Blind Channel Estimation and Data Detection With Unknown Modulation and Coding Scheme

Yu Liu and Fanggang Wang, Senior Member, IEEE

Abstract—Blind signal processing techniques of channel and noise power estimation, modulation and channel coding scheme (MCS) recognition, and data detection have played crucial roles in wireless transmission scenarios, where the receiver cannot obtain a priori information of the incoming signals. Each of these blind signal processing tasks has been extensively studied in the literature. Few works studied the combination of partial tasks jointly. However, to the best of our knowledge, an overall problem that involves all the aforementioned tasks has not been investigated previously. Simply cascading the solution of each individual task is apparently not a suitable approach for the overall problem. To address the above issues, we design a joint blind receiver that jointly improves the performance of parameter estimation, MCS recognition, and data detection. Thus, we refer to it as the blind estimation, recognition, and detection (BERD) receiver. The multipath fading channel, noise power, and MCS of the transmitter are all unknown to the receiver side; the required side information is only the pool of MCS candidates and perfect synchronization. In this BERD receiver, we first propose an expectation-maximization-based channel and noise power estimator block, which solves the non-convex maximum likelihood estimation problem in a tractable way. Then, a soft-information detector and regenerator block is designed to detect the transmitted information bits and regenerate more reliable symbols. The BERD receiver iterates between these two blocks, and consequently, the accuracy of channel estimation is improved and in turn helps the data detection. Finally, a multistage likelihood-based fusion and decision block is proposed to make the final decision on the adopted MCS, the information bits, and the unknown channel information. Numerical results are provided to show the superior performance of the proposed BERD receiver compared to the existing schemes, in terms of data detection, MCS recognition, and channel estimation.

Index Terms—Blind channel estimation, blind data detection, channel encoder recognition, modulation recognition, likelihood-based fusion.

I. INTRODUCTION

The increasing requirements for immensely higher data rate, reliability, and quality of service have brought new technical challenges to wireless communication systems. Moreover, the scarce spectrum resources against the explosive data traffic have become an urgent problem that needs to be solved [1]. Consequently, various standardization organizations have proposed flexible dynamic spectrum access and sharing technologies. With a priori information of the spectrum occupation, they can improve spectrum efficiency and guarantee reliable data detection [2]. The pilot-assisted transmission method is widely adopted to convey a priori information of the spectrum occupation [3], [4]. However, it is inapplicable in two kinds of scenarios, as follows. The first kind of application scenario is non-cooperative communication, where the transceiver does not share the exact pilot design protocol [5]. For instance, in non-cooperative military electronic reconnaissance and defense scenarios, the enemy transmits hostile signals to interfere with other receivers (e.g., warships and early warning aircraft), while such hostile signals must be identified and processed to ensure system security. In such a case, blind signal processing techniques are required to identify and process the hostile signals since the receiver is incapable of getting a priori information of the hostile signal from its pilots. The second kind of application scenario is cooperative communication, where the desired signals are jammed by heterogeneous interference from other operators or even some malicious emitters [6], [7]. For instance, the communication signals from the long term evolution (LTE) and Wi-Fi systems coexist in the 2.4 GHz frequency band and interfere with each other [6]. Such an interference cancellation problem is cumbersome since the heterogeneous systems do not share their pilot transmission protocols, and a priori information of the heterogeneous interference is limited. As discussed in [6], the heterogeneous Wi-Fi interference can be first recognized and detected by using the blind signal processing techniques to determine its unknowns. Further, it is regenerated based on the determined unknowns and canceled from the received signal, and the desired LTE signal can be detected after canceling the interference. In the aforementioned two application scenarios, the unknown prior information includes the channel state information (CSI), noise power, the exact modulation and channel coding scheme (MCS) of the incoming signal (e.g., hostile signals and heterogeneous interference). In light of this, the techniques of blind channel and noise power estimation, MCS recognition, and blind...
data detection provide efficient solutions to determine these unknowns and detect the transmitted information bits [5], [6], [7]. A joint blind receiver is required to simultaneously implement the above techniques. Note that the MCS candidates set, which contains all the possible MCS that the incoming signal may adopt, is typically assumed to be known at the blind receiver, facilitating the blind signal processing. This assumption is reasonable and widely considered since the system type of the incoming signal (e.g., LTE or Wi-Fi system) can be determined at the receiver with the aid of the system recognition [8], and then, its corresponding MCS candidates set (e.g., $M_{\text{LTE}} = \{\text{BPSK, 16-QAM, 64-QAM}\}$ or $M_{\text{Ni-Fi}} = \{\text{BPSK, QPSK, 16-QAM, 64-QAM}\}$) is easily determined.

Each of the above individual blind signal processing techniques has been extensively studied in the literature. Blind channel estimation techniques can be classified into maximum likelihood-based (LB) and moment-based (MB) [8], [9], [10], [11], [12]. The LB rules are usually optimal and approach the minimum variance unbiased estimators [8], [9], [10], whereas the MB ones can provide near-optimal performance with proper design [11], [12]. Modulation recognition methods are categorized into two groups, i.e., LB and feature-based (FB) [5]. To be specific, the LB methods have been thoroughly investigated in additive white Gaussian noise (AWGN) and flat-fading channels, as they yield the optimal classifier in the Bayesian sense [13], [14], [15], [16]. However, they are computationally complex and sensitive to unknown channel conditions. On the other hand, the FB methods have lower complexity and can achieve near-optimal performance when the extracted features are highly distinguishable [17], [18], [19], [20], [21], [22], [23]. Channel encoder recognition methods determine the unknown channel encoder from the output bits of demodulation, with focuses on two types of error-correcting codes, i.e., block and convolutional codes [24], [25], [26], [27], [28]. In addition, most of the existing works concentrate on determining the parameters of the channel encoder [24], [25], [26], [27]; few works recognize its type [28]. On the basis of the above literature, we consider a joint blind signal processing problem, which aims to simultaneously determine all the unknown prior information and detect the information bits of the incoming signal. Obviously, the straightforward recipe is to simply cascade the solution of each individual blind signal processing problem [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]. Nevertheless, this recipe accumulates the inaccuracies of the solutions to the each of the problems, and leads to their performance degradation. Therefore, it is apparently not a suitable solution to an overall problem.

In addition to the previous works, joint problems by combining several blind signal processing tasks have been investigated recently [29], [30], [31], [32], [33], [34], [35]. Modulation recognition and data detection problems were simultaneously studied in [29], which improved the symbol detection performance in the ideal AWGN channel. Moreover, blind channel estimation, modulation recognition, and symbol detection problems were jointly considered in the more practical scenarios, i.e., the flat-fading [30] and multipath channel [31]. They adopted a maximum-likelihood (ML) based channel estimator and a Bayesian equalizer to estimate the CSI, determine the modulated symbols, and recognize the unknown modulation formats [30], [31]. Furthermore, the problems of blind channel estimation, noise power estimation, and channel encoder recognition were investigated in AWGN [32] and flat-fading channel [33]. Based on these, we recently proposed a joint channel estimation, encoder recognition, and data detection scheme in [34], which iterates between an expectation-maximization (EM) based channel estimator and a Bayes detector. It simultaneously estimates the channel amplitude and phase, and recognizes the channel encoder in the multipath fading channel. However, the above-cited joint works only focus on two or three blind signal processing issues when significant side information is available at the receiver [29], [30], [31], [32], [33], [34], [35]. Consequently, how to design a joint blind receiver that simultaneously improves the performance of each problem is of interest.

To fulfill the research gap, we propose a joint blind receiver for a multipath wireless transmission scenario, where the receiver has no prior information of the multipath CSI, noise power, and MCS. The proposed blind receiver jointly estimates the multipath channel and noise power without the use of pilots, recognizes the unknown MCS, and detects the data of interest; accordingly, we refer to it as the blind estimation, recognition, and detection (BERD) receiver. It contains three blocks, including the EM-based channel and noise power estimator (EM-CNE) block, soft-information detector and regenerator (SiDR) block, and multistage LB fusion and decision (MLFD) block. In particular, our main contributions are summarized as follows:

- We propose a joint blind receiver approach - the BERD receiver, which simultaneously enhances the performance of channel estimation, noise power estimation, MCS recognition, and data detection under multipath fading channel conditions. The required side information consists of the MCS candidates. In addition, the proposed BERD approach can be applied to the systems with a single receiver and multiple cooperative ones. To the best of our knowledge, the BERD receiver is the first work investigating the overall problem with such limited side information.
- A SiDR block is designed to enhance the data detection and channel estimation accuracy iteratively. The detector in the SiDR block contains a Bayes equalizer, a soft demodulator, and a soft decoder. Compared to the traditional detector, its primary merit is that the bit errors are partially corrected, and then more reliable modulated symbols are regenerated for future channel estimation. As such, by adopting the EM-CNE block, the accuracy of the multipath channel estimation is improved accordingly, which helps the following data detection in turn. This iterative manner between the SiDR and EM-CNE blocks is the reason why the BERD receiver is a joint design, rather than a simple cascade of the channel estimation and data detection, providing an efficient solution to the joint
blind signal processing problem in a multipath fading channel.

- Numerical results show that the BERD receiver achieves considerable enhancement in the recognition, detection, and estimation performance. Especially, in terms of data detection, it significantly outperforms the cascaded schemes. Moreover, the reduced versions of the original BERD receiver are investigated, showing their performance superiority versus the benchmark schemes.

The remainder of this paper is organized as follows. Section II presents the system model. The proposed BERD receiver is introduced in Section III. The BERD approach for the system with multiple receivers is investigated in Section IV. Numerical results are shown in Section V, and conclusions are drawn in Section VI.

Notation: Throughout this paper, variables, vectors, and matrices are written as italic x, bold italic x, and bold capital italic letters X, respectively. A random variable and its realization are respectively denoted by x and x; a random vector and its realization are respectively denoted by x and x; a random matrix and its realization are respectively denoted by X and X; |x| is the cardinality of set X; p(x) denotes the probability density function (PDF) of continues random variable x, and p(x|y) denotes the conditional PDF of x conditioned on continuous random variable y; E{·} denotes the expectation with respect to (w.r.t.) the random variables in the argument; |c| and ℑ(·) represent the real and imaginary parts of the complex number c, respectively; CN(μ, σ^2) denotes the mean μ and variance σ^2; GF(q) denotes the Galois field of a prime power q. The operators [·]^T, [·]*, (·)H denote transpose, conjugate, and Hermitian of their arguments, respectively; the operator ‖·‖ denotes the ℓ2-norm of the argument; the operator ⊗ is the Kronecker product; the L-by-L identity matrix and L-by-1 identity vector are denoted by I_L and 1_L, respectively; log c and ln c denote the logarithm of a real number c to the base 2 and e, respectively; the imaginary unit is denoted by i = √−1; Z, Z^0+, Z_2, and Z_2^n represent all the integers, all the non-negative integers, the set {0, 1}, and the set with n elements which take the value from Z_2, respectively; T_a^n = {a, a + 1, . . . , b} is the shorthand of the index set from the integer a to b, and a < b.

