A Factor Graph Approach to Exploiting Cyclic Prefix for Equalization in OFDM Systems

Jindan Yang, *Student Member, IEEE,* Qinghua Guo, *Member, IEEE,*
Defeng (David) Huang, *Senior Member, IEEE,* and Sven Nordholm, *Senior Member, IEEE*

Abstract—In OFDM systems, cyclic prefix (CP) insertion and removal enables the use of a set of computationally efficient single-tap equalizers at the receiver. Due to the extra transmission time and energy, the CP causes a loss in both spectrum efficiency and power efficiency. On the other hand, as a repetition of part of the data, the CP brings extra information and can be exploited for detection. Therefore, instead of discarding the CP observation as in the conventional OFDM system, we utilize all the received signals in a soft-input soft-output equalizer of a turbo equalization OFDM system. First, the models for both the CP part and the non-CP part of observation are presented in a Forney-style factor graph (FFG). Then based on the computation rules of the FFG and the Gaussian message passing (GMP) technique, we develop an equalization algorithm. With proper approximation, the complexity of the proposed algorithm is reduced to $\mathcal{O}(2RNlog_2N + 4RGlog_2G + 2RG)$ per data block for R iterations, where N is the length of the data block and G is equal to $P + L - 1$ with P the length of the CP and L the maximum delay spread of the channel. To justify the performance improvement, SNR analysis is provided. Simulation results show that the proposed approach achieves a significant gain over the conventional approach and the turbo equalization system converges within two iterations.

Index Terms—OFDM system, cyclic prefix, iterative receiver, factor graph.

I. INTRODUCTION

RECENTLY, orthogonal frequency division multiplexing (OFDM) [1] has been adopted for a variety of applications and wireless broad-band standards, such as digital audio broadcasting (DAB) [2], digital video broadcasting (DVB) [3], IEEE 802.11a [4], MMAC [5] and HIPERLAN/2 [6]. These systems are based on a cyclic prefix (CP) assisted scheme, where the CP, a repetition of the end of an OFDM symbol, is inserted into the front of the symbol before transmission and discarded after reception.

J. Yang, Q. Guo and D. Huang are with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Perth, WA 6009, Australia (e-mail: 20877549@student.uwa.edu.au; david.huang@uwa.edu.au).

Q. Guo is also with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW 2522, Australia (e-mail: qinghua_guo@uow.edu.au).

S. Nordholm is with the Department of Electrical and Computer Engineering, Curtin University, Perth, WA 6102, Australia (email: s.nordholm@curtin.edu.au).

The role of the CP is to convert linear convolution into circular convolution and turn the frequency selective channel into a set of parallel frequency flat channels. As these parallel channels are unlikely to fade simultaneously, coding and interleaving across them can be used to exploit the frequency diversity. On the other hand, due to the extra transmission time and energy induced by the CP, a loss occurs in both the spectrum efficiency and the power efficiency. Some work has been conducted on avoiding the use of the CP [7]. However, considering the compatibility with the existing systems, a number of papers are dedicated to exploiting the prefix redundancy instead, such as [8]–[10] for the purpose of time-frequency offset estimation, [11], [12] for channel estimation, and [13]–[31] for equalization.

In regards to utilizing the CP in equalization, some studies have been carried out in the context of non-iterative systems. For example, [13] investigates this issue in the single-carrier frequency-domain equalization (SC-FDE) system that has a similar CP structure as in the OFDM system, where two independent estimates are obtained from the two observations of the CP and are combined together based on the maximum ratio combining (MRC) rule. For OFDM systems, [14] and [15] assume that the length of the CP is designed to be longer than the maximum delay spread of the channel, and a portion of the CP, which is free of interference from the preceding OFDM symbol, is used to improve the data estimation. As an extension, [16]–[21] exploit the whole CP through interference cancellation. However, the performance gain of these works is quite limited. With a more sophisticated design of the receiver, [22] and [23] achieve greater improvement in performance, but at the cost of a dramatic increase in complexity.

Meantime, inspired by turbo equalization [24]–[26], it is also possible to exploit the CP in iterative systems, where at the receiver, extrinsic information is exchanged between the equalizer and the decoder in an iterative fashion. For example, [27], [28] utilizes the CP to jointly estimate the channel as well as the data using the expectation-maximization (EM) algorithm. For data detection only, the methods in [14]–[23], [29], [30] can be extended to turbo equalization systems to make use of the CP. However, these methods feature a very high complexity, for instance, the minimum mean square error (MMSE) method in [20] is of cubic complexity. In [31], we propose a low-complexity equalization algorithm for the iterative SC-FDE systems. With complexity of $\mathcal{O}(NlogN)$ per data block per iteration where N is the length of the data block, a gain of 0.7dB is achieved under both the additive white Gaussian noise (AWGN) channel and the inter-symbol

interference (ISI) channel with a 1/4 CP ratio.

In this paper, we consider the coded OFDM systems where turbo equalization is employed. We present the models of both the CP and the non-CP parts of the received signal in a Forney-style factor graph (FFG) [32]. Then based on the computation rules of the FFG and the Gaussian message passing (GMP) technique [32], we develop an equalization algorithm, which exploits the prefix redundancy and enhances the system performance. With proper approximation, we propose a reduced-complexity algorithm without compromising the bit-error-rate (BER) performance. The complexity of the reduced-complexity algorithm is $\mathcal{O}(2RN log_2 N + 4RG log_2 G + 2RG)$ per data block for R iterations, where N is the length of the data block and G is equal to $P + L - 1$ with P the length of the CP and L the maximum delay spread of the channel. In the future, with the aid of interference cancellation among the signals from different transmit antennas, it is also possible to apply the proposed factor graph based approach and its associated approximation in multiple-input multiple-output (MIMO) OFDM systems [33]. We summarize the contributions of this paper as follows.

1) For exploiting the CP in OFDM systems, the equalizers proposed in the literature such as [20] and [21] are at least of $\mathcal{O}(N^3)$ complexity per data block per iteration, while our proposed reduced-complexity equalizer is only of $\mathcal{O}(N log_2 N)$ per data block per iteration, therefore making it desirable for use in practical applications.

2) Our proposed equalizer utilizes the full CP observation after interference cancellation and the tail of each block in the succeeding block (see the grey triangles in Fig. 2(b)) as well, while previous equalizers such as [15], [20] and [21] only make use of the CP or part of the CP for detection. As a result, simulation results show that our proposed equalizer achieves 1dB performance gain over the MMSE equalizers in [20] and [21] and 2dB gain over the MMSE equalizer in [15] for the system setup in Section V.

3) An SNR analysis is provided to verify the performance improvement of our proposed equalizer over the conventional equalizer. Such an analysis for exploiting the CP in OFDM systems, to the best of our knowledge, has not appeared in the literature before.

4) For AWGN channel, we have demonstrated that it is necessary to use an iterative receiver (i.e. turbo receiver) to take full advantage of the CP energy in OFDM systems. A bound of 0.97dB for 1/4 CP ratio is achieved within two iterations.

5) We have also investigated the sensitivity of the proposed equalizer to the quality of the available channel state information (CSI). In the presence of channel estimation errors, our proposed equalizer is robust in terms of BER performance.

The rest of this paper is organized as follows. Section II describes the system models. Section III presents the models in a factor graph and develops a low-complexity equalization algorithm. Section IV analyzes the SNR improvement and Section V shows the numerical results. Conclusions are drawn
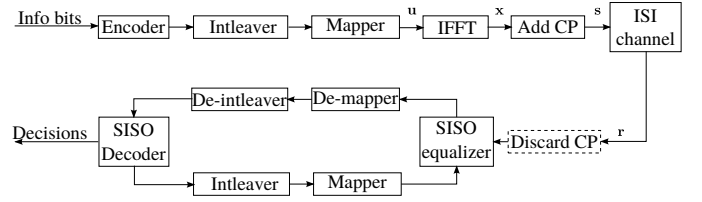

Fig. 1.   A coded OFDM system

in Section VI.

Notations: We use upper (lower) boldface letters to denote matrices (column vectors) and italics to denote scalars. The superscript ^{T} and ^{H} represent the transpose and the conjugate transpose, respectively. $\mathbf{F}_N$ represents a normalized $N \times N$ discrete Fourier transform (DFT) matrix with the (m, n)th element given by $N^{-1/2} e^{-j2\pi mn/N}$, where $j = \sqrt{-1}$. $Diag\,(\mathbf{a})$ is a diagonal matrix with the entries of $\mathbf{a}$ on its diagonal. $(\mathbf{A})_{diag}$ returns a matrix with the diagonal entries of $\mathbf{A}$ on its diagonal and zeros off its diagonal. $\mathbf{I}$ and $\mathbf{0}$ denote an identity matrix and an all-zero vector/matrix, with proper sizes.

II. SYSTEM MODEL

Consider a coded OFDM system as depicted in Fig. 1. The information bit stream is encoded by a channel encoder, then the coded bit stream is interleaved and mapped onto an M-QAM symbol sequence $\mathbf{u}$, $\mathbf{u} = [u_0, u_1, ..., u_{Z-1}]^T$. This sequence is partitioned into K blocks, each of length N, where we assume $Z = KN$. The kth block is denoted by $\mathbf{u}_k = [u_{kN}, u_{kN+1}, ..., u_{(k+1)N-1}]^T$, $k = 0, ..., K - 1$. An N-point IFFT is applied on $\mathbf{u}_k$, producing $\mathbf{x}_k = \mathbf{F}_N^H \mathbf{u}_k$. By copying the last P samples of $\mathbf{x}_k$ to its front, the CP is formed, which we denote by $\tilde{\mathbf{x}}_k = \mathbf{A}\mathbf{x}_k$, where $\mathbf{A} = [\ \mathbf{0}_{P \times (N-P)} \quad \mathbf{I}_{P \times P}\ ]$. For the CP to be effective, $P \geq L - 1$ is required, where L is the maximum delay spread of the multipath channel. The extended block $\mathbf{s}_k = [x_{(k+1)N-P}, ..., x_{(k+1)N-1}, x_{kN}, ..., x_{(k+1)N-1}]^T$, is of length M, $M = N + P$. The symbol sequence $\mathbf{s} = [s_0, s_1, ..., s_{J-1}]^T = [\mathbf{s}_0^T, \mathbf{s}_1^T, ..., \mathbf{s}_{K-1}^T]^T$, where $J = KM$, is transmitted over the channel, and at the receiver side, the sequence $\mathbf{r} = [r_0, r_1, ..., r_{V-1}]^T$, where $V = J + L - 1$, is observed, and

$$r_j = \sum_{l=0}^{L-1} h_{j,l} \cdot s_{j-l} + n_j, \quad j = 0, ..., V - 1 \qquad (1)$$

where $\{h_{j,l}, l = 0, ..., L - 1\}$ is the CSI at time j, and $\{n_j\}$ is the AWGN noise with zero mean and variance $2\sigma^2$. In this paper, we assume that the channel is quasi-static [34], which means that the channel remains the same within the duration of a block but it may vary from block to block.

