Channel Decoding with a Bayesian Equalizer

Luis Salamanca, Juan José Murillo-Fuentes
DTSC University of Seville,
Camino de los Descubrimientos s/n,
41092 Seville, Spain.
{salamanca,murillo}@us.es

Fernando Pérez-Cruz
DTSC University Carlos III in Madrid,
Avda. de la Universidad 30, 28911,
Leganés (Madrid), Spain.
fernando@tsc.uc3m.es

***Abstract*—In this paper we show that, in case of uncertainties during the estimation, overconfident posterior probabilities tend to mislead the performance of soft-decoders. Maximum likelihood (ML) estimates of the channel state information (CSI) make the equalizer to provide overconfident posterior probabilities of the equalized symbols half of the time, that can derail the decoder in case of wrong estimated bits. Thus, as a solution we propose and analyze a Bayesian equalizer that produces more accurate probabilities, because it considers the uncertainties in the estimation. This approach is based on an averaged BCJR over the probability density function of the estimated CSI. We exploit the improvement in the posterior probabilities by feeding the channel decoder with these better estimates. The proposed method exhibits a much better performance compared to the ML-BCJR when a LDPC decoder is considered, as illustrated in the experiments.**

I. Introduction

Communication channels can be characterized by a linear finite impulsive response that either represents the dispersive nature of a physical media or the multiple paths of wireless communications [1]. This representation causes inter-symbol interference (ISI) at the receiver end that can impair the digital communication. Given the channel state information (CSI), the maximum likelihood sequence detector (MLSD) [2], a.k.a. the Viterbi Algorithm, provides the optimal transmitter sequence at the receiver end. And, if we are interested in individual posterior probability estimates for each transmitted bit, we can use the BCJR algorithm [3] that, in a similar fashion to the Viterbi algorithm, is capable of providing bitwise optimal decisions. The CSI is typically acquired from a preamble or mid-amble of known symbols at the receiver [1].

Channel encoders introduce controlled redundancy in the transmitter sequence to be able to correct the errors caused by the channel in the received sequence. Modern channel decoders, such as turbo or low-density parity-check (LDPC) codes [4], need accurate posterior probability estimates to be able to achieve channel capacity [5], and they typically assume independent and identically distributed (i.i.d.) channel realizations with known statistics from which these posterior estimates can be readily computed.

In this paper, we study how the uncertainty in the estimation of the CSI affects the optimal performance of modern channel decoders. If we only consider hard-outputs from the equalization process, it is well known that for long enough training sequences, predictions given by the Viterbi or BCJR algorithms are not severely affected by inaccuracies in the estimate of the CSI, as we are only trying to detect the bit value (shown in this paper). But modern channel decoders, which need accurate posterior probability estimates to operate optimally, can be significantly impaired by these inaccuracies, because we are not only interested in deciding if a bit is a zero or a one, but knowing with what probability it is. In case of inaccuracies, overconfident posterior probabilities provide by the channel equalizer to the decoder can derail the performance of this process, because wrong estimated probabilities leads into bits that are harder to flip. Therefore, we can improve the performance of the decoding stage if somehow we translate the uncertainties in the estimated CSI into less confident posterior probabilities that gave the decoder more freedom to decide over the codeword.

According to the proposed approach, we show that the maximum likelihood (ML) estimates of the CSI, which does not take into account the uncertainties, gives the same number of under and overconfident posterior probability estimates for transmitted bits. We consecutively propose and analyze a Bayesian equalizer, which takes into account the uncertainty in the noise and the uncertainty in the CSI estimate. The Bayesian equalizer underestimates the posterior probabilities compared to the equalizer using ML estimates for the CSI, which improves the performance of LDPC codes. The gains are more significant for high signal to noise ratio, channels with long impulsive responses, and/or short training sequences.

In a previous work, we have shown that accurate posterior probability estimates increase the performance of LDPC decoders [6], although that work focuses on nonlinear channel estimation. In the framework of turbo-receivers [7], some approaches can be found in the literature that incorporate these uncertainties in the iterative process of equalization and decoding. It has been exploited in [8], where the authors use an MMSE to estimate the channel, and in [9], [10], where they do not focus on the optimal estimation of the APP. In [11] we find a proposal to estimate some parameters in a OFDM system to later include them in the decoding.

