The impact of both *a priori* information and channel estimation errors on the MAP equalizer performance *

Noura Sellami[1*], Aline Roumy[2], Inbar Fijalkow [3]

[1] ISECS, Route Menzel Chaker km 0.5, B.P 868, 3018 Sfax, Tunisia

Email: sellami_noura@yahoo.fr, Tel: (216)97059318, Fax: (216)74274088

[2] IRISA-INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France

Email: aline.roumy@irisa.fr, Tel: (33)299847394, Fax: (33)299847171

[3] ETIS, UMR 8051, ENSEA-UCP-CNRS,

6 av. du Ponceau, 95014 Cergy-Pontoise, France

Email: fijalkow@ensea.fr, Tel: (33)130736610, Fax: (33)130736627

Abstract

To combat the effects of intersymbol interference, the optimal equalizer to be used is based on maximum *a posteriori* (MAP) detection. In this paper, we consider the case where the MAP equalizer is fed with *a priori* information on the transmitted data and propose to study analytically their impact on the MAP equalizer performance. We assume that the channel is not perfectly estimated and show that the use of both the *a priori* information and the channel estimate is equivalent to a shift in terms of signal-to-noise ratio (SNR) for which we provide an analytical expression. Simulation results show that the analytical expression approximates well the equalizer behavior.

EDICS:

- PERF Performance Analysis, Optimization, and Limits
- CEQU Channel modeling, estimation, and equalization

*Parts of this paper have been presented at the SPAWC'04 and VTC-FALL'04 conferences. This work was partially supported by the INRIA-DGRSRT program under Grant 05 I 17

1 Introduction

To combat the effects of intersymbol interference, an equalizer has to be used. The optimal equalizer, in the sense of minimum sequence error rate (SER) or bit error rate (BER) is based on maximum *a posteriori* (MAP) detection. We distinguish two criteria, the MAP sequence detection and the MAP symbol detection. When no *a priori* information on the transmitted data is available, MAP detection turns into maximum likelihood (ML) detection. Efficient algorithms exist for MAP sequence detection, for example the SER optimizing Viterbi algorithm [1], and MAP symbol detection, for example the BER optimizing BCJR algorithm [2]. These algorithms are interesting since their complexity grows linearly rather than exponentially with the sequence size.

The performance of the Viterbi equalizer in the presence of additive white Gaussian noise (AWGN) has been studied by Forney [1]. This study assumed that the channel is perfectly known at the receiver and no *a priori* information is provided to the equalizer. However, the receiver does not know perfectly the channel in practice and has to estimate it. In [3], Gorokhov studied the impact of channel estimation errors on the performance of the Viterbi equalizer and showed that it is equivalent to a loss in signal-to-noise ratio (SNR) and evaluated this loss. In [4], we have extended the study to a List-type equalizer prefiltered by the whitened matched filter, in the case of multiple-input multiple-output (MIMO) systems.

In this paper, we consider the case where the MAP equalizer has *a priori* information on the data and uses an imperfect channel estimate. The *a priori* information are provided by another module in the receiver. Our study can be applied, for instance, to a MAP equalizer within a system using a bit interleaved coded modulation at the transmitter, as shown in Figure 1, and a turbo-equalizer at the receiver, as shown in Figure 2. In a turbo-equalizer, during the iterations, the equalizer and the decoder exchange extrinsic information and use them as *a priori* in order to improve their performance [5]. We propose to study analytically the impact of both the *a priori* information and the channel estimation errors on the MAP equalizer performance. We show that it is equivalent to a shift in SNR and we give a closed form of this shift. Here, the channel is estimated using a training sequence [6]. The motivation for considering a training sequence based channel estimator is that the statistics of the channel estimation error has a closed form and depends on the training sequence properties and the transmission noise variance only. However, this analysis also holds for any variance of the channel estimation error. It can therefore be extended to the case of blind iterative algorithms such as the K-means algorithm [7] and the EM algorithm [8] [9].

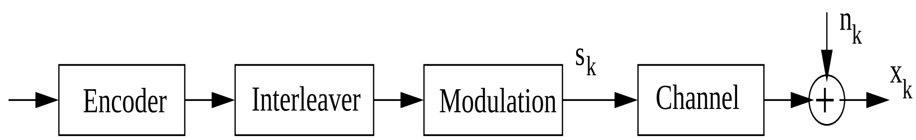

Figure 1: Transmitter structure: bit interleaved coded modulation

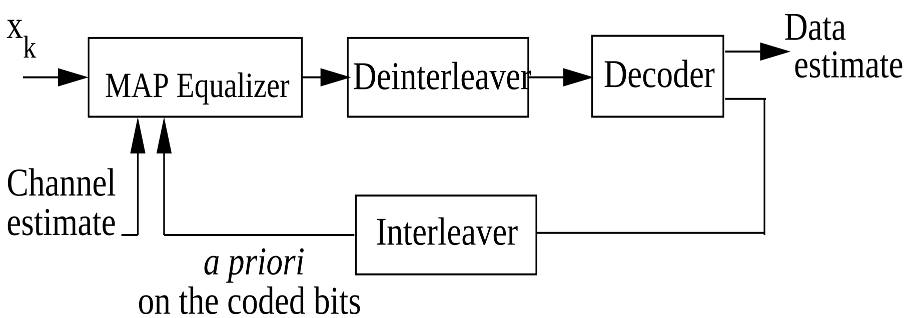

Figure 2: Receiver structure: turbo-equalizer

This work is a first step in the study of the convergence analysis of turbo-equalizers using MAP equalization. Most analyses are based on extrinsic information transfer (EXIT) charts [10][11]. These analyses use generally simulations since it is difficult to study analytically the performance of a MAP equalizer having a large number of states. Actually, analytical studies based on the EXIT function have been performed when the trellis has only two states [12]. The contribution of our paper is to give an analytical study of the MAP equalizer performance when the number of states is greater than two.

The paper is organized as follows. In section 2, we describe the system model. In section 3, we study the impact of the *a priori* information and the channel estimation errors on the equalizer performance. In section 4, we give simulation results.

Throughout this paper scalars and matrices are lower and upper case respectively and vectors are underlined lower case. $(.)^T$ denotes the transposition and I_m is the $m \times m$ identity matrix.

2 System model

We consider a data transmission system over a frequency selective channel. The input information bit sequence is mapped to the symbol alphabet $\mathcal{A}$. For simplicity, we will consider only the BPSK modulation ($\mathcal{A} = \{+1, -1\}$). We assume that transmissions are organized into bursts of T symbols. The channel is supposed to be invariant during one burst. The received baseband signal sampled at the symbol rate at time k is

$$x_k = \sum_{l=0}^{L-1} h_l s_{k-l} + n_k \tag{1}$$

where L is the channel memory and s_k, for $1 - L \leq k \leq T - 1$, are the transmitted symbols. In this expression, n_k are modeled as independent random variables of a real white Gaussian noise with normal probability density function (pdf) $\mathcal{N}(0, \sigma^2)$ where $\mathcal{N}(\alpha, \sigma^2)$ denotes a Gaussian distribution with mean α and variance σ^2. The term h_l is the l^{th} tap gain of the channel, which is assumed to be real valued. Let $\underline{s} = (s_{T-1}, ..., s_{1-L})^T$ be the $(L+T-1)$-long vector of symbols and $\underline{n} = (n_{T-1}, ..., n_0)^T$ be the T-long noise vector. The output of the channel is the T-long vector $\underline{x} = (x_{T-1}, ..., x_0)^T$ defined as

$$\underline{x} = \tau(\underline{h})\underline{s} + \underline{n} \tag{2}$$

where $\tau(\underline{h})$ is a $T \times (T + L - 1)$ Toeplitz matrix with its first row being $(h_0, h_1, ..., h_{L-1}, 0, ..., 0)$ and its first column $(h_0, 0, ..., 0)^T$.

