shift gain than the exponentially decaying channel. Moreover, it is verified that the C-Seq with $L = 6$ exhibits a higher shift gain than that with $L = 3$, but the diversity gain still remains as 3. As a result, if the channel length is not considered in pilot designs, the MLE cannot fully exploit the DOF of the channel. Note that these results can be also verified from the upper bounds.

To confirm our modified criteria for consistent CFO estimation, the results in Fig. 4 are simulated for 5-tap exponentially decaying fading channels with $\epsilon = 30.24$ and the simulation runs of 100 000. In addition, the CRB of the AP-Seq is obtained from [14]. We see that the MLE approaches the CRB at a high SNR, and various consistent sequences exhibit different performance. It should be noted that the large mean square error (MSE) in low- and moderate-SNR regions results from outliers. Similar to the previous simulation results, C-Seq shows reduced outliers than B-Seq and AP-Seq in low- and moderate-SNR regions, whereas A-Seq exhibits the smallest outliers. As a result, we confirm that A-Seq provides the best performance over other sequences for consistent CFO estimation. Hence, it is shown that our modified criteria are also effective for consistent CFO estimation.

V. Conclusion

In this paper, we have derived the average PEP of the MLE and established the relationship of the PEP and the consistent sequence in terms of a diversity gain and a shift gain. Then, we have presented the criteria suitable for the case where the subcarriers are correlated. Simulation results show that the consistent sequences based on our modified criteria produce much reduced outliers over conventional sequences for consistent CFO estimation in frequency-selective fading channels.

References

[1] J. A. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 25, no. 5, pp. 5–14, May 1990.

[2] K. Lee and I. Lee, "Robust pilot designs for consistent frequency offset estimation in OFDM systems," in *Proc. IEEE GLOBECOM*, Dec. 2010, pp. 1–5.

[3] K. Lee, S.-H. Moon, and I. Lee, "Low-complexity leakage-based carrier frequency offset estimation techniques for OFDMA uplink systems," in *Proc. IEEE GLOBECOM*, Dec. 2010, pp. 1–5.

[4] J. Lei and T.-S. Ng, "A consistent OFDM carrier frequency offset estimator based on distinctively spaced pilot tones," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 588–599, Mar. 2004.

[5] Y. Li, H. Minn, N. Al-Dhahir, and A. R. Calderbank, "Pilot designs for consistent frequency-offset estimation in OFDM systems," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 864–877, May 2007.

[6] D. Huang and K. B. Letaief, "Carrier frequency offset estimation for OFDM systems using subcarriers," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 813–823, May 2006.

[7] Q. T. Zhang and D. P. Liu, "A simple capacity formula for correlated diversity Rician fading channels," *IEEE Commun. Lett.*, vol. 6, pp. 481–483, Nov. 2002.

[8] Z. Fang, X. Bao, L. Li, and Z. Wang, "Performance analysis and power allocation for amplify-and-forward cooperative networks over Nakagami-m fading channel," *IEICE Trans. Commun.*, vol. E92-B, no. 3, pp. 1004–1012, Mar. 2009.

[9] Z. Wang and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1389–1398, Aug. 2003.

[10] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[11] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.

[12] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1910–1917, Nov. 2003.

[13] M. Morelli and M. Moretti, "Integer frequency offset recovery in OFDM transmissions over selective channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5220–5226, Dec. 2008.

[14] Y. Li, H. Minn, and J. Zeng, "An average Cramer–Rao bound for frequency offset estimation in frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 9, no. 3, pp. 871–875, Mar. 2010.

