Maximum Likelihood Decoding for Gaussian Noise Channels with Gain or Offset Mismatch

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

DOI
10.1109/LCOMM.2018.2809749

Publication date
2018

Document Version
Accepted author manuscript

Published in
IEEE Communications Letters

Maximum Likelihood Decoding for Gaussian Noise Channels with Gain or Offset Mismatch

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

***Abstract*—Besides the omnipresent noise, other important inconveniences in communication and storage systems are formed by gain and/or offset mismatches. In the prior art, a maximum likelihood (ML) decision criterion has already been developed for Gaussian noise channels suffering from unknown gain and offset mismatches. Here, such criteria are considered for Gaussian noise channels suffering from either an unknown offset or an unknown gain. Furthermore, ML decision criteria are derived when assuming a Gaussian or uniform distribution for the offset in the absence of gain mismatch.**

Index Terms: maximum likelihood decoding, gain mismatch, offset mismatch, non-volatile memories.

I. Introduction

In non-volatile memories, such as floating gate memories, the data is represented by stored charge, which can leak away from the floating gate. This leakage may result in a shift of the threshold voltage of the memory cell. The amount of leakage depends on various physical parameters and, clearly, on the time elapsed between writing and reading the data and the magnitude of the charge. The receiver estimates the mean leakage, but as the estimate is not perfect, a remaining uncertainty that the receiver must account for in the detection algorithm is still present. In the prior art [1], [2], it is assumed that the receiver is completely ignorant of the amount of leakage. Here, however, we will also consider the case that the leakage has a particular distribution.

In general, we can say that dealing with varying offset and/or gain is an important issue in signal processing for modern storage and communication systems. For example, methods to solve these difficulties in flash memories have been discussed in, e.g., [3], [4], and [5]. Also, in optical disc media, the retrieved signal depends on the dimensions of the written features and upon the quality of the light path, which may be obscured by fingerprints or scratches on the substrate, leading to offset and gain variations of the retrieved signal. Throughout this letter, we assume that the offset and gain may change from block to block, but that they do not vary within a block, which typically is the case for applications with relatively small data block lengths. Immink and Weber [1] showed that detectors using the Pearson distance offer immunity to offset and gain mismatch.

Blackburn [2] derived a maximum likelihood criterion for the case that both the gain and the offset are completely unknown, except for the sign of the gain, which is assumed to be positive. Here, ML criteria will be derived for the case that there is no gain mismatch, but there is an unknown offset and vice versa. Unknown in this context means that we do assume a certain range from which the gain/offset takes its values, but that there is no assumption with regard to the probability distribution on this range. To make an assumption on such a distribution could be an appropriate thing to do though, particularly in applications in which the behavior can be predicted to a certain extent. Therefore, we also develop criteria for the cases of Gaussian or uniform offset in the absence of gain mismatch. The main contribution of this letter is a proof that, for Gaussian noise and offset, the maximum likelihood criterion is a weighted average of the well-known Euclidean and Pearson norms, where the weighing coefficients depend on the ratio of the noise and offset variances.

The remainder of this letter is organized as follows. In Section II, we introduce the model under consideration and provide notational convention. Then, in Section III, decision criteria for decoding purposes are discussed. Next, we present maximum likelihood criteria, in Section IV under the assumption that the noise is Gaussian and either the offset or the gain is unknown, and in Section V for the case that the noise and the offset is Gaussian or uniform, while there is no gain mismatch. Finally, in Section VI, we draw conclusions.

