Maximum Likelihood Decoding for Channels With Gaussian Noise and Signal Dependent Offset

Bu, Renfei; Weber, Jos H.; Schouhamer Immink, Kees A.

DOI
10.1109/TCOMM.2020.3026383

Publication date
2021

Document Version
Final published version

Published in
IEEE Transactions on Communications

Maximum Likelihood Decoding for Channels With Gaussian Noise and Signal Dependent Offset

Renfei Bu, Jos H. Weber, *Senior Member, IEEE*, and Kees A. Schouhamer Immink, *Life Fellow, IEEE*

***Abstract*— In many channels, the transmitted signals do not only face noise, but offset mismatch as well. In the prior art, maximum likelihood (ML) decision criteria have already been developed for noisy channels suffering from *signal independent offset*. In this paper, such ML criterion is considered for the case of binary signals suffering from Gaussian noise and *signal dependent offset*. The signal dependency of the offset signifies that it may differ for distinct signal levels, i.e., the offset experienced by the zeroes in a transmitted codeword is not necessarily the same as the offset for the ones. Besides the ML criterion itself, also an option to reduce the complexity is considered. Further, a brief performance analysis is provided, confirming the superiority of the newly developed ML decoder over classical decoders based on the Euclidean or Pearson distances.**

***Index Terms*— Maximum likelihood decoding, Gaussian noise, offset mismatch, signal dependent offset.**

I. Introduction

THE on-going data revolution demands that data is communicated efficiently, stored reliably, and processed robustly. Tackling the problem of data distortions such as noise, intersymbol interference, offset mismatch, fading, clock jitter, etc., is a fundamental and challenging topic in the theory of channel coding. From these, we are interested in channels with noise and offset mismatch.

Perhaps the best known examples of these channels are flash memories. In flash memories, the number of electrons in a cell decreases with time and some cells become defective over time [1]. The amount of charge leakage, which can be modeled as gain and/or offset mismatch, depends on various physical parameters, such as the device temperature, the magnitude of the charge, and the time elapsed between writing and reading the data [2]. In digital optical recording, fingerprints and scratches on the surface of discs result in offset variations of the retrieved signal [3]. For direct conversion receivers, the local oscillator is the primary source of dc-offset [4].

One would like to understand how these channels with offset mismatch differ from the classical noisy ones. There are conceptual connections, as well as important differences between the noise distortion and the offset mismatch. Both of them are considered to be additive quantities to the transmitted or stored signals. However, noise is a symbol-wise distortion, which is usually independently distributed for each symbol, and thus its value changes symbol by symbol. Offset, on the contrary, is a type of block-wise distortion, which remains constant within one block length, then may change to another value and remains constant for another block length, and so on. As a result, while Euclidean distance-based decoding is known to be optimal if the transmitted or stored signals are only disturbed by Gaussian noise [5], it may perform poorly if there is offset as well.

Various methods have been proposed to address the offset mismatch issues. One approach is accomplished by using pilot sequences to estimate the unknown channel offset [6] and then adjust the detector settings to match the actual values of the channel parameters. A drawback of this method is a rather high redundancy. Up to now, various coding techniques have been applied to alleviate the detection in case of channel mismatch; specifically, rank modulation [7], balanced codes [8], and composition check codes [9]. These methods are often considered to be too expensive in terms of redundancy and complexity. For example, the redundancy of a full set of balanced codewords is $O(\log m)$, where m is the number of user bits [10].

A promising decoding technique with asymptotic zero redundancy as the codeword length increases has been proposed in [11], where it is shown that decoders using the Pearson distance have immunity to offset and/or gain mismatch. A study [12] has shown that a digital modulation transceiver based on Pearson distance detection provides excellent error performance for noisy channels with Rayleigh fading. The use of the Pearson distance requires that the set of codewords satisfies several specific properties. Such sets are called Pearson codes, which have been attracting a lot of interest [13]–[16]. In [13], optimal Pearson codes are presented, in the sense of having the largest number of codewords and thus minimum redundancy among all q-ary Pearson codes of fixed length n. Properties of binary Pearson codes are discussed in [14], where the Pearson noise distance is compared to the well-known Hamming distance. A simple systematic Pearson coding scheme, that maps sequences of information symbols generated by a q-ary source to q-ary code sequences, is proposed in [15]. Construction of a particular kind of Pearson codes, i.e., T-constrained codes [11], using a finite state machine, is introduced in [16].

Furthermore, a considerable amount of literature has grown around the theme of detection schemes that tackle the offset mismatch issues. In [17], a decoder has been proposed based

Manuscript received February 14, 2020; revised July 5, 2020 and September 3, 2020; accepted September 18, 2020. Date of publication September 24, 2020; date of current version January 15, 2021. The associate editor coordinating the review of this article and approving it for publication was I. Tal. *(Corresponding author: Renfei Bu.)*

Renfei Bu and Jos H. Weber are with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: r.bu@tudelft.nl; j.h.weber@tudelft.nl).

Kees A. Schouhamer Immink is with Turing Machines Inc., 3016 DK Rotterdam, The Netherlands (e-mail: immink@turing-machines.com).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2020.3026383

on minimizing a weighted sum of Euclidean and Pearson distances. A dynamic threshold detection (DTD) scheme is proposed in [18], where the gain and offset are first estimated. The estimates of the gain and offset are used to re-scale the received signal within its normal range. Then, the re-scaled signal, brought into its standard range, can be forwarded to the final detection/decoding system, where optimum use can be made of the distance properties of the code by applying, for example, the Chase algorithm. A further improvement of DTD by employing a neural network is investigated in [19]. Novel neural network detectors only need to be invoked when the error correction code decoder fails, or periodically when the system is in the idle state.

The above-summarized methods have improved the resilience to the offset mismatch, or even established immunity to it. However, the price paid for this benefit is a higher noise sensitivity in comparison with the Euclidean decoder. It is important and challenging to study the optimal (ML) decoding solutions considering both noise and offset issues. Blackburn has investigated an ML criterion for channels with Gaussian noise and unknown gain and offset mismatch in [20]. In a subsequent study, ML decoding criteria are derived for Gaussian noise channels when assuming various distributions for the offset in the absence of gain mismatch [21]. In [22], an ML decoder is proposed for the channels with the assumption of bounded noise and offset mismatch. A general framework of ML decoding criteria for such channels is summarized in [23].

A common feature of these prior studies is the assumption that the offset is independent of the signal levels. In this paper, we take a different look, and model the offset mismatch as a signal dependent parameter. Specifically, b_0 is the offset for the '0' signal level and b_1 is the offset for the '1' signal level in the binary case we consider. The signal dependent offset model is appropriate in many scenarios. For example, the binary input user data is stored as the two resistance states of a spin-torque transfer magnetic random access memory (STT-MRAM) cell [24]. A signal dependent offset model is reasonable when process variation causes an asymmetric distribution of both the low and high resistance states. The model is also appropriate for modeling the retention of multilevel-cell phase-change memory, which is adversely affected by resistance states dependent drift and noise [25]. Moreover, degradation of the data reliability can be modelled as a signal dependent offset model, for the situation that with the increase of temperature, the low signal level hardly changes, while the high signal level decreases, leading to a drift of the high signal level to the low signal level [26]. Cai *et al.* [27] have proposed and analyzed a k-means clustering technique as a detection method, for channels where the signal dependent offsets are assumed to be uncorrelated stochastic variables with a uniform probability distribution.