Variational-Inference-Based Data Detection for OFDM Systems With Imperfect Channel Estimation

Feng Li, Zongben Xu, and Shihua Zhu

***Abstract*—This paper studies the problem of joint estimation of data and channels for orthogonal frequency-division multiplexing (OFDM) systems using variational inference. The proposed methods are used to combat imperfect channel estimation at the receiver since it can degrade system performance seriously. The proposed methods simplify the maximum *a posteriori* (MAP) scheme based on the theory of variational inference and formulate an optimization problem using variational free energy. The channel state information (CSI) and data are dealt with jointly and iteratively. The proposed schemes offer a variety of solutions for getting soft information when turbo equalization is implemented for coded systems. The effectiveness of the new approach is demonstrated by Monte Carlo simulations.**

***Index Terms*—Channel imperfections, orthogonal frequency-division multiplexing (OFDM), variational inference.**

I. Introduction

Orthogonal frequency-division multiplexing (OFDM) is being investigated as a promising solution of physical layers for current and future wireless communication systems. However, much attention should be paid to several problems that challenge the practical utilization of OFDM, including channel estimation, time-and-frequency offset, phase noise, peak-to-average power ratio, and data detection [1]. In this paper, we consider the problem of data detection with imperfect channel estimation.

Anderson [2] studies the relationship between channel fade and its estimate. The model used in [2] is specialized to three typical channel estimation approaches by Annavajjala *et al.* in [3]. There have been many effects in analyzing OFDM with imperfect channel estimation, e.g., in [4] and [5]. Song *et al.* [4] show us that, compared with other parameter imperfections, inherent channel estimation imperfection is the main reason that degrades the system performance. Krondorf and Fetweiss [5] investigate the Alamouti space–time-coded OFDM performance under receiver impairments, and it is demonstrated that channel uncertainty weights more heavily than carrier frequency offset and I/Q imbalance against system performance. Since imperfect

Manuscript received April 22, 2012; revised August 28, 2012; accepted November 30, 2012. Date of publication December 5, 2012; date of current version March 13, 2013. This work was supported in part by the National Natural Science Foundation of China under Grant 11131006 and Grant 61075054, by the International Cooperation and Exchanges Project under Grant S2010GR0902, and by the National Major Special Projects in Science and Technology of China under Grant 2011ZX03003-002-03. The review of this paper was coordinated by Prof. G. Bauch.

F. Li and S. Zhu are with the Department of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: lf1981@mail.xjtu.edu.cn; szhu@mail.xjtu.edu.cn).

Z. Xu is with the Institute for Information and System Sciences, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: zbxu@mail.xjtu.edu.cn).

Digital Object Identifier 10.1109/TVT.2012.2231972

channel estimation can affect OFDM system performance seriously, we should seek for strategy to overcome it.

The theory basis of our scheme is variational inference, which has been widely employed in many fields, such as applied physics, machine learning, and image processing [6], [7]. Using [8] and [9] for example, many authors have also implemented it in wireless communication systems. Lin and Lim [8] propose a framework of joint data detection for OFDM systems with phase noise. Stefanatos and Katsaggelos [9] propose an algorithm for joint data detection and channel tracking based on the variational Bayes method with the channel being modeled as an autoregressive process. We have studied the iterative detection approach to OFDM with imperfect channel state information (CSI) [10]. The relationship between variational inference and the expectation–maximization (EM) algorithm shows us that there exists a family of choices by treating the unknowns as parameters or hidden variables, separately [8], [11]. The proposed scheme in [10] can be regarded as an EM algorithm with only hidden variables. In this paper, we will study other schemes by treating CSI and data as parameters or hidden variables, separately, and then a framework of iterative detector for OFDM with channel estimation error can be established.

The contributions of this paper are listed as follows. First, we propose algorithms for joint estimation of data and channel for OFDM in an iterative manner based on the theory of variational inference. Second, by choosing proper approximate distributions, we can obtain different schemes. The key parameter of the approximate distributions, e.g., the variance of Gaussian distribution, that indicates the degree of reliability of data estimation can be viewed as the soft information that can be used to construct a turbo structure for coded systems. Finally, compared with the published work in [10], the proposed schemes achieve the same performance with lower complexity.