II. Model and Notation

We assume a simple channel model, which is not only applicable to flash memories, but also to other communication and storage systems. It reads

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

where $\mathbf{x} = (x_1, \ldots, x_n)$ is the transmitted codeword from a codebook $\mathcal{S} \subset \mathbb{R}^n$, where a uniform distribution is assumed, i.e., all codewords are equally likely to be transmitted, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n) \in \mathbb{R}^n$ is the noise vector, where the ν_i are independently normally distributed with mean 0 and standard deviation σ, a and b are real numbers representing the channel gain and offset, respectively, $\mathbf{1}$ is the real all-one vector $(1, \ldots, 1)$ of length n, and $\mathbf{r} \in \mathbb{R}^n$ is the received vector. It is assumed throughout that the transmitted codeword, noise, gain, and offset are all independent of each other. Note that the noise value varies from symbol to symbol, while the gain and offset values are assumed to be constant for all symbols within a codeword. However, the gain and offset may vary from block to block, i.e., the values of a and b when transmitting a codeword may differ from the values in the previous transmission. In case there is no gain mismatch we fix $a = 1$, while in case there is no offset mismatch we fix $b = 0$.

Manuscript submitted to IEEE Communications Letters dd. Aug. 22, 2017, revision dd. Dec. 4, 2017, 2nd revision dd. Jan. 24, 2018, final version dd. Feb. 23, 2018. Jos H. Weber is with Delft University of Technology, The Netherlands, j.h.weber@tudelft.nl. Kees A. Schouhamer Immink is with Turing Machines Inc., Rotterdam, The Netherlands, immink@turing-machines.com.

For any vector $\mathbf{u} \in \mathbb{R}^n$, let $\overline{\mathbf{u}} = (1/n)\sum_{i=1}^n u_i$ denote the average symbol value, let $\sigma_{\mathbf{u}} = \left(\sum_{i=1}^n (u_i - \overline{\mathbf{u}})^2\right)^{1/2}$ denote the unnormalized symbol standard deviation, and let $\| \mathbf{u} \| = \left(\sum_{i=1}^n (u_i)^2\right)^{1/2}$ denote the norm.

For any two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, let $\langle \mathbf{x}, \mathbf{y} \rangle$ denote the standard inner product of $\mathbf{x}$ and $\mathbf{y}$, i.e.,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i = \| \mathbf{x} \| \cdot \| \mathbf{y} \| \cos\theta, \tag{2}$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$, let $d_{\mathrm{E}}(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^n (x_i - y_i)^2\right)^{1/2}$ be the Euclidean distance between $\mathbf{x}$ and $\mathbf{y}$, and let $d_{\mathrm{P}}(\mathbf{x}, \mathbf{y}) = 1 - \rho_{\mathbf{x},\mathbf{y}}$ be the Pearson distance between $\mathbf{x}$ and $\mathbf{y}$, where $\rho_{\mathbf{x},\mathbf{y}} = \left(\sum_{i=1}^n (x_i - \overline{\mathbf{x}})(y_i - \overline{\mathbf{y}})\right) / (\sigma_{\mathbf{x}} \sigma_{\mathbf{y}})$ is the well-known (Pearson) correlation coefficient. Note that the Pearson distance is not a metric in the strict mathematical sense, but in engineering parlance it is still called a 'distance' since it provides a useful measure of similarity between vectors.

III. Decision Criteria

A general decoding technique upon receipt of the vector $\mathbf{r}$ is to choose as the decoder output $\mathbf{x}_{\mathrm{o}}$ the codeword $\hat{\mathbf{x}} \in \mathcal{S}$ which maximizes some probability. This is then often translated into minimizing a distance-based decision criterion $L(\mathbf{r}, \hat{\mathbf{x}})$, i.e., $\mathbf{x}_{\mathrm{o}} = \operatorname{argmin}_{\hat{\mathbf{x}} \in \mathcal{S}} L(\mathbf{r}, \hat{\mathbf{x}})$. Two criteria are said to be equivalent if for any $\mathbf{r}$ a codeword optimizing the one also optimizes the other.

