

# A Low-Complexity Iterative Channel Estimation and Detection Technique for Doubly Selective Channels

Qinghua Guo, *Member, IEEE*, Li Ping, *Senior Member, IEEE*, and Defeng (David) Huang, *Senior Member, IEEE*

**Abstract**—In this paper, we propose a low-complexity iterative joint channel estimation, detection and decoding technique for doubly selective channels. The key to the proposed technique is a segment-by-segment processing strategy under the assumption that the channel is approximately static within a short segment of a data block. Through a virtual zero-padding technique, the proposed segment-by-segment equalization approach inherits the low-complexity advantage of the conventional frequency domain equalization (FDE), but does not need the assistance of guard interval (for cyclic-prefixing or zero-padding), thereby avoiding the spectral and power overheads. Furthermore, we develop a low-complexity bidirectional channel estimator, where the Gaussian message passing (GMP) technique is used to exploit the channel correlation information, and the intermediate channel estimation results in the iterative process are employed to perform inter-tap interference cancellation. Simulation results demonstrate the effectiveness of the proposed detection and channel estimation algorithms.

**Index Terms**—Channel estimation, turbo equalization, doubly selective channels, frequency domain equalization (FDE), cyclic-prefixing, zero-padding, CP reconstruction, Gaussian message passing (GMP).

## I. INTRODUCTION

**S**INGLE-CARRIER block transmission with iterative frequency domain equalization (FDE) [1], [26] is a promising technique to alleviate inter-symbol interference (ISI) in frequency selective channels. By using cyclic prefix (CP), inter-block interference is removed and linear convolution is converted to cyclic convolution in FDE, leading to efficient receiver implementation using the fast Fourier transform (FFT).

Manuscript received October 30, 2008; revised February 20, 2009; accepted May 6, 2009. The associate editor coordinating the review of this paper and approving it for publication was L. Lampe.

Q. Guo was with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China. He is now with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Australia (e-mail: qhguo@ee.uwa.edu.au).

L. Ping is with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China (e-mail: eeliping@cityu.edu.hk).

D. Huang is with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Australia (e-mail: huangdf@ee.uwa.edu.au).

Q. Guo's work was supported by a grant from the Research Grant Council of the Hong Kong SAR, China, under project CityU 117007 and Australian Research Council's *Discovery Projects* funding scheme (project number DP0877616). L. Ping's work was fully supported by a grant from the Research Grant Council of the Hong Kong SAR, China, under project CityU 117007. D. Huang's work was supported by Australian Research Council's *Discovery Projects* funding scheme (project number DP0877616). This paper was presented in part at the IEEE Global Communications Conference, New Orleans, LA, USA, November 30-December 4, 2008.

Digital Object Identifier 10.1109/TWC.2009.081448

However, cyclic-prefixing incurs both power and spectral overheads that can be measured by the ratio of the CP length to the data block length. This ratio is limited by two requirements, i.e., the channel should be approximately static within a block (and so the block length is limited by channel coherence time), and the CP length should be larger than the channel memory length. Hence, the overheads can be large in doubly selective channels (i.e., time-varying ISI channels) when the channel coherence time is short and the channel memory length is large.

Using a shorter CP is a way to reduce the overheads. However, when the CP length is less than the channel memory length, interference between adjacent data blocks is induced, and the assumption of the conversion from linear convolution to cyclic convolution is invalid. Remedies based on CP reconstruction for these problems have been studied in [2]-[6]. The basic idea behind them is to compensate the effect of missing CP through the so-called CP reconstruction operation for the concerned data block after cancelling the interference from the preceding data block. However, these approaches can not sufficiently exploit all available observations because, for detection of each data block, the observations corresponding to the tail symbols of the data block are discarded. In addition, CP reconstruction is based on the estimated tail symbols (which may be very unreliable). Hence, performance degradation may be caused when the interference between adjacent data blocks is severe.

In this paper, we consider the issue of efficient detection without the assistance of CP in the context of doubly selective channels. We assume that a conventional single-carrier transmitter is employed that transmits signal continuously (without inserting any guard intervals) in a data block. To develop an efficient detection approach, we partition the transmitted data block  $\mathbf{x}$  with length  $J$  (which is formed from a codeword after symbol mapping) into a number of short segments  $\{\mathbf{x}_k\}$  (note that such partition does not affect the structure of the transmitted signal), each with length  $M \geq L$  (where  $L$  is the channel memory length), as illustrated in Fig. 1(a), and assume that the channels remain approximately static within a segment but not necessarily within a block.<sup>1</sup> The received

<sup>1</sup>In the proposed approach, since each block is formed from a codeword after symbol mapping, it is better to allow independently processing for each block at the receiver to avoid high latency. For this, a string of zeros is appended to each block (see Fig. 1(a)) to avoid the interference between adjacent blocks. Since the length of block  $\mathbf{x}$  is not restricted by the channel coherence time, it can be large enough to make the spectral loss due to zero-appending negligible.

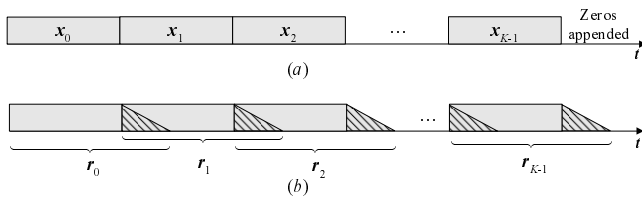


Fig. 1. (a) The transmitted data block  $\mathbf{x}$ , formed from a codeword after symbol mapping, is partitioned into a number of short segments  $\{x_k\}$ . Each segment is assumed to undergo a static channel. (b) The multi-path channel causes interference between adjacent segments at the receiver, as illustrated by the shadowing parts.

signal is shown in Fig. 1(b). Due to the channel delay spread, the observation vector  $\mathbf{r}_k$  corresponding to segment  $x_k$  has a length of  $M + L$ , which covers all the contribution of  $x_k$ . In this paper, we develop an efficient segment-by-segment equalization algorithm based on a cyclic convolution representation of the received signal for each segment through a virtual zero-padding operation. The proposed technique inherits the low-complexity advantage of the conventional FDE technique, but does not need the assistance of CP. Moreover, it overcomes the problems in the CP reconstruction based technique, i.e., it can sufficiently exploit all the received energy (since the detection of  $x_k$  in this technique is based on the entire  $\mathbf{r}_k$  instead of the first  $M$  entries of  $\mathbf{r}_k$  in the CP reconstruction based technique), and avoids the CP reconstruction operation.

Channel estimation is also considered in this paper. The bidirectional trellis based approaches that carry out channel estimation and equalization simultaneously [30]-[32] deliver good performance, but their computational complexities are often intractable. In this paper, we follow the more practical framework of iterative joint channel estimation and detection discussed in [10]-[13], [28], and [29], in which the channel estimator is separated out from the equalizer. In [10]-[13], [28], and [29], various channel estimation approaches such as those based on minimum mean square error (MMSE), least squares (LS), recursive least squares (RLS) and Kalman filtering (KF) have been studied. However, in those works, only a unidirectional recursion is involved in tracking the time-varying channels, and some of them have relatively high complexity. In this paper, we present a bidirectional channel estimation approach, where the recently proposed Gaussian message passing (GMP) technique [15], [16] is used to exploit the channel correlation information, and the intermediate channel estimation results in the iterative process are employed to perform inter-tap interference cancellation. Compared with the unidirectional approaches, the proposed bidirectional estimation approach can exploit the correlation information of time-varying channels more efficiently, which is essential to improve the system performance, as will be demonstrated by the simulation results. The overall complexity of the proposed channel estimation and detection algorithms is  $O(L' + \log_2(M + L))$  per symbol per iteration, where  $L' \leq L + 1$  is the number of nonzero channel taps.

The notations used in this paper are as follows. Lower case letters denote scalars, bold lower case letters denote column vectors, and bold upper case letters denote matrices. The superscript “ $T$ ” denotes transpose, and “ $H$ ” conjugate

transpose. The symbol  $\mathbf{I}$  denotes an identity matrix with proper size. Expectation and (co)variance are represented by  $E(\cdot)$  and  $V(\cdot)$ , respectively. For a complex variable, e.g.,  $x$ , we use  $x^{\text{Re}}$  to denote its real part and  $x^{\text{Im}}$  its imaginary part. In addition, we use  $\text{diag}[\mathbf{a}]$  to denote a diagonal matrix with its diagonal entries given by  $\mathbf{a}$ , and  $(\mathbf{A})_{\text{diag}}$  a diagonal matrix with its diagonal entries given by those of matrix  $\mathbf{A}$ .

## II. PRELIMINARY

In this section, we list the key results of MMSE estimation principle to be used in the following sections, and depict the receiver structure for iterative channel estimation, detection and decoding.

