DISTANCE DISTRIBUTION FOR TURBO-EQUALIZED SYSTEMS OVER STATIC FREQUENCY-SELECTIVE CHANNELS

Abdel-Nasser Assimi, Charly Poulliat, Inbar Fijalkow

ETIS, UMR 8051
ENSEA - University of Cergy-Pontoise - CNRS
6, avenue du Ponceau, 95014 Cergy-Pontoise, France
Email: {abdelnasser.assimi, charly.poulliat, inbar.fijalkow}@ensea.fr

ABSTRACT

In this paper we investigate the Euclidean distance distribution in turbo-equalized systems over static frequency-selective (ISI) channels [1]. We propose a novel approach in evaluating the squared Euclidean distance between transmitted sequences at the output of the ISI channel based on the correlogram of error sequences. By inspecting autocorrelation properties of error sequences, we derive the main statistics of the output Euclidean distance. The proposed method provides a more comprehensive tool to predict system behavior with finite or infinite packet length. We exploit obtained results to evaluate the frame error rate (FER) performance of the system under maximum-likelihood sequence estimation.

Index Terms— turbo-equalization, maximum-likelihood detection, Euclidean distance distribution, intersymbol interference.

1. INTRODUCTION

In digital packet communication systems over frequency-selective channels, turbo-equalization [1] is an efficient technique that combines signal detection and forward error correction (FEC) in iterative scheme leading to considerable gains in intersymbol interference (ISI) mitigation in comparison with systems using separated signal detection and correction. Finite length system performance can be predicted by analytic assessment of the corresponding maximum likelihood (ML) receiver. The performance of the maximum likelihood sequence estimation for ISI channels have been first investigated by Forney [2] for non coded transmission, where an upper bound (often referenced as the Union Bound) was derived based on the Euclidean distance distribution at the output of the ISI channel. This distribution is estimated using a trellis-based approach using the state diagram of the channel for a non coded transmission to calculate single error events weights and their associated multiplicities. The same principle was used in [3] by treating the ISI channel of length L as a concatenation of a rate $1/L$ convolutional encoder with a memoryless nonlinear mapper. In [4], a method is proposed to upperbound the performance of a serially concatenated turbo-coded system assuming uniform interleaving. This approach has been further extended by several authors and applied to the case of turbo-equalized systems [5], [6] in order to evaluate system performance over ISI channels. In [5], the performance analysis of turbo-equalized system was studied over special type of ISI channels which are the partial response channels. In [6], the authors tried to account for general ISI channels in the derivation of the Euclidian distance spectrum using an enumeration technique. They have efficiently applied their method to 2-taps and 3-taps channels, but the extension to channels with more than 3-taps appears as computationally prohibitive. In this paper, we address the problem of the evaluation and the characterization of the distance spectrum for general ISI channels for ML performance evaluation of turbo-equalized systems. To this end, we propose a new method for the evaluation of the Euclidean distance distribution based on the evaluation of the autocorrelation of the error sequence. Using this new approach, we derive a new upper bound on the FER performance of the ML receiver for a turbo-equalized system.

[1] This work was supported by the project "Urbanisme des radio Communication" of the program research "Pôle de compétitivité SYSTEM@TIC".

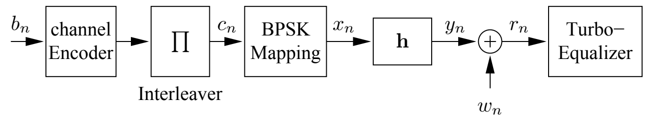

Fig. 1. Turbo-equalized transmission system model

2. SYSTEM MODEL AND MAXIMUM-LIKELIHOOD DECODING

We consider the communication system model in Fig. 1. A sequence of K information bits b_n is encoded into a sequence of N coded bits using rate $R = K/N$ error correcting code. After random interleaving, the interleaved sequence c_n is mapped into a sequence of symbols x_n using BPSK modulation. The modulated symbols x_n are transmitted through a static ISI channel with equivalent discrete time finite impulse response (FIR) given by $\mathbf{h} = (h_0, ..., h_{L-1})$ where all coefficients are assumed to be real-valued. The received signal is modeled as follows:

