

Iterative Receiver Design for Probabilistic Constellation Shaping in ISI Channel

XIANG LI^{ID}, (Graduate Student Member, IEEE), JING WU, (Member, IEEE),
WEI HENG, (Member, IEEE), AND YANG HUANG^{ID}

National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China

Corresponding author: Wei Heng (wheng@seu.edu.cn)

This work was supported by the National Nature Science Foundation of China (NSFC) Projects under Grant 61771132.

ABSTRACT This paper investigates the receiver design for probabilistic constellation shaping signaling over inter-symbol interference channel. The key component performing the constellation shaping is an adjustable distribution matcher, and the probabilistic shaping system is capable to adapt to variable data rates by adjusting the distribution match rate rather than the modulation or coding mode. In this paper, we resort to distinct techniques to derive two iterative receivers operating in time domain. Shaped-BCJR is a trellis-based solution where forward/backward algorithm together with turbo iteration is used to compute the posteriori probability in which the nonuniform a priori symbol probability initializes the calculation. Another receiver is based on linear filtering where expectation propagation provides a Gaussian approximation of the posteriori probability of each symbol. This receiver structure is represented by a factor graph and we detail the derivation of messages exchanging between adjacent nodes. The Bayesian equalization in each EP iteration brings unacceptable computational burden. An efficient block-wise matrix inversion strategy is proposed to tackle this problem, significantly reducing the computational complexity with little performance loss. Simulation results show that the proposed algorithms remarkably outperform LMMSE solutions and traditional EP based algorithms. Proposed matrix inversion strategy can also be used to improve the performance of other filter-type solutions.

INDEX TERMS Probabilistic constellation shaping, distribution match, BCJR, decision feedback equalization, expectation propagation.

I. INTRODUCTION

High-order modulation and flexible channel coding scheme play important roles in optimizing the spectral efficiency (SE) in wired and wireless communication systems. To promote the SE, modulation formats should have a Gaussian-like shape [1]. Huffman code based matcher is proposed in [2] to achieve this goal, however, the codebook must be pre-generated and stored offline. Realtime solution is proposed in [3] where arithmetic coding is used to calculate the codebook online. Recently, Constant Composition Distribution Matcher (CCDM) [4] is proposed to map uniformly distributed data bits to non-uniform amplitudes with a desired distribution. By properly choosing the distribution, shaping gain predicted in theory can be achieved by a practical scheme with a low complexity. Probabilistic Constellation Shaping (PCS) is a multilayer code modulation (CM) scheme implemented by

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wang^{ID}.

combining DM with a systematic binary low-density parity-check (LDPC) code [5]. This system has shown an advantage of flexibility in terms of transmission rate even though fixed forward error correction (FEC) code, modulation format and bandwidth are used. Latest works on optimizing the distribution and practical implements can be found in [6] and [7].

We show the comparison of SE between non-equiprobable shaping QAM, conventional equiprobable shaping QAM and the Shannon limit in Fig. 1. Both CM schemes use LDPC code as FEC and the operating points target at where the frame error rate (FER) is no more than 10^{-2} . The PCS scheme uses only two CM modes, namely, 16-QAM with a rate 3/4 code and 64-QAM with a rate 4/5 code. However, equiprobable constellation scheme uses up to 8 modes, i.e., combining {4, 16, 64}-QAM with rate {3/5, 2/3, 3/4, 4/5, 5/6, 9/10} codes. As shown in the figure, the gap between the operating points of equiprobable scheme and the AWGN channel capacity varies along different SNRs. Two factors contribute to this gap: one is that LDPC code has a finite

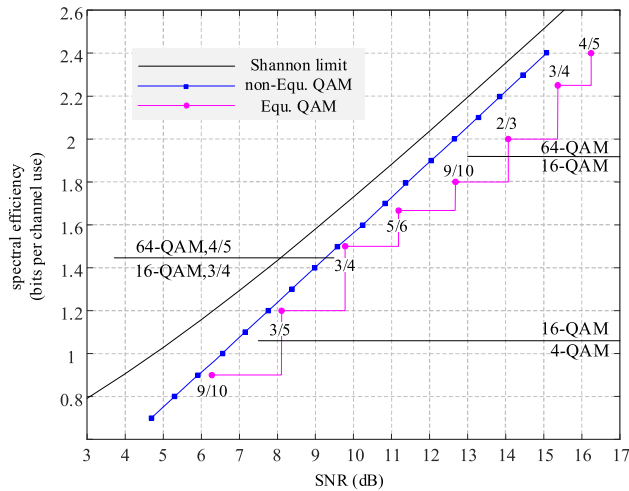


FIGURE 1. Spectral efficiency comparison between probabilistic constellation shaped QAM, conventional equiprobable constellation shaped QAM and the Shannon limit.

length (coding gap), another is due to the equiprobable constellation shape (shaping gap) [5]. To provide a finer granularity and narrow the coding and shaping gap, it needs to increase the number of code rates and modulation formats. Rate-compatible codes are suggested in [8] and [9], however, the system complexity also increases to support a large number of modes. In this work, another approach is adopted, and proposed scheme works with a coarse CM granularity where only one code rate is assigned to a modulation format, e.g., 16-QAM only works with a rate 3/4 code. Compared with the non-shaped scheme, it provides a dense ladder-like shaping gain.

Faster-Than-Nyquist (FTN) is another solution to improve the SE which has attracted great attentions recent years [10]–[13]. FTN was introduced by Mazao in 1970s [14], where the symbols are transmitted in a rate faster than the Nyquist which increases the capacity and introduces the intentional inter-symbol interference (ISI) at the same time. In [11], authors compare the frequency-domain equalization (FDE) and time-domain equalization (TDE) for FTN signaling detection. Simulation results show that TDE outperforms FDE in terms of bit error rate, especially for a low packing factor, while FDE is attractive for its lower computational complexity. In [12], the channel estimation and data detection in doubly-selective channel are jointly designed through FD message-passing. The algorithm is then refined by mitigating the impact of colored noise and non-Gaussian symbol to improve the performance. PCS and FTN signal improve the spectral efficiency from different aspects. FTN signal increases the rate by narrowing the symbol interval while PCS signal through high-order modulation. FTN gets higher information rate than Nyquist case mainly due to the benefit of using the excess pulse bandwidth, however, pulse shaping is beyond our scope. Because the average information per symbol of PCS is less the equiprobable scheme, PCS can be viewed as “Beneath-the-Nyquist” in a sense.

Communication systems suffer from ISI provoked by the dispersive nature of wide-band channel. For a complex communication system, a reliable receiver is needed to maintain robust transmission in ISI channel. Soft or probabilistic channel equalization is a technique to mitigate the ISI [15]. Inspired by turbo code, iterative processing schemes are extended to joint equalizing and decoding via soft-input soft-output messages where a priori information provided by the channel encoder is used to reduce detection errors. The optimal detector is based on the maximum a posteriori (MAP) criterion which can operate near the channel capacity with properly designed coding and detecting scheme [16]. Assuming a perfect knowledge on channel impulse response (CIR), BCJR algorithm [17] works on a trellis network to compute the posteriori probability. But for a channel with long memory or a large signal alphabet, the operational complexity becomes intractable. This motivates the invention of low-complexity detection algorithms, e.g., linear mini-mum mean square error (LMMSE) detection algorithm, in which the discrete-value symbols are directly mapped to exponential variables [18].

Recent years, expectation propagation (EP) has been proposed as a general framework for approximate variational inference through moment matching [19]. Message passing algorithm is a technique solving the MAP problem in the cascade system. Firstly, the jointly distribution of data bits, symbols, channel taps and noise samples are factorized into local functions. After formulating the system factor graph, the factor nodes exchange messages with their neighbors obeying the framework of EP-based message passing rule. Specifically, the codeword log-likelihood ratio (LLR) are exchanged between the channel decoder and demapper, while the extrinsic messages are exchanged between demapper and channel equalizer [13], [20], [21]. EP-based iterative receivers have shown their advantages over conventional turbo-LMMSE schemes in channel decoding, user detection, channel estimation and interference cancellation in terms of error performance, achievable rate and convergence speed [22]–[25].