II. SYSTEM MODEL

Consider a wireless transmission in which the receiver has no prior information of the adopted MCS of the transmitter, multipath CSI, and noise power. This consideration is the same as the first non-cooperative application scenario discussed in Section I. The available side information at the receiver includes the modulation candidate set M and channel coding candidate set C of the transmitter, as well as perfect synchronization. Denote the unknown MCS by θ = {η, ζ} ∈ M × C, where η ∈ M and ζ ∈ C are the modulation format and channel coding of the transmitted signal, respectively. Then, the received signal can be expressed as

\[ r_j = \sum_{\ell \in Z_0^{a-b}} a_\ell e^{j\varphi_\ell} s_{j-\ell} + v_j, \quad j \in T_N^a, \]

where L is the number of the wireless channel and N is the number of the received symbols; a_\ell ≥ 0 and \varphi_\ell ∈ [0, 2\pi) are the unknown channel amplitude and the phase of the lth path, respectively; s_j is the modulated symbol from the unknown constellation S^η. Here, S^η denotes the set of all constellation points corresponding to the modulation format η and one modulated symbol maps to \log_2 |S^η| coded bits of a codeword c ∈ Z_q^n (e.g., one QPSK symbol maps to 2 coded bits of a codeword). Note that the information bit and codeword are assumed to be binary, and this assumption is also available at the receiver side [24]. Define the uncoded information bit sequence with length k as b ∈ Z_k^n, and assume that a (n, k) low-density parity-check (LDPC) code with code rate R = k/n is adopted in the transmission. Then, the codeword c is obtained by encoding b using the generator matrix G ∈ Z_2^{nk×n}, which is expressed as c = G^T b. The generator matrix G corresponds to a unique parity-check matrix H ∈ Z_2^{(n-k)×n}, with HG^T = 0. The noise v_j, j ∈ T_N^a, follows the independently identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian (CSCG) distribution, i.e., v_j ∼ CN(0, σ^2).

The tasks of the proposed BERD receiver are to jointly estimate the multipath CSI, including its amplitude a_\ell and phase \varphi_\ell, l ∈ Z_0^{a-b}, estimate the noise power σ^2, recognize the unknown modulation η and channel coding ζ from a candidates set M × C, and finally detect the transmitted information bits b.

III. THE PROPOSED BERD RECEIVER

In this section, the proposed BERD receiver is introduced, along with each of its blocks.

A. The BERD Approach

The block diagram of the BERD receiver is shown in Fig. 1. It iterates between an EM-CNE block and a SiDR block under each MCS hypothesis. The EM-CNE block is deployed to estimate the unknown parameters (i.e., multipath CSI and noise power), while the SiDR block is designed for the data detection. The proposed SiDR block consists of the Bayes equalizer module, the soft demodulator and decoder module, and the re-encoder and re-modulator module. Furthermore, a MLFD block is devised to make the final decision on the transmitted information bits, adopted MCS, and estimated CSI and noise power.

To start the iteration process between the EM-CNE and SiDR blocks under each MCS hypothesis θ′ ∈ M × C, an initialization stage is required. First, the initial values of the unknown parameters is obtained by using the initialization methods presented in Section III-B. Then, with these initial
parameters, the initial modulated symbols $\hat{s}^{(0)}$ is determined by adopting the proposed SiDR block, according to Section III-C. It is worth mentioning that the initialization of the unknown parameters belongs to the zeroth iteration, while the determination of initial modulated symbols $\hat{s}^{(0)}$ is the start of the first iteration. After the above initialization stage, the essential steps in the iterative process under each MCS hypothesis $\theta' \in \mathcal{M} \times \mathcal{C}$ are summarized as follows: 1) utilizing the received signal $r$ and the re-modulated symbols $\hat{s}^{(0)}$ from previous iteration, EM algorithm is applied to deal with the intractable maximum likelihood estimation (MLE) problem of channel information. Then, the unknown parameters are updated, including both the amplitude $\hat{a}$ and phases $\hat{\phi}$ of the multipath fading channel, and noise power $\hat{\sigma}^2$, which are collectively denoted by $\hat{\beta} = [\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_{L-1}, \hat{\phi}_0, \hat{\phi}_1, \ldots, \hat{\phi}_{L-1}, \hat{\sigma}^2]^T$. The details of the blind channel and noise power estimation in EM-CNE block are provided in Section III-B; 2) given the estimated parameter $\hat{\beta}$, the received signal $r$ is equalized by using the Bayes equalizer to obtain the posterior probability of modulated symbols $\rho^{(l)}$ under current MCS candidate $\theta'$; 3) to further suppress the noise and inter-symbol interference induced by the multipath channel, the soft-output symbols $\hat{\theta}'$ are demodulated and decoded to obtain the possible transmitted information bits $\hat{b}'$; 4) the output bits of the decoder $\hat{b}'$ are re-encoded and re-modulated to regenerate the modulated symbols $\hat{s}^{(l)}$. The details of steps 2)-4) in SiDR block are presented in Section III-C; 5) a stopping criterion is adopted, i.e., the iterative process stops if the mean square error (MSE) of the estimated channel information in the current iteration and the previous one is below a threshold $\varepsilon$, $\Delta \hat{\beta} = \|\hat{\beta}^{(l)} - \hat{\beta}^{(l-1)}\|_2^2 < \varepsilon$ or the iterations exceed the maximum iterations, $I > I_{\text{max}}$. Ultimately, based on the estimated parameters and the detected information bits under each MCS hypothesis $\theta' \in \mathcal{M} \times \mathcal{C}$, the MLFD block makes the final decision on the information bits $\hat{b}$, adopted MCS $\hat{\theta}$, and estimated channel information $\hat{\beta}$. The details of the MLFD block are presented in Section III-D. The overall algorithm is summarized in Algorithm 1.

### B. Blind Channel and Noise Power Estimation

#### 1) EM-CNE Block: Here we propose an algorithm to estimate the unknown parameters in multipath scenario, including the channel gain $a_{\ell}$, the channel phase $\varphi_{\ell}, \ell \in \mathcal{L}_{0}^{L-1}$, and the noise power $\sigma^2$. From Algorithm 1, one can see that the parameters are estimated in each iteration $I$ under each MCS hypothesis $\theta' \in \mathcal{M} \times \mathcal{C}$. Thus, we omit the superscript $\theta'$ and the iteration index $I$ for notational simplicity, and illustrate how to update the multipath channel and noise power in one iteration. To solve this problem, the ML estimator is adopted, which aims to estimate the unknown parameter $\beta = [a_0, a_1, \ldots, a_{L-1}, \varphi_0, \varphi_1, \ldots, \varphi_{L-1}, \sigma^2]^T$ of the likelihood function $p(r|\hat{s};\beta)$. Particularly, the modulated symbols $\hat{s}$ are obtained from the SiDR block, which is presented in Section III-C. Then, the expression of $p(r|\hat{s};\beta)$ is given by

$$p(r|\hat{s};\beta) = \prod_{j \in \mathcal{I}_N} p(r_j|\hat{s}_j;\beta)$$

$$\propto \frac{1}{\sigma^{2N}} \exp \left( -\frac{1}{2} \sum_{j \in \mathcal{I}_N} \left( r_j - \sum_{\ell \in \mathcal{L}_{0}^{L-1}} a_{\ell} e^{i\varphi_{\ell}} \hat{s}_{\ell} - \hat{s}_j \right)^2 \right),$$

where $\hat{s} = [\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_N]^T \in \mathbb{S}^N$ and $\hat{s}_j = [\hat{s}_{j-1}, \hat{s}_{j-2}, \ldots, \hat{s}_{j-L} + \hat{s}_{j-L+1}]^T \in \mathbb{S}^{L-2}$. Consequently, the log-likelihood function $\mathcal{F}(\beta)$ can be expressed as

$$\mathcal{F}(\beta) = \ln p(r|\hat{s};\beta)$$

Note that if $j < \ell$, then $\hat{s}_{j-\ell} = 0$. Considering the memory characteristics of the multipath channel, the received symbol $r_j$ is conditional independent to the other received symbols given $\hat{s}_{j-L+1}^j$ and $\beta$.  

---

**Algorithm 1** The Proposed BERD Receiver

1: \textbf{for} $\theta' \in \mathcal{M} \times \mathcal{C}$ \textbf{do}
2: \hspace{1em} \textbf{Initialization:} $\hat{\beta}^{(0)}$ according to Section III-B;
3: \hspace{1em} \textbf{while} the variation of the estimated channel exceeds $\varepsilon$ and the number of iterations $I$ does not exceed $I_{\text{max}}$ \textbf{do}
4: \hspace{2em} Compute $\rho^{(l)}, \hat{b}^{(l)}$, and $\hat{s}^{(l)}$ in the SiDR block according to Algorithm 3 in Section III-C;
5: \hspace{2em} Update $\hat{\beta}^{(l)}$ in the EM-CNE block according to Algorithm 2 in Section III-B;
6: \hspace{1em} \textbf{end while}
7: \hspace{1em} \textbf{end for}
8: Determine the transmitted information bits $\hat{b}$, adopted MCS $\hat{\theta}$, and estimated channel information $\hat{\beta}$ in the MLFD block according to Algorithm 4 in Section III-D.
\[
\alpha \sum_{j \in \mathcal{I}_N^L} - \frac{1}{\sigma^2} r_j - \sum_{\ell \in \mathcal{I}_0^{L-1}} a_{\ell} e^{j\varphi_{\ell}} \hat{s}_{j-\ell}^2 - 2N \ln \sigma.
\]

Then, the MLE of \( \beta \) is given by
\[
\hat{\beta} = \arg \max_{\beta} \mathcal{F}(\beta).
\]