### A. MMSE Estimation for Gaussian Variables

Consider a standard estimation problem based on the following linear model

$$\mathbf{r} = \mathbf{A}\mathbf{h} + \mathbf{n} \quad (1)$$

where  $\mathbf{r}$  is an observation vector,  $\mathbf{A}$  a system transfer matrix,  $\mathbf{h}$  a vector to be estimated, and  $\mathbf{n}$  a vector of Gaussian noise. We assume that the transfer matrix  $\mathbf{A}$  and the statistics of Gaussian noise  $\mathbf{n}$  are known, and that  $\mathbf{h}$  is Gaussian distributed with *a priori* mean vector and covariance matrix given by  $E(\mathbf{h})$  and  $V(\mathbf{h})$ , respectively. Then the *a posteriori* mean and variance of  $\mathbf{h}$  can be computed as [14]

$$E^p(\mathbf{h}) = E(\mathbf{h}) + (V(\mathbf{h})^{-1} + \mathbf{A}^H V(\mathbf{n})^{-1} \mathbf{A})^{-1} \mathbf{A}^H V(\mathbf{n})^{-1} (\mathbf{r} - \mathbf{A}E(\mathbf{h}) - E(\mathbf{n})) \quad (2a)$$

and

$$V^p(\mathbf{h}) = (V(\mathbf{h})^{-1} + \mathbf{A}^H V(\mathbf{n})^{-1} \mathbf{A})^{-1} \quad (2b)$$

where the superscript “ $p$ ” denotes “*a posteriori*”.

### B. Joint Gaussian (Linear MMSE) Estimation for Binary Variables

The estimation becomes complicated for the following linear model

$$\mathbf{r} = \mathbf{A}\mathbf{x} + \mathbf{n} \quad (3)$$

where the entries of  $\mathbf{x}$  are binary and independent of each other, and the system transfer matrix  $\mathbf{A}$  and the statistics of Gaussian noise  $\mathbf{n}$  are known.

We first assume that all the variables involved in (3) are real. In this case, the estimate is usually given in an entry-by-entry extrinsic logarithm of likelihood ratio (LLR) form as [7]-[9]

$$e(x_j) = \ln \frac{p(\mathbf{r} | x_j = +1)}{p(\mathbf{r} | x_j = -1)} \quad (4)$$

where  $x_j$  is the  $j$ th entry of  $\mathbf{x}$ . The optimal approach to calculating (4) is based on the maximum *a posteriori* probability (MAP) criterion at the cost of high complexity. A low-complexity sub-optimal alternative is the so-called joint Gaussian (JG) approach [19], in which, by focusing on one symbol  $x_j$ , (3) is rewritten in the following signal-plus-distortion form

$$\mathbf{r} = \mathbf{a}_j x_j + \boldsymbol{\xi}_j \quad (5)$$

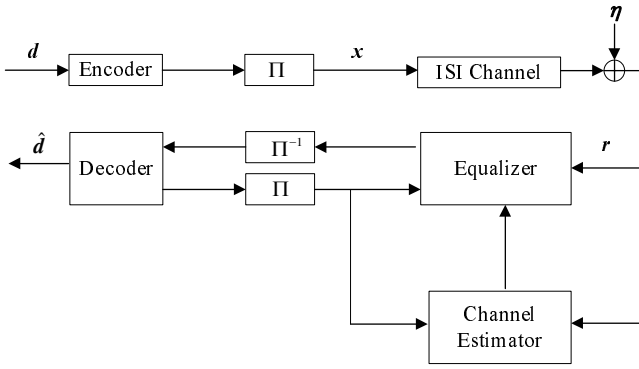


Fig. 2. The transmitter and turbo receiver.  $\Pi$  and  $\Pi^{-1}$  denote interleaver and de-interleaver, respectively.

where  $\mathbf{a}_j$  is the  $j$ th column of  $\mathbf{A}$  and

$$\boldsymbol{\xi}_j = \sum_{j' \neq j} \mathbf{a}_{j'} x_{j'} + \mathbf{n}. \quad (6)$$

By the central limit theorem, the entries of  $\boldsymbol{\xi}_j$  can be approximated as joint Gaussian variables with

$$\mathbf{E}(\boldsymbol{\xi}_j) = \mathbf{E}(\mathbf{r}) - \mathbf{a}_j \mathbf{E}(x_j) \quad (7a)$$

and

$$\mathbf{V}(\boldsymbol{\xi}_j) = \mathbf{V}(\mathbf{r}) - \mathbf{V}(x_j) \mathbf{a}_j \mathbf{a}_j^T \quad (7b)$$

where

$$\mathbf{E}(\mathbf{r}) = \mathbf{A} \mathbf{E}(\mathbf{x}) + \mathbf{E}(\mathbf{n}) \quad (7c)$$

and

$$\mathbf{V}(\mathbf{r}) = \mathbf{A} \text{diag}[\mathbf{V}(x_0), \mathbf{V}(x_1), \dots] \mathbf{A}^T + \mathbf{V}(\mathbf{n}). \quad (7d)$$

Based on the above assumption, (4) can be computed as

$$e(x_j) = \ln \frac{\exp\left[-\frac{1}{2}(\mathbf{r} - \mathbf{a}_j - \mathbf{E}(\boldsymbol{\xi}_j))^T \mathbf{V}(\boldsymbol{\xi}_j)^{-1} (\mathbf{r} - \mathbf{a}_j - \mathbf{E}(\boldsymbol{\xi}_j))\right]}{\exp\left[-\frac{1}{2}(\mathbf{r} + \mathbf{a}_j - \mathbf{E}(\boldsymbol{\xi}_j))^T \mathbf{V}(\boldsymbol{\xi}_j)^{-1} (\mathbf{r} + \mathbf{a}_j - \mathbf{E}(\boldsymbol{\xi}_j))\right]} \quad (8a)$$

$$= 2 \mathbf{a}_j^T \mathbf{V}(\boldsymbol{\xi}_j)^{-1} (\mathbf{r} - \mathbf{E}(\mathbf{r}) + \mathbf{a}_j \mathbf{E}(x_j)) \quad (8a)$$

$$= 2 \frac{\mathbf{a}_j^T \mathbf{V}(\mathbf{r})^{-1} (\mathbf{r} - \mathbf{E}(\mathbf{r}) + \mathbf{a}_j \mathbf{E}(x_j))}{1 - \mathbf{V}(x_j) \mathbf{a}_j^T \mathbf{V}(\mathbf{r})^{-1} \mathbf{a}_j} \quad (8b)$$

$$= 2 \frac{\mathbf{a}_j^T \mathbf{V}(\mathbf{r})^{-1} (\mathbf{r} - \mathbf{A} \mathbf{E}(\mathbf{x}) - \mathbf{E}(\mathbf{n})) + \mathbf{a}_j^T \mathbf{V}(\mathbf{r})^{-1} \mathbf{a}_j \mathbf{E}(x_j)}{1 - \mathbf{V}(x_j) \mathbf{a}_j^T \mathbf{V}(\mathbf{r})^{-1} \mathbf{a}_j}. \quad (8c)$$

The derivation from (8a) to (8b) is based on the matrix inversion lemma [23]. The result in (8) is the same as that in the so-called Linear MMSE approach derived in [7]-[9]. However, the derivation described above is more concise and straightforward.

We now extend the above result to complex systems using quadrature phase shift keying (QPSK) modulation with Gray mapping. Denote  $x_j = x_j^{\text{Re}} + ix_j^{\text{Im}}$ , where  $i = \sqrt{-1}$  and  $\{x_j^{\text{Re}}, x_j^{\text{Im}}\}$  are binary, and define

$$e(x_j) = e(x_j^{\text{Re}}) + ie(x_j^{\text{Im}}) \quad (9)$$

where

$$e(x_j^{\text{Re}}) = \ln \frac{p(\mathbf{r} | x_j^{\text{Re}} = +1)}{p(\mathbf{r} | x_j^{\text{Re}} = -1)} \quad (10a)$$

and

$$e(x_j^{\text{Im}}) = \ln \frac{p(\mathbf{r} | x_j^{\text{Im}} = +1)}{p(\mathbf{r} | x_j^{\text{Im}} = -1)}. \quad (10b)$$

Assume that  $\mathbf{V}(x_j^{\text{Re}}, x_j^{\text{Im}}) = 0$ ,  $\mathbf{V}(x_j^{\text{Re}}) = \mathbf{V}(x_j^{\text{Im}})$ ,  $\mathbf{V}(\mathbf{n}^{\text{Re}}, \mathbf{n}^{\text{Im}}) = \mathbf{0}$ , and  $\mathbf{V}(\mathbf{n}^{\text{Re}}) = \mathbf{V}(\mathbf{n}^{\text{Im}})$ . Then  $e(x_j)$  can be calculated as

$$e(x_j) = 4 \frac{\mathbf{a}_j^H \mathbf{V}(\mathbf{r})^{-1} (\mathbf{r} - \mathbf{A} \mathbf{E}(\mathbf{x}) - \mathbf{E}(\mathbf{n})) + \mathbf{a}_j^H \mathbf{V}(\mathbf{r})^{-1} \mathbf{a}_j \mathbf{E}(x_j)}{1 - \mathbf{V}(x_j) \mathbf{a}_j^H \mathbf{V}(\mathbf{r})^{-1} \mathbf{a}_j} \quad (11)$$

The derivation of (11) is shown in Appendix I. Letting  $\mathbf{e}(\mathbf{x}) = [e(x_0), e(x_1), \dots]^T$ , we can rewrite (11) in a compact vector form as

$$\mathbf{e}(\mathbf{x}) = 4(\mathbf{I} - \mathbf{V}(\mathbf{x})\mathbf{U})^{-1} (\mathbf{A}^H \mathbf{V}(\mathbf{r})^{-1} (\mathbf{r} - \mathbf{A} \mathbf{E}(\mathbf{x}) - \mathbf{E}(\mathbf{n})) + \mathbf{U} \mathbf{E}(\mathbf{x})) \quad (12)$$

where

$$\mathbf{V}(\mathbf{r}) = \mathbf{A} \mathbf{V}(\mathbf{x}) \mathbf{A}^H + \mathbf{V}(\mathbf{n}), \quad (13a)$$

$$\mathbf{V}(\mathbf{x}) = \text{diag}[\mathbf{V}(x_0), \mathbf{V}(x_1), \dots], \quad (13b)$$

and

$$\mathbf{U} = (\mathbf{A} \mathbf{V}(\mathbf{r})^{-1} \mathbf{A}^H)_{\text{diag}}. \quad (13c)$$

### C. Receiver Structure for Iterative Channel Estimation, Detection and Decoding

We follow the framework of iterative channel estimation and detection [10]-[13], [28], [29], where the channel estimator is separated out from the equalizer. The iterative receiver is shown in the lower part of Fig. 2. It consists of three modules: a channel estimator, an equalizer and a decoder. The three modules work in an iterative manner.

