Hybrid Minimum Pearson and Euclidean Distance Detection

Kees A. Schouhamer Immink and Jos H. Weber

Abstract - The reliability of mass storage systems, such as optical data recording and Non-Volatile Memory (Flash), is seriously hampered by uncertainty of the actual value of the offset (drift) or gain (amplitude) of the retrieved signal. The recently introduced minimum Pearson distance detection is immune to unknown offset or gain, but this virtue comes at the cost of a lessened noise margin at nominal channel conditions. We will present a novel hybrid detection method, where we combine the outputs of the minimum Euclidean distance and Pearson distance detectors so that we may trade detection robustness versus noise margin. We will compute the error performance of hybrid detection in the presence of unknown channel mismatch and additive noise.

Key words: Constant composition code, permutation code, flash memory, digital optical recording, Non-Volatile Memory, NVM, mismatch, adaptive equalisation, Pearson distance, Euclidean distance.

I. INTRODUCTION

The receiver of a transmission or storage system is often ignorant of the exact value of the amplitude (gain) and/or offset of the received signal, which depend on the actual, time-varying, conditions of the channel. In wireless communications, for example, the amplitude of the received signal may vary rapidly due to multipath propagation or due to shadowing from obstacles affecting the wave propagation, a phenomenon often called fading. In optical disc recording, both the gain and offset depend on the reflective index of the disc surface and the dimensions of the written features [1]. Fingerprints on optical discs may result in rapid gain and offset variations of the retrieved signal.

In non-volatile memories, such as floating gate memories, the data is represented by stored charge. The stored charge can leak away from the floating gate through the gate oxide or through the dielectric. This leakage may result in a shift of the offset or threshold voltage of the memory cell. The amount of leakage depends on various physical parameters and, clearly, the time elapsed between writing and reading the data. As a result, the offset between different groups of cells may be very different. In such a case, the conventional approach of combatting the effects of offset uncertainty, or channel mismatch, by using an automatic offset or gain control (AGC), which estimates the actual offset or gain from the previously received data, can not be applied, and other methods have to be developed.

In optical disc recording devices and non-volatile memories, constrained codes, specifically dc-free or balanced codes, have been used and/or proposed to counter the detrimental effects of offset and gain mismatch [2]. Jiang et al. [3] addressed a q-ary balanced coding technique, called rank modulation, for circumventing the difficulties with flash memories having aging offset levels. Zhou et al. [4], Immink [5], and Sala et al. [6], [7] investigated constrained codes that enable dynamic reading thresholds in non-volatile memories. Immink & Weber [8] advocated Pearson-distance-based detection, which is intrinsically resistant to offset and gain mismatch. A minimum Pearson distance detector has the virtue of immunity against gain and offset uncertainty, but it comes at a price of a lessened noise margin at nominal channel conditions.

There is no need to strive for immunity where merely a certain amount of detector robustness will fulfill the design criteria. Our paper offers an answer to this desideratum as we will investigate a hybrid detector that combines both the minimum Pearson and Euclidean distance detectors, offering the designer the flexibility of trading the detector robustness on the one hand versus noise margin penalty at nominal channel conditions on the other hand.

The outline of the paper is as follows. We start our presentation in Section II with a description of hybrid minimum Pearson and Euclidean distance detection. In Sections III and IV, we present an analysis of the error performance of the hybrid detector, and a mismatch sensitivity analysis. We will discuss implementation issues in Section V. In Section VI, we will describe our conclusions.

II. DESCRIPTION OF HYBRID MINIMUM PEARSON AND EUCLIDEAN DISTANCE DETECTION

We consider a communication codebook, S, of chosen q-ary codewords \( x = (x_1, x_2, \ldots, x_n) \) over the q-ary alphabet \( \mathbb{Q} = \{0, 1, \ldots, q-1\} \), \( q \geq 2 \), where \( n \), the length of \( x \), is a positive integer. We suppose that the sent codeword, \( x \), is received as the vector \( r = a(x + \nu) + b, r_i \in \mathbb{R} \), where \( x \) is scaled by an unknown factor, called gain, \( a > 0 \), offsetted by an unknown (dc-) offset, \( b \), where \( a \) and \( b \in \mathbb{R} \).
and corrupted by additive noise $\mathbf{\nu} = (\nu_1, \ldots, \nu_n)$, $\nu_i \in \mathbb{R}$ are noise samples with distribution $\mathcal{N}(0, \sigma^2)$. We use the shorthand notation $x + b = (x_1 + b, x_2 + b, \ldots, x_n + b)$.