$$r_n = \sum_{i=0}^{L-1} h_i x_{n-i} + w_n, \quad (1)$$

where w_n is an independent additive white Gaussian noise with variance σ_w^2. At the receiver side, we consider a turbo-equalizer for iterative detection and decoding. The union bound on the FER performance of the concatenated system, assuming uniform interleaver, is the sum of pairwise error probabilities between all pairs of coded sequences $\mathbf{x} = (x_1, \cdots, x_N)$ given in [4], [6] by

$$P_w \leq \sum_{d_E^2} A(d_E^2) \cdot Q\left(\sqrt{\frac{d_E^2}{4\sigma_w^2}}\right), \tag{2}$$

where d_E^2 is the squared Euclidean distance between the transmitted sequence $\mathbf{x}$ and the estimated sequence $\hat{\mathbf{x}}$ at the output of the noiseless ISI channel, $Q(\cdot)$ is the Gaussian error probability function, and $A(d_E^2)$ is the output squared Euclidean distance enumerator of the concatenated system given by

$$A(d_E^2) \triangleq \sum_{d=d_{min}}^{N} A^c(d) \cdot \frac{A_d^{ch}(d_E^2)}{\binom{N}{d} 2^d}, \tag{3}$$

where d_{min} is the minimum free distance of the FEC code, $A^c(d)$ is the output weight enumerator of the FEC code defined as the total number of code error sequences $\mathbf{e} = \hat{\mathbf{x}} - \mathbf{x}$ with Hamming weight $d_H(\mathbf{e}) = d$, and $A_d^{ch}(d_E^2)$ is the input-output Euclidean distance enumerator of the ISI channel, defined as the number of all error sequences with input Hamming weight d and output squared Euclidean distance d_E^2. The squared Euclidean distance d_E^2 between the transmitted sequence $\mathbf{x}$ and the estimated sequence $\hat{\mathbf{x}}$ at the output of the noiseless ISI channel can be calculated as,

$$d_E^2 = \|\mathbf{h} * \hat{\mathbf{x}} - \mathbf{h} * \mathbf{x}\|^2 = \|\mathbf{h} * \mathbf{e}\|^2, \tag{4}$$

which is function of the error sequence $\mathbf{e}$ and will be denoted in the sequel by $d_E^2(\mathbf{e})$. The squared Euclidean distance can be thought as the energy of the filtered error sequence $\mathbf{e}$ by the channel response $\mathbf{h}$. Since the ISI channel disperses the energy of the transmitted sequence over $N+L-1$ symbol periods, it is more accurate to evaluate the squared Euclidean distance over $N+L-1$ samples rather than only N samples as it is the case in some previous works [5], [6]. Moreover, we assume, in the evaluation of (4), that symbols outside sequence period are known by the receiver and, consequently, the corresponding error values are assumed to be identically zeros. For the BPSK mapping scheme, error sequence elements e_n take their values from the ensemble $\{-2, 0, +2\}$ depending on the transmitted elements x_n in such a way that $\mathbf{x} + \mathbf{e}$ is a valid code sequence. The second term in (3) can be interpreted as the conditional probability of having a squared output Euclidean distance d_E^2 given the input weight d of the error sequence. We denote this conditional probability by $P_d(d_E^2)$,

$$P_d(d_E^2) \triangleq \frac{A_d^{ch}(d_E^2)}{\binom{N}{d} 2^d}. \tag{5}$$

We focus in this paper on the characterization of this conditional probability in order to evaluate the upper bound (2). We start by establishing a new formulation for the Euclidean distance in order to assess its statistical characteristics.

