


Clearly, \( \overline{y}_n \in \mathcal{S}(n_k(m+r), k-1) \) and therefore Lemma 5.3 implies that \( m_k(\overline{y}_n) \) is divisible by \( 2k! \) provided \( \mathcal{S}(n_k(m+r), k) \neq \emptyset \). In view of Lemma 2.6, to satisfy the latter condition it would suffice to choose any value of \( n_k \) which is divisible by \( 2^k \). Hence, using Lemma 5.5 we may construct a vector \( \overline{z}_k \in \mathcal{S}(k-1) \) of length

\[
\ell(\overline{z}_k) = O \left( \frac{(m+r)^{\ell+1} \cdot 2^{\frac{k}{2} + O(k)}}{2k!} \right)
\]

such that \( m_k(\overline{z}_k) = -m_k(\overline{y}_n) \). The output of the encoder \( \mathcal{E}(k) \) then consists of a vector \( \overline{u} = (\overline{y}_n | \overline{z}_k) \in \mathcal{S}(k) \).

Let \( \mathcal{C} \) denote the \( k \)-th order spectral-null code at the output of \( \mathcal{E}(k) \), which is the set of all vectors obtained by applying \( \mathcal{E}(k) \) to \( \Phi^n \). Then clearly

\[
\rho(\mathcal{C}) = \ell(\overline{z}_k) + n_k \ell(\overline{z}_{k-1}) + n_k n_{k-1} \ell(\overline{z}_{k-2}) + \cdots = \sum_{i=1}^{k} \ell(\overline{z}_i) \prod_{j=0}^{k-i-1} n_{k-j}
\]

The rate of \( \mathcal{C} \) is given by

\[
R(\mathcal{C}) = \frac{n}{n + \rho(\mathcal{C})} = \frac{1}{1 + \frac{\rho(\mathcal{C})}{n}}
\]

and we have

\[
\frac{\rho(\mathcal{C})}{n} = \sum_{i=1}^{k} \frac{\ell(\overline{z}_i) \prod_{j=0}^{i-1} n_{k-j}}{\prod_{j=1}^{l} n_{j}} = \sum_{i=1}^{k} \frac{\ell(\overline{z}_i)}{\prod_{j=1}^{l} n_{j}}
\]  

(30)

Note that, in view of (29), the value of \( \ell(\overline{z}_i) \) depends on \( n_1, n_2, \ldots, n_{i-1} \) but \emph{not} on \( n_i \). Hence by taking \( n_i \) sufficiently large for all \( i = 1, 2, \ldots, k \) we can make each of the \( k \) terms in (30) as close to zero as desired. Therefore, \( \limsup_{n_1, n_2, \ldots, n_k \to \infty} R(\mathcal{C}) = 1 \) as claimed.

**Acknowledgement.** We would like to thank Tuvi Etzion for helpful discussions.

**References**


where $W(k)$ is given by (28), while $\delta_1 = j_2 - j_1$ and $\delta_2 = j_4 - j_3$. W.l.o.g. we may assume that $W(k-1)$ is positive, otherwise exchange the roles of $\mathbf{u}_{k-1}$ and $-\mathbf{u}_{k-1}$ in the foregoing construction. Clearly $\delta_1$ and $\delta_2$ could not be less than the lengths of $\mathbf{u}_{k-1}$ and $\mathbf{z}_{k-1}$, respectively, but otherwise are arbitrary. Thus we may take $\delta_1$ to be the smallest odd integer $\geq \ell(\mathbf{u}_{k-1})$, such that $W(k-1)\delta_1 > 2\ell(\mathbf{z}_{k-1})$. Since both $\delta_1$ and $W(k-1)$ are odd, we may furthermore take $2\delta_2 = W(k-1)\delta_1 + 1$. In this case we have $m_k(\mathbf{z}) = 2k!$ and thereby the lemma is proved. \[\]

It is clear from (27) and (28) that $\ell(\mathbf{u}_k) = O(2^k)$ and $W(k) = O(2^{k(k+1)/2})$. Substituting this in the construction of Lemma 5.4 we see that $\ell(\mathbf{z}_k) = O(\ell(\mathbf{z}_{k-1})2^k) = 2^{k^2} + O(k)$. Hence we have the following lemma.

**Lemma 5.5.** For any $k \geq 1$ and any integer $t$, there exists a vector $\mathbf{z} \in \mathcal{S}(k)$ of length at most $t \cdot 2^{k^2} + O(k)$ with $m_k(\mathbf{z}) = 2k!$.

*Proof.* Take $\delta_1 \equiv t \pmod{2}$ and $2\delta_2 = W(k-1)\delta_1 + t$ in the construction of Lemma 5.4. \[\]

Lemma 5.5 is the required generalization of Lemma 5.1.

### 5.3. An encoder for $k$-th order spectral-null codes

It is now clear how the encoder of Section 5.1 may be extended for null orders greater than two. Let $\mathcal{E}(k)$ denote such a general encoder from $\Phi^n$ into $\mathcal{S}(k)$. Then $\mathcal{E}(k)$ may be specified recursively as follows. Assume that the length of the information word $\mathbf{x}$ to be encoded is given by $n = n_k n_{k-1} \cdots n_1$, and denote $m = n/n_k = n_{k-1} n_{k-2} \cdots n_1$. First $\mathbf{x}$ is partitioned into $n_k$ blocks $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{n_k}$ of length $m$. Subsequently, each $\mathbf{x}_i$ is mapped into a vector $\mathbf{z}_i \in \mathcal{S}(m + r, k-1)$, where $r$ is the redundancy associated with the encoder $\mathcal{E}(k-1)$ applied to vectors of length $m$. The vectors $\mathbf{z}_i$ are then concatenated, and possibly negated, to obtain the series $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_{n_k}$ where $\mathbf{y}_1 = \mathbf{z}_1$ and

\[
\mathbf{y}_i = \begin{cases} (\mathbf{y}_{i-1} \mid \mathbf{z}_i) & m_k(\mathbf{y}_{i-1}) \cdot m_k(\mathbf{z}_i) \leq 0 \\ (\mathbf{y}_{i-1} \mid -\mathbf{z}_i) & m_k(\mathbf{y}_{i-1}) \cdot m_k(\mathbf{z}_i) > 0 \end{cases}
\]

This ensures that

\[
|m_k(\mathbf{y}_{n_k})| \leq \max_{1 \leq i \leq n_k} |m_k(\mathbf{z}_i)| \leq (m + r)^{k+1}
\]
Using Lemma 2.1, it is easy to see that \( m_i(\vec{u}_k) = 0 \) for all \( i = 0, 1, \ldots, k-1 \) and

\[
m_k(\vec{u}_k) = -k \cdot \left( \ell(\vec{u}_{k-1}) + \frac{1 - (-1)^k}{2} \right) \cdot m_{k-1}(\vec{u}_{k-1})
\]  

(26)

where \( \ell(\vec{u}_k) \) is the length of \( \vec{u}_k \). Further, we have

\[
\ell(\vec{u}_k) = \frac{4 \cdot 2^k + (-1)^{k+1} + 1 - (-1)^k}{3}
\]  

(27)

Substituting this into (26) and solving the recursion, we obtain

\[
m_k(\vec{u}_k) = k! \cdot \prod_{i=1}^{k-1} \left[ \frac{(-1)^i - 4 \cdot 2^i}{3} \right] \equiv k! \cdot W(k)
\]  

(28)

the empty product being 1 by convention. It is easy to see that \( m_k(\vec{u}_k) \) is indeed an odd multiple of \( k! \).

We presently employ the series of ternary vectors \( \vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k \) in order to construct a series of binary vectors \( \vec{z}_1, \vec{z}_2, \ldots, \vec{z}_k \), such that \( \vec{z}_k \in \mathcal{S}(k) \) and \( m_k(\vec{z}) = 2k! \).

**Lemma 5.4.** For any \( k \geq 1 \), there exists a vector \( \vec{z}_k \in \mathcal{S}(k) \) with \( m_k(\vec{z}_k) = 2k! \).

**Proof.** For \( k = 1, 2 \) the lemma follows by considering \( \vec{z}_1 = (-+--) \in \mathcal{S}(4,1) \) and \( \vec{z}_2 = (++-) \in \mathcal{S}(4,2) \). As an induction hypothesis, assume the existence of a vector \( \vec{z}_{k-1} \in \mathcal{S}(k-1) \) such that \( m_{k-1}(\vec{z}_{k-1}) = 2(k-1)! \). We now construct the following ternary vector

\[
\vec{v} = (\ldots, \vec{u}_{k-1}, \ldots, -\vec{u}_{k-1}, \ldots, -\vec{z}_{k-1}, \ldots, \vec{z}_{k-1}, \ldots)
\]

meaning that \( \vec{u}_{k-1} \) and \(-\vec{u}_{k-1}, \) given by (25), are contained in \( \vec{v} \) starting at positions \( j_1 \) and \( j_2 \), while \(-\vec{z}_{k-1} \) and \( \vec{z}_{k-1} \) are contained in \( \vec{v} \) starting at positions \( j_3 \) and \( j_4 \). Let \( \vec{e} \) be an arbitrary ternary vector, such that \( \epsilon_i = 0 \) if and only if \( v_i = 0 \). Note that \( \pm \vec{v} + \vec{e} \) are both binary vectors over \( \Phi \). Hence by Lemma 2.6 there exists a vector \( \vec{y} \in \mathcal{S}(k) \), such that \( \vec{v} + \vec{e} \) is a prefix of \( \vec{y} \). Consider the binary vector \( \vec{z} \) of length \( \ell(\vec{z}) = \ell(\vec{y}) \) obtained from \( \vec{y} \) by changing the prefix \( \vec{v} + \vec{e} \) to \(-\vec{v} + \vec{e} \). From the construction of \( \vec{z} \) and \( \vec{y} \) it follows that

\[
m_i(\vec{z}) = m_i(\vec{y}) + 2 \sum_{l=0}^{i} \left( \binom{i}{l} (j_4 - j_3) m_{i-l}(\vec{z}_{k-1}) - 2 \sum_{l=0}^{i} \left( \binom{i}{l} (j_2 - j_1) m_{i-l}(\vec{u}_{k-1}) \right) \right)
\]

Since all the moments of \( \vec{z}_{k-1} \) and \( \vec{u}_{k-1} \) vanish up to order \( k-2 \), \( m_i(\vec{z}) \) is obviously 0 for \( i = 0, 1, \ldots, k-2 \). Furthermore, substituting \( i = k-1 \) into the above expression we obtain \( m_{k-1}(\vec{z}) = 0 \) and hence \( \vec{z} \in \mathcal{S}(k) \). A similar argument now shows that

\[
m_k(\vec{z}) = 2k \cdot (j_4 - j_3) m_{k-1}(\vec{z}_{k-1}) - 2k \cdot (j_2 - j_1) m_{k-1}(\vec{u}_{k-1}) = 2k! \cdot 2\delta_2 - 2k! \cdot W(k-1) \cdot \delta_1
\]
5.2. Construction of the balancing sequence

It is evident that in order to extend the construction of the previous subsection beyond $k = 2$, we need the analogue of Lemma 5.1 for arbitrary large values of $k > 2$. More specifically, let $x_1, x_2, \ldots, x_s \in S(n, k)$ and let $S = m_k(x_1) \pm m_k(x_2) \pm \cdots \pm m_k(x_s)$ with $|S| \leq \max_{1 \leq i \leq s} |m_1(x_i)|$. Then we have to be able to construct a vector $\vec{z} \in S(k)$ whose length does not depend on $s$, such that $m_k(\vec{z}) = S$. To this end we proceed as follows. First we show that $k! | S$, and furthermore if $S(n, k+1) \neq \emptyset$ then $2k! | S$. Then we show how to construct a vector $\vec{z} \in S(k)$, such that $m_k(\vec{z})$ is any prescribed multiple of $2k!$.

