Minimum Pearson Distance Detection in the Presence of Unknown Slowly Varying Offset

Vitaly Skachek and Kees Schouhamer Immink

***Abstract*—Minimum Pearson Distance (MPD) detection offers resilience against unknown channel gain and varying offset. MPD detection is used in conjunction with a set, S, of 2-ary codewords having specific properties. In this work, we study the properties of the codewords of S, compute the size of S, and derive its redundancy for asymptotically large values of the codeword length n. The redundancy of S is approximately $\frac{3}{2}\log_2 n + \alpha$, where $\alpha = \log_2 \sqrt{\pi/24} = -1.467..$ for n odd and $\alpha = -0.467..$ for n even.**

***Index Terms*—Constant composition code, permutation code, flash memory, optical recording, minimum Pearson distance detection**

I. Introduction

In solid-state memories, such as flash storage, the signals are stored as charges in memory cells. Over time, the charge may partially erode, and, as a result, the retrieved data signal is superimposed on an unknown-time varying drift [1]. In digital optical recording, fingerprints on the surface of a disc result in rapidly offset and gain variations of the retrieved signal. Memory cells closer to hotter areas on the chip may lose their charge faster than cells closer to colder areas.

Various methods have been proposed to combat the detrimental effects of unknown channel parameters. A popular approach is the insertion of *training sequences,* where a known pattern of symbols is embedded in the data stream. Since the channel is time-varying and corrupted by noise or drop-outs, the parameter estimation will, by necessity, be based on an average over a given time-interval, and the estimated values will thus lag behind the actual values of the varying parameters. As a result, the error performance of the detection circuitry will degrade. A more frequent insertion of training sequences may improve the parameter estimation which, however, comes at the cost of decreased payload. It is therefore needed to find alternative methods.

Prior art *Minimum Pearson Distance* (MPD) detection has been advocated since it has innate resistance, or is said to be *immune*, to unknown signal amplitude and offset of the received signal [2], [3]. It is assumed in this prior art that the offset is constant (uniform) for all symbols in the codeword. The main idea is generalized and it is assumed that the offset varies linearly over the codeword symbols. We show that MPD detection in conjunction with a dedicated coding step will result in a detection performance that is immune to slowly varying offsets. We start with some terminology.

We consider a communication codebook, S, of chosen codewords $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ over the alphabet $\mathcal{Q} = \{0, 1\}$, where n, the *length* of $\boldsymbol{x}$, is a positive integer. We assume that the sent codeword, $\boldsymbol{x}$, is received as

$$\boldsymbol{r} = a(\boldsymbol{x} + \boldsymbol{\nu}) + b\mathbf{1} + c\boldsymbol{s}, \tag{1}$$

where $\boldsymbol{r} = (r_1, \ldots, r_n)$, $r_i \in \mathbb{R}$, $\mathbf{1} = (1, \ldots, 1)$, and $\boldsymbol{s} = (1, 2, \ldots, n)$. The basic premises are that $\boldsymbol{x}$ is received with an unknown (positive) attenuation (gain) a, $a > 0$, is offsetted by an unknown varying offset, $b\mathbf{1} + c\boldsymbol{s}$, where a, b, and $c \in \mathbb{R}$, and corrupted by additive Gaussian noise $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)$, $\nu_i \in \mathbb{R}$ are noise samples with distribution $N(0, \sigma^2)$, where $\sigma^2 \in \mathbb{R}$ denotes the variance of the additive noise. The difference between the above and the prior art model in [2] is the term $c\boldsymbol{s}$, which denotes an additional linearly increasing or decreasing offset, whose slope magnitude c is unknown to the receiver. The term, $c\boldsymbol{s}$, makes the problem under consideration significantly more difficult than its prior art counterpart in [2].

We start in Section II with a brief description of the prior art. Section III presents the backbone of the paper, where it is shown how an MPD detector can be used in conjunction with codewords taken from a dedicated codebook, S, in such a way that the detector is immune to gain mismatch and varying offset. Properties of S are collected in Section IV. In Sections V and VI, we count the number of constrained codewords, and show that for asymptotically large n, the redundancy of S is approximately $\frac{3}{2}\log_2 n + \alpha$, where $\alpha = \log_2 \sqrt{\pi/24} = -1.467..$ for n odd and $\alpha = -0.467..$ for n even. The Conclusions section concludes this paper.