A. T-constrained codes

Before we proceed with a description of the new detector, we briefly describe $T$-constrained codes, which are required in conjunction with the Pearson distance metric. $T$-constrained codes, presented in [5] for enabling simple dynamic threshold detection of $q$-ary codewords, consist of $q$-ary $n$-length codewords, where $T$, $0 < T \leq q$, prescribed symbols must appear at least once in a codeword.

Of specific interest in this paper are two types of $T$-constrained codes, which are denoted by $S_1$ and $S_2$. The code $S_1$ comprises codewords wherein the symbol ‘0’ appears at least once. The cardinality of $S_1$ equals [5]

$$|S_1| = q^n - (q - 1)^n, q > 1.$$  

The set $S_2$ contains codewords wherein both the symbols ‘0’ and ‘$q - 1$’ appear at least once. The cardinality of $S_2$ equals [5]

$$|S_2| = q^n - 2(q - 1)^n + (q - 2)^n, q > 1.$$  

For the binary case, $q = 2$, we simply find

$$|S_1| = 2^n - 1,$$

as the all-'1' word is deleted, and

$$|S_2| = 2^n - 2,$$

as both the all-'1' and all-'0' words are deleted.

B. New hybrid detector

We are now in the position to introduce the new hybrid receiver. We will study two situations, where a) we have offset and gain mismatch and b) where we face offset only mismatch. For the first situation, both gain and offset mismatch, Immink & Weber [8] introduced the Pearson distance as it is more suitable to be combined with the Euclidean distance measure.

Throughout we will use the above expression (7) for the Pearson distance as it is more suitable to be combined with the Euclidean distance measure.

For the second situation, offset-only mismatch, Immink & Weber [8] introduced a modified Pearson distance between $r$ and $\hat{x}$, defined by

$$\delta_{p_1}(r, \hat{x}) = \sum_{i=1}^{n} (r_i - \hat{x}_i + \overline{x})^2,$$

where $x$ and $\hat{x}$ are taken from codebook $S_1$.

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

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

We are now in the position to define the new distance metric, $\delta(r, \hat{x})$, a weighted sum of Pearson and Euclidean distance, namely

$$\delta(r, \hat{x}) = \gamma \delta_{e}(r, \hat{x}) + (1 - \gamma)\delta_{p}(r, \hat{x}),$$  

where the weighting parameter $\gamma$, $0 \leq \gamma \leq 1$, is a design parameter. The hybrid receiver outputs the codeword

$$x_o = \arg \min_{\hat{x} \in S} \delta(r, \hat{x}).$$  

Note the following clerical convention for (10): for the gain-and-offset-mismatch case we read $S = S_2$ and $\delta_p(r, \hat{x}) = \delta_{p_2}(r, \hat{x})$, and for the offset-only case, we read $S = S_1$ and $\delta_p(r, \hat{x}) = \delta_{p_1}(r, \hat{x})$.

The parameter $\gamma$ is selected by the system designer. Clearly, for $\gamma = 1$ we obtain a conventional minimum Euclidean distance detector, which offers optimal noise margin, but lacks robustness in the presence of channel mismatch. For $\gamma = 0$, we obtain a minimum Pearson distance detector, which is immune against both offset and gain mismatch, but it comes at a lessened noise margin at nominal channel conditions. For non-zero values of the parameter $\gamma$, the receiver is not immune to channel mismatch, but by a judicious choice of $\gamma$ the designer may balance detection robustness against channel mismatch uncertainty on one hand versus noise margin penalty at nominal channel conditions on the other hand. In the next sections, we will compute the error performance of the hybrid minimum Euclidean/Pearson distance detector, where we will focus on the trade-off relationship between detector robustness and noise margin. We start with the simplest case, offset-only mismatch, which is amenable to analysis.
III. ANALYSIS OF THE HYBRID DETECTOR, OFFSET MISMATCH

In the three subsections of this section, we will analyze the error performance and the tolerance range of the hybrid detector, then we will introduce parity check codes that will improve the noise margin at the cost of a redundancy of $1/n$.