However, (6) is a non-convex problem that is intractable. In addition, since the unknown parameter \( \beta \) is of high-dimension in the multipath scenario, the computational complexity of finding the MLE of \( \beta \) is extremely high without using an efficient solution. Then, the EM algorithm is adopted to solve this MLE problem. A complete data \( Z \) with known density \( p(Z|\hat{s}; \beta) \) is defined in the EM algorithm to facilitate the MLE problem. Note that \( Z \) cannot be observed directly, whereas it correlates with the received signal \( r \) by a non-invertible transformation \( \mathcal{K}(\cdot) \), i.e., \( r = \mathcal{K}(Z) \). In such a setting, the strategy of the EM algorithm is to iteratively perform between two steps instead of directly maximizing \( \ln p(r|\hat{s}; \beta) \), as follows. In the expectation step (E-step), we obtain the lower-bound of \( \ln p(r|\hat{s}; \beta) \) at the current estimate \( \beta[t] \), which is proved to be the conditional expectation of \( \ln p(Z|\hat{s}; \beta) \) given \( r, \hat{s}, \) and \( \beta[t] \) [36]; then, in the maximization step (M-step), this lower-bound is maximized with respect to the unknown parameter to derive a new estimate \( \beta[t + 1] \). The mathematical expressions of E-step and M-step are respectively formulated as

E-step: \[
J(\beta; \hat{\beta}[t]) = \mathbb{E}_{Z|r, \hat{s}, \hat{\beta}[t]} \left[ \ln p(Z|\hat{s}; \beta) | r, \hat{s}, \hat{\beta}[t] \right],
\]

M-step: \[
\hat{\beta}[t + 1] = \arg \max_{\beta} J(\beta; \hat{\beta}[t]),
\]

where \( t \) is the iteration index; Here, \( Z = [z_1, z_2, \ldots, z_N] \in \mathbb{C}^{L \times N} \) and \( z_j = [z_{0,j}, z_{1,j}, \ldots, z_{L-1,j}]^T \in \mathbb{C}^L \), where \( z_{r,j} \) denotes the \( j \)-th complete data of the \( \ell \)-th path. It is worth noting that the choices of \( Z \) and \( \mathcal{K}(\cdot) \) are not unique, while bad choices of \( Z \) and \( \mathcal{K}(\cdot) \) may lead to the invalidation of the EM-based algorithm. Due to the channel memory in our system model, the choice of the complete data under a simple flat-fading channel is no longer suitable for the multipath channel condition [33]. Considering the multipath channel estimation problem in the proposed BERD receiver, the received signal from the multipath channel is the summation of the signals from \( L \) independent paths. Hence, the complete data is devised as the signal from each independent path, i.e.,

\[ z_{\ell,j} = a_{\ell} e^{j\varphi_{\ell}} \hat{s}_{j-\ell} + \nu_{\ell,j}, \quad j \in \mathcal{I}_1^N, \quad \ell \in \mathcal{I}_0^{L-1}, \]

where \( \nu_{\ell,j} \) is an i.i.d. CSCG noise with the power \( \sigma^2_{\ell} = w_{\ell}\sigma^2 \). Define the noise decomposition factor as \( \mathbf{w} = [w_0, w_1, \ldots, w_{L-1}]^T \) with all the elements satisfying \( \sum_{\ell \in \mathcal{I}_0^{L-1}} w_{\ell} = 1 \); thus, the noise element \( \nu_{j} \) satisfies

3By applying the Gibbs’ inequality, it can be proved that the lower-bound on \( \ln p(r|\hat{s}; \beta) \) is tight in each iteration, and the EM algorithm can converge monotonically [36].

4The impact of the noise decomposition factor \( \mathbf{w} \) on the convergence of the EM-CNE algorithm is presented in Section III-B.2.

Fig. 2. Flowchart of the EM-CNE algorithm.

Let \( \mathbf{z}_{\ell,j} = a_{\ell} e^{j\varphi_{\ell}} \hat{s}_{j-\ell} \). Since the modulated symbols \( \hat{s}_j \), \( j \in \mathcal{I}_1^N \), have been determined by the SiDR block as in Section III-C, \( \mathbf{z}_{\ell,j} = a_{\ell} e^{j\varphi_{\ell}} \hat{s}_{j-\ell} \) is the unknown deterministic signal. Then, \( \ln p(Z|\hat{s}; \beta) \) in (7) can be expressed as [37]

\[
\ln p(Z|\hat{s}; \beta) = C_1 - \sum_{j \in \mathcal{I}_1^N} \sum_{\ell \in \mathcal{I}_0^{L-1}} \frac{1}{\sigma^2_{\ell}} |\mathbf{z}_{\ell,j} - \mathbf{\hat{z}}_{\ell,j}|^2,
\]

where \( C_1 = N \sum_{\ell \in \mathcal{I}_0^{L-1}} \frac{1}{\pi \sigma^2_{\ell}} \) is a value that is independent of the blind channel estimation. Then, given \( r, \hat{s}, \) and \( \hat{\beta}[t] \), the conditional expectation of (11) is written as [38]

\[
J(\beta; \hat{\beta}[t]) = C_2 - \sum_{j \in \mathcal{I}_1^N} \sum_{\ell \in \mathcal{I}_0^{L-1}} \frac{1}{\sigma^2_{\ell}} |\mathbf{\hat{z}}_{\ell,j}[t] - a_{\ell} e^{j\varphi_{\ell}} \hat{s}_{j-\ell}|^2,
\]

where \( C_2 = C_1 - NL \) is another value independent of the blind channel estimation; \( \hat{\mathbf{z}}_{\ell,j}[t] \) is the conditional expectation of the \( j \)-th complete data of the \( \ell \)-th path in iteration \( t \). Accordingly, the E-step in (7) and the M-step in (8) can be respectively simplified as

E-step: Compute the complete data

\[
\hat{\mathbf{z}}_{\ell,j}[t] = \mathbf{\hat{z}}_{\ell,j}[t] + w_{\ell} \left( r_j - \sum_{\ell \in \mathcal{I}_0^{L-1}} \mathbf{\hat{z}}_{\ell,j}[t] \right),
\]

M-step: Estimate the channel information

\[
\hat{\beta}[t + 1] = \arg \min_{\beta} \sum_{j \in \mathcal{I}_1^N} \sum_{\ell \in \mathcal{I}_0^{L-1}} \frac{1}{\sigma^2_{\ell}} |\mathbf{\hat{z}}_{\ell,j}[t] - a_{\ell} e^{j\varphi_{\ell}} \hat{s}_{j-\ell}|^2.
\]

It should be noted that by setting the derivative w.r.t. \( a_{\ell} \) in (14) to zero, the updated channel gain \( \hat{\beta}[t + 1] \) is derived in (15). Since the second derivative of (14) w.r.t. \( a_{\ell} \) is a negative definite matrix, the equation (15) is the optimal estimate of \( a_{\ell} \). Then, substituting (15) into (14), we obtain the updated channel phase \( \hat{\varphi}_{\ell}[t + 1] \) in (16) with some straightforward operations. Afterward, the noise power \( \hat{\sigma}^2[t + 1] \) is simply estimated by computing the expectation of the noise element, as formulated in (17). The updated closed-form expressions of
Algorithm 2 EM-CNE Algorithm

1: **Initialization:** Set $t = 0$, $\hat{\beta}[0]$;
2: while the variation of the estimated channel exceeds $\varepsilon$ or the number of iterations $t$ does not exceed $t_{\text{max}}$ do
3: Compute $\hat{z}_{\ell,j}[t]$, $\ell \in I_0^{L-1}$, $j \in I_N^1$, according to (13);
4: Update $\hat{a}_{\ell}[t + 1]$ and $\hat{\varphi}_\ell[t + 1]$, $\ell \in I_0^{L-1}$, according to (15) and (16), respectively;
5: Update $\hat{\sigma}^2[t + 1]$ according to (17);
6: end while
7: Update $\hat{\beta}$ and feedback the updated $\hat{\beta}$ to the SiDR block.

The channel gain $\hat{a}_{\ell}[t + 1]$, channel phase $\hat{\varphi}_\ell[t + 1]$, and noise power $\hat{\sigma}^2[t + 1]$ in iteration $(t + 1)$-th are given by

$$\hat{a}_{\ell}[t + 1] = \frac{1}{P} \sum_{j \in I_N^1} R\left(\hat{s}_{j-1}^e, \hat{z}_{\ell,j}[t]e^{-j\varphi_{\ell}[t+1]}\right), \quad \ell \in I_0^{L-1},$$

$$\hat{\varphi}_\ell[t + 1] = \tan^{-1}\left(\frac{3\sum_{j \in I_N^1} \hat{s}_{j-1}^e \hat{z}_{\ell,j}[t]}{\sqrt{\sum_{j \in I_N^1} \hat{s}_{j-1}^e \hat{z}_{\ell,j}[t]^2}}\right), \quad \ell \in I_0^{L-1},$$

$$\hat{\sigma}^2[t + 1] = \frac{1}{N} \sum_{j \in I_N^1} \left| r_j - \sum_{\ell \in I_0^{L-1}} \hat{a}_{\ell}[t + 1]e^{j\varphi_{\ell}[t+1]} \hat{s}_{j-1}^e \right|^2,$$

where $P = \sum_{j \in I_N^1} |\hat{s}_j|^2$ is the total power of the transmitted symbols. The EM-based algorithm iterates between the E-step in (13) and the M-step in (15), (16), and (17) until the stopping criterion is satisfied, i.e., $\Delta\hat{\beta} = \|\hat{\beta}[t + 1] - \hat{\beta}[t]\|^2 < \varepsilon$ or $t > t_{\text{max}}$, where $\varepsilon$ is the stopping threshold and $t_{\text{max}}$ is the maximum number of iterations. Based on the above derivation, the proposed EM-CNE block designs a proper complete data and decomposes the original intractable MLE problem in (6) into $L$ sub-problems, so that the unknown parameter $\beta$ can be estimated in a tractable way. When the EM-based algorithm converges, the EM-CNE block outputs the estimated $\hat{\beta}$ and feedbacks it to the SiDR block to update the modulated symbol $\hat{s}$, as presented in Section III-C. The flowchart of the proposed EM-CNE algorithm is shown in Fig. 2, and its pseudo-code is summarized in Algorithm 2.