Lemma 5.2. Let $\vec{x} \in S(n, k)$. Then $m_k(\vec{x})$ is divisible by $k!$.

Proof. Let $D(n, k+1)$ be the "systematic" parity-check matrix for $S(n, k+1)$ as in (7) and let $B(k+1)$ be the inverse of $H(k+1, k+1)$, as defined in Lemma 2.3. Further, for any $\vec{x} \in \Phi^n$ let $s_k(\vec{x}) = (s_0(\vec{x}), s_1(\vec{x}), \ldots, s_k(\vec{x}))^t \overset{\text{def}}{=} D(n, k+1)\vec{x}^t$. Now, if $\vec{x} \in S(n, k)$ then obviously $H(n, k+1)\vec{x}^t = (0, 0, \ldots, 0, m_k(\vec{x}))^t$. Hence we have

$$s_k(\vec{x}) = D(n, k+1)\vec{x}^t = B(k+1)H(n, k+1)\vec{x}^t = B(k+1)(0, 0, \ldots, 0, m_k(\vec{x}))^t$$

Thus $s_k(\vec{x}) = b_{k,k}m_k(\vec{x}) = m_k(\vec{x})/k!$ where the second equality follows from (9). Yet, it was shown in Lemma 2.3 that $D(n, k+1)$ is an integer matrix, and hence $s_k(\vec{x})$ must be an integer. It follows that $k!$ divides $m_k(\vec{x})$ for all $\vec{x} \in S(n, k)$. □

Lemma 5.3. If $S(n, k+1) \neq \emptyset$ then $m_k(\vec{x})$ is divisible by $2k!$ for all $\vec{x} \in S(n, k)$.

Proof. As we have seen, if $\vec{x} \in S(n, k)$ then $s_k(\vec{x}) = m_k(\vec{x})/k!$. On the other hand, $s_k(\vec{y}) = 0$ for any $\vec{y} \in S(n, k+1)$. Since $\vec{x} \equiv \vec{y} \pmod{2}$ for all $\vec{x}, \vec{y} \in \Phi^n$, it follows that $s_k(\vec{x}) \equiv s_k(\vec{y}) \equiv 0 \pmod{2}$, provided $S(n, k+1) \neq \emptyset$. In other words, $s_k(\vec{x})/k! = s_k(\vec{y}) \equiv 0 \pmod{2}$, and therefore $2k!$ must divide $m_k(\vec{x})$. □

We point out that there are examples where $S(n, k+1) = \emptyset$ and indeed $\exists \vec{x} \in S(n, k)$ such that $m_k(\vec{x})$ is divisible by $k!$ but not by $2k!$. For instance consider $\mu(1) = (-+, +) \in S(2, 1)$ with $m_1(\mu(1)) = 1$. For a less trivial example take the vector $\vec{z} \in S(20, 3)$ with $m_3(\vec{z}) = 6$, obtained by concatenating $+-+-+-++$ with the complement of its reflection. Similar examples exist for $k = 3$ and $n = 28$. In general, however, it can be verified by induction on $k$ that $m_k(\mu(k)) = (-1)^k 2^{(k-1)/2} k!$, and hence $m_k(\mu(k))$ is divisible by $2k!$ for all $k > 1$. In fact, for $k > 3$, we do not know of any binary sequence $\vec{x} \in S(n, k)$ for which $m_k(\vec{x})$ is an odd multiple of $k!$. On the other hand, it is relatively easy to construct a series of ternary vectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k$ such that $m_i(\vec{u}_k) = 0$ for all $i = 0, 1, \ldots, k-1$ and $m_k(\vec{u}_k)$ is an odd multiple of $k!$ for all $k > 1$: Set $\vec{u}_1 = (-+, +)$ and for $k = 1, 2, \ldots$ define

$$\vec{u}_{k+1} = \begin{cases} (\vec{u}_k \mid -\vec{u}_k) & k \equiv 0 \pmod{2} \\ (\vec{u}_k \mid 0 \mid -\vec{u}_k) & k \equiv 1 \pmod{2} \end{cases} \quad (25)$$

25
Now set $y_1 = x_i$. For $i = 2, 3, \ldots, n_2$ define the vector $y_i$ as follows:

$$y_i = \begin{cases} (y_{i-1} | x_j) & m_1(y_{i-1}) \cdot m_1(x_j) \leq 0 \\ (y_{i-1} | -x_j) & m_1(y_{i-1}) \cdot m_1(x_j) > 0 \end{cases}$$  (23)

Thus $y_{n_2}$ is essentially a concatenation of the vectors $x_1, x_2, \ldots, x_{n_2}$, with some of these vectors being negated. The first position in each such vector indicates whether it has been negated or not. Note that the first position in $y_{n_2}$ remains fixed at +1. Clearly $m_0(y_{n_2}) = 0$. It is also clear from Lemma 2.1 that

$$m_1(y_{n_2}) = m_1(x_1) \pm m_1(x_2) \pm \cdots \pm m_1(x_{n_2})$$  (24)

Furthermore, the simple polarity inversion technique of (23) ensures that the terms in (24) always add-up in such a way that $|m_1(y_{n_2})| \leq \max_{1 \leq i \leq n_2} |m_1(x_j)| \leq (n_1 + 1 + r)^2/4$ where the second equality follows by (22). We shall assume that $n_1 + 1 + r \equiv 0 \mod 4$, in which case $m_1(y_{n_2})$ must be even. In order to complete the encoding we need a vector $z \in S(1)$ with $m_1(z) = -m_1(y_{n_2})$. Such a vector always exists, and has length at most $(n_1 + 1 + r)$ as is shown in the following lemma.

**Lemma 5.1.** Let $n \equiv 0 \mod 4$. Then for any even integer $t$ with $|t| \leq n^2/4$, there exists a vector $z \in S(n, 1)$ with $m_1(z) = t$.

**Proof.** Let $z = (z_1, z_2, \ldots, z_n)$ and assume that either $(++)$ or $(+-)$ is contained in $z$ at position $j$. We may then define $F_j z = (z_1, z_2, \ldots, -z_j, -z_{j+1}, \ldots, z_n)$, where the effect of the flip operator $F_j$ amounts to interchanging the positions of + and − in the coordinates $j$ and $j+1$. Now set $z_0 = (- \cdots - + + \cdots +) \in S(n, 1)$. Clearly $m_0(z_0) = 0$, $m_1(z_0) = n^2/4$, and $(+-)$ is contained in $z_0$ at position $n/2$. For $i = 1, 2, \ldots, n^2/4$ define $z_i = F_j z_{i-1}$, where $j$ is the smallest index such that $(+-)$ is contained in $z_{i-1}$ at position $j$. It is easy to see that as $i$ varies from 0 to $n^2/4$, the first moment of $z_i$ takes on all the even values in the range $+n^2/4$ to $-n^2/4$.

Using the algorithm of Lemma 5.1 we can readily construct a vector $z \in S(n_1 + 1 + r, 1)$ with $m_1(z) = -m_1(y_{n_2})$. The output of the encoder then consists of the sequence $y = (y_{n_2} | z)$, which clearly satisfies $m_0(y) = m_1(y) = 0$.

Let $R(C)$ denote the rate of the second-order spectral-null code $C$ consisting of all the vectors of length $(n_2 + 1)(n_1 + r + 1)$ obtained using this construction. Then obviously,

$$R(C) = \frac{\log_2 |C|}{(n_2 + 1)(n_1 + r + 1)} = \frac{n_1 n_2}{(n_2 + 1)(n_1 + r + 1)}$$

Since $r$ approaches $\log n_1$ as $n_1 \to \infty$, it is easy to see that $\lim_{n_1, n_2 \to \infty} R(C) = 1$. We note that the optimal choice of parameters $n_1, n_2$ in this case is $n_2 = n_1 / \log n_1$, which yields $\rho(C) = O(\sqrt{n \log n})$. 

24
5. A general encoding scheme

In this section we present a recursive encoding scheme for mapping arbitrary sequences over $\Phi$ into spectral-null sequences of order $k$, for any fixed value of $k$. We start in Section 5.1 with the description of this encoding scheme for the special case $k = 2$, which illustrates the basic ideas involved in our construction. The resulting second-order spectral-null codes have higher redundancy than the codes introduced in Section 4. However, unlike the construction of Section 4, the construction of Section 5.1 naturally extends to values of $k$ greater than two. We first prove in Section 5.2 that $m_k(\underline{z})$ is divisible by $k!$ for any $\underline{z} \in S(n, k)$ and, furthermore, that $m_k(\underline{z})$ is divisible by $2k!$ if $S(n, k+1) \neq \emptyset$. Then we show how to construct a vector $\underline{y} \in S(k)$ such that $m_k(\underline{y})$ is any prescribed even multiple of $k!$. These results are employed in Section 5.3 to present a recursive construction of spectral-null codes of order $k$, for any fixed $k$. Furthermore it is shown in Section 5.3 that the rate of these codes approaches $\text{cap}(S(k)) = 1$ as their length goes to infinity.

We point out that while the encoding scheme described herein is fairly simple to implement for the first few values of $k$ (say, $k \leq 4$), it becomes impractical as the order of the null increases. Thus for large values of $k$ our encoder is best regarded as yet another way to prove that $\text{cap}(S(k)) = 1$. Such a proof differs from the existence result of Theorem 2.8, in the sense that it provides an explicit encoder from $\Phi^n$ into $S(k)$ which achieves the capacity. Unlike the enumerative encoding scheme, the proposed encoder features complexity which is polynomial in both $n$ and $k$ (although $n$ has to tend to infinity nonuniformly with respect to $k$ in order for the rate to approach unity).