As shown in Fig. 2, corresponding to $\mathbf{s}_k$, the counterpart block at the receiver $\mathbf{r}_k = [r_{kM}, r_{kM+1}, ..., r_{(k+1)M+L-2}]^T$, is always $L - 1$ symbols longer than $\mathbf{s}_k$ because of the multipath channel. In conventional OFDM systems, the CP is usually removed at the receiver, and only the part $\mathbf{y}_k = [r_{kM+P}, r_{kM+P-1}, ..., r_{(k+1)M-1}]^T$ is used for further processing. This part can be modeled as the circular convolution
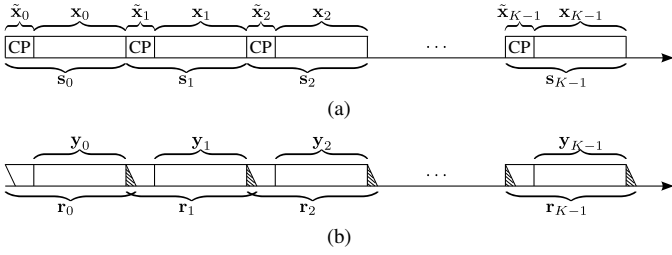
Fig. 2. Data structure. (a) Transmitted signal (b) Received signal

of the data block $\mathbf{x}_k$ and the channel $\mathbf{h}_k$ as follows.

$$\mathbf{y}_k = \mathbf{h}_k \circledast \mathbf{x}_k + \mathbf{n}_k, \quad k = 0, ..., K-1 \tag{2}$$

where "$\circledast$" denotes the circular convolution, $\mathbf{h}_k$ is the multipath channel that $\mathbf{x}_k$ undergoes, $\mathbf{h}_k = [h_k^0, h_k^1, ..., h_k^{L-1}]^T$, and $\mathbf{n}_k$ is the AWGN noise vector with mean $\mathbf{0}$ and covariance $2\sigma^2 \mathbf{I}$.

The rest part of $\mathbf{r}_k$ is a combination of the CP observation and the first $L-1$ received symbols of the next block. It is of length $G = P + L - 1$ and denoted by $\tilde{\mathbf{y}}_k = [r_{kM}, r_{kM+1}, ..., r_{kM+P-1}, r_{(k+1)M}, ..., r_{(k+1)M+L-2}]^T$. Note that $\tilde{\mathbf{y}}_k$ contains the interference from adjacent blocks $\mathbf{s}_{k-1}$ and $\mathbf{s}_{k+1}$, which we denote by $\mathbf{w}_{k-1}^{tail}$ and $\mathbf{w}_{k+1}^{head}$, respectively. Let $tail(\cdot)$ and $head(\cdot)$ be two truncation functions, which return the $L-1$ tail and the $L-1$ head symbols of the sequence in the parentheses, respectively. Then the interference terms $\mathbf{w}_{k-1}^{tail}$ and $\mathbf{w}_{k+1}^{head}$ can be given by

$$\mathbf{w}_{k-1}^{tail} = [tail(\mathbf{h}_{k-1} * \tilde{\mathbf{x}}_{k-1}), \mathbf{0}_{1 \times P}]^T \tag{3}$$

$$\mathbf{w}_{k+1}^{head} = [\mathbf{0}_{1 \times P}, head(\mathbf{h}_{k+1} * \tilde{\mathbf{x}}_{k+1})]^T \tag{4}$$

for $k = 1, ..., K-2$, where "$*$" denotes the linear convolution, and for $k = 0$ and $k = K-1$, we have

$$\mathbf{w}_{-1}^{tail} = \mathbf{0}_{G \times 1}, \qquad \mathbf{w}_K^{head} = \mathbf{0}_{G \times 1}. \tag{5}$$

Modeled as the linear convolutions of the block $\tilde{\mathbf{x}}_k$ and the channel $\mathbf{h}_k$, this part of observation, $\tilde{\mathbf{y}}_k$, can be written as

$$\tilde{\mathbf{y}}_k = \mathbf{h}_k * \tilde{\mathbf{x}}_k + \mathbf{w}_{k-1}^{tail} + \mathbf{w}_{k+1}^{head} + \tilde{\mathbf{n}}_k,$$
$$k = 0, ..., K-1 \tag{6}$$

where $\tilde{\mathbf{n}}_k$ denotes the AWGN noise vector of length G. For notational simplicity, we put the last three terms of (6) together as one error term

$$\mathbf{v}_k = \mathbf{w}_{k-1}^{tail} + \mathbf{w}_{k+1}^{head} + \tilde{\mathbf{n}}_k. \tag{7}$$

As a result, (6) becomes

$$\tilde{\mathbf{y}}_k = \mathbf{h}_k * \tilde{\mathbf{x}}_k + \mathbf{v}_k. \tag{8}$$

Since linear convolution after proper zero-padding can be transformed into circular convolution, and circular convolution in the time domain can be transformed into multiplication in the frequency domain by DFT operation, we can rewrite (2) and (8) as

$$\mathbf{y}_k = \mathbf{F}_N^H \mathbf{H}_k \mathbf{F}_N \mathbf{x}_k + \mathbf{n}_k, \tag{9}$$

$$\tilde{\mathbf{y}}_k = \mathbf{F}_G^H \tilde{\mathbf{H}}_k \mathbf{F}_G \mathbf{Q} \tilde{\mathbf{x}}_k + \mathbf{v}_k \tag{10}$$

where $\mathbf{Q} = \left[\begin{array}{cc} \mathbf{I}_{P \times P} & \mathbf{0}_{P \times (L-1)} \end{array}\right]^T$ is to adapt $\tilde{\mathbf{x}}_k$ for the DFT operation by adding $L-1$ zeros to its end, and

$$\mathbf{H}_k = Diag\left(\sqrt{N} \mathbf{F}_N \hat{\mathbf{h}}_k\right), \quad \text{with } \hat{\mathbf{h}}_k = \left[\begin{array}{cc} \mathbf{h}_k^T & \mathbf{0}_{1 \times (N-L)} \end{array}\right]^T, \tag{11}$$

$$\tilde{\mathbf{H}}_k = Diag\left(\sqrt{G} \mathbf{F}_G \check{\mathbf{h}}_k\right), \quad \text{with } \check{\mathbf{h}}_k = \left[\begin{array}{cc} \mathbf{h}_k^T & \mathbf{0}_{1 \times (G-L)} \end{array}\right]^T. \tag{12}$$

Substituting $\mathbf{x}_k = \mathbf{F}_N^H \mathbf{u}_k$ and $\tilde{\mathbf{x}}_k = \mathbf{A}\mathbf{x}_k$ into (9) and (10), respectively, we have

$$\mathbf{y}_k = \mathbf{F}_N^H \mathbf{H}_k \mathbf{u}_k + \mathbf{n}_k, \tag{13}$$

$$\tilde{\mathbf{y}}_k = \mathbf{F}_G^H \tilde{\mathbf{H}}_k \mathbf{B} \mathbf{u}_k + \mathbf{v}_k \tag{14}$$

with $\mathbf{B} = \mathbf{F}_G \mathbf{Q} \mathbf{A} \mathbf{F}_N^H$. The term $\mathbf{v}_k$ can also be written in a similar form as

$$\mathbf{v}_k = \tilde{\mathbf{n}}_k + \mathbf{C} \mathbf{F}_G^H \tilde{\mathbf{H}}_{k-1} \mathbf{B} \mathbf{u}_{k-1} + \mathbf{C}^H \mathbf{F}_G^H \tilde{\mathbf{H}}_{k+1} \mathbf{B} \mathbf{u}_{k+1} \tag{15}$$

where the matrix $\mathbf{C}$ and its Hermitian transpose $\mathbf{C}^H$ function as $tail(\cdot)$ and $head(\cdot)$, respectively, and

$$\mathbf{C} = \left[\begin{array}{cc} \mathbf{0}_{(L-1) \times P} & \mathbf{I}_{(L-1) \times (L-1)} \\ \mathbf{0}_{P \times P} & \mathbf{0}_{P \times (L-1)} \end{array}\right]. \tag{16}$$

III. FACTOR GRAPH BASED EQUALIZATION ALGORITHMS

In this section, we present the models (13) and (14) in a factor graph, and based on the computation rules of the graph, we develop an equalization algorithm that utilizes all the received signals.

The same notational convention as in [32] is used in this paper for a Forney-style factor graph (FFG), where an edge or a half-edge represents a variable, and a node represents a factor (or a function). Variables in the graph of this paper are all vectors. Messages about these variables are assumed to be with Gaussian distribution, parameterized by their mean vectors and covariance matrices. To update the messages, the Gaussian message passing (GMP) technique in [32] is employed.

Let $\mathbf{x}$ be a vector variable, represented by a directed edge, i.e. an edge with an arrow, in an FFG. We use $\overrightarrow{\mathbf{m}}_\mathbf{x}$ and $\overrightarrow{\mathbf{V}}_\mathbf{x}$ to denote the mean vector and the covariance matrix of the message that flows in the direction of the edge, and we call it the forward message, while $\overleftarrow{\mathbf{m}}_\mathbf{x}$ and $\overleftarrow{\mathbf{V}}_\mathbf{x}$ are for the message in the opposite direction, and we call it the backward message. Moreover, the marginal message which is the product of these two messages, is denoted by $\mathbf{m}_\mathbf{x}$ and $\mathbf{V}_\mathbf{x}$. As in [32], the transformed mean and the weight matrix (e.g. $\mathbf{W}\mathbf{m} = \mathbf{V}^{-1}\mathbf{m}$ and $\mathbf{W} = \mathbf{V}^{-1}$ for the marginal message), are sometimes used as an equivalent pair for the mean vector and the covariance matrix.

A. FFG of model (13) and model (14)

According to model (13) and model (14), we can draw an FFG as illustrated in Fig. 3. The conventional equalizer only uses the $\mathbf{y}_k$ part of the observation, which corresponds to model (13) and the right branch of the graph, whereas the proposed equalizer utilizes both $\mathbf{y}_k$ and $\tilde{\mathbf{y}}_k$, which are related
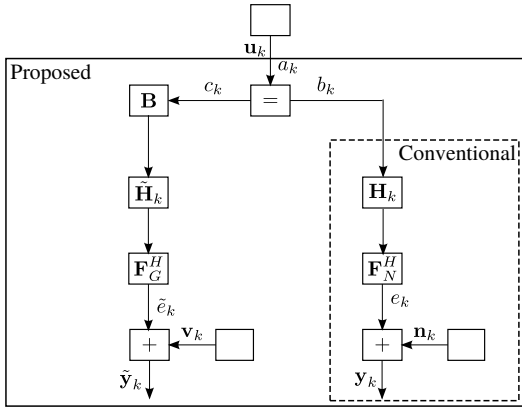
Fig. 3. The kth block of the Forney-style factor graph for both conventional and proposed equalizers in an OFDM system

to models (13) and (14), which are represented by the right branch and the left branch of the graph, respectively.

In each iteration of the receiver, the soft decoder outputs extrinsic log-likelihood ratios (LLRs) about the coded bits, which are then interleaved and converted into probabilities about the data symbols. Characterized by the mean vector $\overrightarrow{\mathbf{m}}_{a_k}$ and the covariance matrix $\overrightarrow{\mathbf{V}}_{a_k}$, these probabilities are fed into the equalizer as its *a priori* information. Then based on the *a priori* information, the observation and the channel information, the equalizer produces extrinsic means $\mathbf{m}_k^{ext}$ and variances $\mathbf{V}_k^{ext}$ about the data symbols. These means and variances are passed onto the de-mapper where they are converted back to LLRs. Then these LLRs are de-interleaved and fed into the soft decoder serving as its *a priori* information.

At the first iteration, there is no information coming from the decoder and the *a priori* information for the equalizer is normally set to $\overrightarrow{\mathbf{m}}_{a_k} = \mathbf{0}$ and $\overrightarrow{\mathbf{V}}_{a_k} = \mathbf{I}$. For other iterations, it is a common practice to model $\overrightarrow{\mathbf{V}}_{a_k}$ as a diagonal matrix, since symbols within a block can be treated as independent of each other due to the use of the interleaver.