The main contributions of this paper are summarized in the following:

- A trellis-based approach for PCS system is developed, applying both forward/backward algorithm and turbo equalization to perform detecting and decoding in an iterative way.
- A filter-type approach is proposed. Based on the system factor graph, we show how the EP could operate in the soft decision feedback equalization (DFE) in multipath channel.
- To reduce the matrix inversion cost, a computation reduced scheme capitalizing on block-wise matrix factorization is proposed for Bayesian equalization.

In our work, PCS reaps the large shaping gain and achieves a remarkable degree of flexibility with respect to the transmission rate. Both trellis-based and filter-type algorithms are

performed in PCS signaling to attain reliable receivers in ISI channel. To the best of our knowledge, no other implementation of iterative receiver for PCS in ISI channel has been reported in literatures so far.

The paper is organized as follows. We first describe the system model to show how the PCS transmitter works in Section II. Inspired by the classical BCJR algorithm, section III is devoted to develop a new turbo receiver where the non-uniform priors are taken into consideration. In Section IV we apply the EP framework to derive the passing messages along the system factor graph. A novel matrix inversion strategy is proposed in Section V to reduce the computational complexity of equalization. Several simulations are included in Section VI to compare the performance between different receivers. Finally, this paper ends with conclusions.

Through the paper, bold lowercase letters are used for vectors, e.g., \mathbf{u} is a $T \times 1$ vector where u_t is its t th entries where $t = 1, \dots, T$. Capital bold letters denote matrices, e.g., an $M \times N$ matrix \mathbf{H} has M rows and N columns. \mathbf{I} is the identity matrix and $\mathbf{0}$ is all zeros matrix. $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ return the expectation and variance value. The probability of x at a is $p(x = a)$ and the probability density function (PDF) is denoted as $p(x)$. $\mathcal{CN}(\mu, \sigma^2)$ represents the circularly-symmetric complex Gaussian distribution of mean μ and variance σ^2 .

II. PRELIMINARIES

A. DISTRIBUTION MATCHING

The CCDM transforms uniformly distributed data bits $\mathbf{u} = [u_0, u_1, \dots, u_{T-1}]$ to amplitudes $\mathbf{a} = [a_0, a_1, \dots, a_{N-1}]$ with the same empirical distribution

$$P(A_i) = n_c(A_i)/N, \quad \text{where } A_i \in \mathcal{A}. \quad (1)$$

The output set is of the equal distance amplitudes $\mathcal{A} = \{1, 3, \dots, 2m - 1\}$ and function $n_c(\cdot)$ counts the occurrences of its input in \mathbf{a} . The DM rate is defined as $R_{DM} = T/N$. At the receiver, it also needs to know the distribution to achieve the best performance [26].

B. PROBABILISTIC CONSTELLATION SHAPING

The distribution of constellations approaching the channel capacity is symmetric around zero. Therefore, the sign bits should be equiprobable distributed, and the labeling of the sign bits C_1 is stochastically independent of the following amplitude bits $C_2 \dots C_m$, where

$$P_C(\mathbf{c}) = \underbrace{P_{C_1}(c_1)}_{\text{codeword}} \underbrace{P_{C_2 \dots C_m}(c_2 \dots c_m)}_{\text{sign}} \underbrace{P_{C_2 \dots C_m}(c_2 \dots c_m)}_{\text{amplitude}}. \quad (2)$$

From [27], parity bits output LDPC encoder preserves approximately uniform distribution even if the input bits are non-uniformly distributed. For PCS, sign bits can be generated by copying parity bits and appending some data bits if needed. Amplitudes are labelled by the binary reflected gray code (BRGC) ASK in real and imaginary dimension, and the transmitted symbol is formed by mapping two real

ASK symbols to one complex QAM symbol. For instance, the PCS-16QAM sign label function $\beta(\cdot)$ and the amplitude label function $\beta(\cdot)$ are defined as

$$\begin{cases} \beta(-1) = 0 \\ \beta(1) = 1 \end{cases} \quad \text{and} \quad \begin{cases} \beta(1) = 0 \\ \beta(3) = 1. \end{cases} \quad (3)$$

Their inverse labelling functions are defined as $\beta^{-1}(\cdot)$ and $\beta^{-1}(\cdot)$, respectively. In Fig. 2 we show the visualization of the probability distribution of PCS-16QAM constellations and the height of each bar indicates the probability of the constellation.

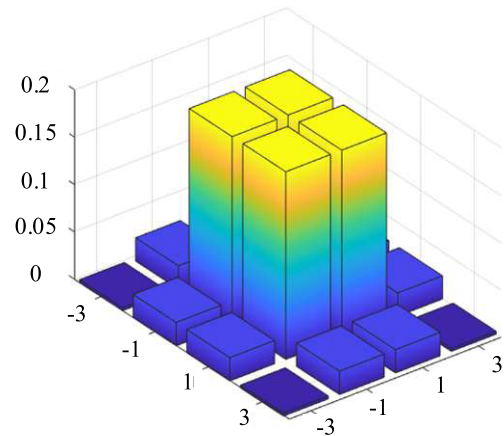


FIGURE 2. Graphical illustration of the probability distribution for PCS-16QAM.

The distribution becomes more “shaped” if the probability of constellations with lower energy increase and thereby decreasing the whole entropy.

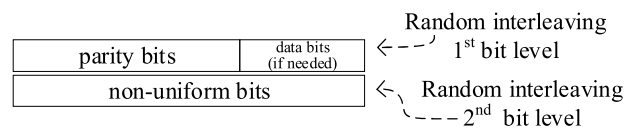


FIGURE 3. Bit-level assignment and interleaving scheme.

C. BIT-LEVEL INTERLEAVING

Since bit levels have different probabilities, interleaving should be implemented by employing several interleavers that independently scramble bits on row basis while the resulting binary frames are mapped to symbols column-wisely. As shown in Fig. 3, the random interleaving scheme operates on two bit levels where 0 and 1 are of different probabilities in each level. Similar scheme can also be found in bit assignment in block-fading channels [28].

D. MULTIPATH CHANNEL

We consider a single user, single carrier and single-input single-output (SISO) transmission scenario. The multipath

channel is modelled as a base-band symbol-space linear filter with L taps:

$$\mathbf{h} = [h_{L-1}, h_{L-2}, \dots, h_0] \quad (4)$$

in which the pulse shaping is also accounted for, and

$$y_k = v_k + w_k = \sum_{l=0}^{L-1} h_l x_{k-l} + w_k \quad (5)$$

where k is the time index and v_k is the noiseless channel output. The signal going through the ISI channel is affected by the thermal noise $w_k \sim \mathcal{CN}(0, \sigma_w^2)$. The transmission can be written in a matrix form:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (6)$$

where \mathbf{H} is a $(K + L - 1) \times (K + 2L - 2)$ matrix whose k^{th} row is $[\mathbf{0}_{k-1}, \mathbf{h}, \mathbf{0}_{K+L-1-k}]$. For the sake of simplicity, the receiver has a perfect knowledge on channel state information. We also assume ideal time and frequency synchronization, and the inter-block inference is ignored.

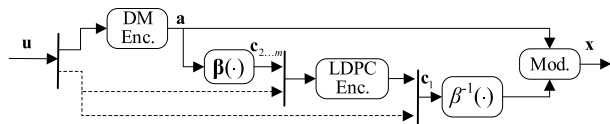


FIGURE 4. Transmitter of the PCS system. Interleaver is omitted for the sake of simplicity.

The diagram of the transmitter of the PCS system is shown in Fig. 4. The dashed lines are needed if there are not enough parity bits, e.g., when the code rate is less than 1/2 for 16QAM or 2/3 for 64QAM. The function Mod. receives amplitudes and signs to form QAM symbols. A significant feature of this structure is that variable data rates are compatible through adjusting the DM rate while the distribution imposed by DM is still preserved at the output.