II. Prior Art

The well-known (squared) Euclidean distance between the received signal vector $\boldsymbol{r}$ and the codeword $\hat{\boldsymbol{x}}$ is defined by

$$\delta_{\mathrm{e}}(\boldsymbol{r}, \hat{\boldsymbol{x}}) = \sum_{i=1}^{n} (r_i - \hat{x}_i)^2. \tag{2}$$

This work is supported in part by the grants PUT405 and IUT2-1 from the Estonian Research Council.

V. Skachek is with the Institute of Computer Science, University of Tartu, J. Liivi 2-216, Tartu 50409, Estonia. E-mail: vitaly.skachek@ut.ee.

K. A. Schouhamer Immink is with Turing Machines Inc, Willemskade 15b-d, 3016 DK Rotterdam, The Netherlands. E-mail: immink@turing-machines.com.

978-1-5090-1805-5/16/$31.00 ©2016 IEEE

A minimum Euclidean distance detector outputs the codeword

$$\boldsymbol{x}_{\rm o} = \arg\min_{\hat{\boldsymbol{x}}\in S} \delta_{\rm e}(\boldsymbol{r}, \hat{\boldsymbol{x}}). \tag{3}$$

Working out (2) gives

$$\begin{aligned}\delta_{\rm e}(\boldsymbol{r}, \hat{\boldsymbol{x}}) &= \sum_{i=1}^{n}(x'_i - \hat{x}_i + b + ci)^2 \\ &= \sum_{i=1}^{n}(x'_i - \hat{x}_i)^2 + (b+ci)^2 \\ &+ 2b\sum_{i=1}^{n} x'_i + 2c\sum_{i=1}^{n} ix'_i \\ &- 2b\sum_{i=1}^{n} \hat{x}_i - 2c\sum_{i=1}^{n} i\hat{x}_i, \end{aligned} \tag{4}$$

where $x'_i = a(x_i + \nu_i)$. The dependence of (4) on the unknown terms has a significant bearing on the error performance. In the prior art, constrained coding methods have been presented that offer a cure for the reported error performance degradation. For example, in [4], constrained codes are advocated, where the codebook S is chosen such that each codeword $\boldsymbol{x} \in S$ satisfies two conditions, namely

$$\sum_{i=1}^{n} x_i = \frac{n}{2} \tag{5}$$

and

$$\sum_{i=1}^{n} ix_i = \frac{n(n+1)}{4}. \tag{6}$$

After substituting the above conditions into (4), it can be verified that the outcome of (3) is independent of the parameters a, b and c. In that case, the detector is said to be (a, b, c)-immune. Note that in case the offset is uniform, $c = 0$, it suffices to impose only the first condition (5) to render the performance (a, b)-immune. A code that satisfies the first condition (5) is called a *balanced* code, while a code that satisfies both conditions (5) and (6) is called a *dc²-balanced* or *second-order spectral-null* code.

Let $N_{\rm dc}(n)$ denote the number of dc-balanced codewords. For n even, we simply obtain

$$N_{\rm dc}(n) = \binom{n}{\frac{n}{2}},$$

and the redundancy, $r_{dc}(n)$, of balanced codes is approximately [5]

$$r_{\rm dc}(n) = n - \log_2 N_{\rm dc}(n) \simeq \frac{1}{2}\log_2 n + 0.326, \quad n \gg 1. \tag{7}$$

Properties and constructions of dc^2-balanced codes were first presented by Immink [4]. Efficient code constructions were presented by, for example, Yang [6], Tallini and Bose [7].

Let the number of dc^2-balanced codewords of length n be denoted by $N_{\rm dc^2}(n)$. It has been found in [4] that $N_{\rm dc^2}(n) = 0$ if $n \mod 4 \neq 0$. The number of dc^2-balanced codewords equals for asymptotically large n [8] (see also Section VI)

$$N_{\rm dc^2}(n) \simeq \frac{4\sqrt{3}}{\pi}\frac{2^n}{n^2}, \quad n \mod 4 = 0. \tag{8}$$

The redundancy of dc^2-balanced codes, denoted by $r_{\rm dc^2}(n)$, is approximated by

$$r_{\rm dc^2}(n) = n - \log_2 N_{\rm dc^2}(n) \simeq 2\log_2 n - 1.141, \quad n \gg 1. \tag{9}$$

The redundancy of dc^2-balanced codes is usually considered to be too high for practical applications. In the next section, we investigate a less redundant option that also guarantees (a, b, c)-immunity.