3. CHANNEL OUTPUT EUCLIDEAN DISTANCE

The squared Euclidean distance between any pair of sequences, having an error sequence $\mathbf{e}$ between them, at the output of noiseless ISI channel is given by

$$d_E^2(\mathbf{e}) = \|\mathbf{h} * \mathbf{e}\|^2 = \sum_{m=1}^{M+L-1} \left| \sum_{\ell=0}^{L-1} h(\ell) e(m-\ell) \right|^2. \tag{6}$$

By developing the squared sum and performing some algebraic computations, we can rewrite the squared Euclidean distance under the form

$$d_E^2(\mathbf{e}) = \sum_{\ell=-L+1}^{L-1} R_h(-\ell) R_e(\ell), \tag{7}$$

where $R_z(\ell)$ are the aperiodic autocorrelation coefficients (AAC) at lag ℓ of the parameter sequence $\mathbf{z}$, and defined by

$$R_z(\ell) \triangleq \sum_{n=1}^{N} z(n) z(n-\ell), \quad |\ell| \leq N-1,$$

with $z_n = 0$ for $n \notin [1, N]$. The equation (7) shows that, for a given channel response, the squared Euclidean distance is fully characterized by the AAC of the error sequence up to lag $L-1$. Consequently, the Euclidean distance inherits all autocorrelation properties. For a real channel response and real modulation alphabet the equation (7) can be rewritten, using the symmetry of the AAC $R_h(-\ell) = R_h(\ell)$, as the sum of two terms

$$d_E^2(\mathbf{e}) = \underbrace{R_h(0) R_e(0)}_{\triangleq \Lambda} + \underbrace{2 \sum_{\ell=1}^{L-1} R_h(\ell) R_e(\ell)}_{\triangleq \Delta}. \tag{8}$$

The first term Λ is the squared Euclidean distance over AWGN channel, whereas the second term Δ expresses its variations due to the presence of the ISI. Let $\mathcal{E}$ be the set of all possible error sequences $\mathbf{e}$. Then, the squared Euclidean distance can be viewed as a discrete random variable taking its values in positive real numbers, and defined over the probability space $\mathcal{E}$. It is expressed as a linear multivariate function that maps the first L autocorrelation coefficients $R_e(\ell)$ for $\ell = 0, \cdots, L-1$ to positive real number. Autocorrelation coefficients are also discrete random variables over the same probability space.

4. CONDITIONAL DISTANCE DISTRIBUTION

The conditional probability $P_d(d_E^2)$ is given by the distribution of d_E^2 over a subset $\mathcal{E}_d$ of $\mathcal{E}$ formed by all error sequences $\mathbf{e} \in \mathcal{E}$ with Hamming weight $d_H(\mathbf{e}) = d$. For a fixed input error weight d the first term in (8) is a constant $\Lambda = 4d$ assuming unit channel gain ($R_h(0) = 1$), while the interference term Δ is a weighted sum of $L-1$ random variables $R_e(\ell)$ for $1 \leq \ell \leq L-1$ taking their values in the ensemble $\mathcal{R} \triangleq \{4k : \ k = -(d-1), \cdots, +(d-1)\}$. Before studying the conditional distance distribution, we derive its average and variance.

4.1. Average and variance

Assuming the independence between error elements e_n, it can be easily shown from the definition of the autocorrelation that the random variables $R_e(\ell)$ are pairwise uncorrelated with zero mean and variance given by,

$$\sigma^2(R_e(\ell)) = 16(N-\ell) \cdot \frac{d(d-1)}{N(N-1)}, \ 1 \leq \ell \leq N-1. \tag{9}$$

The mean of the squared Euclidean distance can be calculated by averaging (8). We obtain

$$\mu(d_E^2) = \Lambda = R_h(0) R_e(0) = 4dE_h, \tag{10}$$

where $E_h \triangleq R_h(0)$ is the channel gain. The non-correlation between the AAC allows us to evaluate the variance of the squared Euclidean distance, which is an important measure of distance dispersion introduced by the ISI channel. This yields to

$$\sigma^2(d_E^2) = 64\frac{d(d-1)}{N(N-1)}\sum_{\ell=1}^{L-1}(N-\ell)R_h^2(\ell), \qquad (11)$$

which shows that the variance for large values of N, in comparison with channel length L, is proportional to $1/N$ which explains the interleaving gain in the system. When N tends to infinity with fixed value of d the variance $\sigma^2(d_E^2)$ tends to zero. This explains the convergence of the performance to AWGN case as it was already shown in [7] under iterative decoding assumptions.