In this paper, we study an ML decoding criterion for channels with Gaussian noise and signal dependent offset. To the best of our knowledge, this is the first paper investigating an ML decoding criterion for such channels. The situation of uniform noise and signal dependent offset is discussed in [28]. The contribution of this work is two-fold. First, an ML criterion for Gaussian distributed noise and signal dependent offsets is derived, where the correlation between different offset random variables is considered as well. Second, we show that an ML criterion in the prior art can be derived as a special case of our decoding criterion that is obtained by letting offsets be identically fully correlated distributed. Further, an option to effectively reduce the complexity is considered for an example with uncorrelated offsets.

The remainder of the paper is organized as follows. In Section II, we review two classical decoding methods and introduce the channel model with noise and signal dependent offset. We present, in Section III, the ML decoding criterion for such channels, where the noise and signal dependent offset are Gaussian distributed. Furthermore, in Section IV, we focus on a special case when the offsets are identical, followed by a complexity discussion in Section V. Finally, a performance evaluation in Section VI and conclusions in Section VII terminate the paper.

II. Noisy Channels With Signal Dependent Offset Mismatch

In this section, we introduce the noisy channel with signal dependent offset mismatch. Further, we review classical decoding criteria that will be compared to the ML decoding criterion to be presented in this paper.

A. Channel Model

Let $\mathcal{Q} = \{0, 1\}$ denote the binary alphabet. We consider transmitting a codeword $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ from a codebook $\mathcal{S} \subseteq \mathcal{Q}^n$. We pursue the binary alphabet since it is the most essential and fundamental case for data transmission or storage. The extension of this channel model to a q-ary alphabet will be investigated in future research.

The transmitted symbols x_i are distorted by additive noise v_i and by signal dependent offsets b_{x_i}. In other words, the received symbols read

$$r_i = x_i + v_i + b_{x_i}, \tag{1}$$

for $i = 1, \ldots, n$. The offset b_{x_i} takes one of two values, b_0 or b_1, depending on the value of x_i. Let

$$\mathbf{b_x} = (b_{x_1}, b_{x_2}, \ldots, b_{x_n})$$

denote the offset vector when $\mathbf{x}$ is transmitted. For example, $\mathbf{b_x} = (b_0, b_1, b_0, b_1)$ if $\mathbf{x} = (0, 1, 0, 1)$. These two values, b_0 and b_1, which neither the transmitter nor the receiver knows, may vary from one transmitted codeword to the next one, but they do not vary within a codeword length of n. The received vector when a codeword $\mathbf{x}$ is transmitted is

$$\mathbf{r} = \mathbf{x} + \mathbf{v} + \mathbf{b_x}, \tag{2}$$

where the noise vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ is such that the v_i are i.i.d. Gaussian random variables with zero mean and variance σ^2, i.e., $v_i \sim \mathcal{N}(0, \sigma^2)$. The probability density function of $\mathbf{v}$ is

$$\phi(\mathbf{v}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-v_i^2/(2\sigma^2)}. \tag{3}$$

Throughout the transmitted codeword, the noise is independent of the offsets. Note that the noise value varies from symbol to symbol, while the offsets b_0 and b_1 are assumed to be constant for all '0' and '1' symbols within a codeword, respectively. We assume that the offsets, b_0 and b_1, are Gaussian distributed with mean 0 and variances β_0^2 and β_1^2, respectively. The probability density functions of b_0 and b_1 are

$$\zeta(b_i) = \frac{1}{\beta_i\sqrt{2\pi}} e^{-b_i^2/(2\beta_i^2)}, \quad (4)$$

where $i = 0, 1$. The correlation between b_0 and b_1 is

$$\rho = \frac{\mathrm{cov}(b_0, b_1)}{\beta_0\beta_1},$$

where $\mathrm{cov}(b_0, b_1)$ is the covariance of b_0 and b_1. The correlation is bounded by $-1 \leq \rho \leq 1$. We have $\rho = 0$ when b_0 and b_1 are uncorrelated, $\rho = 1$ when they are fully correlated, and $\rho = -1$ when they are completely anti-correlated.

B. Decoding Criteria

In this subsection, we discuss two decision criteria for decoding purposes, which are the optimal decoding criteria for variants of channel (2).

1) Euclidean Distance-Based Decoding (ED): A well-known decoding criterion upon receipt of the vector $\mathbf{r}$ is to choose a codeword $\hat{\mathbf{x}} \in \mathcal{S}$, which minimizes the (squared) Euclidean distance between the received vector $\mathbf{r}$ and candidate codeword $\hat{\mathbf{x}}$. In other words, a Euclidean decoder outputs

$$\arg\min_{\hat{\mathbf{x}}\in\mathcal{S}} \delta_E(\mathbf{r}, \hat{\mathbf{x}}),$$

where

$$\delta_E(\mathbf{r}, \hat{\mathbf{x}}) = \sum_{i=1}^{n} (r_i - \hat{x}_i)^2. \quad (5)$$

It is known to be ML with regard to handling Gaussian noise, i.e., $\mathbf{r} = \mathbf{x} + \mathbf{v}$, but not optimal in situations which require resistance towards offset mismatch.

2) Modified Pearson Distance-Based Decoding (PD): Immink and Weber [11] have advocated a modified Pearson distance-based measure instead of the conventional Euclidean distance for improving the resistance towards offset mismatch. Define the vector average for any vector $\mathbf{u} \in \mathbb{R}^n$ as

$$\bar{\mathbf{u}} = \frac{1}{n}\sum_{i=1}^{n} u_i.$$

The modified Pearson distance, $\delta'_P(\mathbf{r}, \hat{\mathbf{x}})$, between the received vector $\mathbf{r}$ and a codeword $\hat{\mathbf{x}} \in \mathcal{S}$, is defined by

$$\delta'_P(\mathbf{r}, \hat{\mathbf{x}}) = \sum_{i=1}^{n} (r_i - \hat{x}_i + \bar{\hat{\mathbf{x}}})^2. \quad (6)$$

Hence, (6) applies the squared Euclidean distance on codewords which are normalized by subtracting their vector average value from each coordinate. A Pearson decoder chooses a codeword minimizing this distance, i.e.,

$$\arg\min_{\hat{\mathbf{x}}\in\mathcal{S}} \delta'_P(\mathbf{r}, \hat{\mathbf{x}}).$$

The use of the Pearson distance demands that the set of codewords satisfies certain special properties. Those that can be used for the decoder based on (6) are investigated in [13].

It can easily be verified that the minimization of $\delta'_P(\mathbf{r}, \hat{\mathbf{x}})$ is immune to offset that does not depend on the signal values. This offset, denoted by b, is constant within a word of length n, but it may vary word by word. In the case where there is no noise and only an unknown offset, which is signal independent, i.e., $\mathbf{r} = \mathbf{x} + b\mathbf{1}$, where $\mathbf{1}$ is the all-one vector, a Pearson decoder is the ML choice. However, in the situation where there is noise as well or the offset is signal dependent, its performance deteriorates [11].

References [20]–[23] have discussed ML decoding criteria for noisy channels with offset, which is the same for all signal levels. In contrast, this paper is concerned with ML decoding methods for noisy channels with signal dependent offset.

III. ML Decoding Criterion for Noisy Channels With Signal Dependent Offset

If a vector $\mathbf{r}$ is received, optimum decoding must determine a codeword $\hat{\mathbf{x}} \in \mathcal{S}$ maximizing $\mathrm{P}(\hat{\mathbf{x}}\,|\mathbf{r})$. If all codewords are equally likely to be sent, then, by Bayes' theorem, this scheme is equivalent to maximizing $\mathrm{P}(\mathbf{r}\,|\hat{\mathbf{x}})$, that is, the probability density value of the received vector $\mathbf{r}$ given the candidate codeword $\hat{\mathbf{x}}$.