In this paper, we assume standard soft-in soft-out (SISO) decoding algorithms for the decoder (e.g., the BCJR decoding algorithm [20] for a convolutional code), and focus on the design of low-complexity algorithms for the channel estimator and the equalizer, which will be detailed in Sections III and IV, respectively.

## III. SEGMENT-BY-SEGMENT EQUALIZATION WITH VIRTUAL ZERO-PADDING

### A. System Model

Consider the following baseband discrete-time time-varying (complex) ISI channel model

$$r_j = \sum_{l=0}^L h_{j,l} \cdot x_{j-l} + \eta_j \quad (14)$$

where  $\{x_j, j = 0, 1, \dots, J-1\}$  are the transmitted data symbols formed from the interleaved output of a forward error correction (FEC) encoder with Gray mapping and QPSK modulation, whose *a priori* information (in the form of mean and variance) at the receiver can be computed using the feedbacks from the decoder [7]-[9],  $\{r_j\}$  the observations,  $\{\eta_j\}$  the samples of additive white Gaussian noise (AWGN) with zero mean and variance  $2\sigma^2$ , and  $\{h_{j,l}, l = 0, 1, \dots, L\}$  the channel state information (CSI) at time index  $j$ .

To develop an efficient detection algorithm, as mentioned in Introduction, we partition the transmitted signal  $\mathbf{x}$  into  $K$

short segments  $\{\mathbf{x}_k = [x_{kM}, x_{kM+1}, \dots, x_{(k+1)M-1}]^T, k = 0, 1, \dots, K-1\}$  (assuming that  $J = KM$ ), each with length  $M$ , as shown in Fig. 1. We use a static channel to approximate the channel undergone by  $\mathbf{x}_k$ . Due to the delay spread, the observation vector  $\mathbf{r}_k = [r_{kM}, r_{kM+1}, \dots, r_{(k+1)M+L-1}]^T$ , corresponding to  $\mathbf{x}_k$ , has a length of  $N = M + L$ , which can be represented as

$$\mathbf{r}_k = \mathbf{h}_k * \mathbf{x}_k + \mathbf{y}_{k-1}^{tail} + \mathbf{y}_{k+1}^{head} + \boldsymbol{\eta}_k + \boldsymbol{\delta}_k \quad (15)$$

with

$$\mathbf{y}_{k-1}^{tail} = [tail(\mathbf{h}_{k-1} * \mathbf{x}_{k-1}), 0, \dots, 0]^T \quad (16a)$$

and

$$\mathbf{y}_{k+1}^{head} = [0, \dots, 0, head(\mathbf{h}_{k+1} * \mathbf{x}_{k+1})]^T \quad (16b)$$

where “ $*$ ” denotes linear convolution,  $\mathbf{y}_{k-1}^{tail}$  and  $\mathbf{y}_{k+1}^{head}$  denote the interference from the adjacent segments  $\mathbf{x}_{k-1}$  and  $\mathbf{x}_{k+1}$ , respectively,  $\boldsymbol{\eta}_k$  denotes the AWGN vector,  $tail(\cdot)$  and  $head(\cdot)$  in (16) denote two truncation functions that return the tail and head parts (with length  $L$ ) of the sequence in the parentheses, respectively. The error vector  $\boldsymbol{\delta}_k$  in (15) is induced since the channel is actually time-varying within each segment. However, it is negligible when the segments are short enough and the channel variation is not too rapid. With this assumption, we have

$$\mathbf{r}_k = \mathbf{h}_k * \mathbf{x}_k + \mathbf{n}_k \quad (17)$$

where

$$\mathbf{n}_k = \mathbf{y}_{k-1}^{tail} + \mathbf{y}_{k+1}^{head} + \boldsymbol{\eta}_k. \quad (18)$$

### B. Segment-by-Segment Equalization with Virtual Zero-Padding

We first assume perfect CSI  $\mathbf{h}_k$  available at the receiver (and we will discuss the imperfect CSI case in Section V). By appending a proper number of zeros to  $\mathbf{x}_k$  and  $\mathbf{h}_k$ , we define the following two vectors with length  $N = M + L$ :

$$\tilde{\mathbf{h}}_k = [\mathbf{h}_k^T, \underbrace{0, \dots, 0}_{M-1 \text{ replicas}}]^T \quad (19a)$$

and

$$\tilde{\mathbf{x}}_k = [\mathbf{x}_k^T, \underbrace{0, \dots, 0}_{L \text{ replicas}}]^T. \quad (19b)$$

Then (17) can be rewritten as

$$\mathbf{r}_k = \tilde{\mathbf{h}}_k \otimes \tilde{\mathbf{x}}_k + \mathbf{n}_k \quad (20)$$

where “ $\otimes$ ” denotes the cyclic convolution. As a result, the above operation for  $\mathbf{h}_k$  and  $\mathbf{x}_k$  transforms the linear convolution in (17) into the cyclic convolution in (20).

Define  $\mathbf{F}$  as the normalized discrete Fourier transform (DFT) matrix with size  $N \times N$  (i.e., the  $(m, n)$ th element of  $\mathbf{F}$  is given by  $\mathbf{F}(m, n) = N^{-1/2} e^{-i2\pi mn/N}$ ). According to the property of cyclic convolution and (20), we have

$$\sqrt{N}\mathbf{F}\mathbf{r}_k = (\sqrt{N}\mathbf{F}\tilde{\mathbf{h}}_k) \bullet (\sqrt{N}\mathbf{F}\tilde{\mathbf{x}}_k) + \sqrt{N}\mathbf{F}\mathbf{n}_k \quad (21)$$

where “ $\bullet$ ” denotes the element-wise product of two vectors. Denote the DFT of  $\tilde{\mathbf{h}}_k$  as

$$[g_{k,0}, g_{k,1}, \dots, g_{k,N-1}]^T = \sqrt{N}\mathbf{F}\tilde{\mathbf{h}}_k \quad (22)$$

and define the following diagonal matrix

$$\mathbf{G}_k = \text{diag}[g_{k,0}, g_{k,1}, \dots, g_{k,N-1}]. \quad (23)$$

Then (21) can be rewritten in a matrix form as

$$\mathbf{r}_k = \underbrace{\mathbf{F}^H \mathbf{G}_k \mathbf{F}}_{\mathbf{A}_k} \tilde{\mathbf{x}}_k + \mathbf{n}_k \quad (24)$$

where the mean vector and covariance matrix of  $\mathbf{n}_k$  are given by

$$\mathbf{E}(\mathbf{n}_k) = \mathbf{E}(\mathbf{y}_{k-1}^{tail}) + \mathbf{E}(\mathbf{y}_{k+1}^{head}) \quad (25a)$$

and

$$\mathbf{V}(\mathbf{n}_k) = \mathbf{V}(\mathbf{y}_{k-1}^{tail}) + \mathbf{V}(\mathbf{y}_{k+1}^{head}) + 2\sigma^2\mathbf{I}. \quad (25b)$$

We adopt the following two approximations which lead to considerable complexity reduction:<sup>2</sup>

$$\mathbf{V}(\tilde{\mathbf{x}}_k) \approx v\mathbf{I} \quad (26a)$$

and

$$\mathbf{V}(\mathbf{n}_k) \approx \alpha_k\mathbf{I} \quad (26b)$$

where  $v$  is the average of the variances of  $\{x_j\}$ , and  $\alpha_k$  is the average of diagonal elements of matrix  $\mathbf{V}(\mathbf{n}_k)$ , which can be calculated as

$$\alpha_k = 2\sigma^2 + vN^{-1} \sum_{l=0}^L (l|h_{k-1}^l|^2 + (L-l)|h_{k+1}^l|^2). \quad (27)$$

Based on (24) and (26), we have

$$\mathbf{V}(\mathbf{r}_k) = v\mathbf{A}_k\mathbf{A}_k^H + \mathbf{V}(\mathbf{n}_k) = \mathbf{F}^H (v\mathbf{G}_k\mathbf{G}_k^H + \alpha_k\mathbf{I})\mathbf{F}. \quad (28)$$

It can be shown that

$$(\mathbf{A}_k^H \mathbf{V}(\mathbf{r}_k)^{-1} \mathbf{A}_k)_{diag} = u_k \mathbf{I} \quad (29)$$

where

$$u_k = \frac{1}{N} \sum_{n=0}^{N-1} |g_{k,n}|^2 (v|g_{k,n}|^2 + \alpha_k)^{-1}. \quad (30)$$

Based on (28), (29) and the result of (12), the extrinsic LLRs for the entries in  $\mathbf{x}_k$  can be computed as

$$e(\mathbf{x}_k) = 4(1 - vu_k)^{-1} \mathbf{S} (\mathbf{F}^H \mathbf{G}_k^H \mathbf{D}^{-1} (\mathbf{z}_k - \mathbf{G}_k \mathbf{F} \mathbf{E}(\tilde{\mathbf{x}}_k) - \mathbf{F} \mathbf{E}(\mathbf{n}_k)) + u_k \mathbf{E}(\tilde{\mathbf{x}}_k)) \quad (31)$$

where

$$\mathbf{S} = [\mathbf{I}_{M \times M}, \mathbf{0}]_{M \times N}, \quad (32a)$$

$$\mathbf{D} = v\mathbf{G}_k\mathbf{G}_k^H + \alpha_k\mathbf{I}, \quad (32b)$$

$$\mathbf{z}_k = \mathbf{F}\mathbf{r}_k, \quad (32c)$$

and  $\mathbf{G}_k$  is a diagonal matrix (see (23)).