5.1. An alternative encoder for second-order spectral-null codes

Let $\underline{v}$ be the sequence over $\Phi$ which is to be encoded into $S(2)$, and further assume that the length of $\underline{v}$ is $n = n_1n_2$, where $n_1$ is odd. We first partition this sequence as $\underline{v} = (\underline{v}_1 | \underline{v}_2 | \cdots | \underline{v}_{n_2})$, where $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_{n_2} \in \Phi^{n_1}$, and then extend each $\underline{v}_i$ by an extra coordinate fixed at $+1$ to obtain $\underline{v}'_i = (+ | \underline{v}_i)$ of length $n_1 + 1$. Subsequently, each $\underline{v}'_i$ is encoded into a vector $\underline{x}_i \in S(n_1+1+r, 1)$, such that the first position in each $\underline{x}_i$ remains $+1$. This may be accomplished in a number of ways. To be specific, we assume that one of the algorithms of Knuth [17] is employed, in which case $\underline{x}_i = (\underline{v}'_i \cdot \underline{w}_i | \underline{w}_j)$, where $\cdot$ stands for bit by bit multiplication,

$$\underline{u}_j = \left( \begin{array}{cccc} + & + & \cdots & + \end{array} \right)_{j} \overbrace{\begin{array}{c} n_1+1-j \end{array}} \ldots \vdots \begin{array}{c} + \end{array}$$

for some index $j$ such that $\underline{v}'_i \cdot \underline{u}_i \in S(n_1+1, 1)$, and $\underline{w}_j \in S(r, 1)$ is a representation of $j$. Note that in this case $r = \log_2 n_1 + O(1)$.
Example. Assume that \( n = 26 \), in which case \( m = 5 \) and \( h = 16 \). Further assume that the vector \( \underline{y} \in \Phi^{26} \) to be encoded is given by

\[
\underline{y} = (- - - + + + + - - - - + + - - + - - - + + + )
\]

Then after step A1 we have

\[
\underline{x} = (- - - + + + + - - - - + + 0 0 0 0 + 0 - - + 0 - - - + + + )
\]

with \( \sigma_1(\underline{x}) = 64 \). Applying the procedure of step A2 yields \( l_0 = -14 \) and we have

\[
\underline{x} = ( + + + - + + + + - - - - + + 0 0 0 0 + 0 - - + 0 - - - + + + )
\]

with \( S = \sigma_1(\underline{x}; -14) = 2 \leq 16 \). The binary representation of the integer \( S + 2^4 - 1 = 17 \) is given by \( 1 + 2^4 \). Hence \((b_0, b_1, b_2, b_3) = (0, 0, 0, 1)\), and after step A3 we have

\[
\underline{x} = ( + + - + + + + - - - - + + 0 + + + - - - - + + + + + )
\]

with \( \sigma_1(\underline{x}) = 0 \) as desired. Now \( \sigma_0(\underline{x}) = 5 \). Applying step B1 yields \( i_0 = 2 \) with

\[
\underline{x} = ( + + - + + + + - - - - + + - 0 + + + - - - - + + + + + )
\]

and \( \sigma_0(\underline{x}, 2) = 1 \). Thus at step B2 we set \( x_0 = -1 \), which produces the final encoded vector

\[
\underline{x} = ( + + - + + + + - - - - + + - + - - - + + + + + + )
\]

with \( \underline{x} \in \mathcal{S}(32, 2) \). The binary representations of \( l_0 + h = 2 \) and \( i_0 = 2 \) are now encoded recursively and appended to \( \underline{x} \).

\[ \square \]

It is clear from the foregoing description that, presented with an arbitrary vector \( \underline{y} \) of length \( n \), our algorithm will encode \( \underline{y} \) into a second-order spectral-null sequence of length \( n + 3m + O(\log m) \), where \( m \) is the smallest integer such that \( n + m + 1 \leq 2^m \). Obviously \( m = O(\log n) \). Hence the redundancy of the second-order spectral-null code \( \mathcal{C} \) which is the image of the proposed encoder is given by \( \rho(\mathcal{C}) = 3 \log_2 n + O(\log \log n) \).
Without loss of generality assume that \( S \) is nonnegative, or else apply the following argument to \(-S\). The binary expansion of an odd integer \( S + 2^{m-1} - 1 \) may be written as

\[
S + 2^{m-1} - 1 = 1 + \sum_{s=0}^{m-2} b_s 2^s + 1
\]

where \( b_s \in \{0, 1\} \) for \( s = 0, 1, \ldots, m - 2 \). Substituting \( 2^{m-1} - 1 = \sum_{s=0}^{m-2} 2^s \) in the above expression, we obtain

\[
S = 1 + \sum_{s=0}^{m-2} (2b_s - 1) 2^s.
\]

Hence we set \( x_{-1} = +1 \) and \( x_{2^s - 1} = 1 - 2b_s \) for \( s = 0, 1, \ldots, m - 2 \).

Having encoded \( y \) into a vector \( x \) which satisfies the condition \( \sigma_1(x) = 0 \), we now apply the second phase of the algorithm which ensures that \( \sigma_0(x) = \sum_{j=-h}^{h-1} x_j = 0 \).

**Phase B: Balancing \( \sigma_0(x) \)**

**Step B1.** Call an index \( i \) qualifying if \( x_i = x_{-i} \). For increasing values of qualifying indices \( i \geq 1 \) flip the signs of both \( x_i \) and \( x_{-i} \), and let \( \sigma_0(x; i) \) denote the value of \( \sigma_0(x) \) just prior to flipping \( x_i \) and \( x_{-i} \). Proceed until \( |\sigma_0(x; i)| = 1 \), and let \( i_0 \) denote the (smallest) index \( i \) for which this condition is met.

**Step B2.** If \( \sigma_0(x; i_0) = +1 \) set \( x_0 = -1 \). Otherwise, set \( x_0 = +1 \).

To verify that the condition of step B1 is indeed met for some \( i < h \), notice that \( \sigma_0(x; i) \) is odd for every qualifying index \( i \) and that \( \sigma_0(x; i) \) increases or decreases by either 0 or 4 at each sign flip. Flipping the signs of \( x_i \) and \( x_{-i} \) for all qualifying indices \( i \) in the range \( 1 \leq i < h \) will result in negating the initial value of \( \sigma_0(x) \). Hence the condition of step B1 must be met for some \( i < h \).

Note that Phase B of our algorithm essentially consists of one of the algorithms in Knuth’s paper [17], applied only to those positions in \( x \) where the two (reflected) halves of \( x \) agree. Such a process guarantees that the value of \( \sigma_1(x) \) will not be affected by the sign flippings performed in Phase B. Thus at the output of Phase B we have a vector \( x \in \Phi^{2h} \) such that \( \sigma_0(x) = \sigma_1(x) = 0 \). It therefore follows that \( x \in S(2h, 2) \), by Lemma 2.1 of Section 2.

The final phase of our algorithm may be specified recursively as follows.

**Phase C: Encoding the indices**

**Step C1.** Apply Phase A and Phase B recursively to the binary representations of \( l_0 + h \) and \( i_0 \) which were computed in steps A2 and B1, respectively. Concatenate the resulting vector with \( x \) as the final output of the encoder.
Let $n$ be a positive integer and let $m$ be the smallest integer such that $n + m + 1 \leq 2^m$.
We further assume that $n + m + 1$ is divisible by 4, or else we may increase $n$ by at most 3
to meet this condition. Thus, let $h$ be the even integer $(m + n + 1)/2$. We now show how to
encode an arbitrary sequence $y$ in $\Phi^n$ into a word of $S(2h, 2)$.

As a first phase, we encode $y$ into a vector $\underline{x} = (x_{-h} x_{-h+1} \ldots x_0 x_1 \ldots x_{h-1})$ over $\Phi$ which
satisfies the equation $\sigma_1(\underline{x}) \equiv \sum_{j=-h}^{h-1} j x_j = 0$. Note that $\sigma_1(\underline{x})$ is essentially the
first moment of $\underline{x}$ with respect to the matrix $H(2h, 2; -(h+1))$. The encoding procedure may
be specified as follows.

**Phase A: Balancing $\sigma_1(\underline{x})$**

**Step A1.** Assign the entries of $\underline{y}$ to the entries $x_j$, where $j$ ranges over all integers
between $-h$ and $h - 1$ that are not equal to $-1, 0, 1, 2, 4, \ldots 2^{m-2}$. For the time
being, set the $m + 1$ unassigned entries of $x_j$ to zero.

**Step A2.** For increasing values of $l = -h, -h+1, \ldots$, flip the sign of $x_l$. Let $\sigma_1(\underline{x}; l)$
denote the value of $\sigma_1(\underline{x})$ just prior to flipping the sign of $x_l$. Proceed until the
absolute value of $\sigma_1(\underline{x}; l)$ is not greater than $h$, and let $l_0$ denote the (smallest) index $l$
for which this condition is met. Set $l_0 = h$ if the whole vector $\underline{x}$ was negated.

**Step A3.** For $j = -1, 1, 2, 4, \ldots 2^{m-2}$ set the entries $x_j$ to $+1$ or $-1$ so that the
resulting overall sum $\sigma_1(\underline{x}) = \sum_{j=-h}^{h-1} j x_j$ is zero.

We point out that the value of $x_0$ has not been set in the above procedure, neither does
its value affect $\sigma_1(\underline{x}) = \sum_{j=-h}^{h-1} j x_j$.

We presently show that the foregoing algorithm will always find an index $l < h$ for which
$|\sigma_1(\underline{x}; l)| \leq h$. Indeed, flipping the signs of every entry in $\underline{x}$ negates $\sigma_1(\underline{x})$ with respect to
its initial value, that is $\sigma_1(\underline{x}; h) = -\sigma_1(\underline{x}; -h)$. Thus there exists an $l$ such that
$\sigma_1(\underline{x}; l) \cdot \sigma_1(\underline{x}; l+1) \leq 0$. Furthermore, $|\sigma_1(\underline{x}; l+1) - \sigma_1(\underline{x}; l)| \leq 2h$ for all $l = -h, -h+1, \ldots h-1$.
Hence we must reach in step A2 an index $l_0$ for which $|\sigma_1(\underline{x}; l_0)| \leq h$.

Next, we show how to compute the entries $x_j$ for $j = -1, 1, 2, 4, \ldots 2^{m-2}$ in step A3 of
the algorithm. Denote $S = \sigma_1(\underline{x}; l_0)$. First note that the value of $\sigma_1(\underline{x}; l)$ is always even.
This is due to the fact that $j(x_j - x_{-j})$ is even for every $0 < j < h$, and so is $-h x_{-h}$.
It remains to prove that every even integer $S$ in the range $-2^{m-1} \leq S \leq 2^{m-1}$ can be
written in the form

$$S = -x_{-1} + \sum_{i=0}^{m-2} x_{2^i} 2^i$$

where $x_{-1}, x_1, x_2, \ldots x_{2^{m-2}} \in \Phi$. 

20
4. An encoder for second-order spectral-null codes

While the previous two sections are devoted to the study of various properties of $\mathcal{S}(n, k)$, in this and the next section we present specific encoders into subsets of $\mathcal{S}(n, k)$, viz. spectral-null codes of order $k$. In this section an encoding scheme for second-order spectral-null codes is described. In the next section we present an alternative encoding scheme, which extends to spectral-null codes of any fixed order.