Throughout this paper we propose a novel Bayesian solution that considers the uncertainties in the estimated CSI, and provides to the decoder underconfident (compared to ML-BCJR) posterior probabilities that ease the correct decoding. As an alternative to an iterative receiver, we consider the performance of both low-density parity-check (LDPC) codes [12],

[13] and BER-optimal BCJR algorithm as equalizer [3]. Both techniques ensure near Shannon-limit results. Furthermore, modern machine learning techniques and inference in graphs afford the lack of huge computational complexity of these algorithms yielding to efficient and near-optimal algorithms to equalize [14] and decode [15], [16]. The LDPC decoder very much benefits from this approach, exhibiting gains of 1 dB with respect to the ML-BCJR solution.

The paper is organized as follows. In Section II we describe the structure of the general communication system proposed and approach the ML estimation. The Bayesian equalization technique is presented in Section III. Thus, experimental results of Section IV shows the performance of our method. Finally, in Section V the results obtained are summarized and future work about our proposal is presented.

II. System Model

We consider the discrete-time dispersive communication system depicted in Fig. 1. The channel $H(z)$ is completely specified by the CSI, i.e. $\mathbf{h} = [h_0, h_1, \ldots, h_L]^\top$, where L is the length of the channel. We model the values of the channel $\mathbf{h}$ as independent, zero-mean and unit-variance Gaussians (Rayleigh fading).

A block of K message bits, $\mathbf{m} = [m_1, m_2, \ldots, m_K]^\top$, is encoded with a rate $R = K/N$ to obtain the codeword $\mathbf{b} = [b_1, b_2, \ldots, b_N]^\top$ that is transmitted over the channel using a BPSK modulation:

$$x_i = \mathbf{b}_i^\top \mathbf{h} + w_i, \tag{1}$$

where $\mathbf{b}_i = [b_i, b_{i-1}, \ldots, b_{i-L+1}]^\top$ and w_i is additive white Gaussian noise (AWGN) with variance σ_n^2. Thus, the received sequence is $\mathbf{x} = [x_1, x_2, \ldots, x_N]^\top$.

We first transmit a preamble with n known bits, $\mathcal{D} = \{x_i^*, b_i^*\}_{i=1}^n$, that are used to estimate the unknown channel at the receiver. Then, we transmit the codeword $\mathbf{b}$.

The ML criterion is widely considered for the task of estimation:

$$\hat{\mathbf{h}}_{ML} = \arg\max_{\mathbf{h}} p(\mathbf{x}^*|\mathbf{b}^*, \mathbf{h}). \tag{2}$$

Once we have estimated the channel coefficients with the preamble, we apply the BCJR algorithm to obtain the posterior probability estimates for each transmitted bit:

$$p(b_i = b|\mathbf{x}, \hat{\mathbf{h}}_{ML}) \qquad i = 1 \ldots N. \tag{3}$$

Finally we decode the received word using the LDPC decoder to obtain a maximum a posteriori estimate for $\mathbf{m}$.

The LDPC decoder relies on the estimates in (3) being accurate and, if they are not, the decoding might not finish or might return an incorrect codeword. In Fig. 2 we show $p(b_i = b|\mathbf{x}, \hat{\mathbf{h}}_{ML})$ versus $p(b_i = b|\mathbf{x}, \mathbf{h})$ for $b = 1$, a channel with no memory (only one tap) and five different training sequences with the same length (that give us five different estimates of the CSI).