2) Convergence of the EM-CNE Algorithm: In this section, the convergence of this algorithm is discussed, which is directly related to the performance of the overall BERD receiver. Refer to [39], the EM algorithm has been proved to converge monotonically as the iteration proceeds. Note that a noise decomposition factor $w$ is introduced to define the complete data in the proposed EM-CNE algorithm. In light of this, we particularly clarify and prove the impact of $w$ on the convergence rate/result of the channel estimation here, as presented in Lemma 1.

**Lemma 1:** The choice of the noise decomposition factor $w$ affects the convergence rate of the channel amplitude, channel phase, and noise power, while it does not change their convergence results.

Proof: See Appendix A. \hspace{1cm} □

**Remark 1:** Different from the general intuitions, the choice of $w$ relevant to the complete data $\hat{z}_{\ell,j}$ in the E-step has no impact on the convergence result of the channel information. This crucial discovery guarantees the convergence and effectiveness of the proposed BERD receiver, which means no matter how to choose $w$, the convergence of the proposed scheme is the same.

3) Initialization of the EM-CNE Algorithm: The result achieved by the EM algorithm highly depends on the initialization stage [40]. Thus, we first determine the initial value of the unknown parameter $\beta$. As we consider a $L$-path channel, the dimension of the initial values is $2L + 1$, which renders the classical methods [30], [40], [42] inapplicable as their complexity is exponential w.r.t. $L$.

To facilitate the initialization with lower complexity, we adopt two initialization methods. First, we initialize $\beta$ by its true value with some bias [31], [35]. The initial channel gain, channel phase, and noise power are assumed to be uniformly distributed in the regions $(0, 0_+ + \delta_{a}], [\varphi_\ell - 0_+ + \delta_{\varphi}, \varphi_{\ell} + 0_+ + \delta_{\varphi}]$, and $(0, \sigma_0^2 + \delta_\sigma)$, respectively, where $\delta_{a}$, $\delta_{\varphi}$, and $\delta_\sigma$ are the maximum biases for the unknown parameters, respectively. Note that these bias bounds determine how close the initial points are to the true values, whereas the assumption of uniform distribution guarantees fairness in choosing different initial points. Second, we apply a modified fourth-order moment-based method [19], [31] to initialize the unknown multipath channel and adopt the coarse grid search [30] to initialize the noise power. The fourth-order moment of the received signal is defined as $m_4^r(k_1, k_2, k_3, k_4) = \frac{1}{4} \sum_{j \in I_N^1} r_j^{k_1} r_j^{k_2} r_j^{k_3} r_j^{k_4}$, where $k_i \in \mathbb{Z}^0_+$, $i = 1, 2, \ldots, 4$, and the multipath channel coefficient $\hat{h}_\ell$ of the $\ell$-path is estimated by [19] and [31]

$$\hat{h}_\ell = \frac{m_4^r(k, \kappa, \kappa, \ell)}{m_4^r(k, \kappa, \kappa, \kappa)}, \quad \ell \in I_0^{L-1}.$$ 

Without loss of generality, the leading path with $\kappa = 0$ is assumed to be the dominant path [19], [31]. Furthermore, the coarse grid search algorithm [30] initializes the noise power $\sigma_0^2$ by finding the MLE of the log-likelihood function in (5), with a noise power parameter space $(0, \frac{1}{N} \sum_{j \in I_N^1} |r_j|^2)$ and a search step size $\alpha$. By plugging the initialization of $\hat{h}_\ell$, $\ell \in I_0^{L-1}$, and the noise power of each coarse grid into (5), the initialization of $\sigma_0^2$ is determined by (6). Regarding the above two initialization methods, the first one is considered for analyzing the system performance theoretically, revealing the impact of initialization on the BERD receiver, whereas the second one is deployed to verify the system performance practically.

C. SiDR Block

In this section, we introduce the SiDR block, which is composed of the Bayes equalizer, soft demodulator, soft decoder, soft bit decision module, re-encoder and re-modulator module, as presented in Fig. 3. Particularly, the former four

---

5Typical values of $L$ can be 3 [19], 4 [18], [41], or 6 [31], etc.
modules in the SiDR block are first designed to determine the unknown information bits, which is the ultimate goal of the overall BERD receiver. Then, the re-encoder and re-modulator module in the SiDR block is introduced to obtain the modulated symbols $s$, which are required to estimate the unknown channel information in (15), (16) and (17). In the following, we introduce the detailed detection and regeneration processes.

First, we employ the Bayes equalizer to equalize the multipath channel. Define $\rho_{m,j}$ as the posterior probability of the constellation point $m$ in $\mathcal{S}$ given the $j$-th received symbol and the previous $L - 1$ modulated symbols. Then, $\rho_{m,j}$ is expressed as

$$
\rho_{m,j} = p(s_j = m | r_j, \hat{s}^{-L+1}_j ; \hat{\beta}) = p(r_j | s_j = m, \hat{s}^{-L+1}_j ; \hat{\beta}) \cdot p(s_j = m | \hat{s}^{-L+1}_j ; \hat{\beta}) \sum_{m' \in \mathcal{S}} p(r_j | s_j = m', \hat{s}^{-L+1}_j ; \hat{\beta}) \cdot p(s_j = m' | \hat{s}^{-L+1}_j ; \hat{\beta}).
$$

The equalization of (20) follows from the Bayes rule. Moreover, $p(r_j | s_j = m, \hat{s}^{-L+1}_j ; \hat{\beta})$ is proportional to

$$
p(r_j | s_j = m, \hat{s}^{-L+1}_j ; \hat{\beta}) \propto \frac{1}{\sigma^2} \exp \left( - \frac{1}{\sigma^2} | r_j - f(s_j = m) |^2 \right), \quad \mu_m \in \mathcal{S}, \ j \in \mathcal{I}_1^N,
$$

and

$$
f(s_j = m) = \hat{a}_0 e^{i \hat{\phi}_0} m + \sum_{\ell \in \mathcal{I}_1^{-L+1}} \hat{a}_\ell e^{i \hat{\phi}_\ell} \hat{s}_{j-\ell}.
$$

Note that, to compute $\rho_{m,j}$ in (20), $\hat{s}^{-L+1}_j = [\hat{s}_{j-L+1}, \hat{s}_{j-L+2}, \ldots, \hat{s}_{j-1}]^T$ is needed, which can be determined by

$$
\hat{s}_{j-\ell} = \sum_{m \in \mathcal{S}} \mu_{m} \rho_{m,j-\ell}, \quad j \in \mathcal{I}_1^N, \ \ell \in \mathcal{I}_1^{-L+1}.
$$

Hereafter, we adopt a soft demodulator to recover the symbols $\hat{s} = [\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_N]^T$ by using the output $\rho = [\rho_{1,1}, \rho_{2,1}, \ldots, \rho_{\log|\mathcal{S}|, 1}, \rho_{1,2}, \ldots, \rho_{\log|\mathcal{S}|, N}]^T$ from the Bayes equalizer. In general, a constellation point $m \in \mathcal{S}$ corresponds to $\log|\mathcal{S}|$ coded bits. Define the coded bits as $\mathbf{C} = [c_1, c_2, \ldots, c_N] \in \mathbb{Z}_{2^{\log|\mathcal{S}|}} \times \mathbb{N}$ where $c_j = [c_{j,1}, c_{j,2}, \ldots, c_{j,\log|\mathcal{S}|}]^T \in \mathbb{Z}_{2^{\log|\mathcal{S}|}}$; each $c_j$ maps to a constellation point in $\mathcal{S}$. To describe the soft demodulation explicitly, we define the constellation set $\mathcal{A}_g \subseteq \mathcal{S}, \ g \in \mathcal{T}_1^{\log|\mathcal{S}|}$, which contains all the constellation points with $c_{j,g} = 0, \ g \in \mathcal{T}_1^{\log|\mathcal{S}|}$. Then, the soft demodulator outputs the posterior probability log-likelihood ratio (LLR) $\lambda_{j,g}$ as

$$
\lambda_{j,g} = \ln \frac{p(c_{j,g} = 0 | r_j, \hat{s}^{-L+1}_j ; \hat{\beta})}{p(c_{j,g} = 1 | r_j, \hat{s}^{-L+1}_j ; \hat{\beta})} \sum_{\mu_m \in \mathcal{A}_g} \rho_{m,j}, \quad j \in \mathcal{I}_1^N, \ g \in \mathcal{T}_1^{\log|\mathcal{S}|}.
$$

After the soft demodulation, the output $\lambda_{\mathbf{out}} = [\lambda_{1,1}, \lambda_{2,1}, \ldots, \lambda_{\log|\mathcal{S}|, 1}, \lambda_{1,2}, \ldots, \lambda_{\log|\mathcal{S}|, N}]^T$ serves as the input of the soft decoder. Define the information bits as $\mathbf{b} = [b_1, b_2, \ldots, b_k]^T \in \mathbb{Z}_2^k$. Then, the output posterior probability LLR $\xi = [\xi_1, \xi_2, \ldots, \xi_k]^T$ of the soft decoder is denoted by

$$
\xi_l = \ln \frac{p(b_l = 0 | \Theta', \lambda_{\mathbf{out}})}{p(b_l = 1 | \Theta', \lambda_{\mathbf{out}})}, \quad l \in \mathcal{I}_1^k
$$

where $\Theta'$ is the parity-check relation in the hypothesis $\Theta'$. Let $\lambda_{\mathbf{in}} = [\lambda_{1,1}, \lambda_{2,1}, \ldots, \lambda_{\log|\mathcal{S}|, N}]^T$. The elements in $\lambda_{\mathbf{in}}$ equal to the first $k$ elements in $\lambda_{\mathbf{out}}$. Furthermore, the updated extrinsic message $\Lambda_{\mathbf{in}} = [\Lambda_{1,1}, \ldots, \Lambda_{k,1}]^T$ is given by

$$
\Lambda_{l,j} = \xi_l - \lambda_{l,j}^\mathbf{in}, \quad l \in \mathcal{I}_1^k.
$$

After a hard decision of the soft bits, the information bits $\hat{\mathbf{b}} = [\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_k]^T$ are obtained as

$$
\hat{b}_l = \begin{cases} 
0 \quad \text{if} \quad \lambda_{l,0}^\mathbf{in} > 0 \\
1 \quad \text{otherwise} \end{cases}, \quad l \in \mathcal{I}_1^k.
$$

Finally, the detected information bits $\hat{\mathbf{b}}$ input to the re-encoder and the re-modulator module to regenerate the modulated symbols $\hat{s}$. If the stopping criterion of the overall BERD receiver is not satisfied, the updated $\hat{s}$ is fed to the EM-CNE block to update the estimate of the channel information $\hat{\beta}$ in the next iteration. It is worth mentioning that compared to the traditional detectors, the primary merit of the proposed SiDR block is that the information bit is determined and the bit errors can be partially corrected by the former four modules in the SiDR block, and then, more reliable modulated symbols are regenerated for future channel estimation. The proposed SiDR block is summarized in Algorithm 3.