When the channel is known and no *a priori* information is provided to the equalizer, the data estimate according to the MAP sequence criterion (or to the ML criterion since there is no *a priori)* is given by

$$\underline{\hat{s}} = \arg\min_{\underline{u}} \left(\|\underline{x} - \tau(\underline{h})\underline{u}\| : \underline{u} \in \mathcal{A}^{T+L-1} \right). \tag{3}$$

Now, we consider a particular error event characterized by its length m [1]. Thus, we suppose that there exists an interval of size m such that all the symbols of $\underline{\hat{s}}$ are different from the corresponding symbols of $\underline{s}$ while the preceding symbol and the following one are the same for $\underline{s}$ and $\underline{\hat{s}}$. Define $\underline{s}_m$ and $\underline{\hat{s}}_m$ to be the vectors of symbols corresponding to this interval and the vector of errors $\underline{e}_m = \underline{\hat{s}}_m - \underline{s}_m$. A subevent $\mathcal{E}_m$ of the error event is that "$\underline{\hat{s}}_m$ is better than $\underline{s}_m$" in the sense of the ML metric

$$\mathcal{E}_m : \|\underline{x}_m - \tau_m(\underline{h})\underline{\hat{s}}_m\| \leq \|\underline{x}_m - \tau_m(\underline{h})\underline{s}_m\| \tag{4}$$

where $\underline{x}_m$ is the subvector of $\underline{x}$ and $\tau_m(\underline{h})$ is the block of $\tau(\underline{h})$ corresponding to the error interval. The probability $P(\mathcal{E}_m)$ of $\mathcal{E}_m$ is given by [1]:

$$P(\mathcal{E}_m) = Q\left(\frac{\|\underline{\varepsilon}_m\|}{2\sigma}\right) \tag{5}$$

where $\underline{\varepsilon}_m = \tau_m(\underline{h})\underline{e}_m$ and $Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty \exp(-\frac{y^2}{2})dy$. Let Σ_m be the set of all possible error events of length m. Then, the probability, $P(\Sigma_m)$, that any error event is of length m is bounded by the sum of the probabilities of the subevents $\mathcal{E}_m$

$$P(\Sigma_m) \leq \sum_{\mathcal{E}_m} P(\mathcal{E}_m). \tag{6}$$

Let d_{min} be the channel minimum distance [1]. Because of the exponential decrease of the Gaussian distribution function, the overall probability of error $P(\Sigma) \leq \sum_m P(\Sigma_m)$ will be dominated at high SNR by the term involving the minimum value $d_{\min}$ of $\|\underline{\varepsilon}_m\|$. Thus,

$$P(\Sigma) \simeq Q\left(\frac{d_{\min}}{2\sigma}\right). \tag{7}$$

Our goal is to find an approximation of $P(\Sigma)$ when the equalizer is fed with *a priori* information and an imperfect channel estimation.

3 Performance analysis

We want to evaluate the impact of both the *a priori* information and the channel estimation errors on the MAP equalizer performance. The study will be done here for the equalizer using the MAP sequence criterion. It holds for the MAP symbol equalizer using the BCJR algorithm [2] since the two equalizers have almost the same performance as shown in [13, page 814].

The channel is estimated by using a training sequence of length $T_0 + L - 1$. Since the channel noise is additive, white and Gaussian, the ML estimator based on the observations generated by the training sequence consists in minimizing the mean square error between the signal received during the emission of the TS and its noiseless counterpart (the filtered version of the training sequence), which is the least squares estimator [6]. Let $\underline{\tilde{s}} = (s_{T_0-1}, ..., s_{1-L})^T$ be the vector of training symbols and $H_L(\underline{\tilde{s}})$ be the $T_0 \times L$ Hankel matrix having the first column $(s_{T_0-1}, \cdots, s_0)^T$ and the last row $(s_0, \cdots, s_{1-L})$. The output of the channel corresponding to the training sequence is the T_0-long vector $\underline{\tilde{x}}$ given by

$$\underline{\tilde{x}} = H_L(\underline{\tilde{s}})\,\underline{h} + \underline{\tilde{n}} \tag{8}$$

where $\underline{h}= (h_0, \cdots, h_{L-1})^T$ and $\underline{\tilde{n}}$ is the corresponding noise vector. The least squares channel estimate $\underline{\hat{h}} = \left(\hat{h}_0, \cdots, \hat{h}_{L-1}\right)^T$ is given by

$$\begin{aligned} \underline{\hat{h}} &= \arg\min_{\underline{f}} \left\|\underline{\tilde{x}} - H_L(\underline{\tilde{s}})\,\underline{f}\right\|^2 \\ &= \left(H_L(\underline{\tilde{s}})^T H_L(\underline{\tilde{s}})\right)^{-1} H_L(\underline{\tilde{s}})^T \underline{\tilde{x}}. \end{aligned} \tag{9}$$

Hence, we get,

$$\underline{\hat{h}} - \underline{h} \sim \mathcal{N}\left(0, \sigma^2 \left(H_L(\underline{\tilde{s}})^T H_L(\underline{\tilde{s}})\right)^{-1}\right). \tag{10}$$

We assume here that the training sequence has ideal autocorrelation and crosscorrelation properties, i.e. $H_L(\underline{\tilde{s}})^T H_L(\underline{\tilde{s}}) = T_0 I_L$. The estimates $\hat{h}_l$ of the tap gains h_l, for $0 \leq l \leq L-1$, are thus modeled as $\hat{h}_l = h_l + \sigma_e k_l$, where k_l are independent Gaussian random variables with zero mean and variance 1 and $\sigma_e = \frac{\sigma}{\sqrt{T_0}}$.

Moreover, we suppose that the *a priori* observations at the input of the equalizer are modeled as the outputs of an additive white Gaussian noise (AWGN) channel with zero mean and variance σ_a^2. These *a priori* observations are

$$z_k = s_k + w_k \tag{11}$$

where $w_k \sim \mathcal{N}(0, \sigma_a^2)$, for $1 - L \leq k \leq T - 1$. Thus, the *a priori* Log Likelihood Ratios, fed back for instance from the decoder in a turbo-equalizer, can be modeled as independent and identically distributed random variables with the conditional pdf $\mathcal{N}(\frac{2s_k}{\sigma_a^2}, \frac{4}{\sigma_a^2})$. This assumption is classically taken in the analyses of iterative receivers [10][11].

Proposition 1 *Suppose we are given a frequency selective channel with L taps and an AWGN with variance σ^2. Assume that a priori observations are available and that they can be modeled as the outputs of an AWGN channel with noise variance σ_a^2. The estimates $\hat{h}_l$ of the channel tap gains h_l, for $0 \leq l \leq L-1$, are modeled as $\hat{h}_l = h_l + \sigma_e k_l$, where k_l are independent Gaussian random variables with zero mean and variance 1. Then, at high SNR, the MAP equalizer using the a priori information and the channel estimate can be approximated by the MAP equalizer having no a priori information and a perfect channel knowledge but with an equivalent signal-to-noise ratio*

$$S\tilde{N}R = SNR \frac{d'^2}{d_{\min}^2}(1 + \frac{4m'\mu^2}{d'^2}) \left(1 + \frac{L\rho^2}{1 + \frac{4m'\mu^2}{d'^2}}\right)^{-1} \tag{12}$$

where SNR is the signal-to-noise ratio of the transmission, $\mu = \frac{\sigma}{\sigma_a}$ and $\rho = \frac{\sigma_e}{\sigma}$. The quantities m' and d' are defined as $(m', d') = \arg\max_{m, \|\varepsilon_m\|} P(\mathcal{E}_m)$ with

$$P(\mathcal{E}_m) = Q\left(\frac{\sqrt{\|\varepsilon_m\|^2 + 4m\mu^2}}{2\sigma}\left(1 + \frac{1}{1 + \frac{4m\mu^2}{\|\varepsilon_m\|^2}} L\rho^2\right)^{-1/2}\right). \tag{13}$$

The proof of Proposition 1 is given in the Appendix. The proof is divided into three parts. First, the probability of an error subevent of length m, $P(\mathcal{E}_m)$, is derived and then upper bounded. Finally, the overall probability of error, $P(\Sigma)$, is calculated in order to find an approximation of the equivalent

SNR.