One way of encoding an arbitrary vector $\underline{y} = (y_1, y_2, \ldots, y_m)$ of length $m = n - \lfloor \rho(\mathcal{S}(n, k)) \rfloor$ over the alphabet $F = \{0, 1\}$ into a vector $\underline{x} \in \mathcal{S}(n, k)$ is by means of enumerative coding. For instance, assume that all the elements of $\mathcal{S}(n, k)$ have been arranged in lexicographic order, and a 1-1 map $b : F^m \rightarrow \{0, 1, \ldots, 2^m - 1\}$ has been established, say $b(\underline{y}) = \sum_{i=1}^{m} y_i 2^{i-1}$. Then the enumerative encoder, presented with the sequence $\underline{y}$, encodes $\underline{y}$ into a vector $\underline{x} \in \mathcal{S}(n, k)$ whose rank in the lexicographic ordering is equal to $b(\underline{y})$. In particular, such an enumerative encoder for $\mathcal{S}(n, 2)$ was proposed in [12].

We note that the enumerative encoding technique can be, in principle, extended for $k \geq 2$. This, however, would require pre-computing and storing a prohibitively large amount of information. For an integer vector $\underline{z} = (z_0, z_1, \ldots, z_{k-1}) \in \mathbb{Z}^k$ let

$$\mathcal{S}(n, k; \underline{z}) = \{ \underline{z} \in \Phi^n : H(n, k) \underline{z}^t = \underline{z} \}$$

Then the enumerative encoding algorithm requires the knowledge of the (nonzero) values of $|\mathcal{S}(l, k; \underline{z})|$ for all $\underline{z} \in \mathbb{Z}^k$ and all $l = 1, 2, \ldots, n$. These values may be pre-computed using dynamic programming. However, for any fixed $k$, the $i$-th entry $z_i$ of $\underline{z}$ in $\mathcal{S}(l, k; \underline{z})$ may range over $\Theta(i^{i+1})$ values\(^1\). Hence, for each $l$ we may end up with $\Theta(l^{k+1}/2)$ nonzero values $|\mathcal{S}(l, k; \underline{z})|$, and therefore the number of nonzero values we would need to compute and store in this way is $\Theta(n^{k(k+1)/2} + 1)$. This makes the enumerative method quite impractical even for small values of $k$.

In this section we concentrate on the case $k = 2$. The enumerative coding technique we have just outlined will require us to pre-compute and store $\Theta(n^4)$ values of $|\mathcal{S}(l, 2; \underline{z})|$, and the redundancy of the encoded set of words of $\mathcal{S}(n, 2)$ thus obtained is $\Theta(\log n)$. We now present an alternative encoding algorithm for $k = 2$ which requires $O(n \log n)$ bit operations for encoding, without any pre-computation, and whose resulting redundancy is still $\Theta(\log n)$. In a way, this algorithm may be regarded as a generalization of one of the algorithms in Knuth’s paper [17] for the case $k = 2$.

\(^1\)Here $\Theta(f(n))$ denotes a function in $n$ which is bounded from below and above by $c_1 \cdot f(n)$ and $c_2 \cdot f(n)$, respectively, for some constants $c_1$ and $c_2$ independent of $n$. 
solutions \( \{n/2\} \in S(n, 1) \) and \( \{3n/4, n/4\} \in S(n, 2) \) correspond to the Kronecker product of the vector \((+ \cdots + +)\) of appropriate length with the Morse sequences \( \mu(1) \in S(2, 1) \) and \( \mu(2) \in S(4, 2) \), respectively.

Application of Lemma 3.1 to the case \( k = 3 \) shows that the lower bound of \( k \) on the number of sign changes is not always tight. In particular, after some straightforward algebra, the conditions (18) on sign change positions reduce to the equation

\[
8t_3^2n^2 - 8t_3n^3 + n^4 = 0
\]

The first sign change position \( t_3 \) must be of the form \( \alpha n \), for some rational number \( \alpha \). The condition that \( \alpha \) must therefore satisfy is

\[
8\alpha^2 - 8\alpha + 1 = 0
\]

This quadratic equation has the two solutions

\[
\alpha = \frac{2 \pm \sqrt{2}}{4}
\]

neither of which is rational. It follows that there is no element of \( S(n, 3) \) with only three sign changes.

Knowing that the Morse sequence \( \mu(3) \in S(8, 3) \) has 5 sign changes, one might naturally inquire if 5 sign changes is the minimum number among the sequences with a null of order 3. If we assume that a sequence with 4 sign change positions \( n > t_1 > t_2 > \ldots > t_4 > 0 \) has a third-order null, and proceed as before, we obtain the following relations expressing \( t_1, t_2, \) and \( t_3 \) in terms of \( t_4 \):

\[
\begin{align*}
t_1 &= a + b + c \\
t_2 &= a + 2b \\
t_3 &= a + b - c
\end{align*}
\]

where

\[
\begin{align*}
a &= \frac{n + 2t_4}{2} \\
b &= \frac{2t_4^2}{n - 4t_4} \\
c &= \frac{(8(n - 4t_4)^3n + (4t_4)^4)^{\frac{1}{2}}}{8(n - 4t_4)}
\end{align*}
\]

The last expression translates into the condition

\[
32\beta^4 - 64\beta^3 + 48\beta^2 - 12\beta + 1 = 2\gamma^2
\]

where \( t_4 = \beta n, \beta \neq 1/4 \) and \( \gamma \) is a rational number. We conjecture that there is no such pair of rational numbers \( \beta, \gamma \), implying that 5 sign changes are necessary for a binary sequence with a third-order null. The conjecture is supported by the results of a search that tested all values of \( \beta = p/q \), with \( (p, q) = 1 \) and \( q \leq 2048 \).
The sums corresponding to degrees $s = 0, 1, \ldots, k$ have been shown to be zero, in view of (20) and (21). Since $f_{k,k+1} = 1/(k+1)$, we conclude

$$m_k(\varphi) = \frac{\text{sgn}(x_n)}{k+1} \left( n^{k+1} - 2 \sum_{j=0}^{i} (-1)^{j-1} t_j^{k+1} \right),$$

completing the induction step and the proof of the lemma. $\blacksquare$

For small values of the null order $k$, the conditions (18) and (19) on the sign-change positions may be used to determine elements $\varphi \in \mathcal{S}(n,k)$, as well as to find bounds on the first non-zero moment $m_k(\varphi)$ for such $\varphi$.

For example, when $k = 1$ we have

$$n - 2t_1 + \cdots + (-1)^{t_l} = 0$$

implying that $t_1 \geq n/2$. Consequently, for any $\varphi \in \mathcal{S}(n,1)$, we have the bound

$$|m_1(\varphi)| \leq \frac{1}{2} \left( n^2 - 2 \left( \frac{n}{2} \right)^2 \right) = \frac{n^2}{4} \quad (22)$$

We recall that a sequence with a $k$-th order null must have at least $k$ sign changes (cf. [15]). If we restrict attention to sequences in $\mathcal{S}(n,1)$ with precisely one sign change, the conditions of Lemma 3.1 produce the unique solution $t = \{n/2\}$, which attains both the lower bound on the number of sign changes and the upper bound (22) on $m_1(\varphi)$.

For $k = 2$, we may solve for a sequence $\varphi \in \mathcal{S}(n,2)$ having exactly two sign changes. The conditions are

$$n - 2t_1 + 2t_2 = 0$$

$$n^2 - 2t_1^2 + 2t_2^2 = 0$$

From these equations we easily derive $t = \{3n/4, n/4\}$, which leads to

$$m_2(\varphi) = \frac{n^3}{16}$$

if we assume $x_n = +1$. Note that if $\bar{\varphi} \in \mathcal{S}(n,2)$ has more than two sign changes, then $m_2(\bar{\varphi}) \leq m_2(\varphi)$. This follows from the observation that there must exist sign change positions $i < j$ such that $y_i = -1$ and $y_j = +1$. If we transpose the symbols $y_i$ and $y_i+1$, and then transpose the symbols $y_j$ and $y_j+1$, it is easily checked that the resulting sequence $\bar{\varphi}'$ is again in $\mathcal{S}(n,2)$ and $m_2(\bar{\varphi}') > m_2(\bar{\varphi})$. Clearly, this procedure terminates when $\varphi'$ has only two sign changes. Hence the solution $t = \{3n/4, n/4\}$ again achieves both the lower bound on the number of sign changes and the upper bound on $m_2(\varphi)$. Note that the two
Proof. We proceed by induction. For $k = 1$ we have
\[
m_0(\mathbf{x}) = \text{sgn}(x_n) \left[ n - 2t_1 + \cdots + (-1)^i2t_i \right]
\]
Moreover,
\[
m_1(\mathbf{x}) = \text{sgn}(x_n) \left[ f_1(n) - 2f_1(t_1) + \cdots + (-1)^i2f_i(t_i) \right]
\]
\[
= \text{sgn}(x_n) \left[ \frac{1}{2}(n^2 - 2t_1^2 + \cdots + (-1)^2t_1^2) + \frac{1}{2}(n - 2t_1 + \cdots + (-1)^i2t_i) \right]
\]
Therefore, when $\mathbf{x} \in S(n, 1)$ we have
\[
m_1(\mathbf{x}) = \frac{\text{sgn}(x_n)}{2} \left( n^2 - 2t_1^2 + \cdots + (-1)^i2t_i^2 \right)
\]
as desired.

Now assume that the lemma holds for $k - 1$. Thus $\mathbf{x} \in S(n, k - 1)$ if and only if
\[
n^{r+1} - 2 \sum_{j=1}^{i} (-1)^{j-1}t_j^{r+1} = 0 \quad \text{for } r = 0, 1, \ldots, k - 2 \quad (20)
\]
and furthermore for any such $\mathbf{x}$ we have
\[
m_{k-1}(\mathbf{x}) = \frac{\text{sgn}(x_n)}{k} \left( n^k - 2 \sum_{j=1}^{i} (-1)^{j-1}t_j^k \right) \quad (21)
\]
Clearly $\mathbf{x} \in S(n, k)$ iff $\mathbf{x} \in S(n, k - 1)$ and $m_{k-1}(\mathbf{x}) = 0$. By induction hypothesis the former condition is equivalent to (20), while the latter condition is equivalent to
\[
n^k - 2 \sum_{j=1}^{i} (-1)^{j-1}t_j^k = 0
\]
To complete the proof we write the polynomials $f_k(n)$ in the form
\[
f_k(n) = \sum_{s=0}^{k+1} f_{k,s}n^s
\]
and rewrite the moment $m_k(\mathbf{x})$ as
\[
m_k(\mathbf{x}) = \text{sgn}(x_n) \left( \sum_{s=0}^{k+1} f_{k,s}n^s - 2 \sum_{s=0}^{k+1} f_{k,s}t_1^s + \cdots + (-1)^i2 \sum_{s=0}^{k+1} f_{k,s}t_i^s \right).
\]
Grouping terms of equal degree, we find
\[
m_k(\mathbf{x}) = \text{sgn}(x_n) \sum_{s=0}^{k+1} f_{k,s} \left( n^s - 2 \sum_{j=0}^{i} (-1)^{j-1}t_j^s \right)
\]
3. On sign changes in spectral-null sequences

The characterizations of $S(n, k)$ in the previous section may be recast into a form involving only the positions $i$ where the component values in a vector $\underline{x} = (x_1, x_2, \ldots, x_n) \in S(n, k)$ change sign, that is $x_{i+1} = -x_i$. We shall denote these positions by the sign-change list $t = \{t_1, t_2, \ldots, t_l\}$, where $n > t_1 > t_2 > \ldots > t_l > 0$.