III. Detection using the Pearson Distance

Immink and Weber [2] presented a coding and detection technique that is immune to unknown gain and (uniform) offset, that is (a, b) immunity, which is based on the Pearson distance. The Pearson distance, $\delta(\boldsymbol{r}, \hat{\boldsymbol{x}})$, between the vectors $\boldsymbol{r}$ and $\hat{\boldsymbol{x}}$ is defined by

$$\delta(\boldsymbol{r}, \hat{\boldsymbol{x}}) = 1 - \rho_{\boldsymbol{r},\hat{\boldsymbol{x}}}, \tag{10}$$

where

$$\rho_{\boldsymbol{r},\hat{\boldsymbol{x}}} = \frac{\sum_{i=1}^{n}(r_i - \overline{r})(\hat{x}_i - \overline{\hat{x}})}{\sigma_r \sigma_{\hat{x}}} \tag{11}$$

is the well-known *(Pearson) correlation coefficient*, and we define two quantities, namely the *average symbol value* of $\hat{\boldsymbol{x}}$

$$\overline{\hat{x}} = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i, \tag{12}$$

and the (unnormalized) *symbol value variance* of $\hat{\boldsymbol{x}}$

$$\sigma_{\hat{x}}^2 = \sum_{i=1}^{n}(\hat{x}_i - \overline{\hat{x}})^2. \tag{13}$$

It is assumed that both codewords $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$ are taken from a judiciously chosen codebook S, whose properties are explained later. A Minimum Pearson Distance (MPD) detector outputs the codeword

$$\boldsymbol{x}_{\rm o} = \arg\min_{\hat{\boldsymbol{x}}\in S} \delta(\boldsymbol{r}, \hat{\boldsymbol{x}}). \tag{14}$$

Working out (10) yields

$$\begin{aligned}\delta(\boldsymbol{r}, \hat{\boldsymbol{x}}) &= 1 - \frac{1}{\sigma_r\sigma_{\hat{x}}}\sum_{i=1}^{n}(x'_i + b' + ci)(\hat{x}_i - \overline{\hat{x}}) \\ &= 1 - \frac{1}{\sigma_r\sigma_{\hat{x}}}\sum_{i=1}^{n} x'_i(\hat{x}_i - \overline{\hat{x}}) \\ &- \frac{1}{\sigma_r\sigma_{\hat{x}}}\sum_{i=1}^{n}(b' + ci)(\hat{x}_i - \overline{\hat{x}}), \end{aligned} \tag{15}$$

where $x'_i = a(x_i + \nu_i)$ and $b' = b - \overline{r}$. In the process of minimization (14) the terms independent of $\hat{\boldsymbol{x}}$ are irrelevant. The relevant $(b, c, \hat{\boldsymbol{x}})$-dependent term of $\delta(\boldsymbol{r}, \hat{\boldsymbol{x}})$ equals

$$\sum_{i=1}^{n}(b'+ci)(\hat{x}_i-\overline{\hat{x}}) = b'\sum_{i=1}^{n}(\hat{x}_i-\overline{\hat{x}})+c\sum_{i=1}^{n} i(\hat{x}_i-\overline{\hat{x}}). \quad (16)$$

Clearly, the b'-dependent term of (16) is nil since by definition (12) we have

$$\sum_{i=1}^{n}(\hat{x}_i - \overline{\hat{x}}) = \sum_{i=1}^{n}\hat{x}_i - n\overline{\hat{x}} = 0.$$

The c-dependent term of (16) can be neutralized if all codewords, $\hat{\boldsymbol{x}} \in S$, satisfy

$$\sum_{i=1}^{n} i(\hat{x}_i - \overline{\hat{x}}) = 0,$$

or

$$\sum_{i=1}^{n} i\hat{x}_i = \overline{\hat{x}}\sum_{i=1}^{n} i = \frac{1}{2}n(n+1)\overline{\hat{x}}.$$