We can see in Fig. 2 that the predictions for each bit, if we only consider if the bit is a zero or a one, are quite accurate. Nevertheless, these posterior probability estimates

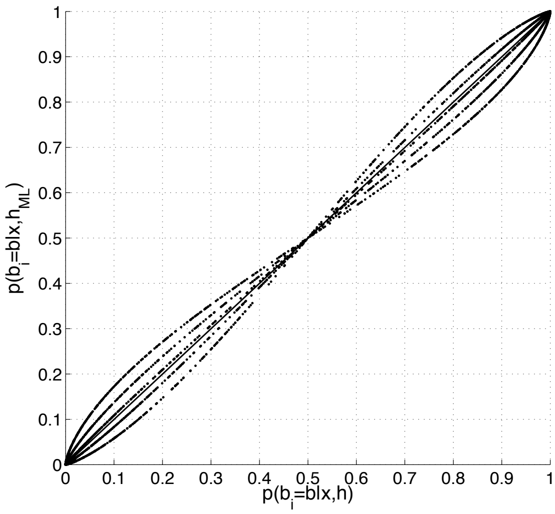

Fig. 2. Calibration curve for the ML-BCJR.

are overconfident half of the time, when $\hat{h}_{ML} > h$. This could be especially pernicious, because a bit with high confidence for a zero or a one is hard to overrule if it is incorrect, which can easily derail the LDPC decoder

III. Bayesian Equalization

We have exposed above how the BCJR, assuming a ML estimation, gives overconfident predictions half of the time. However, this criterion does not assume the uncertainties in the CSI. Therefore, and due to the BCJR takes into account optimally only the uncertainty about the noise, in this section we propose a Bayesian equalizer that takes into account both the uncertainty about the noise and the uncertainty about the CSI estimate. We compute the posterior probability as:

$$p(b_i = b|\mathbf{x}, \mathcal{D}) = \int p(b_i = b|\mathbf{x}, \mathbf{h}) p(\mathbf{h}|\mathcal{D}) d\mathbf{h}, \tag{4}$$

where

$$\begin{aligned} p(\mathbf{h}|\mathcal{D}) &= \frac{p(\mathbf{h}) p(\mathbf{x}^*|\mathbf{b}^*, \mathbf{h})}{p(\mathbf{x}^*|\mathbf{b}^*)} \\ &= \frac{p(\mathbf{h}) \prod_{i=1}^n p(x_i{}^*|\mathbf{b}_i{}^*, \mathbf{h})}{p(x_1{}^*, \ldots, x_n{}^*|b_1{}^*, \ldots, b_n{}^*)}, \end{aligned} \tag{5}$$

is the posterior probability for the CSI, given the likelihood (Gaussian noise) and the prior (Rayleigh fading).

The marginalization of $\mathbf{h}$ in (4) provides at least as accurate posterior probabilities as ML-BCJR because it includes all the information of $p(\mathbf{h}|\mathcal{D})$. To illustrate this idea we have to notice that, if the training sequence is large enough, this Gaussian posterior tends to a multidimensional delta centered at its mean, whose value tends to the real value of $\mathbf{h}$, as $\hat{\mathbf{h}}_{ML}$. Thus, the result of the BCJR algorithm assuming a ML estimation is quite similar to the result of (4). In case of uncertainty in the CSI, the performance of the ML-BCJR is misled due to inaccuracies in the estimation. Conversely, the Bayesian equalizer in (4) considers the information of both the mean and variance of the posterior of $\mathbf{h}$, and hence all

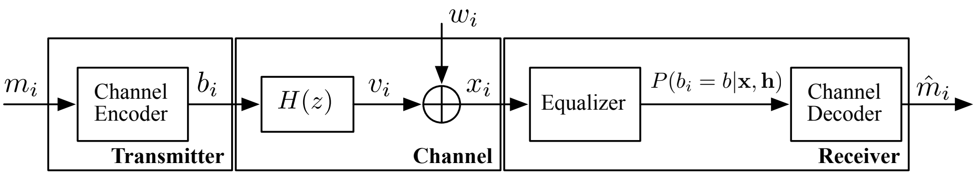

Fig. 1. System model.

the uncertainty in the estimation of the CSI, providing more accurate posterior probabilities.