When the channel is perfectly estimated, ρ is 0, which leads to the following proposition.

Proposition 2 *When the channel is perfectly known at the receiver, then at high SNR the MAP equalizer using the a priori information can be approximated by the MAP equalizer having no a priori information and a perfect channel knowledge but with an equivalent signal-to-noise ratio*

$$S\tilde{N}R = SNR\frac{d'^2}{d_{\min}^2}(1+\frac{4m'\mu^2}{d'^2}) \tag{14}$$

where m' and d' are defined as $(m',d') = \arg\max_{m,\|\underline{\varepsilon}_m\|} P(\mathcal{E}_m)$ with

$$P(\mathcal{E}_m) = Q\left(\frac{\sqrt{\|\underline{\varepsilon}_m\|^2+4m\mu^2}}{2\sigma}\right). \tag{15}$$

We propose in the sequel to refine the previous results given in propositions 1 and 2. We consider different cases according to the values of μ and of the channel memory, and determine the error sequence that maximizes $P(\mathcal{E}_m)$. We then give a closed form of $S\tilde{N}R$.

Corollary 3 *When the channel memory is not too large (less than 6) and μ sufficiently smaller than 1, the equivalent signal-to-noise ratio is given by:*

$$S\tilde{N}R = SNR(1+\frac{8\mu^2}{d_{\min}^2})\left(1+\frac{L\rho^2}{1+\frac{8\mu^2}{d_{\min}^2}}\right)^{-1} \tag{16}$$

Proof of Corollary 3: Let us consider first the case of perfect channel knowledge. By definition, $\|\underline{\varepsilon}_m\|^2 \geq d_{\min}^2$. Generally, in the MAP equalizer, errors occur in packets. This is still true here since the *a priori* information are not very reliable ($\mu << 1$). Thus, we do not consider isolated errors since they occur rarely and we assume that $m \geq 2$. Then, a lower bound for $\sqrt{\|\underline{\varepsilon}_m\|^2+4m\mu^2}$ is given by

$$bound(\mu^2) = \sqrt{d_{\min}^2+8\mu^2}. \tag{17}$$

We observed that this bound is reached for channels with memory L less than 6, since for these channels the error sequence allowing to attain the minimum distance is generally of length $m = 2$ (see the channels in [14] for instance). Then, an upper bound for $P(\mathcal{E}_m)$ is $Q\left(\frac{\sqrt{d_{\min}^2+8\mu^2}}{2\sigma}\right)$.

When the channel is not perfectly estimated, the probability of the error event (13), given in Proposition 1, can be rewritten as

$$P(\mathcal{E}_m) = Q\left(\frac{\|\underline{\varepsilon}_m\|^2 + 4m\mu^2}{2\sigma\sqrt{\|\underline{\varepsilon}_m\|^2 L\rho^2 + \|\underline{\varepsilon}_m\|^2 + 4m\mu^2}}\right). \tag{18}$$

We assume here that $L << T_0$ which is equivalent to $L\rho^2 << 1$. In this case, $\|\underline{\varepsilon}_m\|^2 L\rho^2$ is negligible compared to $\|\underline{\varepsilon}_m\|^2$. Then, (18) is almost equal to (15) and the error sequence that maximizes (15) will also maximize (18). In the case of short channels, this sequence is the one allowing to attain the minimum distance and is of length $m = 2$. Thus, we can consider that the quantity obtained by calculating (18) taking $m = 2$ and $\|\underline{\varepsilon}_m\|^2 = d_{\min}^2$ is an upper bound for (18). Thus, the overall probability of error can be approximated by:

$$P(\Sigma) \simeq Q\left(\frac{d_{\min}}{2\sigma}\sqrt{1+\frac{8\mu^2}{d_{\min}^2}}\left(1+\frac{L\rho^2}{1+\frac{8\mu^2}{d_{\min}^2}}\right)^{-1/2}\right). \tag{19}$$

The expression of the error probability given in (19) can then be viewed as the one given in (7) with the equivalent signal-to-noise ratio (16).

Corollary 4 *When μ is sufficiently larger than 1, the equivalent signal-to-noise ratio is given by*

$$S\tilde{N}R = SNR\frac{4}{d_{\min}^2}(1+\mu^2)\left(1+\frac{L\rho^2}{1+\mu^2}\right)^{-1} \tag{20}$$

Proof of Corollary 4: When μ is high, $\sigma >> \sigma_a$, most of the *a priori* observations are very reliable and have more influence on the detection than the channel observations. Since the *a priori* information are independent, the errors will not occur in packets. Actually, in this case, isolated errors corresponding to the few non reliable *a priori* observations will occur and will dominate the overall probability of error. Thus, the overall probability of error can be approximated by the upper bound of $P(\mathcal{E}_m)$ obtained by replacing m by 1 and $\|\underline{\varepsilon}_m\|^2$ by 4 in (13),

$$P(\Sigma) \simeq Q\left(\frac{\sqrt{1+\mu^2}}{\sigma}\left(1+\frac{1}{1+\mu^2}L\rho^2\right)^{-1/2}\right). \tag{21}$$

The expression of the error probability given in (21) can then be viewed as the one given in (7) with the equivalent signal-to-noise ratio (20).

4 Simulation results

In this section, we propose to test for the validity of the analytical results given previously. In the simulations, the modulation used is the BPSK and the channel is assumed to be constant. We plot the Bit Error Rate (BER) curves with respect to the SNR, for different values of the ratio $\mu = \frac{\sigma}{\sigma_a}$. Each curve is obtained while the ratio μ is kept constant. The solid lines indicate the equalizer performance obtained by simulations. The dotted lines are obtained by shifting the curve corresponding to the case with no *a priori* and with a perfect channel knowledge ($\mu = 0, \rho = 0$) by the values of the SNR shifts calculated in section 3. We consider two cases according to the length of the channel. We also consider the case where the channel is overestimated.

4.1 Case of short channels ($L \leq 6$)

In our simulations, we consider the following channels:

- Channel3: $(0.5; 0.71; 0.5)$
- Channel5: $(0.29; 0.50; 0.58; 0.50; 0.29)$.

Table.1 shows the values of the minimum error distance $d_{\min}$ and the minimum distance input error sequence for the channels of interest [14].

	Channel3	Channel5
$d_{\min}$	1.5308	1.0532
Error sequence	$(2, -2)$	$(2, -2)$

Table.1

Figures 3 and 4 show the BER curves for different values of the ratio $\mu = \frac{\sigma}{\sigma_a}$ when the channel is assumed to be perfectly known at the receiver ($\rho = 0$). For Channel 3, when $\mu < 0.66$, we use the result of Corollary 3, hence the SNR shift in dB is $10 \log_{10} \left(1 + \frac{8\mu^2}{d_{\min}^2}\right)$. When $\mu \geq 0.66$, we use the result of Corollary 4 and the SNR shift is $10 \log_{10} \left(\frac{4}{d_{\min}^2}(1 + \mu^2)\right)$. For Channel 5, when $\mu < 1$, we use Corollary 3, otherwise, we use Corollary 4. We notice that the theoretical curves (dotted lines) approximate well the BER.