III. SHAPED-BCJR RECEIVER DESIGN

A. OPTIMAL DETECTION

The optimal receiver computed the estimation \hat{u}_t by minimizing the bit error probability, which is equivalent to

$$\hat{u}_t = \arg \max_{u \in \{0,1\}} p(u_t = u | \mathbf{y}). \quad (7)$$

The posteriori probability $p(u_t = u | \mathbf{y})$ can be obtained by marginalizing over variables in the sequential posteriori probability $p(\mathbf{u} | \mathbf{y})$ as

$$\begin{aligned} p(u_t = u | \mathbf{y}) &= \sum_{\mathbf{u} \in \{0,1\}^K, u_t = u} p(\mathbf{u} | \mathbf{y}) \\ &= \sum_{\mathbf{u} \in \{0,1\}^K, u_t = u} p(\mathbf{y} | \mathbf{u}) p(\mathbf{u}) / p(\mathbf{y}). \end{aligned} \quad (8)$$

Since codewords and binary sources are matched one-by-one, it is very convenient to work with codeword LLR than the bits probability. LLR is defined as

$$\mathcal{L}(c_k | \mathbf{y}) = \ln \frac{p(c_k = 0 | \mathbf{y})}{p(c_k = 1 | \mathbf{y})}. \quad (9)$$

Then the decoding rule can be equivalently written as

$$c_k = \begin{cases} 0, & \mathcal{L}(c_k | \mathbf{y}) \geq 0 \\ 1, & \mathcal{L}(c_k | \mathbf{y}) < 0. \end{cases} \quad (10)$$

Interleaving enables the independence of code-word bits c_k , which yields

$$p(\mathbf{c}) = \prod_{k=0}^{K-1} p(c_k). \quad (11)$$

Then the posteriori probability can be further decomposed into extrinsic LLR and a priori LLR as

$$\begin{aligned} \mathcal{L}(c_k | \mathbf{y}) &= \ln \frac{\sum_{\mathbf{a}:c_k=0} p(\mathbf{y} | \mathbf{c}) \prod_{n=0}^{K-1} p(c_n)}{\sum_{\mathbf{a}:c_k=1} p(\mathbf{y} | \mathbf{c}) \prod_{n=0}^{K-1} p(c_n)} \\ &= \ln \frac{\sum_{\mathbf{a}:c_k=0} p(\mathbf{y} | \mathbf{a}) \prod_{n=0, n \neq k}^{K-1} p(c_n)}{\sum_{\mathbf{a}:c_k=1} p(\mathbf{y} | \mathbf{a}) \prod_{n=0, n \neq k}^{K-1} p(c_n)} + \ln \frac{p(c_k = 0)}{p(c_k = 1)} \\ &= \underbrace{\mathcal{L}_{\text{ext}}(c_k | \mathbf{y})}_{\text{extrinsic LLR}} + \underbrace{\mathcal{L}_a(c_k)}_{\text{a priori LLR}} \end{aligned} \quad (12)$$

where extrinsic LLR $\mathcal{L}_{\text{ext}}(c_k | \mathbf{y})$ represents the information on c_k from received samples \mathbf{y} and c_n for all $n \neq k$, and a priori LLR $\mathcal{L}_a(c_k)$ represents the available a priori information on c_k .

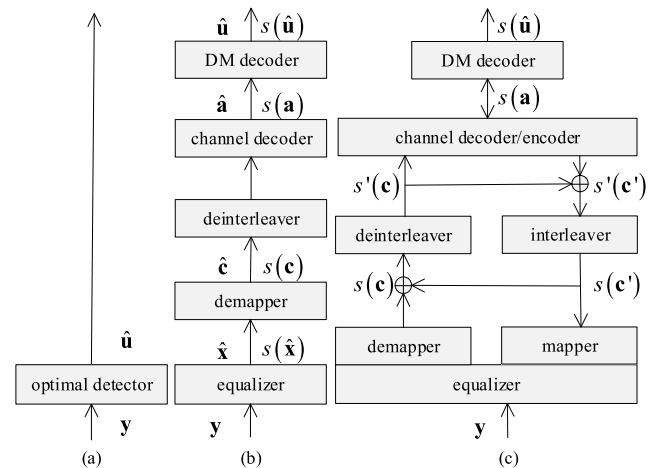


FIGURE 5. Three receiver structures: a) optimal receiver, b) Separated equalization and decoding (left: hard decision, right: soft decision), c) Turbo receiver.

B. SHAPED-BCJR

Optimal detection has an intractable computational complexity of order $\mathcal{O}(2^K)$. A general approach to reduce that cost is separating the detection problem into two subproblems: equalizing and decoding. Three different structures are shown in Fig. 5, in which the components communicate with their neighbors using either hard estimates $\hat{\mathbf{x}}, \hat{\mathbf{c}}, \hat{\mathbf{a}}$ and $\hat{\mathbf{u}}$ from

their alphabets or corresponding soft estimates $s(\hat{\mathbf{x}})$, $s(\mathbf{c})$, $s(\mathbf{a})$ and $s(\hat{\mathbf{u}})$, respectively. When considering the subproblem of equalization, MAP algorithm estimates symbol x_k from its alphabets Ξ as follows

$$\hat{x}_k = \arg \max_{x \in \mathcal{X}} p(x_k = x | \mathbf{y}). \quad (13)$$

This can be efficiently processed using the BCJR algorithm when the ISI channel has a trellis structure with a sufficiently small number of states.

Consider the Proakis-B channel [29] which has an impulse response length of $L + 1 = 3$, and the tapped delay line contains $L = 2$ memories. Corresponding to the possible content of the channel memory, the channel trellis has totally 4^L states in each dimension of 16QAM alphabets. We denote $\mathcal{S} = \{r_1, r_2, \dots, r_{4L}\}$ the set of all the possible states, and the state of the channel is a random variable from the set $s_k \in \mathcal{S}$ at each time instance. Given current state s_k , the next state s_{k+1} only lies in four possible states depending on the input $+3, +1, -1$ or -3 fed into the channel. Therefore, the state evolution can be depicted in a trellis form as shown in Fig. 6. Each path across the trellis corresponds to the sequence of input symbols, and any branch of the trellis can be fully characterized by a four-element tuple $(i, j, x_{i,j}, v_{i,j})$, where index pair (i, j) indicates the valid branch. For the trellis defined in Fig. 6, the set of index pair is

$$\mathcal{I} = \{(0, 0), (0, 4), (0, 8), (0, 12), (1, 0), (1, 4), \dots, (15, 15)\}, \quad (14)$$

where $x_{i,j}$ denotes the channel input when the state transforms from r_i to r_j , and $v_{i,j}$ is the corresponding noise-free channel output. This trellis representation will then be used to compute the posteriori probability $p(x_k | \mathbf{y})$.

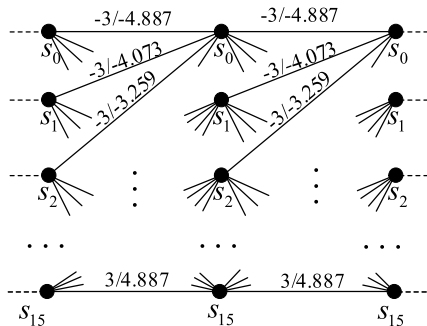


FIGURE 6. Trellis for Proakis-B channel and 16QAM modulation format.

Random variable x_k is assumed to be independently distributed, so $p(\mathbf{x})$ can be fully factorized into $\prod_{k=0}^{K-1} p(x_k)$. Then $p(x_k | \mathbf{y})$ can be computed efficiently using the forward/backward algorithm [25] along the paths contained in the branches. The computation is based on the decomposition of the joint probability $p(s_k, s_{k+1}, \mathbf{y})$ as

$$p(s_k, s_{k+1}, \mathbf{y}) = p(\mathbf{y}) p(s_k, s_{k+1} | \mathbf{y}). \quad (15)$$

Sequence \mathbf{y} can be separated into casual and non-causal samples as

$$p(s_k, s_{k+1}, \mathbf{y}) = p(s_k, s_{k+1}, y_0, \dots, y_{k-1}, y_k, y_{k+1}, \dots, y_{K-1}). \quad (16)$$

Applying the joint probability decomposition chain rule $p(\alpha, \beta) = p(\alpha) \times p(\beta | \alpha)$, $p(s_k, s_{k+1}, \mathbf{y})$ can be further decomposed into

$$\underbrace{p(s_k, y_1, \dots, y_{k-1})}_{\mathcal{F}_k(s_k)} \underbrace{p(s_{k+1}, y_k | s_k)}_{\mathcal{R}_k(s_k, s_{k+1})} \underbrace{p(y_{k+1}, \dots, y_K | s_{k+1})}_{\mathcal{B}_{k+1}(s_{k+1})}. \quad (17)$$