**Algorithm 3 SiDR Algorithm**

1. Compute $\rho$ according to (20);
2. Compute $\lambda_{\mathbf{out}}$ according to (25);
3. Compute $\xi$ according to (26), update $\lambda_{\mathbf{in}}$ by (27), and compute information bits $\hat{\mathbf{b}}$ by (28);
4. Regenerate $\hat{s}$ in the re-encoder and re-modulator module, and feedback $\hat{s}$ to the EM-CNE block if the stopping criterion of the BERD receiver is not satisfied.

**D. MLFD Block**

We propose a MLFD block to determine the information bits $\mathbf{b}$, the adopted MCS $\theta = [\eta, \zeta]$, and the unknown channel $\Theta^\mathbf{in}$.

---

For LDPC encoder, the $k$ information bits are encoded into $n$ coded bits, and then, are mapped to $N = \frac{n}{\log|\mathcal{S}|}$ modulated symbols. In addition, the outputs $\xi_l, l \in \mathcal{I}_1^k$, of the soft decoder are determined by using the belief propagation (BP) algorithm in [43].
information $\beta$, as shown in Fig. 1(b). The MLFD block aims to extract the recognition features with high distinguishability, so as to improve the accuracy of the final decision. The MCS hypothesis $\theta^\prime = \{\eta^\prime, \zeta^\prime\} \in \mathcal{M} \times \mathcal{C}$ is used as the superscript of the decision metrics in this block. Note that the full name of MLDF refers to multistage LB fusion and decision. The definition of multistage and fusion in the proposed MLFD block are explained as follows. The multistage indicates that the MLFD block determines the adopted MCS, transmitted information bits, and channel information in multiple stages. Particularly, the MLFD block first decides on the modulation format, and then, makes the channel coding decision. After the above two stages, the information bits and the channel information are correspondingly determined. The fusion means that when multiple cooperative receivers are considered in the system model (as presented in Section IV), a fusion center involved in the MLFD block collects all the estimated parameters from the multiple receivers, as well as the detected symbols. Then, the decision metrics can be calculated by this fusion center. In the following, we introduce the decision process.

1) Modulation Decision: The log-likelihood function $\mathcal{F}(\tilde{\eta}^\prime, \zeta^\prime), \eta^\prime \in \mathcal{M}, \zeta^\prime \in \mathcal{C},$ in (5) is used as a decision metric and the ML algorithm is adopted for recognition, i.e.,

$$\tilde{\eta}^\prime = \arg \max_{\eta^\prime \in \mathcal{M}} \mathcal{F}(\tilde{\eta}^\prime, \zeta^\prime), \quad \zeta^\prime \in \mathcal{C}. \quad (29)$$

Hence, using $\tilde{\eta}^\prime$, $\zeta^\prime \in \mathcal{C}$, we make the decision on $\tilde{\eta}$ by majority vote in $\mathcal{M}$. Hereafter, we need to recognize the channel coding scheme in $\mathcal{C}$ given the modulation $\tilde{\eta}$.

2) Channel Coding Decision: Different from the traditional cumulative LLR adopted for channel coding recognition, the channel coding decision metric is designed as the average LLR of the syndrome $a$ posteriori probability (APP), aiming to extract more distinguishable recognition features and enhance the recognition performance. To derive the average LLR, the definition of the syndrome is firstly provided. We consider the channel coding $\zeta^\prime \in \mathcal{C}$ and denote a non-zero vector $\pi_{\tilde{\eta}, \zeta^\prime}$ as the indices of the non-zero entries in the $i$-th row of the parity-check matrix $H_{\tilde{\eta}, \zeta^\prime}$, i.e.,

$$\pi_{\tilde{\eta}, \zeta^\prime} = [\pi_{\tilde{\eta}, \zeta^\prime}(1), \pi_{\tilde{\eta}, \zeta^\prime}(2), \ldots, \pi_{\tilde{\eta}, \zeta^\prime}(N_i)]^T, \quad 1 \leq \pi_{\tilde{\eta}, \zeta^\prime}(1) < \pi_{\tilde{\eta}, \zeta^\prime}(2) < \ldots < \pi_{\tilde{\eta}, \zeta^\prime}(N_i) \leq n,$$

where $N_i$ is the number of the non-zero elements in the $i$-th row of $H_{\tilde{\eta}, \zeta^\prime}$. Then, the syndrome is expressed as

$$\tilde{c}_{\pi_{\tilde{\eta}, \zeta^\prime}}(1) \oplus \tilde{c}_{\pi_{\tilde{\eta}, \zeta^\prime}}(2) \oplus \ldots \oplus \tilde{c}_{\pi_{\tilde{\eta}, \zeta^\prime}}(N_i) = 0, \quad \pi_{\tilde{\eta}, \zeta^\prime} \in \mathcal{I}_i^{n-\pi_{\tilde{\eta}, \zeta^\prime}}. \quad (30)$$

If and only if $\tilde{\eta} = \eta$ and $\zeta^\prime = \zeta$, the relation in (30) holds. Furthermore, based on (30), we define the LLR of the syndrome APP for the $i$-th parity-check bit as $\gamma_{i, \zeta^\prime}$ and specify it by the following theorem, which is used to derive the average LLR metric.

**Theorem 1:** Given the modulation $\eta$ and a $(n, k)$ LDPC code $\zeta$, the LLR of the syndrome APP for the $i$-th parity-check bit is denoted by

$$\gamma_{i, \zeta^\prime} = 2 \tanh^{-1} \prod_{\tau \in \mathcal{I}_i^{n-k}} \tanh \left( \frac{1}{2} \psi_{\pi_{\eta, \zeta^\prime}}(\tau) \right), \quad i \in \mathcal{I}_i^{n-k}, \quad (31)$$

$$\begin{cases} 1: \text{Decide the possible modulation candidates according to (29).} \\ 2: \text{If the modulation candidate having the majority vote in } \mathcal{M} \text{ is unique, i.e., } \eta, \text{ then} \\ 3: \text{Compute } \gamma_{i, \zeta^\prime}, \ i \in \mathcal{I}_i^{n-k}, \zeta^\prime \in \mathcal{C}, \text{ from Theorem 1.} \\ 4: \text{Compute } F_{\tilde{\eta}, \zeta^\prime}, \zeta^\prime \in \mathcal{C}, \text{ from (32).} \\ 5: \text{The channel coding } \zeta \text{ is determined from (33).} \\ 6: \text{Compute } F_{\tilde{\eta}^\prime}, \zeta^\prime \in \mathcal{C}, \text{ using Theorem 1, and determine MCS } \theta^\prime \text{ from (33).} \\ 7: \text{End if} \\ 9: \text{If } \tilde{\eta} = \beta, \text{ then } \beta^\prime \text{ are determined correspondingly.} \end{cases}$$

where $\psi_{\pi_{\eta, \zeta^\prime}}(\tau) \in \mathcal{P}$ is the LLR of the posterior probability of the $\pi_{\eta, \zeta^\prime}(\tau)$-th coded bit in a codeword, and $\psi = [\psi_1, \psi_2, \ldots, \psi_n]^T$ is equal to $\mathcal{L}_{\text{out}}$, which is obtained from (25).

**Proof:** See Appendix B.

From Theorem 1, the average LLR of the syndrome APP $F_{\tilde{\eta}, \zeta^\prime}$ for the channel coding decision is calculated by

$$F_{\tilde{\eta}, \zeta^\prime} = \frac{1}{n-k} \sum_{i \in \mathcal{I}_i^{n-k}} \gamma_{i, \zeta^\prime}, \quad \zeta^\prime \in \mathcal{C}. \quad (32)$$

Then, the decision on the channel coding is made as

$$\tilde{\zeta} = \arg \max_{\zeta^\prime \in \mathcal{C}} F_{\tilde{\eta}, \zeta^\prime}. \quad (33)$$

Occasionally, the modulation decision step cannot determine a modulation format if the modulation candidate having the majority vote is not unique; then, we directly employ the average LLR $F_{\tilde{\eta}, \zeta^\prime}, \zeta^\prime \in \mathcal{M} \times \mathcal{C}$, in (32) as decision metric for both modulation and channel coding recognition. Thus, the final decision on the adopted MCS $\theta$ is made among all the MCS candidates according to (33).

The final decision on information bits is made by $\tilde{\beta}$ correspondingly. Furthermore, the multipath channel and the noise power are decided as $\beta^\prime$. The MLFD block is summarized in Algorithm 4.

**IV. BERD APPROACH FOR MULTIPLE RECEIVERS**

Here, the proposed BERD approach is extended to the system with multiple receivers, which cooperatively enhances the performance of data detection, MCS recognition, and channel estimation.

Assume that the number of the receivers is $K$ and the received signal at $q$-th receiver is $r_{q,j} = [r_{q,1}, r_{q,1}, \ldots, r_{q,N}]^T$; then, the $j$-th received symbol at $q$-th receiver $r_{q,j}$ is given by

$$r_{q,j} = \sum_{\ell \in \mathcal{I}_q^{K-1}} a_{q, \ell} e^{j \varphi_{q, \ell}} s_j - \ell + v_{q,j}, \quad q \in \mathcal{I}_1^K, \quad j \in \mathcal{I}_1^{N}, \quad (34)$$

where $a_{q, \ell} > 0$ and $\varphi_{q, \ell} \in [0, 2\pi]$ are the unknown channel amplitude and phase of the $\ell$-th path at the $q$-th receiver; $v_{q,j}$ is the noise at the $q$-th receiver which follows a CSCG distribution, i.e., $v_{q,j} \sim \mathcal{C}N(0, \sigma_v^2)$. In particular, with the cooperation of multiple receivers, the tasks of
the BERD approach are to detect the information bits b, recognize the MCS $\theta$, and estimate the unknown channel information $\hat{\beta} = [\hat{\beta}_1, \ldots, \hat{\beta}_K]$ where $\hat{\beta}_q = [\hat{a}_{q,0}, \hat{a}_{q,1}, \ldots, \hat{a}_{q,L-1}, \varphi_{q,0}, \varphi_{q,1}, \ldots, \varphi_{q,L-1}, \sigma_q^2]^T$.