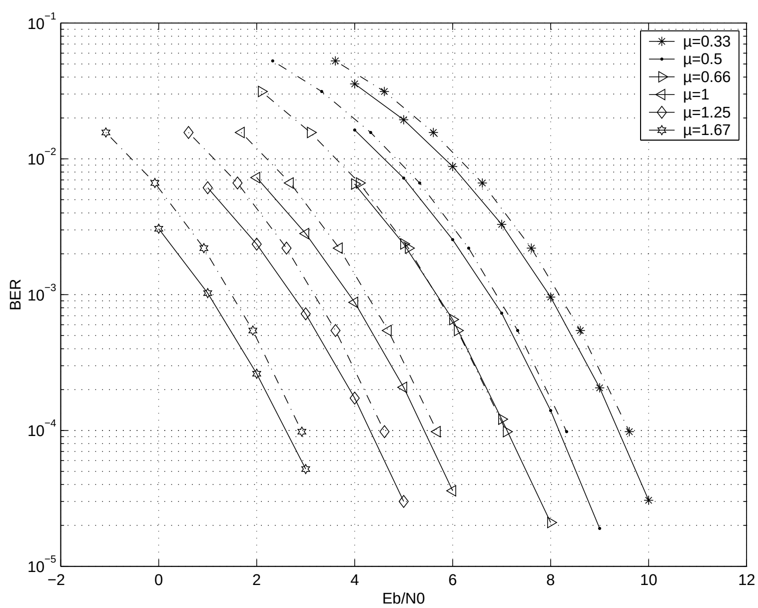

Figure 3: BER versus SNR: comparison of the equalizer performance (solid curves) and the theoretical performance (dotted curves) for Channel3, when the channel is perfectly estimated.

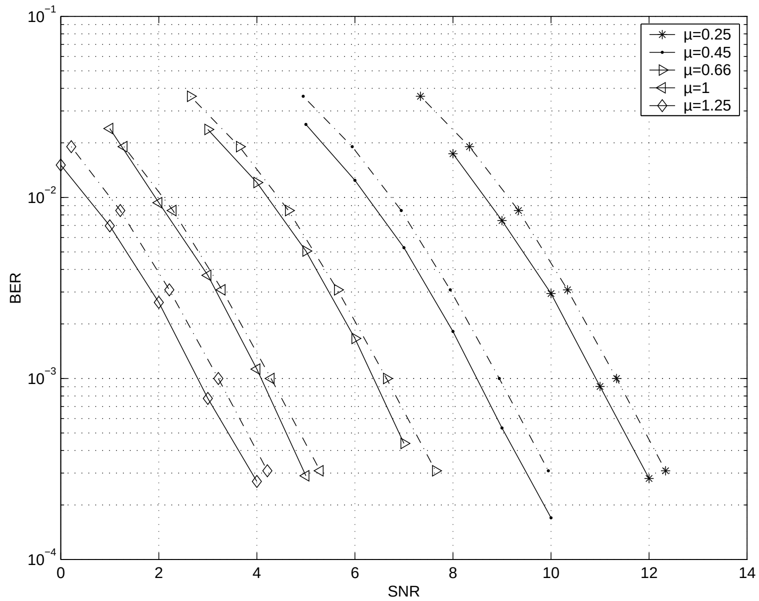

Figure 4: BER versus SNR: comparison of the equalizer performance (solid curves) and the theoretical performance (dotted curves) for Channel5, when the channel is perfectly estimated.

Figures 5 and 6 show the BER curves for different values of μ and for $\rho = \frac{\sigma_e}{\sigma} = 0.3$. This value of the ratio ρ corresponds to $T_0 = 11$. For $\mu < 0.83$, we use the expression of the SNR shift given in Corollary 3. For larger μ, we use the expression given in Corollary 4. We notice that the theoretical curves approximate well the BER.

Figure 7 shows the BER curves for Channel3, for different values of μ and for $\rho = 0.4$ and $\rho = 0.2$. These values of the ratio ρ correspond respectively to $T_0 = 6$ and $T_0 = 25$. We notice that also for these values of ρ, the approximation of the SNR shift given by the analytical expressions is still accurate.

4.2 Case of long channels

We consider now a long channel, Channel7, with impulse response $(0.18; 0.32; 0.48; 0.53; 0.48; 0.32; 0.18)$. The error sequence allowing to reach the minimum distance is $(2, -2, -2, 2, 2, -2)$ and $d_{min} = 0.7283$ [14]. However, this sequence does not maximize the probability of an error event $P(\mathcal{E}_m)$, since its

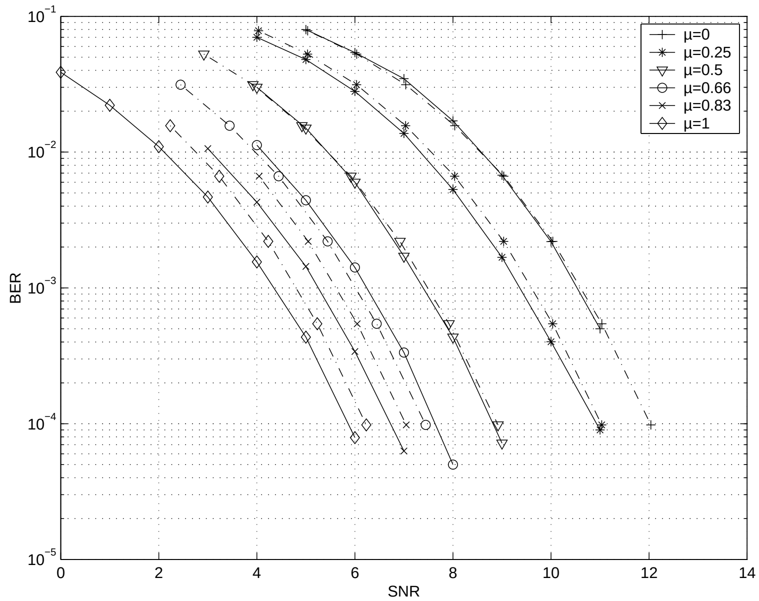

Figure 5: BER versus SNR: comparison of the equalizer performance and the theoretical performance when $\rho = \frac{\sigma_e}{\sigma} = 0.3$, for different values of μ, for Channel3.

length m is equal to 6. In this case, an exhaustive search shows that the error sequence maximizing $P(\mathcal{E}_m)$ is (+2,-2) corresponding to $m' = 2$ and $d' = 0.8$. Figure 8 shows the BER curves for different values of μ when the channel is assumed to be perfectly known at the receiver ($\rho = 0$). For $\mu < 1$, we use the result of Proposition 1, hence the SNR shift is $10\log_{10}\left(\frac{d'^2}{d^2_{\min}}(1+\frac{4m'\mu^2}{d'^2})\right)$. When $\mu = 1$, we use the result of Corollary 4 and the SNR shift is $10\log_{10}\left(\frac{4}{d^2_{\min}}(1+\mu^2)\right)$. We notice that the theoretical curves approximate well the BER.

Figure 9 shows the BER curves for different values of the ratio μ and for $\rho = 0.3$. The approximation holds also in this case.