*Remarks:*

1. The matrix inversion involved in (31) is trivial since the matrix  $\mathbf{D}$  is diagonal.
2. The product of matrix  $\mathbf{F}$  (or  $\mathbf{F}^H$ ) and a vector can be efficiently realized using FFT.

<sup>2</sup>The approximation (26a) was first implicitly used in [27] to reduce detection complexity, and later adopted in many other works such as [7], [18], and [29]. The impact of approximation (26b) gradually reduces when  $v$  approaches 0 with the iteration.

3. The term “ $-\mathbf{F}\mathbf{E}(\mathbf{n}_k)$ ” in (31) provides soft cancellation of the interference from the adjacent segments  $k+1$  and  $k-1$  in the frequency domain (see (18) for the definition of  $\mathbf{n}_k$ ). Here,  $\mathbf{E}(\mathbf{n}_k) = \mathbf{E}(\mathbf{y}_{k-1}^{tail}) + \mathbf{E}(\mathbf{y}_{k+1}^{head})$  can be efficiently calculated using FFT because

$$\begin{aligned} \mathbf{E}(\mathbf{y}_{k-1}^{tail}) &= \mathbf{E}\left([\text{tail}(\mathbf{h}_{k-1} * \mathbf{x}_{k-1}), 0, \dots, 0]^T\right) \\ &= \mathbf{E}\left([\text{tail}(\tilde{\mathbf{h}}_{k-1} \otimes \tilde{\mathbf{x}}_{k-1}), 0, \dots, 0]^T\right) \\ &= [\text{tail}(\mathbf{F}^H \mathbf{G}_{k-1} \mathbf{F} \mathbf{E}(\tilde{\mathbf{x}}_{k-1})), 0, \dots, 0]^T \end{aligned} \quad (33a)$$

and similarly

$$\mathbf{E}(\mathbf{y}_{k+1}^{head}) = [0, \dots, 0, \text{head}(\mathbf{F}^H \mathbf{G}_{k+1} \mathbf{F} \mathbf{E}(\tilde{\mathbf{x}}_{k+1}))]^T. \quad (33b)$$

Moreover, the term “ $\mathbf{G}_k \mathbf{F} \mathbf{E}(\tilde{\mathbf{x}}_k)$ ” is involved in (31), which indicates that  $\mathbf{G}_k \mathbf{F} \mathbf{E}(\tilde{\mathbf{x}}_k)$  can be shared by  $e(\mathbf{x}_{k-1})$  (for calculating  $\mathbf{E}(\mathbf{y}_k^{head})$ ),  $e(\mathbf{x}_k)$ , and  $e(\mathbf{x}_{k+1})$  (for calculating  $\mathbf{E}(\mathbf{y}_k^{tail})$ ).

4. The complexity involved in (31) is  $O(\log_2 N)$  per symbol.  
5. In this approach, the detection of  $\mathbf{x}_k$  is based on entire  $\mathbf{r}_k$ , and the operation of CP reconstruction is avoided.

#### IV. LOW-COMPLEXITY BIDIRECTIONAL CHANNEL ESTIMATION

##### A. System Model

In this paper, we adopt the widely used wide sense stationary uncorrelated scattering (WSSUS) model [21], [22], in which the channel taps are independent of each other and the autocorrelation function for each tap is the zeroth order Bessel function of the first kind. The power of the  $l$ th tap is denoted by  $q^l$  ( $\{q^l\}$  is often called the power delay profile). Since the channel is assumed to be static within each short segment, we use the following segment-level first-order autoregressive (AR(1)) to approximately characterize the time-varying channel [17]:

$$h_k^l = \beta h_{k-1}^l + w_k^l, \quad l = 0, 1, \dots, L \quad (34)$$

where  $h_k^l$  denotes the  $l$ th channel tap corresponding to segment  $\mathbf{x}_k$ ,  $\beta$  is a constant, and  $w_k^l$  is zero mean white Gaussian noise with power  $p^l = q^l(1 - \beta^2)$ . The parameter  $\beta$  can be determined as in [17]. Equation (34) can be represented by a Forney-style factor graph [15], [16] shown in Fig. 3. When the information about  $h_{k-1}^l$  (in the form of mean and variance) is available, we can infer the information about  $h_k^l$ , and vice versa, based on the GMP technique detailed in [15] and [16].

Now return to (14). Similarly to the previous section, we have the following signal model after ignoring the channel variation within each short segment:<sup>3</sup>

$$\check{\mathbf{r}}_k = \mathbf{B}_k \mathbf{h}_k + \check{\boldsymbol{\eta}}_k \quad (35)$$

where the observation vector  $\check{\mathbf{r}}_k = [r_{kM}, r_{kM+1}, \dots, r_{(k+1)M-1}]^T$ ,  $\mathbf{B}_k$  (with size  $M \times (L+1)$ )

<sup>3</sup>In Section III, the detection of  $\mathbf{x}_k$  is based on  $\mathbf{r}_k$  (with length  $N = M+L$ ) due to the effect of delay spread. In contrast, here we use  $\check{\mathbf{r}}_k$  (with length  $M$ ) to estimate  $\mathbf{h}_k$  to remove the impact from adjacent segments.

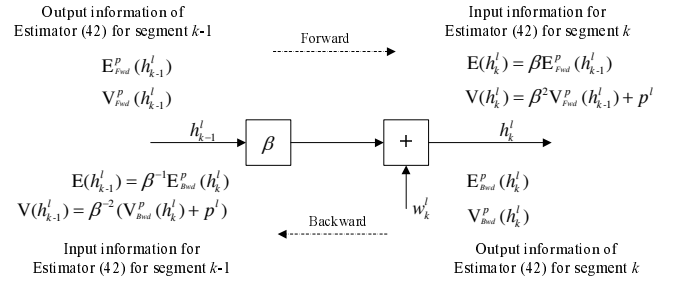


Fig. 3. Exploiting correlation of time-varying channels via bidirectional Gaussian message passing.

is a Toeplitz matrix with the  $l$ th column given by

$$\mathbf{b}_{k,l} = [x_{kM-l}, x_{kM-l+1}, \dots, x_{(k+1)M-l-1}]^T \quad (36)$$

(in (36),  $x_j = 0$  if  $j < 0$ ), and  $\check{\boldsymbol{\eta}}_k$  is the corresponding AWGN vector.

##### B. Low-Complexity Bidirectional Channel Estimation

We focus on one tap  $h_k^l$  and rewrite (35) as

$$\check{r}_k = \mathbf{b}_{k,l} h_k^l + \zeta_{k,l} \quad (37)$$

where

$$\zeta_{k,l} = \sum_{l' \neq l} \mathbf{b}_{k,l'} h_{k,l'} + \check{\eta}_k \quad (38)$$

represents the noise and interference from other taps. The mean vector and covariance matrix of  $\zeta_{k,l}$  are given by

$$\mathbf{E}(\zeta_{k,l}) = \mathbf{E}(\check{r}_k) - \mathbf{b}_{k,l} \mathbf{E}(h_k^l) \quad (39a)$$

and

$$\mathbf{V}(\zeta_{k,l}) = \mathbf{V}(\check{r}_k) - \mathbf{V}(h_k^l) \mathbf{b}_{k,l} \mathbf{b}_{k,l}^H \quad (39b)$$

with

$$\mathbf{E}(\check{r}_k) = \sum_{l'=0}^L \mathbf{b}_{k,l'} \mathbf{E}(h_{k,l'}^l) \quad (39c)$$

and

$$\mathbf{V}(\check{r}_k) = \sum_{l'=0}^L \mathbf{V}(h_{k,l'}^l) \mathbf{b}_{k,l'} \mathbf{b}_{k,l'}^H + 2\sigma^2 \mathbf{I}. \quad (39d)$$

In general,  $\mathbf{V}(\zeta_{k,l})$  is a full matrix. To reduce the computational complexity in the following developed estimation algorithm, we approximate  $\mathbf{V}(\zeta_{k,l})$  using its diagonal part<sup>4</sup>

$$\mathbf{D}_{k,l} = (\mathbf{V}(\zeta_{k,l}))_{diag}. \quad (40)$$

Based on the above approximation, the MMSE estimator (2) and the model (37), we have

$$\mathbf{V}^p(h_k^l) \approx \frac{\mathbf{V}(h_k^l)}{\mathbf{V}(h_k^l) \mathbf{b}_{k,l}^H \mathbf{D}_{k,l}^{-1} \mathbf{b}_{k,l} + 1} \quad (41a)$$

<sup>4</sup>This approximation can be justified as follows. At the beginning of the process of the iterative channel estimation and detection,  $\mathbf{V}(\zeta_{k,l})$  is weakly dominated by its diagonal elements. However, the channel estimates become more and more accurate with iteration, and hence  $\mathbf{E}(\zeta_{k,l})$  also becomes more and more accurate (with respect to the true value of  $\zeta_{k,l}$ ), which are cancelled out from  $\check{r}_k$  (see the term “ $\check{r}_k - \mathbf{E}(\zeta_{k,l})$ ” in (41b)). This means that the covariance matrix  $\mathbf{V}(\zeta_{k,l})$  will be gradually dominated by its diagonal elements due to AWGN, and hence the impact of this approximation gradually reduces with the iteration. On the other hand, the start-up of the iterative process is critical. In Section V, we will discuss a way to start the iterative process using superimposed training.