This computation of the posterior probabilities does not yield a significant improvement in the BER after the equalizer, as we illustrate in Section IV, because for detection we only consider if the probability is lower or higher than 0.5. However, the main advantage of the Bayesian equalizer is that the performance of the soft-decoder is improved with this new estimation of the probabilities, that translates into better results in terms of BER at the output of the decoder. To illustrate this idea we can compare the calibration curves for ML-BCJR (Fig. 2) and Bayesian equalization (Fig. 3). In Fig. 3 we show $p(b_i = b|\mathbf{x}, \mathcal{D})$ versus $p(b_i = b|\mathbf{x}, \mathbf{h})$ for $b = 1$, a channel with no memory and the same training sequences as in Fig. 2. The posterior probabilities obtained assuming a ML estimation are roughly half of the time overconfident and half of the time underconfident, as seen in Fig. 2, because ML is an unbiased estimator of the CSI. For the LDPC decoder, while underconfident estimates can be easily dealt with, these overconfident posterior probability estimates can derail the decoding process, because a bit with high confident in its value cannot be easily flipped in case it is incorrect. The Bayesian equalizer incorporates the uncertainty in the CSI estimate. Its zero-mean Gaussian prior tends to underestimate the CSI. Besides, the averaging over all possibles values of $p(\mathbf{h}|\mathcal{D})$ gives mostly underconfident posterior probability estimates, as we can see in Fig. 3. Therefore, the LDPC decoder does not have a strong preference for the received bits and they can be more easily flipped, if necessary, which improves the performance of the decoder in this case of uncertainties in the estimation of the channel.

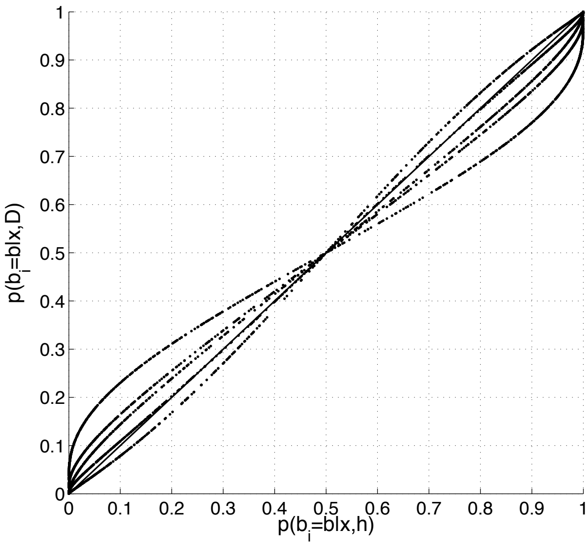

Fig. 3. Calibration curve for Bayesian equalization.

A. Computation of the solution

To deal with the problem of solving (4), we propose to use Monte Carlo to obtain an approximate result. To compute $p(b_i = b|\mathbf{x}, \mathcal{D})$ we consider the following steps:

1) Calculate the posterior of the channel: in (5) the numerator is the product of the likelihood $p(\mathbf{x}^*|\mathbf{b}^*, \mathbf{h})$ and the prior of $\mathbf{h}$. In the proposed system, both are Gaussians distributed as:

$$p(\mathbf{x}^*|\mathbf{b}^*, \mathbf{h}) \sim \mathcal{N}((\mathbf{B}^*)^\top \mathbf{h}, \sigma_n^2 \mathbf{I}), \quad (6)$$

$$p(\mathbf{h}) \sim \mathcal{N}(0, \mathbf{I}), \quad (7)$$

where $\mathbf{B}^* = [\mathbf{b}_1{}^*, \mathbf{b}_2{}^*, \ldots, \mathbf{b}_n{}^*]$ and has dimension $L \times n$, and with no loss of generality we assume that the variance of the coefficients of the channel is equal to one. The expressions for the mean and the covariance matrix of the posterior when both terms are Gaussians are straightforward [17, Chapter 10, Section 6]. These expressions can be particularized for our system as:

$$\overline{\mathbf{h}}_{\mathbf{h}|\mathcal{D}} = (\mathbf{I} + \mathbf{B}^*(\mathbf{B}^*)^\top \sigma_n^{-2})^{-1} \mathbf{B}^* \sigma_n^{-2} \mathbf{x}^*, \quad (8)$$

$$C_{\mathbf{h}|\mathcal{D}} = (\mathbf{I} + \mathbf{B}^*(\mathbf{B}^*)^\top \sigma_n^{-2})^{-1}. \quad (9)$$