A. Error performance

The hybrid detector errs, see (11), if there is at least one codeword $\hat{x} \neq x, \hat{x} \in S_1$, that satisfies

$$\delta(r, \hat{x}) < \delta(r, x).$$

Using (10), we obtain

$$\gamma(\delta_e(r, \hat{x}) - \delta_e(r, x)) + (1 - \gamma)(\delta_p_1(r, \hat{x}) - \delta_p_1(r, x)) < 0.$$  \hspace{1cm} (12)

Working out the Euclidean part, using (9), we obtain

$$\delta_e(r, \hat{x}) - \delta_e(r, x) = 2 \sum_{i=1}^{n} e_i \nu_i + \sum_{i=1}^{n} e_i (e_i + 2b),$$  \hspace{1cm} (13)

and for the Pearson part, using (8), we obtain

$$\delta_p_1(r, \hat{x}) - \delta_p_1(r, x) = 2 \sum_{i=1}^{n} (e_i - \bar{e}) \nu_i + \sum_{i=1}^{n} (e_i - \bar{e})^2,$$  \hspace{1cm} (14)

where, for clerical convenience, we introduce the short-hand notation

$$e_i = x_i - \hat{x}_i$$  \hspace{1cm} (15)

and

$$\bar{e} = \bar{x} - \bar{\hat{x}}.$$  \hspace{1cm} (16)

Note that an interfering $b$-dependent term,

$$2b \sum_{i=1}^{n} e_i = 2bn\bar{e},$$

is only present in (13), the Euclidean part of (12). The Pearson part (14) of (12) is, as expected, independent of the offset $b$. We summarize inequality (12) by

$$\zeta_0 + \sum_{i=1}^{n} \zeta_i \nu_i < 0,$$  \hspace{1cm} (17)

where

$$\zeta_0 = \gamma \sum_{i=1}^{n} e_i (e_i + 2b) + (1 - \gamma) \sum_{i=1}^{n} (e_i - \bar{e})^2, $$  \hspace{1cm} (18)

and

$$\zeta_i = 2\gamma e_i + 2(1 - \gamma)(e_i - \bar{e}), \quad 1 \leq i \leq n.$$  \hspace{1cm} (19)

The noise samples $\nu_i$ are supposed to be drawn from $N(0, \sigma)$, so that the probability that $\delta(r, \hat{x}) < \delta(r, x)$ equals

$$Q \left( \frac{d_1(x, \hat{x})}{2 \sigma} \right),$$

where the $Q$-function is defined by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-u^2/2} du,$$

and

$$d_1(x, \hat{x}) = \frac{2\alpha_1}{\sqrt{\beta_1}},$$  \hspace{1cm} (20)

where

$$\alpha_1 = \zeta_0$$  \hspace{1cm} (21)

and

$$\beta_1 = \sum_{i=1}^{n} \zeta_i^2.$$  \hspace{1cm} (22)

Working out (18), (19), and (22) yields

$$\alpha_1 = \gamma \sum_{i=1}^{n} e_i (e_i + 2b) + (1 - \gamma) \sum_{i=1}^{n} (e_i - \bar{e})^2$$

$$= -(1 - \gamma) n\bar{e}^2 + 2\gamma bn\bar{e} + \sum_{i=1}^{n} e_i^2,$$  \hspace{1cm} (23)

and

$$\beta_1 = \sum_{i=1}^{n} [2\gamma e_i + 2(1 - \gamma)(e_i - \bar{e})]^2$$

$$= 4 \sum_{i=1}^{n} e_i^2 - (1 - \gamma) \bar{e}^2$$

$$= -4(1 - \gamma)^2 n\bar{e}^2 + 4 n \sum_{i=1}^{n} e_i^2.$$  \hspace{1cm} (24)

Note that

$$\sum_{i=1}^{n} e_i^2 = \delta_e(x, \hat{x})$$

equals the Euclidean distance between $x$ and $\hat{x}$.

