example 3). □

Table 2 presents the true values of $\mathrm{red}(\mathcal{C}(m,r))$ for $m \leq 5$. It can be seen that the estimates in Table 1 for $m \leq 5$ are rather close to their true counterparts in Table 2.

$\frac{m}{r}$	1	2	3	4	5
1	1	1.42	1.87	2.35	2.84
2		3	5	8.21	12.38
3			7	12.68	22.89
4				15	28.42
5					31

Table 2: Values of red($\mathcal{C}(m,r)$ for $m \leq 5$.

6 Encoding scheme for $\mathcal{C}(m,2)$

In this section we present an efficient encoding algorithm of unconstrained binary sequences into $\mathcal{C}(m,2)$. The encoding algorithm requires redundancy $O(m^2)$.

For the sake of clarity, we break the description of our algorithm into two pieces. We first describe a so-called balancing procedure that transforms every word $\mathbf{y} \in \Phi^{2^m}$ into a word $\mathbf{x} \in \mathcal{C}(m,2)$ using a set of operations that can be represented by m^2+m bits. Then we show how to embed these redundancy bits into the generated word, thus allowing reconstructing $\mathbf{y}$ out of $\mathbf{x}$.

We start with the balancing procedure. Let $\mathbf{y}$ be an arbitrary word in Φ^{2^m}. As a first phase, we transform $\mathbf{y}$ into a word $\mathbf{x} = [\, x_{\mathbf{u}} \,]_{\mathbf{u} \in F^m} \in \mathcal{C}(m,1)$ by using Knuth's algorithm [13]; namely, we flip the signs of consecutive entries in $\mathbf{y}$ until the resulting word, $\mathbf{x}$, is in $\mathcal{C}(m,1)$. The index of the last flipped entry can be represented by m bits.

Let $\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_{m-1}$ denote the unit vectors in F^m and let $\mathbf{u} \in F^m$ be an index for which $x_{\mathbf{u}} = -x_{\mathbf{u} \oplus \mathbf{e}_i}$. Flipping the signs of $x_{\mathbf{u}}$ and $x_{\mathbf{u} \oplus \mathbf{e}_i}$ (which is the same as switching between their values) will increase or decrease the value of $(H_{\mathrm{RM}}(m,2)\,\mathbf{x}')_{\mathbf{e}_i}$ by 2. On the other hand, the values $(H_{\mathrm{RM}}(m,2)\,\mathbf{x}')_{\mathbf{e}_j}$ for $j \neq i$, as well as $(H_{\mathrm{RM}}(m,2)\,\mathbf{x}')_{\mathbf{0}}$, will remain unchanged.

The following is a balancing procedure for transforming a word $\mathbf{x} \in \mathcal{C}(m,1)$ into a word in $\mathcal{C}(m,2)$.

- For $i = 0, 1, \ldots, m-1$ do:
  - For increasing values of indices $\mathbf{u}$ such that $\mathbf{u} \leq \mathbf{u} \oplus \mathbf{e}_i$, switch between the values

of $x_{\mathbf{u}}$ and $x_{\mathbf{u}\oplus\mathbf{e}_i}$ until $(H_{\mathrm{RM}}(m,2)\,\mathbf{x}')_{\mathbf{e}_i} = 0$.

– Let $\mathbf{u}_i$ denote the last index $\mathbf{u}$ for which the value $x_{\mathbf{u}}$ was changed during the switching process.

Note that if we switch between $x_{\mathbf{u}}$ and $x_{\mathbf{u}\oplus\mathbf{e}_i}$ for *all* indices $\mathbf{u} \leq \mathbf{u} \oplus \mathbf{e}_i$, then $(H_{\mathrm{RM}}(m,2)\,\mathbf{x}')_{\mathbf{e}_i}$ will be negated compared to its original value. Hence, the balancing procedure is always guaranteed to reach a word $\mathbf{x}$ for which $(H_{\mathrm{RM}}(m,2)\,\mathbf{x}')_{\mathbf{e}_i} = 0$. Now, representing the $\mathbf{u}_i$'s requires m^2 bits; therefore, the balancing transformation of $\mathbf{y} \in \Phi^{2^m}$ into $\mathbf{x} \in \mathcal{C}(m,2)$ can be represented by $m^2 + m$ bits.