4.2. Conditional distance distribution

Since d_E^2 is a weighted sum of related random variables, the complete characterization of $P_d(d_E^2)$ requires the knowledge of the joint probability of autocorrelation coefficients. Let $\mathbf{R_e} = (R_e(1), \cdots, R_e(L-1))$ denotes the vector formed by the first $L-1$ out-of-phase correlation lags. Let V_R denotes the set of all possible values of $\mathbf{R_e}$. The conditional probability can be calculated as,

$$P_d(d_E^2 = x) = \sum_{\mathbf{R_e}\in V_R : d_E^2 = x} P_J(\mathbf{R_e}),\ x > 0, \qquad (12)$$

where $P_J(\mathbf{R_e})$ denote the joint probability of the autocorrelation vector $\mathbf{R_e}$. Unfortunately, to the author's knowledge there is no closed-form expression for the joint probability of AAC [8] and, in general, it is very difficult to be evaluated, specially for large packet size. To overcome this problem, we resort to bounding techniques on the joint probability as it will be shown in the next section. However, for simple ISI channels with only two non-zero tap coefficients, as it is the case for some partial response channels, $\mathbf{h} = (h_0, 0, \cdots, 0, h_{L-1})$ for some $L > 1$ only a single AAC $R_e(L-1)$ will be implied in the evaluation of the Euclidean distance. In this case, the corresponding marginal probability mass function (p.m.f), denoted by P_{L-1}, determines the output Euclidean distance distribution of the channel. Marginal p.m.f can be determined as follows. We begin with the p.m.f of $R_e(1)$. For this, we define a new variable $S_1 = \sum_{i=1}^{N-1}|z_{1,i}|$ where $z_{1,i} = e_i e_{i+1}/4$. The introduced variable S_1 gives the number of non-zero terms in the autocorrelation definition. We evaluate the p.m.f of $R_e(1)$ as the marginal probability of $R_e(1)$ conditionally to S_1,

$$\Pr(R_e(1)) = \sum_{s=0}^{d-1}\Pr\left(R_e(1)|S_1 = s)\right)\Pr(S_1 = s). \qquad (13)$$

First, we note that S_1 is also the number of consecutive non-zero elements in the error sequence $\mathbf{e}$. An exact enumeration of the number of consecutive non-zero elements in an error sequence of length N with Hamming weight d leads to the following hyper-geometric p.m.f. of S_1

$$\Pr(S_1 = s) = \frac{\binom{d-1}{s}\binom{N+d-1}{d-s}}{\binom{N}{d}}. \qquad (14)$$

Second, for a given value of $S_1 = s$, we have $R_e(1) = 4(s-2n)$ where n is the number of negative elements in $\mathbf{e}$. Since there is $\binom{s}{n}$ different possibilities out of 2^s to select negative elements from s non-zero i.i.d elements We find the conditional probability $\Pr(R_e(1)|S_1 = s))$, after defining $k = s - 2n$, is

$$\Pr(R_e(1) = 4k|S_1 = s)) = \frac{\binom{s}{n}}{2^s} = 2^{-s}\binom{s}{\frac{s-|k|}{2}}, \qquad (15)$$

with the convention $\binom{i}{j} = 0$ for non integer values of j. By substituting equations (14) and (15) in (13) we obtain the p.m.f of $R_e(1)$,

$$P_1(4k) = \frac{1}{\binom{N}{d}}\sum_{s=|k|}^{d-1}2^{-s}\binom{d-1}{s}\binom{N-d+1}{d-s}\binom{s}{\frac{s-|k|}{2}}. \qquad (16)$$

Similar analysis can be applied for higher lags and we can show that when $N >> L$ (which is the practical case) all AAC can be considered as identically distributed with the same p.m.f given by (16). This result can be predicted from the expression of the variance of AAC (9) which presents a very small variations with lag ℓ for large packet length N. With this, we find the same results in [5], [6] for the distance distribution of 2-taps channels with a slight difference related to the difference in defining the output Euclidean distance as we noticed before.