The word error rate (WER) over all coded sequences $x$ is upperbounded by

$$\text{WER} < \frac{1}{|S|} \sum_{x \in S} \sum_{\hat{x} \neq x} Q \left( \frac{d_1(x, \hat{x})}{2 \sigma} \right).$$ \hspace{1cm} (26)

Figure 1 shows results of computer simulations (dotted lines) and computations invoking (26) for $q = 2, n = 8$, and $\gamma = 0, \gamma = 0.25,$ and $\gamma = 1$ for the matched case $b = 0$. The signal-to-noise-ratio (SNR), is defined by

$$\text{SNR}(\text{dB}) = -20 \log_{10} \sigma.$$  \hspace{1cm} (27)

We notice that in the ideal case, $b = 0$, the Euclidean detector, $\gamma = 1$, outperforms both the Pearson and hybrid detector. The situation is completely changed, however, in Figure 2, where results are shown of computer simulations (dotted lines) for the same parameters as used in Figure 1, but now at an offset $b = 0.3$. The diagram demonstrates that at an offset, $b = 0.3$, the traditional Euclidean distance detector has thrown in the towel as, on the average, one in five codewords are detected erroneously, while the performance of the hybrid detector is showing robustness to cope with unpredictable offset tolerances. In the next subsection, we will quantify the tolerance range of the offset $b$. 

---

$Q(x)$ is the cumulative distribution function of the standard normal distribution. The term $d_1(x, \hat{x})$ is the distance between $x$ and $\hat{x}$ measured in the Euclidean distance. The $Q$-function gives the probability that a random variable from a standard normal distribution is less than $x$. The $\alpha_1$ and $\beta_1$ terms are used to calculate the probability of error in the presence of noise.
B. Tolerance range

At asymptotically high signal-to-noise ratios, the word error rate is overbounded by [8]

$$\text{WER} < \frac{2n(q-1)}{q} Q \left( \frac{d_{1,\text{min}}}{2\sigma} \right), \quad (28)$$

where the minimum distance between any pair of codewords in $S_1$, $d_{1,\text{min}}$, is

$$d_{1,\text{min}} = \min_{\substack{x,x' \in S_1 \setminus \{x\} \neq x'}} d_1(x,x') = \min_{\substack{x,x' \in S_1 \setminus \{x\} \neq x'}} \frac{2\alpha_1}{\sqrt{\beta_1}}. \quad (29)$$

We readily find that unity-Euclidean-distance error events show the least $d_{1,\text{min}}$. Then in the worst case scenario, $\sum c_i = 1$ and $b\bar{e} = -b/n$, and we obtain, using (23) and (25), that

$$d_{1,\text{min}} = \frac{1 - \frac{1-\gamma^2}{n}}{\sqrt{1 - \frac{1-\gamma^2}{n}}}.$$

Note that $d_{1,\text{min}}$ is independent of the alphabet size $q$. By varying the value of $\gamma$, we select an error performance that lies between that of a minimum Euclidean and a minimum Pearson distance detector. For the ideal case, $b = 0$, we find that the asymptotic coding gain, $\Gamma_1(\gamma)$, equals

$$\Gamma_1(\gamma) = \frac{1 - \frac{1-\gamma}{n}}{\sqrt{1 - \frac{1-\gamma^2}{n}}}. \quad (31)$$

The term $2\gamma|b|$ in (30) shows how the offset mismatch affects the error performance. The offset mismatch, $b$, must lie within the tolerance range specified by

$$|b| \leq b_w(\gamma), \quad (32)$$

denotes the maximum offset that the receiver can tolerate just before collapsing. Clearly the maximum allowable offset, $b_w(\gamma)$, varies between $b_w(\gamma) = 0.5$ for Euclidean distance detection ($\gamma = 1$), and $b_w(\gamma) \to \infty$ for Pearson distance detection ($\gamma = 0$). By invoking (31) and (33), we can compute the trade-off function for other values of the design parameter $\gamma$. We plotted in Figure 3 the parametric equation of the maximum offset that the receiver can tolerate, $b_w(\gamma)$, versus coding gain, $20\log \Gamma_1(\gamma)$ (dB), for $n = 6, 8$ and 10. The curves show the crucial trade-off function between coding gain and offset tolerance range. For example, for $n = 8$, the designer may buy for an insurance premium of 0.4 dB coding loss at nominal channel conditions, a maximum offset tolerance improvement by a factor of around six with respect to conventional Euclidean distance detection. The receiver is completely immune to offset mismatch for the price of a coding loss of $10\log(7/8) \approx 0.58$ dB at nominal channel conditions.

