Turbo EP-Based Equalization: A Filter-Type Implementation

Irene Santos, Juan José Murillo-Fuentes, Eva Arias-de-Reyna, *Senior Member, IEEE*, and Pablo M. Olmos

***Abstract*— We propose a novel filter-type equalizer to improve the solution of the linear minimum-mean squared-error (LMMSE) turbo equalizer, with computational complexity constrained to be quadratic in the filter length. When high-order modulations and/or large memory channels are used, the optimal BCJR equalizer is unavailable, due to its computational complexity. In this scenario, the filter-type LMMSE turbo equalization exhibits a good performance compared to other approximations. In this paper, we show that this solution can be significantly improved by using expectation propagation (EP) in the estimation of the *a posteriori* probabilities. First, it yields a more accurate estimation of the extrinsic distribution to be sent to the channel decoder. Second, compared to other solutions based on EP, the computational complexity of the proposed solution is constrained to be quadratic in the length of the finite impulse response. In addition, we review the previous EP-based turbo equalization implementations. Instead of considering default uniform priors, we exploit the outputs of the decoder. Some simulation results are included to show that this new EP-based filter remarkably outperforms the turbo approach of the previous versions of the EP algorithm and also improves the LMMSE solution, with and without turbo equalization.**

***Index Terms*— Expectation propagation (EP), linear MMSE, low-complexity, turbo equalization, ISI, filter-type equalizer.**

I. Introduction

MANY digital communication systems need to transmit over channels that are affected by inter-symbol interference (ISI). The equalizer produces a probabilistic estimation of the transmitted data given the vector of observations [1]. Significant improvements are found when the previous estimation is given to a probabilistic channel decoder [2]. Equalization can be done in the frequency domain to avoid complexity problems associated with the inverse of covariance matrices [3]. In addition, feeding the equalizer back again with the output of the decoder, iteratively, yields a turbo-equalization scheme that significantly reduces the overall error rate [4]–[6].

The BCJR algorithm [7] performs optimal turbo equalization under the maximum a posteriori (MAP) criterion. It provides a posteriori probability (APP) estimations given some *a priori* information about the transmitted data. However, its complexity grows exponentially with the length of the channel and the constellation size, becoming intractable for few taps and/or multilevel constellations. In this situation, approximated BCJR solutions, such as [8]–[11], can be used. They are based on a search over a simplified trellis with only M_e states, yielding a complexity which is linear in this number of states. However, the performance of these approaches is quite dependent on the channel realization and the order of the constellation used. In addition, these approximated BCJR solutions degrade rapidly if the number of survivor paths does not grow according to the total number of states. For these reasons, filter-based equalizers are preferred [12].

A quite extended filter type equalizer in the literature is based on the well-known linear minimum-mean squared-error (LMMSE) algorithm [5], [13], [14]. This LMMSE filter is an appealing alternative where the BCJR is computationally unfeasible due to its robust performance with linear complexity in the frame length, `N`, and quadratic dependence with the window length, `W`. From a Bayesian point of view, the LMMSE algorithm obtains a Gaussian extrinsic distribution by replacing the discrete prior distribution of the transmitted symbols with a Gaussian prior.

A more accurate estimation for the extrinsic distribution can be obtained by replacing the prior distributions with approximations of the probability distribution. This can be done by means of the expectation propagation (EP) algorithm. The EP approach projects the approximated posterior distribution into the family of Gaussians by matching its moments iteratively with the ones of the true posterior. This algorithm has been already successfully applied to multiple-input multiple-output (MIMO) systems [15] and low-density parity-check (LDPC) channel decoding [16], [17], among others. It has been also applied to turbo equalization in a message passing approach as a way to incorporate into the BP algorithm the discrete information coming from the channel decoder [18], [19]. These message passing methods reduce to the LMMSE estimation if no turbo equalization is employed. A different approach is proposed in [20] and [21] under the name of block EP (BEP) where, rather than applying EP after

Manuscript received September 7, 2017; revised January 10, 2018 and March 27, 2018; accepted April 25, 2018. Date of publication May 2, 2018; date of current version September 14, 2018. This work was partially funded by Spanish government (Ministerio de Economía y Competitividad TEC2016-78434-C3-{2-3}-R and Juan de la Cierva Grant No. IJCI-2014-19150) and by the European Union (FEDER). The associate editor coordinating the review of this paper and approving it for publication was M. Morelli. *(Corresponding author: Irene Santos.)*

I. Santos, J. J. Murillo-Fuentes, and E. Arias-de-Reyna are with the Departamento de Teoría de la Señal y Comunicaciones, Escuela Técnica Superior de Ingeniería, Universidad de Sevilla, 41092 Sevilla, Spain (e-mail: irenesantos@us.es; murillo@us.es; earias@us.es).

P. M. Olmos is with the Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, 28911 Madrid, Spain, and also with the Departamento de Teoría de la Señal y Comunicaciones, Instituto de Investigación Sanitaria Gregorio Marañón, 28007 Madrid, Spain (e-mail: olmos@tsc.uc3m.es).

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

Digital Object Identifier 10.1109/TCOMM.2018.2832202

the channel decoder, it is used within the equalizer to better approximate the posterior, outperforming previous solutions.

The computational complexity of previous EP-based equalizers is large for long frame lengths or memories of the channel. Due to its block implementation, the complexity of the BEP is quadratic in the frame length, becoming intractable for large frames [21]. To overcome this drawback, a smoothing EP (SEP) implementation is proposed in [22], but its complexity is cubic with the memory of the channel. Furthermore, due to their iterative procedure, their computational load is roughly S times the one of the LMMSE counterparts, where S is the number of iterations used in the EP algorithm, typically around 10 [15], [20], [21]. Besides, in both, BEP and SEP, uniform discrete priors are assumed for the constellation of the modulations when computing the EP approximations, even within the turbo equalization iterations, while the use of information from the decoder remains unexplored.

The results developed in this paper focus on improving these previous EP-based equalizers [21], [22] both in computational complexity and performance. First, we improve the prior information used in the equalizer once the turbo procedure starts, forcing the true discrete prior to be non-uniform in contrast to the uniform priors used by previous EP-based approaches. As a result, we achieve a performance improvement. Second, the computational complexity of the EP algorithm is reduced to roughly a third part of that in [21], by optimizing the choice of EP parameters. Third, and most important, a new filter-type EP solution is designed. This solution is constrained to have linear complexity in the frame length and quadratic in the filter length, i.e., it is endowed with the same complexity order than the LMMSE filter.

The novel EP-based filter proposed outperforms the LMMSE algorithm with a robust behavior to changes in the constellation size and the channel realization, as the BEP and SEP approaches do [21], [22]. In the experiments included, we show that the EP filter solution greatly improves the LMMSE solution with and without turbo equalization, specifically we have 2 dB gains for a BPSK, 5 dB for the 8-PSK and 6-13 dB for 16 and 64-QAM, respectively. In comparison with previous EP approaches, the EP filter matches their performance with BPSK constellations, and outperforms them with gains of 2 dBs for 8-PSK and 4-5 dBs for 16 and 64-QAM. We study the extrinsic information transfer (EXIT) charts [5], [23] of our proposal for a BPSK, where the EP-based filter achieves the same performance as the BEP.

The scope of this paper encompasses time domain equalization. Frequency domain equalization has received a lot of attention as it usually achieves a complexity reduction for the block-wise processing [3], [14], [24], [25]. For this reason, derivation of a frequency domain counterpart for the proposed EP based turbo-equalizer remains as a future research line. Another promising research route is the application to MIMO with channels with memory [3], [26], [27].

The paper is organized as follows. We first describe in Section II the model of the communication system at hand. Section III is devoted to develop a new implementation of the EP-based equalizer considering non-uniform priors and studies the optimal values for the parameters. In Section IV, we review the formulation for the LMMSE filter in turbo equalization and describe the novel EP filter-type solution proposed. In Section VI, we include several simulations to compare both EP and LMMSE approaches. We end with conclusions.

Through the paper, we denote the i-th entry of a vector $\mathbf{u}$ as u_i, its complex conjugate as $\mathbf{u}^*$ and its Hermitian transpose as $\mathbf{u}^{\mathrm{H}}$. We define $\delta(u_i)$ as the delta function that takes value one if $u_i = 0$ and zero in other case. We use $\mathcal{CN}(\mathbf{u} : \boldsymbol{\mu}, \Sigma)$ to denote a normal distribution of a random proper complex vector $\mathbf{u}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

II. System Model

The model of the communication system is depicted in Fig. 1, including turbo equalization at the receiver. There are three main blocks: transmitter, channel and turbo receiver.

A. Transmitter

The information bit sequence, $\mathbf{a} = [a_1, ..., a_K]^\top$ where $a_i \in \{0, 1\}$, is encoded into the coded bit vector $\mathbf{b} = [b_1, ..., b_V]^\top$ with a code rate equal to $R = K/V$. After permuting the bits with an interleaver, the codeword $\mathbf{c} = [c_1, ..., c_V]^\top$ is partitioned into $\mathtt{N}$ blocks of length $Q = \log_2(\mathtt{M})$, $\mathbf{c} = [\mathbf{c}_1, ..., \mathbf{c}_{\mathtt{N}}]^\top$ where $\mathbf{c}_k = [c_{k,1}, ..., c_{k,Q}]$, and modulated with a complex $\mathtt{M}$-ary constellation $\mathcal{A}$ of size $|\mathcal{A}| = \mathtt{M}$. These modulated symbols, $\mathbf{u} = [u_1, ..., u_{\mathtt{N}}]^\top$, where each component $u_k = \mathcal{R}(u_k) + j\mathcal{I}(u_k) \in \mathcal{A}$, are transmitted over the channel. Hereafter, transmitted symbol energy and energy per bit are denoted as E_s and E_b, respectively.