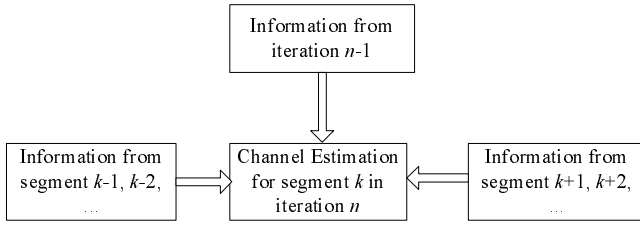


Fig. 4. Relevant information used for channel estimation in segment  $k$  at iteration  $n$ .

and

$$\mathbf{E}^p(h_k^l) \approx \mathbf{E}(h_k^l) + \frac{\mathbf{V}(h_k^l) \mathbf{b}_{k,l}^H \mathbf{D}_{k,l}^{-1} (\tilde{\mathbf{r}}_k - \mathbf{E}(\zeta_{k,l}) - \mathbf{b}_{k,l} \mathbf{E}(h_k^l))}{\mathbf{V}(h_k^l) \mathbf{b}_{k,l}^H \mathbf{D}_{k,l}^{-1} \mathbf{b}_{k,l} + 1}. \quad (41b)$$

After some manipulations, we can rewrite (41) as

$$\mathbf{V}^p(h_k^l) \approx \left( \frac{1}{\mathbf{V}(h_k^l)} + \mathbf{b}_{k,l}^H \mathbf{D}_{k,l}^{-1} \mathbf{b}_{k,l} \right)^{-1} \quad (42a)$$

and

$$\mathbf{E}^p(h_k^l) \approx \mathbf{V}^p(h_k^l) \left( \frac{\mathbf{E}(h_k^l)}{\mathbf{V}(h_k^l)} + \mathbf{b}_{k,l}^H \mathbf{D}_{k,l}^{-1} (\tilde{\mathbf{r}}_k - \mathbf{E}(\zeta_{k,l})) \right). \quad (42b)$$

In (42),  $\mathbf{E}(h_k^l)$  and  $\mathbf{V}(h_k^l)$  denote the *a priori* mean and variance of the concerned tap  $h_k^l$ , and  $\mathbf{E}(\zeta_{k,l})$  and  $\mathbf{D}_{k,l}$  are related to noise-plus-interference. The complexity involved in computing (42) is low since  $\mathbf{D}_{k,l}$  is diagonal. In the following, we will discuss how to determine the values of  $\{\mathbf{E}(h_k^l), \mathbf{V}(h_k^l)\}$  and  $\{\mathbf{E}(\zeta_{k,l}), \mathbf{D}_{k,l}\}$ .

The information that can be used for channel estimation in segment  $k$  at iteration  $n$  is shown in Fig. 4. Specifically, the *a posteriori* means and variances about  $\{h_k^l, l' \neq l\}$  calculated in the last iteration can be used as  $\{\mathbf{E}(h_k^{l'}), \mathbf{V}(h_k^{l'}), l' \neq l\}$  to update  $\mathbf{E}(\zeta_{k,l})$  and  $\mathbf{D}_{k,l}$  in (42) based on (39) and (40), i.e., to perform inter-tap interference cancellation. In contrast, the *a posteriori* information about  $h_k^l$  itself calculated in the last iteration should not be used as  $\mathbf{E}(h_k^l)$  and  $\mathbf{V}(h_k^l)$  in (42) to estimate  $h_k^l$ , since the *a priori* information should be *extrinsic* according to the turbo principle. On the other hand, the *extrinsic a priori* information of  $h_k^l$  can come from its adjacent segments (i.e., segments  $k-1$  and  $k+1$ ) through the use of GMP based on the AR model (34). This motivates us to propose a bidirectional channel estimation approach with the following three steps (see Fig. 3).

### 1. Forward Recursion

Assume that  $\{\mathbf{E}_{Fwd}^p(h_{k-1}^l), \mathbf{V}_{Fwd}^p(h_{k-1}^l)\}$  are available, where the subscript “*Fwd*” denotes the direction of forward recursion. Set  $\mathbf{E}(h_k^l) = \beta \mathbf{E}_{Fwd}^p(h_{k-1}^l)$  and  $\mathbf{V}(h_k^l) = \beta^2 \mathbf{V}_{Fwd}^p(h_{k-1}^l) + p^l$ , and then compute  $\{\mathbf{E}_{Fwd}^p(h_k^l), \mathbf{V}_{Fwd}^p(h_k^l)\}$  using (42).

### 2. Backward Recursion

Assume that  $\{\mathbf{E}_{Bwd}^p(h_k^l), \mathbf{V}_{Bwd}^p(h_k^l)\}$  are available, where the subscript “*Bwd*” denotes the direction of backward recursion. Set  $\mathbf{E}(h_{k-1}^l) = \beta^{-1} \mathbf{E}_{Bwd}^p(h_k^l)$  and  $\mathbf{V}(h_{k-1}^l) = \beta^{-2} (\mathbf{V}_{Bwd}^p(h_k^l) + p^l)$ , and then compute  $\{\mathbf{E}_{Bwd}^p(h_{k-1}^l), \mathbf{V}_{Bwd}^p(h_{k-1}^l)\}$  using (42).

### 3. Combining the Forward and Backward Information

By combining the forward and backward information, the final estimates  $\{\mathbf{E}_{Final}^p(h_k^l), \mathbf{V}_{Final}^p(h_k^l)\}$  can be represented as

$$\mathbf{V}_{Final}^p(h_k^l) = \left( \frac{1}{\mathbf{V}_{Fwd}^p(h_k^l)} + \frac{1}{\beta^{-2} (\mathbf{V}_{Bwd}^p(h_{k+1}^l) + p^l)} \right)^{-1} \quad (43a)$$

and

$$\mathbf{E}_{Final}^p(h_k^l) = \left( \frac{\mathbf{E}_{Fwd}^p(h_k^l)}{\mathbf{V}_{Fwd}^p(h_k^l)} + \frac{\beta^{-1} \mathbf{E}_{Bwd}^p(h_{k+1}^l)}{\beta^{-2} (\mathbf{V}_{Bwd}^p(h_{k+1}^l) + p^l)} \right) \mathbf{V}_{Final}^p(h_k^l) \quad (43b)$$

Equation (43) is the application of the Gaussian Message combining rule [15], [16].

*Remarks:*

1. To accelerate the convergence of the algorithm, we can adopt a serial tap-by-tap estimation schedule, i.e., first estimating  $\{h_k^0, \forall k\}$  via the above described forward and backward recursions, and then estimating  $\{h_k^1, \forall k\}$ , and so on (until  $\{h_k^L, \forall k\}$ ). In this process,  $\{\mathbf{E}(\zeta_{k,l}), \mathbf{D}_{k,l}\}$  are calculated based on the most updated  $\{\mathbf{E}_{Final}^p(h_k^l), \mathbf{V}_{Final}^p(h_k^l)\}$  (i.e., some of them are computed in the current iteration). It is similar to the serial schedule for multi-user detection discussed in [19], and also  $\{\mathbf{E}(\zeta_{k,l}), \mathbf{D}_{k,l}\}$  can be efficiently updated with complexity  $O(M)$ .
2. Both the forward and backward recursions involve computing (42) with different *a priori* information about  $h_k^l$ . Note that, the terms “ $\mathbf{b}_{k,l}^H \mathbf{D}_{k,l}^{-1} (\tilde{\mathbf{r}}_k - \mathbf{E}(\zeta_{k,l}))$ ” and “ $\mathbf{b}_{k,l}^H \mathbf{D}_{k,l}^{-1} \mathbf{b}_{k,l}$ ” in (42) are independent of the *a priori* information, and hence they can be shared by the forward and backward recursions, which is the advantage of using (42) over (41).
3. It can be shown that the complexity of the bidirectional channel estimation approach is  $O((L+1)M)$  (i.e.,  $O(L+1)$  per symbol). If the ISI channel is sparse, the complexity reduces to  $O(L')$  per symbol, where  $L' \leq L+1$  is the number of nonzero channel taps.
4. The discussion can be extended to the case of high-order AR models. For example, we can use a second-order AR (AR(2)) model to get better performance with slightly increased complexity (see the simulation results in Section VI).