Known choices for $L(\mathbf{r}, \hat{\mathbf{x}})$ are based on the (squared) Euclidean distance $d_{\mathrm{E}}(\mathbf{x}, \mathbf{y})$, i.e.,

$$L^{\mathrm{E}}(\mathbf{r}, \hat{\mathbf{x}}) = \sum_{i=1}^n (r_i - \hat{x}_i)^2 = \| \mathbf{r} - \hat{\mathbf{x}} \|^2, \tag{3}$$

or on the Pearson distance $d_{\mathrm{P}}(\mathbf{x}, \mathbf{y})$, i.e.,

$$L^{\mathrm{P}}(\mathbf{r}, \hat{\mathbf{x}}) = \sum_{i=1}^n \left(r_i - \frac{\hat{x}_i - \overline{\hat{\mathbf{x}}}}{\sigma_{\hat{\mathbf{x}}}} \right)^2. \tag{4}$$

Actually, the latter expression is not equal to the Pearson distance between $\mathbf{r}$ and $\hat{\mathbf{x}}$, but, as shown in [1], minimizing (4) is equivalent to minimizing $d_{\mathrm{P}}(\mathbf{r}, \hat{\mathbf{x}})$. It was also shown in [1], that in case there is no gain mismatch, i.e., $a = 1$, a suitable simpler decision criterion is obtained by removing the division by $\sigma_{\hat{\mathbf{x}}}$ from (4), i.e.,

$$L^{\mathrm{P}'}(\mathbf{r}, \hat{\mathbf{x}}) = \sum_{i=1}^n \left(r_i - \hat{x}_i + \overline{\hat{\mathbf{x}}} \right)^2. \tag{5}$$

Of particular interest from a performance perspective are maximum likelihood (ML) decoders, which choose the codeword of maximum probability given the received vector. Since we assume that all codewords are equally likely, it follows from Bayes' rule that this is equivalent to maximizing the probability density value $f(\mathbf{r}|\hat{\mathbf{x}})$ of the received vector given the codeword, i.e., $\mathbf{x}_{\mathrm{o}} = \operatorname{argmax}_{\hat{\mathbf{x}} \in \mathcal{S}} f(\mathbf{r}|\hat{\mathbf{x}})$. Taking the logarithm and inverting the sign we obtain the frequently-used equivalent decision rule $\mathbf{x}_{\mathrm{o}} = \operatorname{argmin}_{\hat{\mathbf{x}} \in \mathcal{S}} -\log(f(\mathbf{r}|\hat{\mathbf{x}}))$. Hence, we have

$$L^{\mathrm{ML}}(\mathbf{r}, \hat{\mathbf{x}}) = -\log(f(\mathbf{r}|\hat{\mathbf{x}})).$$

In case the gain a and offset b are known, while the noise is normally distributed with mean zero and variance σ^2, an ML decoder will choose the codeword $\hat{\mathbf{x}}$ that maximizes $\phi((\mathbf{r} - b\mathbf{1})/a - \hat{\mathbf{x}})$, where

$$\phi(\boldsymbol{\nu}) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\nu_i^2/(2\sigma^2)},$$

or, equivalently, that minimizes $\sum_{i=1}^n \left(\frac{r_i - b}{a} - \hat{x_i}\right)^2 = L^{\mathrm{E}}((\mathbf{r} - b\mathbf{1})/a, \hat{\mathbf{x}})$. This gives the well-known fact that, in case both the gain and the offset are known by the receiver, the Euclidean criterion from (3) is ML when applying fixed shifting and scaling operations on the received vector before using the criterion. On the other hand, if there are unknown gain and offset, it has been demonstrated that the Pearson criterion from (4) may perform much better than the Euclidean distance between $\mathbf{r}$ and $\hat{\mathbf{x}}$ [1].

IV. Unknown Gain or Offset

Blackburn [2] investigated the case that both the gain a and the offset b are fully unknown, except for the sign of the gain, which is assumed to be positive. Since knowledge of the offset and gain is lacking, the best thing to do, upon receipt of a vector $\mathbf{r}$, is to set, for any candidate codeword $\hat{\mathbf{x}}$, the gain and the offset in such a way that the resulting noise is minimized. Hence, in order to achieve ML decoding, the criterion to be maximized over all codewords is $\max_{a,b \in \mathbb{R}: a>0} \phi((\mathbf{r} - b\mathbf{1})/a - \hat{\mathbf{x}})$, which was shown to be equivalent to the minimization of the explicit criterion

$$L^{\mathrm{B}}(\mathbf{r}, \hat{\mathbf{x}}) = \begin{cases} \sigma_{\hat{\mathbf{x}}}^2 (1 - \rho_{\mathbf{r},\hat{\mathbf{x}}}^2) & \text{if } \rho_{\mathbf{r},\hat{\mathbf{x}}} > 0, \\ \sigma_{\hat{\mathbf{x}}}^2 & \text{otherwise.} \end{cases}$$