B. Channel

The channel is completely specified by the CIR, i.e., $\mathbf{h} = [h_1, ..., h_{\mathtt{L}}]^\top$, where $\mathtt{L}$ is the number of taps, and is corrupted with AWGN whose noise variance, σ_w^2, is known. Each k-th entry of the complex received signal $\mathbf{y} = [y_1, ..., y_{\mathtt{N}+\mathtt{L}-1}]^\top$ is given by

$$y_k = \sum_{j=1}^{\mathtt{L}} h_j u_{k-j+1} + w_k = \mathbf{h}^\top \mathbf{u}_{k:k-\mathtt{L}+1} + w_k, \qquad (1)$$

where $w_k \sim \mathcal{CN}(w_k : 0, \sigma_w^2)$ and $u_k = 0$ for $k < 1$ and $k > \mathtt{N}$.

C. Turbo Receiver

When no information is available from the channel decoder, the posterior probability of the transmitted symbol vector $\mathbf{u}$ given the whole vector of observations $\mathbf{y}$ yields

$$p(\mathbf{u}|\mathbf{y}) \ \propto \ p(\mathbf{y}|\mathbf{u})p(\mathbf{u}) \qquad (2)$$

where, assuming equiprobable symbols, the prior would be given by

$$p(\mathbf{u}) = \frac{1}{\mathtt{M}} \prod_{k=1}^{\mathtt{N}} \sum_{u \in \mathcal{A}} \delta(u_k - u). \qquad (3)$$

This prior matches with the definition given in [21] but, as explained below, it is just valid before the turbo procedure.

In a turbo architecture the equalizer and decoder iteratively exchange information for the same set of received

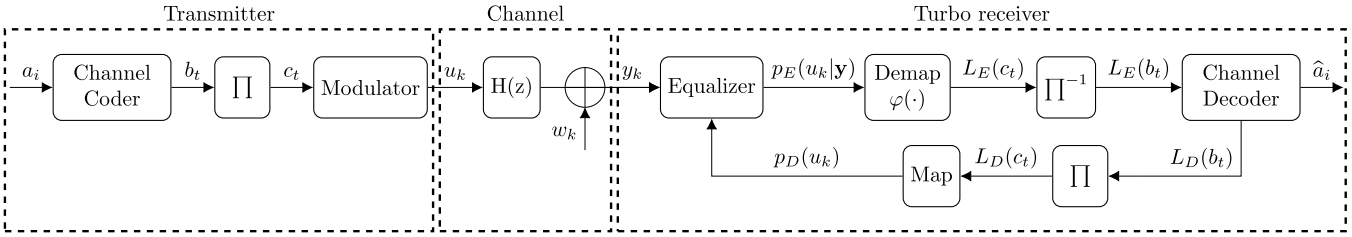

Fig. 1. System model.

symbols [5], [14]. Traditionally, this exchange of information is done in terms of extrinsic probabilities in order to improve convergence and avoid instabilities. The extrinsic information at the output of the equalizer (see Fig. 1), $p_E(u_k|\mathbf{y})$, is computed so as to meet the turbo principle [13]. These probabilities, $p_E(u_k|\mathbf{y})$, are approximated when the optimal solution is intractable. We will denote the approximation by $q_E(u_k)$.

The extrinsic distributions are demapped,

$$L_E(c_{k,j}) = \log \frac{\sum_{u_k \in \mathcal{A}|c_{k,j}=0} p_E(u_k|\mathbf{y})}{\sum_{u_k \in \mathcal{A}|c_{k,j}=1} p_E(u_k|\mathbf{y})}, \tag{4}$$

deinterleaved and given to the decoder as extrinsic log-likelihood ratios, $L_E(b_t)$. The channel decoder computes an estimation of the information bits, $\widehat{\mathbf{a}}$, along with the extrinsic LLRs on the coded bits, computed as

$$L_D(b_t|L_E(\mathbf{b})) = \log \frac{p(b_t=0|L_E(\mathbf{b}))}{p(b_t=1|L_E(\mathbf{b}))} - L_E(b_t). \tag{5}$$

These extrinsic LLRs are interleaved, mapped again and given to the equalizer as updated priors, $p_D(\mathbf{u}|L_E(\mathbf{b}))$, which are computed as

$$p_D(u_k|L_E(\mathbf{b})) = \sum_{u \in \mathcal{A}} \delta(u_k - u) \prod_{j=1}^{Q} p_D(c_{k,j} = \varphi_j(u)|L_E(\mathbf{b})), \tag{6}$$

with $\varphi_j(u)$ denoting the j-th bit associated to the demapping of symbol u. This process is repeated iteratively for a given maximum number of iterations, T, or until convergence. Note that in Fig. 1 we have included the computation of the extrinsic information within the equalization and channel decoding blocks. Note also that, once the turbo procedure starts, the prior in (2) is conditioned on the input at the channel decoder and the symbols are not equiprobable anymore. In this situation, the posterior distribution computed by the equalizer is given by

$$p(\mathbf{u}|\mathbf{y}) \;\propto\; p(\mathbf{y}|\mathbf{u}) \prod_{k=1}^{\mathtt{N}} p_D(u_k|L_E(\mathbf{b})), \tag{7}$$

The true posterior distribution in (2) and (7) has complexity proportional to $\mathtt{M}^{\mathtt{L}}$. When this complexity becomes intractable, we will approximate it, denoting it as $q(\mathbf{u})$. In the following, we omit the dependence on the input at the decoder, $L_E(\mathbf{b})$, to keep the notation uncluttered in the rest of the paper. It is also uncluttered in Fig. 1.

III. Non-Uniform BEP Turbo Equalizer

EP [28]–[32] is a technique in Bayesian machine learning that approximates a (non-exponential) distribution with an exponential distribution whose moments match the true ones. In this paper, we focus on computing a Gaussian approximation for the posterior in (7), which is clearly non Gaussian due to the product of discrete priors in (6). As introduced in [21], this is done by iteratively updating an approximation within the Gaussian exponential family by replacing the non Gaussian prior terms in (2) by a product of Gaussians,[1] i.e.,

$$\begin{aligned} q^{[\ell]}(\mathbf{u}) &\propto p(\mathbf{y}|\mathbf{u}) \prod_{k=1}^{\mathtt{N}} \tilde{p}_D^{[\ell]}(u_k) \\ &= \mathcal{CN}\big(\mathbf{y} : \mathbf{H}\mathbf{u}, \sigma_w^2 \mathbf{I}\big) \prod_{k=1}^{\mathtt{N}} \mathcal{CN}\Big(u_k : m_k^{[\ell]}, \eta_k^{[\ell]}\Big) \end{aligned} \tag{8}$$

The marginalization of the resulting approximated Gaussian posterior distribution for the k-th transmitted symbol and ℓ-th EP iteration yields

$$q^{[\ell]}(u_k) \sim \mathcal{CN}\Big(u_k : \mu_k^{[\ell]}, s_k^{2[\ell]}\Big) \tag{9}$$

where

$$\begin{aligned} \mu_k^{[\ell]} &= m_k^{[\ell]} \\ &\quad + \eta_k^{[\ell]} \mathbf{h_k}^{\mathrm{H}} \left(\sigma_w^2 \mathbf{I} + \mathbf{H}\mathrm{diag}(\boldsymbol{\eta}^{[\ell]})\mathbf{H}^{\mathrm{H}}\right)^{-1} (\mathbf{y} - \mathbf{H}\mathbf{m}^{[\ell]}), \end{aligned} \tag{10}$$

$$s_k^{2[\ell]} = \eta_k^{[\ell]} - \eta_k^{2[\ell]} \mathbf{h_k}^{\mathrm{H}} \left(\sigma_w^2 \mathbf{I} + \mathbf{H}\mathrm{diag}(\boldsymbol{\eta}^{[\ell]})\mathbf{H}^{\mathrm{H}}\right)^{-1} \mathbf{h_k}, \tag{11}$$

$\mathbf{H}$ is the $\mathtt{N}+\mathtt{L}-1 \times \mathtt{N}$ channel matrix given by

$$\mathbf{H} = \begin{bmatrix} h_1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h_{\mathtt{L}} & & \ddots & 0 \\ 0 & \ddots & & h_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h_{\mathtt{L}} \end{bmatrix} \tag{12}$$

and $\mathbf{h_k}$ is the k-th column of $\mathbf{H}$ (see Appendix A for the demonstration). At this point it is interesting to remark that (9)-(11) are completely equivalent to [21, eqs. (15)–(17)]. Here we developed the values of the mean and variance for each symbol while in [21] they were computed in block form.

[1]Note that in [21] we used an alternative expression for (8) (an exponential distribution with parameters $\gamma_k = m_k/\eta_k$ and $\Lambda_k = 1/\eta_k$).

The current description is simpler because we only include the elements of the covariance matrix that are used during the execution of the algorithm, excluding the non-diagonal elements.