2) Produce random samples from the posterior: with the vector of means and the covariance matrix, we can exactly sample to obtain G random samples.
3) Solve the BCJR algorithm: the posterior probability of each transmitted bit is computed for the G different samples of $p(\mathbf{h}|\mathcal{D})$. These G different values of $p(b_i = b|x_1, \ldots, x_N, \mathbf{h})$ average the posterior probability of each transmitted bit over all possible cases of $\mathbf{h}$:

$$p(b_i = b|x_1, \ldots, x_N, \mathcal{D}) \approx$$
$$\approx \frac{1}{G} \sum_{j=1}^{G} p(b_i = b|x_1, \ldots, x_N, \mathbf{h}_j), \quad (10)$$

that is an approximate result of (4).

As discussed before, if we use these posterior probabilities as inputs to a soft-decoder, the system achieves better performance, especially in case of inaccurate estimations of the CSI, i.e. the length of the training sequence is short compared with the number of channel taps. However, this solution is computationally demanding, because we have to calculate

G times the BCJR algorithm, whose complexity increases exponentially with L.

IV. Simulation Results

We use Monte Carlo to obtain the solution of (4), our proposed Bayesian Equalization. To illustrate the better performance of this technique, we compare its bit error rate curves with the ones of the ML-BCJR, before an after the decoder. In all the experiments we consider the following scenario:

- Test sequence of 500 random bits encoded with a regular LDPC code (3,6) of rate $1/2$.
- Up to 10000 frames of 1000 bits are transmitted over the channel. We set a halt condition of 100 wrong decoded frames to avoid unnecessary computations.
- Between frames of information bits, a training sequence of n uncoded bits is transmitted to estimate the channel.
- Every frame of test bits, and its associated training sequence, is sent over the same channel, whose taps are Rayleigh distributed. We consider that the channel coherence time is greater than the duration of the frame, i.e. the channel does not change during this time. Furthermore, in our experiments we take the same value for the taps of the channel during all transmitted frames. For the channel with three taps -$L = 3$- these values are:

$$H(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2},$$

  and for $L = 6$:

$$H(z) = 0.1600 + 0.5450z^{-1} - 0.6720z^{-2} + \\ +0.2560z^{-3} + 0.0950z^{-4} - 0.3890z^{-5}.$$

- The prior of every tap has zero mean and variance equal to 1.

We first depict in Fig. 4 the BER at the output of the equalizer. As predicted, the difference between BER curves is negligible and decreases with the length of the training sequence, since the improvement of the Bayesian equalizer is based on the estimation of the posterior probabilities.

Conversely, it is the decoder the process that takes advantage of these better estimations of the posterior probabilities provided by the Bayesian equalizer, improving its performance and achieving better results in terms of BER.

In Fig. 5 we can see the results for different lengths of the training sequence for a channel with $L = 3$. As we observe in the figure, at higher values of SNR and worst estimations of the channel, due to shorter training sequences, the difference between SNR values of both methods for a certain bit error rate increases. This yields to an approximate gain of 0.5 dB around a SNR of 5.5 dB, and 1 dB in the range of $9 - 10$ dB. BER curve assuming a perfect knowledge of the channel is included too. This sets a lower bound of performance for the system.

As we can conclude from these results, if a worst estimation involves a greater gain, as the uncertainties in the CSI estimation increase, our method has to provide a better performance, and the difference between both curves has to become greater.