In this section, we present similar criteria for other important cases with regard to the assumptions on the gain a and offset b. In particular, we consider the situations in which there is either no gain mismatch or no offset mismatch.

A. Unknown offset, no gain mismatch

First, we assume there is no gain mismatch, i.e., $a = 1$, but there is an unknown offset b. For the offset we assume that it takes its values within a certain range, specifically $b_1 \leq b \leq b_2$, but we do not make any further assumptions on the distribution on this interval. In this case, the criterion to maximize is

$$\max_{b \in \mathbb{R}: b_1 \leq b \leq b_2} \phi(\mathbf{r} - b\mathbf{1} - \hat{\mathbf{x}}), \tag{6}$$

which leads to the following explicit result.

Theorem 1. *In case $a = 1$ and the unknown offset b is assumed to be restricted to a range $b_1 \leq b \leq b_2$, ML decoding is achieved by minimizing*

$$L^{\mathrm{MLb}}(\mathbf{r}, \hat{\mathbf{x}}) = \begin{cases} L^{\mathrm{E}}(\mathbf{r} - b_1\mathbf{1}, \hat{\mathbf{x}}) & \textit{if } \overline{\mathbf{r}} - \overline{\hat{\mathbf{x}}} < b_1, \\ L^{\mathrm{E}}(\mathbf{r} - b_2\mathbf{1}, \hat{\mathbf{x}}) & \textit{if } \overline{\mathbf{r}} - \overline{\hat{\mathbf{x}}} > b_2, \\ L^{\mathrm{E}}(\mathbf{r} - (\overline{\mathbf{r}} - \overline{\hat{\mathbf{x}}})\mathbf{1}, \hat{\mathbf{x}}) & \textit{otherwise.} \end{cases}$$

Proof: Note that maximizing (6) is equivalent to minimizing the smallest squared Euclidean distance from the codeword $\hat{\mathbf{x}}$

to $\mathcal{L} = \{\mathbf{r} - c\mathbf{1} | b_1 \leq c \leq b_2\}$, which is a piece of the line $\mathcal{L}' = \{\mathbf{r} - c\mathbf{1} | c \in \mathbb{R}\}$ in $\mathbb{R}^n$. The point on $\mathcal{L}'$ closest to $\hat{\mathbf{x}}$ is $\mathbf{p} = \mathbf{r} - (\overline{\mathbf{r}} - \overline{\hat{\mathbf{x}}})\mathbf{1}$. Hence, the point on $\mathcal{L}$ closest to $\hat{\mathbf{x}}$ is $\mathbf{r} - b_1\mathbf{1}$ if $\overline{\mathbf{r}} - \overline{\hat{\mathbf{x}}} < b_1$, $\mathbf{r} - b_2\mathbf{1}$ if $\overline{\mathbf{r}} - \overline{\hat{\mathbf{x}}} > b_2$, and $\mathbf{p}$ otherwise, which gives the stated result. ∎

By letting $b_1 \to -\infty$ and $b_2 \to \infty$, we obtain the criterion

$$L^{\mathrm{E}}(\mathbf{r} - (\overline{\mathbf{r}} - \overline{\hat{\mathbf{x}}})\mathbf{1}, \hat{\mathbf{x}}) = \sum_{i=1}^{n} (r_i - \hat{x}_i + \overline{\hat{\mathbf{x}}})^2 - n\overline{\mathbf{r}}^2$$

for the case there is no knowledge of the offset at all (just $b \in \mathbb{R}$). Since the last term is irrelevant in the optimization process, we conclude that this expression is equivalent to the (modified) Pearson criterion $L^{\mathrm{P}'}(\mathbf{r}, \hat{\mathbf{x}})$ from (5), which thus achieves ML decoding in this case.