C. Parity check codes

For certain applications it may be desirable to improve the noise margin at the expense of code redundancy. A parity check $T$-constrained code, denoted by $S_{p,1}$, consists of codewords, $x$, taken from $S_1$ that additionally satisfy the parity check

$$\left( \sum_{i=1}^n x_i \right) \mod q = a, \quad (34)$$

where $a \in Q$ is an integer chosen by the code designer, for example, to maximize the code size. For the best choice of $a$, the code size, $|S_{p,1}|$, is at least

$$|S_{p,1}| \geq \left\lceil \frac{|S_1|}{q} \right\rceil.$$
For the binary case, we have from (1) that $|S_1| = 2^n - 1$. Then choosing $a = (n + 1) \mod 2$, the code size equals $|S_{p,1}| = 2^{n-1}$, so that the rate of the parity check code is exactly $1 - 1/n$. We easily compute that the minimum distance, $d_{p,\text{min}}$, between any pair of codewords equals

$$d_{p,\text{min}} = \sqrt{2 \frac{1 - 2^{-\gamma}}{n} - 2\gamma|b|}.$$

(35)

The coding gain, $\Gamma_{p,1}(\gamma)$, under matched conditions, $b = 0$, equals

$$\Gamma_{p,1}(\gamma) = \sqrt{2 \frac{1 - 2^{-\gamma}}{n}}.$$  

(36)

Figure 4 shows a comparison of the error performance of offset-resistant detection with and without extra parity check. For asymptotically large SNR, we may, for $n = 12$ and $\gamma = 0$, expect to gain a factor $\Gamma_{p,1}(0)/\Gamma_1(0) = \sqrt{20/11} (= 2.6 \text{ dB})$ in noise margin with respect to the non-parity case. We deduct from the diagram that the gain is around 1.9 dB in the SNR region shown.

The receiver will fail if, see (35),

$$|b| \geq b_{p,w}(\gamma),$$

(37)

where

$$b_{p,w}(\gamma) = \frac{1}{2\gamma} \left(1 - 2^{-\gamma} \frac{1}{n}\right)$$

(38)

denotes the maximum offset that the detector can tolerate before collapsing. We conclude that the parity check receiver is slightly more susceptible to offset mismatch than a standard receiver, see (33). We plotted in Figure 5 the parametric equation of the maximum offset that the receiver can tolerate, $b_{p,w}(\gamma)$, versus coding gain, $20 \log \Gamma_{p,1}(\gamma)$, for $n = 6$, 8, and 10. In the next section, we will discuss the performance of the hybrid detector when it faces both offset and gain uncertainty.

IV. ERROR PERFORMANCE ANALYSIS OF THE HYBRID DETECTOR, OFFSET AND GAIN MISMATCH

The hybrid receiver will fail if there is at least one codeword $\tilde{x} \neq x, \tilde{x} \in S$, see (10) and (11), such that

$$\gamma(\delta_x(r, \tilde{x}) - \delta_x(r, x)) + (1 - \gamma)(\delta_{p_{1,2}}(r, \tilde{x}) - \delta_{p_{1,2}}(r, x)) < 0.$$  

(39)

We have, using (9),

$$\delta_x(r, \tilde{x}) - \delta_x(r, x) = \sum_{i=1}^{n}(x_i - \tilde{x}_i)^2 + 2\alpha \sum_{i=1}^{n}(x_i - \tilde{x}_i)t_i + t,$$

(40)

where the interference term

$$t = t_1 + t_2.$$
consists of two terms, namely
\[ \nu_1 = 2b \sum_{i=1}^{n} (x_i - \hat{x}_i) \]
and
\[ \nu_2 = 2(a - 1) \sum_{i=1}^{n} x_i(x_i - \hat{x}_i). \]

Further, using (7), we have
\[ \delta_p(r, \hat{x}) - \delta_p(r, \hat{x}) = 2a \sum_{i=1}^{n} x_i(a_i - \hat{a}_i) + 2a \sum_{i=1}^{n} \nu_i (a_i - \hat{a}_i) \quad (41) \]
where
\[ a_i = \frac{x_i - \bar{x}}{\sigma_x} \quad (42) \]
and
\[ \hat{a}_i = \frac{\tilde{x}_i - \bar{x}}{\sigma_x}. \quad (43) \]