Based on the channel model (2), we define the total distortion as

$$\mathbf{d}_{\mathbf{x}} = \mathbf{v} + \mathbf{b}_{\mathbf{x}} = \mathbf{r} - \mathbf{x}.$$

Then the probability density function of $\mathbf{d}_{\mathbf{x}}$ is given by

$$\psi(\mathbf{d}_{\mathbf{x}}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(\mathbf{d}_{\mathbf{x}} - \mathbf{b}_{\mathbf{x}})\zeta(b_0)\zeta(b_1)\,\mathrm{d}b_0\,\mathrm{d}b_1. \quad (7)$$

An ML decoder should maximize $\mathrm{P}(\mathbf{r}\,|\hat{\mathbf{x}})$, which is equivalent to choosing a codeword maximizing the probability density function (7), that is,

$$\arg\max_{\hat{\mathbf{x}}\in\mathcal{S}} \psi(\mathbf{d}_{\hat{\mathbf{x}}}).$$

In the situation of zero-mean Gaussian noise samples and signal dependent offset, $\mathbf{d}_{\hat{\mathbf{x}}}$ has a multivariate Gaussian distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}_{\hat{\mathbf{x}}}$. Since the noise is independent of the offsets and the correlation coefficient between the offset values is ρ, $\mathbf{\Sigma}_{\hat{\mathbf{x}}}$ is an $n \times n$ matrix with the (i, j)th entry specified by

$$\mathbf{\Sigma}_{\hat{\mathbf{x}}}(i,j) = \begin{cases} \beta_0^2 & \text{if } i \neq j, i, j \in \hat{X}^{(0)}, \\ \beta_1^2 & \text{if } i \neq j, i, j \in \hat{X}^{(1)}, \\ \sigma^2 + \beta_0^2 & \text{if } i = j \in \hat{X}^{(0)}, \\ \sigma^2 + \beta_1^2 & \text{if } i = j \in \hat{X}^{(1)}, \\ \rho\beta_0\beta_1 & \text{otherwise}, \end{cases} \quad (8)$$

where $\hat{X}^{(1)}$ and $\hat{X}^{(0)}$ are index sets, indicating the positions of the ones and zeroes in $\hat{\mathbf{x}}$, respectively. Thus, the probability density function of $\mathbf{d}_{\hat{\mathbf{x}}}$ is

$$\psi(\mathbf{d}_{\hat{\mathbf{x}}}) = \frac{\exp(-\mathbf{d}_{\hat{\mathbf{x}}}\mathbf{\Sigma}_{\hat{\mathbf{x}}}^{-1}\mathbf{d}_{\hat{\mathbf{x}}}^{\mathrm{T}}/2)}{\sqrt{(2\pi)^n(\det\mathbf{\Sigma}_{\hat{\mathbf{x}}})}}, \quad (9)$$

where $\mathbf{\Sigma}_{\hat{\mathbf{x}}}^{-1}$ is the inverse matrix of $\mathbf{\Sigma}_{\hat{\mathbf{x}}}$ and $\det\mathbf{\Sigma}_{\hat{\mathbf{x}}}$ is the determinant of $\mathbf{\Sigma}_{\hat{\mathbf{x}}}$.

Before working out this expression, we first introduce some further notations. Let ω denote the weight of $\hat{\mathbf{x}}$, i.e., the size of set $\hat{X}^{(1)}$. Clearly, the size of set $\hat{X}^{(0)}$ is $n - \omega$. According to $\hat{X}^{(1)}$ and $\hat{X}^{(0)}$, we can cluster symbols of the received vector. Specifically, symbols of the received vector at positions where the values of $\hat{x}_i$ are 0 are grouped in one category, and the rest of the symbols form another category. We focus on the average values of these two categories and define two quantities, namely, the average value of received symbols in the '1' positions of $\hat{\mathbf{x}}$, i.e.,

$$\overline{\mathbf{r}^{(1)}} = \frac{1}{\omega} \sum_{i \in \hat{X}^{(1)}} r_i, \tag{10}$$

and the average value of received symbols in the '0' positions of $\hat{\mathbf{x}}$, i.e.,

$$\overline{\mathbf{r}^{(0)}} = \frac{1}{n-\omega} \sum_{i \in \hat{X}^{(0)}} r_i. \tag{11}$$

Finally, α_0 and α_1 are the ratios of noise variance to offset variances, i.e.,

$$\alpha_0 = \sigma^2/\beta_0^2, \tag{12}$$
$$\alpha_1 = \sigma^2/\beta_1^2. \tag{13}$$

When the channels suffer from the signal dependent offsets, the positions of the zeroes and the ones in a candidate codeword $\hat{\mathbf{x}}$ are important factors for the ML decoding. The weight of the codeword is a critical issue as well. Attention is drawn to these factors in the following theorem, where we present the ML decoding criterion for noisy channels with signal dependent offsets.

Theorem 1: In the case that the i.i.d. noise $v_i \sim \mathcal{N}(0, \sigma^2)$, the offsets $b_0 \sim \mathcal{N}(0, \beta_0^2)$, $b_1 \sim \mathcal{N}(0, \beta_1^2)$, and the correlation between b_0 and b_1 is ρ, ML decoding is achieved by minimizing

$$\log\eta + \frac{1}{\sigma^2}\Bigg[\delta_E(\mathbf{r},\hat{\mathbf{x}}) - \frac{\alpha_0 + (1-\rho^2)(n-\omega)}{\eta}\omega^2(\overline{\mathbf{r}^{(1)}}-1)^2 - \frac{\alpha_1 + (1-\rho^2)\omega}{\eta}(n-\omega)^2\overline{\mathbf{r}^{(0)}}^2 - \frac{2\rho\sqrt{\alpha_0\alpha_1}}{\eta}\omega(n-\omega)(\overline{\mathbf{r}^{(1)}}-1)\overline{\mathbf{r}^{(0)}}\Bigg], \tag{14}$$

over all candidate codewords $\hat{\mathbf{x}} \in \mathcal{S}$, where $\eta = \alpha_1\alpha_0 + \omega\alpha_0 + (n-\omega)\alpha_1 + \omega(n-\omega)(1-\rho^2)$.

Before proving this theorem, an example of the covariance matrix (8) is given here.

Example 2: Consider a candidate codeword $\hat{\mathbf{x}} = (1,0,0,1)$ and a received vector $\mathbf{r}$. Then

$$\mathbf{d}_{\hat{\mathbf{x}}} = \mathbf{r} - \hat{\mathbf{x}} = (r_1 - 1, r_2, r_3, r_4 - 1).$$

For this example codeword, we have $\hat{X}^{(1)} = \{1,4\}$ and $\hat{X}^{(0)} = \{2,3\}$. Thus, based on (8), the covariance matrix of $\mathbf{d}_{\hat{\mathbf{x}}}$ is

$$\boldsymbol{\Sigma}_{\hat{\mathbf{x}}} = \begin{bmatrix} \sigma^2+\beta_1^2 & \rho\beta_0\beta_1 & \rho\beta_0\beta_1 & \beta_1^2 \\ \rho\beta_0\beta_1 & \sigma^2+\beta_0^2 & \beta_0^2 & \rho\beta_0\beta_1 \\ \rho\beta_0\beta_1 & \beta_0^2 & \sigma^2+\beta_0^2 & \rho\beta_0\beta_1 \\ \beta_1^2 & \rho\beta_0\beta_1 & \rho\beta_0\beta_1 & \sigma^2+\beta_1^2 \end{bmatrix}. \tag{15}$$

Further, the average value of received symbols in the '1' positions is

$$\overline{\mathbf{r}^{(1)}} = (r_1 + r_4)/2$$

and the average value of received symbols in the '0' positions is

$$\overline{\mathbf{r}^{(0)}} = (r_2 + r_3)/2.$$

Thus the ML measurement for $\hat{\mathbf{x}} = (1,0,0,1)$ according to Theorem 1 is

$$\log\eta + \frac{1}{\sigma^2}\Bigg[\delta_E(\mathbf{r},\hat{\mathbf{x}}) - \frac{\alpha_0 + 2(1-\rho^2)}{\eta}(r_1+r_4-2)^2 - \frac{\alpha_1 + 2(1-\rho^2)}{\eta}(r_2+r_3)^2 - \frac{2\rho\sqrt{\alpha_0\alpha_1}}{\eta}(r_1+r_4-2)(r_2+r_3)\Bigg],$$

where $\eta = \alpha_1\alpha_0 + 2\alpha_0 + 2\alpha_1 + 4(1-\rho^2)$.