The procedure of the BERD for multiple receivers can be summarized as follows. First, the multiple receivers individually estimate the multipath channel and noise power $\hat{\beta}_q$, $k \in \mathbb{I}_k^K$, in the EM-CNE block, which follows Algorithm 2. Then, the SiDR block uses the estimation $\hat{\beta}_q = [\hat{\beta}_1, \ldots, \hat{\beta}_K]$ from the multiple receivers to cooperatively detect the information bits $\hat{b}$ and regenerate the modulated symbols $\hat{s}$. Moreover, the algorithm iterates between the EM-CNE block and the SiDR block until the stopping criterion is satisfied. Finally, by utilizing $\hat{\beta}_q$, $\hat{b}$, and $\hat{s}$ determined under each MCS hypothesis $\theta' \in \mathcal{M} \times \mathcal{C}$, the MLFD block makes the final decision on the information bits, MCS, and channel information.

For the SiDR block in the system with multiple receivers, the information bits $\hat{b}$ and the modulated symbols $\hat{s}$ are determined in a cooperative manner. The likelihood probability $p(r_{j,\cdot}|\hat{s}_j = \mu_m, \hat{s}_{j-L+1;\cdot}; \hat{\beta})$ in (21) is rewritten as

$$p(r_{j,\cdot}|\hat{s}_j = \mu_m, \hat{s}_{j-L+1;\cdot}; \hat{\beta}) \propto \prod_{q \in \mathbb{I}_k^K} \sigma_q^2 \exp \left( \sum_{q \in \mathbb{I}_k^K} -\frac{1}{\sigma_q^2} r_{q,j} - f_q(\hat{s}_j = \mu_m) \right)^2,$$

(35)

with $r_{j,\cdot} = [r_{1,j}, r_{2,j}, \ldots, r_{K,j}]^T \in \mathbb{C}^K$ as the $j$-th received symbols of $K$ receivers. Let $\hat{\beta} = [\hat{\beta}_q, f_q(\hat{s}_j = \mu_m)]$ is derived using (22) at each receiver. Then, the output posterior probability $\rho$ of the Bayes equalizer is determined by plugging (35) into (20). After the Bayes equalizer, the methods to obtain the output of the soft demodulator $\hat{\lambda}$, the output of the soft decoder $\hat{\xi}$, the detected information bits $\hat{b}$, and the regenerated symbols $\hat{s}$ are the same as those in Section III-C.

For the case of multiple receivers, the decision-making idea is similar as for the MLFD block presented in Section III-D, but here just say clearly what it is different. In the modulation decision, the likelihood function $F(\hat{\beta})$ in (29) is computed in each MCS hypothesis $\theta' = \{\eta', \zeta'\} \in \mathcal{M} \times \mathcal{C}$, which is rewritten as

$$F(\hat{\beta}) = \sum_{q \in \mathbb{I}_k^K} \sum_{j \in \mathbb{I}_k^K} -\frac{1}{\sigma_q^2} r_{q,j} - \sum_{\ell \in \mathbb{I}_l} \hat{a}_{q,\ell} e^{i \varphi_{q,\ell}} \hat{s}_{j-\ell} \right|^2$$

$$- 2N \sum_{q \in \mathbb{I}_k^K} \ln \sigma_q. (36)$$

The decision on the channel coding $\hat{\zeta}$, information data bits $\hat{b}^\theta$, and channel information $\hat{\beta}^\theta$ performs as the methods in Section III-D.

V. NUMERICAL RESULTS

In this section, we first quantitatively assess the MCS recognition and data detection performance of the overall BERD receiver, considering recognition error probability $P_r$ and bit error rate (BER) as performance metrics. Then, the MCS recognition and data detection performance for two reduced versions of the original BERD problem are evaluated. Finally, the channel and noise power estimation performance of the BERD receiver is further presented, taking into account MSE as performance metric.

A. Simulation Setup

We consider the modulation candidate set as $\mathcal{M} = \{\text{QPSK}, \text{8-PSK}, \text{16-QAM}\}$ [31], while the LDPC channel coding candidate set $\mathcal{C}$ is defined according to the IEEE 802.11ac standard [32]. Then, three code lengths $n = 648, 1296, \text{and} 1944$ are considered in $\mathcal{C}$, and each code length corresponds to four code rates $R = \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \text{and} \frac{5}{6}$. The numbers of channel paths and receivers are $L = 6$ and $K = 5$, respectively. Without loss of generality, the main coefficient of multipath channel is set to $1$, i.e., $h_{k,0} = a_k e^{i \gamma_{k,0}} = 1$, and the remaining coefficients follow the CSCG distribution with $\epsilon_{k}^2 = 0.1$, $k \in \mathbb{I}_k^K$ [18], [19], [31]. The signal-to-noise ratio (SNR) at the $k$-th receiver is defined as $1 + (L-1)\epsilon_{k}^2$. Two initialization methods mentioned in Section III-B.3 are considered for $\beta$. One uses true value of it plus bias, and the maximum bias set is $\delta_0 = 0.1, \delta_\varphi = \frac{\pi}{180}, \delta_{\varphi_0} = 0.1$ [35], which is defined as err1. The other employs the fourth-order moment-based estimates and coarse grid search with the search step size $\alpha = 0.1$. Note that the first initialization method is devised for analyzing the system performance in theoretical simulations, revealing how superior the proposed BERD receiver is when with a better initialization method, whereas the second one is adopted to verify the system performance practically. In addition, we set $I_{\text{max}} = 50, I_{\text{max}} = 50$, and $\varepsilon = 10^{-3}$, along with $\beta_{\text{BP}} = 30$ as the number of iterations of BP algorithm [43]. All numerical results are obtained by running 10000 Monte-Carlo trails.

B. MCS Recognition and Data Detection of BERD Receiver

Figs. 4 and 5 respectively investigate the performance of the BERD receiver with different channel coding candidate sets $\mathcal{C}$, in terms of both MCS recognition and data detection. To be specific, $\mathcal{C}$ contains the encoders with code rates $(\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{5}{6})$ and the code length is $n = 648$ in Fig. 4; while it contains the encoders with code lengths $(648, 1296, 1944)$ and the code rate is $R = \frac{5}{6}$ in Fig. 5 [32]. Moreover, the number of the received symbols per receiver used in Figs. 4 and 5 is $N = 648$ and $N = 3888$, respectively. For comparison, we also consider two existing benchmark schemes, which solve the overall problem by cascading its solution: 1) the first scheme is designed for the multipath scenario, which combines the iterative approaches in [31] and [34]; 2) the second is designed for the single-path flat-fading scenario, which cascades the schemes in [30] and [33]. Note that in the benchmark scheme, we first adopt the approach in [31] or [30] to jointly detect the modulated symbols and estimate the unknown parameters, and then, classify the modulation format. On the basis of the detected symbols and estimated

7Note that, for the second comparison scheme [30], [33], we only adopt the fourth-order moment-based initialization since $\beta$ is treated as a single-path flat-fading channel during estimation, which is not matched to the dimension of the true value.
channels, the channel coding format is further recognized by directly adopting the encoder classifier in [34] or [33], while the information bits are detected based on the BP detector. Meanwhile, the BERD receiver with the perfect CSI and noise power serves as the MCS recognition performance upper bound in Figs. 4 (a) and 5 (a), while that with the perfect CSI, noise power, and known MCS $\theta$ is treated as the data detection performance upper bound in Figs. 4 (b) and 5 (b). Note that the above assumptions of the perfect CSI, noise power, and known MCS $\theta$ are only adopted to revealing an ideal performance upper bound of the proposed BERD receiver, while it is not employed in real communication scenarios.

In Figs. 4 and 5, we observe that compared to the benchmark schemes [30], [31], [33], and [34], the MCS recognition and data detection performance of the BERD receiver with different initialization methods are simultaneously improved. Especially, the superiority of the BERD receiver in terms of BER is extremely significant in the SNR region with $P_e \leq 10\%$. This is attributed to the iterative manner between the EM-CNE block and the SiDR block, which iteratively enhances the MCS recognition performance and decreases the BER. Moreover, it is obvious that the benchmark scheme [30], [33] designed for the flat-fading channel cannot achieve acceptable performance in the multipath scenario. Compared to the performance upper bounds in Figs. 4 and 5, when applying a good initialization err1 in the BERD receiver, the losses in $P_e$ and BER are respectively reduced to within 0.3 dB at 10% and within 0.5 dB at 10$^{-3}$. The curves with err1 reflect the positive impact of a good initialization on the BERD receiver; thus, even if they provide an optimistic performance, they are still of interest. However, when adopting the fourth-order moment-based initialization, $P_e$ of the BERD receiver has certain gaps compared to the upper bounds, and the BER decreases slowly or has a floor in the high SNR region in Figs. 4 and 5. This is due to the fact that the EM-CNE block cannot accurately estimate the channel information when with poor initialization, and in turn, the SiDR block is unable to correct the bit errors effectively. As expected, the performance of the BERD receiver can be enhanced with better initialization methods, whereas initialization is another substantial research topic and beyond the scope of this paper.

C. MCS Recognition and Data Detection of Reduced BERD Receiver

Although the proposed BERD receiver reveals superiority in solving the overall blind signal processing problem
in Figs. 4 and 5, it is still essential to demonstrate its scalability/effectiveness on addressing the reduced problem composed of partial blind signal processing tasks. As such, the recognition and detection performance for the two reduced versions of the original BERD receiver are presented here. One reduced version is designed for the sub-tasks of modulation recognition, data detection, multipath channel estimation, and noise power estimation with known channel coding $\zeta$, as shown in Fig. 6. The other one is applied for addressing the channel coding recognition, data detection, multipath channel estimation, and noise power estimation problems with known modulation format $\eta$, as shown in Fig. 7. In addition, $\mathcal{C}$ is the same as Fig. 4, $\zeta$ in Fig. 6 is randomly selected from $\mathcal{C}$, and $\eta$ in Fig. 7 is randomly chosen from $\mathcal{M}$. Meanwhile, the reduced BERD receivers in Figs. 6 and 7 are respectively compared to the benchmark schemes [30], [31] and [33], [34], using different initialization methods.