For the conventional equalizer when applied in an iterative system, the *a priori* information $(\overrightarrow{\mathbf{m}}_{a_k}, \overrightarrow{\mathbf{V}}_{a_k})$ is of no use in the calculation of the extrinsic information $(\mathbf{m}_k^{ext}, \mathbf{V}_k^{ext})$. All the sub-carriers are independent of each other in the frequency domain, and the *a priori* information of one symbol does not contribute to the extrinsic information of the others in the same block. However, for the proposed equalizer, the *a priori* information is very useful. Details are provided in the following two subsections, where the message passing process is described and points a_k, b_k, c_k, e_k and $\tilde{e}_k$ in Fig. 3 are used to indicate the location of the message in discussion.

B. Message passing for the conventional equalizer

1) As the noise is with zero mean and $2\sigma^2$ variance, following the rules (II.10) and (II.8) in [32], the backward message at point e_k is given by

$$\overleftarrow{\mathbf{m}}_{e_k} = \mathbf{y}_k, \quad \overleftarrow{\mathbf{V}}_{e_k} = \beta \mathbf{I}_N \qquad (17)$$

where $\beta = 2\sigma^2$. Applying the rules (III.6) and (III.5) in [32], we have the backward message at point b_k as

$$\overleftarrow{\mathbf{W}}_{b_k} \overleftarrow{\mathbf{m}}_{b_k} = \beta^{-1} \mathbf{H}_k^H \mathbf{F}_N \mathbf{y}_k, \qquad (18)$$

$$\overleftarrow{\mathbf{W}}_{b_k} = \beta^{-1} \mathbf{H}_k^H \mathbf{H}_k \qquad (19)$$

where $\overleftarrow{\mathbf{W}}_{b_k} \overleftarrow{\mathbf{m}}_{b_k}$ is the transformed mean vector, and $\overleftarrow{\mathbf{W}}_{b_k} = \overleftarrow{\mathbf{V}}_{b_k}^{-1}$ is the weight matrix. This message can also be written in the mean/covariance form as

$$\overleftarrow{\mathbf{m}}_{b_k} = \mathbf{H}_k^{-1} \mathbf{F}_N \mathbf{y}_k, \qquad (20)$$

$$\overleftarrow{\mathbf{V}}_{b_k} = \left[\beta^{-1} \mathbf{H}_k^H \mathbf{H}_k \right]^{-1}. \qquad (21)$$

2) Combining the backward message at point b_k with the *a priori* message of $\mathbf{u}_k$, we have the *a posteriori* message or marginal message of $\mathbf{u}_k$ as

$$\mathbf{W}_{\mathbf{u}_k} \mathbf{m}_{\mathbf{u}_k} = \overleftarrow{\mathbf{W}}_{b_k} \overleftarrow{\mathbf{m}}_{b_k} + \overrightarrow{\mathbf{V}}_{a_k}^{-1} \overrightarrow{\mathbf{m}}_{a_k}, \qquad (22)$$

$$\mathbf{W}_{\mathbf{u}_k} = \overleftarrow{\mathbf{W}}_{b_k} + \overrightarrow{\mathbf{V}}_{a_k}^{-1}. \qquad (23)$$

To extract the extrinsic information of each symbol in $\mathbf{u}_k$, we use the relations (14) and (15) in [35]. Since $\overleftarrow{\mathbf{V}}_{b_k}$ and $\overrightarrow{\mathbf{V}}_{a_k}$ are diagonal, the extrinsic variances and means of the data symbols in the kth block, $k = 0, ..., K-1$, can be written as

$$\mathbf{V}_k^{ext} = \left[\mathbf{W}_{\mathbf{u}_k} - \overrightarrow{\mathbf{V}}_{a_k}^{-1} \right]^{-1} = \overleftarrow{\mathbf{V}}_{b_k}, \qquad (24)$$

$$\mathbf{m}_k^{ext} = \mathbf{V}_k^{ext} \left[\mathbf{W}_{\mathbf{u}_k} \mathbf{m}_{\mathbf{u}_k} - \overrightarrow{\mathbf{V}}_{a_k}^{-1} \overrightarrow{\mathbf{m}}_{a_k} \right] = \overleftarrow{\mathbf{m}}_{b_k}. \qquad (25)$$

Note that from (24) and (25), the extrinsic information is exactly the same as the backward message at point b_k, and that there is no term associated with the *a priori* information in the expression of the extrinsic information, i.e. the *a priori* information does not contribute to the extrinsic information in the conventional equalizer case.

Once the extrinsic means and variances about the symbols are available, we can use (16) in [35] to calculate the LLRs about the coded bits. The equalizer presented here is the same as the conventional soft-output equalizer mentioned in [33].

C. Message passing for the proposed equalizer

1) As shown in Fig. 3, the right branch of the graph for the proposed equalizer is exactly the same as the graph for the conventional equalizer. Therefore, the forward message at point c_k is the same as the *a posteriori* message of $\mathbf{u}_k$ in the conventional equalizer. Using (22), (23) and (18), (19), we have

$$\overrightarrow{\mathbf{W}}_{c_k} \overrightarrow{\mathbf{m}}_{c_k} = \overrightarrow{\mathbf{V}}_{a_k}^{-1} \overrightarrow{\mathbf{m}}_{a_k} + \beta^{-1} \mathbf{H}_k^H \mathbf{F}_N \mathbf{y}_k, \qquad (26)$$

$$\overrightarrow{\mathbf{W}}_{c_k} = \overrightarrow{\mathbf{V}}_{a_k}^{-1} + \beta^{-1} \mathbf{H}_k^H \mathbf{H}_k, \qquad (27)$$

or in the mean/covariance form as

$$\overrightarrow{\mathbf{m}}_{c_k} = \overrightarrow{\mathbf{V}}_{c_k} \left(\overrightarrow{\mathbf{V}}_{a_k}^{-1} \overrightarrow{\mathbf{m}}_{a_k} + \beta^{-1} \mathbf{H}_k^H \mathbf{F}_N \mathbf{y}_k \right), \qquad (28)$$

$$\overrightarrow{\mathbf{V}}_{c_k} = \left(\overrightarrow{\mathbf{V}}_{a_k}^{-1} + \beta^{-1} \mathbf{H}_k^H \mathbf{H}_k \right)^{-1}. \qquad (29)$$

2) Using rules (II.10) and (II.8) in [32], we have the backward message at point $\tilde{e}_k$ as

$$\overleftarrow{\mathbf{m}}_{\tilde{e}_k} = \tilde{\mathbf{y}}_k - \overrightarrow{\mathbf{m}}_{\mathbf{v}_k}, \quad \overleftarrow{\mathbf{V}}_{\tilde{e}_k} = \overrightarrow{\mathbf{V}}_{\mathbf{v}_k}. \quad (30)$$

The mean vector of $\mathbf{v}_k$ in (15) is updated using $\overrightarrow{\mathbf{m}}_{c_{k-1}}$ and $\overrightarrow{\mathbf{m}}_{c_{k+1}}$, which are the latest mean vectors about $\mathbf{u}_{k-1}$ and $\mathbf{u}_{k+1}$,

$$\overrightarrow{\mathbf{m}}_{\mathbf{v}_k} = \mathbf{C}\mathbf{F}_G^H \boldsymbol{\xi}_{k-1} + \mathbf{C}^H \mathbf{F}_G^H \boldsymbol{\xi}_{k+1} \quad (31)$$

with $\boldsymbol{\xi}_k = \tilde{\mathbf{H}}_k \mathbf{B} \overrightarrow{\mathbf{m}}_{c_k}$. Using (15), we approximate the covariance matrix of $\mathbf{v}_k$ by

$$\overrightarrow{\mathbf{V}}_{\mathbf{v}_k} = \lambda_k \mathbf{I}_G \quad (32)$$

where

$$\lambda_k = 2\sigma^2 + \frac{1}{G} \sum_{l=0}^{L-1} \left[\gamma_{k+1}(L-1-l) \left| h_{k+1}^l \right|^2 + \gamma_{k-1} l \left| h_{k-1}^l \right|^2 \right], \quad (33)$$

and

$$\gamma_k = \frac{1}{N} \sum_{i=0}^{N-1} (\overrightarrow{\mathbf{W}}_{c_k}^{-1})_{i,i}. \quad (34)$$

Based on rules (III.6) and (III.5) in [32], the backward message at point c_k is

$$\overleftarrow{\mathbf{W}}_{c_k} \overleftarrow{\mathbf{m}}_{c_k} = \lambda_k^{-1} \mathbf{B}^H \tilde{\mathbf{H}}_k^H \mathbf{F}_G \left(\tilde{\mathbf{y}}_k - \overrightarrow{\mathbf{m}}_{\mathbf{v}_k} \right), \quad (35)$$

$$\overleftarrow{\mathbf{W}}_{c_k} = \lambda_k^{-1} \boldsymbol{\Psi}_k \quad (36)$$

where $\boldsymbol{\Psi}_k = \mathbf{B}^H \tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k \mathbf{B}$.

3) To obtain the *a posteriori* message at point c_k, we first have the covariance matrix as

$$\mathbf{V}_{c_k} = \left(\overrightarrow{\mathbf{W}}_{c_k} + \overleftarrow{\mathbf{W}}_{c_k} \right)^{-1}$$
$$= \left(\overrightarrow{\mathbf{V}}_{a_k}^{-1} + \beta^{-1} \mathbf{H}_k^H \mathbf{H}_k + \lambda_k^{-1} \boldsymbol{\Psi}_k \right)^{-1}. \quad (37)$$

For the mean vector, by substituting rule (I.1) into rule (I.6) in [32], we get $\mathbf{m} = \overrightarrow{\mathbf{m}} + \mathbf{V}(\overleftarrow{\mathbf{W}}\overleftarrow{\mathbf{m}} - \overleftarrow{\mathbf{W}}\overrightarrow{\mathbf{m}})$. Using (35) and (36), the *a posteriori* mean vector at point c_k is given by

$$\mathbf{m}_{c_k} = \overrightarrow{\mathbf{m}}_{c_k} + \mathbf{V}_{c_k} \left(\overleftarrow{\mathbf{W}}_{c_k} \overleftarrow{\mathbf{m}}_{c_k} - \overleftarrow{\mathbf{W}}_{c_k} \overrightarrow{\mathbf{m}}_{c_k} \right)$$
$$= \overrightarrow{\mathbf{m}}_{c_k} + \lambda_k^{-1} \mathbf{V}_{c_k} \mathbf{B}^H \tilde{\mathbf{H}}_k^H$$
$$\cdot \left[\mathbf{F}_G \left(\tilde{\mathbf{y}}_k - \overrightarrow{\mathbf{m}}_{\mathbf{v}_k} \right) - \tilde{\mathbf{H}}_k \mathbf{B} \overrightarrow{\mathbf{m}}_{c_k} \right]. \quad (38)$$

Then according to rules (II.5) and (II.6) in [32], we have the *a posteriori* message of $\mathbf{u}_k$ as

$$\mathbf{m}_{\mathbf{u}_k} = \mathbf{m}_{c_k}, \quad \mathbf{V}_{\mathbf{u}_k} = \mathbf{V}_{c_k}. \quad (39)$$

4) To calculate the extrinsic information of each data symbol, we extract the *a posteriori* mean/variance of the symbol, and subtract its *a priori* mean/variance using the relationships (14) and (15) in [35] as

$$\mathbf{V}_k^{ext} = \left\{ \left[(\mathbf{V}_{\mathbf{u}_k})_{diag} \right]^{-1} - \overrightarrow{\mathbf{V}}_{a_k}^{-1} \right\}^{-1}, \quad (40)$$

$$\mathbf{m}_k^{ext} = \mathbf{V}_k^{ext} \left\{ \left[(\mathbf{V}_{\mathbf{u}_k})_{diag} \right]^{-1} \mathbf{m}_{\mathbf{u}_k} - \overrightarrow{\mathbf{V}}_{a_k}^{-1} \overrightarrow{\mathbf{m}}_{a_k} \right\}. \quad (41)$$

TABLE I
THE PROPOSED EQUALIZATION ALGORITHMS WITH AND WITHOUT APPROXIMATION

A. Initialization:
1) The blocks $\mathbf{y}_k$, $\tilde{\mathbf{y}}_k$, $k = 0, ..., K-1$, are observed. The channel state information $\mathbf{h}_k$, $k = 0, ..., K-1$ and the noise variance $\beta = 2\sigma^2$ are either assumed to be known or made available by estimation.
2) At each iteration, the *a priori* information $(\overrightarrow{\mathbf{m}}_{a_k}, \overrightarrow{\mathbf{V}}_{a_k})$ is fed into the equalizer and the following steps are executed to calculate the extrinsic information $(\mathbf{m}_k^{ext}, \mathbf{V}_k^{ext})$. Variables associated with indices -1 and K are treated as zero.