4.3. An upper bound on distance distribution

In order to evaluate the union bound in (2), we upperbound the conditional distance distribution by upper bounding the joint probability of AAC based on the knowledge of the corresponding marginal p.m.f. It is shown in [9] that attainable autocorrelation values must lay within a convex region, denoted by V_R, defined by all L-uple vectors $\mathbf{R_e} = (R_e(0), \cdots, R_e(L-1))$ with positive-definite autocorrelation matrix M_L defined by,

$$M_L = \begin{bmatrix} R_e(0) & R_e(1) & \cdots & R_e(L-1) \\ R_e(1) & R_e(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_e(1) \\ R_e(L-1) & \cdots & R_e(1) & R_e(0) \end{bmatrix}.$$

Mathematically, V_R is defined by

$$V_R \triangleq \{\mathbf{R_e} \in \mathcal{R}^L : \det(M_L) > 0\}. \qquad (17)$$

The joint probability of the autocorrelation vector $\mathbf{R_e}$ denoted by P_J, knowing marginal p.m.f P_k, can be upper bounded by the minimum of all marginal probability as follows,

$$P_J(\mathbf{R_e}) \leq \min\left[P_1(R_e(1)), \cdots, P_{L-1}(R_e(L-1))\right], \qquad (18)$$

for $\mathbf{R_e} \in V_R$ and zero otherwise. This bound is the best known bound on the joint probability if no other information is available about the joint probability. Further investigations for additional information about joint probability may result in a tighter upper bound.

5. MINIMUM FREE EUCLIDEAN DISTANCE

Now, we show that the proposed approach allows us to calculate the minimum free Euclidean distance by solving a minimization problem under constraints. Minimizing the squared Euclidean distance over valid autocorrelation values (V_R) gives the minimum Euclidean distance $d_{E,min}$ (for sufficiently large frame size, see [10]). Because the Euclidean distance is a linear function of autocorrelation coefficients, the minimum will be achieved for some point on the boundary region of valid autocorrelation values V_R. The convexity

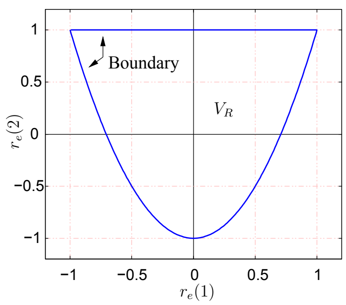

Fig. 2. The region V_R of attainable values for $r_e(1)$ and $r_e(2)$.

of V_R ensure the convergence of the minimization algorithm to the actual minimum free distance of the channel. We show the determination of $d_{E,min}$ by an example. The FIR of the Proakis-B channel is $\mathbf{h} = (0.408,\ 0.817,\ 0.408)$ of length $L = 3$. The AAC values for positive lags are $R_h(0) = 1$, $R_h(1) = 0.666$ and $R_h(2) = 0.166$. The output Euclidean distance is evaluated using (8) as,

$$d_E^2 = R_e(0) + 1.33R_e(1) + 0.33R_e(2). \tag{19}$$

The determinant of the autocorrelation matrix $\det(M_3)$ expressed in terms of the normalized AAC values $r_e(\ell) \triangleq R_e(\ell)/R_e(0)$ as,

$$\det(M_3) = R_e^3(0)[1 - r_e(2)][1 + r_e(2) - 2r_e^2(1)]. \tag{20}$$

The boundary can be found by making $\det(M_3) = 0$ which yields to following solution (see Fig 2)

$$r_e(2) = 2r_e^2(1) - 1, \quad \text{or} \quad r_e(2) = 1. \tag{21}$$

Minimizing the Euclidean distance in (19) with the constraints in (21) leads to the following solution $R_1 = -4(d-1)$ and $R_2 = 4(d-2)$ for which we have $d_{E,min}^2 = 2.66$ for any value of error weight d. This minimum value is attained by any error sequence with consecutive error elements of alternating signs.