After substituting (12), we obtain the principal condition

$$2\sum_{i=1}^{n} i\hat{x}_i = (n+1)\sum_{i=1}^{n}\hat{x}_i. \quad (17)$$

If all codewords $\hat{\boldsymbol{x}} \in S$ satisfy (17) then the remaining term of (15) is

$$1 - \frac{1}{\sigma_r\sigma_{\hat{x}}}\sum_{i=1}^{n} x'_i(\hat{x}_i - \overline{\hat{x}}) \ .$$

is independent of a, b, and c. The gain term, a, in both σ_r and x'_i, is merely a scaling factor not affecting the outcome of (14) (note the assumption is $a > 0$). We conclude that if all codewords $\boldsymbol{x} \in S$ satisfy condition (17) that a minimum Pearson distance detector using (14) is (a, b, c)-immune.

By rearranging the terms of (17), we obtain

$$\sum_{i=1}^{n}\left(i - \frac{n+1}{2}\right)\hat{x}_i = 0. \quad (18)$$

We infer from the above, borrowing terminology from mechanics, that the codewords in S have their *center of mass* at $(n+1)/2$. Note that for odd n, the value of the center bit, $x_{(n+1)/2}$, does not affect the outcome of (17).

The cardinality of S is denoted by $N(n) = |S|$. It is easily verified that both the all-'1' and the all-'0' words satisfy (17). From the teachings of [2], we infer that for unambiguous MPD detection, we must bar the all-'0' and all-'1' codewords, so that the number of available codewords is $N(n) - 2$.

IV. Properties of S

Below we have listed properties of the set S and its codewords.

1) Property 1: Let the inverse of the symbol x_i be denoted by $\tilde{x}_i = 1 - x_i$. The vector $\tilde{\boldsymbol{x}}$ is the inverse of $\boldsymbol{x}$ if all symbols are inverted. We simply find that $\boldsymbol{x} \in S \implies \tilde{\boldsymbol{x}} \in S$. So that we infer that $N(n) \bmod 2 = 0$.

2) Property 2: Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and let $\boldsymbol{x}_\mathrm{r} = (x_n, x_{n-1}, \ldots, x_1)$ denote the reverse of $\boldsymbol{x}$. We simply find using (18) that $\boldsymbol{x} \in S \implies \boldsymbol{x}_\mathrm{r} \in S$.

3) Property 3: Let n be odd, and $\boldsymbol{x} \in S$. Assume that $\tilde{\boldsymbol{x}}$ agrees with $\boldsymbol{x}$ on all $\tilde{x}_i$, $i \neq (n+1)/2$, and $\tilde{x}_{(n+1)/2} = 1 - \hat{x}_{(n+1)/2}$. Then, $\tilde{\boldsymbol{x}} \in S$. This follows from the fact that (18) does not depend on $\hat{x}_{(n+1)/2}$. Since $|S| \geq 2$ for all $n \geq 1$, we conclude that for odd n, the minimum distance of S equals unity. For even n, however, from (18) we infer that if $x_i = x_{n-i}$ for all $i \in \{1, 2, \ldots, n/2\} \implies \boldsymbol{x} \in S$. There are $2^{n/2}$ such codewords, where the first $n/2$ bits are taken arbitrarily, and the last $n/2$ bits are fixed as $x_i = x_{n-i}$. It is also easy to see that the minimum distance of S is exactly two, since $(100\cdots001)$ and $(000\cdots000)$ are in S. On the other hand, the minimum distance of S is not unity, since $\boldsymbol{x}$ satisfies (18).

4) Property 4: Assume that n is even. The left-hand side of (17) is even, but since $(n+1)$ is odd, $\sum_{i=1}^{n}\hat{x}_i$ must be even, and therefore any $\boldsymbol{x} \in S$ contains an even number of ones.