Remark 5 *The bound given in (6) and the analytical expressions of the error probabilities $P(\mathcal{E}_m)$ we derive in different cases are close at high SNR and loose at low SNR [1]. However, in this paper our aim is to estimate the shift in performance due to the use of the a priori information and the channel estimate and not to estimate the performance. To obtain an expression of this shift, we compare the analytical expression of $P(\mathcal{E}_m)$ when the channel is perfectly estimated and no a priori is available*

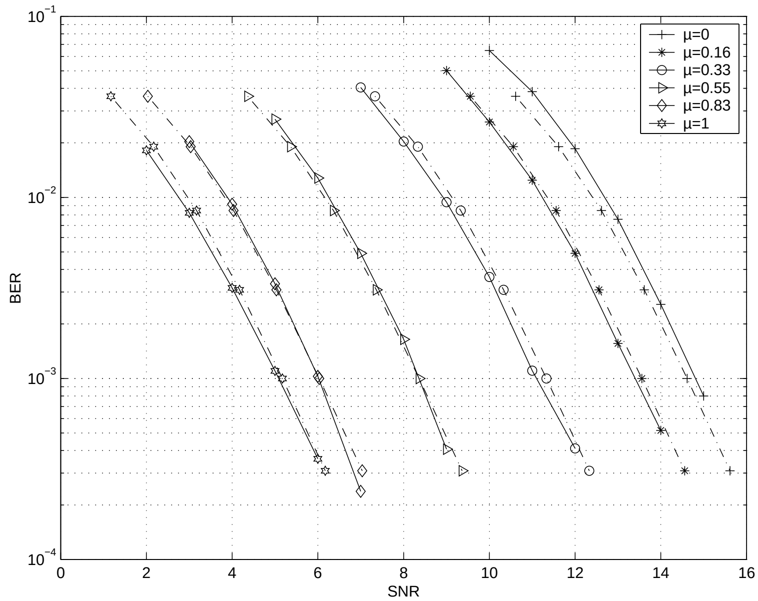

Figure 6: BER versus SNR: comparison of the equalizer performance and the theoretical performance when $\rho = \frac{\sigma_e}{\sigma} = 0.3$, for different values of μ, for Channel5.

with the expression obtained when the channel is estimated and a priori information are available. These analytical approximations have the same behaviors, i.e. they are close at high SNR and loose at low SNR. Therefore, the estimation of the shift remains accurate at low SNR.

4.3 Case of overestimated channel

Until this point, we assumed that the channel length is perfectly known at the channel estimator. Here, we consider the more realistic case where the channel is overestimated. We denote L_1 the estimated channel length. The estimates $\hat{h}_l$ of the tap gains h_l are such as $\hat{h}_l = h_l + \sigma_e k_l$, for $0 \leq l \leq L-1$ and $\hat{h}_l = \sigma_e k_l$, for $L-1 \leq l \leq L_1$, where k_l are independent Gaussian random variables with zero mean and variance 1 and $\sigma_e = \frac{\sigma}{\sqrt{T_0}}$. One can easily check that all the results given in this paper remain valid while replacing L by L_1. Figure 10 shows the BER curves for different values of μ for Channel3, for $\rho = 0.3$, when the channel is overestimated with $L_1 = 6$. The theoretical curves are obtained here by using (16) and (20) and replacing $L = 3$ by $L_1 = 6$. We notice that the theoretical

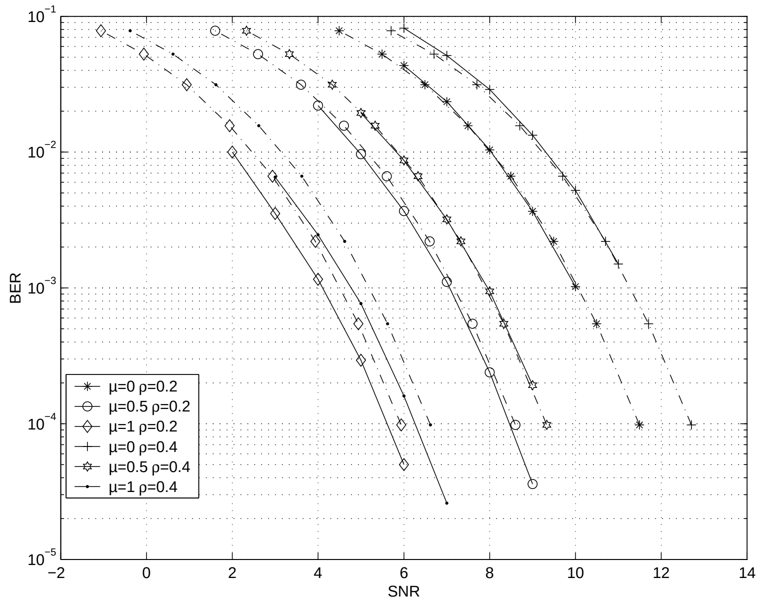

Figure 7: BER versus SNR: comparison of the equalizer performance and the theoretical performance when $\rho = \frac{\sigma_e}{\sigma} \in \{0.4, 0.2\}$, for different values of μ, for Channel3.

curves approximate well the curves obtained by simulations. It is worth mentioning that there exist methods to eliminate the pure noisy estimated channel taps as proposed in [15]. However, we do not consider such methods since it is not the purpose of this paper.

5 Conclusion

In this paper, we consider a MAP equalizer fed with *a priori* information, as in a turbo equalizer, and with a channel estimate obtained by using a training sequence. We propose to study analytically the impact of both the *a priori* information and the channel estimation errors on the MAP equalizer performance. We give an approximation of the error probability which allows us to find an expression of the shift in terms of the SNR due to the use of the *a priori* information and the channel estimate. Simulation results show that the analytical expressions give a good approximation of the equalizer performance. The aim of this work is to obtain in the future the analytical convergence analysis of turbo equalizers using MAP equalization.

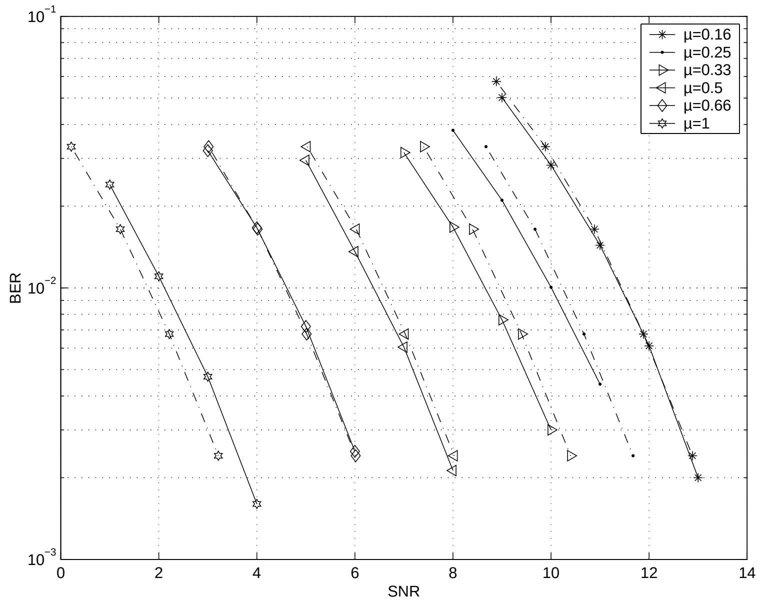

Figure 8: BER versus SNR: comparison of the equalizer performance (solid curves) and the theoretical performance (dotted curves) for Channel7, when the channel is perfectly estimated.

6 Appendix: Proof for Proposition 1

The proof is divided into three parts. First, we derive the probability of an error subevent of length m, $P(\mathcal{E}_m)$, and then upper bound it. Finally, we calculate the overall probability of error, $P(\Sigma)$, in order to find an approximation of the equivalent SNR.

We recall that the output of the channel during a burst is the T-long vector $\underline{x} = (x_{T-1}, ..., x_0)^T$ defined as

$$\underline{x} = \tau(\underline{h})\underline{s} + \underline{n} \tag{22}$$

where $\underline{s} = (s_{T-1}, ..., s_{1-L})^T$ is the $(L+T-1)$-long vector of transmitted symbols, $\underline{n} = (n_{T-1}, ..., n_0)^T$ is the T-long noise vector and $\tau(\underline{h})$ is the $T \times (T+L-1)$ Toeplitz channel matrix.