B. Unknown gain, no offset mismatch

Next, we assume there is no offset mismatch ($b = 0$) but there is a gain a, of which we only assume that it is within the range $0 < a_1 \leq a \leq a_2$. In this case, the criterion to maximize is $\max_{a \in \mathbb{R}: a_1 \leq a \leq a_2} \phi(\mathbf{r}/a - \hat{\mathbf{x}})$, which leads to the following explicit result.

Theorem 2. *In case $b = 0$ and the unknown gain a is assumed to be restricted to a range $0 < a_1 \leq a \leq a_2$, ML decoding is achieved by minimizing*

$$L^{\mathrm{MLa}}(\mathbf{r}, \hat{\mathbf{x}}) = \begin{cases} L^{\mathrm{E}}(\mathbf{r}/a_1, \hat{\mathbf{x}}) & \text{if } \langle \mathbf{r}, \hat{\mathbf{x}} \rangle > \frac{1}{a_1} \parallel \mathbf{r} \parallel^2, \\ L^{\mathrm{E}}(\mathbf{r}/a_2, \hat{\mathbf{x}}) & \text{if } \langle \mathbf{r}, \hat{\mathbf{x}} \rangle < \frac{1}{a_2} \parallel \mathbf{r} \parallel^2, \\ \parallel \hat{\mathbf{x}} \parallel^2 - \left(\frac{\langle \mathbf{r}, \hat{\mathbf{x}} \rangle}{\parallel \mathbf{r} \parallel} \right)^2 & \text{otherwise.} \end{cases}$$

Proof: Note that maximizing $\max_{a \in \mathbb{R}: a_1 \leq a \leq a_2} \phi(\mathbf{r}/a - \hat{\mathbf{x}})$ is equivalent to minimizing the smallest squared Euclidean distance d^2 from the codeword $\hat{\mathbf{x}}$ to $\mathcal{L} = \{c\mathbf{r} | 1/a_2 \leq c \leq 1/a_1\}$, which is a piece of the line $\mathcal{L}' = \{c\mathbf{r} | c \in \mathbb{R}\}$ in $\mathbb{R}^n$. Let θ be the angle between $\hat{\mathbf{x}}$ and $\mathbf{r}$. The point on $\mathcal{L}'$ closest to $\hat{\mathbf{x}}$ is $\mathbf{p} = p\mathbf{r}$ with $p = (\parallel \hat{\mathbf{x}} \parallel \cos\theta) / \parallel \mathbf{r} \parallel = \langle \mathbf{r}, \hat{\mathbf{x}} \rangle / \parallel \mathbf{r} \parallel^2$, where the latter equality follows from (2). Hence, the point on $\mathcal{L}$ closest to $\hat{\mathbf{x}}$ is $\mathbf{r}/a_1$ if $p > 1/a_1$, $\mathbf{r}/a_2$ if $p < 1/a_2$, and $\mathbf{p}$ otherwise. This implies that the smallest squared Euclidean distance from $\hat{\mathbf{x}}$ to $\mathcal{L}$ is as given by $L^{\mathrm{MLa}}(\mathbf{r}, \hat{\mathbf{x}})$, where the expression in the "otherwise" case follows from $d^2 = d_{\mathrm{E}}^2(\hat{\mathbf{x}}, \mathbf{p}) = \parallel \hat{\mathbf{x}} \parallel^2 \sin^2\theta = \parallel \hat{\mathbf{x}} \parallel^2 (1 - \cos^2\theta) = \parallel \hat{\mathbf{x}} \parallel^2 - \left(\frac{\langle \mathbf{r}, \hat{\mathbf{x}} \rangle}{\parallel \mathbf{r} \parallel} \right)^2$, which concludes the proof. ∎