6. NUMERICAL RESULTS

In Fig. 3 we show the calculated upper bound from (2) for the Proakis-B channel. On the same figure we traced the lower bound on code FER performance over AWGN which is also a lower bound on the coded system over any ISI channel. We compare obtained bounds to the simulated performance using a maximum *a posteriori* turbo-equalizer using the Backward-Forward algorithm with 5 turbo-iterations. The gap between the two bounds at medium SNR is about 1dB, and the upper bound diverges for at high SNR values. This is due to the poor quality of the joint probability bound.

7. CONCLUSIONS

In this paper, we present a new approach to evaluate the output Euclidean distance of a general ISI channel. we determine the main statistical characteristics of the output Euclidean distance including mean, variance, minimum distance, and distance distribution. We apply obtained results to evaluate an upper bound on the FER of the ML receiver for turbo-equalized systems. Our approach allows for better understanding of the impact of system parameters on the performance behavior. It is a general framework and can be extended to time varying channels with known statistics. It would be of interest to extend this approach to high order modulations. with a given mapping scheme.

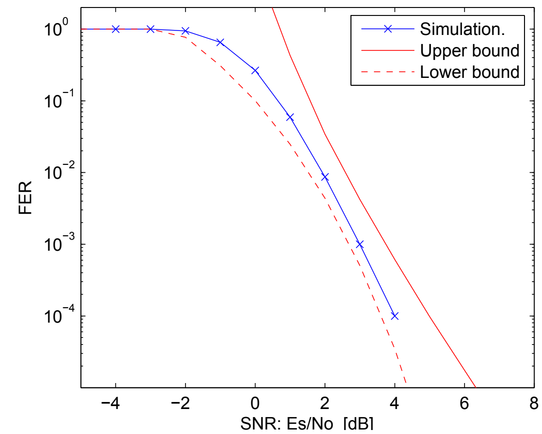

Fig. 3. Lower and upper bounds on the FER of turbo-equalized Proakis-B channel with packet size $N = 256$ using convolutional code (7,5) in comparison with simulated results.

8. REFERENCES

[1] C. Douillard, A. Picart, P. Didier, M. Jzquel, C. Berrou, and A. Glavieux, "Iterative correction of intersymbol interference: turbo-equalization," *European Trans. Telecommun.*, vol. 6, pp. 507–512, 1995.

[2] G. Forney Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 18, pp. 363–378, 1972.

[3] S. A. Raghavan, J. K. Wolf, and L. B. Milstein, "On the performance evaluation of ISI channels," *IEEE Trans. Inf. Theory*, vol. 39, pp. 957–965, 1993.

[4] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: performance analysis, design, and iterative decoding," *IEEE Trans. Inf. Theory*, vol. 44, pp. 909–926, 1998.

[5] M. Oberg and P. H. Siegel, "Performance analysis of turbo-equalized partial response channels," *IEEE Trans. Commun.*, vol. 49, pp. 436–444, 2001.

[6] J. Li, K. R. Narayanan, and C. N. Georghiades, "An efficient algorithm to compute the euclidean distance spectrum of a general intersymbol interference channel and its applications," *IEEE Trans. Commun.*, vol. 52, pp. 2041–2046, 2004.

[7] N. Sellami, A. Roumy, and I. Fijalkow, "A proof of convergence of the MAP turbo-detector to the AWGN case," *To appear in IEEE Trans. Signal Processing*, 2007.

[8] S.M. Kay, A.H. Nuttall, and P.M. Baggenstoss, "Multidimensional probability density function approximations for detection, classification, and model order selection," *IEEE Trans. Signal Process.*, vol. 49, pp. 2240–2252, 2001.

[9] M. H. Quenouille, "The joint distribution of serial correlation coefficients," *The Annals of Mathematical Statistics*, vol. 20, pp. 561–571, 1949.

[10] A. Steinhardt and J. Makhoul, "On the autocorrelation of finite-length sequences," *IEEE Trans. ASSP*, vol. 33, pp. 1516–1520, 1985.