We next show how to incorporate these $m^2 + m$ bits into the first $2^t = O(m^2)$ locations of $\mathbf{x}$, namely, in the entries $x_{\mathbf{u}}$ indexed by $\mathbf{u} < \mathbf{e}_t$. (We remark that the requirement that the number of redundancy entries in $\mathbf{x}$ be a power of 2 can be relaxed; however, for the sake of clarity we chose to have this slight loss of generality in the description of the encoding algorithm.) To this end, set the first 2^t entries of $\mathbf{x}$ initially to zero. It is easy to verify that the balancing procedure we have described will still work correctly for all $i < t$; that is, the first t executions of the outer loop of the algorithm will generate a word $\mathbf{x}$ with $(H_{\mathrm{RM}}(m,2)\,\mathbf{x}')_{\mathbf{e}_i} = 0$ for $i = 0, 1, \ldots, t-1$. Furthermore, the zero entries in $\mathbf{x}$ will remain unaffected.

On the other hand, for $i \geq t$, we will have indices $\mathbf{u}$ for which exactly one of the entries, $x_{\mathbf{u}}$ or $x_{\mathbf{u}\oplus\mathbf{e}_i}$, is zero. Since the zero entries of $\mathbf{x}$ are not to be affected by the balancing procedure, we need to consider in the switching process only indices $\mathbf{u}$ for which $\mathbf{e}_t \leq \mathbf{u} \leq \mathbf{u} \oplus \mathbf{e}_i$. This constraint, however, requires us also to initially have $\sum_{\mathbf{u}<\mathbf{e}_{i+1}} x_{\mathbf{u}} = 0$, or else we might not reach the equality $(H_{\mathrm{RM}}(m,2)\,\mathbf{x}')_{\mathbf{e}_i} = 0$. Therefore, we will need to allocate more than m bits to represent the somewhat more complex transformation which is initially required to obtain $\mathbf{x}$ out of $\mathbf{y}$. Such a transformation is included in the following encoding procedure.

- Using Knuth's algorithm, transform the unconstrained word $\mathbf{y} \in \Phi^{2^m - 2^t}$ into a word $\mathbf{x} = [\,x_{\mathbf{u}}\,]_{\mathbf{u}\in F^m}$ in which the first 2^t entries are zero, the other entries are in Φ, and $\sum_{\mathbf{e}_i \leq \mathbf{u} < \mathbf{e}_{i+1}} x_{\mathbf{u}} = 0$ for $i = t, t+1, \ldots, m-1$.
- Let $\mathbf{z}$ denote the $t + (t+1) + \ldots + (m-1) < m^2/2$ bits that represent such a transformation.

- For $i = 0, 1, \ldots, m-1$ do:
  - For increasing values of indices $\mathbf{u}$ such that $\mathbf{e}_t \leq \mathbf{u} \leq \mathbf{u} \oplus \mathbf{e}_i$, switch between the values of $x_{\mathbf{u}}$ and $x_{\mathbf{u} \oplus \mathbf{e}_i}$ until $(H_{\mathrm{RM}}(m, 2)\, \mathbf{x}')_{\mathbf{e}_i} = 0$.
  - Let $\mathbf{u}_i$ denote the last index $\mathbf{u}$ for which the value $x_{\mathbf{u}}$ was changed during the switching process.
- Encode the bits of $\mathbf{z}$ and the $\mathbf{u}_i$'s recursively into a word $\mathbf{c} \in \mathcal{C}(t, 2)$ and embed $\mathbf{c}$ into the first 2^t locations of $\mathbf{x}$.