Note that the above term (41), unlike the Euclidean term (40), is not affected by intersymbol interference. Using (40) and (41), we may rewrite inequality (39) by
\[ \zeta_0 + \sum_{i=1}^{n} \zeta_i \nu_i < 0, \quad (44) \]
where
\[ \zeta_0 = \gamma \nu + \gamma \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 + 2a(1-\gamma) \sum_{i=1}^{n} x_i(a_i - \hat{a}_i) \quad (45) \]
and
\[ \zeta_i = 2a\gamma (x_i - \hat{x}_i) + 2a(1-\gamma)(a_i - \hat{a}_i), \quad 1 \leq i \leq n. \quad (46) \]

The noise samples \( \nu_i \) are drawn from \( N(0, \sigma) \), so that the probability that an error is made, that is, if condition (44) holds, equals
\[ Q \left( \frac{d_2(x, \hat{x})}{2\sigma} \right), \]
where
\[ d_2(x, \hat{x}) = \frac{2\alpha_2}{\sqrt{\beta_2}} \quad (47) \]
and
\[ \alpha_2 = \zeta_0 \quad (48) \]
and
\[ \beta_2 = \sum_{i=1}^{n} \zeta_i^2. \quad (49) \]

Working out the above yields
\[ \beta_2 = \sum_{i=1}^{n} \zeta_i^2 = 4a^2[\gamma^2 \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 + 2(1-\gamma) \sum_{i=1}^{n} (x_i - \hat{x}_i)(a_i - \hat{a}_i) + (1-\gamma)^2 \sum_{i=1}^{n} (a_i - \hat{a}_i)^2]. \quad (50) \]

Reference [8] teaches that
\[ \sum_{i=1}^{n} x_i(a_i - \hat{a}_i) = \sum_{i=1}^{n} (x_i - \bar{x})(a_i - \hat{a}_i) \]
and
\[ \sum_{i=1}^{n} x_i(a_i - \hat{a}_i) = \left( \frac{x_i - \bar{x}}{\sigma_x} \right) \sum_{i=1}^{n} (x_i - \bar{x}) = \sigma_x (1 - \rho_{x, \hat{x}}). \quad (51) \]

where \( \rho_{x, \hat{x}} \) is the Pearson correlation coefficient between \( x \) and \( \hat{x} \), defined by (4). Similarly,
\[ \sum_{i=1}^{n} (a_i - \hat{a}_i)^2 = 2(1 - \rho_{x, \hat{x}}) \quad (52) \]
and
\[ \sum_{i=1}^{n} (x_i - \hat{x}_i)(a_i - \hat{a}_i) = -\sigma_x (1 - \rho_{x, \hat{x}}) \quad (53) \]

Then substitution yields
\[ \alpha_2 = \gamma \delta_x(x, \hat{x}) + \gamma \nu + 2a(1-\gamma)\sigma_x (1 - \rho_{x, \hat{x}}) \quad (54) \]
and
\[ \beta_2 = 4a^2[\gamma^2 \delta_x(x, \hat{x}) + 2\gamma(1-\gamma)(\sigma_x + \sigma_{\hat{x}})(1 - \rho_{x, \hat{x}}) + 2(1-\gamma)^2(1 - \rho_{x, \hat{x}})]. \quad (55) \]

Figure 6 shows results of computer simulations (dotted lines) and computations invoking (26) replacing \( d_1(x, \hat{x}) \) by \( d_2(x, \hat{x}) \), for the matched case \( (\nu = 0) \), \( q = 2 \), \( n = 10 \), and \( \gamma = 0, \gamma = 0.25 \), and \( \gamma = 1 \). In the matched case, \( a = 1 \) and \( b = 0 \), the coding gain, \( \Gamma_2(\gamma) \), defined by
\[ \Gamma_2(\gamma) = \min_{x, \hat{x} \neq S_2} \frac{2\alpha_2}{\sqrt{\beta_2}} \quad \text{for} \quad x, \hat{x} \in S_2 \quad (56) \]
In this section, we will study the effects of channel mismatch on the error performance of the hybrid Pearson/Euclidean distance detector. In order to illustrate the sensitive analysis, we have simulated the error performance of the hybrid receiver. We have selected the parameters $q = 3$, $n = 6$, gain $a = 1.06$, and offset $b = 0.06$ at an SNR of 18 dB. Results are plotted in Figure 8, where we notice that by selecting $\gamma = 0.2$ we minimize the word error rate.