Let $f_k(n)$ denote the sum of $k$-th powers of consecutive integers,

$$f_k(n) = \sum_{j=1}^{n} j^k$$

It is well-known [16, p.499] that $f_k(n)$ is an integer polynomial of degree $k+1$. Specifically,

$$f_k(n) = \frac{n^{k+1}}{k+1} + \frac{n^k}{2} + B_1 \frac{kn^{k-1}}{2!} + B_2 \frac{k(k-1)(k-2)n^{k-3}}{4!} + \cdots$$

where $B_i$ is the $i$-th Bernoulli number, and the series terminates at the $n^2$ term if $k$ is odd, or the $n$ term if $k$ is even. For example, recalling that $B_1 = 1/6$ and $B_2 = -1/30$, (cf. [16, p.615]) we find:

$$f_1(n) = \frac{n^2}{2} + \frac{n}{2} = \frac{n(n+1)}{2}$$
$$f_2(n) = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} = \frac{n(n+1)(2n+1)}{6}$$
$$f_3(n) = \frac{n^4}{4} + \frac{n^3}{3} + \frac{n^2}{4} = \frac{n^2(n+1)^2}{4}$$
$$f_4(n) = \frac{n^5}{5} + \frac{n^4}{2} + \frac{n^3}{3} - \frac{n}{30} = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}$$

For a sequence $\underline{x}$ with sign-change positions $\{t_1, t_2, \ldots, t_l\}$, we can rewrite the moments $m_k(\underline{x}) = \sum_{j=1}^{n} j^k x_j$ in the form

$$m_k(\underline{x}) = \text{sgn}(x_n) \left( [f_k(n) - f_k(t_1)] - [f_k(t_1) - f_k(t_2)] + \cdots + (-1)^l f_k(t_l) \right)$$

for all $k \geq 0$. This clearly reduces to

$$m_k(\underline{x}) = \text{sgn}(x_n) \left( f_k(n) - 2f_k(t_1) + \cdots + (-1)^l 2f_k(t_l) \right)$$

When $\underline{x} \in S(n, k)$, the expression above translates the vanishing moment conditions of (4) into simple conditions on the sign-change positions, as shown in the following lemma.

**Lemma 3.1.** Let $\underline{x}$ be a sequence in $\Phi^n$ with sign-change list $t = \{t_1, t_2, \ldots, t_l\}$. Then $\underline{x} \in S(n, k)$ if and only if

$$n^{s+1} - 2 \sum_{j=1}^{l} (-1)^{j-1} t_j^{s+1} = 0 \quad \text{for } s = 0, 1, \ldots, k - 1. \quad (18)$$

Furthermore,

$$m_k(\underline{x}) = \frac{\text{sgn}(x_n)}{k+1} \left( n^{k+1} - 2 \sum_{j=1}^{l} (-1)^{j-1} t_j^{k+1} \right). \quad (19)$$
Both [14] and [15] make use of a well-known result from number theory, — the Prouhet-Tarry problem. Suppose that \( A = \{a_1, a_2, \ldots, a_s\} \) and \( B = \{b_1, b_2, \ldots, b_s\} \) are two disjoint sets of distinct positive integers and consider the system of \( k \) equations
\[
a_i^1 + a_i^2 + \cdots + a_i^s = b_i^1 + b_i^2 + \cdots + b_i^s \quad \text{for} \quad i = 0, 1, \ldots, k - 1 \quad (17)
\]

Then the Prouhet-Tarry problem asks for the least value of \( s \) for which (17) has a solution. We shall use \( P(k) \) to denote this value of \( s \). The proof of the lower bound on \( d(S(n, k)) \) in [15] is based on the lower bound \( P(k) \geq k \). Herein we employ the upper bound on \( P(k) \) to derive an upper bound on \( d(S(n, k)) \).

**Lemma 2.9.** (Prouhet-Tarry [10, p.329])

\[
P(k) \leq \frac{k(k - 1)}{2} + 1
\]

**Proof.** Fix \( k \) and let \( s = \frac{1}{2}k(k - 1) + 1 \). For sufficiently large integer \( N \) we have, \( V(N, s) > s^kN^{s-1} \) where the LHS of the inequality is the number of distinct subsets of \( \{1, 2, \ldots, N\} \) of size \( \leq s \) and the RHS is an upper bound on the number of distinct values of the vector \( (S_0, S_1, \ldots, S_{k-1}) \), where \( S_i = \sum_{a \in A} a^i \) and \( A \) ranges over all the subsets of \( \{1, 2, \ldots, N\} \) of size \( \leq s \). The lemma now follows by counting arguments. \( \blacksquare \)

**Theorem 2.10.** For any fixed \( k \) and any sufficiently large \( n \) that is divisible by \( 2^k \),
\[
d(S(n, k)) \leq 2P(k) \leq k(k - 1) + 2.
\]

**Proof.** Set \( s = P(k) \) and let \( A = \{a_1, a_2, \ldots, a_s\} \) and \( B = \{b_1, b_2, \ldots, b_s\} \) be the two solutions of (17) guaranteed by Lemma 2.9. Further, take \( N \) to be an integer greater than any of the elements in \( A \cup B \). Consider the vector \( \underline{y} = (y_1, y_2, \ldots, y_N) \in \Phi^N \), where \( y_a = -1 \) if \( a \in A \) and \( y_a = 1 \) otherwise. Let \( \underline{z} = (z_1, z_2, \ldots, z_N) \in \Phi^N \) be a similar vector with respect to the set \( B \). Clearly \( y_i = z_i \) for all the positions \( i \) that are not in \( A \cup B \), and therefore the distance between \( \underline{y} \) and \( \underline{z} \) is at most \( 2P(k) \leq k(k - 1) + 2 \). In view of Lemma 2.6, there exists a vector \( \underline{z} \in S(N2^k, k) \) which contains \( \underline{y} \) as its \( N \)-prefix. Replacing this prefix by \( \underline{z} \), we obtain another vector in \( S(N2^k, k) \) at distance \( k(k - 1) + 2 \) from \( \underline{z} \). Since \( N \) must be sufficiently large, but is otherwise arbitrary, the same argument applies to all sufficiently large values of \( n \) divisible by \( 2^k \). \( \blacksquare \)

**Remark.** It is known [11, p.567] that \( P(k) = k \) for all \( k \leq 10 \). Hence it follows from Theorem 2.10 that for \( k \leq 10 \) the minimum distance of \( S(n, k) \) is exactly \( 2k \) for infinitely many values of \( n \).
Finding the minimum of (13) subject to the constraint (14) yields

\[ |S(n, k)| \geq \frac{|S(h, k-1)|^2}{|A(h, k)|} \]  

Taking logarithms of both sides in (15) we obtain,

\[ \rho(S(n, k)) \leq 2 \rho(S(h, k-1)) + \log_2 |A(h, k)|. \] (16)

Obviously, \(|A(h, k)| \leq 1 + 2 \sum_{j=1}^{h} j^{k-1} \leq n^k\) whenever \(k \geq 2\). Substituting this upper bound on \(A(h, k)\) into (16), and writing \(n = q2^m\) for some \(m \geq k\), yields

\[ \rho(S(q2^m, k)) \leq 2 \rho(S(q2^{m-1}, k-1)) + k (m + \log_2 q). \]

Taking into account that \(\rho(S(n, 0)) = 0\) for all \(n\), we can solve the above recursion to show that,

\[ \rho(S(q2^m, k)) \leq \sum_{i=0}^{k-1} 2^i (k - i)(m - i + \log_2 q) \]

\[ = O\left((2^k - 1)(\log_2 q + m - k + 1)\right), \]

as claimed. \[\qed\]

**Remark.** Theorem 2.8 can be slightly improved by way of using better estimates for the size of \(A(n, k)\) and observing that the sizes of \(S(n, k; a)\) depend on \(a\). In particular, an improvement can be obtained by taking into account that \(k!\) must divide \(a\) for every \(a \in A(n, k)\), as will be shown in Section 5.2. However such arguments will not get rid of the \(2^k\) term in the bound of the Theorem 2.8, and are therefore omitted.

**Remark.** Referring to Table 1, it is clear that Theorem 2.8 does not cover the entire range of values of \(n\) and \(k\) for which \(S(n, k) \neq \emptyset\). For instance, taking \(n = 12\) and \(k = 3\) we see that \(2^k\) does not divide \(n\), and yet \(S(n, k) = S(12, 3) \neq \emptyset\).

### 2.5. Bounds on the minimum distance

It is shown in [14],[15] that the minimum distance of \(S(n, k)\) is bounded from below by \(2k\). In this subsection we present an upper bound on the minimum distance of \(S(n, k)\) for infinitely many values of \(n\), establishing equation (3).
The following theorem is essentially a sphere-packing upper bound on the cardinality of the set $S(n, k)$.

**Theorem 2.7.** For all $n \geq 1$,

$$\rho(S(n, k)) \geq (k - 1) \log_2 n - (k - 0.5) \log_2 k - 1.5 \geq (k - 1)(\log_2 n - k + 1).$$

**Proof.** It is shown in [14],[15] that the minimum distance of $S(n, k)$ is at least $2k$. Thus by the sphere-packing bound [19, Ch.1] we have

$$\log_2 |S(n, k)| \leq n - \log_2 V(n, k-1)$$

where $V(n, k) = \sum_{i=0}^{k} \binom{n}{i}$ denotes the volume of the Hamming sphere of radius $k$ in $\Phi^n$. The theorem now follows by the Stirling approximation [19, p.310].

The following theorem is a nonconstructive lower bound on the cardinality of $S(n, k)$, which implies in particular that $\text{cap}(S(n, k)) = 1$. In Section 5 we present a construction of spectral-null codes which attains the capacity. However, the existence result of Theorem 2.8 provides a much better bound on the redundancy of $S(n, k)$.

**Theorem 2.8.** For all $n \geq 1$ such that $2^k|n$,

$$\rho(S(n, k)) \leq O\left((2^k - 1) (\log_2 n - k + 1)\right).$$

**Proof.** The result is obviously true for $k = 0, 1$. Hence we hereafter assume that $k > 1$, in which case $n$ is even. Write $n = 2h$ and let $S(h, k; a)$ denote the set of all vectors $\underline{z}$ in $S(h, k-1)$, such that $m_{k-1}(\underline{z}) = a$ for some fixed integer $a$. Further, let $A(h, k)$ denote the set of all the integers $a$ for which $S(h, k; a)$ is nonempty.