The terms $\mathcal{F}_k(s_k)$ and $\mathcal{B}_{k+1}(s_{k+1})$ can be computed via the forward/backward recursion:

$$\begin{cases} \mathcal{F}_k(s) = \sum_{s' \in \mathcal{S}} \mathcal{F}_{k-1}(s') \mathcal{R}_{k-1}(s', s) \\ \mathcal{B}_k(s) = \sum_{s' \in \mathcal{S}} \mathcal{B}_{k+1}(s') \mathcal{R}_k(s, s') \end{cases} \quad (18)$$

with $\mathcal{F}_0(s) = p(s = s_0)$, which corresponds to the shaped distribution. This initial state is different from the classical BCJR algorithm. $\mathcal{B}_K(s)$ is initialized as 1 for all $s \in \mathcal{S}$. The transition probability $\mathcal{R}_k(r_i, r_j)$ can be further decomposed as

$$\mathcal{R}_k(r_i, r_j) = p(s_{k+1} | s_k) p(y_k | s_{k+1}, s_k) \quad (19)$$

where the value of $p(s_{k+1} | s_k)$ is governed by channel input $x_{i,j}$ which is non-uniformly distributed, and $p(y_k | v_k = v_{i,j})$ depends on the corresponding noiseless channel output $v_{i,j}$. The transition probability $\mathcal{R}_k(r_i, r_j)$ is zero if the its index is not in \mathcal{I} , i.e.,

$$\mathcal{R}_k(r_i, r_j) = \begin{cases} p(x_k = x_{i,j}) p(y_k | v_k = v_{i,j}) & (i, j) \in \mathcal{I} \\ 0 & (i, j) \notin \mathcal{I}. \end{cases} \quad (20)$$

Since we have $y_k = v_k + n_k$, from the noise distribution, $p(y_k | v_k = v_{i,j})$ is given by

$$p(y_k | v_k) = \exp(-|y_k - v_k|^2 / \sigma^2) / \sqrt{\pi \sigma^2}. \quad (21)$$

A distinct difference of shaped-BCJR is that the symbol probability is imposed by the DM rather than uniformly distributed. After above preparations, the posteriori probability $p(x_k = x | \mathbf{y})$ can be obtained by margining the joint probability $p(s_k, s_{k+1} | \mathbf{y})$ over all the possible branches that correspond to the channel input where $x_k = x$ as

$$P(x_k = x | \mathbf{y}) = \sum_{s_k=x} P(s_k = r_i, s_{k+1} = r_j | \mathbf{y}). \quad (22)$$

Then from (15) and (22), we have the conditional LLR of the bit c_k as

$$\mathcal{L}(c_k | \mathbf{y}) = \ln \frac{\sum_{x_{i,j}:c_k=0} \mathcal{F}_k(r_i) \mathcal{R}_k(r_i, r_j) \mathcal{B}_{k+1}(r_j)}{\sum_{x_{i,j}:c_k=1} \mathcal{F}_k(r_i) \mathcal{R}_k(r_i, r_j) \mathcal{B}_{k+1}(r_j)}. \quad (23)$$

Finally, the codeword estimates \hat{c}_k can be recovered from $\mathcal{L}(c_k | \mathbf{y})$.

C. TURBO EQUALIZATION

In the shaped-BCJR based symbol detection algorithm, a priori knowledge on the probability of each symbol along the trellis is initialized by the DM empirical distribution. The forward/backward evolution does not have extra information to calculate Eq. (15) but solely relies on the observations. Commonly, the performance can be improved if a priori information is available. It is natural to feed the information provided by the encoder as the priori information and the decoder can also use the posteriori probability provided by the equalizer to improve performance. This is the main motivation of turbo equalization. However, feeding back $\mathcal{L}(c_k|\mathbf{y})$ contains $\mathcal{L}_a(c_k)$, which is a direct positive feedback. In order to create feedback that is not too strong and avoid too fast convergency, extrinsic information is usually more practical. Therefore, only extrinsic information is fed back both in equalization and decoding in Fig. 5. Note that equalizing and encoding may introduce extra correlation between adjacent entries, interleaving helps to suppress the correlation between neighbors.

The major drawback of this trellis-based algorithm is that its computational complexity grows exponentially with the number of stored trellis states. It will become intractable for high order modulations or multiple channel taps. So, filter-type approaches are more preferred as high order modulation is concerned in PCS system.

IV. EP-BASED DECISION FEEDBACK RECEIVER DESIGN

This section focuses on the design of a filter-type receiver that approximates the posteriori probabilities using EP-based message passing algorithm along the system factor graph.

A. EXPECTATION PROPAGATION FRAMEWORK

EP can be viewed as an extension of loopy belief propagation where variable nodes (VNs) are assumed to lie in the exponential distribution families [31]. The messages exchanged between VNs and factor nodes (FNs) can be characterized by brief distributions. This enables approximating the posteriori PDF $p(\mathbf{c}|\mathbf{y})$ in a fully-factorized iterative way. The basic rule of updating the messages between FN F and VN \mathbf{v} at its i^{th} entry is as follows

$$\begin{cases} m_{\mathbf{v} \rightarrow F}(v_i) = \prod_{G \neq F} m_{G \rightarrow \mathbf{v}}(v_i) \\ m_{F \rightarrow \mathbf{v}}(v_i) = \text{Proj}_{Q_{v_i}} [q_F(v_i)] / m_{\mathbf{v} \rightarrow F}(v_i) \end{cases} \quad (24)$$

where $\text{Proj}_{Q_{v_i}}(\cdot)$ is the well-known *Kullback-Leibler* projection to the target probability distribution Q_{v_i} . The belief $q_F(v_i)$ is an approximation of the marginal of the true posteriori $p(v_i)$, which can be obtained by combining the factors on FN F with messages from neighbor VNs

$$q_F(v_i) = \sum_{\mathbf{v} \setminus v_i} f_F(\mathbf{v}) \prod_{v_i} m_{\mathbf{v} \rightarrow F}(v_i) \quad (25)$$

where $\mathbf{v} \setminus v_i$ denotes the set of VNs without v_i . This projection to exponential families is the *moment matching*, which significantly simplifies the messages calculation [19].

Symbol VNs are assumed to be multivariate circularly symmetric Gaussian distributed with diagonal covariance matrices. Interleaving reduces the correlation between nearby symbols and allows the neglect of non-diagonal entries in the covariance matrix. Therefore, the approximated posteriori distribution can be factorized into independent Gaussians, i.e., the message on x_i can be fully defined by a mean and a variance. Codeword VNs are assumed to be Bernoulli distributed, their messages can be represented by bit-level LLR.

B. FACTOR GRAPH MODEL

The optimal receiver satisfies the MAP criterion by maximizing the posteriori probability, where the posteriori PDF can be factorized as

$$p(\mathbf{u}|\mathbf{y}) = p(\mathbf{u}, \mathbf{c}, \mathbf{x}|\mathbf{y}) \propto \underbrace{p(\mathbf{y}|\mathbf{x})}_{\text{channel}} \underbrace{p(\mathbf{x}|\mathbf{c})}_{\text{mapping}} \underbrace{p(\mathbf{c}|\mathbf{u})}_{\text{coding}}, \quad (26)$$

in which $p(\mathbf{x}|\mathbf{c}) = \prod_{k=0}^{K-1} p(x_k|c_k)$ is the memoryless mapping and $p(\mathbf{c}|\mathbf{u})$ is the overall coding scheme, i.e., from binary sources to codewords. Channel factor in Eq. (26) represents the relationship between the received samples and the transmitted symbols. This posteriori PDF results in the factor graph shown in Fig. 7, and posteriori probabilities will be estimated through messages passing algorithm iteratively.