We assume that the equalizer has at its input a set of *a priori* observations

$$z_k = s_k + w_k \tag{23}$$

where $w_k \sim \mathcal{N}(0, \sigma_a^2)$, for $1 - L \leq k \leq T - 1$.

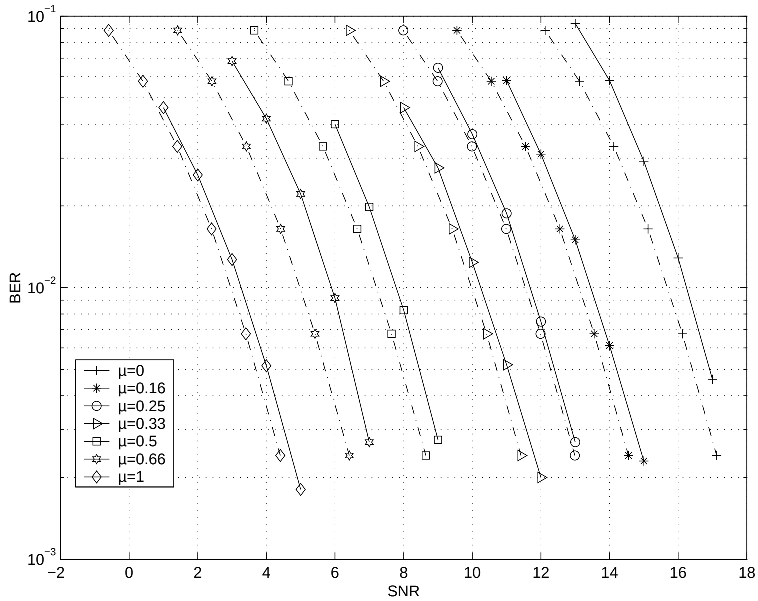

Figure 9: BER versus SNR: comparison of the equalizer performance and the theoretical performance when $\rho = \frac{\sigma_e}{\sigma} = 0.3$, for different values of μ, for Channel7.

We denote $\underline{h}= (h_0, \cdots, h_{L-1})^T$ the vector of true channel parameters, $\hat{\underline{h}}= \left(\hat{h}_0, \cdots, \hat{h}_{L-1}\right)^T$ its estimate and $\Delta\underline{h}=\hat{\underline{h}}-\underline{h}$. We suppose here that a perfect training sequence of length $T_0 + L - 1$ is used and then for $0 \le l \le L-1$, $\hat{h}_l = h_l + \sigma_e k_l$, where k_l are modeled as independent Gaussian random variables with zero mean and variance 1 and $\sigma_e = \frac{\sigma}{\sqrt{T_0}}$.

6.1 Proof-part1: $P(\mathcal{E}_m)$

The MAP equalizer is fed here with an imperfect channel estimate $\hat{\underline{h}}$. In standard works on equalization, the MAP equalizer assumes perfect knowledge of the channel and replaces $\underline{h}$ by $\hat{\underline{h}}$. In [16], the MAP equalization algorithm has been rederived for imperfect channel knowledge, taking into account that $\hat{\underline{h}}$ is different from $\underline{h}$. In this paper, we use the approximation replacing $\underline{h}$ by $\hat{\underline{h}}$. Hence, taking into account the *a priori* information and using $\hat{\underline{h}}$ instead of $\underline{h}$, the *a posteriori* probability

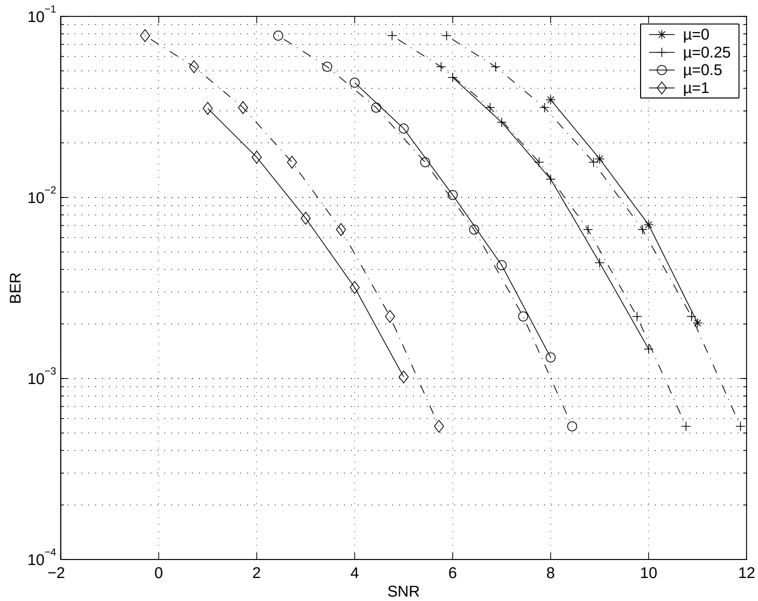

Figure 10: BER versus SNR: comparison of the equalizer performance and the theoretical performance for Channel3 when $\rho = 0.3$ and the estimated channel length is $L_1 = 6$, for different values of μ.

of the sequence $\underline{s}$ is given by

$$p(\underline{s}|\underline{x}, \underline{z}, \underline{h} = \underline{\hat{h}}) \propto \exp\left(-\frac{\left\|\underline{x} - \tau(\underline{\hat{h}})\underline{s}\right\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|\underline{z} - \underline{s}\|^2}{2\sigma_a^2}\right) \tag{24}$$

where $\underline{z}= (z_{T-1}, ..., z_{1-L})^T$ and $\tau(\underline{\hat{h}})$ is a $T \times (T + L - 1)$ Toeplitz matrix associated to $\underline{\hat{h}}$. The data estimate according to the MAP sequence criterion is then given by

$$\underline{\hat{s}} = \arg\min_{\underline{u}} \left(\left\|\underline{x} - \tau(\underline{\hat{h}})\underline{u}\right\|^2 + \frac{\sigma^2}{\sigma_a^2} \|\underline{z} - \underline{u}\|^2 : \underline{u} \in \mathcal{A}^{T+L-1} \right) \tag{25}$$

Now, we consider a particular error event characterized by its length m. We suppose that there exists an interval of size m such that all the symbols of $\underline{\hat{s}}$ are different from the corresponding symbols of $\underline{s}$ while the preceding symbol and the following one are the same for $\underline{s}$ and $\underline{\hat{s}}$. Let $\underline{s}_m = (s_{m+t_0-1}, \cdots, s_{1+t_0-L})$ be the vector of transmitted symbols and $\underline{\hat{s}}_m = (\hat{s}_{m+t_0-1}, \cdots, \hat{s}_{1+t_0-L})$ be the vector of estimated symbols corresponding to the error interval such as for $t_0 \leq k \leq m+t_0-1$, $\hat{s}_k \neq s_k$ and for $t_0 - L + 1 \leq k \leq t_0 - 1$, $\hat{s}_k = s_k$. Let $\underline{e}_m = \underline{\hat{s}}_m - \underline{s}_m$ be the vector of errors and

$\underline{\varepsilon}_m = \tau_m(\underline{h})\underline{e}_m$. A subevent $\mathcal{E}_m$ of the error event of length m is that $\hat{\underline{s}}_m$ is better than $\underline{s}_m$ in the sense of the MAP sequence metric

$$\mathcal{E}_m : \left\|\underline{x}_m - \tau_m(\hat{\underline{h}})\hat{\underline{s}}_m\right\|^2 + \frac{\sigma^2}{\sigma_a^2}\left\|\underline{z}_m - \hat{\underline{s}}_m\right\|^2 \leq \left\|\underline{x}_m - \tau_m(\hat{\underline{h}})\underline{s}_m\right\|^2 + \frac{\sigma^2}{\sigma_a^2}\left\|\underline{z}_m - \underline{s}_m\right\|^2 \tag{26}$$

where $\underline{z}_m$ is the subvector of $\underline{z}$ and $\tau_m(\hat{\underline{h}})$ is the $m \times (m+L-1)$ block of $\tau(\hat{\underline{h}})$ corresponding to the error interval.