B. Calculations of $\overrightarrow{\mathbf{W}}_{c_k}$, $\overrightarrow{\mathbf{m}}_{c_k}$, λ_k and $\boldsymbol{\xi}_k$:
1) **for** $k = 0, ..., K-1$
   a) $\overrightarrow{\mathbf{W}}_{c_k} = \overrightarrow{\mathbf{V}}_{a_k}^{-1} + \beta^{-1} \mathbf{H}_k^H \mathbf{H}_k$
   b) $\gamma_k = \frac{1}{N} \sum_{i=0}^{N-1} (\overrightarrow{\mathbf{W}}_{c_k}^{-1})_{i,i}$
   c) $\overrightarrow{\mathbf{m}}_{c_k} = \overrightarrow{\mathbf{W}}_{c_k}^{-1} \left[\overrightarrow{\mathbf{V}}_{a_k}^{-1} \overrightarrow{\mathbf{m}}_{a_k} + \beta^{-1} \mathbf{H}_k^H \mathbf{F}_N \mathbf{y}_k \right]$
   end for
2) **for** $k = 0, ..., K-1$
   a) $\lambda_k = \beta + \frac{1}{G} \sum_{l=0}^{L-1} \left[\gamma_{k+1}(L-1-l) \left| h_{k+1}^l \right|^2 + \gamma_{k-1} l \left| h_{k-1}^l \right|^2 \right]$
   b) $\boldsymbol{\xi}_k = \tilde{\mathbf{H}}_k \mathbf{B} \overrightarrow{\mathbf{m}}_{c_k}$
   end for

C1. Exact calculations of $\mathbf{V}_k^{ext}$ and $\mathbf{m}_k^{ext}$:
   for $k = 0, ..., K-1$
   a) $\mathbf{V}_{\mathbf{u}_k} = \left[\overrightarrow{\mathbf{W}}_{c_k} + \lambda_k^{-1} \boldsymbol{\Psi}_k \right]^{-1}$
   b) $\mathbf{m}_{\mathbf{u}_k} = \overrightarrow{\mathbf{m}}_{c_k} + \lambda_k^{-1} \mathbf{V}_{\mathbf{u}_k} \mathbf{B}^H \tilde{\mathbf{H}}_k^H$
      $\cdot \left[\mathbf{F}_G \left(\tilde{\mathbf{y}}_k - \mathbf{C}\mathbf{F}_G^H \boldsymbol{\xi}_{k-1} - \mathbf{C}^H \mathbf{F}_G^H \boldsymbol{\xi}_{k+1} \right) - \boldsymbol{\xi}_k \right]$
   c) $\mathbf{V}_k^{ext} = \left\{ \left[(\mathbf{V}_{\mathbf{u}_k})_{diag} \right]^{-1} - \overrightarrow{\mathbf{V}}_{a_k}^{-1} \right\}^{-1}$
   d) $\mathbf{m}_k^{ext} = \mathbf{V}_k^{ext} \left\{ \left[(\mathbf{V}_{\mathbf{u}_k})_{diag} \right]^{-1} \mathbf{m}_{\mathbf{u}_k} - \overrightarrow{\mathbf{V}}_{a_k}^{-1} \overrightarrow{\mathbf{m}}_{a_k} \right\}$
   end for

C2. Approximated calculations of $\mathbf{V}_k^{ext}$ and $\mathbf{m}_k^{ext}$:
   for $k = 0, ..., K-1$
   a) $\mathbf{V}_k^{ext} = \left[\beta^{-1} \mathbf{H}_k^H \mathbf{H}_k + \lambda_k^{-1} \boldsymbol{\Phi}_k \right]^{-1}$
   b) $\mathbf{m}_k^{ext} = \mathbf{V}_k^{ext} \left\{ \beta^{-1} \mathbf{H}_k^H \mathbf{F}_N \mathbf{y}_k + \lambda_k^{-1} \boldsymbol{\Phi}_k \overrightarrow{\mathbf{m}}_{c_k} + \lambda_k^{-1} \mathbf{B}^H \tilde{\mathbf{H}}_k^H \right.$
      $\left. \cdot \left[\mathbf{F}_G \left(\tilde{\mathbf{y}}_k - \mathbf{C}\mathbf{F}_G^H \boldsymbol{\xi}_{k-1} - \mathbf{C}^H \mathbf{F}_G^H \boldsymbol{\xi}_{k+1} \right) - \boldsymbol{\xi}_k \right] \right\}$
   end for

As shown in (39) and (37), different from the conventional equalizer, the *a posteriori* covariance matrix $\mathbf{V}_{\mathbf{u}_k}$ for the proposed equalizer is not diagonal, and the *a posteriori* variances of the data symbols are given by the diagonal entries of $\mathbf{V}_{\mathbf{u}_k}$. Due to the existence of the non-zero off-diagonal entries in $\mathbf{V}_{\mathbf{u}_k}$, the *a posteriori* information of a symbol includes the *a priori* information of other symbols in the data block, and even after subtracting the *a priori* information of its own (see (40) and (41)), the *a priori* information of other symbols is still contained in the extrinsic information of the symbol.

We summarize the proposed algorithm as Parts A, B, and C1 in Table I. Note that this equalizer features a higher latency than the conventional equalizer and some of the works in the literature such as [15], [20] and [29]. However, such a latency is intrinsic to a turbo receiver due to the use of the interleaver/de-interleaver. In a turbo equalization system, the whole sequence associated with a codeword must be processed

TABLE II
COMPLEXITY OF THE PROPOSED ALGORITHM IN SECTION III.C AND THE PROPOSED REDUCED-COMPLEXITY ALGORITHM IN SECTION III.D

The Proposed Algorithm		The Proposed Reduced-complexity Algorithm	
$\mathbf{H}_k, \tilde{\mathbf{H}}_k$	$\mathcal{O}(Nlog_2 N + Glog_2 G)$	$\mathbf{H}_k, \tilde{\mathbf{H}}_k$	$\mathcal{O}(Nlog_2 N + Glog_2 G)$
$\boldsymbol{\Psi}_k$	$\mathcal{O}(N^3)$	$\boldsymbol{\Phi}_k$	$\mathcal{O}(Nlog_2 N + Glog_2 G)$
$\mathbf{H}_k^H \mathbf{F}_N \mathbf{y}_k$	$\mathcal{O}(Nlog_2 N + N)$	$\mathbf{H}_k^H \mathbf{F}_N \mathbf{y}_k$	$\mathcal{O}(Nlog_2 N + N)$
$\boldsymbol{\xi}_k$	$\mathcal{O}(RNlog_2 N + RGlog_2 G + RG)$	$\boldsymbol{\xi}_k$	$\mathcal{O}(RNlog_2 N + RGlog_2 G + RG)$
$\mathbf{V}_{\mathbf{u}_k}$	$\mathcal{O}(RN^3)$	$\mathbf{m}_k^{ext}$	$\mathcal{O}(RNlog_2 N + 3RGlog_2 G + RG)$
$\mathbf{m}_{\mathbf{u}_k}$	$\mathcal{O}(RN^2 + RNlog_2 N + 3RGlog_2 G + RG)$		

before it can be passed on to the equalizer or the decoder.

Note also that Fig. 3 is cycle-free since it only depicts the kth block of the whole Forney-style factor graph. If we consider all the k blocks, $k = 0, ..., K - 1$, the dependence of $\mathbf{v}_k$ on $\mathbf{u}_{k-1}$ and $\mathbf{u}_{k+1}$ in model (15) results in cycles in the graph. Thus, in order to use the GMP technique (which is a special case of the sum-product algorithm [36]), some scheduling schemes regarding the message passing must be employed. [36] suggests a serial scheduling scheme and a flooding scheduling scheme. The algorithms shown in Table I can be treated as a hybrid scheduling scheme, which passes messages within a block in a serial manner and passes messages between blocks in a flooding manner.

D. The proposed reduced-complexity algorithm

In general, for an N-point block, the complexity of the FFT operation is $\mathcal{O}(NlogN)$, while matrix multiplication is $\mathcal{O}(N^3)$, matrix-vector multiplication is $\mathcal{O}(N^2)$ and matrix inversion is $\mathcal{O}(N^3)$. Therefore, it is desirable to avoid the matrix multiplication, matrix-vector multiplication and matrix inversion and use FFT or other means of implementation instead. As seen from (36), in the proposed algorithm there is a full matrix $\boldsymbol{\Psi}_k$, and its calculation involves not only FFT operation but also matrix multiplication. Moreover, because of the existence of $\boldsymbol{\Psi}_k$, matrix inversion occurs in the calculation of $\mathbf{V}_{\mathbf{u}_k}$ (see C1.a in Table I), and because the matrix $\mathbf{V}_{\mathbf{u}_k}$ is full, matrix-vector multiplication is involved in the calculation of $\mathbf{m}_{\mathbf{u}_k}$ (see C1.b in Table I) too. Since the calculation of $\mathbf{m}_{\mathbf{u}_k}$ and $\mathbf{V}_{\mathbf{u}_k}$ needs to be done every block and in every iteration, a very high computational demand is incurred.

To reduce the complexity, we make the following approximation

$$\mathbf{V}_{\mathbf{u}_k} = \left[\overrightarrow{\mathbf{W}}_{c_k} + \lambda_k^{-1} \boldsymbol{\Phi}_k \right]^{-1} \quad (42)$$

where $\boldsymbol{\Phi}_k = \left(\boldsymbol{\Psi}_k \right)_{diag}$. Since $\overrightarrow{\mathbf{W}}_{c_k}$ is diagonal, $\mathbf{V}_{\mathbf{u}_k}$ becomes a diagonal matrix too. However, its diagonal entries are still different from the exact values (i.e. $\left[\left(\overrightarrow{\mathbf{W}}_{c_k} + \lambda_k^{-1} \boldsymbol{\Psi}_k \right)^{-1} \right]_{diag}$), due to the matrix inversion.