V. Counting using generating functions

We use *generating functions* for enumerating the number, $N(n) = |S|$, of codewords $\boldsymbol{x}$ that satisfy (17). To that end, let the generating function $g(y)$ be defined by the formal power series

$$g(y) = \sum_{i=0}^{\infty} g_i y^i.$$

Let $[y^n]g(y)$ denote the *extraction* of the coefficient of y^n in the formal power series $g(y) = \sum g_i y^i$, that is, define

$$[y^n]\left(\sum_{i=0}^{\infty} g_i y^i\right) = g_n.$$

In a similar fashion, we define a bi-variate generating function

$$\hat{g}(x, y) = \sum_{i,j}\hat{g}_{i,j}x^i y^j.$$

Define [4], [9]

$$h_n(x, y) = (1+xy)(1+xy^2)\ldots(1+xy^n). \quad (19)$$

Then $[x^{i_0}y^{j_0}]h_n(x, y)$ equals the number of n-sequences that satisfy the conditions

$$\sum_{i=1}^{n} x_i = i_0 \text{ and } \sum_{i=1}^{n} ix_i = j_0.$$

The number of n-sequences that satisfy (17) is given by

$$N(n) = \sum_{i=0}^{n}\left[x^i y^{\frac{i(n+1)}{2}}\right]h_n(x, y).$$

TABLE I
SIZE OF CODEBOOK, $N(n)$, AND $N_{\text{DC}^2}(n)$.

n	$N(n)$	$N_{\text{dc}^2}(n)$
4	4	2
5	8	0
6	8	0
7	20	0
8	18	8
9	52	0
10	48	0
11	152	0
12	138	58

We also have

$$N_{\text{dc}^2}(n) = [x^{\frac{n}{2}} y^{\frac{n(n+1)}{4}}] h_n(x, y), \; n \mod 4 = 0.$$

As

$$h_n(x, y) = h_{n-1}(x, y)(1 + xy^n),$$

we can write down a recursive relation for the coefficients of $h_n(x, y)$ [4], [9], where for clerical convenience we use the notation

$$C_m(i, j) = [x^i y^j] h_m(x, y).$$

For $m = 1, \ldots, n$, $i = 0, \ldots, m$ and $j = 0, \ldots, m(m+1)/2$

$$C_m(i, j) = C_{m-1}(i, j) + C_{m-1}(i - 1, j - m)$$

with initial conditions $C_0(0, 0) = 1$ and $C_0(i, j) = 0$ for any $(i, j) \neq (0, 0)$. Table I shows results of computations of $N(n)$ and $N_{\text{dc}^2}(n)$ versus n.

VI. ASYMPTOTIC BEHAVIOR

Using statistical arguments, we compute an approximation to the redundancy of the new codes for asymptotically large values of n. To that end, we define the stochastic variables

$$s = x_1 + x_2 + \ldots + x_n,$$

and

$$p = x_1 + 2x_2 + \ldots + nx_n,$$

where x_i, $1 \leq i \leq n$, are seen as i.i.d. two-valued random variables whose numerical outcomes '0' or '1' are equally likely. The expectations are $E[x_i^2] = E[x_i] = 1/2$, and $E[x_i x_j] = 1/4$, where $E[.]$ denotes the expected value operator. In case n is large, the central limit theorem states that the stochastic variables s and p, which are obtained by summing a large number, n, of independent stochastic variables, will show a two-dimensional Gaussian distribution. Then, for asymptotically large n, the number of n-sequences, denoted by $\varphi(s, p)$, is given by [10]

$$\varphi(s, p) \approx \frac{2^n}{2\pi\sigma_s\sigma_p\sqrt{1 - \rho^2}} e^{-\frac{f(s,p)}{2(1-\rho^2)}}, \tag{20}$$

where

$$f(s, p) = \left(\frac{s - \mu_s}{\sigma_s}\right)^2 + \left(\frac{p - \mu_p}{\sigma_p}\right)^2 - \frac{2\rho(s - \mu_s)(p - \mu_p)}{\sigma_s\sigma_p}. \tag{21}$$

The parameters μ_s and μ_p denote the average of s and p, σ_s^2 and σ_p^2 denote the variance of s and p, while ρ denotes the (linear) correlation between s and p. The five parameters are computed below.
We simply find

$$\mu_s = E[s] = E[x_1 + \ldots + x_n] = nE[x_i] = \frac{n}{2}. \tag{22}$$

Since

$$\begin{aligned} E[s^2] &= E[(x_1 + \ldots + x_n)^2] \\ &= \sum_{i=1}^{n} E[x_i^2] + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E[x_i x_j] \\ &= \frac{n}{2} + \frac{n(n-1)}{4}, \end{aligned}$$