We may assume without loss of generality that $\hat{\mathbf{x}}$ is rearranged such that $\hat{x}_1 \geq \hat{x}_2 \geq \ldots \geq \hat{x}_n$, as long as $\mathbf{r}$ and thus $\mathbf{d}_{\hat{\mathbf{x}}} = \mathbf{r} - \hat{\mathbf{x}}$ are rearranged according to the same permutation as $\hat{\mathbf{x}}$. We will do so throughout the rest of this section, since it allows a more convenient representation of the covariance matrix. In the example as just presented, one possible permutation of $\hat{\mathbf{x}}$ is

$$(\hat{x}_1, \hat{x}_4, \hat{x}_2, \hat{x}_3) = (1,1,0,0).$$

Then the corresponding representation of $\mathbf{d}_{\hat{\mathbf{x}}}$ is

$$(r_1 - 1, r_4 - 1, r_2, r_3),$$

and its covariance matrix yields

$$\boldsymbol{\Sigma}_{\hat{\mathbf{x}}} = \left[\begin{array}{cc|cc} \sigma^2+\beta_1^2 & \beta_1^2 & \rho\beta_0\beta_1 & \rho\beta_0\beta_1 \\ \beta_1^2 & \sigma^2+\beta_1^2 & \rho\beta_0\beta_1 & \rho\beta_0\beta_1 \\ \hline \rho\beta_0\beta_1 & \rho\beta_0\beta_1 & \sigma^2+\beta_0^2 & \beta_0^2 \\ \rho\beta_0\beta_1 & \rho\beta_0\beta_1 & \beta_0^2 & \sigma^2+\beta_0^2 \end{array}\right]. \tag{16}$$

Proof of Theorem 1: We start the evaluation of (9) by considering the covariance matrix $\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}$ given in (8). Since the entries of this matrix are assigned values according to two index sets $\hat{X}^{(1)}$ and $\hat{X}^{(0)}$, each $\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}$ is interpreted being subdivided into four blocks, that is,

$$\boldsymbol{\Sigma}_{\hat{\mathbf{x}}} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix},$$

where $\mathbf{A}$ is an $\omega \times \omega$ matrix with all entries on the main diagonal equal to $\sigma^2 + \beta_1^2$ and all other entries equal to β_1^2, $\mathbf{D}$ is an $(n-\omega)\times(n-\omega)$ matrix with all entries on the main diagonal equal to $\sigma^2 + \beta_0^2$ and all other entries equal to β_0^2, and $\mathbf{B}$, $\mathbf{C}$ are matrices of sizes $\omega \times (n-\omega)$ and $(n-\omega)\times\omega$, respectively, with all entries equal to $\rho\beta_0\beta_1$. An example of such a block structure can be found in (16).

If $\mathbf{A}$ and $\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ are non-singular, then the inverse and determinant of $\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}$ are [29, pp. 107-108]

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{X}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{X}^{-1} \\ -\mathbf{X}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{X}^{-1} \end{bmatrix}, \tag{17}$$

and

$$\det\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = (\det\mathbf{A})(\det\mathbf{X}), \tag{18}$$

where $\mathbf{X} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$. We first show that $\mathbf{A}$ and $\mathbf{X}$ are non-singular matrices, and then use the above formulas to calculate the inverse matrix and determinant of $\boldsymbol{\Sigma}_{\hat{\mathbf{x}}}$.

The determinant and the inverse matrix of $\mathbf{A}$ are

$$\det \mathbf{A} = \sigma^{2(\omega-1)}(\sigma^2 + \omega\beta_1^2)$$

and

$$\mathbf{A}^{-1} = \begin{bmatrix} s & t & \cdots & t \\ t & s & & \vdots \\ \vdots & & \ddots & t \\ t & \cdots & t & s \end{bmatrix},$$

where $s = \frac{\sigma^2+(\omega-1)\beta_1^2}{\sigma^2(\sigma^2+\omega\beta_1^2)}$ and $t = \frac{-\beta_1^2}{\sigma^2(\sigma^2+\omega\beta_1^2)}$. $\mathbf{A}$ is a non-singular matrix since $\det \mathbf{A} \neq 0$.

Next we investigate the matrix $\mathbf{X}$. After simple calculations, $\mathbf{X}$ can be written down as

$$\mathbf{X} = \begin{bmatrix} \sigma^2+\gamma & \gamma & \cdots & \gamma \\ \gamma & \sigma^2+\gamma & & \vdots \\ \vdots & & \ddots & \gamma \\ \gamma & \cdots & \gamma & \sigma^2+\gamma \end{bmatrix},$$

where $\gamma = \frac{\alpha_1+(1-\rho^2)\omega}{\alpha_1+\omega}\beta_0^2$. The determinant of $\mathbf{X}$ is

$$\det \mathbf{X} = \frac{\sigma^{2(n-\omega-1)}\beta_0^2\eta}{\alpha_1+\omega},$$

where $\eta = \alpha_0\alpha_1 + \omega\alpha_0 + (n-\omega)\alpha_1 + \omega(n-\omega)(1-\rho^2)$. The matrix $\mathbf{X}$ is singular only when $\eta = 0$. However, $\alpha_0\alpha_1 > 0$ implies that $\eta > 0$ which means that $\mathbf{X}$ is a non-singular matrix. Then the inverse matrix of $\mathbf{X}$ is

$$\mathbf{X}^{-1} = \begin{bmatrix} g & h & \cdots & h \\ h & g & & \vdots \\ \vdots & & \ddots & h \\ h & \cdots & h & g \end{bmatrix},$$

where $g = \frac{\sigma^2+(n-\omega-1)\gamma}{\sigma^2(\sigma^2+(n-\omega)\gamma)}$ and $h = \frac{-\gamma}{\sigma^2(\sigma^2+(n-\omega)\gamma)}$.

Since $\mathbf{A}$ and $\mathbf{X}$ are non-singular, we use (17) and (18) to calculate the determinant and the inverse matrix of $\mathbf{\Sigma}_{\hat{\mathbf{x}}}$. Let γ' be $\frac{\alpha_0+(n-\omega)(1-\rho^2)}{\alpha_0+(n-\omega)}\beta_1^2$. Then we have

$$\mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{X}^{-1}\mathbf{C}\mathbf{A}^{-1} = \begin{bmatrix} g' & h' & \cdots & h' \\ h' & g' & & \vdots \\ \vdots & & \ddots & h' \\ h' & \cdots & h' & g' \end{bmatrix},$$

where $g' = \frac{\sigma^2+(\omega-1)\gamma'}{\sigma^2(\sigma^2+\omega\gamma')}$ and $h' = \frac{-\gamma'}{\sigma^2(\sigma^2+\omega\gamma')}$. We also have that

$$-\mathbf{A}^{-1}\mathbf{B}\mathbf{X}^{-1} \text{ and } -\mathbf{X}^{-1}\mathbf{C}\mathbf{A}^{-1}$$

are two matrices of sizes $\omega \times (n-\omega)$ and $(n-\omega) \times \omega$, respectively, with all entries equal to