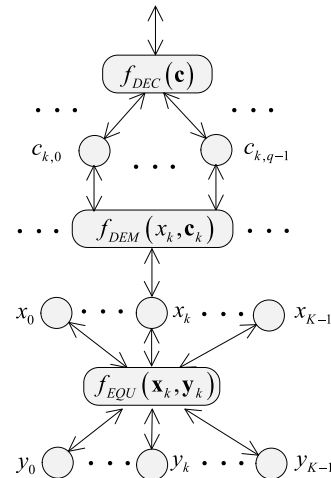


FIGURE 7. Factor graph representation of the posteriori PDF.

C. BAYESIAN EQUALIZATION

From Bayesian's perspective, the channel inputs are modelled as random variables with Gaussian priori PDF $\mathcal{CN}(\mathbf{x}^d, \mathbf{v}^d)$, and the inputs \mathbf{x} and outputs \mathbf{y} are assumed to be jointly Gaussian. Hence, the Bayesian estimation of posteriori PDF $\mathcal{CN}(\boldsymbol{\mu}^e, \boldsymbol{\Gamma}^e)$ has an explicit expression [32]:

$$\begin{cases} \boldsymbol{\mu}^e = \mathbf{x}^d + \mathbf{V}^d \mathbf{H}^T (\mathbf{H}^T \mathbf{V}^d \mathbf{H} + \sigma_w^2 \mathbf{I})^{-1} (\mathbf{y} - \mathbf{H} \mathbf{x}^d) \\ \boldsymbol{\Gamma}^e = \mathbf{V}^d - \mathbf{V}^d \mathbf{H}^T (\mathbf{H}^T \mathbf{V}^d \mathbf{H} + \sigma_w^2 \mathbf{I})^{-1} \mathbf{H} \mathbf{V}^d. \end{cases} \quad (27)$$

Matrix operations in Eq. (27) need a lot of computations, especially the computational complexity of matrix inversion

is usually unaffordable. For a more general implementation, windowed process is adopted by applying a sliding window $[k - p, k + d]$, where k is the window index, p and d are the number of pre-cursor and post-cursor samples, respectively. Then Eq. (27) can be rewritten as

$$\begin{cases} \mu_k^e = x_k^d + v_k^d \mathbf{h}_k^T \left(\mathbf{H}_k^T \mathbf{V}_k^d \mathbf{H}_k + \sigma_w^2 \mathbf{I} \right)^{-1} \left(\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^d \right) \\ \gamma_k^e = v_k^d \left(1 - v_k^d \mathbf{h}_k^T \left(\mathbf{H}_k^T \mathbf{V}_k^d \mathbf{H}_k + \sigma_w^2 \mathbf{I} \right)^{-1} \mathbf{h}_k \right). \end{cases} \quad (28)$$

By selecting p and d larger than the finite length of channel taps, Eq. (28) transforms the matrix inversion into equivalent element-wise operations. Finally, we have

$$q_{EQU}(x_k) \propto \mathcal{CN}(\mu_k^e, \gamma_k^e). \quad (29)$$

D. MESSAGE EXCHANGING RULE

In this section, we detail the messages exchanged along the paths in the considered factor graph. The messages arriving at the VN x_k are independent Gaussians, we have

$$\begin{cases} m_{EQU \rightarrow x}(x_k) \propto \mathcal{CN}(x_k^e, v_k^e) \\ m_{DEM \rightarrow x}(x_k) \propto \mathcal{CN}(x_k^d, v_k^d), \end{cases} \quad (30)$$

and the messages across the VN $c_{k,j}$ are Bernoulli distributed and represented by bit LLR as

$$\begin{cases} m_{DEC \rightarrow c}(c_{k,j}) \propto \mathcal{L}_a(c_{k,j}) \\ m_{DEM \rightarrow c}(c_{k,j}) \propto \mathcal{L}_e(c_{k,j}) \end{cases} \quad (31)$$

where $\mathcal{L}_a(\cdot)$ and $\mathcal{L}_e(\cdot)$ denote a priori and extrinsic LLR, respectively. In the factor graph of Fig. 7, each VN only connects with a pair of distinct FNs and the message output VN is only characterized by its input message, e.g., $m_{DEM \rightarrow c}(c_{k,j}) = m_{c \rightarrow DEC}(c_{k,j})$.

1) MESSAGES FROM DEC TO DEM

DEC node generates a priori information $\mathcal{L}_a(\mathbf{c})$ and feed them to DEM nodes. The DEM node uses the priori LLR to compute the priori probability on $x_k = \alpha$ as

$$p_k(\alpha) \propto \prod_{j=0}^{q-1} e^{-\varphi^{-1}(\alpha) \mathcal{L}_a(c_{k,j})}, \quad \forall \alpha \in \mathcal{X} \quad (32)$$

where $\varphi^{-1}(\cdot)$ denotes the constellation mapping. Above is the Probability Mass Function (PMF) corresponding to the message

$$m_{DEM \rightarrow x}(x_k) = f_{DEM}(x_k, \mathbf{c}_k) m_{c \rightarrow DEM}(\mathbf{c}_k), \quad (33)$$

which will be used to compute $q_{DEM}(x_k)$.

2) MESSAGES FROM EQU TO DEM

Based on the approximate posteriori probability of x_k at EQU, the message output the equalizer can be derived using the message passing rule of Eq. (24) as

$$\begin{aligned} m_{EQU \rightarrow x}(x_k) &= q_{EQU}(x_k) / m_{x \rightarrow EQU}(x_k) \\ &= q_{EQU}(x_k) / m_{DEM \rightarrow x}(x_k). \end{aligned} \quad (34)$$

Accordingly, the message can be extracted via the Gaussian density division

$$x_k^e = \frac{\mu_k^e v_k^d - x_k^d \gamma_k^e}{v_k^d - \gamma_k^e}, \quad v_k^e = \frac{v_k^d \gamma_k^e}{v_k^d - \gamma_k^e}. \quad (35)$$

Gaussian division may lead to a negative variance. This unstable value can be directly replaced by its modulus or parameterized by a damping or mixing factor [33].

3) MESSAGES FROM DEM TO EQU

The demapper calculates its approximate posteriori on x_k by combining messages from connected VNs

$$\begin{aligned} q_{DEM}(x_k) &= m_{DEM \rightarrow x}(x_k) m_{x \rightarrow DEM}(x_k) \\ &= \sum_{\mathbf{d}_k} f_{DEM}(x_k, \mathbf{c}_k) m_{c \rightarrow DEC}(\mathbf{c}_k) m_{EQU \rightarrow x}(x_k) \end{aligned} \quad (36)$$

which is also the posteriori PMF on the elements x_k over the constellation set \mathcal{X} . Using Eq. (30) and (32), it can be denoted as

$$\mathcal{D}_k(\alpha) \propto \exp\left(-|\alpha - x_k^e|^2 / v_k^e\right) p_k(\alpha), \quad \forall \alpha \in \mathcal{X}. \quad (37)$$

We make a small modification on computing $\mathcal{D}(\alpha)$ by replacing simply product (SP) with geometric mean (GM) over its inputs

$$\mathcal{D}_k(\alpha) \propto \sqrt{\exp\left(-|\alpha - x_k^e|^2 / v_k^e\right) p_k(\alpha)}, \quad \forall \alpha \in \mathcal{X}. \quad (38)$$

We term it DFE-GM and Eq. (37) DFE-SP. A major difference is that geometric mean ‘‘normalizes’’ the inputs, i.e., no range dominates the weighting and a given percentage change in each of its inputs has equal effect on the geometric mean product. The simulation results in Sec. VI show its advantages. Through moment matching, it is projected as

$$\begin{aligned} \mu_k^d &= \mathbb{E}[\mathcal{D}_k(\alpha)] = \sum_{\alpha \in \mathcal{X}} \alpha \mathcal{D}_k(\alpha), \\ \gamma_k^d &= \mathbb{V}[\mathcal{D}_k(\alpha)] = \sum_{\alpha \in \mathcal{X}} |\alpha|^2 \mathcal{D}_k(\alpha) - \left| \mu_k^d \right|^2. \end{aligned} \quad (39)$$

Based on the approximate posteriori on x_k from Bayesian equalization, the message to the equalizer $m_{x \rightarrow EQU}(x_k)$ is calculated according to Eq. (24) as

$$x_k^d = \frac{\mu_k^d v_k^e - x_k^e \gamma_k^d}{v_k^e - \gamma_k^d}, \quad v_k^d = \frac{v_k^e \gamma_k^d}{v_k^e - \gamma_k^d}. \quad (40)$$

This Gaussian density division corresponds to the extrinsic information to the Bayesian equalizer which ‘‘partially’’ removes the message $m_{x \rightarrow DEM}(x_k)$. However, it is completely removed in BP.