It thus follows that we can encode $\mathbf{y} \in \Phi^{2^m - 2^t}$ into $\mathbf{x} \in \mathcal{C}(m, 2)$ whenever the redundancy 2^t is at least $\frac{3}{2}m^2 + O(\log_2^2 m)$.

We finally point out that in [17], an efficient algorithm is described for encoding arbitrary binary sequences into $\mathcal{S}(n, 2)$. Therefore, as far as the spectral and distance properties are concerned, there is no advantage in using $\mathcal{C}(m, 2)$ instead of $\mathcal{S}(2^m, 2)$, since both have minimum distance 4, whereas the latter has smaller redundancy. This is not the case, however, for larger values of r, where the minimum distance of $\mathcal{C}(m, r)$ is much larger than that of $\mathcal{S}(2^m, r)$. Yet, generalizing the encoding algorithm which was described in this section for larger values of r is still an open problem.

Acknowledgment

The author thanks Tuvi Etzion, Paul Siegel, and Alexander Vardy for helpful discussions.

Appendix

Let the real function $\eta : [0, 1] \to [0, 1]$ be defined by $\eta(t) = ((1-t)\log_2(1-t) + (1+t)\log_2(1+t))/2$. The following Lemma is the well-known Chernoff bound (see, for instance, [14, p. 310]).

Lemma 4. *Let δ be a real number in the interval $[0, 1]$. The number of words $\mathbf{x} = [x_0\, x_1\, \ldots\, x_{n-1}]$ over Φ with $\sum_{j=0}^{n-1} x_j \geq \delta n$ is at most $2^{n\,(1-\eta(\delta))}$.*

Lemma 5. *For $m > r \geq 0$, let δ be a rational number of the form $k/2^{m-r-1}$ for some nonnegative integer $k \leq 2^{m-r-1}$. Assume further that there exists a real number α in the interval $(0,1)$ such that the following inequality is satisfied:*

$$2^{m-1}\eta(\delta) \;\geq\; \mathrm{red}(\mathcal{C}(m{-}1,r{-}1)) \;+\; \log_2 \tbinom{m-1}{r-1} \;+\; 1 \;-\; \log_2 \alpha \;. \tag{6}$$

Then,

$$\mathrm{red}(\mathcal{C}(m,r)) \leq 2\,\mathrm{red}(\mathcal{C}(m{-}1,r{-}1)) + \tbinom{m-1}{r-1} \log_2(\delta\, 2^{m-r} - 1) - 2\,\log_2(1-\alpha) \;. \tag{7}$$

Proof. Fix U to be a subset $\{i_1, i_2, \ldots, i_{r-1}\}$ of $\{0, 1, \ldots, m{-}2\}$ where $i_1 < i_2 < \cdots < i_{r-1}$. For a vector $\mathbf{z} = [z_0\, z_1\, \ldots\, z_{m-2}] \in F^{m-1}$, we denote by $\mathbf{z}_U$ the vector $[z_{i_1}\, z_{i_2}\, \ldots\, z_{i_{r-1}}]$ in F^{r-1}; that is, $\mathbf{z}_U$ is the subvector of $\mathbf{z}$ which is indexed by U. Let $\mathbf{x} = [\, x_{\mathbf{z}}\,]_{\mathbf{z}\in F^{m-1}}$ be a word in $\Phi^{2^{m-1}}$. We denote by $\sigma(\mathbf{x})$ the vector in $\mathcal{Z}^{2^{r-1}}$ whose components are given by

$$(\sigma(\mathbf{x}))_{\mathbf{a}} \;=\; \sum_{\mathbf{z}\in F^{m-1}\,:\,\mathbf{z}_U=\mathbf{a}} x_{\mathbf{z}} \qquad \text{for every } \mathbf{a}\in F^{r-1}.$$