With the approximation of (42), the full matrix inversion originally involved in C1.a in Table I is turned into the computation of scalar reciprocals, and with $\mathbf{V}_{\mathbf{u}_k}$ approximated to be a diagonal matrix, the matrix-vector multiplication that previously exists in C1.b in Table I is also reduced to scalar

multiplications, i.e.

$$\mathbf{m}_{\mathbf{u}_k} = \overrightarrow{\mathbf{m}}_{c_k} + \lambda_k^{-1} \left[\overrightarrow{\mathbf{W}}_{c_k} + \lambda_k^{-1} \boldsymbol{\Phi}_k \right]^{-1} \mathbf{B}^H \tilde{\mathbf{H}}_k^H$$
$$\cdot \left[\mathbf{F}_G \left(\tilde{\mathbf{y}}_k - \mathbf{C} \mathbf{F}_G^H \boldsymbol{\xi}_{k-1} - \mathbf{C}^H \mathbf{F}_G^H \boldsymbol{\xi}_{k+1} \right) - \boldsymbol{\xi}_k \right]. \quad (43)$$

Furthermore, the calculation of $\boldsymbol{\Phi}_k$ can be accomplished by a fast approach presented in Appendix A, where full matrix multiplication is avoided and only one G-point FFT and one N-point FFT are required.

Substituting (42) and (43) into C1.c and C1.d in Table I, we can write the extrinsic variances and means of the data symbols as

$$\mathbf{V}_k^{ext} = \left[\beta^{-1} \mathbf{H}_k^H \mathbf{H}_k + \lambda_k^{-1} \boldsymbol{\Phi}_k \right]^{-1} \quad (44)$$

$$\mathbf{m}_k^{ext} = \mathbf{V}_k^{ext} \Big\{ \beta^{-1} \mathbf{H}_k^H \mathbf{F}_N \mathbf{y}_k + \lambda_k^{-1} \boldsymbol{\Phi}_k \overrightarrow{\mathbf{m}}_{c_k} + \lambda_k^{-1} \mathbf{B}^H \tilde{\mathbf{H}}_k^H$$
$$\cdot \left[\mathbf{F}_G \left(\tilde{\mathbf{y}}_k - \mathbf{C} \mathbf{F}_G^H \boldsymbol{\xi}_{k-1} - \mathbf{C}^H \mathbf{F}_G^H \boldsymbol{\xi}_{k+1} \right) - \boldsymbol{\xi}_k \right) \Big] \Big\}. \quad (45)$$

The proposed reduced-complexity equalization algorithm is summarized as Parts A, B and C2 in Table I.

To evaluate the reduction of complexity due to the use of approximation (42), we count the complex multiplications required for processing one data block for R iterations between the equalizer and the decoder. In Table II, a breakdown list of the complexity is provided for the proposed algorithm with and without approximation. There are three types of computation that significantly contribute to the overall complexity.

1) $\mathbf{H}_k$, $\tilde{\mathbf{H}}_k$ and $\boldsymbol{\Psi}_k$ for the proposed algorithm without approximation, and $\mathbf{H}_k$, $\tilde{\mathbf{H}}_k$ and $\boldsymbol{\Phi}_k$ for the proposed reduced-complexity algorithm, need to be calculated once per CSI update;
2) $\mathbf{H}_k^H \mathbf{F}_N \mathbf{y}_k$ needs to be calculated once per received signal block for both with and without approximation;
3) $\boldsymbol{\xi}_k$, $\mathbf{V}_{\mathbf{u}_k}$ and $\mathbf{m}_{\mathbf{u}_k}$ for the proposed algorithm without approximation, and $\boldsymbol{\xi}_k$ and $\mathbf{m}_k^{ext}$ for the proposed reduced-complexity algorithm, need to be calculated once per iteration.

Without approximation (42), the complexity of the proposed algorithm is dominated by cubic terms, i.e. $\mathcal{O}(RN^3 + N^3)$ per data block for R iterations, while with approximation, its overall complexity is dominated by the third type and only a few FFTs are needed, i.e. $\mathcal{O}(2RNlog_2 N + 4RGlog_2 G + 2RG)$ per data block for R iterations.

IV. SNR Performance Analysis

To enable the SNR analysis, we rewrite $\mathbf{m}_k^{ext}$ into the following form

$$\mathbf{m}_k^{ext} = \mathbf{u}_k + \boldsymbol{\zeta}_k \qquad (46)$$

where $\mathbf{u}_k$ is the kth transmitted data block, and $\boldsymbol{\zeta}_k$ is the estimation error vector.

To distinguish from the mean and covariance of a Gaussian message in an FFG, we use $\mathrm{E}(\boldsymbol{\zeta}_k)$ and $\mathrm{Var}(\boldsymbol{\zeta}_k)$ to denote the mean vector and covariance matrix of $\boldsymbol{\zeta}_k$. When $\mathrm{E}(\boldsymbol{\zeta}_k)$ is zero, the SNR of the ith sub-carrier in the kth block is given by

$$\mathrm{SNR}_{k,i} = \frac{P_{k,i}}{\mathrm{Var}(\boldsymbol{\zeta}_k)_{i,i}}, \quad i = 0, ..., N-1, \quad k-0, ..., K-1 \qquad (47)$$

where $P_{k,i}$ is the power of the data symbol and $\mathrm{Var}(\boldsymbol{\zeta}_k)_{i,i}$ is the variance of the estimate of the data symbol. Note that only the diagonal entries of $\mathrm{Var}(\boldsymbol{\zeta}_k)$ are of our interest.

A. SNR analysis for the proposed reduced-complexity algorithm

To rewrite (45) in the form of (46), we
1) substitute (13), (14) into (45),
2) replace $\mathbf{v}_k$ with $\mathbf{v}_k = \overrightarrow{\mathbf{m}}_{\mathbf{v}_k} + \boldsymbol{\omega}_k$ where $\boldsymbol{\omega}_k \sim (\mathbf{0}, \overrightarrow{\mathbf{V}}_{\mathbf{v}_k})$,
3) replace $\overrightarrow{\mathbf{m}}_{a_k}$ with $\overrightarrow{\mathbf{m}}_{a_k} = \mathbf{u}_k + \boldsymbol{\epsilon}_k$ where $\boldsymbol{\epsilon}_k \sim (\mathbf{0}, \overrightarrow{\mathbf{V}}_{a_k})$.

As a result, the estimation error $\boldsymbol{\zeta}_k$ is obtained as

$$\boldsymbol{\zeta}_k = \left[\beta^{-1}\mathbf{H}_k^H\mathbf{H}_k + \lambda_k^{-1}\boldsymbol{\Phi}_k\right]^{-1}\left[\beta^{-1}(\mathbf{I}+\boldsymbol{\Theta}_k)\mathbf{H}_k^H\mathbf{F}_N\mathbf{n}_k \right.$$
$$\left. + \boldsymbol{\Theta}_k\overrightarrow{\mathbf{V}}_{a_k}^{-1}\boldsymbol{\epsilon}_k + \lambda_k^{-1}\mathbf{B}^H\tilde{\mathbf{H}}_k^H\mathbf{F}_G\boldsymbol{\omega}_k\right] \qquad (48)$$

where

$$\boldsymbol{\Theta}_k = \lambda_k^{-1}(\boldsymbol{\Phi}_k - \boldsymbol{\Psi}_k)(\overrightarrow{\mathbf{V}}_{a_k}^{-1} + \beta^{-1}\mathbf{H}_k^H\mathbf{H}_k)^{-1}. \qquad (49)$$

Since $\mathrm{E}(\mathbf{n}_k) = \mathbf{0}_{N\times1}$, $\mathrm{E}(\boldsymbol{\omega}_k) = \mathbf{0}_{G\times1}$ and $\mathrm{E}(\boldsymbol{\epsilon}_k) = \mathbf{0}_{N\times1}$, the error vector $\boldsymbol{\zeta}_k$ has zero means, i.e. $\mathrm{E}(\boldsymbol{\zeta}_k) = \mathbf{0}_{N\times1}$. As shown in Appendix B, we have the diagonal entries of $\mathrm{Var}(\boldsymbol{\zeta}_k)$ for the proposed reduced-complexity algorithm as follows.

$$\mathrm{Var}(\boldsymbol{\zeta}_k)_{diag}^{prop}$$
$$= (\beta^{-1}\mathbf{H}_k^H\mathbf{H}_k + \lambda_k^{-1}\boldsymbol{\Phi}_k)^{-1} + \lambda_k^{-2}(\beta^{-1}\mathbf{H}_k^H\mathbf{H}_k + \lambda_k^{-1}\boldsymbol{\Phi}_k)^{-1}$$
$$\cdot \left[(\boldsymbol{\Phi}_k - \boldsymbol{\Psi}_k)(\overrightarrow{\mathbf{V}}_{a_k}^{-1} + \beta^{-1}\mathbf{H}_k^H\mathbf{H}_k)^{-1}(\boldsymbol{\Phi}_k - \boldsymbol{\Psi}_k)\right]_{diag}$$
$$\cdot (\beta^{-1}\mathbf{H}_k^H\mathbf{H}_k + \lambda_k^{-1}\boldsymbol{\Phi}_k)^{-1} \qquad (50)$$

B. SNR analysis for the conventional algorithm

To rewrite (25) in the form of (46), we substitute (13) into (25) and get

$$\mathbf{m}_k^{ext} = \mathbf{u}_k + \underbrace{\mathbf{H}_k^{-1}\mathbf{F}_N\mathbf{n}_k}_{\boldsymbol{\zeta}_k}. \qquad (51)$$

Since $\mathrm{E}(\mathbf{n}_k) = \mathbf{0}_{N\times1}$, the estimation error term $\boldsymbol{\zeta}_k$ has zero means, i.e. $\mathrm{E}(\boldsymbol{\zeta}_k) = \mathbf{0}_{N\times1}$. The covariance matrix of $\boldsymbol{\zeta}_k$ is

$$\mathrm{Var}(\boldsymbol{\zeta}_k)^{conv} = [\beta^{-1}\mathbf{H}_k^H\mathbf{H}_k]^{-1} \qquad (52)$$
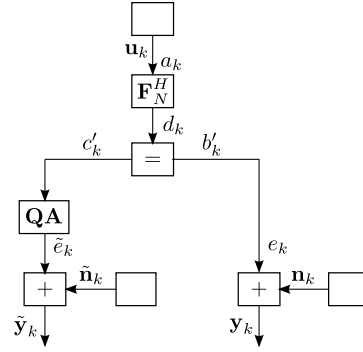
and it is diagonal.

Fig. 4. The kth block of the Forney-style factor graph for the proposed equalizer over AWGN channel.

C. SNR improvement

The SNR improvement of the proposed reduced-complexity algorithm over the conventional algorithm can be evaluated for each iteration by substituting (50) and (52) into

$$\frac{\mathrm{SNR}_{k,i}^{prop}}{\mathrm{SNR}_{k,i}^{conv}} = \frac{P_{k,i}/\mathrm{Var}(\boldsymbol{\zeta}_k)_{i,i}^{prop}}{P_{k,i}/\mathrm{Var}(\boldsymbol{\zeta}_k)_{i,i}^{conv}} = \frac{\mathrm{Var}(\boldsymbol{\zeta}_k)_{i,i}^{conv}}{\mathrm{Var}(\boldsymbol{\zeta}_k)_{i,i}^{prop}}. \qquad (53)$$

As with iterations the *a priori* covariance matrix $\overrightarrow{\mathbf{V}}_{a_k}$ gets very small in the high SNR regime, (50) becomes

$$\mathrm{Var}(\boldsymbol{\zeta}_k)_{diag}^{prop} = (\beta^{-1}\mathbf{H}_k^H\mathbf{H}_k + \lambda_k^{-1}\boldsymbol{\Phi}_k)^{-1}. \qquad (54)$$

and the SNR gain for the ith sub-carrier in the kth block is given by

$$\frac{\mathrm{SNR}_{k,i}^{prop}}{\mathrm{SNR}_{k,i}^{conv}} = \frac{\mu_{k,i} + \tau_{k,i}}{\mu_{k,i}} = 1 + \frac{\tau_{k,i}}{\mu_{k,i}} \qquad (55)$$

where $\mu_{k,i} = \beta^{-1}\left|(\mathbf{H}_k)_{i,i}\right|^2$, $\tau_{k,i} = \lambda_k^{-1}(\boldsymbol{\Phi}_k)_{i,i}$, for $i = 0, ..., N-1$, $k = 0, ..., K-1$.