Recognition error probability and BER of the reduced BERD receivers with different initializations also outperform the comparison ones, especially in terms of BER, as presented in Figs. 6 and 7. The considerable improvement on the recognition and detection performance is also attributed to the iterative manner between the EM-CNE block and SiDR block in the reduced BERD receiver. Consequently, this reveals that the iterative manner in the proposed BERD receiver provides an efficient solution to the joint blind signal processing problems, for both the original and the reduced ones.

### D. Channel and Noise Power Estimation of BERD Receiver

MSE performance of the multipath channel and noise power estimation in the BERD receiver is evaluated in Fig. 8. For comparison, we also plot the MSE of the benchmark scheme [31], [34], zero forcing (ZF), and linear minimum mean squared error (LMMSE) pilot-based channel estimation methods, i.e., exploit all the transmitted data bits as pilots. Moreover, $\mathcal{C}$ is the same as Fig. 4.

According to Fig. 8 (a), the CSI estimation performance of the BERD receiver with different initialization methods is better than that of the benchmark scheme [31], [34] in the low SNR region, and is comparable in the high SNR region. However, the BER gain of the BERD receiver in Fig. 4 (b) is consistently significant. This also implies that the proposed SiDR block outperforms the conventional detector, even with similar CSI estimation performance. Furthermore, compared to the ZF and LMMSE methods, the BERD receiver has certain MSE gains in specific situations. The MSE of CSI has a $6\text{dB}$ gain at $10^{-2}$ against the LMMSE method when with good initialization err1, while its loss is within $3\text{dB}$ at $10^{-2}$ compared to the ZF one when with fourth-order moment-based initialization. According to Fig. 8 (b), the MSE of noise power estimation achieved with different initialization methods is nearly the same for both BERD receiver and benchmark scheme [31], [34], which means the noise power estimation is not sensitive to the initialization. Moreover, the MSE performance has a deterioration in the high SNR region in Figs. 8 (a) and (b). This may lie in that the MLE problem of the multipath CSI in formula (6) is dominated by noise when SNR is low, and thus, it contains less local extrema and is not sensitive to the initialization bias. On the contrary, the MLE problem is dominated by the transmitted signal in the high SNR region, along with more local extrema and is more sensitive to the initialization bias [31], [35]. Consequently, even the initialization is not far from the true value of $\beta$, the estimation result $\hat{\beta}$ in the high SNR region is still more likely to be trapped in the local extrema, which leads to the MSE deterioration.

### E. Computational Complexity Analysis

The computational complexity of the proposed Algorithms 1-4 are respectively analyzed here. In Algorithm 2, the computational complexity of the EM-CNE block comes from the complete data calculation in (13), multipath channel estimation in (15) and (16), as well as noise power estimation in (17). Thus, the total complexity of the EM-CNE block is of order $O(t_{\text{em}}KN(12L + L^2))$, with $t_{\text{em}} \leq t_{\text{max}}$ as its number of iterations. Additionally, the complexity of the err1 and fourth-order moment-based initialization methods are of order $O(\frac{4KL}{T} \sum_{I \in T} |r_I|^2)$ and $O(8KLN)$, respectively. In Algorithm 3, the complexity of the SiDR block involves calculating the posterior probability of the modulated symbols in (20), detecting the information bits according to...
to (25)-(28), and regenerating the symbols by utilizing the re-encoder and re-modulator. Under MCS hypothesis $\theta' \in \mathcal{M} \times \mathcal{C}$, the total computational complexity of the SiDR block is of order $O(2t_{bp} \log |S_\theta'|^2 N^2)$, where $|S_\theta'|$ is the cardinality of constellation set $S_\theta'$. In Algorithm 4, the computational complexity of the MLFD block is induced by the final decision on the modulation format and channel coding, respectively given in (29) and (33). The maximum computational complexity of the MLFD block is of order $O(|M| K N^2 + P_{\theta'} \in C \log |\hat{\eta}| N_{\theta'} N)$. Consequently, according to the iterative process in Algorithm 1, the total computational complexity of the proposed BERD receiver is of order $O\left(\sum_{\theta' \in \mathcal{M} \times \mathcal{C}} (2I_{berd} t_{bp} \log |S_\theta'|^2 N^2 + \log |S_\theta'| N_{\theta'} N)\right)$, where $I_{berd} \leq I_{\text{max}}$ is the number of iterations of the overall BERD receiver. As for the BERD receiver with the perfect CSI and noise power, which serves as the MCS recognition performance upper bound in Figs. 4 (a) and 5 (a), its computational complexity is of order $O(2t_{bp} \log |S_\theta'|^2 N^2)$. As can be observed, the computational complexity of the original BERD receiver is higher than the above performance upper bounds, while they are still in the same order.

The computational complexity of the benchmark scheme [31], [34] are accessed for comparison. Since the benchmark scheme [31], [34] is the cascade of the solutions in [31] and [34], its total computational complexity is obtained by adding that of [31] and [34]. The scheme in [31] classifies the modulation format, detects the modulated symbols, and estimates the unknown parameters. Its computational complexity of is of order $O(|M| K I_{\text{n}} N^2)$, where $I_{\text{n}} \leq I_{\text{max}}$ is the number of iterations. The scheme in [34] involves the channel coding recognition and information bits detection. Its computational complexity is of order $O\left(\sum_{\mathcal{C}' \in C} \log |S_\theta'| N_{\theta'} N\right)$. Therefore, the total complexity of the benchmark scheme [31], [34] is of order $O(|M| K I_{\text{n}} N^2 + \sum_{\mathcal{C}' \in C} \log |S_\theta'| N_{\theta'} N)$.

Through straightforward mathematical operations, it can be seen that the proposed BERD receiver has a higher computational complexity than the benchmark scheme. This implies that the simultaneous improvement of data detection, MCS recognition, and parameter estimation is at the cost of...
complexity. However, the computational complexity is still in the same order as the comparison one, yet acceptable.

VI. CONCLUSION

This paper proposes a joint blind receiver approach, referred to as BERD, which can be applied in both the single receiver and the multiple cooperative ones. By iterating between the EM-CNE and SiDR blocks and making the final decision in the MLFD block, the BERD receiver jointly estimates the multipath channel and noise power, recognizes the MCS, and detects the data of interest. Numerical results indicate that the recognition error probability, BER of detection, and MSE of estimation in the BERD receiver are considerably improved compared to the benchmark schemes, especially in terms of BER. Further, the reduced versions of the BERD receiver are investigated. Numerical results show that the reduced BERD receivers also achieve outstanding recognition and detection performance. This implies that the iterative manner in the BERD receiver can be successfully applied to the reduced joint blind signal processing cases. However, the proposed BERD receiver is dedicated to the LDPC codes and multipath channel conditions with perfect synchronization. Therefore, in further work, we intend to explore a joint blind approach by extending the BERD receiver to a more complicated propagation environment, such as imperfect synchronization, frequency selectivity, and Doppler shift. Moreover, other channel encoders and larger MCS candidate sets will be considered.

APPENDIX A

CONVERGENCE OF THE EM-CNE ALGORITHM

First, we prove the impact of $w$ on the convergence rate of the channel information in Lemma 1. The EM algorithm utilizes the estimates of the previous iteration to update the new estimates of the unknown parameters by iterating between the E-step and the M-step. Thus, a mapping is defined as $\tilde{\beta}[t + 1] = \mathcal{G}(\tilde{\beta}[t])$, where $\mathcal{G}(\cdot)$ is a continuous function. Note that $\mathcal{G}(\cdot)$ can find a stationary point when the EM algorithm converges, i.e., $\tilde{\beta} = \mathcal{G}(\tilde{\beta})$. In our proposed algorithm, $\tilde{\beta}$ may be a local extremum or a global optimal solution, which depends on its initialization. Then, the Taylor’s series expansion of $\mathcal{G}(\cdot)$ w.r.t. $\tilde{\beta}$ can be expressed as [36]

$$\mathcal{G}(\tilde{\beta}[t]) = \mathcal{G}(\tilde{\beta}) + U(\tilde{\beta}[t] - \tilde{\beta}),$$

(37)

where $U = \frac{\partial \mathcal{G}(\beta)}{\partial \beta}|_{\beta = \tilde{\beta}}$. By adopting the mapping function $\tilde{\beta}[t + 1] = \mathcal{G}(\tilde{\beta}[t]),$ (37) can be rewritten as

$$\tilde{\beta}[t + 1] = \mathcal{G}(\tilde{\beta}[t]) = \mathcal{G}(\tilde{\beta}) + U(\tilde{\beta}[t] - \tilde{\beta}).$$

(38)

From [36], we know that the convergence rate $u_c$ of the estimates in the EM algorithm is defined as the largest eigenvalue $\delta_{\text{max}}$ of $U$, i.e., $u_c = \delta_{\text{max}}$. In the following, we define $\hat{h} = [h_0, h_1, \ldots, h_{L-1}]^T$ to simplify the expression of $U$, where $h_{\ell} = a_{\ell e} e^{j\phi_{\ell}}$. Then, formulas (15) and (16) are rewritten as

$$\hat{h}_{\ell}[t + 1] = \frac{1}{P} \sum_{j \in I^N_{\ell}} \hat{s}_{\ell, j}^* \hat{z}_{\ell, j}[t].$$

(39)

Substituting the complete data in (13) into (39), we have

$$\hat{h}_{\ell}[t + 1] = \hat{h}_{\ell}[t] + \frac{1}{P} \sum_{j \in I^N_{\ell}} w_{\ell}(\hat{s}_{\ell, j}^* r_{\ell})$$

$$- \hat{s}_{\ell} \sum_{\ell \in I^N_{\ell}} \hat{h}_{\ell}[t].$$

(40)

By some manipulation, (40) can be simplified as

$$\hat{h}[t] = \left( I_L - \frac{1}{P}(w \otimes 1_L^T) \hat{S} S^H \right)^H \hat{h}[t + 1]$$

$$= (I_L - U) \hat{h} - \frac{1}{P}(w \otimes 1_L^T) \hat{S} S^H (w \otimes 1_L^T) \hat{S} r,$$

(41)

where $\hat{S} \in S^{L \times N}$ represents the transmitted symbol matrix, and the $\ell$-th row of $S$ is the transmitted signal passing through the $\ell$-th path, which has been determined in the SiDR block in Section III-C. Substituting (41) into (38), we have

$$\left( I_L - U \left( I_L - \frac{1}{P}(w \otimes 1_L^T) \hat{S} S^H \right)^H \right) \hat{h}[t + 1]$$

$$= (I_L - U) \hat{h} - \frac{1}{P}(w \otimes 1_L^T) \hat{S} S^H (w \otimes 1_L^T) \hat{S} r.$$

(42)

From the mapping function $\hat{h}[t + 1] = \mathcal{G}(\hat{h}[t])$ and $U = \frac{\partial \mathcal{G}(\hat{h}[t])}{\partial \hat{h}[t]}|_{\hat{h}[t]=\hat{h}}$, we obtain $U$ as

$$U = I_L - \frac{1}{P}(w \otimes 1_L^T) \hat{S} S^H.$$  (44)

The convergence rate $u_c$ is the largest eigenvalue of $U$, which is related to $w$. Hence, we proof Lemma 1 that the noise decomposition factor $w$ has impact on $u_c$.