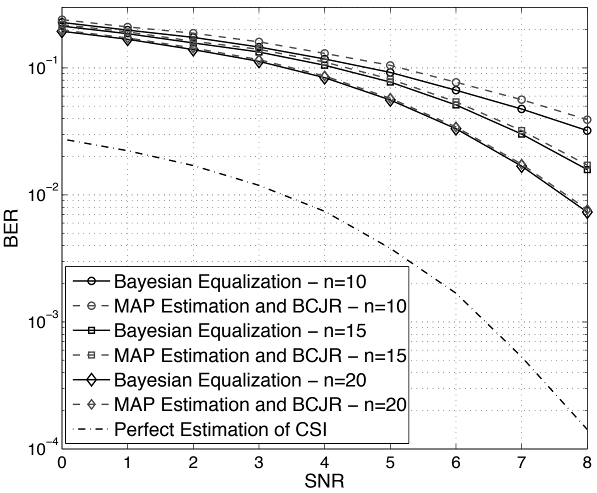

Fig. 4. BER performance for Bayesian equalizer (solid lines) and ML-BCJR (dashed lines) before the decoder, for a channel with 6 taps and different lengths of the training sequence, $n = 10$ (○), $n = 15$ (□) and $n = 20$ (◇). Dash-dotted line illustrates performance assuming perfect CSI.

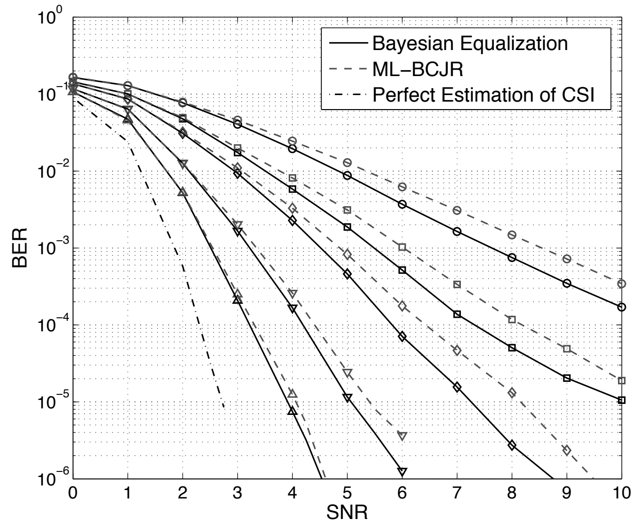

Fig. 5. BER performance for Bayesian equalizer (solid lines) and ML-BCJR (dashed lines) after the LDPC decoder, for a channel with 3 taps and different lengths of the training sequence, $n = 10$ (○), $n = 15$ (□), $n = 20$ (◇), $n = 35$ (▽) and $n = 60$ (△). Dash-dotted line illustrates performance assuming perfect CSI.

In Fig. 6 we illustrate this for a channel with $L = 6$. We can see how the difference between BER curves increases compared to the BER for a channel with 3 taps (Fig. 5), especially for smaller values of n and greater SNR. For this channel we achieve a gain of 1 dB in the range of $7 - 8$ dB.

V. Conclusion and Future Work

The ML-BCJR equalizer, that make use of the ML estimates of the channel state information (CSI), provides the same number of over and underconfident posterior probabilities of each transmitted symbol. Specially when the channel is hard to estimate, these overconfident predictions can assign a value

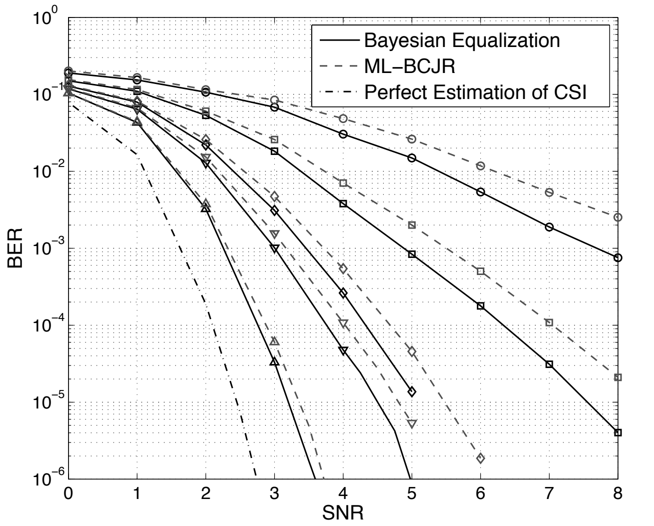

Fig. 6. BER performance for Bayesian equalizer (solid lines) and ML-BCJR (dashed lines) after the decoder, for a channel with 6 taps and different lengths of the training sequence, $n = 15$ ($\circ$), $n = 25$ ($\square$), $n = 40$ ($\diamond$), $n = 50$ ($\bigtriangledown$) and $n = 90$ ($\bigtriangleup$). Dash-dotted line illustrates performance assuming perfect CSI.