1) SNR improvement over AWGN channel: For the AWGN channel, we have $\mathbf{H}_k = \mathbf{I}_N$, $\tilde{\mathbf{H}}_k = \mathbf{I}_G$, and $\lambda_k = \beta = 2\sigma^2$, and the matrix $\boldsymbol{\Psi}_k$ is reduced to $\mathbf{F}_N\mathbf{A}^H\mathbf{A}\mathbf{F}_N^H$ with its diagonal entries all equal to the CP ratio $r_{CP} = P/N$. The proposed reduced-complexity algorithm improves the SNR of the ith sub-carrier in the kth block by

$$\frac{\mathrm{SNR}_{k,i}^{prop}}{\mathrm{SNR}_{k,i}^{conv}} = 1 + r_{CP} \qquad (56)$$

for $i = 0, ..., N-1$ and $k = 0, ..., K-1$. For example, when $r_{CP} = 1/4$, we have $\mathrm{SNR}_{k,i}^{prop}/\mathrm{SNR}_{k,i}^{conv} = 1+1/4 = 0.97$dB.

Note that the proposed algorithm enhances the SNRs of all the sub-carriers equally by $10log_{10}(1+r_{CP})$ dB, and thus the overall system performance gain is also $10log_{10}(1+r_{CP})$ dB.

Note also that $10log_{10}(1 + r_{CP})$ dB is the upper bound of the gain that the r_{CP}-ratio CP can offer over the AWGN channel, and iterations are essential for achieving such a bound. Although there is no ISI between data blocks for the AWGN channel, the use of the CP observation introduces mutual dependence between the symbols within a data block. From (13) and (14) with $\mathbf{H}_k = \mathbf{I}_N$ and $\tilde{\mathbf{H}}_k = \mathbf{I}_G$ over the AWGN channel, we can draw the factor graph for the proposed equalizer as Fig. 4. Since the covariance matrix $\overrightarrow{\mathbf{V}}_{a_k}$
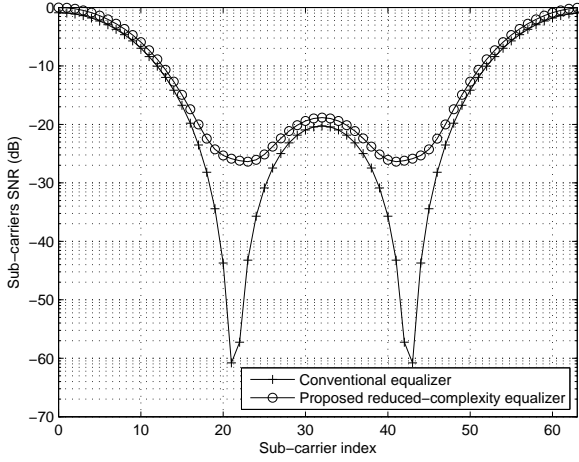
Fig. 5. Sub-carrier SNR improvement over Proakis C channel. Sub-carrier SNRs are normalized by the maximum sub-carrier SNR of both equalizers.
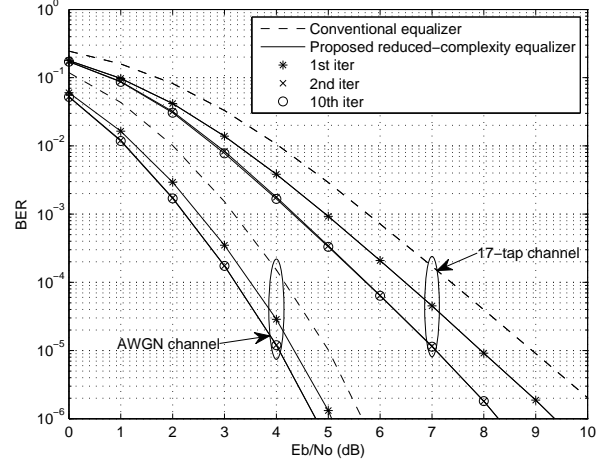

Fig. 6. BER performance of the conventional equalizer and the proposed reduced-complexity equalizer over the AWGN channel and the random 17-tap channel.

is diagonal (see Section III.A), the *a priori* message of $\mathbf{u}_k$ indicates that the symbols within a block are independent of each other. However, the backward message of $\mathbf{u}_k$ indicates some dependence between the symbols as its associated weight matrix is not diagonal. Details are as follows.

1) From (17), the backward message at point b'_k has covariance matrix $\overleftarrow{\mathbf{V}}_{b'_k} = \beta \mathbf{I}_N$.
2) Using (30) with $\overrightarrow{\mathbf{m}}_{\mathbf{v}_k} = \mathbf{0}$ and rule (III.5) in [32], the backward message at point c'_k has weight matrix $\overleftarrow{\mathbf{W}}_{c'_k} = \beta^{-1}\mathbf{A}^H\mathbf{A}$, and it is a diagonal matrix but not a scaled identity matrix.
3) Using rule (II.1) in [32], the backward message at point d_k has weight matrix $\overleftarrow{\mathbf{W}}_{d_k} = \beta^{-1}\left(\mathbf{A}^H\mathbf{A} + \mathbf{I}_N\right)$, and it is also a diagonal matrix but not a scaled identity matrix.
4) Using rule (III.5) in [32], the backward message at point a_k has weight matrix $\overleftarrow{\mathbf{W}}_{a_k} = \beta^{-1}\left(\mathbf{F}_N\mathbf{A}^H\mathbf{A}\mathbf{F}_N^H + \mathbf{I}_N\right)$, and it is a full matrix (specifically, a Toeplitz matrix).

The non-zero off-diagonal entries of $\overleftarrow{\mathbf{W}}_{a_k}$ show that there is mutual dependence between the symbols within a data block. As a result, the extrinsic information that the equalizer produces is relevant to the *a priori* information fed back from the decoder, and that leads to performance improvement of the turbo equalization system during the iteration process.

If the proposed reduced-complexity algorithm is employed (i.e. with approximation (42)), the extrinsic covariance matrix $\mathbf{V}_k^{ext}$ does not contain any term related to the *a priori* message of $\mathbf{u}_k$ anymore (see C2.a in Table I), but the extrinsic mean vector $\mathbf{m}_k^{ext}$ still does (see C2.b in Table I). As will be shown in Fig. 6, with $r_{CP} = 1/4$, the proposed reduced-complexity equalizer achieves only 0.7dB gain over the conventional equalizer after the first iteration, and two iterations are needed to reach the bound of gain 0.97dB.

Note that if the IFFT node (i.e. the $\mathbf{F}_N^H$ node) is removed, Fig. 4 becomes the factor graph of the SC-FDE system [31] (where the CP is also being exploited). In that case, there is no performance improvement with iteration for the AWGN channel since the weight matrix $\overleftarrow{\mathbf{W}}_{a_k} = \beta^{-1}\left(\mathbf{A}^H\mathbf{A} + \mathbf{I}_N\right)$

of the backward message of $\mathbf{u}_k$ is a diagonal matrix and no mutual dependence is introduced.

2) SNR improvement over ISI channels: For ISI channels, we use Proakis C channel [37] as an example to show how the SNR is improved by the proposed algorithm. This channel has 5 taps in the time domain, $\mathbf{h} = [0.227, 0.46, 0.688, 0.46, 0.227]^T$, and in the frequency domain, it has two deep fades as we can see from Fig. 5. In conventional systems, it is hard to recover data from sub-carriers in these fades. However, by adding another branch in the FFG (the left branch in Fig. 3), the proposed algorithm can obtain extra information from the CP observation, which enhances the SNRs of the sub-carriers in the fades by as much as 30dB (see Fig. 5). Other sub-carriers' SNRs are also improved at the same time but only marginally. The unevenness of the SNR gains over different sub-carriers makes it hard to predict the overall system performance gain under ISI channels. However, as an example we will simulate a convolutional coded OFDM system over the Proakis C channel and show the numerical results in the next section.

V. SIMULATION RESULTS

The simulation settings are as follows. A rate-1/2 convolutional code $(23, 35)_8$, a S-random interleaver, a 4QAM modulator with Gray mapping and the BCJR-based decoder [38] are employed. Each data sequence consists of 64 blocks, i.e. $K = 64$, and each data block has 64 symbols before CP insertion, i.e. $N = 64$. With 1/4 ratio, the CP length is 16, i.e. $P = 16$. For each E_b/N_0 point, at least 10^7 sequences are simulated to get an average BER.

There are three types of channels simulated, the AWGN channel, the Proakis C channel, and a random 17-tap channel. For the former two, the channel is fixed and every block experiences the same channel. For the third type, 17 complex coefficients are generated independently as the channel taps for each block, i.e. $L = 17$, and these coefficients vary from block to block. For all the three types, the energy of the channel is
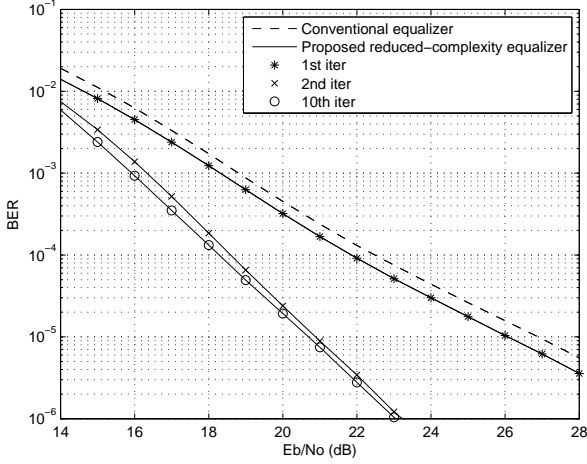
Fig. 7. BER performance of the conventional equalizer and the proposed reduced-complexity equalizer over Proakis C channel.

Fig. 8. BER performance of the proposed reduced-complexity algorithm with different qualities of CSI over the random 17-tap channel.

normalized to 1, i.e. $\frac{1}{L}\sum_{l=0}^{L-1}\left|h_k^l\right|^2 = 1$, for $k = 0, ..., K-1$. Moreover, uniform power delay profile applies for the random 17-tap channel.

A. Performance with perfect CSI

In Fig. 6, the BER performance of the proposed reduced-complexity algorithm is compared with that of the conventional algorithm (i.e. without using the CP) under the AWGN channel and the random 17-tap channel. As the performance of the conventional algorithm does not improve with iterations, only the BER of its first iteration is plotted. For the proposed algorithm, we can see that the BER decreases dramatically within the first two iterations, but after that the improvement is very small. Compared with the conventional algorithm at the BER of 10^{-5}, the proposed algorithm achieves 0.97dB gain under AWGN channel and 2dB gain under the random 17-tap channel after convergence. Note that as discussed in Section IV.C.1, 0.97dB is the upper bound of the SNR gain that the 1/4-ratio CP can offer under the AWGN channel.

Fig. 7 shows the improvement that the proposed reduced-complexity algorithm provides under the Proakis C channel. This channel has two deep fades in the frequency domain as shown in Fig. 5. By exploiting the extra information in the CP, SNRs of the sub-carriers in the deep fades are enormously enhanced, and after two iterations the proposed algorithm obtains 6dB gain at the BER of 10^{-5} compared with the conventional algorithm.