For each $a \in A(h, k)$ define the set $B(n, k; a)$ as follows

$$B(n, k; a) \overset{\text{def}}{=} \left\{ (\underline{z}, \underline{y}) \in \Phi^n : \underline{z} \in S(h, k; a) \text{ and } \underline{y} \in S(h, k; -a) \right\}$$

It follows from Lemma 2.1 that $B(n, k; a) \subseteq S(n, k)$ for every $a \in A(h, k)$. Furthermore, $|B(n, k; a)| = |S(h, k; a)|^2$ since $|S(h, k; a)| = |S(h, k; -a)|$, and $B(n, k; a) \cap B(n, k; b) = \emptyset$ whenever $a \neq b$ since the sets $S(h, k; a)$ form a partition of $S(h, k-1)$. Hence we have,

$$|S(n, k)| \geq \sum_{a \in A(h, k)} |S(h, k; a)|^2 \tag{13}$$

and

$$\sum_{a \in A(h, k)} |S(h, k; a)| = S(h, k-1) | \tag{14}$$

12
Applying the construction of Lemma 2.6 to the vector \( \mathbf{z} = (+) \in \Phi^1 \) produces, for \( l \to \infty \), the infinite binary Morse sequence

\[
+ - - + - + - + - - + + - - + - - + - + - + - + - + - + \ldots
\]

which is well-known in symbolic dynamics [8],[21]. We shall denote by \( \mu(k) \) the truncation of this sequence to its first \( 2^k \) positions. Then it follows from Lemma 2.6 in conjunction with Table 1 that for all \( k \leq 5 \), the Morse sequence \( \mu(k) \) is the shortest spectral-null sequence of order \( k \). Furthermore, for \( k \leq 5 \) the two Morse sequences, \( \mu(k) \) and its complement, are the only elements in \( \mathcal{S}(2^k, k) \).

It is tempting to ask whether the two properties of the Morse sequence exhibited in Table 1 extend beyond \( k = 5 \). Thus,

**Question.** Is \( \mu(k) \) the shortest spectral-null sequence of order \( k \), for all \( k \)?

**Question.** Is it true that \( \mathcal{S}(2^k, k) = \{ \mu(k), \overline{\mu(k)} \} \) for all \( k \)?

While the first question remains open, the answer to the second question is, surprisingly, negative. For \( k = 6 \), the set \( \mathcal{S}(2^6, 6) \) contains sequences other than the Morse sequence, e.g., the word

\[
++- - - + - + - + - + + + - - + - - + - - + - + - + - +
\]

concatenated with its reflection. The construction of Lemma 2.6 may now be applied to this sequence to show that \( \mathcal{S}(2^k, k) \neq \{ \mu(k), \overline{\mu(k)} \} \) for all \( k \geq 6 \).

### 2.4. Bounds on redundancy

Let \( \mathcal{S}(k) = \bigcup_{n \geq 1} \mathcal{S}(n, k) \) denote the (infinite) set of finite words over \( \Phi \) with a \( k \)-th order spectral null at zero frequency. The set \( \mathcal{S}(k) \) may be thought of as the set of all sequences admitted by the binary-input spectral-null channel of order \( k \). The capacity of a spectral-null channel of order \( k \) is then defined by

\[
\operatorname{cap}(\mathcal{S}(k)) = \limsup_{n \to \infty} \frac{\log_2 |\mathcal{S}(n, k)|}{n}
\]

It was noted in [15], using arguments based on canonical finite-state transition diagrams, that the capacity of a \( k \)-th order spectral-null channel should be equal to unity for any fixed order \( k \). In this subsection we prove that indeed \( \operatorname{cap}(\mathcal{S}(k)) = 1 \). Furthermore, we provide upper and lower bounds on \( |\mathcal{S}(n, k)| \), or equivalently on \( \rho(\mathcal{S}(n, k)) \), establishing the stronger claim of equation (2).
Table 1: Redundancy of $S(n,k)$ for $n \leq 32$ with $4 \mid n$.

Several observations are imminent from Table 1. In particular, it readily follows from the table that the condition of Theorem 2.5 is not sufficient for $S(n,k)$ to be nonempty. For example, taking $n = 16$, $k = 5$, and $m = \lceil \log_2 k \rceil + 1 = 3$ we see that $2^m$ divides $n$, and yet $S(n,k) = S(16,5) = \emptyset$. On the other hand, it follows from the table that $S(4q,3) \neq \emptyset$ for $q = 2, 3$. Since $x_1 \in S(n_1,k)$ and $x_2 \in S(n_2,k)$ can always be concatenated to produce $(x_1 | x_2) \in S(n_1 + n_2, k)$, we deduce from Table 1 that $S(4q,3) \neq \emptyset$ for $q \geq 2$ and $S(4q,2) \neq \emptyset$ for $q \geq 1$. Furthermore, it can be verified by computer search that $S(8q,5) \neq \emptyset$ for $q = 4, 5, 6, 7$. Hence, $S(8q,5) \neq \emptyset$ for $q \geq 4$ and $S(8q,4) \neq \emptyset$ for $q \geq 2$.

We also point out that for $k \geq 1$ and $n \leq 32$, the minimum distance of a nonempty set $S(n,k)$ equals $2k$, except when the redundancy is $n - 1$. In the latter case the two words in $S(n,k)$ are complements of each other and, therefore, the minimum distance is $n$.

The following two facts particularly stand out in Table 1. For all $k \leq 5$:

- The smallest integer $n$ for which $S(n,k) \neq \emptyset$ is $n = 2^k$.
- The set $S(2^k,k)$ contains exactly 2 sequences.

We now show that these two sequences are (the truncations of) the binary Morse sequence (cf. [8],[21]) and its complement. The following lemma is slightly more general.

**Lemma 2.6.** For any $k \geq 0$ and any vector $x \in \Phi^n$ there exists a vector $y \in S(n2^k, k)$ containing $x$ as its prefix.

*Proof.* Set $x_1 = (x | -x)$, where $-x$ is the negation (or complement) of $x$. Obviously, $x_1 \in S(2n,1)$. For $l \geq 1$ define recursively $x_{l+1} = (x_{l} | -x_l)$. Evidently $x_l \in \Phi^{2^l}$. Furthermore, Lemma 2.1 implies (by induction on $l$) that $x_l \in S(n2^l,l)$. The lemma now follows by taking $y = x_k$. \[\square\]
By assumption, \(2^{t-1} \leq 2^{m-1} \leq k\). Thus, substituting \(2^{t-1} - 1\) for \(i\) in the foregoing expression, and subsequently applying the Lucas equality (Lemma 2.4), we obtain

\[
\sum_{j=1}^{n} \binom{2^{t-1} - 1}{j} \equiv \prod_{i=0}^{t-2} \binom{j_i}{1} \equiv \prod_{i=1}^{n} \binom{j_i}{1} \equiv 0 \pmod{2}
\]

where \(j = \sum_{i=0}^{t-1} j_i 2^i\) is the binary representation of \(j\). Evidently the parenthesized product in (11) is nonzero if and only if

\[
j_0 = j_1 = \cdots = j_{t-2} = 1
\]

It therefore follows from (11) that the total number of integers \(j\) between 1 and \(n\) satisfying (12) must be even. Let \(1 < \alpha_1 < \alpha_2 < \cdots < \alpha_{2h} < n\) denote all such integers, and let \(\alpha_i = \sum_{i=0}^{t-1} a_i 2^i\) be the binary representation of \(\alpha_i\). Then clearly \(a_i, i-1 = 1\) iff \(i\) is even, and in particular \(a_{2h, i-1} = 1\). Yet by induction hypothesis \(a_{2h} = n-1\). This implies that \(n_{t-1} = 0\), and the theorem is thereby proved. \(\blacksquare\)

Let \(M(k)\) denote the largest integer \(M\) such that \(2^M\) divides the length of any (nonempty) spectral-null code of order \(k\). Then Theorem 2.5 provides a lower bound on \(M(k)\),

\[
M(k) \geq \lceil \log_2 k \rceil + 1
\]

We shall see in the next subsection that this bound is tight for \(k = 1, 2, 3, 4, 5\). Whether this bound is tight in general remains an open question. Equivalently,

**Question.** Is it true that \(S(2^m q, 2^m - 1) \neq \emptyset\) for any fixed \(m\) and sufficiently large odd \(q\)?

Note that it would suffice to exhibit the existence of \(S(2^m q, 2^m - 1)\) for one particular odd \(q\), for any given \(m\). Since \(S(2^{2m-1}, 2^m - 1) \neq \emptyset\), as will be shown in the sequel, by the conductor theorem of Frobenius (cf. [23, p.376]) we would have \(S(2^m q, 2^m - 1) \neq \emptyset\) for infinitely many sufficiently large odd \(q\).

### 2.3. Spectral-null codes of short length and Morse sequences

The cardinality of the set \(S(n, k)\) for small values of \(n\) may be easily calculated, using for instance the generating function of [14] or direct enumeration. Table 1 below lists the values of \(\rho(S(n, k))\) for lengths \(n \leq 32\) that are divisible by 4. Empty entries in the table correspond to empty sets \(S(n, k)\). A similar table which lists the cardinality of \(S(n, k)\) for \(n \leq 32\) may be found in [14].
2.2. A constraint on the length of $S(n,k)$

It is obvious that $S(n,1) \neq \emptyset$ only if $n$ is even, and it is well-known [14] that $S(n,2) \neq \emptyset$ only if $n$ is divisible by 4. How does this constraint on the length of a (nonempty) spectral-null code of order $k$ extend to values of $k$ greater than two? In particular is it true that as the order $k$ of the null increases, the length of a spectral-null code of order $k$ must be divisible by increasing powers of 2? In this subsection we settle the latter question affirmatively.

First we need the following lemma.

**Lemma 2.4.** (Lucas equality [18]) Let $p$ be a prime and let $a, b$ be nonnegative integers whose $p$-ary representations are given by

$$a = a_{m-1}p^{m-1} + \cdots + a_1p + a_0 = \sum_{i=0}^{m-1} a_ip^i$$

$$b = b_{m-1}p^{m-1} + \cdots + b_1p + b_0 = \sum_{i=0}^{m-1} b_ip^i$$

where $a_i, b_i \in \{0,1,\ldots, p-1\}$ for all $i$. Then

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a_{m-1} \\ b_{m-1} \end{pmatrix} \cdots \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} \pmod{p}$$

In particular for $p = 2$ we have

$$\begin{pmatrix} a \\ b \end{pmatrix} \equiv \prod_{i=0}^{m-1} \begin{pmatrix} a_i \\ b_i \end{pmatrix} \pmod{2}$$

where $a_i, b_i \in \{0,1\}$. Thus $\left(\frac{a}{b}\right)$ is odd if and only if $b_i \leq a_i$ for all $i = 0,1,\ldots m-1$.