V. Offset with a Known Distribution

Thus far, we have assumed that the gain and offset mismatches are either absent (i.e., $a = 1$ and/or $b = 0$) or completely unknown, except for the ranges from which they can take their values. However, a more practical assumption may be that the receiver has some knowledge about the amount of gain and/or offset to be expected. Therefore, it is useful to also consider scenarios in which we assume certain distributions for the gain and/or offset. In this section, we will do so for the case that there is no gain mismatch. Hence, we set $a = 1$ and let b have a specified probability density function ζ with mean μ and variance β^2, while the noise values ν_i will still be assumed to be Gaussian distributed with mean 0 and variance σ^2. Let α denote the ratio of the noise and offset variances, i.e., $\alpha = \sigma^2/\beta^2$. Since a receiver can subtract $\mu\mathbf{1}$ from $\mathbf{r}$ in case the expected offset value μ is not equal to zero, we may assume $\mu = 0$ without loss of generality, which we will do throughout the rest of this section.

Since $a = 1$, the model from (1) reduces to $\mathbf{r} = \mathbf{x} + \boldsymbol{\nu} + b\mathbf{1} = \mathbf{x} + \mathbf{d}$, where $\mathbf{d} = \boldsymbol{\nu} + b\mathbf{1}$ can be seen as the total disturbance. The probability density function of $\mathbf{d}$ is denoted by ψ, which satisfies

$$\psi(\mathbf{d}) = \int_{-\infty}^{\infty} \phi(\mathbf{d} - b\mathbf{1}) \zeta(b) \mathrm{d}b, \tag{7}$$

i.e., it is the convolution of the probability density functions of the noise and the offset. In order to achieve ML decoding, we need to maximize $\psi(\mathbf{r} - \hat{\mathbf{x}})$ over all candidate codewords. Next, we will investigate this criterion in case the offset is assumed to be Gaussian or uniform.

A. Gaussian Offset

When investigating $\psi(\mathbf{r} - \hat{\mathbf{x}})$ in case of independent zero-mean Gaussian noise samples with variance σ^2 and Gaussian offset with mean 0 and variance β^2, note that the disturbance $\mathbf{d} = \boldsymbol{\nu} + b\mathbf{1}$ has then a multivariate Gaussian distribution with mean vector $\mathbf{0}$ and covariance matrix S, where S is the $n \times n$ matrix with all entries on the main diagonal equal to $\sigma^2 + \beta^2$ and all other entries equal to β^2. Hence, the probability density function of the disturbance is $\psi(\mathbf{d}) = \frac{\exp\left(-\mathbf{d}\mathrm{S}^{-1}\mathbf{d}^{\mathrm{T}}/2\right)}{\sqrt{(2\pi)^{\mathrm{n}}|\mathrm{S}|}}$, where S^{-1} is the inverse matrix of S, which is found to be the $n \times n$ matrix with all entries on the main diagonal equal to g and all other entries equal to h, where

$$g = \frac{\sigma^2 + (n-1)\beta^2}{\sigma^2(\sigma^2 + n\beta^2)} \text{ and } h = \frac{-\beta^2}{\sigma^2(\sigma^2 + n\beta^2)}. \tag{8}$$

Remember that the Euclidean criterion from (3) is ML in case there is no offset mismatch ($b = 0$), while the (modified) Pearson criterion from (5) is ML in case the offset is completely unknown. In the next theorem, we show that the ML criterion in case the offset has a normal distribution is in fact a weighted average of these two criteria. A hybrid method using a combination of the Euclidean and Pearson measures for detection purposes was already studied in [6] in a heuristic way. Here, we present the optimal balance between the two measures for a Gaussian offset.