we have

$$\sigma_s^2 = E[(s - \mu_s)^2] = E[s^2] - \mu_s^2 = \frac{n}{4}. \tag{23}$$

Similarly

$$\mu_p = E[p] = (1 + 2 + \ldots + n)E[x_i] = \frac{n(n+1)}{4}. \tag{24}$$

Since

$$\begin{aligned} E[p^2] &= \sum_{i=1}^{n} i^2 E[x_i^2] + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} ijE[x_i x_j] \\ &= n\left(\frac{n^3}{16} + \frac{5n^2}{24} + \frac{3n}{16} + \frac{1}{24}\right), \end{aligned}$$

we have

$$\sigma_p^2 = E[(p - \mu_p)^2] = \frac{n(n+1)(2n+1)}{24}. \tag{25}$$

We have

$$\begin{aligned} E[s.p] &= E[(x_1 + \ldots + x_n)(x_1 + 2x_2 \ldots + nx_n)] \\ &= \frac{n+1}{4}\sum_{i=1}^{n} i = \frac{n(n+1)^2}{8}, \end{aligned}$$

so that

$$\rho^2 = \frac{(E[s.p] - \mu_s\mu_p)^2}{\sigma_s^2\sigma_p^2} = \frac{3}{2}\frac{n+1}{2n+1}. \tag{26}$$

The number of dc^2-balanced codewords, $N_{\text{dc}^2}(n)$, for asymptotically large n, $n \mod 4 = 0$, can be found by substituting

$$s = \frac{n}{2} = \mu_s$$

and

$$p = \frac{n(n+1)}{4} = \mu_p,$$

into (20). Then

$$N_{\mathrm{dc}^2}(n) \approx \varphi(\mu_s, \mu_p) \approx \frac{2^n}{2\pi\sigma_s\sigma_p\sqrt{1-\rho^2}},$$

so that

$$\begin{aligned} r_{\mathrm{dc}^2}(n) &= n - \log_2 N_{\mathrm{dc}^2}(n) \\ &\approx \log_2 2\pi\sigma_s\sigma_p\sqrt{1-\rho^2} \\ &\approx 2\log_2 n - \log_2 \frac{4\sqrt{3}}{\pi}, \quad n \gg 1. \end{aligned}$$

The above result was earlier obtained in [8], [11].

The computation of $N(n)$ for asymptotically large n is slightly more involved. We find, using (20),

$$\begin{aligned} N(n) &\approx \sum_{s=0}^{n} \varphi\left(s, \frac{(n+1)s}{2}\right) \qquad (27) \\ &\approx \frac{2^n}{2\pi\sigma_s\sigma_p\sqrt{1-\rho^2}} \sum_{\substack{s=0 \\ s(n+1) \mod 2=0}}^{n} e^{-\frac{f\left(s, \frac{(n+1)s}{2}\right)}{2(1-\rho^2)}} \\ &\approx N_{\mathrm{dc}^2}(n) \sum_{\substack{s=0 \\ s(n+1) \mod 2=0}}^{n} e^{-\frac{f\left(s, \frac{(n+1)s}{2}\right)}{2(1-\rho^2)}}. \end{aligned}$$

We obtain after substituting $p = s(n+1)/2$

$$\left(\frac{p-\mu_p}{\sigma_p}\right)^2 = \frac{3(n+1)}{2(2n+1)} \frac{(s-\frac{n}{2})^2}{\frac{n}{4}} = \rho^2 \left(\frac{s-\mu_s}{\sigma_s}\right)^2.$$

After working out (21) and (26), we have

$$\frac{f\left(s, \frac{n+1}{2}s\right)}{2(1-\rho^2)} = \frac{(s-\frac{n}{2})^2}{\frac{n}{2}}.$$

Then, since for asymptotically large n,

$$\lim_{n\to\infty} \sum_{s=0}^{n} e^{-\frac{(s-\frac{n}{2})^2}{\frac{n}{2}}} \approx \int_0^\infty e^{-\frac{(s-\frac{n}{2})^2}{\frac{n}{2}}} ds = \sqrt{\frac{\pi n}{2}},$$

we find for n odd, as in (27) we add the terms for all $s \geq 0$,

$$\begin{aligned} N(n) &\approx N_{\mathrm{dc}^2}(n)\sqrt{\frac{\pi n}{2}} \\ &= \frac{2^n}{n^{3/2}}\sqrt{\frac{24}{\pi}}, \quad n \gg 1. \qquad (28) \end{aligned}$$