The mean and variance parameters in (8) are initialized with the statistics from the channel decoder as

$$m_k^{[1]} = \sum_{u\in\mathcal{A}} u \cdot p_D(u_k = u), \tag{13}$$

$$\eta_k^{[1]} = \sum_{u\in\mathcal{A}} (u - m_k^{[1]})^*(u - m_k^{[1]}) \cdot p_D(u_k = u). \tag{14}$$

Then, they are updated in parallel and iteratively by matching the moments of the following distributions

$$q_E^{[\ell]}(u_k)p_D(u_k) \overset{\text{moment matching}}{\longleftrightarrow} q_E^{[\ell]}(u_k)\mathcal{CN}\Big(u_k : m_k^{[\ell+1]}, \eta_k^{[\ell+1]}\Big) \tag{15}$$

where $q_E^{[\ell]}(u_k)$ is an extrinsic marginal distribution computed as[2]

$$q_E^{[\ell]}(u_k) = q^{[\ell]}(u_k)/\tilde{p}_D^{[\ell]}(u_k) = \mathcal{CN}\Big(u_k : z_k^{[\ell]}, v_k^{2[\ell]}\Big) \tag{16}$$

where

$$z_k^{[\ell]} = \frac{\mu_k^{[\ell]}\eta_k^{[\ell]} - m_k^{[\ell]}s_k^{2[\ell]}}{\eta_k^{[\ell]} - s_k^{2[\ell]}}, \tag{17}$$

$$v_k^{2[\ell]} = \frac{s_k^{2[\ell]}\eta_k^{[\ell]}}{\eta_k^{[\ell]} - s_k^{2[\ell]}}. \tag{18}$$

Note that this equalizer differs from the one in [21] because we used different definitions for the true prior, $p_D(u_k)$. In the current manuscript, we considered non-uniform and discrete priors, given by (6), during the moment matching procedure in the equalizer, while in [21] uniform priors as in (3) were considered by default even after the turbo procedure. To increase the accuracy of the algorithm, a damping procedure follows the moment matching in (15). We have defined an algorithm, described in Algorithm 1, called *Moment Matching and Damping* that runs these two procedures.

A. The nuBEP Algorithm

Algorithm 2 contains a detailed description of the whole EP procedure, where S is the number of EP iterations while T is the number of turbo iterations. Note that the difference with the approach in [21] lies in the definition of the prior distribution used during the moment matching procedure. In [21], we use an uniform distribution (denoted with the indicator function), forcing the same *a priori* probability for the symbols, regardless of the information fed back from the decoder, during the moment matching employed in the equalizer even after the turbo procedure starts. In this paper, we refine the definition of the prior used in the moment matching of EP algorithm as in (6), considering non-uniform priors once the turbo procedure starts. For this reason, we named this algorithm non-uniform BEP (nuBEP) turbo equalizer.

[2]Note that in this paper we used $q_E^{[\ell]}(u_k)$ to denote the extrinsic marginal distribution, while in [21] we denoted as $q^{[\ell]\setminus k}(u_k)$ and called it cavity marginal function.

Algorithm 1 Moment Matching and Damping

Given inputs: $\mu_{p_k}^{[\ell]}, \sigma_{p_k}^{2[\ell]}, z_k^{[\ell]}, v_k^{2[\ell]}, m_k^{[\ell]}, \eta_k^{[\ell]}$

1) Run moment matching: Set the mean and variance of the unnormalized Gaussian distribution

$$q_E^{[\ell]}(u_k) \cdot \mathcal{CN}\Big(u_k : m_{k,new}^{[\ell+1]}, \eta_{k,new}^{[\ell+1]}\Big) \tag{19}$$

equal to $\mu_{p_k}^{[\ell]}$ and $\sigma_{p_k}^{2[\ell]}$, to get the solution

$$\eta_{k,new}^{[\ell+1]} = \frac{\sigma_{p_k}^{2[\ell]}v_k^{2[\ell]}}{v_k^{2[\ell]} - \sigma_{p_k}^{2[\ell]}}, \tag{20}$$

$$m_{k,new}^{[\ell+1]} = \eta_{k,new}^{[\ell+1]}\left(\frac{\mu_{p_k}^{[\ell]}}{\sigma_{p_k}^{2[\ell]}} - \frac{z_k^{[\ell]}}{v_k^{2[\ell]}}\right). \tag{21}$$

2) Run damping: Update the values as

$$\eta_k^{[\ell+1]} = \left(\beta\frac{1}{\eta_{k,new}^{[\ell+1]}} + (1-\beta)\frac{1}{\eta_k^{[\ell]}}\right)^{-1}, \tag{22}$$

$$m_k^{[\ell+1]} = \eta_k^{[\ell+1]}\left(\beta\frac{m_{k,new}^{[\ell+1]}}{\eta_{k,new}^{[\ell+1]}} + (1-\beta)\frac{m_k^{[\ell]}}{\eta_k^{[\ell]}}\right). \tag{23}$$

if $\eta_k^{[\ell+1]} < 0$ **then**

$$\eta_k^{[\ell+1]} = \eta_k^{[\ell]}, \quad m_k^{[\ell+1]} = m_k^{[\ell]}. \tag{24}$$

end if

Output: $\eta_k^{[\ell+1]}, m_k^{[\ell+1]}$

B. On the Election of EP Parameters

The moment matching condition explained in (15) determines the optimal operation point found by the EP approximation. By repeating this procedure, we allow to find a stationary solution for the operation point. In order to avoid instabilities and control the accuracy and speed of convergence, some EP parameters are introduced. These parameters are the number of EP iterations (S), a minimum allowed variance (ϵ) and a damping factor (β). Based on recent studies, these EP parameters can be further optimized [33]–[35]. Following the guidelines in those papers and after extensive experimentation, in the general case we found out that instabilities can be controlled by setting[3] $\epsilon = 1e^{-8}$. Regarding the accuracy of the algorithm, it is convenient to start with a conservative value of the damping parameter β in Algorithm 2. The value $\beta = 0.1$ forces our algorithm to move slowly towards the EP solution. Once the turbo procedure starts, we let the damping parameter grow in order to speed up the achievement of the EP solution, reducing the value of S from 10 in [21] to 3. A simple rule for determination of β that fulfills this requirements and leads to good performance is an exponential growth with a saturation value of 0.7, i.e., $\beta = \min(\exp^{t/1.5}/10), 0.7)$, where $t \in [0, T]$ is the number of the current turbo iteration. With this criterion the number of EP iterations after the turbo procedure starts is reduced to $S = 3$,

[3]Parameters have been chosen to optimize turbo equalization [35].

Algorithm 2 nuBEP Turbo Equalizer

Initialization: Set $p_D(u_k) = \frac{1}{\mathtt{M}} \sum_{u \in \mathcal{A}} \delta(u_k - u)$ for $k = 1, \ldots, \mathtt{N}$
for $t = 1, ..., T$ **do**
1) Compute the mean $m_k^{[1]}$ and variance $\eta_k^{[1]}$ given by (13) and (14), respectively.
for $\ell = 1, ..., S$ **do**
for $k = 1, ..., \mathtt{N}$ **do**
2) Compute the k-th extrinsic distribution as in (16), i.e.,

$$q_E^{[\ell]}(u_k) = \mathcal{CN}\Big(u_k : z_k^{[\ell]}, v_k^{2[\ell]}\Big) \tag{25}$$

where $z_k^{[\ell]}$ and $v_k^{2[\ell]}$ are given by (17) and (18), respectively.
3) Obtain the distribution $\widehat{p}^{[\ell]}(u_k) \propto q_E^{[\ell]}(u_k) p_D(u_k)$ and estimate its mean $\mu_{p_k}^{[\ell]}$ and variance $\sigma_{p_k}^{2[\ell]}$. Set a minimum allowed variance as $\sigma_{p_k}^{2[\ell]} = \max(\epsilon, \sigma_{p_k}^{2[\ell]})$.
4) Run the moment matching and damping procedures by executing Algorithm 1.
end for
end for
5) With the values $m_k^{[S+1]}, \eta_k^{[S+1]}$ obtained after the EP algorithm, calculate the extrinsic distribution $q_E(u_k)$.
6) Demap the extrinsic distribution and compute the extrinsic LLR, $L_E(c_{k,j})$, by means of (4).
7) Run the channel decoder to output $p_D(u_k)$
end for
Output: Deliver $L_E(c_{k,j})$ to the channel decoder for $k = 1, \ldots, \mathtt{N}$ and $j = 1, \ldots, Q$

hence reducing the computational complexity by more than a third.

IV. Filter-Type Turbo Equalization

A. LMMSE Filter

In this subsection we review the formulation of the LMMSE-based filter [5], [13], [14], modified to allow for unnormalized transmitted energy and a different computation of the extrinsic distribution. The LMMSE-based filter [5], [13], [14] estimates one symbol per k-th iteration, u_k, given a $\mathtt{W}$-size window of observations, $\mathbf{y_k} = [y_{k-\mathtt{W}_2}, ..., y_{k+\mathtt{W}_1}]^\top$, where $\mathtt{W} = \mathtt{W}_1 + \mathtt{W}_2 + 1$. This procedure differs from [36], where each transmitted symbol is estimated given the whole vector of observations, $\mathbf{y}$. The LMMSE equalizer approximates the prior for each symbol, $p_D(u_k)$, as a Gaussian

$$p_D(u_k) \approx \tilde{p}_D(u_k) = \mathcal{CN}(u_k : m_k, \eta_k)\,, \tag{26}$$

where the mean, m_k, and variance, v_k, are *a priori* statistics for each transmitted symbol, given by (13) and (14), respectively. For the first iteration of the turbo equalization no *a priori* information is available and a suitable initialization is $m_k = 0$, $\eta_k = E_s$, which boils down to $m_k = 0$, $\eta_k = 1$ when normalizing the energy [5], [13], [14]. Given the current prior and the channel impulse response (CIR), the LMMSE filter computes a Gaussian approximation of the *posterior* probability of each symbol. When a turbo scheme is used, the equalizer and decoder exchange *extrinsic* information [6]. Through the turbo equalization iterations, the *a priori* statistics in (26) are updated with the information fed back from the channel decoder.