4) MESSAGES FROM DEM TO DEC

Message from DEM to DEC is calculated using the approximate posteriori on the VN $d_{k,j}$ as bit LLR

$$\mathcal{L}_e(d_{k,j}) = \ln \sum_{\alpha \in \mathcal{X}_j^0} \mathcal{D}_k(\alpha) - \ln \sum_{\alpha \in \mathcal{X}_j^1} \mathcal{D}_k(\alpha) - \mathcal{L}_a(d_{k,j}). \quad (41)$$

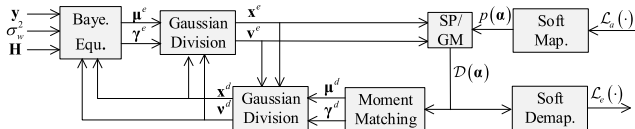


FIGURE 8. EP based receiver structure.

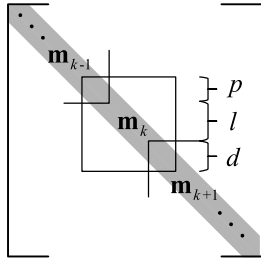


FIGURE 9. Illustration of banded matrix \mathbf{M} and its submatrix \mathbf{m}_k .

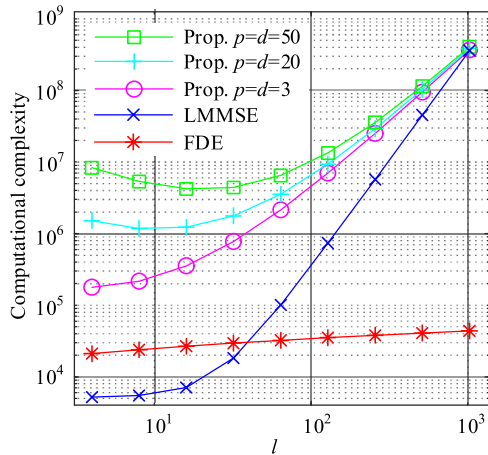


FIGURE 10. Complexity comparison of FDE and TDE.

This message is obtained by marginalizing on k^{th} symbol's j^{th} bit, and \mathcal{X}_j^p is the set of symbols whose j^{th} bit is p . Bit-level LLR calculation is followed by a bit-level assignment before fed to the FEC decoder.

The receiver structure is shown in Fig. 8, interleaving and bit assignment are omitted for the representation simplicity. Unlike turbo iteration where mild extrinsic information is always feedback, this receiver feeds approximate a priori information to the equalizer to calculate its posteriori probabilities because most of the symbol posteriori information has already been removed in Eq. (40).

V. BLOCK-WISE MATRIX INVERSION USING CHOLESKY FACTORIZATION

Bayesian equalization requires excessive computational costs, and symbol-wise matrix inversion contributes to the major part of the computation. Therefore, an efficient calculation method is in need. In this section, we consider a block-type matrix inversion strategy. In detail, we define

$$\mathbf{M} = \left(\mathbf{H}^T \mathbf{V}^d \mathbf{H} + \sigma_w^2 \mathbf{I} \right) \quad (42)$$

which is a banded matrix with equal lower and upper bandwidth. As shown in Fig. 9, we define its submatrix \mathbf{m}_k which

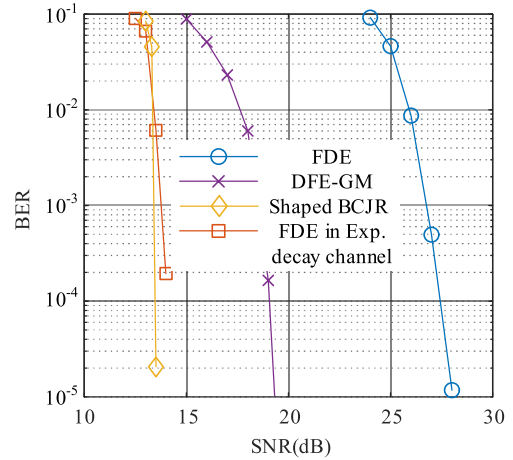


FIGURE 11. BER versus SNR for proposed algorithms and algorithm with FDE.

is formed by taking serial entries, namely $(kl+1)^{\text{th}} \sim (kl+l+p+d)^{\text{th}}$ row and column, from the original matrix. We still use p and d denoting the number of pre-cursor and post-cursor samples. The block size is $(p+l+d)$ where l is a positive variable that affects the algorithm granularity and computational burden. Then the Bayesian equalization can be written as

$$\begin{cases} \mu_k^e = \mathbf{x}_k^d + \mathbf{V}_k^d \mathbf{H}_k^T \mathbf{m}_k^{-1} (\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^d) \\ \gamma_k^e = \mathbf{V}_k^d - \mathbf{V}_k^d \mathbf{H}_k^T \mathbf{m}_k^{-1} \mathbf{H}_k \mathbf{V}_k^d \end{cases} \quad (43)$$

where \mathbf{x}_k^d , \mathbf{V}_k^d , μ_k^e and γ_k^e are the corresponding k^{th} submatrices of \mathbf{x}^d , \mathbf{V}_d , μ^e and γ^e , respectively, and

$$\mathbf{m}_k = \left(\mathbf{H}_k^T \mathbf{V}_k^d \mathbf{H}_k + \sigma_w^2 \mathbf{I}_k \right). \quad (44)$$

Clearly, Eq. (27) and (28) are two special cases of Eq. (44). When l reduces to one, this scheme degrades to typical element-wise filtering, and when l is set to K , the algorithm is equivalent to Eq. (28).

From the definition, we can see that \mathbf{m}_k is symmetric and positively defined and it can be factored as

$$\mathbf{m}_k = \mathbf{R}^T \mathbf{R} \quad (45)$$

where \mathbf{R} is an upper triangular with positive diagonal elements. The matrix inversion can be solved by Cholesky decomposition in a recursive way.

Then we give a brief complexity analysis of proposed algorithm, linear MMSE and FDE [12]. The Cholesky factorization has a computational complexity of $W^3/3+2W^2$ where W is the matrix order. Denoting l , L and N as the block size, channel length and total symbols, totally $(2L+4(l+p+d+1)) \times N + (l+p+d)^2 \times N/3$ operations are required in proposed algorithm during the block-wise equalization. The FDE involves the calculation of N_{FFT} -point FFT/IFFT which has a complexity of $\mathcal{O}(N_{\text{FFT}} \times \log(N_{\text{FFT}}))$. Taking the prior knowledge in calculation, the total operations required in FDE is $4N \times \log(N_{\text{FFT}}) + 15N$. For LMMSE equalization, the complexity is $(N \times l^2)/3 + (L+1) \times N$.