Let $\mu = \frac{\sigma}{\sigma_a}$, $\underline{y}= (x_{T-1}, x_{T-2}, \cdots, x_0, \mu z_{T-1}, \cdots, \mu z_{1-L})^T$, $M = \left((\tau(\underline{h}))^T, \mu I_{T+L-1}\right)^T$ a $(2T+L-1) \times (T+L-1)$ matrix and $\underline{b}= (n_{T-1}, n_{T-2}, \cdots, n_0, \mu w_{T-1}, \cdots, \mu w_{1-L})^T$. Using (22) and (23), we can write

$$\underline{y} = M\underline{s} + \underline{b}. \tag{27}$$

Assuming perfect channel knowledge, i.e. using $\hat{\underline{h}}$ instead of $\underline{h}$, the data estimate according to the MAP sequence criterion is given by,

$$\hat{\underline{s}} = \arg\min_{\underline{u}} \left(\left\|\underline{y} - \hat{M}\underline{u}\right\|^2 : \underline{u} \in \mathcal{A}^{T+L-1}\right) \tag{28}$$

where $\hat{M} = \left(\left(\tau(\hat{\underline{h}})\right)^T, \mu I_{T+L-1}\right)^T$. Hence, (26) is equivalent to

$$\mathcal{E}_m : \left\|\underline{y}_m - \hat{M}_m\hat{\underline{s}}_m\right\|^2 \leq \left\|\underline{y}_m - \hat{M}_m\underline{s}_m\right\|^2 \tag{29}$$

where $\underline{y}_m$ is the $(2m+L-1) \times 1$ subvector of $\underline{y}$ corresponding to the error interval and

$$\hat{M}_m = \left(\left(\tau_m(\hat{\underline{h}})\right)^T, \mu I_{m+L-1}\right)^T \tag{30}$$

is a $(2m+L-1) \times (m+L-1)$ matrix.

The event (29) is equivalent to

$$\left\|\hat{M}_m\underline{e}_m\right\|^2 \leq 2\left(\underline{e}_m^T\hat{M}_m^T\left(\underline{y}_m - \hat{M}_m\underline{s}_m\right)\right). \tag{31}$$

Let $M_m = \left((\tau_m(\underline{h}))^T, \mu I_{m+L-1}\right)^T$, $M_m(\Delta\underline{h}) = \hat{M}_m - M_m$ and $\underline{b}_m = \underline{y}_m - M_m\underline{s}_m$, then we obtain

$$\left\|\hat{M}_m\underline{e}_m\right\|^2 \leq 2\left(\underline{e}_m^T\hat{M}_m^T\underline{b}_m - \underline{e}_m^T\hat{M}_m^T M_m(\Delta\underline{h})\,\underline{s}_m\right). \tag{32}$$

This inequality contains products of random variables (for instance $\left\|\hat{M}_m\underline{e}_m\right\|^2$). Therefore, the computation of the probability of the error event is difficult. It can be simplified by noticing that $\hat{M}_m\underline{e}_m$ converges in quadratic mean to $M_m\underline{e}_m$ as T_0 tends to infinity. The following step is to show this convergence. First, we have,

$$\tau_m(\Delta\underline{h})\underline{e}_m \sim \mathcal{N}(0, \|\underline{e}_m\|^2 \frac{\sigma^2}{T_0} I_m)$$

where $\tau_m(\Delta\underline{h}) = \tau_m(\hat{\underline{h}}) - \tau_m(\underline{h})$. Since the modulation used is a BPSK, the m first components of $\underline{e}_m$ are equal to ± 2 and the others are equal to zero, we obtain, $\|\underline{e}_m\|^2 = 4m$. Thus, for finite m, the random process $X_{T_0} = \tau_m(\Delta\underline{h})\underline{e}_m$ converges in quadratic mean to the vector zero as T_0 tends to infinity. From the definition of $\hat{M}_m$ (30), and since μ is independent of T_0, it follows that $\hat{M}_m\underline{e}_m$ converges in quadratic mean to $M_m\underline{e}_m$. Moreover, for all continuous function f, $f\left(\hat{M}_m\underline{e}_m\right)$ converges in quadratic mean to $1_{f(M_m\underline{e}_m)>0}$, where $1_{()}$ is the indicator function. Finally, the event (32) is well approximated by

$$\|M_m\underline{e}_m\|^2 \leq 2\left(\underline{e}_m^T M_m^T \underline{b}_m - \underline{e}_m^T M_m^T M_m\left(\Delta\underline{h}\right)\underline{s}_m\right). \tag{33}$$

Notice that $M_m\left(\Delta\underline{h}\right)\underline{s}_m$ can not be neglected. Indeed,

$$\tau_m(\Delta\underline{h})\underline{s}_m \sim \mathcal{N}(0, \|\underline{s}_m\|^2 \frac{\sigma^2}{T_0} I_m),$$

where $\|\underline{s}_m\|^2 = (m+L-1)^2$ for BPSK. When T_0 and L tend to infinity with the constraint that the ratio $\frac{L^2}{T_0}$ is finite and non zero, the random process $Y_{T_0} = \tau_m(\Delta\underline{h})\underline{s}_m$ converges in quadratic mean to a random variable with a non zero variance.

From $M_m = \left(\left(\tau_m(\underline{h})\right)^T, \mu I_{m+L-1}\right)^T$, we obtain,

$$\begin{aligned}\|M_m\underline{e}_m\|^2 &= \|\tau_m(\underline{h})\underline{e}_m\|^2 + 4m\mu^2 \\ &= \|\underline{\varepsilon}_m\|^2 + 4m\mu^2,\end{aligned} \tag{34}$$

since the m first components of $\underline{e}_m$ are equal to ± 2 and the others are equal to zero.

Hence, (33) can be rewritten as

$$\|\underline{\varepsilon}_m\|^2 + 4m\mu^2 \leq 2\left(\underline{e}_m^T M_m^T \underline{b}_m - \underline{e}_m^T M_m^T M_m\left(\Delta\underline{h}\right)\underline{s}_m\right)$$

Since the lower block of $M_m\left(\Delta\underline{h}\right)$ is equal to zero, we obtain

$$\underline{e}_m^T M_m^T M_m\left(\Delta\underline{h}\right)\underline{s}_m = \underline{\varepsilon}_m^T \tau_m(\Delta\underline{h})\underline{s}_m. \tag{35}$$