**Theorem 2.5.** The set $S(n,k)$ is nonempty only if $n$ is divisible by $2^m$ for $m \geq \lfloor \log_2 k \rfloor + 1$.

**Proof.** Let $S(n,k) \neq \emptyset$. We shall prove by induction on $t = 0,1,\ldots m$ that $2^t \mid n$. Clearly, this holds for $t = 0$. As induction hypothesis assume that $1 \leq t \leq m$ and that $2^{t-1} \mid n$. Let $n = \sum_{i=0}^{t-1} n_i 2^i$ be the binary representation of $n$. By assumption we have $n_0 = n_1 = \cdots = n_{t-2} = 0$, and it would suffice to show that $n_{t-1} = 0$ as well. Let $\underline{x} = (x_1, x_2, \ldots x_n)$ be a vector in $S(n,k)$. Then by Lemma 2.2 we have $\sum_{j=1}^{n} \left(\begin{pmatrix} j \\ i \end{pmatrix} x_j \equiv 0 \right)$ for $i = 0,1,\ldots k-1$. Since $x_j \equiv 1 \pmod{2}$ for all $j$, evaluating the above modulo 2 yields

$$\sum_{j=1}^{n} \left(\begin{pmatrix} j \\ i \end{pmatrix} \equiv 0 \pmod{2} \right) \quad \text{for} \quad i = 0,1,\ldots k-1$$
The matrix \( D(n,k) \) is a "systematic" parity-check matrix for \( S(n,k) \), which allows to express the first \( k \) positions of any vector \( \mathbf{z} \in S(n,k) \) as a function of the last \( n-k \) positions. In fact, since any \( k \times k \) submatrix of \( H(n,k) \) is a nonsingular Vandermonde matrix, any \( k \) positions in \( \mathbf{z} \) may be expressed in terms of the remaining \( n-k \) positions in a manner similar to that of Lemma 2.3. Furthermore, we have for \( j \geq k \)

\[
d_{i,j} = (-1)^{k-i-1} \binom{j}{i} \binom{j-i-1}{k-i-1} = (-1)^{k-i-1} \binom{k}{i} \binom{j}{j-i} \frac{k-i}{j-i}
\]

Thus the first \( k \) columns of \( D(n,k) \) form the identity matrix, while the last \( n-k \) columns of \( D(n,k) \) form a generalized Cauchy matrix (cf. [22]) with integer entries. All these properties of spectral-null codes resemble to a certain extent the properties of Reed-Solomon codes [19, Ch.11].

Lemma 2.3 will be employed in Section 5.2 to show that the \( k \)-th moment of every \( \mathbf{z} \) in \( S(n,k) \) is divisible by \( k! \). This property is crucial for the construction of high-order spectral-null codes presented in Section 5.
Then the foregoing lemma will be employed in the next subsection to show that the length of degree $j$ for $p_i$ applies to the polynomial $f_i(X) = X^i$. Substituting $f_i(X) = X^i$ into (8) and evaluating at $X = 1, 2, \ldots n$ we obtain

$$j^i = b_{i,0}(j)_0 + b_{i,1}(j)_1 + b_{i,2}(j)_2 + \cdots + b_{i,i}(j)_i$$

for $j = 1, 2, \ldots n$, where $b_{i,0}, b_{i,1}, b_{i,2}, \ldots b_{i,i}$ are constants and $b_{i,i} \neq 0$. It follows that the $i$-th row of $H(n,k)$ is a linear combination of the first $i$ rows of $Q(n,k)$. Hence, $H(n,k) = LQ(n,k)$ for some $k \times k$ lower triangular matrix $L$. Evidently,

$$\det L = \prod_{i=0}^{k-1} b_{i,i} \neq 0$$

since $b_{i,i} \neq 0$ for all $i$. This implies that the matrix $L$ is nonsingular and therefore the null spaces of $H(n,k)$ and $Q(n,k)$ are equal. $\blacksquare$

The foregoing lemma will be employed in the next subsection to show that the length of a (nonempty) spectral-null code of order $k$ must be divisible by $2^m$ for $m \geq \lfloor \log_2 k \rfloor + 1$.

**Lemma 2.3.** Let

$$D(n,k) = \begin{bmatrix}
1 & 0 & \cdots & 0 & (-1)^{k-1}(k)_0 & (-1)^{k-1}(k+1)_0 & \cdots & (-1)^{k-1}(n-1)_0 & (n-2)_{k-1} \\
0 & 1 & \cdots & 0 & (-1)^{k-2}(k)_1 & (-1)^{k-2}(k+1)_1 & \cdots & (-1)^{k-2}(n-1)_1 & (n-3)_{k-2} \\
0 & 0 & \cdots & 0 & (-1)^{k-3}(k)_2 & (-1)^{k-3}(k+1)_2 & \cdots & (-1)^{k-3}(n-1)_2 & (n-4)_{k-3} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & (-1)^0(k)_{k-1} & (-1)^0(k+1)_{k-1} & \cdots & (-1)^0(n-1)_{k-1} & (n-k-1)_{0}
\end{bmatrix}$$

Then

$$S(n,k) = \left\{ \mathbf{z} \in \Phi^n : D(n,k)\mathbf{z}^i = \mathbf{0} \right\} = \left\{ \mathbf{z} \in \Phi^n : (-1)^{k-i-1} \sum_{j=k}^{n-1} \binom{j}{i} \binom{j-i-1}{k-i-1} x_{j+1} = x_{i+1} \text{ for } i = 0, 1, \ldots k-1 \right\}$$
Lemma 2.1. For any constant $c \in \mathbb{R}$, let

$$H(n, k; c) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 + c & 2 + c & \cdots & n + c \\ (1 + c)^2 & (2 + c)^2 & \cdots & (n + c)^2 \\ \vdots & \vdots & \ddots & \vdots \\ (1 + c)^{k-1} & (2 + c)^{k-1} & \cdots & (n + c)^{k-1} \end{bmatrix}$$

Then

$$S(n, k) = \left\{ \mathbf{x} \in \Phi^n : H(n, k; c) \mathbf{x}^i = \mathbf{0} \right\} = \left\{ \mathbf{x} \in \Phi^n : \sum_{j=1}^{n} (j + c)^i x_j = 0 \text{ for } i = 0, 1, \ldots, k-1 \right\}$$

**Proof.** Using binomial coefficients, write $(j + c)^i = \sum_{l=0}^{i} \binom{i}{l} c^i j^{i-l} = j^i + ic j^{i-1} + \cdots + ic^{i-1} j + c^i$. This shows that the $i$-th row of $H(n, k; c)$ is a linear combination of the first $i$ rows of $H(n, k)$ with real coefficients (the coefficient of the $i$-th row being 1). Hence there exists a lower triangular $k \times k$ matrix $L$, having 1’s on its main diagonal, such that $H(n, k; c) = LH(n, k)$. Since $L$ is obviously nonsingular, the null spaces of $H(n, k; c)$ and $H(n, k)$ must be equal. The lemma now follows by (6). □

The foregoing lemma shows that the spectral properties of a vector $\mathbf{x} \in \Phi^n$ are position (or time, or shift) invariant. That is a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a $k$-th order spectral-null sequence if and only if any shifted version of $\mathbf{x}$, say $(x_1+c, x_2+c, \ldots, x_n+c)$, is also a $k$-th order spectral-null sequence. This property will be of importance in Section 2.4 and also in Sections 4 and 5.

Lemma 2.2. Let

$$Q(n, k) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \binom{1}{1} & \binom{2}{1} & \cdots & \binom{n}{1} \\ \binom{1}{2} & \binom{2}{2} & \cdots & \binom{n}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \binom{1}{k-1} & \binom{2}{k-1} & \cdots & \binom{n}{k-1} \end{bmatrix}$$

Then

$$S(n, k) = \left\{ \mathbf{x} \in \Phi^n : Q(n, k) \mathbf{x}^i = \mathbf{0} \right\} = \left\{ \mathbf{x} \in \Phi^n : \sum_{j=1}^{n} \binom{j}{i} x_j = 0 \text{ for } i = 0, 1, \ldots, k-1 \right\}$$
Several necessary and sufficient conditions for a vector \( \mathbf{x} \) to be a \( k \)-th order spectral-null sequence are given in [5],[14],[15],[20], among other works. In particular it is shown in [14],[15] that \( \mathbf{x} \in \Phi^n \) is in \( S(n,k) \) if and only if all the moments of \( \mathbf{x} \) up to order \( k-1 \) vanish. For an integer \( k \geq 0 \) and any real vector \( \mathbf{x} = (x_1, x_2, \ldots x_n) \) the \( k \)-th order moment of \( \mathbf{x} \) is defined by \( m_k(\mathbf{x}) \overset{\text{def}}{=} \sum_{j=1}^{n} j^k x_j \). Thus by [14],[15],

\[
S(n,k) = \left\{ \mathbf{x} \in \Phi^n : m_i(\mathbf{x}) = \sum_{j=1}^{n} j^i x_j = 0 \text{ for } i = 0, 1, \ldots k-1 \right\}
\]  

(4)

The foregoing definition of \( S(n,k) \) may be restated in terms of the "parity-check" matrix

\[
H(n,k) \overset{\text{def}}{=} \left[ j^i \right]_{i=0, j=1}^{k-1, n} = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
1 & 2 & \cdots & n \\
1^2 & 2^2 & \cdots & n^2 \\
\vdots & \vdots & \ddots & \vdots \\
1^{k-1} & 2^{k-1} & \cdots & n^{k-1}
\end{bmatrix}
\]

(5)

Evidently, equation (4) is equivalent to

\[
S(n,k) = \left\{ \mathbf{x} \in \Phi^n : H(n,k) \mathbf{x}^t = 0 \right\}
\]

Let \( V(n,k) \) be the null space of \( H(n,k) \), that is the vector space over the real field \( \mathbb{R} \) of dimension \( n-k \) consisting of all vectors \( \mathbf{y} \in \mathbb{R}^n \) satisfying \( H(n,k) \mathbf{y}^t = 0 \). Then, obviously, \( S(n,k) = V(n,k) \cap \Phi^n \). It therefore follows that if \( M \) is any \( k \times n \) matrix with entries from \( \mathbb{R} \) such that the null space of \( M \) is equal to that of \( H(n,k) \), then

\[
S(n,k) = \left\{ \mathbf{x} \in \Phi^n : M\mathbf{x}^t = 0 \right\}
\]

(6)

We now specifically indicate three such matrices:

\[
H(n,k;c) \overset{\text{def}}{=} \left[ (j+c)^i \right]_{i=0, j=1}^{k-1, n} \\
Q(n,k) \overset{\text{def}}{=} \left[ \left( \frac{j}{i} \right) \right]_{i=0, j=1}^{k-1, n} \\
D(n,k) \overset{\text{def}}{=} \left[ d_{i,j} \right]_{i=0, j=0}^{k-1, n-1}
\]

where

\[
d_{i,j} = \begin{cases}
1 & \text{if } j = i \\
0 & \text{if } j < k \text{ and } j \neq i \\
(-1)^{k-i-1} \left( \frac{j}{i} \right) \left( \frac{j-i}{i-i} \right) & \text{if } j \geq k
\end{cases}
\]

(7)

Substituting these matrices into (6) leads to three equivalent characterizations of \( S(n,k) \) which will prove to be useful in the sequel.
The remaining two sections are devoted to constructions of block spectral-null codes of length \( n \) and order \( k \), for increasing values of \( k \). Using enumerative encoding [12],[13] it is fairly easy to encode an arbitrary binary sequence of length \( m = n - 0.5 \log_2 n - O(1) \), regarded as an \( m \)-bit representation of an integer \( N \), into a vector \( \mathbf{x}_N \in \mathcal{S}(n,1) \) indexed by \( N \) according to the standard lexicographic order on \( \mathcal{S}(n,1) \). In [17] Knuth described a simpler encoding method into a subset of \( \mathcal{S}(n,1) \) whose redundancy is about twice the redundancy of \( \mathcal{S}(n,1) \). See also [1],[2],[9]. In Section 4 we present an algorithm for encoding arbitrary sequences into a subset of \( \mathcal{S}(n,2) \), which is in some sense a generalization of the encoding method of Knuth [17]. The redundancy of the resulting second-order spectral-null code is bounded from above by \( 3 \log_2 n + O(\log \log n) \).