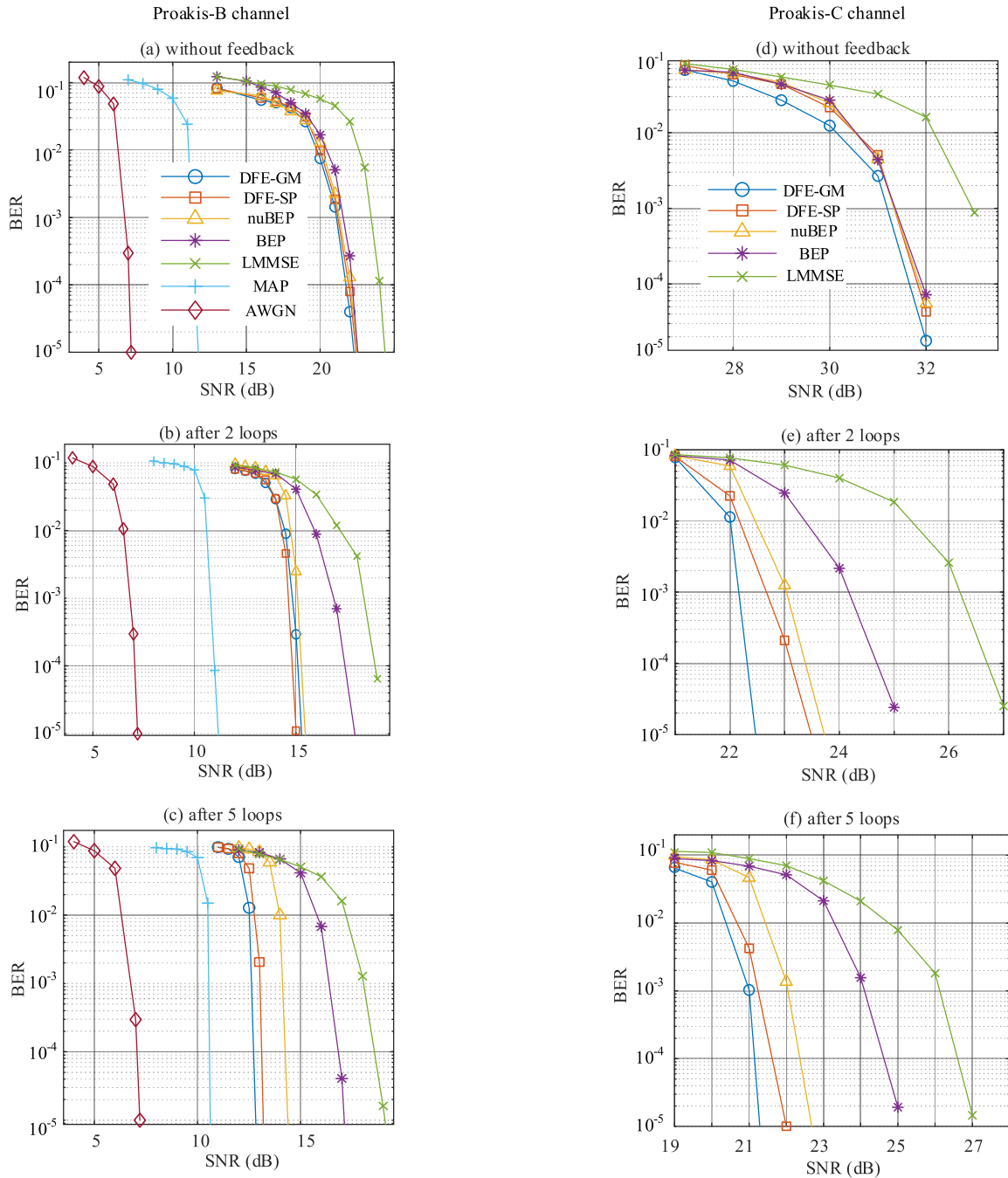


FIGURE 12. BER versus SNR for different algorithms.

In Fig. 10 we show the computational complexity comparison of different block sizes where up to $N = 2^{10}$ symbols are considered. LMMSE has the lowest computational complexity when the block size is small. Once initialized, the LMMSE detector requires no update during the computation, so the initialization becomes the major computation with the growing of l . Benefited from the rapid FFT computation, single-tap FDE saves a lot of computations, especially when a large block size is adopted. As depicted in the figure, proposed block-wise matrix inversion strategy saves up to

70% computational burden than the symbol-wise implement ($l = 1$), and the most computational efficient block size varies with different overlapping samples.

VI. PERFORMANCE EVALUATION

In this section, we show the performance of DFE-SP, DFE-GM, nuBEP, B-EP [34] and LMMSE for different scenarios. Shaped-BCJR and AWGN are also presented in Proakis-B channel as lower bound. From [35], MMSE families show similar performance as LMMSE, so other MMSE approaches

are not included in simulation. We also do not include EP-F because it exhibits almost identical performance with nuBEP for each iteration when 16QAM is used [36]. A detailed comparison of BCJR families is reviewed in [30] and the shaped-BCJR runs in Proakis-B channel with a tolerable computational complexity, but it is not included in Proakis-C channel simulations.

We average the simulation over 10^4 random frames with PCS-16QAM. DVB-S2 LDPC code of rate 3/4 is used as FEC code and belief propagation performs decoding with a maximum of 10 iterations. The pre- and post-cursor samples are both set to 20, and the block size is set to 10. Coefficients are [0.407, 0.815, 0.407] and [0.227, 0.46, 0.688, 0.46, 0.227] for Proakis-B and -C channel, respectively.

BER performance of shaped-BCJR, DFE-GM and FDE in Proakis-B channel are given in Fig. 11 where the DM rate is 0.8. FDE in 64-order exponential-decay channel is also presented as reference while the performance of shaped-BCJR is not included due to the unaffordable quantities of trellis states. It can be observed that DFE-GM has a performance gap of almost 6 dB with respect with the shaped-BCJR at $BER=10^{-4}$, and FDE falls behind the DFE-GM about 8 dB. Simulation results show that TDEs have substantially superior performance to the FDE in Proakis-B channel which contains a strong dominant component.

In Fig. 12 we depict the BER performance versus SNR with two channel configurations where the DM rate is 0.5. At first iteration EP based approaches have similar performance which are about 2 dB better than the LMMSE. After 5 iterations, BEP keeps about 2 dB better than LMMSE. Proposed algorithm with SP has an improvement of 5 dB with respect to the LMMSE and 1 dB compared to the nuBEP algorithm. When the GM is applied, another 0.5 dB further gain can be obtained in both channel configurations, and the gap with shaped-BCJR narrows to about 2 dB in Proakis-B channel.

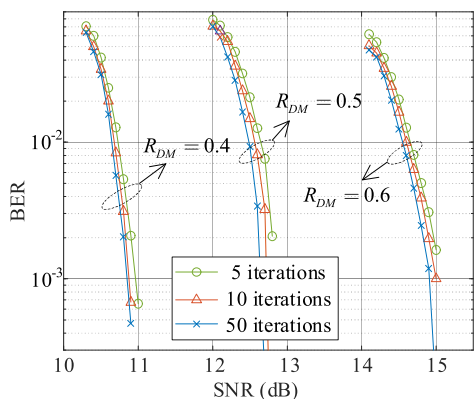


FIGURE 13. BER versus SNR for different DM rates and LDPC iterations.

In Fig. 13, we compare the BER performance of DFE-GM with different DM rates and LDPC iterations for PCS-16QAM in Proakis-B channel. It can be observed that higher SNR enables a larger DM rate for reliable transmission, on the

other hand, transceiver could adjust the DM rate to maximum the system throughput or reliability. The performance can also be improved by increasing the number of LDPC iterations, but the effect is limited. It improves only 0.1dB by increasing the number of LDPC iterations from 10 to 50, which means five times the decoding computational burden.

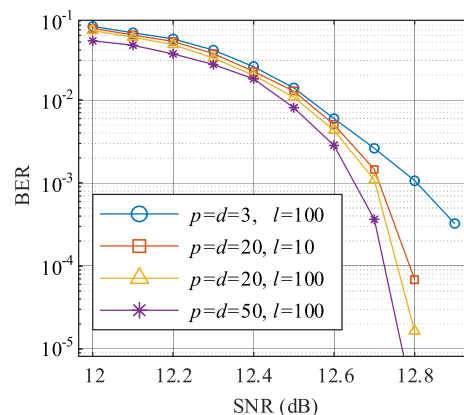


FIGURE 14. BER versus SNR for different block sizes and overlapping samples.

Fig. 14 shows the BER performance of DFE-GM with different block sizes and overlapping samples where the DM rate is 0.5. We observe that block size does not affect the BER performance significantly. From 10 to 100, it only improves 0.01 dB when there are 20 overlapping samples. Note that the behavior of proposed algorithm could be improved by increasing the pre- and post-cursor samples, because a more accurate inverse matrix could be obtained. From 3 to 20, about 0.1 dB gain could be achieved at 10^{-3} BER, and another 0.05 dB could be achieved if the number of overlapping samples extends to 50, however, the computational burden nearly doubled according to Fig. 10.