Theorem 3. *In case $a = 1$ and the offset b is assumed to be normally distributed, ML decoding is achieved by minimizing*

$$\frac{\alpha}{n + \alpha} L^{\mathrm{E}}(\mathbf{r}, \hat{\mathbf{x}}) + \frac{n}{n + \alpha} L^{\mathrm{P}'}(\mathbf{r}, \hat{\mathbf{x}}). \tag{9}$$

Proof: By taking the logarithm of $\psi(\mathbf{r} - \hat{\mathbf{x}})$, inverting the sign, and ignoring irrelevant terms and factors, we find that maximizing this function is equivalent to minimizing

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (r_i - \hat{x}_i) (S^{-1})_{i.j} (r_j - \hat{x}_j) =$$

$$g \sum_{i=1}^{n} (r_i - \hat{x}_i)^2 + h \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} (r_i - \hat{x}_i)(r_j - \hat{x}_j) =$$

$$(g - h) \sum_{i=1}^{n} (r_i - \hat{x}_i)^2 + hn^2 (\overline{\mathbf{r}} - \overline{\hat{\mathbf{x}}})^2.$$

TABLE I
WER FOR $\mathcal{S}^*$ WITH VARIOUS SETTINGS FOR THE GAUSSIAN NOISE AND GAUSSIAN OFFSET AND VARIOUS DECODERS.

σ (noise)	β (offset)	α (σ^2/β^2)	(3) (Euclidean)	(5) (Pearson)	(9) (ML)
0.2	1	0.04	0.318	0.031	0.029
0.2	0.2	1	0.026	0.031	0.009
0.3	0.2	2.25	0.064	0.130	0.054
0.3	0.01	900	0.025	0.130	0.025

TABLE II
WER FOR $\mathcal{S}^*$ WITH VARIOUS SETTINGS FOR THE GAUSSIAN NOISE AND UNIFORM OFFSET AND VARIOUS DECODERS.

σ (noise)	β (offset)	(3) (Eucl.)	(5) (Pear.)	(9) ($\mathrm{ML_{Gauss}}$)	(10) (Th. 1)	(11) (ML)
0.2	1	0.362	0.031	0.029	0.027	0.026
0.2	0.2	0.019	0.031	0.009	0.010	0.008
0.3	0.2	0.064	0.130	0.054	0.062	0.054
0.3	0.01	0.025	0.130	0.025	0.025	0.025

Dividing by $g-h$ and substituting the g and h values from (8) gives $L^{\mathrm{E}}(\mathbf{r},\hat{\mathbf{x}}) - \frac{n^2}{\alpha+n}(\overline{\mathbf{r}}-\overline{\hat{\mathbf{x}}})^2$. Substituting $n(\overline{\mathbf{r}}-\overline{\hat{\mathbf{x}}})^2 = L^{\mathrm{E}}(\mathbf{r},\hat{\mathbf{x}}) - L^{\mathrm{P}'}(\mathbf{r},\hat{\mathbf{x}})$, which follows from (5) when adding an irrelevant term $n\overline{\mathbf{r}}^2$, gives (9). ∎

Note that in the offset dominant regime, i.e., $\beta \gg \sigma$ and thus α being very small, (9) essentially reduces to the modified Pearson criterion $L^{\mathrm{P}'}(\mathbf{r},\hat{\mathbf{x}})$ from (5). On the other hand, note that in the noise dominant regime, i.e., $\beta \ll \sigma$ and thus α being very large, (9) essentially reduces to the Euclidean criterion $L^{\mathrm{E}}(\mathbf{r},\hat{\mathbf{x}})$. Furthermore, it can be observed that for any value of α, the criterion tends more and more towards the offset-resistant Pearson distance when n is increasing. This can be explained by noting that the standard deviation per dimension is $\sqrt{n\sigma^2}/n = \sigma/\sqrt{n}$ for the noise, while it is constant at β for the offset.

The above-mentioned findings are illustrated in Table I, where simulated word error rate (WER) results are shown for the codebook $\mathcal{S}^* = \{(0,0,0),(1,1,0),(1,0,1),(0,1,1)\}$ of length $n=3$ and size 4, in combination with different decoders and various choices for the noise and offset standard deviations. Note that in case neither the noise nor the offset is strongly dominating the other, the ML decoder from (9) is clearly outperforming both the Euclidean decoder and the Pearson decoder.