B. Performance with channel estimation error

This set of simulations are to investigate the sensitivity of the proposed algorithm to the quality of the available CSI. Three scenarios, with severe, moderate and mild channel estimation error for the random 17-tap channel, are simulated, and their performance is compared with that of the perfect CSI scenario in Fig. 8.

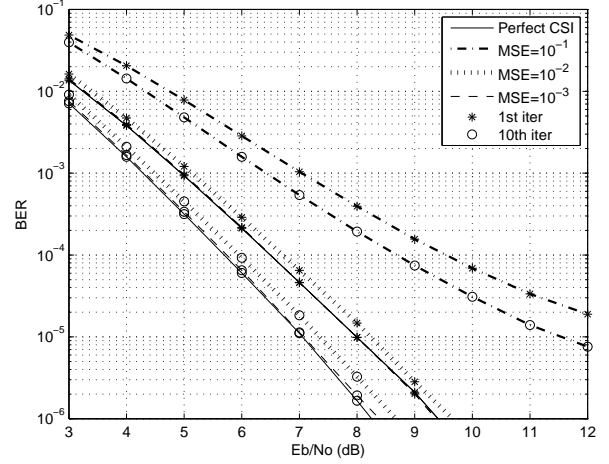As in [39], the estimate for the lth tap of the channel is modeled as $\hat{h}^l = h^l + n^l$, $l = 0, ..., L-1$, where h^l denotes the true channel tap coefficient, $\hat{h}^l$ the estimate and n^l the estimation error. The error n^l is assumed to be complex Gaussian with zero-mean and σ_l^2 variance, and statistically independent for different l's. Since uniform power delay profile is applied in the 17-tap channel, the power of the error for different taps is assumed to be equal, i.e. $\sigma_l^2 = \sigma^2$, $l = 0, ..., L-1$. The mean-square error (MSE) of the channel estimation is defined as $\sum_{l=0}^{L-1}|\hat{h}^l - h^l|^2$, and therefore in our case it is equal to $L\sigma^2$. Three different MSE values, 10^{-1}, 10^{-2} and 10^{-3}, are simulated, respectively, as the severe, moderate and mild channel estimation error scenarios.

As shown in Fig. 8, at the BER of 10^{-5}, there is a performance loss of around 4.5dB for the severe channel estimation error scenario. When the MSE reduces to 10^{-2}, the loss is reduced to within 0.5dB, and for the mild channel estimation error scenario, we can hardly see any performance loss.

C. Performance comparison with [15], [20] and [21]

In this set of simulations, we apply the methods in [15], [20] and [21] in turbo equalization systems and compare their performance with our proposed reduced-complexity algorithm.

The work [15] uses the 'clear' part of the CP observation, and under the settings of this paper, i.e. $P = 16$ and $L = 17$, there is no 'clear' symbols left to be exploited, therefore, the scheme in [15] is reduced to the conventional scheme. As shown in Fig. 9, over the random 17-tap channel, the MMSE solution from [15] performs the same as the conventional system (see Fig. 6), and its least square (LS) solution leads to an error floor due to the error propagation.

The works [20] and [21] make use of the first P symbols of the vector $\tilde{\mathbf{y}}_k$ after interference cancellation, while our proposed algorithm also utilizes the last $L-1$ symbols of $\tilde{\mathbf{y}}_k$ after interference cancellation, therefore exploiting more information about the CP part of data. Fig. 9 compares the performance of both the LS and MMSE methods in [20] and [21] with that of our proposed reduced-complexity algorithm,
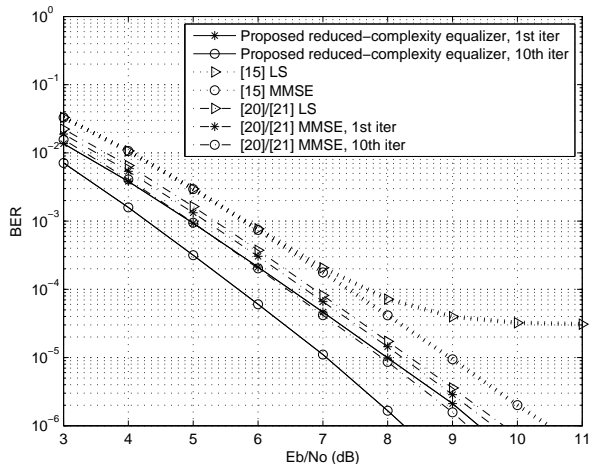
Fig. 9. Performance comparison between the proposed reduced-complexity algorithm and some previous works over the random 17-tap channel.

Fig. 10. Performance comparison between the proposed reduced-complexity algorithm and the *PerfIC-NoApprox* case over the random 17-tap channel.

where perfect interference cancellation is carried out using the true data for [20] and [21], but not for our proposed algorithm. As we can see, over the random 17-tap channel, the LS method in [20] and [21] has similar performance as the first iteration of their MMSE method. With iterations, the MMSE method in [20] and [21] improves, but its convergence performance can only reach that of the first iteration of our proposed algorithm, and there is a gap of around 1dB between their convergence performance and ours at the BER of 10^{-5}.

D. Performance comparison with an ideal case

This set of simulations are to evaluate the performance loss due to the use of approximations (32) and (42). To this end, we compare the proposed reduced-complexity algorithm with an ideal case, the *PerfIC-NoApprox* case, where the following two hypotheses are employed.

1) Perfect interference cancellation is performed, i.e. instead of using (31) and (32), $\overrightarrow{\mathbf{m}}_{\mathbf{v}_k}$ and $\overrightarrow{\mathbf{V}}_{\mathbf{v}_k}$ are calculated as follows.

$$\overrightarrow{\mathbf{m}}_{\mathbf{v}_k} = \mathbf{C}\mathbf{F}_G^H\tilde{\mathbf{H}}_k\mathbf{B}\mathbf{u}_{k-1} + \mathbf{C}^H\mathbf{F}_G^H\tilde{\mathbf{H}}_k\mathbf{B}\mathbf{u}_{k+1}, \quad (57)$$

$$\overrightarrow{\mathbf{V}}_{\mathbf{v}_k} = 2\sigma^2\mathbf{I}_G, \quad (58)$$

where the true data $\mathbf{u}_k$ is assumed to be perfectly known. As a result, each block is interference free from its adjacent blocks; there is no cycle in the graph; and there is no approximation to $\overrightarrow{\mathbf{V}}_{\mathbf{v}_k}$.

2) The approximation of (42) is disabled, i.e. the exact algorithm (i.e. Parts A, B and C1 in Table I) is executed.

Fig. 10 shows that the difference between the *PerfIC-NoApprox* case and the proposed reduced-complexity algorithm only exists in the first iteration, and that the convergence performance of these two cases is the same.

VI. CONCLUSION

We have proposed a factor graph approach to exploiting the CP for equalization in OFDM systems. The models of both the
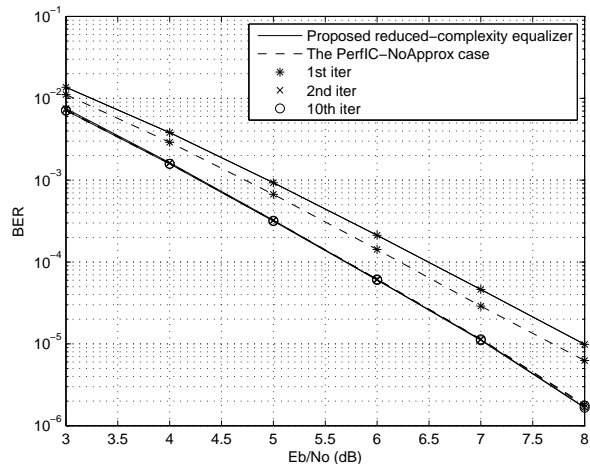
CP part and non-CP part of the received signal are presented in an FFG, and an equalization algorithm is developed based on the computation rules of the FFG and the GMP technique. Proper approximation is made to reduce the overall complexity and SNR analysis is carried out to justify the improvement. Simulation results show that significant gains are achieved by using the proposed algorithm and the turbo equalization system converges within two iterations.

APPENDIX A
A FAST APPROACH TO CALCULATING $\mathbf{\Phi}_k$

From the approximation (42), we have

$$\mathbf{\Phi}_k = \left(\mathbf{B}^H\tilde{\mathbf{H}}_k^H\tilde{\mathbf{H}}_k\mathbf{B}\right)_{diag} \quad (59)$$

with $\mathbf{B} = \mathbf{F}_G\mathbf{Q}\mathbf{A}\mathbf{F}_N^H$. As $\tilde{\mathbf{H}}_k^H\tilde{\mathbf{H}}_k$ is diagonal, the matrix $\mathbf{\Delta}_1 = \mathbf{F}_G^H\tilde{\mathbf{H}}_k^H\tilde{\mathbf{H}}_k\mathbf{F}_G$ is Hermitian circulant and its first column is given by

$$\mathbf{b} = \frac{1}{\sqrt{G}}\mathbf{F}_G^H\mathbf{a} \quad (60)$$

where the vector $\mathbf{a}$ contains all the diagonal entries of $\tilde{\mathbf{H}}_k^H\tilde{\mathbf{H}}_k$. Then we define matrix $\mathbf{\Delta}_2 = \mathbf{Q}^H\mathbf{\Delta}_1\mathbf{Q}$, which is an Hermitian Toeplitz matrix

$$\mathbf{\Delta}_2 = \begin{bmatrix} b_0 & b_1^* & \cdots & b_{P-2}^* & b_{P-1}^* \\ b_1 & b_0 & \ddots & b_{P-3}^* & b_{P-2}^* \\ \vdots & b_1 & \cdots & \ddots & \vdots \\ b_{P-2} & \vdots & \ddots & \vdots & b_1^* \\ b_{P-1} & b_{P-2} & \cdots & b_1 & b_0 \end{bmatrix} \quad (61)$$

where the superscript $*$ denotes the conjugate operation and the first column $[b_0, b_1, ..., b_{P-1}]^T$ is the first P entries of vector $\mathbf{b}$. Therefore, the matrix $\mathbf{\Delta}_3 = \mathbf{A}^H\mathbf{\Delta}_2\mathbf{A}$ takes the

following form

$$\mathbf{\Delta}_3 = \begin{bmatrix} \mathbf{0}_{(N-P)\times(N-P)} & \mathbf{0}_{(N-P)\times P} \\ \mathbf{0}_{P\times(N-P)} & \mathbf{\Delta}_2 \end{bmatrix}. \tag{62}$$

The jth diagonal entry of $\mathbf{\Phi}_k = \left(\mathbf{F}_N\mathbf{\Delta}_3\mathbf{F}_N^H\right)_{diag}$ is calculated as

$$\mathbf{f}_j^H\mathbf{\Delta}_3\mathbf{f}_j \tag{63}$$
$$=\frac{1}{N}\Big[Pb_0 + (P-1)b_1 e^{-2\pi i\frac{j}{N}} + (P-2)b_2 e^{-2\pi i\frac{2j}{N}}$$
$$+ ... + b_{P-1}e^{-2\pi i\frac{(P-1)j}{N}} + (P-1)b_1^* e^{2\pi i\frac{j}{N}}$$
$$+ (P-2)b_2^* e^{2\pi i\frac{2j}{N}} + ... + b_{P-1}^* e^{2\pi i\frac{(P-1)j}{N}}\Big]$$
$$=\frac{1}{N}\sum_{n=0}^{P-1}(P-n)b_n e^{-2\pi i\frac{nj}{N}} + \frac{1}{N}\sum_{n=1}^{P-1}(P-n)b_n^* e^{2\pi i\frac{nj}{N}}$$

where $\mathbf{f}_j$ is the jth column of $\mathbf{F}_N^H$, $j = 0,...,N-1$. The first term in (63) can be obtained by taking the FFT of the sequence $\boldsymbol{\eta}$ and the second term is just the conjugate of the first term, and

$$\boldsymbol{\eta} = \begin{cases} (P-n)b_n, & n=0,...,P-1 \\ 0, & n=P,...,N-1 \end{cases} \tag{64}$$

Thus the diagonal entries of $\mathbf{\Phi}_k$ can be calculated as

$$Diag\left(\mathbf{\Phi}_k\right) = \frac{1}{\sqrt{N}}\mathbf{F}_N\boldsymbol{\eta} + \left(\frac{1}{\sqrt{N}}\mathbf{F}_N\boldsymbol{\eta}\right)^* - \frac{P}{N}b_0^*\mathbf{I}_{N\times1}. \tag{65}$$

The proposed approach of calculating $\mathbf{\Phi}_k$ only needs one G-point FFT for (60) and one N-point FFT for (65).