Rather than computing the posterior distribution as in (7), the LMMSE filter [5] considers the a posteriori probabilities with respect to the estimated transmitted symbol, $\widehat{u}_k$. For this reason, and to keep the same notation than in [5], we will denote the approximated posterior as $q(u_k|\widehat{u}_k)$. With this posterior distribution in mind, the extrinsic probability at the output of the LMMSE filter can be computed as

$$q_E(u_k|\widehat{u}_k) = \frac{q(u_k|\widehat{u}_k)}{\tilde{p}_D(u_k)}. \tag{27}$$

This distribution is Gaussian and can be derived from the extrinsic distribution of the estimated symbol computed in [5], as shown in Appendix B, yielding

$$q_E(u_k|\widehat{u}_k) = \mathcal{CN}\big(u_k : z_k, v_k^2\big) \tag{28}$$

where

$$z_k = \frac{\mathbf{c_k}^{\mathrm{H}}(\mathbf{y_k} - \mathbf{H_w}\mathbf{m_k} + m_k \mathbf{h_w})}{\mathbf{c_k}^{\mathrm{H}}\mathbf{h_w}}, \tag{29}$$

$$v_k^2 = \frac{\mathbf{c_k}^{\mathrm{H}}\mathbf{h_w} E_s (1 - \mathbf{h_w}^{\mathrm{H}}\mathbf{c_k})}{(\mathbf{c_k}^{\mathrm{H}}\mathbf{h_w})^2}, \tag{30}$$

and, in turn,

$$\mathbf{c_k} = \Big(\mathbf{\Sigma_k} + (E_s - \eta_k)\mathbf{h_w}\mathbf{h_w}^{\mathrm{H}}\Big)^{-1} E_s \mathbf{h_w}, \tag{31}$$

$$\mathbf{H_w} = \begin{bmatrix} h_{\mathtt{L}} & \ldots & h_1 & & \mathbf{0} \\ & \ddots & & \ddots & \\ & & \ddots & & \ddots \\ \mathbf{0} & & & h_{\mathtt{L}} & \ldots \; h_1 \end{bmatrix} \tag{32}$$

is the $\mathtt{W} \times (\mathtt{W} + \mathtt{L} - 1)$ channel matrix, $\mathbf{h_w}$ is the $(\mathtt{W}_2 + \mathtt{L})$-th column of $\mathbf{H_w}$ and

$$\mathbf{m_k} = [m_{k-\mathtt{L}-\mathtt{W}_2+1}, ..., m_{k+\mathtt{W}_1}]^\top, \tag{33}$$

$$\mathbf{V_k} = \mathrm{diag}(\eta_{k-\mathtt{L}-\mathtt{W}_2+1}, ..., \eta_{k+\mathtt{W}_1}), \tag{34}$$

$$\mathbf{\Sigma_k} = \sigma_w^2 \mathbf{I} + \mathbf{H_w}\mathbf{V_k}\mathbf{H_w}^{\mathrm{H}}. \tag{35}$$

The computational complexity is dominated by (31), which has to be recomputed every k-th iteration. Hence, the complexity is $\mathcal{O}(\mathtt{NW}^2)$. This complexity can be further reduced by relying on some approximations proposed in [5] and [14].

B. EP Filter (EP-F)

A novel EP filter-type is developed in this subsection to improve the accuracy and performance of the LMMSE-based filter explained above. As explained in Subsection IV-A, if the LMMSE filter is run, the prior of each symbol is approximated by a Gaussian with the statistics given by the decoder, i.e., with mean and variance given by (13) and (14), respectively. By using the EP algorithm we approximate the posterior distribution with a Gaussian family. Since the posterior distribution includes the true discrete priors, we take into account the discrete nature of symbols.

Algorithm 3 EP-F

Initialization: Set $p_D(u_k) = \frac{1}{\mathtt{M}} \sum_{u \in \mathcal{A}} \delta(u_k - u)$ for $k = 1, \ldots, \mathtt{N}$

for $t = 1, ..., T$ **do**

1) Compute the mean $m_k^{[1]}$ and variance $\eta_k^{[1]}$ given by (13) and (14), respectively.

for $\ell = 1, ..., S$ **do**

for $k = 1, ..., \mathtt{N}$ **do**

2) Compute the k-th extrinsic distribution as in (28), i.e.,

$$q_E^{[\ell]}(u_k|\widehat{u}_k) = \mathcal{CN}\Big(u_k : z_k^{[\ell]}, v_k^{2[\ell]}\Big) \tag{37}$$

where $z_k^{[\ell]}$ and $v_k^{2[\ell]}$ are given by (29) and (30), respectively.

3) Obtain the distribution $\widehat{p}^{[\ell]}(u_k) \propto q_E^{[\ell]}(u_k|\widehat{u}_k) p_D(u_k)$ and estimate its mean $\mu_{p_k}^{[\ell]}$ and variance $\sigma_{p_k}^{2[\ell]}$. Set a minimum allowed variance as $\sigma_{p_k}^{2[\ell]} = \max(\epsilon, \sigma_{p_k}^{2[\ell]})$.

4) Run the moment matching and damping procedures by executing Algorithm 1.

end for

end for

5) With the values $m_k^{[S+1]}, \eta_k^{[S+1]}$ obtained after the EP algorithm, calculate the extrinsic distribution $q_E(u_k|\widehat{u}_k)$ in (28).

6) Demap the extrinsic distribution and compute the extrinsic LLR, $L_E(c_{k,j})$, by means of (4).

7) Run the channel decoder to output $p_D(u_k)$

end for

Output: Deliver $L_E(c_{k,j})$ to the channel decoder for $k = 1, \ldots, \mathtt{N}$ and $j = 1, \ldots, Q$

At every iteration of the EP algorithm, ℓ, we approximate the product of priors of individual symbols in (7) as a product of $\mathtt{N}$ Gaussians, $\tilde{p}_D^{[\ell]}(u_k) = \mathcal{CN}\Big(u_k : m_k^{[\ell]}, \eta_k^{[\ell]}\Big)$, whose parameters (means and variances) are adjusted to find a better approximation, $q^{[\ell]}(\mathbf{u}) \propto p(\mathbf{y}|\mathbf{u}) \prod_{k=1}^{\mathtt{N}} \tilde{p}_D^{[\ell]}(u_k)$, to the true posterior. Similarly to (27)-(28), for each k-th symbol, we first compute the current extrinsic distribution,

$$q_E^{[\ell]}(u_k|\widehat{u}_k) = \frac{q^{[\ell]}(u_k|\widehat{u}_k)}{\tilde{p}_D^{[\ell]}(u_k)}. \tag{36}$$

Now, a more accurate posterior distribution can be obtained by finding a new Gaussian approximation, $\tilde{p}_D^{[\ell+1]}(u_k)$, to match the moments of $q_E^{[\ell]}(u_k|\widehat{u}_k)\tilde{p}_D^{[\ell+1]}(u_k)$ and $q_E^{[\ell]}(u_k|\widehat{u}_k)p_D(u_k)$, as in (15). With these new values for the mean, $m_k^{[\ell+1]}$, and variance, $\eta_k^{[\ell+1]}$, we can recompute a new extrinsic distribution $q_E^{[\ell+1]}(u_k|\widehat{u}_k)$, which is more accurate than the one in (28). The final extrinsic distribution delivered to the decoder is the one obtained after the last iteration of the EP algorithm, following (36).

We denote this new algorithm as EP-filter (EP-F). Algorithm 3 is a detailed description of its implementation. Note that the main difference between Algorithm 2 and Algorithm 3 lies in the computation of the extrinsic distribution, i.e., equations (25) and (37). The computational complexity is also dominated by (31), which has to be computed for each symbol and each ℓ-th iteration. Hence, the complexity is S times the LMMSE complexity, i.e. $\mathcal{O}(S\mathtt{NW}^2)$, where S is the number of iterations of the EP-F. At this point, it is interesting to remark that the approximations proposed in [5] and [14] to further reduce the complexity cannot be applied when the EP is used. The reason is that these approximations remove (at some points) the prior variance computed by the decoder, setting it to one.

V. Relation to Previous Approaches

A. Update of the Priors

We improve the prior information used in the equalizer once the turbo procedure starts, forcing the true discrete prior to be non-uniform in contrast to the uniform priors used by previous EP-based approaches.

In previous proposals [20]–[22], the probabilities from the channel decoder, $p_D(u_k)$, were used to initialize, at the beginning of every iteration of the turbo-equalization, the product of Gaussians that in the EP approximation replaces the product of priors, $\tilde{p}_D^{[1]}(u_k)$. But when the moment matching was performed in the EP algorithm, i.e.,

$$q_E^{[\ell]}(u_k)\mathbb{I}_{u_k \in \mathcal{A}} \overset{\text{moment matching}}{\longleftrightarrow} q_E^{[\ell]}(u_k)\tilde{p}_D^{[\ell+1]}(u_k), \tag{38}$$

the true priors used were uniformly distributed following

$$\mathbb{I}_{u_k \in \mathcal{A}} = \frac{1}{\mathtt{M}} \sum_{u \in \mathcal{A}} \delta(u_k - u). \tag{39}$$

In the current proposal, we keep the initialization of the Gaussians in every step of the turbo-equalization, $\tilde{p}_D^{[1]}(u_k)$, but also *propose to replace the uniform priors in* (39) *by non-uniform ones* in the moment matching step, as explained in (15), i.e.,

$$q_E^{[\ell]}(u_k)p_D(u_k) \overset{\text{moment matching}}{\longleftrightarrow} q_E^{[\ell]}(u_k)\tilde{p}_D^{[\ell+1]}(u_k), \tag{40}$$

where

$$p_D(u_k) = \sum_{u \in \mathcal{A}} \delta(u_k - u) \prod_{j=1}^{Q} p_D(c_{k,j} = \varphi_j(u)), \tag{41}$$

Note that the different definition of priors -(39) in previous proposals, (41) in this manuscript- is the difference between the currently proposed nuBEP algorithm and the BEP in [21], with remarkable improvements.