B. Uniform Offset

In case the offset has a non-Gaussian distribution, it may be difficult to convert $\psi(\mathbf{r}-\hat{\mathbf{x}})$ into an explicit distance measure. As an alternative criterion to be considered, we can take one in the spirit of the previous section. Since we now know the offset distribution ζ, we change (6) into

$$\max_{b:\zeta(b)>0} \zeta(b)\phi(\mathbf{r}-b\mathbf{1}-\hat{\mathbf{x}}). \tag{10}$$

If ζ is Gaussian, then maximizing this criterion is equivalent to minimizing (9), hence it is ML. However, this may not be the case for other offset distributions. If ζ is uniform with mean 0 and standard deviation β, i.e., the offset is uniformly distributed on the interval $[-\beta\sqrt{3},\beta\sqrt{3}]$, then maximizing (10) is equivalent to minimizing the criterion from Theorem 1 with $b_1 = -\beta\sqrt{3}$ and $b_2 = \beta\sqrt{3}$. Remember that this is ML in case of unknown offset, since setting the offset such that the noise is minimized is obviously the best thing to do for the codeword selection process in this situation. But if we assume that the offset is uniformly distributed, then we need to choose upon receipt of a vector $\mathbf{r}$ a codeword $\hat{\mathbf{x}}$ such that $\psi(\mathbf{r}-\hat{\mathbf{x}})$ is maximized in order to achieve ML decoding. As can be seen from (7), this means that the *average* noise should be minimized, which may lead to another outcome than the result based on (10). In conclusion, it follows that maximizing

$$\int_{-\beta\sqrt{3}}^{\beta\sqrt{3}} \phi(\mathbf{r}-b\mathbf{1}-\hat{\mathbf{x}})\mathrm{d}b \tag{11}$$

over all codewords $\hat{\mathbf{x}}$ achieves ML decoding in case of a uniformly distributed offset.

Numerical illustrations are provided in Table II, where simulated WER results are shown, again for $\mathcal{S}^* = \{(0,0,0),(1,1,0),(1,0,1),(0,1,1)\}$, in combination with different decoders and various choices for the noise and offset standard deviations. Note that the decoders from (9) and (10) both show close to ML performance for the considered cases.

VI. CONCLUSIONS

Maximum likelihood decision criteria for Gaussian noise channels with gain and/or offset mismatch have been presented for four cases: (i) known gain and unknown offset, (ii) unknown gain and known offset, (iii) known gain and Gaussian distributed offset, and (iv) known gain and uniformly distributed offset. Most noteworthy, it was found that, in case (iii), the ML criterion is a weighted average of the (squared) Euclidean distance and the (modified) Pearson distance between the received vector and the candidate codeword. The weighing coefficients depend on the ratio of the variances of the noise and the offset. They reflect the trade-off between the immunity to offset mismatch of Pearson distance based detection and the higher noise resistance of Euclidean distance based detection.

REFERENCES

[1] K. A. S. Immink and J. H. Weber, "Minimum Pearson Distance Detection for Multi-Level Channels with Gain and/or Offset Mismatch", *IEEE Trans. Inf. Theory,* vol. 60, no. 10, pp. 5966–5974, Oct. 2014.

[2] 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, March 2016.

[3] 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, June 2009.

[4] F. Sala, K. A. S. Immink, and L. Dolecek, "Error Control Schemes for Modern Flash Memories: Solutions for Flash Deficiencies", *IEEE Cons. Electron. Mag.,* vol. 4, no.1, pp. 66–73, Jan. 2015.

[5] H. Zhou, A. Jiang, and J. Bruck, "Error-Correcting Schemes with Dynamic Thresholds in Nonvolatile Memories", *Proceedings IEEE Int. Symp. on Inf. Theory (ISIT)*, St. Petersburg, Russia, pp. 2143–2147, July 2011.

[6] K. A. S. Immink and J. H. Weber, "Hybrid Minimum Pearson and Euclidean Distance Detection", *IEEE Trans. Commun.,* vol. 63, no. 9, pp. 3290–3298, Sept. 2015.