Thus,

$$\|\underline{\varepsilon}_m\|^2 + 4m\mu^2 \leq 2\left(\underline{e}_m^T M_m^T \underline{b}_m - \underline{\varepsilon}_m^T \tau_m(\Delta\underline{h})\underline{s}_m\right) \tag{36}$$

where M_m, $\underline{e}_m$, $\underline{\varepsilon}_m$ and $\underline{s}_m$ are deterministic and $M_m\left(\Delta\underline{h}\right)$ and $\underline{b}_m$ are random. Let's define $H_L(\underline{s}_m)$ the Hankel matrix such as $H_L(\underline{s}_m)\Delta\underline{h}= \tau_m(\Delta\underline{h})\underline{s}_m$. We obtain,

$$\|\underline{\varepsilon}_m\|^2 + 4m\mu^2 \leq \chi_s \tag{37}$$

where $\chi_s = 2\left(\underline{e}_m^T M_m^T \underline{b}_m - \underline{\varepsilon}_m^T \tau_m(\Delta\underline{h})\underline{s}_m\right) \sim \mathcal{N}(0, \Delta_s)$ with

$$\begin{aligned}\Delta_s &= E\left(\chi_s \chi_s^T\right)\\ &= 4\underline{e}_m^T M_m^T E\left(\underline{b}_m \underline{b}_m^T\right) M_m \underline{e}_m + 4\underline{\varepsilon}_m^T H_L(\underline{s}_m) E\left(\Delta\underline{h}\Delta\underline{h}^T\right) H_L(\underline{s}_m)^T \underline{\varepsilon}_m\\ &- 4\underline{\varepsilon}_m^T H_L(\underline{s}_m) E\left(\Delta\underline{h}\underline{b}_m^T\right) M_m \underline{e}_m - 4\underline{e}_m^T M_m^T E\left(\underline{b}_m \Delta\underline{h}^T\right) H_L(\underline{s}_m)^T \underline{\varepsilon}_m.\end{aligned} \tag{38}$$

Assuming that $\Delta\underline{h}$ is independent from $\underline{b}_m$ and since $\Delta\underline{h}\sim \mathcal{N}(0, \sigma_e^2 I_L)$ and $\underline{b}_m \sim \mathcal{N}(0, \sigma^2 I_{2m+L-1})$, we obtain,

$$\Delta_s = 4\sigma^2 \underline{e}_m^T M_m^T M_m \underline{e}_m + 4\sigma_e^2 \underline{\varepsilon}_m^T H_L(\underline{s}_m) H_L(\underline{s}_m)^T \underline{\varepsilon}_m.$$

From the law of large numbers, $H_L(\underline{s}_m)H_L(\underline{s}_m)^T$ tends to LI_m as L tends to infinity. Hence, by using (34), $\Delta_s = 4\sigma^2\left(\|\underline{\varepsilon}_m\|^2 + 4m\mu^2\right) + 4\sigma_e^2 L\|\underline{\varepsilon}_m\|^2$, and we obtain that the probability of the error event

$P(\mathcal{E}_m)$ is given by

$$\begin{aligned} P(\mathcal{E}_m) &= \frac{1}{\sqrt{2\pi\Delta_s}} \int_{\|\varepsilon_m\|^2+4m\mu^2}^{+\infty} \exp\left(-\frac{x^2}{2\Delta_s}\right) dx \\ &= Q\left(\frac{\|\varepsilon_m\|^2 + 4m\mu^2}{\sqrt{\Delta_s}}\right) \\ &= Q\left(\frac{\sqrt{\|\varepsilon_m\|^2 + 4m\mu^2}}{2\sigma}\left(1 + \frac{1}{1+\frac{4m\mu^2}{\|\varepsilon_m\|^2}} L\rho^2\right)^{-1/2}\right) \end{aligned} \tag{39}$$

where $\rho = \frac{\sigma_e}{\sigma}$.

6.2 Proof-part2: upper bound for $P(\mathcal{E}_m)$:

In order to find an approximation of $P(\Sigma)$, the overall probability of error, we now want to find an upper bound for $P(\mathcal{E}_m)$. Actually, at high SNR, this term will dominate the sum of the probabilities of the error events (because of the exponential decrease of the Gaussian distribution function). We consider in the following the error sequence which maximizes $P(\mathcal{E}_m)$. Let m' and d' be respectively its length and its norm (after convolution with the channel). Then,

$$P(\mathcal{E}_m) \leq Q\left(\frac{d'}{2\sigma}\sqrt{1 + \frac{4m'\mu^2}{d'^2}}\left(1 + \frac{L\rho^2}{1+\frac{4m'\mu^2}{d'^2}}\right)^{-1/2}\right). \tag{40}$$

6.3 Proof-part3: $P(\Sigma)$

As in the case without *a priori* and perfect channel knowledge, at high SNR, the overall probability of error $P(\Sigma)$ can be approximated by

$$P(\Sigma) \simeq Q\left(\frac{d'}{2\sigma}\sqrt{1 + \frac{4m'\mu^2}{d'^2}}\left(1 + \frac{L\rho^2}{1+\frac{4m'\mu^2}{d'^2}}\right)^{-1/2}\right). \tag{41}$$

Thus, the expression of the error probability given in (41) can be seen as the one given in (7) with an equivalent signal-to-noise ratio

$$\tilde{SNR} = SNR\frac{d'^2}{d_{\min}^2}(1+\frac{4m'\mu^2}{d'^2})\left(1+\frac{L\rho^2}{1+\frac{4m'\mu^2}{d'^2}}\right)^{-1}. \quad (42)$$

References

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

[2] L.R.Bahl, J.Cocke, F.Jelinek, and J.Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory,* vol. IT-32, pp.284-287, March 1974.

[3] A.Gorokhov, "On the performance of the Viterbi equalizer in the presence of channel estimation errors," *IEEE Signal Process. Letters,* vol. 5, no. 12, pp. 321-324, December 1998.

[4] N.Sellami, I.Fijalkow, and S.Perreau, "Performance analysis of a List-type equalizer over estimated MIMO frequency selective channels, " *EUSIPCO'02*, Toulouse, France, September 2002.

[5] C.Douillard, M.Jézéquel, C.Berrou, A.Picart, P.Didier, and A.Glavieux, "Iterative correction of intersymbol interference: turbo-equalization," *European Trans. Telecommun.,* vol. 6, no. 5, pp. 507-511, 1995.

[6] S.Crozier, D.Falconer, and S.Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *IEE Proceedings,* vol. 138, pp.371-378, August 1991.

[7] B.H.Juang and L.R.Rabiner, "The segmental K-means algorithm for estimating the parameters of hidden Markov models," *IEEE Trans. on Acoustics, Speech, and Signal Processing,* vol. 38, no. 9, pp.1639-1641, September 1990.

[8] A.P.Dempster, N.M.Laird, and D.R.Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical society, Series B (Methodological),* vol. 39, no. 1, 1977.

[9] G.K. Kaleh and R. Vallet, ”Joint parameter estimation and symbol detection for linear or nonlinear unknown channels,” *IEEE Trans. on Communications,* vol. 42, no. 7, pp. 2406-2413, July 1994.

[10] S.Ten Brink, “Convergence of iterative decoding,” *IEEE Electronic Letters,* vol. 35, pp.806-808, May 1999.

[11] M.Tüchler, R.Koetter, and A.Singer, “Turbo equalization: principles and new results,” *IEEE Trans. on Comm.*, vol. 50, no. 5, pp. 754-767, May 2002.

[12] A.Roumy, S.Guemghar, G.Caire, and S.Verdu, “Design methods for irregular repeat accumulate codes,” *IEEE Trans. on Inf. Theory*, pp. 1711-1727, August 2004.

[13] S. Benedetto and E. Biglieri, *Principles of Digital Transmission with Wireless Applications,* NewYork: Kluwer/Plenum, 1999.

[14] W.Ser, K.Tan, and K.Ho, “A new method for determining ”unknown” worst-case channels for maximum-likelihood sequence estimation,” *IEEE Trans. on Comm.*, vol. 46, no. 2, pp. 164-168, February 1998.

[15] S.Lasaulce, P.Loubaton, E.Moulines, and S.Buljore, “Training-baised channel estimation and de-noising in the UMTS TDD mode, ” *IEEE Vehicular Technology Conf. (VTC-Fall)*, Atlantic City, USA, October 2001.

[16] M.Tüchler and M.Mecking, “Equalization for non-ideal channel knowledge, ” *Conf. on Inf. Sciences and Systems*, The Johns Hopkins University, March 2003.