B. Parameter Optimization

The computational complexity of the EP algorithm is reduced to roughly a third part of that in [21], by optimizing the choice of EP parameters. In particular, we propose some new values for ϵ and β, that control numerical instabilities in the EP updates, and S, the number of iterations of the EP

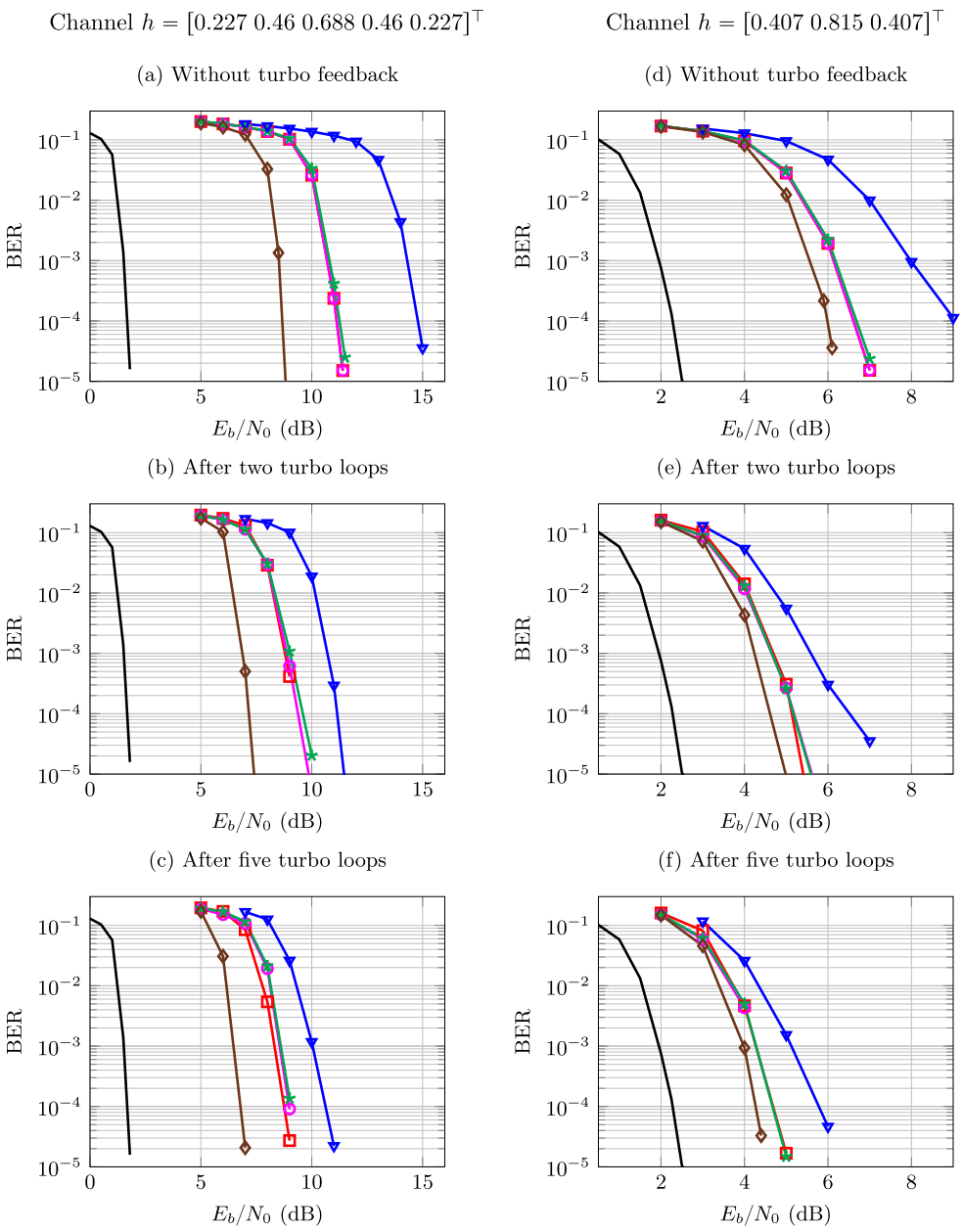

Fig. 2. BER along E_b/N_0 for BEP [21] (▫), nuBEP (○), EP-F (✻), block-LMMSE (▽) and BCJR (◇) turbo equalizers, BPSK, codewords of $V = 4096$ bits (a)-(c) and $V = 1024$ bits (d)-(f) and two different channel responses. Black lines represent the AWGN bound.

equalizer. The parameters proposed in this paper reduce the number of iterations in turbo equalization to $S = 3$, rather than the $S = 10$ iterations that were used in [21].

C. Filter-Type Solution

The new filter-type EP solution proposed is constrained to have linear complexity in the frame length and quadratic in the filter length, i.e., it is endowed with the same complexity order than the LMMSE filter. This complexity is not quadratic with the block length as the one of the BEP [21] nor cubic with the window length as complexity of the SEP [22].

D. Equalization Solved With EP

Regarding the EP-based equalizers proposed by other authors, the approach in [18] and [19] should be mentioned. These proposals deal just with how to pass information between the channel decoder and the LMMSE equalizer. Our proposal first focuses on the EP based equalization, performed independently of the turbo iterations. Therefore the approaches are quite different. Issues such as how to use the priors in the moment matching within the EP equalizer or the damping do not arise in these proposals where the improvement is related only to the handling of probabilities between blocks.

VI. Simulation Results

In this section, we compare the performance of both the block LMMSE and EP-F equalizers for different scenarios. We also include the performance of the BEP [21] and the AWGN bound as references. Note that the MMSE filter [5]

TABLE I
COMPLEXITY COMPARISON BETWEEN ALGORITHMS.

Algorithm	Complexity per turbo iteration
BCJR	$\mathtt{NM}^{\mathtt{L}}$
BEP	$10\mathtt{LN}^2$
block-LMMSE	$\mathtt{LN}^2$
SEP	$10\mathtt{NW}^3$
LMMSE filter	$\mathtt{NW}^2$
nuBEP	$3\mathtt{LN}^2$
EP-F	$3\mathtt{NW}^2$

has not been included in the simulations since the block LMMSE exhibits equal or better performance than any filtering approaches based on the LMMSE algorithm. We did not include the SEP algorithm since it exhibits the same performance as the block implementation, as shown in [22]. We also include the nuBEP approach to illustrate the quite improved behavior when using non-uniform priors at each EP iteration, even reducing from 10 to 3 the number of iterations of the EP approach. The EP parameters have been selected as explained in Subsection III-B, both for the nuBEP and EP-F methods. For a full performance comparison with BCJR approximations, such as M-BCJR [8], M*-BCJR [10], RS-BCJR [9], NZ and NZ-OS [11], please see [21]. In Table I we include a detailed comparison of the complexity of all the simulated algorithms. Above we include the computational complexity of previous algorithms in [21] (BEP) and [22] (SEP), the block and filter implementation of the LMMSE and BCJR approaches. Below we provide the complexity for the new approaches in this paper, i.e., the proposed nuBEP and EP-F. Parameter $\mathtt{W}$ is typically around two times the length of the channel, $\mathtt{L}$. Here, we simulate the scenarios in [13] and [14], using the same channel responses and modulations. Other modulations are also considered. The absolute value of LLRs given to the decoder is limited to 5 in order to avoid very confident probabilities. We use a (3,6)-regular LDPC of rate 1/2, and belief propagation as decoder with a maximum of 100 iterations. The window length in the filtered approach is set to $\mathtt{W} = \mathtt{W}_1 + \mathtt{W}_2 + 1$, where $\mathtt{W}_1 = 2\mathtt{L}$ and $\mathtt{W}_2 = \mathtt{L} + 1$ as suggested in [14].

In the following, we first include a section to analyze the performance of our approach in a low complexity scenario with BPSK modulation, similarly to [14]. The optimal BCJR algorithm can be run in this scenario with a low enough computational complexity and is used as bound. Next, we include a section to analyze the behavior of the algorithms in a large complexity scenario, where we use high-order modulations such as 8-PSK, 16-QAM and 64-QAM.

A. BPSK Scenario

In Fig. 2 we include the BER, averaged over 10^4 random frames, for the LMMSE, BEP [21], nuBEP, EP-F and BCJR equalizers with a BPSK modulation and two different channel responses and lengths of encoded words: $\mathbf{h} = [0.227\ 0.46\ 0.688\ 0.46\ 0.227]^\top$ and $V = 4096$ bits in Fig. 2 (a)-(c) and $h = [0.407\ 0.815\ 0.407]^\top$ and $V = 1024$ bits in Fig. 2 (d)-(f). The channel responses were selected following the simulations in [13] and [14]. The performances of block-algorithms, BEP and nuBEP, are very similar to the equivalent forward filtering approach. When the nuBEP algorithm is applied, 2 and 1.5 dBs gains are obtained compared to LMMSE approach in the turbo scenario, for the two simulated scenarios, respectively. The EP-F exhibits a performance similar to that of the nuBEP.

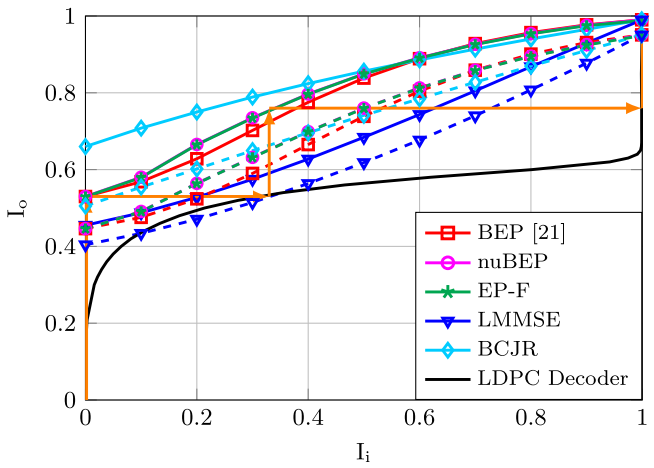

Fig. 3. EXIT charts for the decoder, the BEP [21], nuBEP, EP-F, LMMSE and BCJR equalizers with 7 (dashed) and 9 (solid) dB of E_b/N_0, BPSK modulation and codewords of $V = 4096$.