near to one to a wrong estimated bit, which degrades the performance of the decoder due to these bits are harder to flip. Therefore, since the probabilities are of major importance in the channel decoding stage, we have proposed and analyzed a Bayesian approach to solve this problem. This approach is based on an averaged BCJR over the probability density function of the estimated posterior of the CSI. Therefore, as we consider the uncertainties in the estimation, this method is optimal as we are interested in the posterior probabilities. If we provide a soft-decoder, such as LDPC, with these estimates, we have an improved estimation of the transmitted codeword. As illustrated in the experiments, the proposed Bayesian equalizer exhibits a much better performance compared to the ML-BCJR, in terms of bit error rate after the LDPC channel decoder.

Despite both the BCJR algorithm and LDPC decoding can be computed efficiently using machine learning algorithms applied on graphs, the drawback of this proposal is its computational complexity, because we have to compute several times the results for the BCJR algorithm to average the integral (4). Alternative graphical representation or inference algorithms that capture the essence of this approach may yield sub-optimal but less demanding solution for Bayesian equalization proposed. Furthermore, other systems and channels can be considered.

Acknowledgment

This work was partially funded by Spanish government (Ministerio de Educación y Ciencia TEC2006-13514-C02-{01,02}/TCM and TEC2009-14504-C02-{01,02}, Consolider-Ingenio 2010 CSD2008-00010), and the European Union (FEDER).

References

[1] J. G. Proakis, *Digital Communications*, 5th ed. New York, NY: McGraw-Hill, 2008.

[2] D. Forney, "The Viterbi algorithm," *IEEE Proceedings*, vol. 61, no. 2, pp. 268–278, Mar. 1973.

[3] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. on Inform. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.

[4] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, March 2008.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New Jersey, USA: John Wiley & Sons, 2006.

[6] P. M. Olmos, J. Murillo-Fuentes, and F. Perez-Cruz, "Joint nonlinear channel equalization and soft LDPC decoding with Gaussian processes," *IEEE Trans. on Signal Processing*, 2010, in press.

[7] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glevieux, "Iterative correction of intersymbol interference: Turbo-equalization," *European Trans. on Telecomm.*, vol. 6, no. 5, pp. 507–511, Sept.-Oct. 1995.

[8] L. M. Davis, I. B. Collings, and P. Hoeher, "Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fast-fading channels," *IEEE Trans. on Commun.*, vol. 49, no. 12, pp. 2106–2114, Dec. 2001.

[9] R. Otnes and M. Tuchler, "On iterative equalization, estimation, and decoding," in *IEEE Int. Conf. on Commun.*, vol. 4, May 2003, pp. 2958–2962.

[10] X. Wang and R. Chen, "Blind Turbo equalization in Gaussian and impulsive noise," *IEEE Trans. on Vehicular Technology*, vol. 50, no. 4, pp. 1092–1105, Jul. 2001.

[11] B. Lu and X. Wang, "Bayesian blind Turbo receiver for coded OFDM systems with frequency offset and frequency-selective fading," in *IEEE Int. Conf. on Commun.*, vol. 1, Apr.-May 2002, pp. 44–48.

[12] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, and V. Stemann, "Efficient erasure correcting codes," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 569–584, Feb. 2001.

[13] S. Chung, D. Forney, T. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the shannon limit," *IEEE Commun. Letters*, vol. 5, no. 2, pp. 58–60, Feb. 2001.

[14] F. R. Kschischang, B. I. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[15] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, and V. Stemann, "Improved low-density parity-check codes using irregular graphs and belief propagation," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 585–598, Feb. 2001.

[16] T. Richardson and R. Urbanke, "Efficient encoding of low-density parity-check codes," *IEEE Trans. on Inform. Theory*, vol. 47, no. 2, pp. 638–656, Feb. 2001.

[17] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation theory*. New York, NY: Prentice-Hall, 1993, vol. 1.