Let $\mathbf{u} \in F^{m-1}$ be the characteristic vector of U and suppose that $\mathbf{x}$ is a word in $\mathcal{C}(m{-}1,r{-}1)$. First, it is easy to verify that $(H_{\mathrm{RM}}(m{-}1,r)\,\mathbf{x}')_{\mathbf{u}} = (\sigma(\mathbf{x}))_{\mathbf{1}}$. Second, for every $\mathbf{v} < \mathbf{u}$ we have

$$\begin{aligned}
\sum_{\mathbf{a}\in F^{r-1}} (-1)^{\mathbf{a}\cdot\mathbf{v}_U} (\sigma(\mathbf{x}))_{\mathbf{a}} \;&=\; \sum_{\mathbf{a}\in F^{r-1}} (-1)^{\mathbf{a}\cdot\mathbf{v}_U} \sum_{\mathbf{z}\in F^{m-1}\,:\,\mathbf{z}_U=\mathbf{a}} x_{\mathbf{z}} \\
&=\; \sum_{\mathbf{z}\in F^{m-1}} (-1)^{\mathbf{z}\cdot\mathbf{v}} x_{\mathbf{z}} \\
&=\; (\mathcal{H}(m{-}1,r{-}1)\,\mathbf{x}')_{\mathbf{v}} \;=\; 0
\end{aligned}$$

i.e.,

$$\mathcal{H}(r{-}1,r{-}1)\,\sigma(\mathbf{x})' \;=\; \mathbf{0}\;.$$

Hence, $\sigma(\mathbf{x})$ is a multiple of $\mu(r{-}1)$ and, as such,

$$|(\sigma(\mathbf{x}))_{\mathbf{a}}| \;=\; |(\sigma(\mathbf{x}))_{\mathbf{1}}| \;=\; |(H_{\mathrm{RM}}(m{-}1,r)\,\mathbf{x}')_{\mathbf{u}}| \quad \text{for every } \mathbf{a}\in F^{r-1}\;.$$

For the fixed set U and for $\delta = k/2^{m-r-1} \in [0,1]$, let $Q(\mathbf{u},\delta)$ be the set defined by

$$Q(\mathbf{u},\delta) = \left\{ \pm\mathbf{x} \in \Phi^{2^{m-1}} \mid (\sigma(\mathbf{x}))_{\mathbf{a}}\,(-1)^{\mathrm{wt}(\mathbf{a})} \;\geq\; \delta\, 2^{m-r} \quad \text{for every } \mathbf{a}\in F^{r-1} \right\}\;.$$

In particular, if $\mathbf{x} \in \mathcal{C}(m{-}1, r{-}1) \cap Q(\mathbf{u}, \delta)$, then

$$|(\sigma(\mathbf{x}))_{\mathbf{a}}| \;=\; |(H_{\mathrm{RM}}(m{-}1, r)\,\mathbf{x}')_{\mathbf{u}}| \geq \delta\, 2^{m-r}$$

for every $\mathbf{a} \in F^{r-1}$. By Lemma 4 and the definition of $Q(\mathbf{u}, \delta)$, we have

$$|Q(\mathbf{u}, \delta)| \;\leq\; 2 \cdot \left(2^{2^{m-r}\,(1-\eta(\delta))}\right)^{2^{r-1}} \;=\; 2 \cdot 2^{2^{m-1}\,(1-\eta(\delta))}\;.$$

Hence, by (6) it follows that

$$|Q(\mathbf{u}, \delta)| \;\leq\; \frac{\alpha\,|\mathcal{C}(m{-}1, r{-}1)|}{\binom{m-1}{r-1}}\;.$$