In Fig. 3 we include the EXIT charts of the BEP [21], nuBEP, EP-F, LMMSE and BCJR for the channel response $\mathbf{h} = [0.227\ 0.46\ 0.688\ 0.46\ 0.227]^\top$ as in [5] and [13], BPSK modulation with $E_b/N_0 = 9$ (solid) and 7 dB (dashed). The EXIT chart of the LDPC encoder of $2048/4096$ and $R = 1/2$ used is also depicted (solid). The horizontal and vertical axis depict the mutual information at the input, I_i, and the output, I_o, respectively. We use arrows to show the evolution of the mutual information along the turbo iterations for $E_b/N_o = 9$ dB. Vertical (horizontal) arrows indicate the improvement in the mutual information each time the equalizer (channel decoder) is executed. When no *a priori* information is given to the decoder, i.e., $I_i = 0$, both BEP and EP-F provide a higher value for the mutual information at the output, I_o, than the LMMSE approach, i.e., they start from a more accurate estimation even before the turbo equalization. This greatly improves the performance as it enlarges the gap between the equalizer and the channel decoder EXIT curves. It can be seen that the LMMSE approach will fail when $E_b/N_o = 7$ dB, because both curves intersect.

Note that the wide EXIT tunnel from the equalizer to the LDPC decoder is suggesting that the code is not optimum in terms of capacity [37]. An optimal code in this sense would exhibit an EXIT chart near below the one of the equalizer and above the curve of the $1/2$ rate LDPC code used. The design of this code for the channel equalization response is out of the scope of this paper and remains as a future line of research.

B. Large Complexity Scenario

In Fig. 4 we simulate the same scenario of Fig. 2, but using an 8-PSK modulation rather than a BPSK. It can be

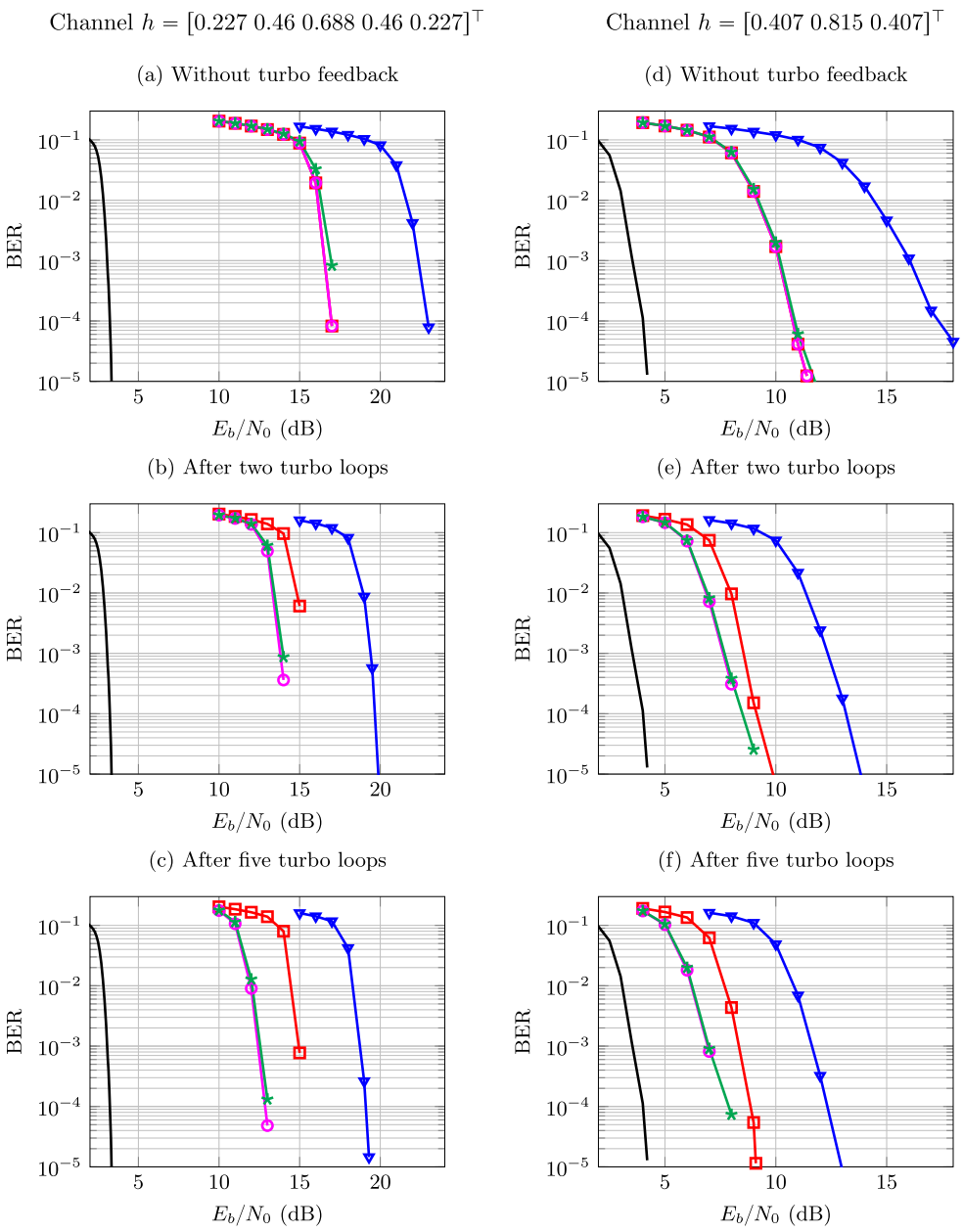

Fig. 4. BER along E_b/N_0 for BEP [21] (▫), nuBEP (○), EP-F (⋆) and block-LMMSE (▽) turbo equalizers, 8-PSK, codewords of $V = 4096$ (a)-(c) and $V = 1024$ (d)-(f) and two different channel responses. Black lines represent the AWGN bound.

observed that after increasing the order of the modulation, the EP-F approach presented in this work performs identically as its block counterpart, greatly improving the performance of the LMMSE algorithm before and after the turbo procedure. An improvement of the BER of the EP-F with respect to the one of the BEP approach in [21], after turbo equalization, can also be observed.

In Fig. 5 we depict the BER performance after five turbo loops for channels $\mathbf{h} = [0.227\ \ 0.46\ \ 0.688\ \ 0.46\ \ 0.227]^\top$ in (a) and $\mathbf{h} = [0.407\ 0.815\ 0.407]^\top$ in (b) with different modulations. We use solid lines to represent a 64-QAM constellation and dashed lines for a 16-QAM. We sent codewords of length $V = 4096$ in both scenarios. It can be observed that the performance of the EP-F matches with the one of its block implementation proposed in this paper (nuBEP) when a 16-QAM is used. However, the EP-F approach slightly degrades with a 64-QAM, where the block nuBEP gets the most accurate performance. Note that the behavior of the EP-F could be improved by increasing the length of the filter, yielding the performance of its block implementation. We have a remarkable improvement of 3-5 dB with respect to the BEP in [21] and of 7-13 dB compared to the LMMSE algorithm.

For the sake of completeness, we include Fig. 6 to show how the BER changes along the turbo iterations and different block lengths at $E_b/N_0 = 13$ dB for an 8-PSK and $\mathbf{h} = [0.227\ \ 0.46\ \ 0.688\ \ 0.46\ \ 0.227]^\top$. The nuBEP algorithm is represented in (a) and the filter approach EP-F in (b). It can be observed that BER is higher for shorter codes and it

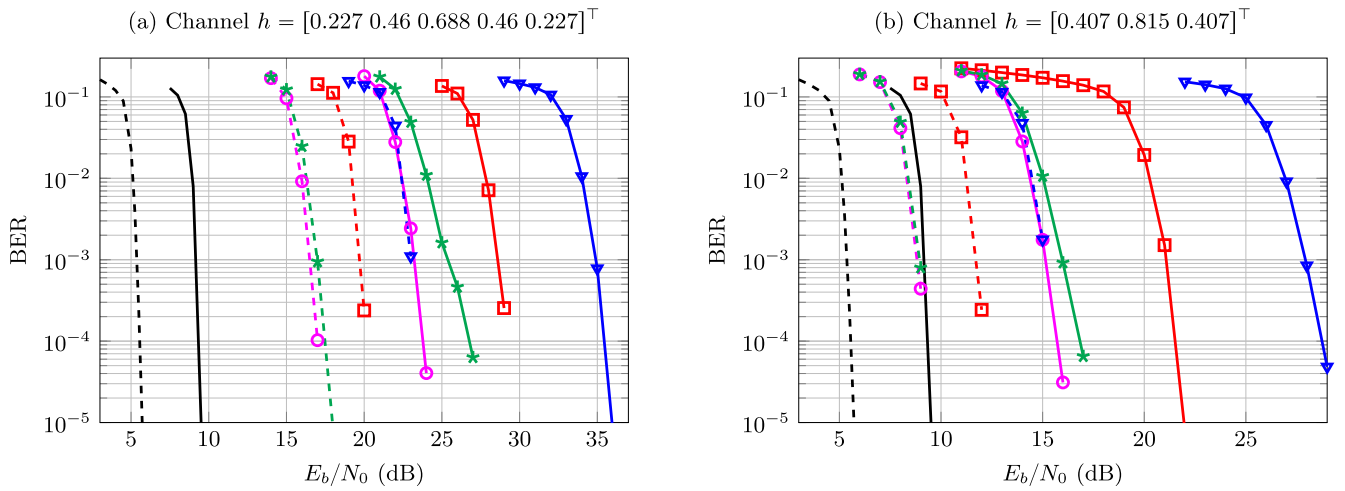

Fig. 5. BER along E_b/N_0 for BEP [21] (▫), nuBEP (○), EP-F (✳) and block-LMMSE (▽) turbo equalizers after five turbo loops, 64-QAM (solid lines) and 16-QAM (dashed lines), codewords of $V = 4096$ and two different channel responses. Black lines represent the AWGN bound.