Figure 7 shows, for the matched case, results of computer searches, where we have evaluated (54), (55), and (56) for $n = 12$ and various values of $q$. We may notice that, as expected, by varying the value of $\gamma$, we select an error performance that lies between that of a pure minimum Euclidean and a minimum Pearson distance detector. In the next subsection, we will examine how a $\gamma$ choice affects the maximum tolerance range of gain and offset.

A. Mismatch Sensitivity Analysis

In this section, we will study the effects of channel mismatch on the error performance of the hybrid Pearson/Euclidean distance detector. In order to illustrate the sensitive analysis, we have simulated the error performance of the hybrid receiver. We have selected the parameters $q = 3$, $n = 6$, gain $a = 1.06$, and offset $b = 0.06$ at an SNR of 18 dB. Results are plotted in Figure 8, where we notice that by selecting $\gamma = 0.2$ we minimize the word error rate.

The receiver will completely fail if

$$\min_{x, \hat{x} \in S_2 \atop x \neq \hat{x}} \alpha_2 < 0,$$

or

$$\min_{x, \hat{x} \in S_2 \atop x \neq \hat{x}} \{\gamma \delta_e(x, \hat{x}) + 2a(1-\gamma)\sigma_x(1-\rho_{x,\hat{x}}) + \gamma \ell\} < 0,$$  \hspace{1cm} (57)

where

$$\ell = 2b \sum (x_i - \hat{x}_i) + 2(a-1) \sum x_i(x_i - \hat{x}_i).$$

Inequality (57) defines a tolerance area in the $(a, b)$ space where the receiver will fail. The inequality is not amenable to analysis, so we relied on computer searches for relatively small $q$ and $n$. Figure 9 shows results of our findings, evaluating (57), for $q = 6$, $n = 12$, and selected values of $\gamma$. The diagram shows tetragons, where the receiver will fail if the gain and offset parameters, $(a, b)$, are outside the tetragons. As expected, the tolerance range is increasing with decreasing values of the weighting factor $\gamma$. The increased tolerance range comes at a price of a coding loss at nominal channel conditions, see Figure 7, of 0.39 dB for $\gamma = 0.25$, and 1.01 dB for $\gamma = 0.1$, respectively.

B. Parity check codes

In the same vein as for the offset-only case, we define $T$-constrained parity check codes, which are denoted by $S_{p,2}$. For the binary case, $q = 2$, we are able to construct a parity check rate, $1 - 1/n$, code for even $n$. We define the coding gain of parity check codes, $\Gamma_{p,2}$, by

$$\Gamma_{p,2} = \min_{x, \hat{x} \in S_{p,2} \atop x \neq \hat{x}} \frac{2\alpha_2}{\sqrt{\beta_2}}$$  \hspace{1cm} (58)

Figure 10 shows results of computations using (54), (55), and (58), where we have selected the same parameters as in Figure 7.
V. IMPLEMENTATION ISSUES

The proposed distance measure, (10), a weighted sum of two distance measures, is more complicated to evaluate than a conventional Euclidean distance. We have studied the implementation issues of the offset-only case. After combining (9), (5), and (8), we rewrite the distance, (10), as

\[
\delta(\mathbf{r}, \mathbf{x}) = \sum_{i=1}^{n} r_i^2 - 2 \sum_{i=1}^{n} r_i x_i + \sum_{i=1}^{n} x_i^2 + (1 - \gamma)(-n \bar{x}^2 + 2\bar{r} \sum_{i=1}^{n} r_i). \tag{59}
\]

The two right-hand terms are the extra load for the receiver with respect to traditional Euclidean detection, which is embodied by the first three terms. Before we further discuss the complexity issue of the distance measure, we will first discuss how many evaluations of the distance are needed per received word. We tacitly assumed that the receiver computes the distance metric for all codewords in \( S \) before it gives its verdict on the codeword sent. This implies an exponential increase in the computational load with codeword length \( n \).

However, the codebook \( S \) can be divided into \( K \) constant composition codes. A constant composition code is a set of codewords where the numbers of occurrences of the symbols within a codeword is the same for each codeword. Immink & Weber [8] showed that we can reduce the number of distance computations to the number of constant composition codes, \( K \). For example, in case \( q = 2 \), the codebook \( S_1 \) is the union of \( K = n \) constant composition (weight) codes, so that instead of \( 2^n \) only \( n \) distance computations are required.