VII. CONCLUSION

In this paper, we show how PCS with LDPC as FEC could work in ISI channel. PCS is a promising technique to optimize the SE, and enables the system compatible with a wide variety of source rates. In order to obtain a reliable receiver, we design two receivers based on different detection algorithms: MAP detection through shaped-BCJR algorithm and filter-type iterative detection using EP approximation. The major advantage of EP-based algorithms lies in the fact that they perform better than the LMMSE-based algorithms, at the same time, computational complexity does not grow exponentially with the channel states as opposed to most MAP-based algorithms.

First, the forward/backward algorithm together with turbo equalization is used to derive the posteriori probability. The algorithm is adapted by exploiting the shaped probability of each symbol, resulting in the shaped-BCJR algorithm. Second, we detail the messages exchanged along connected nodes in the factor graph. Proposed solution iteratively finds the best Gaussian distribution that approximates the true

posteriori of target symbols. This solution presents quite an improved performance compared to previous approaches after several iterations. Finally, we present a novel matrix inversion strategy to cut down on computational complexity with the help of block-wise decomposition.

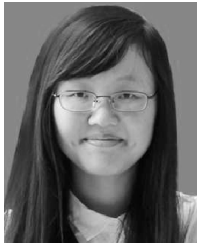
From numerical simulations, we demonstrate that the proposed techniques achieve better performance than LMMSE or conventional EP algorithms. Proposed matrix inversion scheme saves more than half of the computational burden than the element-wise equalization. In this paper, we deal with the detection in SISO channel, applications in MIMO channel and multi-user detection remain to be explored.

REFERENCES

- [1] R. Dar, M. Feder, A. Mecozzi, and M. Shtaf, "On shaping gain in the nonlinear fiber-optic channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 2794–2798.
- [2] G. Ungerboeck, "Huffman shaping," in *Codes, Graphs, and Systems*, R. Blahut and R. Koetter, Eds. New York, NY, USA: Springer, 2002, ch. 17, pp. 299–313.
- [3] N. Cai, S.-W. Ho, and R. W. Yeung, "Probabilistic capacity and optimal coding for asynchronous channel," in *Proc. IEEE Inf. Theory Workshop*, Tahoe City, CA, USA, Sep. 2007, pp. 54–59.
- [4] P. Schulte and G. Bocherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.
- [5] G. Bocherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.
- [6] G. Bocherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 230–244, Jan. 15, 2019.
- [7] A. Josten, B. Baeuerle, B. I. Bitachon, G. Stalder, D. Hillerkuss, and J. Leuthold, "400G probabilistic shaped PDM-64QAM synchronization in the frequency domain," *IEEE Photon. Technol. Lett.*, vol. 31, no. 9, pp. 697–700, May 1, 2019.
- [8] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, Apr. 1988.
- [9] T.-Y. Chen, K. Vakilinia, D. Divsalar, and R. D. Wesel, "Protograph-based raptor-like LDPC codes," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1522–1532, May 2015.
- [10] J. Fan, S. Guo, X. Zhou, Y. Ren, G. Y. Li, and X. Chen, "Faster-than-Nyquist signaling: An overview," *IEEE Access*, vol. 5, pp. 1925–1940, 2017.
- [11] S. Li, W. Yuan, J. Yuan, B. Bai, D. Wing Kwan Ng, and L. Hanzo, "Time-domain vs. frequency-domain equalization for FTN signaling," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9174–9179, Aug. 2020.
- [12] Q. Shi, N. Wu, X. Ma, and H. Wang, "Frequency-domain joint channel estimation and decoding for faster-than-Nyquist signaling," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 781–795, Feb. 2018.
- [13] W. Yuan, N. Wu, A. Zhang, X. Huang, Y. Li, and L. Hanzo, "Iterative receiver design for FTN signaling aided sparse code multiple access," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 915–928, Feb. 2020.
- [14] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell Syst. Tech. J.*, vol. 54, pp. 1450–1462, Oct. 1975.
- [15] C. Douillard, M. Jézéquel, C. Berrou, D. Electronique, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo-equalization," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 507–511, Sep. 1995.
- [16] M. V. Eyuboglu and G. D. Forney, "Trellis precoding: Combined coding, precoding and shaping for intersymbol interference channels," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 301–314, Mar. 1992.
- [17] 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.
- [18] 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.
- [19] T. Minka, "A family of algorithms for approximate Bayesian inference," Ph.D. dissertation, MIT, Cambridge, MA, USA, Jan. 2001.
- [20] J. Dai, K. Niu, and J. Lin, "Iterative Gaussian-approximated message passing receiver for MIMO-SCMA system," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 3, pp. 753–765, Jun. 2019.
- [21] N. Wu, W. Yuan, H. Wang, Q. Shi, and J. Kuang, "Frequency-domain iterative message passing receiver for faster-than-Nyquist signaling in doubly selective channels," *IEEE Wireless Commun. Lett.*, vol. 5, no. 6, pp. 584–587, Dec. 2016.
- [22] P. M. Olmos, J. J. Murillo-Fuentes, and F. Perez-Cruz, "Tree-structure expectation propagation for LDPC decoding over the BEC," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3354–3377, Jun. 2013.
- [23] G. Yao, G. Yang, J. Hu, and C. Fei, "High-efficiency expectation propagation detector for high-order massive MIMO systems," *IEEE Access*, vol. 7, pp. 125225–125239, Sep. 2019.
- [24] X. Tan, Y.-L. Ueng, Z. Zhang, X. You, and C. Zhang, "A low-complexity massive MIMO detection based on approximate expectation propagation," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7260–7272, Aug. 2019.
- [25] S. Şahin, A. M. Cipriano, C. Poulliat, and M.-L. Boucheret, "Iterative equalization with decision feedback based on expectation propagation," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4473–4487, Oct. 2018.
- [26] L. Schmalen, "Probabilistic constellation shaping: Challenges and opportunities for forward error correction," in *Proc. Opt. Fiber Commun. Conf.*, Mar. 2018, paper M.3.C.1.
- [27] G. Bocherer and R. Mathar, "Operating LDPC codes with zero shaping gap," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2011, pp. 330–334.
- [28] R. Knopp and P. A. Humblet, "On coding for block fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 189–205, Jan. 2000.
- [29] J. G. Proakis, *Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill, Nov. 2007.
- [30] R. Koetter, A. Singer, and M. Tüchler, "Turbo equalization: An iterative equalization and decoding technique for coded data transmission," *IEEE Sig. Process. Mag.*, vol. 21, no. 1, pp. 67–80, Jan. 2004.
- [31] T. P. Minka, "Divergence measures and message passing," Microsoft Res., Cambridge, U.K., Tech. Rep. MSR-TR-2005-173, Dec. 2005.
- [32] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
- [33] P. Sun, C. Zhang, Z. Wang, C. N. Manchon, 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.
- [34] 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. 2015, pp. 379–383.
- [35] 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.
- [36] I. Santos, J. J. Murillo-Fuentes, E. Arias-de-Reyna, and P. M. Olmos, "Turbo EP-based equalization: A filter-type implementation," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 4259–4270, Sep. 2018.



XIANG LI (Graduate Student Member, IEEE) was born in 1990. He received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. His research interests include wireless transmission, wireless networks, and signal processing and its applications.



communications and cyber-physical systems.

JING WU (Member, IEEE) received the B.S. degree in electronic science and technology and the M.S. degree in communication and information system from the Nanjing University of Posts and Telecommunications, China, in 2013 and 2016, respectively. She is currently pursuing the Ph.D. degree with the Department of Information Science and Engineering, Southeast University, China. Her research interests include statistical signal processing and its applications in wireless



YANG HUANG was born in 1997. He received the B.S. degree in information engineering from Southeast University, Nanjing, China, where he is currently pursuing the M.S. degree with the National Mobile Communications Research Laboratory. His research interests include cloud computing and big data, machine learning, and wireless signal processing.

...



National Mobile Communications Research Laboratory, Southeast University. His research interests include wireless communication theory, cognitive radio, and signal processing.

WEI HENG (Member, IEEE) was born in 1965. He received the B.S. degree in information engineering from the Huazhong University of Science and Technology, in 1987, and the M.S. degree in signal and information processing and the Ph.D. degree in communications and information system from Southeast University, in 1990 and 1997, respectively. From 1997 to 2002, he was a Research Scholar with The Catholic University of America. Since 2002, he has been a Professor with the