$$\frac{-\rho}{\beta_0\beta_1\eta}.$$

Since the logarithm function is strictly increasing on the positive real numbers and ψ is a positive function, ML decoding can also be achieved by maximizing the logarithm of (9), i.e.,

$$\log\psi(\mathbf{d}_{\hat{\mathbf{x}}}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det \mathbf{\Sigma}_{\hat{\mathbf{x}}}) - \frac{1}{2}\mathbf{d}_{\hat{\mathbf{x}}}\mathbf{\Sigma}_{\hat{\mathbf{x}}}^{-1}\mathbf{d}_{\hat{\mathbf{x}}}^{\mathrm{T}},$$

rather than maximizing (9) itself. By inverting the sign and ignoring irrelevant terms, we find that maximizing $\log\psi(\mathbf{d}_{\hat{\mathbf{x}}})$ is equivalent to minimizing

$$\log\left(\det\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}\right) + \sum_{i=1}^{n}\sum_{j=1}^{n} d_i \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1}_{ij} d_j, \quad (19)$$

where d_i is the i-th term in the vector $\mathbf{d}_{\hat{\mathbf{x}}}$. By applying (18), the first part of (19) is

$$\begin{aligned} \log\left(\det\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}\right) &= \log(\det(A)\det(X)) \\ &= \log\eta + \log(\sigma^{2(n-2)}\beta_1^2\beta_0^2). \quad (20) \end{aligned}$$

By ignoring the last term, that is irrelevant to the optimization process (independent of $\hat{\mathbf{x}}$), we have that minimizing (20) is equivalent to minimizing $\log\eta$. The second part of (19) is more complicated, but we can use $g - h = g' - h' = 1/\sigma^2$ to simplify several terms. Since the average value of the first ω symbols of $\underline{\mathbf{d}_{\hat{\mathbf{x}}}}$ is $\overline{\mathbf{r}^{(1)}} - 1$ and the average value of the other symbols is $\overline{\mathbf{r}^{(0)}}$, we have

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{j=1}^{n} d_i \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1}_{ij} d_j \\ &= \left[(g'-h')\sum_{i=1}^{\omega}(d_i)^2 + h'\omega^2(\overline{\mathbf{r}^{(1)}}-1)^2\right] \\ &\quad + \left[(g-h)\sum_{i=\omega+1}^{n}(d_i)^2 + h(n-\omega)^2(\overline{\mathbf{r}^{(0)}})^2\right] \\ &\quad - \frac{2\rho}{\beta_0\beta_1\eta}\sum_{i=1}^{\omega}\sum_{j=\omega+1}^{n} d_i d_j \\ &= \frac{1}{\sigma^2}\Bigg[\delta_E(\mathbf{r},\hat{\mathbf{x}}) - \frac{\alpha_0+(1-\rho^2)(n-\omega)}{\eta}\omega^2(\overline{\mathbf{r}^{(1)}}-1)^2 \\ &\quad - \frac{\alpha_1+(1-\rho^2)\omega}{\eta}(n-\omega)^2\overline{\mathbf{r}^{(0)}}^2 \\ &\quad - \frac{2\rho\sqrt{\alpha_0\alpha_1}}{\eta}\omega(n-\omega)(\overline{\mathbf{r}^{(1)}}-1)\overline{\mathbf{r}^{(0)}}\Bigg]. \quad (21) \end{aligned}$$

Finally, combining (20) and (21), we can conclude that maximizing $\psi(\mathbf{d}_{\hat{\mathbf{x}}})$ is equivalent to minimizing (14), as required. □

In the next section, we take a look at a special case where the offsets are identical.

IV. Special Case Study

We consider a special situation of (2), where b_0 and b_1 are identical, i.e., $b_0 = b_1 = b$. This signal independent offset b is still assumed to be Gaussian distributed with zero mean and variance β^2. By definitions (12) and (13), the ratio of the noise and the offset variances is identical, i.e., $\alpha_0 = \alpha_1 = \alpha = \sigma^2/\beta^2$. The received vector is given by

$$\mathbf{r} = \mathbf{x} + \mathbf{v} + b\mathbf{1}. \quad (22)$$

In Subsection V.A of [21], an ML decoding criterion for channel model (22) is given. Here, we present this criterion in the following corollary and show that it appears as a special case of the result in Theorem 1.

Corollary 3: In the case of i.i.d. noise $v_i \sim \mathcal{N}(0,\sigma^2)$ *and offset* $b \sim \mathcal{N}(0,\beta^2)$*, an ML decoding criterion for channel model* (22) *is achieved by minimizing a weighted* *combination of the Euclidean distance* (5) *and the modified Pearson distance* (6)*, i.e.,*

$$\frac{\alpha}{n+\alpha}\delta_E(\mathbf{r},\hat{\mathbf{x}})+\frac{n}{n+\alpha}\delta'_P(\mathbf{r},\hat{\mathbf{x}}), \tag{23}$$

over all candidate codewords $\hat{\mathbf{x}}$.

Proof: Since the offset values for transmitted zeroes and ones are both equal to b, an ML criterion for (22) can be derived from Theorem 1 by substituting $\rho = 1$ and $\alpha_0 = \alpha_1 = \alpha$, which gives that the expression to be minimized is

$$\log\eta+\frac{1}{\sigma^2}\Bigg[\delta_E(\mathbf{r},\hat{\mathbf{x}})-\frac{\alpha\omega^2}{\eta}(\overline{\mathbf{r^{(1)}}}-1)^2-\frac{\alpha(n-\omega)^2}{\eta}\overline{\mathbf{r^{(0)}}}^2 -\frac{2\alpha\omega(n-\omega)}{\eta}(\overline{\mathbf{r^{(1)}}}-1)\overline{\mathbf{r^{(0)}}}\Bigg]$$
$$=\log\eta+\frac{1}{\sigma^2}\left[\delta_E(\mathbf{r},\hat{\mathbf{x}})-\frac{(\omega(\overline{\mathbf{r^{(1)}}}-1)+(n-\omega)\overline{\mathbf{r^{(0)}}})^2}{\alpha+n}\right]$$
$$=\log\eta+\frac{1}{\sigma^2}\left[\delta_E(\mathbf{r},\hat{\mathbf{x}})-\frac{n^2}{\alpha+n}(\bar{\mathbf{r}}-\bar{\hat{\mathbf{x}}})^2\right],$$

where $\eta = \alpha^2 + n\alpha$. Ignoring the irrelevant term $\log\eta$ and dividing by $1/\sigma^2$ gives $\delta_E(\mathbf{r},\hat{\mathbf{x}})-\frac{n^2}{\alpha+n}(\bar{\mathbf{r}}-\bar{\hat{\mathbf{x}}})^2$. Substituting $n(\bar{\mathbf{r}}-\bar{\hat{\mathbf{x}}})^2=\delta_E(\mathbf{r},\hat{\mathbf{x}})-\delta'_P(\mathbf{r},\hat{\mathbf{x}})+n\bar{\mathbf{r}}^2$, which follows from (5) and (6), and ignoring the irrelevant term $-n^2\bar{\mathbf{r}}^2/(\alpha+n)$, gives (23). □

Recall that the Euclidean decoder (5) is ML in the situation that there is no offset, while the modified Pearson decoder (6) is ML when signals suffer from identical offset. Here, the ML decoding is shown to be a balanced combination of these two criteria. In the offset dominant regime, i.e., $\beta \gg \sigma$ and thus α being very small, (23) essentially reduces to the modified Pearson criterion from (6). On the other hand, in the noise dominant regime, i.e., $\beta \ll \sigma$ and thus α being very large, (23) essentially reduces to the Euclidean criterion (5) [21].