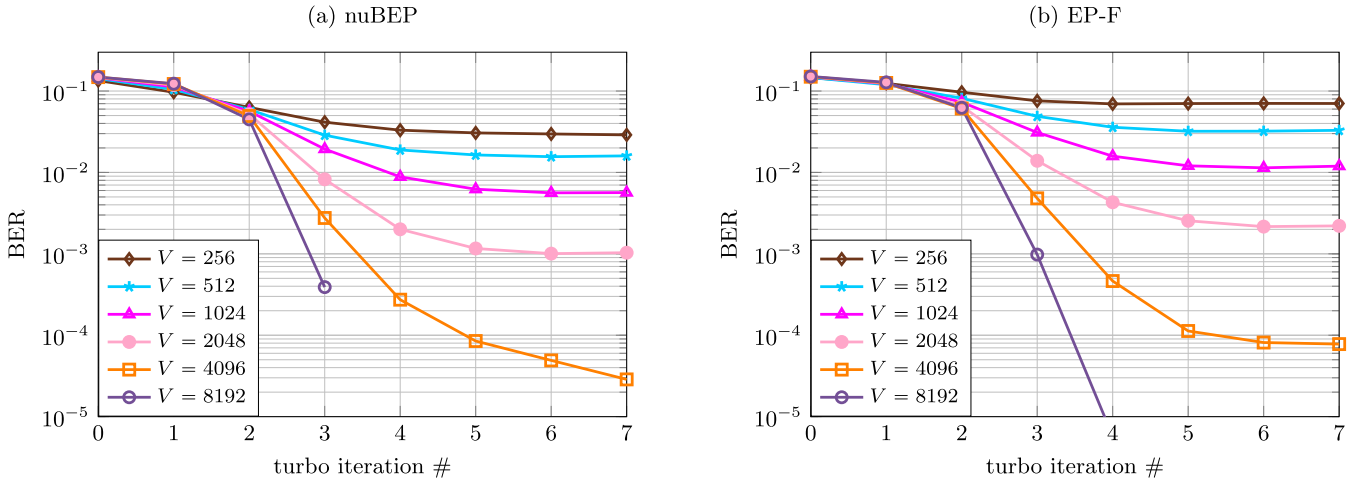

Fig. 6. BER of nuBEP (a) and EP-F (b) turbo equalizers at $E_b/N_0 = 13$ dB for several turbo iterations and lengths of encoded words. 8-PSK modulation and the channel response $\mathbf{h} = [0.227\ 0.46\ 0.688\ 0.46\ 0.227]^\top$ were used.

improves when the code length is increased, as expected. Also, the BER does not significantly improve after the fifth turbo iteration. In the view of these results we stopped running our turbo equalizers after five turbo iterations in the experiments presented above.

VII. Conclusion

In [21], we presented a novel equalizer based on expectation propagation (EP). This solution presents quite an improved performance compared to previous approaches in the literature, both for hard, soft and turbo detection. The solution was presented as a block-wise solution and it was therefore denoted as block-EP (BEP). The major advantage of the BEP lies in the fact that its computational complexity does not grow exponentially with the constellation size and channel memory, as opposed to most equalizers, which are unfeasible for moderate values of these parameters. However, it exhibits a quadratic increase with the size of the transmitted word, V. To avoid this problem, filter-type equalizers are usually preferred [12]. For this reason, we proposed a smoothing EP (SEP) equalizer in [22]. However, the SEP has a computational complexity cubic in the channel length, $\mathtt{L}$. Both BEP and SEP equalizers make use of a moderate feedback in the sense that an initial uniform discrete prior is assumed at the beginning of each execution of the EP algorithm, even after the turbo procedure has started. In this paper, we first propose a design to include the non-uniform discrete nature of the priors from the decoder in the EP algorithm, which amounts to a stronger feedback, quite outperforming the previous BEP and SEP approaches. Second, we develop a reduced-complexity approach by proposing better values of the EP parameters. The resulting algorithm has been denoted as nuBEP, and it significantly outperforms the BEP reducing the computational complexity to less than the third part. Finally, we adapt the EP block equalizer to the filter-type form, emulating the Wiener MMSE filter-type [14]. Therefore, we mimic the structure of the filter-type MMSE equalizer. The EP is used to better approximate the posteriors of a windowed version of the inputs, shifted for every new output estimate. As a result, we present a novel solution dealing with $\mathtt{W}$ inputs at a time

and with quadratic computational complexity in $\mathtt{W}$. This novel solution, the EP-F, despite the reduction in the computational complexity, exhibits a performance in terms of BER quite close to that of its block counterpart, the nuBEP. Furthermore, it remarkably improves the performance of the LMMSE turbo-equalizer, with same complexity order in terms of $\mathtt{L}$ and V. In the included experiments, for channels usually used as benchmarks in the literature, gains in the range 5-13 dB are reported for 8-PSK, 16-QAM and 64-QAM modulations.

One of the main benefits of this new proposal is to reduce the computational complexity, reducing it to be of quadratic order with the filter length. Other approaches, such as those solutions working on the frequency domain [3], could be investigated to achieve this goal. In this paper we face the equalization in single-input single-output channels, the application to MIMO channels with memory [27] remains unexplored.

APPENDIX A
PROOF OF (10) AND (11)

In [21], the posterior distribution used for BEP is

$$q^{[\ell]}(\mathbf{u}) \sim \mathcal{CN}\Big(\mathbf{u} : \boldsymbol{\mu}^{[\ell]}, \boldsymbol{\Sigma}^{[\ell]}\Big) \tag{42}$$

where

$$\boldsymbol{\mu}^{[\ell]} = \boldsymbol{\Sigma}^{[\ell]}(\sigma_w^{-2}\mathbf{H}^{\mathrm{H}}\mathbf{y} + \mathrm{diag}(\boldsymbol{\eta}^{[\ell]})^{-1}\mathbf{m}^{[\ell]}), \tag{43}$$

$$\boldsymbol{\Sigma}^{[\ell]} = \Big(\sigma_w^{-2}\mathbf{H}^{\mathrm{H}}\mathbf{H} + \mathrm{diag}(\boldsymbol{\eta}^{[\ell]})^{-1}\Big)^{-1}. \tag{44}$$

By a direct application of the Woodbury identity, equation (44) can be rewritten as

$$\boldsymbol{\Sigma}^{[\ell]} = \mathrm{diag}(\boldsymbol{\eta}^{[\ell]}) - \mathrm{diag}(\boldsymbol{\eta}^{[\ell]})\mathbf{H}^{\mathrm{H}}\mathbf{C}^{-1}\mathbf{H}\mathrm{diag}(\boldsymbol{\eta}^{[\ell]}) \tag{45}$$

where

$$\mathbf{C} = \mathbf{H}\mathrm{diag}(\boldsymbol{\eta}^{[\ell]})\mathbf{H}^{\mathrm{H}} + \sigma_w^2\mathbf{I}. \tag{46}$$

The k-th diagonal element of (45) yields (11). Regarding (43), it can be divided into two terms

$$\boldsymbol{\mu}^{[\ell]} = \underbrace{\boldsymbol{\Sigma}^{[\ell]}\sigma_w^{-2}\mathbf{H}^{\mathrm{H}}\mathbf{y}}_{T_1} + \underbrace{\boldsymbol{\Sigma}^{[\ell]}\mathrm{diag}(\boldsymbol{\eta}^{[\ell]})^{-1}\mathbf{m}^{[\ell]}}_{T_2}. \tag{47}$$

We apply the following identity [31],

$$(\mathbf{A}^{-1} + \mathbf{B}^{\mathrm{H}}\mathbf{D}^{-1}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{H}}\mathbf{D}^{-1} = \mathbf{A}\mathbf{B}^{\mathrm{H}}(\mathbf{B}\mathbf{A}\mathbf{B}^{\mathrm{H}} + \mathbf{D})^{-1} \tag{48}$$

to the first term, T_1, in (47), yielding

$$T_1 = \mathrm{diag}(\boldsymbol{\eta}^{[\ell]})\mathbf{H}^{\mathrm{H}}\mathbf{C}^{-1}\mathbf{y}. \tag{49}$$

Now, we replace (45) into the second term, T_2, in (47), obtaining

$$T_2 = \mathbf{m}^{[\ell]} - \mathrm{diag}(\boldsymbol{\eta}^{[\ell]})\mathbf{H}^{\mathrm{H}}\mathbf{C}^{-1}\mathbf{H}\mathbf{m}^{[\ell]}. \tag{50}$$

By replacing (49) and (50) into (47), we finally get

$$\boldsymbol{\mu}^{[\ell]} = \mathbf{m}^{[\ell]} + \mathrm{diag}(\boldsymbol{\eta}^{[\ell]})\mathbf{H}^{\mathrm{H}}\mathbf{C}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{m}^{[\ell]}), \tag{51}$$

whose k-th element is given by (10).