For n even, as in (27), we add the terms for even $s \geq 0$,

$$\begin{aligned} N(n) &\approx N_{\mathrm{dc}^2}(n)\sqrt{\frac{\pi n}{8}} \\ &= \frac{2^n}{n^{3/2}}\sqrt{\frac{6}{\pi}}, \quad n \gg 1. \qquad (29) \end{aligned}$$

For asymptotically large n, we obtain for the redundancy $r(n)$ of S

$$r(n) = n - \log_2 N(n) \approx \frac{3}{2}\log_2 n + \alpha, \qquad (30)$$

where $\alpha = \log_2 \sqrt{\pi/24} = -1.467..$ for n odd and $\alpha = -0.467..$ for n even. We conclude that the redundancy of the new codes is smaller than that of dc^2-balanced codes, but larger than that of dc-balanced codes.

VII. Conclusions

We have presented a minimum Pearson distance (MPD) detector that is immune to unknown slowly varying offset. We have formulated a constraint and enumerated the number of codewords that can be used in conjunction with MPD detection. For asymptotically large n, the redundancy $r(n) \approx \frac{3}{2}\log_2 n + \alpha$, where $\alpha = \log_2 \sqrt{\pi/24} = -1.467..$ for n odd and $\alpha = -0.467..$ for n even. The redundancy of the new codes is smaller than that of dc^2-balanced codes, but larger than that of dc-balanced codes. The minimum Hamming distance of the new codes for even n is equal to two and for odd n equal to unity. The redundancy of the new codes for n odd is around one bit less than that for n even.

References

[1] H. Pozidis, T. Mittelholzer, N. Papandreou, T. Parnell, and M. Stanisavljevic, "Phase Change Memory Reliability: A Signal Processing and Coding Perspective," *IEEE Trans. on Magnetics,* vol. MAG-51, No. 4, April 2015

[2] K. A. S. Immink and J. H. Weber, "Minimum Pearson Distance Detection for Multi-Level Channels with Gain and/or Offset Mismatch," *IEEE Trans. Inform. Theory,* vol. IT-60, pp. 5966-5974, Oct. 2014.

[3] K. A. S. Immink and J. H. Weber, "Hybrid Minimum Pearson and Euclidean Distance Detection," *IEEE Trans. Commun.,* vol. COM-63, 2015.

[4] K. A. S. Immink, "Spectrum Shaping with Binary DC^2-constrained Channel Codes," *Philips J. Res.,* vol. 40, pp. 40-53, 1985.

[5] J. H. Weber and K. A. S. Immink, "Knuth's Balancing of Codewords Revisited," *IEEE Trans Inform Theory,* vol. 56, no. 4, pp. 1673-1679, 2010.

[6] C. N. Yang, "Design of Efficient Second-Order Spectral-Null Codes," *IEEE Trans. Inform. Theory,* vol. IT-51, no. 4, pp. 1580-1584, April 2005.

[7] L. G. Tallini and B. Bose, "On Efficient High-Order Spectral-Null Codes," *IEEE Trans. Inform. Theory,* vol. IT-45, no. 7, pp. 2594-2601, Nov. 1999.

[8] H. Prodinger, "On the Number of Partitions of $\{1, \ldots, n\}$ into Two Sets of Equal Cardinalities and Equal Sums," Canad. Math. Bull., Vol. 25 (2), pp. 238-241, 1982.

[9] Y. Xin and I. J. Fair, "Algorithms to Enumerate Codewords for DC^2-constrained Channels," *IEEE Trans. Inform. Theory,* vol. IT-47, no. 7, pp. 3020-3025, Nov. 2001.

[10] P. Flajonet and R. Sedgewick, *Analytic Combinatorics*, ISBN 978-0-521-89806-5, Cambridge University Press, 2009.

[11] R. M. Roth, P. H. Siegel, and A. Vardy, "Higher-Order Spectral-Null Codes: Constructions and Bounds," *IEEE Trans. Inform. Theory,* vol. IT-40, pp. 1826-1840, Nov. 1994.