V. COMPLEXITY REDUCTION

Minimization of (14) by an exhaustive search over all candidate codewords $\hat{\mathbf{x}} \in \mathcal{S}$ may be too complex for large codebooks. In [11], it has been shown that the number of computations in order to minimize (6) can be significantly reduced by considering a structured codebook, which is the union of a number of constant composition codes, and applying Slepian's algorithm [30]. Such complexity reductions can be explored for the setting under consideration here as well. Below we will describe an example for a particular case of channel model (2). Similar results can also be obtained for the general case, but the corresponding formulas are more complicated and less readable.

We assume that the offsets b_0 and b_1 are identically distributed with zero means and variances β^2, and that they are uncorrelated random variables, i.e., $\rho = 0$. By setting $\alpha_0 = \alpha_1 = \alpha$ and $\rho = 0$ in Theorem 1, the ML decoding criterion for such a situation is thus established by minimizing

$$\delta_{\text{ML}}(\mathbf{r},\hat{\mathbf{x}}) =\log(\alpha+\omega)+\log(\alpha+n-\omega) +\frac{1}{\sigma^2}\left[\delta_E(\mathbf{r},\hat{\mathbf{x}})-\frac{\omega^2}{\alpha+\omega}(\overline{\mathbf{r}^{(1)}}-1)^2-\frac{(n-\omega)^2}{\alpha+n-\omega}\overline{\mathbf{r}^{(0)}}^2\right], \tag{24}$$

where ω is the weight of the candidate codeword $\hat{\mathbf{x}}$.

Let $\mathcal{S}_\omega$ denote the set as

$$\mathcal{S}_\omega=\{\mathbf{x}\in\mathcal{Q}^n:\sum_{i=1}^n x_i=\omega\},\omega=0,\ldots,n.$$

Note that each of these $\mathcal{S}_\omega$ contains *all* the vectors of length n and a particular weight ω. The codebook $\mathcal{S}$ under consideration is the union of $|V|$ of such sets, i.e.,

$$\mathcal{S}=\bigcup_{\omega\in V}\mathcal{S}_\omega,$$

where the index set $V \subseteq \{0,\ldots,n\}$. Note that the index set V is of size at most $n+1$.

By working out (5), (10), and (11), we obtain

$$\delta_E(\mathbf{r},\hat{\mathbf{x}})=\sum_{i=1}^n(r_i-\hat{x}_i)^2=\sum_{i=1}^n r_i^2+n\bar{\hat{\mathbf{x}}}-2\sum_{i=1}^n r_i\hat{x}_i,$$
$$\overline{\mathbf{r}^{(1)}}=\frac{\sum_{i=1}^n r_i\hat{x}_i}{n\bar{\hat{\mathbf{x}}}},$$

and

$$\overline{\mathbf{r}^{(0)}}=\frac{n\bar{\mathbf{r}}-\sum_{i=1}^n r_i\hat{x}_i}{n-n\bar{\hat{\mathbf{x}}}},$$

where the first equation is the squared Euclidean distance between $\mathbf{r}$ and $\hat{\mathbf{x}}$, and the second and the third equations are the average value of received symbols in the positions where $\hat{\mathbf{x}}$ has ones and zeroes, respectively.

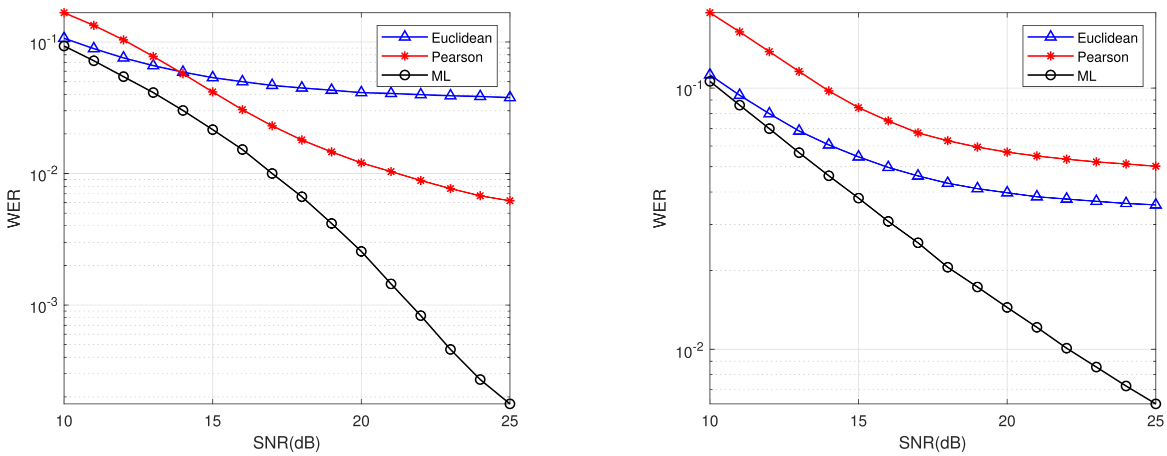
Now, we first investigate the minimization of (24) over all candidate codewords $\hat{\mathbf{x}} \in \mathcal{S}_\omega$ for a fixed value of $\omega \in V$. For any $\hat{\mathbf{x}}$ in $\mathcal{S}_\omega$ we have

$$\delta_{\text{ML}}(\mathbf{r},\hat{\mathbf{x}}) =\log(\alpha+\omega)+\log(\alpha+n-\omega) +\frac{1}{\sigma^2}\Bigg[\sum_{i=1}^n r_i^2+\omega-2\sum_{i=1}^n r_i\hat{x}_i -\frac{\omega^2}{\alpha+\omega}\left(\frac{\sum_{i=1}^n r_i\hat{x}_i}{\omega}-1\right)^2 -\frac{(n-\omega)^2}{\alpha+n-\omega}\left(\frac{n\bar{\mathbf{r}}-\sum_{i=1}^n r_i\hat{x}_i}{n-\omega}\right)^2\Bigg],$$
$$=\log(\alpha+\omega)+\log(\alpha+n-\omega) +\frac{1}{\sigma^2}\left[\sum_{i=1}^n r_i^2+\omega-d(\mathbf{r},\hat{\mathbf{x}})\right], \tag{25}$$

where

$$d(\mathbf{r},\hat{\mathbf{x}})=\left(\frac{1}{\alpha+\omega}+\frac{1}{\alpha+n-\omega}\right)\left(\sum_{i=1}^n r_i\hat{x}_i\right)^2 +\left(\frac{2\alpha}{\alpha+\omega}-\frac{2n\bar{\mathbf{r}}}{\alpha+n-\omega}\right)\sum_{i=1}^n r_i\hat{x}_i +\frac{\omega^2}{\alpha+\omega}+\frac{n^2\bar{\mathbf{r}}^2}{\alpha+n-\omega}. \tag{26}$$

Since ML decoding is a minimization process of (25), we may ignore irrelevant terms and delete scaling constants. Note that ω equals the number of ones in $\hat{\mathbf{x}}$ and thus it does not depend on the specific positions of ones in $\hat{\mathbf{x}}$. The only degree of freedom the decoder has for minimizing (25) is permuting the



(a) $\rho = 0.75$ (b) $\rho = 0.15$

Fig. 1. WER versus SNR of $\mathcal{S}^*$ for channels with Gaussian noise and signal dependent offsets, that have standard deviations $\beta_0 = 0.2$, $\beta_1 = 0.3$, and correlation (a) $\rho = 0.75$; (b) $\rho = 0.15$.

symbols in $\hat{\mathbf{x}}$ to maximize $d(\mathbf{r}, \hat{\mathbf{x}})$. Hence, for the value of ω under consideration, the ML decoding result is

$$\mathbf{x}_{o,\omega} = \arg\max_{\hat{\mathbf{x}} \in \mathcal{S}_\omega} d(\mathbf{r}, \hat{\mathbf{x}}).$$