The proof now continues along the same lines as that of Theorem 3, except that we ignore sets $\mathcal{C}(m{-}1, r; \mathbf{s})$ which correspond to vectors $\mathbf{s}$ containing at least one entry with absolute value $\geq \delta\, 2^{m-r}$. This allows us effectively to reduce the set $A(m{-}1, r)$ into a set $\tilde{A}(m{-}1, r)$ which contains only vectors $\mathbf{s}$ whose entries all have absolute values less than $\delta\, 2^{m-r}$. Since $\delta\, 2^{m-r}$ is assumed to be an even integer, we have $|\tilde{A}(m{-}1, r)| \leq (\delta\, 2^{m-r} - 1)^{\binom{m-1}{r-1}}$ and the inequality (2) becomes

$$|\mathcal{C}(m, r)| \;\geq\; \frac{|\mathcal{C}(m{-}1, r{-}1) - \cup_{\mathbf{u}} Q(\mathbf{u}, \delta)|^2}{|\tilde{A}(m{-}1, r)|} \;\geq\; \frac{(1-\alpha)^2\,|\mathcal{C}(m{-}1, r{-}1)|^2}{(\delta\, 2^{m-r} - 1)^{\binom{m-1}{r-1}}}\;,$$

which, in turn, is equivalent to (7). □

We remark that $|\mathcal{C}(m, r)| > 2\binom{m}{r}$ whenever $m > r > 0$, and, therefore, there always exist α and δ in the specified range such that (6) is satisfied. We can now minimize the right-hand side of (7) over all such values of α and δ. In fact, it suffices to consider only values α that satisfy (6) with equality for a given $\delta = k/2^{m-r-1}$. This is essentially how the entries in Table 1 were computed.

Lemma 6. *For $m \geq r > 0$,*

$$\mathrm{red}(\mathcal{C}(m, r)) \;\leq\; 2\,\mathrm{red}(\mathcal{C}(m{-}1, r{-}1)) \;+\; \tfrac{1}{2}\binom{m-1}{r-1}\,(m \;+\; c_0\, r \log_2(m{+}1))$$

for some absolute constant c_0.

Proof. In view of the bound (4), it suffices to prove the claim for the range $r \leq c_1 \cdot m/\log_2(m{+}1)$, where hereafter c_i denotes a positive absolute constant. Fix α to, say, $\frac{1}{2}$ and let δ' be a positive real such that

$$2\,\log_2\delta' = \log_2\Big(\mathrm{red}(\mathcal{C}(m{-}1,r{-}1)) + 1 + \log_2\tbinom{m-1}{r-1} - \log_2\alpha\Big) - m + 2 - \log_2\log_2 e.$$

By Theorem 3 we have

$$\log_2\delta' \;\leq\; -\frac{m}{2} \;+\; c_2\cdot r\log_2(m{+}1)\;,$$

and, therefore, by choosing a sufficiently small constant c_1, we can guarantee that δ' falls within the interval $(0,1)$. Let δ be the smallest rational of the form $k/2^{m-r-1}$ which is greater than δ'. Noting that $\eta(\delta) \geq \delta^2(\log_2 e)/2$, the inequality (6) is thus satisfied. We now plug the value of $\log_2\delta \leq \log_2\delta' + c_3$ into the following weaker version of (7),

$$\mathrm{red}(\mathcal{C}(m,r)) \leq 2\,\mathrm{red}(\mathcal{C}(m{-}1,r{-}1)) + \tbinom{m-1}{r-1}\,(m + \log_2\delta) + c_4\;,$$

thus yielding the desired result. □

Proof of Theorem 4. Let c_0 be as in Lemma 6 and let $\epsilon(m,r)$ stand for $(c_0\,r/m)\log_2(m{+}1)$. We prove the inequality

$$\mathrm{red}(\mathcal{C}(m,r)) \;\leq\; (m/2)\,V(m,r)\,(1+\epsilon(m,r))$$

by induction on r. The inequality obviously holds when $r=0$, so we assume from now on that $m \geq r > 0$.