To further illustrate the statement that the noise decomposition factor $w$ has no impact on the convergence result of channel information, we substitute the E-step in (13) into (39), and (39) can be rewritten as

$$\hat{h}_{\ell}[t + 1] = \frac{1}{P} \sum_{j \in I^N_{\ell}} \hat{s}_{\ell, j}^* \left( z_{\ell, j}[t] + w_{\ell} \left( r_{\ell} - \sum_{\ell \in I^N_{\ell}} \hat{z}_{\ell, j}[t] \right) \right).$$

(45)

By some manipulation, we obtain

$$\hat{h}_{\ell}[t + 1] = \hat{h}_{\ell}[t] + \frac{w_{\ell}}{P} \sum_{j \in I^N_{\ell}} \hat{s}_{\ell, j}^* \left( r_{\ell} - \sum_{\ell \in I^N_{\ell}} \hat{z}_{\ell, j}[t] \right).$$

(46)

It has been proved that when the EM algorithm converges, we have $\|\hat{h}_{\ell}[t + 1] - \hat{h}_{\ell}[t]\|^2 \to 0$. Then, as seen from (46), the impact of the noise on the channel estimation becomes smaller, i.e., $\hat{s}_{\ell, j}^* (r_{\ell} - \sum_{\ell \in I^N_{\ell}} \hat{z}_{\ell, j}[t]) \to 0$, as the iteration proceeds. This indicates that the choice of $w$ has no impact on the convergence result of channel information. Hence, Lemma 1 is concluded from (44) and (46).
APPENDIX B
PROOFS OF THEOREM 1

We prove the LLR of the syndrome APP stated in Theorem 1 here. Considering two i.i.d. Bernoulli random variables \(x_1\) and \(x_2\), the probability of taking \(x_1 \oplus x_2 = 0\) is written as

\[
p(x_1 \oplus x_2 = 0) = p(x_1 = 0)p(x_2 = 0) + (1 - p(x_1 = 0))(1 - p(x_2 = 0))
\]

Then, the LLR metric of \(x_1 \oplus x_2\) is derived as

\[
\mathcal{L}(x_1 \oplus x_2) = \ln \frac{1 + e^{\mathcal{L}(x_1) + \mathcal{L}(x_2)}}{e^{\mathcal{L}(x_1)} + e^{\mathcal{L}(x_2)}}.
\]

Furthermore, for the i.i.d. Bernoulli random variables \(x_j, j \in I_N\), \(\mathcal{L}(x_1 \oplus x_2 \oplus \ldots \oplus x_N)\) can be obtained by adopting the inductive methods

\[
\mathcal{L}(x_1 \oplus x_2 \oplus \ldots \oplus x_N) = \ln \prod_{j \in I_N} \left( \frac{e^{\mathcal{L}(x_j)} + 1}{e^{\mathcal{L}(x_j)} - 1} \right) - 1. \]

By utilizing the function \(\tanh \frac{x}{2}\), \(\mathcal{L}(x_1 \oplus x_2 \oplus \ldots \oplus x_N)\) is rewritten as

\[
\mathcal{L}(x_1 \oplus x_2 \oplus \ldots \oplus x_N) = 2\tanh^{-1} \prod_{j \in I_N} \tanh \frac{1}{2} \mathcal{L}(x_j).
\]

To further prove the LLR of the syndrome APP stated in Theorem 1, we first derive the posterior probability LLR of the coded bit \(c_{j,g}\), which is denoted by

\[
\mathcal{L}(c_{j,g}|r_j, s^{N-1}_{L-1}; \beta) = \lambda^{\text{out}}_{j,g}, \quad j \in I_1^N, \quad g \in I_1^{\log |S|}.
\]

Since we assume perfect synchronization, the relation between the codeword \(\mathbf{e}\) and the coded bits \(c_{j,g}, j \in I_1^N, g \in I_1^{\log |S|}\), is \(\mathbf{e}_{1,1}, \mathbf{e}_{2,1}, \ldots, \mathbf{e}_{N,1} = \mathbf{c}_{1,1}, \mathbf{c}_{1,2}, \ldots, \mathbf{c}_{N,1}\). In addition, \(\psi = \lambda^{\text{out}}\). Given the modulation \(\eta\) and a \((n, q)\) LDPC code \(\zeta\), from (30), (50), and (51), the LLR of the syndrome APP of the \(i\)-th parity-check bit is obtained by

\[
\gamma_i = \mathcal{L}(\mathbf{c}_{N-q+1}^{(1)} \oplus \cdots \oplus \mathbf{c}_{N-q+1}^{(N)} | r, s^{N-q}; \beta^{N-q})
\]

\[
= 2\tanh^{-1} \prod_{\tau \in I_1^{N-q}} \tanh \frac{1}{2} \psi_{i,q}(\tau), \quad i \in I_1^{N-q}. \]

Therefore, Theorem 1 is concluded.

ACKNOWLEDGMENT

The authors gratefully thank Dr. Octavia A. Dobre and Dr. Mohamed Marey for insightful comments and suggestions on the manuscript.
[23] F. Wang, O. A. Dobre, C. Chan, and J. Zhang, “Fold-based Kolmogorov–Smirnov modulation classifier,” IEEE Signal Process. Lett., vol. 23, no. 7, pp. 1003–1007, Jul. 2016.

[24] A. Bonvard, S. Houcke, R. Gautier, and M. Marazin, “Classification based on Euclidean distance distribution for blind identification of error correcting codes in noncooperative contexts,” IEEE Trans. Signal Process., vol. 66, no. 10, pp. 2572–2583, May 2018.

[25] T. Xia and H.-C. Wu, “Joint blind frame synchronization and encoder identification for low-density parity-check codes,” IEEE Commun. Lett., vol. 18, no. 2, pp. 352–355, Feb. 2014.

[26] Z. Wu, L. Zhang, Z. Zhong, and R. Liu, “Blind recognition of LDPC codes over candidate set,” IEEE Commun. Lett., vol. 24, no. 1, pp. 11–14, Jan. 2020.

[27] L. Huang, W. Chen, E. Chen, and H. Chen, “Blind recognition of k/n rate convolutional encoders from noisy observation,” J. Syst. Eng. Electron., vol. 28, no. 2, pp. 235–243, Apr. 2017.

[28] O. Ozdemir, R. Li, and P. K. Varshney, “Hybrid maximum likelihood modulation classification using multiple radios,” IEEE Commun. Lett., vol. 17, no. 10, pp. 1889–1892, Oct. 2013.

[29] J. Zhang, D. Cabric, F. Wang, and Z. Zhong, “Cooperative modulation classification for multipath fading channels via expectation-maximization,” IEEE Trans. Wireless Commun., vol. 16, no. 10, pp. 6698–6711, Oct. 2017.

[30] R. Swaminathan and A. S. Madhukumar, “Classification of error correcting codes and estimation of interleaver parameters in a noisy transmission environment,” IEEE Trans. Broadcast., vol. 63, no. 3, pp. 463–478, Sep. 2017.

[31] E. Kazikli, B. Dulek, and S. Gezici, “Optimal joint modulation classification and symbol decoding,” IEEE Trans. Wireless Commun., vol. 18, no. 5, pp. 2623–2638, May 2019.

[32] T. Xia and H.-C. Wu, “Joint blind frame synchronization and encoder identification for low-density parity-check codes,” IEEE Commun. Lett., vol. 18, no. 2, pp. 352–355, Feb. 2014.

[33] Z. Wu, L. Zhang, Z. Zhong, and R. Liu, “Blind recognition of LDPC codes over candidate set,” IEEE Commun. Lett., vol. 24, no. 1, pp. 11–14, Jan. 2020.

[34] L. Huang, W. Chen, E. Chen, and H. Chen, “Blind recognition of k/n rate convolutional encoders from noisy observation,” J. Syst. Eng. Electron., vol. 28, no. 2, pp. 235–243, Apr. 2017.

[35] O. Ozdemir, B. Dulek, and S. Gezici, “Optimal joint modulation classification with multiple sensors in asynchronous linear modulation classification with multiple sensors in asynchronous linear modulation classification with multiple sensors,” IEEE Trans. Wireless Commun., vol. 14, no. 11, pp. 6389–6400, Nov. 2015.

[36] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” J. Roy. Stat. Soc., B, Methodol., vol. 39, no. 1, pp. 1–22, Sep. 1977.

[37] H. L. Trees, Detection Estimation Modulation Theory, Part I, New York, NY, USA: Wiley, 1968.

[38] A. Gelb, Applied Optimal Estimation. Cambridge, MA, USA: MIT Press, 1974.

[39] G. J. McLachlan and T. Krishnan, The EM Algorithm and Extensions. Hoboken, NJ, USA: Wiley, 2008.

[40] D. Karlis and E. Xekalaki, “Choosing initial values for the EM algorithm for finite mixtures,” Comput. Statist. Data Anal., vol. 41, nos. 3–4, pp. 577–590, Jan. 2003.

[41] Z. Xing and Y. Gao, “A modulation classification algorithm for multipath signals based on cepstrum,” IEEE Trans. Instrum. Meas., vol. 69, no. 7, pp. 4742–4752, Jul. 2020.

[42] M. Lavielle and E. Moulines, “A simulated annealing version of the EM algorithm for non-Gaussian deconvolution,” Statist. Comput., vol. 7, no. 4, pp. 229–236, Dec. 1997.

[43] J. Hagenauer, E. Offer, and L. Papke, “Iterative decoding of binary block and convolutional codes,” IEEE Trans. Inf. Theory, vol. 42, no. 2, pp. 429–445, Mar. 1996.