Note from (26) that $d(\mathbf{r}, \hat{\mathbf{x}})$ can be regarded as a quadratic function of $\sum_{i=1}^n r_i \hat{x}_i$. Since the graph of the quadratic function is a convex parabola, the maximum value of (26) occurs when $\sum_{i=1}^n r_i \hat{x}_i$ is minimal or maximal among all $\hat{\mathbf{x}} \in \mathcal{S}_\omega$. However, which of these two options leads to the maximum value is not apparent from the expression. Therefore, both the maximum and the minimum values of $\sum_{i=1}^n r_i \hat{x}_i$ are considered. Slepian [30] showed that the value of $\sum_{i=1}^n r_i \hat{x}_i$ can be maximized by matching the largest symbol of $\mathbf{r}$ with the largest symbol of $\hat{\mathbf{x}}$, the second largest symbol of $\mathbf{r}$ with the second largest symbol of $\hat{\mathbf{x}}$, etc. On the other hand, the value of $\sum_{i=1}^n r_i \hat{x}_i$ can be minimized by matching the largest symbol of $\mathbf{r}$ with the smallest symbol of $\hat{\mathbf{x}}$, the second largest symbol of $\mathbf{r}$ with the second smallest symbol of $\hat{\mathbf{x}}$, etc. Hence, only two codewords from $\mathcal{S}_\omega$ under consideration need to be evaluated. 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. For every $\omega \in V$, decide $\mathbf{x}_{o,\omega}$ by maximizing the values of $d(\mathbf{r}, \hat{\mathbf{x}} \in \mathcal{S}_\omega)$ over two candidate codewords, $(\underbrace{1, \ldots, 1}_{\omega}, 0, \ldots, 0)$ and $(0, \ldots, 0, \underbrace{1, \ldots, 1}_{\omega})$ for a fixed ω.

For the complete codebook $\mathcal{S}$, ML decoding of the received vector, $\mathbf{r}$, can thus be accomplished by finding $\mathbf{x}_{o,\omega}$ as described above for all $\omega \in V$, and then minimizing $\delta_{\text{ML}}(\mathbf{r}, \mathbf{x}_{o,\omega})$ over the $|V|$ candidates, i.e.,

$$\mathbf{x}_o = \arg\min_{\omega \in V} \ \delta_{ML}(\mathbf{r}, \mathbf{x}_{o,\omega}).$$

The number of codewords to be evaluated in minimizing (24) is $|\mathcal{S}|$, which is impractical for larger $\mathcal{S}$, since it tends to grow exponentially with n. By using the method presented in this section, the number of words that needs to be evaluated is reduced to only $2|V|$, i.e., twice the number of constant composition codes that constitute the codebook $\mathcal{S}$. The index set V is of size at most $n+1$. Thus, the number of evaluations grows at most linearly with n. For large codebooks, this is a significant reduction compared to an exhaustive search among all the candidate codewords. It should be mentioned that in order to apply this method, the received vector needs to be sorted and the resulting ML codeword needs to be inversely permuted accordingly. However, sorting is well known to have only moderate computational complexity in terms of the length of the vector n, e.g., $O(n \log n)$ symbol swaps. We can thus conclude that the overall complexity is still considerably lower than for an exhaustive search.

VI. Performance Evaluation

In this section, we investigate the word error rate (WER) performance of three decoders, namely, Euclidean distance-based decoding (5), Pearson distance-based decoding (6), and ML decoding (14). Simulated WER results are shown for the example codebook

$$\mathcal{S}^* = \{(0,0,0), (1,1,0), (1,0,1), (0,1,1)\}.$$

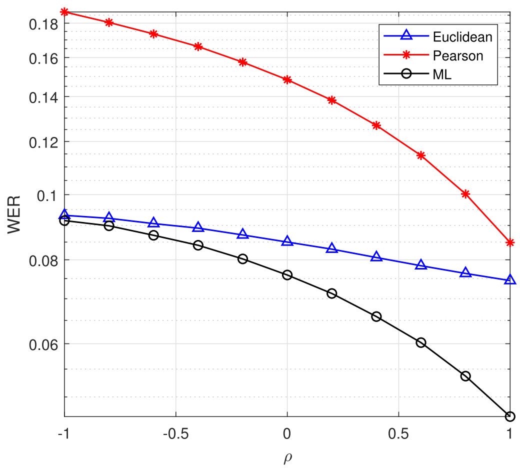
This simple codebook suffices to demonstrate some important WER characteristics of the proposed method (14) in comparison with the traditional methods (5) and (6).

A. WER Versus Signal to Noise Ratio (SNR)

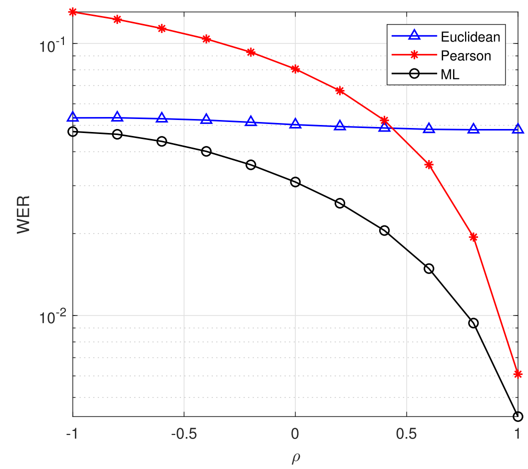
Figure 1 shows the results of simulations for a range of noise levels, where the signal to noise ratio (SNR) is defined as

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

In both Figs. 1a and 1b, each point was the result of 10,000 trials. The signal dependent offsets b_0 and b_1 have zero means and standard deviations $\beta_0 = 0.2$ and $\beta_1 = 0.3$, respectively. The correlation ρ between b_0 and b_1 is set to be 0.75 in Fig. 1a and 0.15 in Fig. 1b. It can be observed from these figures that the performance of Euclidean distance-based decoding is close to ML decoding when the value of SNR is small, for both $\rho = 0.75$ and $\rho = 0.15$. For high correlation, Pearson distance-based decoding outperforms Euclidean distance-based decoding at large SNR values, see Fig. 1a for $\rho = 0.75$. However, for low correlation, the performance of Pearson distance-based decoding is worse than that of Euclidean distance-based decoding, as illustrated in Fig. 1b for $\rho = 0.15$. ML decoding is always better than both of them



(a) SNR = 12 dB



(b) SNR = 17 dB

Fig. 2. WER versus ρ of $\mathcal{S}^*$ for channels with Gaussian noise and signal dependent offsets, that have standard deviations $\beta_0 = 0.2$, $\beta_1 = 0.3$, and signal to noise ratio (a) SNR = 12 dB; (b) SNR = 17 dB.

as expected. By comparing the two figures, it can be seen that the offset correlation, ρ, plays a crucial role in the WER performances. The next subsection is concerned with WER versus ρ.

B. WER Versus Correlation Between Offsets

The simulation results for a range of ρ values are shown in Fig. 2. Each point was the result of 10,000 trials, in the situation that the signal dependent offsets b_0 and b_1 are still assumed to have zero means and standard deviations $\beta_0 = 0.2$ and $\beta_1 = 0.3$, respectively. Here, SNR is set to be 12 dB in Fig. 2a and 17 dB in Fig. 2b.

WERs of all three decoding criteria decrease when ρ increases from -1 to 1. Note that the value of ρ has only little effect on the performance of Euclidean distance-based decoding, but that it has a high impact on the results of Pearson distance-based decoding and ML decoding, especially when SNR is equal to 17 dB. The WER of ML decoding is always better than that of the other two decoders as expected. These results are in accordance with our earlier observations, which showed that Pearson distance-based decoding outperforms Euclidean distance-based decoding when ρ is close to 1 and SNR is large, and that the ML decoding surpasses both of them.