Let a constant composition code be denoted by \( S_{w_i} \), \( 1 \leq i \leq K \). Each constant composition code, \( S_{w_i} \), is characterized by one codeword of the code, called pivot word, denoted by \( \mathbf{x}_{p_i} \), \( 1 \leq i \leq K \). The pivot word \( \mathbf{x}_{p_i} \) is, by definition, the largest word in the lexicographical ordering (the largest symbols first) of the codewords in \( S_{w_i} \).

Thus, for a pivot word we have \( \mathbf{x}_{p_i} \in S_{w_i} \) and \( \mathbf{x}_{p_i} = (x_{p_{i,1}}, x_{p_{i,2}}, \ldots, x_{p_{i,n}}) \), where \( x_{p_{i,1}} \geq x_{p_{i,2}} \geq x_{p_{i,3}} \ldots \geq x_{p_{i,n}} \), \( 1 \leq i \leq K \). The remaining codewords in \( S \) consist of all distinct sequences that can be formed by permuting the order of the \( n \) symbols of each pivot word \( \mathbf{x}_{p_i} \), \( 1 \leq i \leq K \).

Detection of the received vector, \( \mathbf{r} \), is accomplished by invoking the following two-step procedure:

1) The \( n \) symbols of the received word, \( \mathbf{r} \), are sorted, largest to smallest, in the same way as taught in Slepian’s prior art [9].

2) Compute the distance, using (59), for all \( K \) pivot words. The receiver decides that a word taken from \( S_{w_i} \), \( 1 \leq k \leq K \), was sent in case the pivot word \( \mathbf{x}_{p_i} \) is at minimum distance to the received vector \( \mathbf{r} \). Now that we have ascertained that the sent word is a member of \( S_{w_i} \), we can decode the received word by Slepian’s procedure for single permutation codes.

The sorting of the symbols of the received vector, \( \mathbf{r} \), can be accomplished by an routine picked from a wealth of available algorithms, whose choice depends on \( n \) and the available processor hardware. The complexity grows worst-case with \( n^2 \) and on average with \( n \log n \) [10]. In an embodiment of the second part of the above procedure, using a micro-processor, for example, the receiver has tabulated and stored in memory the \( K \) pivot words. In addition, the receiver has, for each pivot word, pre-computed during initialization the terms, see (5),

\[
A_j = \frac{1}{n} \sum_{i=1}^{n} x_{p_{j,i}}, \quad 1 \leq j \leq K
\]

and, see (59),

\[
B_j = \sum_{i=1}^{n} x_{p_{j,i}}^2 - n(1 - \gamma)A_j^2, \quad 1 \leq j \leq K,
\]

which are independent of \( \mathbf{r} \). The receiver computes for each received codeword

\[
C = 2 \sum_{i=1}^{n} r_i
\]

and for each of the \( K \) pivot words, see (59),

\[
D_j = -2 \langle \langle \mathbf{r}, \mathbf{x}_{p_j} \rangle \rangle + B_j + CA_j, \quad 1 \leq j \leq K,
\]

where \( \langle \langle \mathbf{u}, \mathbf{v} \rangle \rangle \) denotes the inner product of the vectors \( \mathbf{u} \) and \( \mathbf{v} \), and \( \mathbf{r} \) denotes the sorted received vector. Note that we do not need to compute the term \( \sum r_i^2 \), since it is irrelevant. As we may observe from the above, the main computational load is the computation of the inner product. This is also the case for Euclidean detection, and we conclude that hybrid detection is only slightly more involved than Euclidean distance detection.
VI. Conclusions

We have presented a hybrid detection method of $q$-ary $n$-length codewords, where the distance measure is a weighted sum of the Euclidean and Pearson distance. The hybrid detector offers the virtue of trading noise margin penalty at nominal channel conditions versus robustness against gain and offset uncertainty. We have computed the error performance of the hybrid detection method in the presence of offset and gain uncertainty and additive noise. We have computed the maximum allowable range of gain and offset versus the penalty in noise margin loss at nominal conditions. We have introduced parity-check $T$-constrained codes, and have shown that for binary codes, $q = 2$, with a code rate $1 - 1/n$, we may gain around 2-3 dB in noise margin with respect to regular $T$-constrained codes. We have discussed implementation issues related to the computational load of the hybrid detector.

References