APPENDIX B
THE DERIVATION OF $\mathrm{Var}(\boldsymbol{\zeta}_k)_{diag}^{prop}$

Using (48) and $\mathrm{E}(\boldsymbol{\zeta}_k) = \mathbf{0}$, the covariance matrix of $\boldsymbol{\zeta}_k$ is given by

$$\mathrm{Var}(\boldsymbol{\zeta}_k) = \mathrm{E}(\boldsymbol{\zeta}_k\boldsymbol{\zeta}_k^H)$$
$$= \mathbf{V}_k^{ext}\Big[\beta^{-1}\left(\mathbf{I}+\mathbf{\Theta}_k\right)\mathbf{H}_k^H\mathbf{H}_k\left(\mathbf{I}+\mathbf{\Theta}_k\right)^H$$
$$+ \mathbf{\Theta}_k(\overrightarrow{\mathbf{V}}_{a_k}^{-1})^H\mathbf{\Theta}_k^H + \lambda_k^{-1}\mathbf{\Psi}_k\Big]\left(\mathbf{V}_k^{ext}\right)^H \tag{66}$$

where the variance matrices $\mathrm{Var}(\mathbf{n}_k) = \beta\mathbf{I}_N$, $\mathrm{Var}(\boldsymbol{\epsilon}_k) = \overrightarrow{\mathbf{V}}_{a_k}$ and $\mathrm{Var}(\boldsymbol{\omega}_k) = \lambda_k\mathbf{I}_G$ are used, and the cross-covariance matrices between any two of $\mathbf{n}_k$, $\boldsymbol{\epsilon}_k$ and $\boldsymbol{\omega}_k$ are treated as zero since they are independent of each other. Meantime, we have $\mathbf{\Psi}_k = \mathbf{\Psi}_k^H$, $\mathbf{\Phi}_k = \mathbf{\Phi}_k^H$, and $\left(\mathbf{V}_k^{ext}\right)^H = \mathbf{V}_k^{ext}$ and $\left(\overrightarrow{\mathbf{V}}_{a_k}^{-1}\right)^H = \overrightarrow{\mathbf{V}}_{a_k}^{-1}$. After simplification, (66) can be rewritten as

$$\mathrm{Var}(\boldsymbol{\zeta}_k) = \mathbf{V}_k^{ext} + \mathbf{V}_k^{ext}\Big[\lambda_k^{-1}\mathbf{\Theta}_k(\mathbf{\Phi}_k-\mathbf{\Psi}_k) + \beta^{-1}\mathbf{\Theta}_k\mathbf{H}_k^H\mathbf{H}_k$$
$$- \overrightarrow{\mathbf{V}}_{a_k}^{-1}\mathbf{\Theta}_k^H\Big]\mathbf{V}_k^{ext}. \tag{67}$$

Since $\mathbf{V}_k^{ext}$ is diagonal and $\left(\mathbf{\Theta}_k\right)_{diag} = \mathbf{0}$ ($\mathbf{\Phi}_k$ and $\mathbf{\Psi}_k$ have the same diagonal entries), the two terms $\beta^{-1}\mathbf{\Theta}_k\mathbf{H}_k^H\mathbf{H}_k$ and $\overrightarrow{\mathbf{V}}_{a_k}^{-1}\mathbf{\Theta}_k^H$ have no contribution to the diagonal entries of

$\mathrm{Var}(\boldsymbol{\zeta}_k)$ at all. After removing them, the following diagonal matrix is obtained.

$$\mathrm{Var}(\boldsymbol{\zeta}_k)_{diag} = \mathbf{V}_k^{ext} + \lambda_k^{-1}\mathbf{V}_k^{ext}\Big[\mathbf{\Theta}_k\left(\mathbf{\Phi}_k-\mathbf{\Psi}_k\right)\Big]_{diag}\mathbf{V}_k^{ext}. \tag{68}$$

Substituting (44) into (68), we have the diagonal entries of $\mathrm{Var}(\boldsymbol{\zeta}_k)$ for the proposed algorithm as (50).

REFERENCES

[1] H. Rohling, *OFDM: Concepts for Future Communication Systems.* Springer, 2011.
[2] *Radio broadcasting systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers*, European Telecommunication Standard ETS 300 401, Feb. 1995.
[3] U. Reimers, "DVB-T: the COFDM-based system for terrestrial television," *Electronics & Communication Engineering Journal*, vol. 9, no. 1, pp. 28–32, 1997.
[4] A. Doufexi, S. Armour, M. Butler, A. Nix, D. Bull, J. McGeehan, and P. Karlsson, "A comparison of the HIPERLAN/2 and IEEE 802.11a wireless LAN standards," *IEEE Commun. Mag.*, vol. 40, no. 5, pp. 172–180, 2002.
[5] *Multimedia mobile access communication systems (MMAC)*, Std. [Online]. Available: http://www.arib.or.jp/mmac/e/index.htm
[6] J. Khun-Jush, P. Schramm, U. Wachsmann, and F. Wenger, "Structure and performance of the HIPERLAN/2 physical layer," in *IEEE Vehicular Technology Conference (VTC), 1999*, vol. 5, pp. 2667–2671.
[7] M. de Courville, P. Duhamel, P. Madec, and J. Palicot, "Blind equalization of OFDM systems based on the minimization of a quadratic criterion," in *IEEE International Conference on Communications (ICC), 1996*, vol. 3, pp. 1318–1322.
[8] J. Van de Beek, M. Sandell, and P. Borjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Processing*, vol. 45, no. 7, pp. 1800–1805, 1997.
[9] M. Morelli and U. Mengali, "An improved frequency offset estimator for OFDM applications," in *Communication Theory Mini-Conference, 1999*, pp. 106–109.
[10] U. Tureli, H. Liu, and M. Zoltowski, "OFDM blind carrier offset estimation: ESPRIT," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1459–1461, 2000.
[11] X. Wang and K. Liu, "Performance analysis for adaptive channel estimation exploiting cyclic prefix in multicarrier modulation systems," *IEEE Trans. Commun.*, vol. 51, no. 1, pp. 94–105, 2003.
[12] G. Al-Rawi, T. Al-Naffouri, A. Bahai, and J. Cioffi, "Exploiting error-control coding and cyclic-prefix in channel estimation for coded OFDM systems," *IEEE Commun. Lett.*, vol. 7, no. 8, pp. 388–390, 2003.
[13] B. Devillers, J. Louveaux, and L. Vandendorpe, "Exploiting cyclic prefix for performance improvement in single carrier systems," in *IEEE 7th Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2006*, pp. 1–5.
[14] G. Bottomley and L. Wilhelmsson, "Recovering signal energy from the cyclic prefix in OFDM," *IEEE Trans. Veh. Technol.*, vol. 57, no. 5, pp. 3205–3211, 2008.
[15] A. Tarighat and A. Sayed, "An optimum OFDM receiver exploiting cyclic prefix for improved data estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2003*, vol. 4, pp. IV–217.
[16] L. Vangelista, M. Rotoloni, and A. Morello, "Improved data detection exploiting full cyclic prefix for the evolution of DVB-T," in *IEEE International Wireless Communications and Mobile Computing Conference (IWCMC), 2008*, pp. 1070–1074.
[17] K. Hueske and J. Götze, "Improving OFDM data estimation by overlapping based cyclic prefix reuse," in *14th International OFDM Workshop (InOWo), 2009*, Hamburg, Germany.
[18] G. Garbo and S. Mangione, "An OFDM receiver exploiting multipath diversity," in *IEEE Mosharaka International Conference on Communications, Computers and Applications (MIC-CCA), 2008*, pp. 65–70.
[19] G. Garbo and S. Mangione, "An improved detection technique for cyclic-prefixed OFDM," *Journal of Networks*, vol. 5, no. 7, pp. 759–765, 2010.
[20] G. Garbo and S. Mangione, "An improved receiver architecture for cyclic-prefixed OFDM," in *IEEE Wireless Communications and Networking Conference (WCNC), 2009*, pp. 1–6.

[21] A. Quadeer, "Enhanced equalization in OFDM systems using cyclic prefix," in *IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS), 2010*, pp. 40–44.

[22] T. Palenik and P. Farkas, "Exploiting redundancy in an OFDM SDR receiver," *International Journal of Digital Multimedia Broadcasting*, vol. 2009, 2009.

[23] T. Palenik and P. Farkas, "Exploiting cyclic prefix redundancy in OFDM to improve performance of Tanner: graph based decoding," *Analog Integrated Circuits and Signal Processing*, vol. 69, no. 2, pp. 143–152, 2011.

[24] A. Glavieux, C. Laot, and J. Labat, "Turbo equalization over a frequency selective channel," in *Proc. Int. Symp. Turbo Codes*, 1997, pp. 96–102.

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

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

[27] G. Al-Rawi, T. Al-Naffouri, A. Bahai, and J. Cioffi, "An iterative receiver for coded OFDM systems over time-varying wireless channels," in *IEEE International Conference on Communications (ICC), 2003*, vol. 5, pp. 3371–3376.

[28] T. Al-Naffouri, "An EM-based forward-backward Kalman filter for the estimation of time-variant channels in OFDM," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3924–3930, 2007.

[29] T. Al-Naffouri and A. Quadeer, "Cyclic prefix based enhanced data recovery in OFDM," *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3406–3410, 2010.

[30] T. Al-Naffouri and A. Quadeer, "Blind maximum-likelihood data recovery in OFDM," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2008*, pp. 2829–2832.

[31] J. Yang, Q. Guo, D. Huang, and S. Nordholm, "Exploiting cyclic prefix in Turbo FDE systems using factor graph," in *IEEE Wireless Communications and Networking Conference (WCNC), 2013*.

[32] H. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. Kschischang, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, 2007.

[33] T.-D. Chiueh, P.-Y. Tsai, and I.-W. Lai, *Baseband Receiver Design for Wireless MIMO-OFDM Communications*, 2nd ed.   Wiley, 2012.

[34] A. Hero and T. Marzetta, "Cutoff rate and signal design for the quasi-static rayleigh-fading space-time channel," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2400–2416, 2001.

[35] Q. Guo and D. Huang, "A concise representation for the soft-in soft-out LMMSE detector," *IEEE Commun. Lett.*, vol. 15, no. 99, pp. 1–3, 2011.

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

[37] J. G. Proakis, *Digital Communications*, 4th ed.   New York: McGraw-Hill, 2001.

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

[39] V. Tarokh, A. Naguib, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 199–207, 1999.