APPENDIX B
PROOF OF (29) AND (30)

In [5], the extrinsic distribution of the estimated symbol is computed as

$$q_E(\widehat{u}_k|u_k) \sim \mathcal{CN}\big(\widehat{u}_k : u_k\mathbf{c_k}^{\mathrm{H}}\mathbf{h}_{\mathtt{W}}, \sigma_k^2\big) \tag{52}$$

where

$$\widehat{u}_k = \mathbf{c_k}^{\mathrm{H}}(\mathbf{y_k} - \mathbf{H}_{\mathtt{W}}\mathbf{m_k} + m_k\mathbf{h}_{\mathtt{W}}), \tag{53}$$

$$\sigma_k^2 = \mathbf{c_k}^{\mathrm{H}}\mathbf{h}_{\mathtt{W}}E_s(1 - \mathbf{h}_{\mathtt{W}}{}^{\mathrm{H}}\mathbf{c_k}), \tag{54}$$

$\mathbf{c_k}$ is given by (31) and $\mathbf{h}_{\mathtt{W}}$ is the $(\mathtt{W}_2 + \mathtt{L})$-th column of $\mathbf{H}_{\mathtt{W}}$ defined in (32). Note that we generalized the expressions in [5] to consider a symbol energy of E_s. If we set $E_s = 1$, we obtain exactly the formulation in [5]. Instead of the extrinsic distribution of the estimated symbol, we use in our formulation the extrinsic distribution of the true symbol, which can be computed from (52) as

$$q_E(u_k|\widehat{u}_k) \sim \mathcal{CN}\big(u_k : z_k, v_k^2\big) \tag{55}$$

where

$$z_k = \frac{\widehat{u}_k}{\mathbf{c_k}^{\mathrm{H}}\mathbf{h}_{\mathtt{W}}}, \tag{56}$$

$$v_k^2 = \frac{\sigma_k^2}{(\mathbf{c_k}^{\mathrm{H}}\mathbf{h}_{\mathtt{W}})^2}, \tag{57}$$

yielding the formulation in (29) and (30).

REFERENCES

[1] S. Haykin, *Communication Systems*, 5th ed. Hoboken, NJ, USA: Wiley, 2009.

[2] L. Salamanca, J. J. Murillo-Fuentes, and F. Perez-Cruz, "Bayesian equalization for LDPC channel decoding," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2672–2676, May 2012.

[3] J. Karjalainen, N. Veselinovic, K. Kansanen, and T. Matsumoto, "Iterative frequency domain joint-over-antenna detection in multiuser MIMO," *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3620–3631, Oct. 2007.

[4] C. Douillard, M. Jézéquel, C. Berrou, P. Didier, and A. Picart, "Iterative correction of intersymbol interference: Turbo-equalization," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 507–512, 1995.

[5] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, May 2002.

[6] R. Koetter, A. Singer, and M. Tuchler, "Turbo equalization," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 67–80, Jan. 2004.

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

[8] V. Franz and J. B. Anderson, "Concatenated decoding with a reduced-search BCJR algorithm," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 186–195, Feb. 1998.

[9] G. Colavolpe, G. Ferrari, and R. Raheli, "Reduced-state BCJR-type algorithms," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 848–859, May 2001.

[10] M. Sikora and D. Costello, "A new SISO algorithm with application to turbo equalization," in *Proc. Int. Symp. Inf. Theory, (ISIT)*, Sep. 2005, pp. 2031–2035.

[11] D. Fertonani, A. Barbieri, and G. Colavolpe, "Reduced-complexity BCJR algorithm for turbo equalization," *IEEE Trans. Commun.*, vol. 55, no. 12, pp. 2279–2287, Dec. 2007.

[12] C. Berrou, Ed., *Codes and Turbo Codes* (Collection IRIS). Paris, France: Springer, 2010.

[13] M. Tuchler, A. C. Singer, and R. Koetter, "Minimum mean squared error equalization using *a priori* information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.

[14] M. Tuchler and A. C. Singer, "Turbo equalization: An overview," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 920–952, Feb. 2011.
[15] J. Cespedes, P. M. Olmos, M. Sanchez-Fernandez, and F. Perez-Cruz, "Improved performance of LDPC-coded MIMO systems with EP-based soft-decisions," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2014, pp. 1997–2001.
[16] P. M. Olmos, J. J. Murillo-Fuentes, and F. Pérez-Cruz, "Tree-structure expectation propagation for LDPC decoding over the BEC," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3354–3377, Jun. 2013.
[17] L. Salamanca, P. M. Olmos, F. Perez-Cruz, and J. J. Murillo-Fuentes, "Tree-structured expectation propagation for LDPC decoding over BMS channels," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4086–4095, Oct. 2013.
[18] J. Hu, H. Loeliger, J. Dauwels, and F. Kschischang, "A general computation rule for lossy summaries/messages with examples from equalization," in *Proc. 44th Allerton Conf. Commun., Control, Comput.*, Sep. 2006, pp. 27–29.
[19] P. Sun, C. Zhang, Z. Wang, C. N. Manchón, and B. H. Fleury, "Iterative receiver design for ISI channels using combined belief- and expectation-propagation," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1733–1737, Oct. 2015.
[20] I. Santos, J. J. Murillo-Fuentes, and P. M. Olmos, "Block expectation propagation equalization for ISI channels," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Aug./Sep. 2015, pp. 379–383.
[21] I. Santos, J. J. Murillo-Fuentes, R. Boloix-Tortosa, E. Arias-de-Reyna, and P. M. Olmos, "Expectation propagation as turbo equalizer in ISI channels," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 360–370, Jan. 2017.
[22] I. Santos, J. J. Murillo-Fuentes, E. Arias-de-Reyna, and P. M. Olmos, "Probabilistic equalization with a smoothing expectation propagation approach," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2950–2962, May 2017.
[23] K. Kansanen and T. Matsumoto, "An analytical method for MMSE MIMO turbo equalizer EXIT chart computation," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 59–63, Jan. 2007.
[24] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58–66, Apr. 2002.
[25] M. Tuchler and J. Hagenauer, "'Turbo equalization using' frequency domain equalizers," in *Proc. Allerton Conf. Commun., Control, Comput.*, Oct. 2000, pp. 1234–1243.
[26] N. Veselinovic, T. Matsumoto, and M. Juntti, "Iterative pdf estimationbased multiuser diversity detection and channel estimation with unknown interference," *EURASIP J. Appl. Signal Process.*, vol. 6, pp. 872–882, Jan. 2005.
[27] L. Xiao *et al.*, "Time-domain turbo equalization for single-carrier generalized spatial modulation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5702–5716, Sep. 2017.
[28] T. P. Minka, "A family of algorithms for approximate Bayesian inference," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2001.
[29] T. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 17th Conf. Uncertainty Artif. Intell. (UAI)*, 2001, pp. 362–369.
[30] M. Seeger, "Expectation propagation for exponential families," Dept. EECS, Univ. California Berkeley, Berkeley, CA, USA, Tech. Rep., 2005.
[31] C. M. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag, 2006.
[32] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 501–505.
[33] M. Opper and O. Winther, "Expectation consistent approximate inference," *J. Mach. Learn. Res.*, vol. 6, pp. 2177–2204, Dec. 2005.
[34] J. Cespedes, "Approximate inference in massive MIMO scenarios with moment matching techniques," Ph.D. dissertation, Dept. Teoría Señal Comunicaciones, Univ. Carlos III Madrid, Madrid, Spain, Jan. 2017.
[35] J. Céspedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Probabilistic MIMO symbol detection with expectation consistency approximate inference," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3481–3494, Apr. 2018.
[36] J. Wu and Y. R. Zheng, "Low complexity soft-input soft-output block decision feedback equalization," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Nov. 2007, pp. 3379–3383.
[37] Q. Xie, Z. Yang, J. Song, and L. Hanzo, "EXIT-chart-matching-aided near-capacity coded modulation design and a BICM-ID design example for both Gaussian and Rayleigh channels," *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1216–1227, Mar. 2013.

Irene Santos received the M.Sc. degrees in telecommunication and electrical engineering from the Universidad de Sevilla, in 2013 and 2014, respectively, where she is currently pursuing the Ph.D. degree with the Department of Signal Theory and Communications. She has been a Visiting Researcher with Stony Brook University, Stony Brook, NY, USA, and also with the Universidad Carlos III de Madrid. Her principal research interests are focused on Bayesian inference techniques and their applications to digital communication systems.

Juan José Murillo-Fuentes received the Degree in telecommunication engineering from the Universidad de Sevilla, in 1996, and the Ph.D. degree in telecommunication engineering from the Universidad Carlos III de Madrid, Spain, in 2001. Since 2016, he has been a Full Professor with the Universidad de Sevilla. His research interests lie in algorithm development for signal processing and machine learning and their applications to digital communication systems and learning from images.

Eva Arias-de-Reyna (M'07–SM'16) received the M.Sc. and Ph.D. degrees in telecommunication engineering from the Universidad de Sevilla, Spain, in 2001 and 2007, respectively. In 2002, she joined the Department of Signal Theory and Communications, Universidad de Sevilla, where she is currently an Associate Professor. Her current research interests include machine learning, communication receiver design, localization techniques, and ultra wideband technology.

Pablo M. Olmos received the M.Sc. and Ph.D. degrees in telecommunication engineering from the Universidad de Sevilla, in 2008 and 2011, respectively. He held a visiting researcher position with Princeton University, Princeton, NJ, USA; EPFL; Notre Dame University, Notre Dame, IN, USA; ENSEA; and Bell Labs, Murray Hill, NJ, USA. He is currently an Assistant Professor with the Universidad Carlos III de Madrid. His research interests range from approximate inference methods for Bayesian machine learning to information theory and digital communications.