By Lemma 6 and the induction hypothesis we have

$$\begin{aligned}\mathrm{red}(\mathcal{C}(m,r)) \;&\leq\; 2\,(m/2)\,V(m{-}1,r{-}1)\,(1+\epsilon(m{-}1,r{-}1))\\ &\qquad +\;\tfrac{1}{2}\tbinom{m-1}{r-1}\,(m\;+\;c_0\,r\,\log_2(m{+}1))\;.\end{aligned}$$

Now, $\epsilon(m,r) \geq \epsilon(m{-}1,r{-}1)$ and, therefore,

$$\begin{aligned}\mathrm{red}(\mathcal{C}(m,r)) \;&\leq\; (m/2)\cdot 2\,V(m{-}1,r{-}1)\,(1+\epsilon(m,r)) + (m/2)\tbinom{m-1}{r-1}\,(1+\epsilon(m,r))\\ &=\; (m/2)\,V(m,r)\,(1+\epsilon(m,r))\;,\end{aligned}$$

as claimed. □

References

[1] S. Al-Bassam, B. Bose, *On balanced codes, IEEE Trans. Inform. Theory,* IT-36 (1990), 406–408.

[2] N. Alon, E.E. Bergmann, D. Coppersmith, A.M. Odlyzko, *Balancing sets of vectors, IEEE Trans. Inform. Theory,* IT-34 (1988), 128–130.

[3] A.M. Barg, *Incomplete sums, DC-constrained codes, and codes that maintain synchronization, Designs, Codes, and Cryptography,* 3 (1993), 105–116.

[4] A.M. Barg, S.N. Lytsin, *DC-constrained codes from Hadamard matrices, IEEE Trans. Inform. Theory,* IT-37 (1991), 801–807.

[5] M. Blaum, *A (16,9,6,5,4) error-correcting DC-free block code, IEEE Trans. Inform. Theory,* IT-34 (1988), 138–141.

[6] E. Eleftheriou, R. Cideciyan, *On codes satisfying Mth order running digital sum constraints, IEEE Trans. Inform. Theory,* IT-37 (1991), 1294–1313.

[7] T. Etzion, *Constructions of error-correcting DC-free block codes, IEEE Trans. Inform. Theory,* IT-36 (1990), 899–905.

[8] H.C. Ferreira, *Lower bounds on the minimum Hamming distance achievable with runlength constrained or DC-free block codes and the synthesis of a (16,8), $D_{\min} = 4$, DC-free block code, IEEE Trans. Magn.,* MAG-20 (1984), 881–883.

[9] W.H. Gottschalk, G.A. Hedlung, *Topological Dynamics, Colloquium Publications of the AMS,* 36, American Math. Society, Providence, Rhode Island, 1955.

[10] H.D.L. Hollmann, K.A. Schouhamer Immink, *Performance of efficient balanced codes, IEEE Trans. Inform. Theory,* IT-37 (1991), 913–918.

[11] L.K. Hua, *Introduction to Number Theory,* Springer, Berlin, 1982.

[12] R. Karabed, P.H. Siegel, *Matched spectral-null codes for partial-response channels, IEEE Trans. Inform. Theory,* IT-37 (1991), 818–855.

[13] D.E. Knuth, *Efficient balanced codes, IEEE Trans. Inform. Theory,* IT-32 (1986), 51–53.

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

[15] C.M. Monti, G.L. Pierobon, *Codes with a multiple spectral null at zero frequency, IEEE Trans. Inform. Theory,* IT-35 (1989), 463–472.

[16] R.M. Roth, G.M. Benedek, *Interpolation and approximation of sparse multivariate polynomials over $GF(2)$, SIAM J. Comput.,* 20 (1991), 291–314.

[17] R.M. Roth, P.H. Siegel, A. Vardy, *High-order spectral-null codes: Constructions and bounds, IEEE Trans. Inform. Theory,* to appear.

[18] K.A. Schouhamer Immink, *Coding Techniques for Digital Recorders,* Prentice-Hall, London, 1991.

[19] K.A. Schouhamer Immink, G. Beenker, *Binary transmission codes with higher order spectral zeros at zero frequency, IEEE Trans. Inform. Theory,* IT-33 (1987), 452–454.

[20] H. van Tilborg, M. Blaum, *On error-correcting balanced codes, IEEE Trans. Inform. Theory,* IT-35 (1989), 1091–1095.