## V. DISCUSSIONS ON THE ITERATIVE PROCESS

### A. Equalization with Channel Uncertainty

The proposed channel estimator in Section IV provides the channel estimates in the form of *a posteriori* means and variances  $\{\mathbf{E}_{Final}^p(h_k^l), \mathbf{V}_{Final}^p(h_k^l)\}$ , as shown in (43). To deal with the uncertainty of the channel estimate, we express the channel tap as

$$h_k^l = \mathbf{E}_{Final}^p(h_k^l) + \delta_{h_k^l} \quad (44)$$

where  $\delta_{h_k^l}$  is a Gaussian random variable with mean 0 and variance  $\mathbf{V}_{Final}^p(h_k^l)$ . Define  $\hat{\mathbf{h}}_k =$

$[E_{Final}^p(h_k^0), E_{Final}^p(h_k^1), \dots, E_{Final}^p(h_k^L)]^T$ . According to (17) and (44), we have

$$\mathbf{r}_k = \hat{\mathbf{h}}_k * \mathbf{x}_k + \mathbf{n}'_k \quad (45)$$

where the distortion due to the uncertainty of channel estimates is absorbed into  $\mathbf{n}'_k$ . Assuming that all the random variables involved in  $\mathbf{n}'_k$  are independent of each other and following the discussion in Section III.B, we can perform detection based on (45).

### B. Channel Estimation with Imperfect Matrix $\mathbf{B}_k$

In the iterative process, the means  $\{E(x_j)\}$  and variances  $\{V(x_j)\}$  about  $\{x_j\}$  are available. Similarly to the above discussion, we have the following signal model based on (37)

$$\check{\mathbf{r}}_k = \hat{\mathbf{b}}_{k,l} h_k^l + \zeta'_{k,l} \quad (46)$$

where  $\hat{\mathbf{b}}_{k,l}$  is constructed based on  $\{E(x_j)\}$ , and the distortion due to the uncertainty of data estimates is absorbed into  $\zeta'_{k,l}$ . Assuming that all the random variables involved in  $\zeta'_{k,l}$  are independent of each other and following the discussion in Section IV.B, we can perform channel estimation based on (46).

### C. Start up the Iterative Process Using Pilot Signals

We start up the iterative process by setting up matrices  $\{\mathbf{B}_k\}$  using pilot symbols (since no feedbacks from the decoder are available at the beginning of the iterative process). Denote the pilot sequence by  $\mathbf{t} = [t_0, t_1, \dots, t_{J-1}]^T$ . We assume that  $t_j = \sqrt{P/2}(t_j^{\text{Re}} + it_j^{\text{Im}})$ , and  $t_j^{\text{Re}}$  (and  $t_j^{\text{Im}} \in \{+1, -1\}$ ) is randomly generated, where  $P$  denotes the power of the pilot signal. In this paper, we adopt the so-called superimposed training approach [25], i.e., the pilot symbols are added to the data symbols, which only incurs power loss (and no spectral loss). In this case, the transmitted signal can be represented as  $\mathbf{x} = \mathbf{t} + \mathbf{c}$ , where  $\mathbf{c} = [c_0, c_1, \dots, c_{J-1}]^T$  denotes the data signal. Hence, in the first iteration,  $E(x_j) = E(t_j) + E(c_j) = t_j$  and  $V(x_j) = V(c_j) = 1$  (the energy of one QPSK symbol is normalized to 1). Then the iterative process can be started. Note that, in the process of equalization, some extra straightforward operations should be included to remove the contribution from the pilot signal.

### D. Steps of the Iterative Process

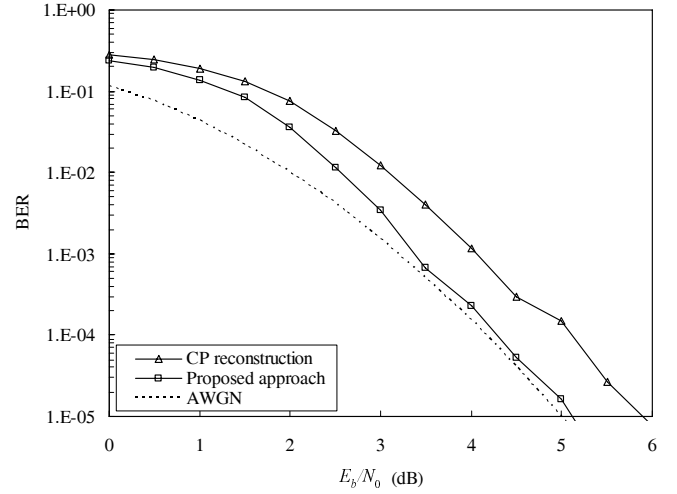
The steps of the iterative process are concluded as follows:

Step 1. Using the pilot signal and the feedbacks from the decoder (in the first iteration,  $E(c_j) = 0$  and  $V(c_j) = 1$ ), the channel estimator perform a serial tap-by-tap channel estimation through the forward and backward recursions based on (42) and (43).

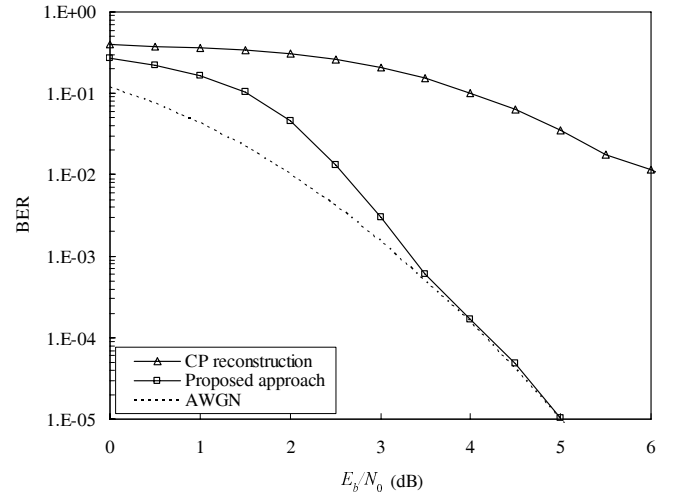
Step 2. Based on the channel estimates and the feedbacks from the decoder, equalization is performed using (31).

Step 3. The decoder refines the data estimates based on coding constraints, and the results will be used by the channel estimator and the equalizer in the next iteration. Then go to Step 1.

Hard decisions on the information bits are made by the decoder during the final iteration.



(a) Results based on modified channel power delay profile type B in HIPERLAN/2 [24].



(b) Results based on uniform channel power delay profile.

Fig. 5. Performance of the proposed approach and the CP reconstruction based approach in 17-tap quasi-static ISI channels with normalized energy. The segment length is 64 and the number of iterations is 10.

## VI. SIMULATION RESULTS

In the simulations, a rate-1/2 convolutional code with generator  $(23, 35)_8$  and QPSK modulation with Gray mapping are employed, and the number of iterations is always 10.

### A. Static ISI Channels with Perfect CSI at the Receiver

To facilitate the comparison of our proposed segment-by-segment equalization approach and the known performance bound, static ISI channels are assumed. We set the length of information bits to be 4096, and so is the block length (i.e., length of  $\mathbf{x}$ ). The segment length (i.e., the length of  $\mathbf{x}_k$ ) is set to be 64, and hence each block consists of 64 segments. The number of channel taps is 17. In each channel realization, the 17 coefficients  $\{h^l\}$  are independently generated according to distributions  $\{h^l \sim \mathbf{N}(0, q^l)\}$  (where  $\mathbf{N}(0, q^l)$  denotes a complex Gaussian distribution with mean 0 and variance  $q^l$ ), and remain constant for all the segments in a block. Moreover, the channel energy is normalized to 1 (i.e.,  $\sum_{l=0}^{16} |h^l|^2 = 1$ ) for each channel realization. As we know, the performance of

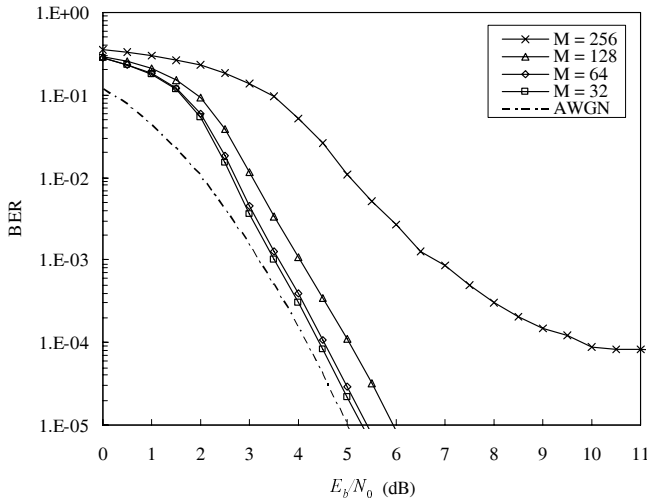


Fig. 6. Impact of different segment lengths on system performance with known CSI. The number of channel taps is 17. The number of iterations is 10.

the system over such ISI channels is bounded by that over an AWGN channel [7].

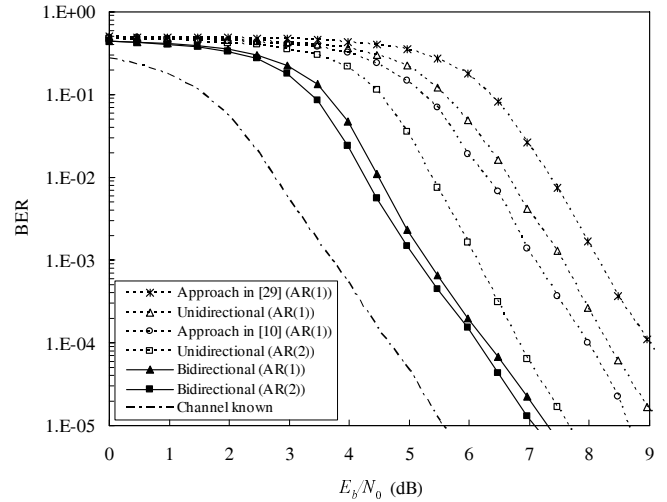
The simulation results are shown in Fig. 5(a) and (b), where a channel power delay profile type B in HIPERLAN/2 [24]<sup>5</sup> and a uniform channel power delay profile are adopted, respectively. The performance of CP reconstruction based approach is also included for comparison. From this figure, it can be found that the proposed technique can almost achieve the performance bound at relatively high  $E_b/N_0$  range, which implies that ISI and inter-segment interference are almost eliminated. In contrast, the CP reconstruction based approach can not approach this bound closely due to the energy loss. The performance of the CP reconstruction based approach is worse for the uniform channel power delay profile than for the channel power delay profile in [24] because the former channel power delay profile causes severer interference between adjacent segments.