In Section 5 we present a general encoding scheme into a subset of \( \mathcal{S}(n,k) \) for any fixed order \( k \). First we describe in Section 5.1 an alternative algorithm for encoding arbitrary sequences into a subset of \( \mathcal{S}(n,2) \). The redundancy of the resulting second-order spectral-null codes is substantially greater than \( 3 \log_2 n + O(\log \log n) \). Hence these codes are inferior to the second-order spectral-null codes introduced in Section 4. Nevertheless, the rate of these codes still approaches unity as \( n \to \infty \). Furthermore, unlike the construction of Section 4, the construction of Section 5.1 lends itself to generalization for values of \( k \) greater than 2. One of the key ingredients required for such generalization is the existence of an algorithm which, given a certain vector \( \mathbf{x} \in \mathcal{S}(n_1,k) \), produces a vector \( \mathbf{y} \) such that \( (\mathbf{x},\mathbf{y}) \in \mathcal{S}(n_2,k+1) \) for some \( n_2 > n_1 \), where \( (\cdot,\cdot) \) denotes concatenation. Such an algorithm is derived in Section 5.2. Finally, in Section 5.3 we employ this algorithm to describe a recursive encoding scheme for spectral-null codes of length \( n \) and any fixed order \( k \). It is also shown in Section 5.3 that these codes are asymptotically optimal, that is their rate approaches unity as \( n \to \infty \).

### 2. Bounds on the parameters of \( \mathcal{S}(n,k) \)

In this section we derive several equivalent presentations of the set \( \mathcal{S}(n,k) \). Using one of these presentations we show that \( \mathcal{S}(n,k) \) is nonempty only if the length \( n \) satisfies a certain constraint. A table of \( \rho(\mathcal{S}(n,k)) \) is provided for all admissible \( n \leq 32 \) and \( k \leq 5 \). The rest of the section is devoted to our main results herein — namely, upper and lower bounds on the cardinality and minimum distance of \( \mathcal{S}(n,k) \).
This work has two main objectives. The first is to study the properties of the set $S(n, k)$, and in particular derive upper and lower bounds on its cardinality and minimum distance. The second is to provide constructions of block codes that are subsets of $S(n, k)$ with rate approaching unity as $n \to \infty$.

The case $k = 0$ corresponds to unconstrained sequences and is therefore trivial. Thus $S(n, 0) = \Phi^n$ with $\rho(S(n, 0)) = 0$ and $d(S(n, 0)) = 1$. The set $S(n, 1)$ consists of all balanced, or DC-free, vectors with (cf. [13],[17])

$$\rho(S(n, 1)) = 0.5 \log_2 n + O(1)$$
$$d(S(n, 1)) = 2$$

for all even $n$. Indeed $S(n, 1)$ is empty if $n$ is odd.

In general, no explicit expressions are presently known for the redundancy and minimum distance of $S(n, k)$ for $k \geq 2$. However, in the next section we shall derive bounds on these parameters. We start in Section 2.1 with several equivalent presentations of the set $S(n, k)$, which will turn out to be instrumental in the sequel. Using one of these presentations in conjunction with the Lucas equality [18], we show in Section 2.2 that $S(n, k) \neq \emptyset$ only if $n$ is divisible by $2^m$ for $m = [\log_2 k] + 1$. Next, we present in Section 2.3 a table which gives the redundancy of $S(n, k)$ for several small values of $n$ and $k$. This table exhibits an interesting relationship between spectral-null codes and the so-called Morse sequences (cf. [8],[21]). In Section 2.4 we derive lower and upper bounds on the cardinality of $S(n, k)$, showing that

$$(k - 1) (\log_2 n - k + 1) \leq \rho(S(n, k)) \leq O\left((2^k - 1) (\log_2 n - k + 1)\right) \quad (2)$$

for all $n$ divisible by $2^k$. Finally, in Section 2.5 we employ a well-known result from number theory (the Prouhet-Tarry problem cf. [10],[11]) to show that the minimum distance of $S(n, k)$ is bounded by

$$2k \leq d(S(n, k)) \leq k(k - 1) + 2 \quad (3)$$

for all sufficiently large $n$ divisible by $2^k$.

In Section 3 we exhibit yet another presentation of the set $S(n, k)$ in terms of the positions of sign changes in every vector $x \in S(n, k)$. An interesting feature of this presentation is that it characterizes $S(n, k)$ as the set of all integer solutions of a certain system of Diophantine equations, without the additional constraint that these solutions belong to the binary alphabet $\Phi = \{+1, -1\}$, which is usually implicit in all other definitions. Using this characterization of $S(n, k)$ we show that the lower bound of $k$ on the number of sign changes in a $k$-th order spectral-null sequence, given in [15], is not tight.
1. Introduction

Let $\Phi$ denote the bipolar binary alphabet $\{+1,-1\}$. We shall often use $+$ and $-$ as a shorthand notation for $+1$ and $-1$. For a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \Phi^n$, the discrete Fourier transform of $\mathbf{x}$ is given by $X(\omega) = \sum_{i=1}^{n} x_i e^{-j\omega i}$. Assume that $\mathbf{x}$ is such that

$$\frac{d^i X(\omega)}{d\omega^i} \bigg|_{\omega=0} = 0 \quad \text{for} \quad i = 0, 1, \ldots, k - 1$$

and let $C$ be a subset of $\Phi^n$ consisting of vectors satisfying (1). If codewords of $C$ are randomly concatenated to form an infinite sequence, then the power spectral density $F(\omega)$ of the concatenated sequence satisfies (cf. [14],[15])

$$\frac{d^i F(\omega)}{d\omega^i} \bigg|_{\omega=0} = 0 \quad \text{for} \quad i = 0, 1, \ldots, 2k - 1$$

The set $C$ is therefore said to be a spectral-null code of order $k$. The individual vector $\mathbf{x}$ satisfying (1) is said to be a $k$-th order spectral-null sequence. We let $S(n,k) \subseteq \Phi^n$ denote the set of all $k$-th order spectral-null sequences of length $n$. Any subset of $S(n,k)$ is a spectral-null code $C$ of length $n$ and order $k$. We shall refer to $\rho(C) = n - \log_2 |C|$ as the redundancy of $C$. The minimum distance $d(C)$ of $C$ is the minimum Hamming distance between any two distinct words in $C$.

Codes consisting of words with prescribed spectral-null properties have been extensively studied over the years. For instance, there is a vast body of literature on the first-order spectral-null codes, commonly known as DC-free codes, see for example [1],[2],[3],[4],[6],[7],[9],[13],[17],[24]. High-order spectral-null codes, that is subsets of $S(n,k)$ for $k > 1$, have been recently considered in several works [13],[14],[5],[15] for various applications. In particular, high-order spectral-null codes have been found useful for achieving a better rejection of the low-frequency components than is possible with the conventional DC-free codes [13],[14], and for enhancing the error-correction capability of codes used in partial-response channels [15],[5].

The problem of analyzing and synthesizing high-order spectral-null codes, has been dealt with in a number of papers recently [12],[14],[15],[20]. Some of the constructions [15],[20] are based on approaching the set $S(n,k)$, when $n$ goes to infinity, by sets of sequences generated by all possible walks on certain labeled directed graphs. However, as the order of the spectral null increases, these graphs quickly become prohibitively complex. An alternative enumerative encoding scheme for $S(n,2)$ was proposed in [12]. Still, there is no general construction of block codes that are subsets of $S(n,k)$ 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 have yet to be explored.
High-Order Spectral-Null Codes: Constructions and Bounds

Ron M. Roth
Computer Science Department
Technion — Israel Institute of Technology
Haifa 32000, Israel

Paul H. Siegel
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120

Alexander Vardy*
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120

Abstract

Let \( S(n,k) \) denote the set of all sequences of length \( n \) over the alphabet \( \{+1,-1\} \), having a \( k \)-th order spectral-null at zero frequency. A subset of \( S(n,k) \) is a spectral-null code of length \( n \) and order \( k \). Upper and lower bounds on the cardinality of \( S(n,k) \) are derived. In particular we prove that \( O(2^k \log_2 n) \geq n - \log_2 |S(n,k)| \geq O(k \log_2 n) \) for infinitely many values of \( n \). On the other hand we show that if \( n \) is not divisible by \( 2^m \) for \( m = \lfloor \log_2 k \rfloor + 1 \), then \( S(n,k) \) is empty. Furthermore, bounds on the minimum Hamming distance \( d \) of \( S(n,k) \) are provided, showing that \( 2k \leq d \leq k(k-1)+2 \) for infinitely many \( n \). We also investigate the minimum number of sign changes in a sequence \( x \in S(n,k) \) and provide an equivalent definition of \( S(n,k) \) in terms of the positions of these sign changes. An efficient algorithm for encoding arbitrary information sequences into a second-order spectral-null code of redundancy \( 3 \log_2 n + O(\log \log n) \) is presented. Furthermore, we prove that the first nonzero moment of any sequence in \( S(n,k) \) is divisible by \( k! \) and then show how to construct a sequence with a spectral null of order \( k \) whose first nonzero moment is any even multiple of \( k! \). This leads to an encoding scheme for spectral-null codes of length \( n \) and any fixed order \( k \), with rate approaching unity as \( n \to \infty \).

Keywords: input-constrained channels, spectral-null codes, spectral encoders.

*Research supported in part by the Rothschild Fellowship administered by the Rothschild Foundation.