VII. Conclusion

In this paper, we have studied channels that are not only distorted by Gaussian noise, but also by another important channel impairment, *offset*. The offset is assumed to be dependent on the signal levels. A maximum likelihood (ML) decoding criterion has been derived for such channels, to improve and strengthen the resilience to Gaussian noise and signal dependent offset. We have shown that a previous result on ML decoding in the case that the offsets are identical, i.e., signal independent, appears as a special case of our proposed criterion. For codebooks consisting of the union of constant weight sets, it has been shown that significant complexity reductions can be obtained. Finally, the superiority of the presented ML decoder over classical decoders has been illustrated by a brief performance analysis for a simple code.

References

[1] D. Ajwani, I. Malinger, U. Meyer, and S. Toledo, "Characterizing the performance of flash memory storage devices and its impact on algorithm design," in *Proc. Int. Workshop Experim. Efficient Algorithms*. Berlin, Germany: Springer, May 2008, pp. 208–219.

[2] F. Sala, K. A. S. Immink, and L. Dolecek, "Error control schemes for modern flash memories: An overview," *IEEE Consum. Electron.*, vol. 4, no. 1, pp. 66–73, 2015.

[3] G. Bouwhuis, J. Braat, A. Huijser, J. Pasman, G. van Rosmalen, and K. A. S. Immink, in *Principles of Optical Disc Systems*. Boston, MA, USA: Adam Hilger, 1985.

[4] R. Svitek and S. Raman, "DC offsets in direct-conversion receivers: Characterization and implications," *IEEE Microw. Mag.*, vol. 6, no. 3, pp. 76–86, Sep. 2005.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2012.

[6] K. A. S. Immink, "Coding schemes for multi-level flash memories that are intrinsically resistant against unknown gain and/or offset using reference symbols," *Electron. Lett.*, vol. 50, no. 1, pp. 20–22, Jan. 2014.

[7] A. Jiang, R. Mateescu, M. Schwartz, and J. Bruck, "Rank modulation for flash memories," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2659–2673, Jun. 2009.

[8] H. Zhou, J. Anxiao, and J. Bruck, "Balanced modulation for nonvolatile memories," 2012, *arXiv:1209.0744*. [Online]. Available: http://arxiv.org/abs/1209.0744

[9] K. A. S. Immink and K. Cai, "Composition check codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 249–256, Jan. 2018.

[10] J. H. Weber and K. A. Schouhamer Immink, "Knuth's balanced codes revisited," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1673–1679, Apr. 2010.

[11] K. A. S. Immink and J. H. Weber, "Minimum pearson distance detection for multilevel channels with gain and/or offset mismatch," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5966–5974, Oct. 2014.

[12] M. S. Chandan and P. G. Poddar, "Minimum pearson distance based detection for data transmission over fading wireless channels," in *Proc. 2nd IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, May 2017, pp. 2149–2153.

[13] J. H. Weber, K. A. Schouhamer Immink, and S. R. Blackburn, "Pearson codes," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 131–135, Jan. 2016.

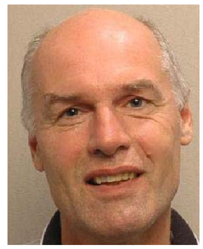
[14] J. H. Weber and K. A. Schouhamer Immink, "Properties of binary pearson codes," in *Proc. Int. Symp. Inf. Theory Appl. (ISITA)*, Oct. 2018, pp. 605–609.

[15] J. H. Weber, T. G. Swart, and K. A. Schouhamer Immink, "Simple systematic pearson coding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 385–389.

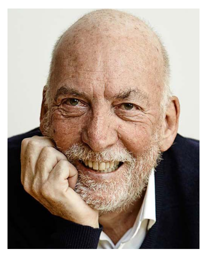
[16] C. Cao and I. Fair, "Capacity-approaching variable-length pearson codes," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1310–1313, Jul. 2018.

[17] K. A. Schouhamer Immink and J. H. Weber, "Hybrid minimum pearson and Euclidean distance detection," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3290–3298, Sep. 2015.
[18] K. A. Schouhamer Immink, K. Cai, and J. H. Weber, "Dynamic threshold detection based on pearson distance detection," *IEEE Trans. Commun.*, vol. 66, no. 7, pp. 2958–2965, Jul. 2018.
[19] Z. Mei, K. Cai, and X. Zhong, "Neural network-based dynamic threshold detection for non-volatile memories," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
[20] S. R. Blackburn, "Maximum likelihood decoding for multilevel channels with gain and offset mismatch," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1144–1149, Mar. 2016.
[21] J. H. Weber and K. A. Schouhamer Immink, "Maximum likelihood decoding for Gaussian noise channels with gain or offset mismatch," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1128–1131, Jun. 2018.
[22] R. Bu and J. H. Weber, "Decoding criteria and zero WER analysis for channels with bounded noise and offset," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
[23] R. Bu and J. H. Weber, "Maximum likelihood decoding for multi-level cell memories with scaling and offset mismatch," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
[24] K. Cai and K. A. S. Immink, "Cascaded channel model, analysis, and hybrid decoding for spin-torque transfer magnetic random access memory," *IEEE Trans. Magn.*, vol. 53, no. 11, pp. 1–11, Nov. 2017.
[25] H. Pozidis *et al.*, "A framework for reliability assessment in multi-level phase-change memory," in *Proc. IEEE Int. Memory Workshop*, May 2012, pp. 143–146.
[26] X. Kou, J. Schmalhorst, A. Thomas, and G. Reiss, "Temperature dependence of the resistance of magnetic tunnel junctions with MgO barrier," *Appl. Phys. Lett.*, vol. 88, no. 21, May 2006, Art. no. 212115.
[27] K. Cai and K. S. Immink, "Detection of noisy and corrupted data using clustering techniques," in *Proc. Int. Symp. Inf. Theory Appl. (ISITA)*, Oct. 2018, pp. 135–138.
[28] R. Bu, J. H. Weber, and K. A. Schouhamer Immink, "Maximum likelihood decoding for channels with uniform noise and signal dependent offset," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 706–710.
[29] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
[30] D. Slepian, "Permutation modulation," *Proc. IEEE*, vol. 53, no. 3, pp. 228–236, Mar. 1965.

Renfei Bu received the B.Sc. degree in communication engineering from Hunan University, in 2015, and the M.Sc. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2017. She is currently pursuing the Ph.D. degree with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, The Netherlands. Her research interests include signal processing and channel coding.



Jos H. Weber (Senior Member, IEEE) received the M.Sc. degree (Hons.) in mathematics and the Ph.D. and M.B.T. degrees from the Delft University of Technology, Delft, The Netherlands, in 1985, 1989, and 1996, respectively. Since 1985, he has been with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, where he is currently an Associate Professor with the Department of Applied Mathematics. He was a Visiting Researcher with the University of California at Davis, CA, USA, the Tokyo Institute of Technology, Japan, the University of Johannesburg, South Africa, EPFL, Switzerland, and SUTD, Singapore. His main research interest includes the area of channel coding.



Kees A. Schouhamer Immink (Life Fellow, IEEE) founded Turing Machines Inc., in 1998, where he currently holds the position of the President. He has contributed to digital video, audio, and data recording products, such as compact disc, DVD, and Blu-ray disc. He received the Personal Emmy Award in 2004, the 2017 IEEE Medal of Honor, the 1999 AES Gold Medal, the 2004 SMPTE Progress Medal, the 2014 Eduard Rhein Prize for Technology, and the 2015 IET Faraday Medal. He received also the Honorary Doctorate from the University of Johannesburg in 2014.