**B. Doubly Selective Channels**

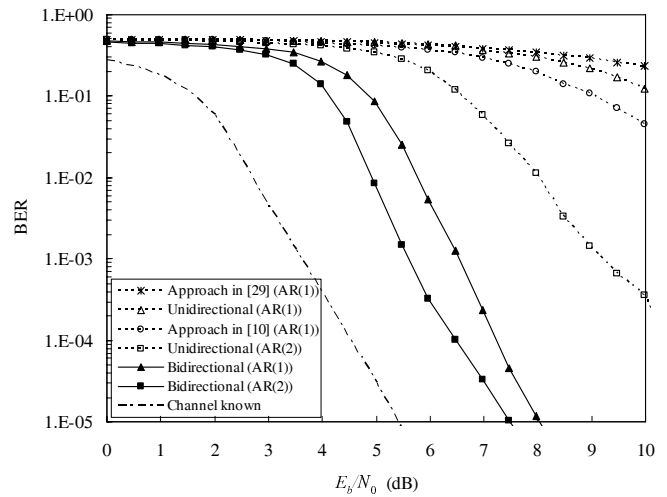
The WSSUS channel model [21], [22] is adopted, and the channels vary symbol by symbol (i.e., they vary within each segment). We set carrier frequency  $f_0 = 2$  GHz, symbol duration  $T_s = 2.5\mu s$ , and mobile speed  $v_m = 280$  km/hr. Hence the Doppler spread  $f_D = 518.52$  Hz. The power delay profile is selected as  $q^l = \exp(-0.1l)$ . For each channel realization, the average channel energy is normalized to 1 (i.e.,  $J^{-1} \sum_{j=0}^{J-1} \sum_{l=0}^L |h_{j,l}|^2 = 1$ ). The information bit length is 2048, and so is the block length.

We first test the proposed approach with CSI available at the receiver (i.e., the exact CSI corresponding to the middle of each segment is known at the receiver), and examine the impact of different segment lengths on system performance. Fig. 6 shows the system performance with segment length  $M = 32, 64, 128,$  and  $256$  (i.e., correspondingly  $f_D M T_s = 0.041, 0.083, 0.166,$  and  $0.331$ , respectively) when the number of channel taps is 17. The performance of the system over an

<sup>5</sup>Since there are only 16 taps in [24], we set the power of the 17th tap to zero in the simulation.



(a) The number of channel taps is 9.



(b) The number of channel taps is 17.

Fig. 7. Performance of the proposed approach and the approaches in [10] and [29]. The power loss of about 1 dB due to the use of the pilot signal is included in  $E_b/N_0$ . The number of iterations is 10, and the segment length is 64.

AWGN channel is also shown for reference.<sup>6</sup> It can be found from Fig. 6 that the system performance degrades with the increase of segment length. Performance is relatively poor at  $M = 128$  compared with  $M = 64$  and  $32$ . An error floor occurs at  $M = 256$ . This is because the algorithm assumes a static ISI channel within a segment (i.e., the error vector  $\delta_k$  in (15) is ignored). It can also be found that the system has similar performance when  $M = 32$  and  $64$ . This indicates that, the assumption of static channel within each segment is reasonable when  $M \leq 64$ . Here, we should note that, if the conventional FDE is employed with  $M = 64$ , the use of CP will incur power loss of at least about 1 dB and a considerable spectral loss of more than 20%.

We next examine the system performance with estimated CSI. We set  $M = 64$  at which the channel can be regarded

<sup>6</sup>We should note that it is only a loose lower bound for the performance of the system over the time-varying channels with normalized average channel energy 1 due to the fluctuation of instantaneous channel energy, which is different from the case of time-invariant ISI channel in Section VIA.

as approximately static within a segment. The power ratio of the superimposed pilot signal to data signal is 1/4, resulting in power loss of about 1 dB. The performance of the proposed approach is shown in Fig. 7, where the power loss of about 1 dB due to the use of pilot signal has been included in  $E_b/N_0$ . It can be found that if the power loss is ignored, the performance gap between the system with the proposed bidirectional channel estimation algorithm and that with channel known is within 2dB using AR(1) model, and it is within 1dB using AR(2) model, for the time-varying channels with both 9 taps and 17 taps. The performance of the unidirectional KF-based approaches with AR(1) in [10] and [29] (whose complexities are  $O((L+1)^2)$  and  $O(\log_2(M+L))$ , respectively), are also included for comparison. It can be found that, among all the unidirectional approaches based on AR(1) channel model, the approach in [10] delivers the best performance but at the cost of highest complexity, and the approach in [29] suffers from performance loss due to the approximation involved in performing KF in the frequency domain. It can also be found that the proposed bidirectional approach outperforms all the unidirectional approaches for all scenarios. Their performance gaps become very significant when the number of channel taps is large.

## VII. CONCLUSION

We have proposed a low-complexity iterative joint channel estimation, equalization and decoding technique for doubly selective channels. The key to the proposed technique is a segment-by-segment processing strategy. Through a virtual zero-padding technique, the proposed segment-by-segment equalizer inherits the low-complexity advantage of the conventional FDE technique. On the other hand, it does not resort to CP, thereby avoids the power and spectral overheads of the conventional FDE technique. To effectively exploit the correlation characteristic of time-varying channels, we have also developed a low-complexity bidirectional channel estimation algorithm by using the recently proposed Gaussian message passing technique. Simulation results have demonstrated the effectiveness of the proposed equalization and channel estimation algorithms.

### APPENDIX A

#### DERIVATION OF EQUATION (11)

Define

$$\bar{\mathbf{x}} = \begin{bmatrix} \mathbf{x}^{\text{Re}} \\ \mathbf{x}^{\text{Im}} \end{bmatrix}, \quad \bar{\mathbf{r}} = \begin{bmatrix} \mathbf{r}^{\text{Re}} \\ \mathbf{r}^{\text{Im}} \end{bmatrix}, \quad \bar{\mathbf{n}} = \begin{bmatrix} \mathbf{n}^{\text{Re}} \\ \mathbf{n}^{\text{Im}} \end{bmatrix}, \quad (47a)$$

and

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A}^{\text{Re}} & -\mathbf{A}^{\text{Im}} \\ \mathbf{A}^{\text{Im}} & \mathbf{A}^{\text{Re}} \end{bmatrix}. \quad (47b)$$

Then a complex model given by (3) is equivalent to the following real one

$$\bar{\mathbf{r}} = \bar{\mathbf{A}}\bar{\mathbf{x}} + \bar{\mathbf{n}}. \quad (48)$$

From (47), we can see that  $x_j = \bar{x}_j + i\bar{x}_{I+j}$ , where  $I$  is the length of  $\mathbf{x}$ , and  $\bar{x}_j$  is the  $j$ th element in  $\bar{\mathbf{x}}$ . Let us focus on  $\bar{x}_j$ , and treat all the other bits as interference. Then (48) can be rewritten as

$$\bar{\mathbf{r}} = \bar{\mathbf{a}}_j\bar{x}_j + \bar{\boldsymbol{\xi}}_j \quad (49)$$

where  $\bar{\mathbf{a}}_j$  is the  $j$ th column of  $\bar{\mathbf{A}}$ , and  $\bar{\boldsymbol{\xi}}_j = \sum_{j' \neq j} \bar{\mathbf{a}}_{j'}\bar{x}_{j'} + \bar{\mathbf{n}}$ , which can be approximated as a joint Gaussian random variable. We assume that  $V(x_j^{\text{Re}}, x_j^{\text{Im}}) = 0$ ,  $V(x_j^{\text{Re}}) = V(x_j^{\text{Im}})$ ,  $V(\mathbf{n}^{\text{Re}}, \mathbf{n}^{\text{Im}}) = \mathbf{0}$ , and  $V(\mathbf{n}^{\text{Re}}) = V(\mathbf{n}^{\text{Im}})$  (which means that  $V(x_j^{\text{Re}}) = V(x_j^{\text{Im}}) = 0.5V(x_j)$  and  $V(\mathbf{n}^{\text{Re}}) = V(\mathbf{n}^{\text{Im}}) = 0.5V(\mathbf{n})$ ), then the covariance matrix of  $\bar{\boldsymbol{\xi}}_j$  is given by

$$V(\bar{\boldsymbol{\xi}}_j) = V(\bar{\mathbf{r}}) - \frac{1}{2}V(x_j)\bar{\mathbf{a}}_j\bar{\mathbf{a}}_j^T \quad (50)$$

with

$$V(\bar{\mathbf{r}}) = \frac{1}{2} \left( \sum_{j=0}^{2I-1} V(x_{(j)_I})\bar{\mathbf{a}}_j\bar{\mathbf{a}}_j^T + \mathbf{C} \right) \quad (51a)$$

and

$$\mathbf{C} = \begin{bmatrix} V(\bar{\mathbf{n}}) & \mathbf{0} \\ \mathbf{0} & V(\bar{\mathbf{n}}) \end{bmatrix} \quad (51b)$$

where  $(\cdot)_I$  denotes the modulo operation with respect to  $I$ . Similarly to the real case, we can obtain the extrinsic LLR of  $\bar{x}_j$  as follows:

$$e(\bar{x}_j) = 2 \frac{\bar{\mathbf{a}}_j^T V(\bar{\mathbf{r}})^{-1} (\bar{\mathbf{r}} - E(\bar{\mathbf{r}}) + \bar{\mathbf{a}}_j E(\bar{x}_j))}{1 - 2^{-1} V(\bar{x}_j) \bar{\mathbf{a}}_j^T V(\bar{\mathbf{r}})^{-1} \bar{\mathbf{a}}_j}. \quad (52)$$

We can verify that

$$\begin{aligned} e(x_j) &= e(\bar{x}_j) + ie(\bar{x}_{j+I}) \\ &= 4 \frac{\mathbf{a}_j^H V(\mathbf{r})^{-1} (\mathbf{r} - E(\mathbf{r}) + \mathbf{a}_j E(x_j))}{1 - V(x_j) \mathbf{a}_j^H V(\mathbf{r})^{-1} \mathbf{a}_j} \end{aligned} \quad (53)$$

based on the fact that, for an invertible complex matrix  $\mathbf{B}$ , the inverse of matrix  $\bar{\mathbf{B}}$  defined as

$$\bar{\mathbf{B}} = \begin{bmatrix} \mathbf{B}^{\text{Re}} & -\mathbf{B}^{\text{Im}} \\ \mathbf{B}^{\text{Im}} & \mathbf{B}^{\text{Re}} \end{bmatrix} \quad (54)$$

is given by

$$\bar{\mathbf{B}}^{-1} = \begin{bmatrix} (\mathbf{B}^{-1})^{\text{Re}} & -(\mathbf{B}^{-1})^{\text{Im}} \\ (\mathbf{B}^{-1})^{\text{Im}} & (\mathbf{B}^{-1})^{\text{Re}} \end{bmatrix}. \quad (55)$$

Equation (11) follows from (53).

### ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their helpful comments to improve the quality of this paper.

### REFERENCES

- [1] M. Tüchler and J. Hagenauer, "Turbo equalization using frequency domain equalizers," in *Proc. Allerton Conf.*, Monticello, IL, USA, Oct. 2000, pp. 1234-1243.
- [2] D. Kim and G. Stüber, "Residual ISI cancellation for OFDM with applications to HDTV broadcasting," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1590-1599, Oct. 1998.
- [3] Y. Li, S. McLaughlin, and D. G. M. Cruickshank, "Bandwidth efficient single carrier systems with frequency domain equalization," *Electron. Lett.*, vol. 41, no. 15, pp. 857-858, July 2005.
- [4] A. Gusmão, P. Torres, R. Dinis, and N. Esteves, "Code-assisted SC/FDE block transmission with reduced cyclic prefix overhead," in *Proc. IEEE/IEEE Int. Symp. Commun. Theory Appl.*, July 2005, pp. 398-402.
- [5] A. Gusmão, P. Torres, R. Dinis, and N. Esteves, "A turbo FDE technique for reduced-CP SC-based block transmission systems," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 16-20, Jan. 2007.

- [6] T. Hwang and Y. Li, "Iterative cyclic prefix reconstruction for coded single-carrier systems with frequency-domain equalization (SC-FDE)," in *Proc. IEEE Veh. Technol. Conf.*, Apr. 2003, vol. 3, pp. 1841-1845.
- [7] M. Tüchler, R. Koetter, and A. C. Singer, "Turbo equalization: principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754-767, May 2002.
- [8] M. Tüchler, A. C. Singer, and R. Koetter, "Minimum mean squared error equalization using a priori information," *IEEE Trans. Signal Processing*, vol. 50, no.3, pp. 673-683, Mar. 2002.
- [9] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046-1061, July 1999.
- [10] S. Song, A. C. Singer, and K. Sung, "Soft input channel estimation for turbo equalization," *IEEE Trans. Signal Processing*, vol. 52, no. 10, pp. 2885-2894, Oct. 2004.
- [11] H. Schoeneich and P. A. Hoeher, "Iterative pilot-layer aided channel estimation with emphasis on interleave-division multiple access systems," *EURASIP J. Applied Signal Processing*, vol. 2006, pp. 1-15, 2006.
- [12] R. Otnes and M. Tüchler, "Iterative channel estimation for turbo equalization of time-varying frequency-selective channels," *IEEE Trans. Commun.*, vol. 3, no. 6, pp. 1918-1923, Nov. 2004.
- [13] K. Ishihara, K. Takeda, and F. Adachi, "Iterative channel estimation for frequency-domain equalization of DSSS signals," *IEICE Trans. Commun.*, vol. E90-B, pp. 1171-1180, May 2007.
- [14] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Prentice-Hall PTR, 1993.
- [15] H. -A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Mag.*, vol. 21, no. 1, pp. 28-41, Jan. 2004.
- [16] H. -A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295-1322, June 2007.
- [17] Q. Dai and E. Shwedyk, "Detection of bandlimited signals over frequency selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 941-950, Feb./Mar./Apr. 1994.
- [18] Q. Guo, X. Yuan, and L. Ping, "Multi-user detection techniques for potential 3GPP long term evolution (LTE) schemes," *Lecture Notes in Electrical Engineering*, vol. 1, pp. 77-86, June 2007.
- [19] L. Liu, W. K. Leung, and L. Ping, "Simple chip-by-chip multiuser detection for CDMA systems," in *Proc. IEEE Veh. Technol. Conf.*, Apr. 2003, pp. 2157-2161.
- [20] 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, Mar. 1974.
- [21] W. C. Jakes, *Microwave Mobile Communications*, New York: John Wiley & Sons, 1974.
- [22] J. I. Smith, "A computer generated multipath fading simulation for mobile radio," *IEEE Trans. Veh. Technol.*, vol. VT-24, no. 3, pp. 39-40, Aug. 1975.
- [23] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.
- [24] Z. Wang, X. Ma, and G. B. Giannakis, "OFDM or single-carrier block transmissions?" *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 380-394, Mar. 2004.
- [25] A. G. Orozco-Lugo, M. M. Lara, and D. C. McLernon, "Channel estimation using implicit training," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 240-254, Jan. 2004.
- [26] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58-66, Apr. 2002.
- [27] A. Glavieux, C. Laot, and J. Labat, "Turbo equalization over a frequency selective channel," in *Proc. Int. Symp. Turbo Codes*, pp. 96-102, Brest, France, Sept. 1997, pp. 96-102.
- [28] H. Liu and P. Schniter, "Iterative frequency-domain channel estimation and equalization for single-carrier transmissions without cyclic-prefix," in *Proc. IEEE 41st Annual Conf. Inform. Sciences Systems*, Mar. 2007, pp. 829-834.
- [29] H. Liu and P. Schniter, "Iterative frequency-domain channel estimation and equalization for single-carrier transmissions without cyclic-prefix," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3686-3691, Oct. 2008.
- [30] A. Anastasopoulos and K. M. Chugg, "Adaptive soft-input soft-output algorithms for iterative detection with parametric uncertainty," *IEEE Trans. Commun.*, vol. 48, no. 10, pp. 1638-1648, Oct. 2000.
- [31] L. Davis, I. Collings, and P. Hoeher, "Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fast-fading channels," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2106-2114, Dec. 2001.
- [32] A. Hansson and T. Aulin, "Generalized APP detection of continuous phase modulation over unknown ISI channels," *IEEE Trans. Commun.*, vol. 53, no. 10, pp. 1615-1619, Oct. 2005.



**Qinghua Guo** (S'07 - M'08) received the B.E. degree in electronic engineering, M.E. degree in signal and information processing from Xidian University, Xi'an, China, in 2001 and 2004, respectively, and the Ph.D. degree in electronic engineering from City University of Hong Kong, Hong Kong, in 2008. He is now with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Australia. His research interests include statistical signal processing and wireless communications.



**Li Ping** (S'87-M'91-SM'06) received his Ph.D. degree at Glasgow University in 1990. He lectured at Department of Electronic Engineering, Melbourne University, from 1990 to 1992, and worked as a member of research staff at Telecom Australia Research Laboratories from 1993 to 1995. He has been with the Department of Electronic Engineering, City University of Hong Kong, since January 1996, where he is now a chair professor. His research interests are communications systems and coding theory. Dr. Li Ping was awarded a British Telecom-Royal Society Fellowship in 1986, the IEE J J Thomson premium in 1993 and a Croucher Senior Research Fellowship in 2005.



**Defeng (David) Huang** (M'01 - S'02 - M'05 - SM'07) received the B. E. E. E. and M. E. E. E. degree in electronic engineering from Tsinghua University, Beijing, China, in 1996 and 1999, respectively, and the Ph.D. degree in electrical and electronic engineering from the Hong Kong University of Science and Technology (HKUST), Kowloon, Hong Kong, in 2004. Currently, he is an associate professor with School of Electrical, Electronic and Computer Engineering at the University of Western Australia. Dr. Huang serves as an Editor for the

IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.