Notation: $\mathbf{A}^T$, $\mathbf{A}^*$, and $\mathbf{A}^H$ denote transpose, conjugate, and conjugate transpose of matrix $\mathbf{A}$. The kth-row lth-column element of matrix $\mathbf{A}$ is written as $\mathbf{A}(k,l)$. $\mathrm{diag}(\mathbf{X})$ is the diagonal matrix with the same diagonal elements of matrix $\mathbf{X}$, and $\mathrm{diag}(\mathbf{x})$ is the diagonal matrix created with vector $\mathbf{x}$. $|\mathbf{A}|$ is the determinant of matrix $\mathbf{A}$. $\mathbf{0}$ is the matrix or the vector with all of its elements being zero. $\mathrm{tr}(\mathbf{A})$ is the trace of matrix $\mathbf{A}$. $\mathbf{A}_L$ denotes the matrix consisted of the first L columns of matrix $\mathbf{A}$. $\mathbf{I}_N$ is the $N \times N$ identity matrix. $p(\mathbf{x})$ denotes the probability density function of $\mathbf{x}$. Let $\mathcal{CN}(\mathbf{Q}, \mathbf{V})$ denote the circularly symmetric complex Gaussian random vectors with mean $\mathbf{Q}$ and covariance matrix $\mathbf{V}$. For an N-dimensional circularly symmetric complex Gaussian random vector $\mathbf{x}$, $\mathcal{CN}(\mathbf{Q}, \mathbf{V}) = (1/\pi^N|\mathbf{V}|)e^{\{-(\mathbf{x}-\mathbf{Q})^H\mathbf{V}^{-1}(\mathbf{x}-\mathbf{Q})\}}$. Expectation is denoted by $\mathbf{E}[\cdot]$.

II. System Description

Assuming that there are N subcarriers and that each of them is used for effective information data transmission. Without loss of generality, the cyclic prefix (CP) is long enough to avoid intersymbol interference. We consider a quasi-static fading channel model of which the channel impulse response (CIR) remains constant during every OFDM symbol and varies block by block. We ignore the index of the OFDM block for brevity in the following. Let $\mathbf{r} = [r_0, r_1, \ldots, r_{N-1}]^T$ with r_n be the nth complex baseband sampled received signal in the time domain after removing CP. $\mathbf{s} = [s_0, s_1, \ldots, s_{N-1}]^T$ denotes the modulated symbol vector in the frequency domain. Each element of $\mathbf{s}$ is assumed to be independent identically distributed and belongs to a finite 2-D constellation set, e.g., quadratic-amplitude modulation. Let $\mathbf{F}$ denote the N-point unitary discrete Fourier transform (DFT) matrix with $\mathbf{F}(k,l) = (1/\sqrt{N})\exp\{-j2\pi kl/N\}$, k, $l = 0, 1, \ldots, N-1$. Then, $\mathbf{F}^H$ is the N-point inverse DFT matrix. Let the number of paths of the wireless channel be L. $\mathbf{h} = [h_0, h_1, \ldots, h_{L-1}]^T$, where h_l is the CIR of the lth path. Define $\mathbf{v} = [v_0, v_1, \ldots, v_{N-1}]^T$ with v_n as the frequency domain channel response at the nth subcarrier and can be written as $\mathbf{v} = \mathbf{F}_L\mathbf{h}$. Define $\mathbf{H} = \mathrm{diag}(\mathbf{v})$; then, $\mathbf{r}$ can be expressed as

$$\mathbf{r} = \mathbf{F}^H\mathbf{H}\mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathbf{n} = [n_0, n_1, \ldots, n_{N-1}]^T$ are samples of complex white Gaussian noise with variance $\sigma^2/2$ per dimension. When CSI at the receiver side is inaccurate, we need to equalize the channel distortion to detect the information symbol $\mathbf{s}$.

III. Joint Detection Via Variational Inference

A. Variational Inference

The theory of variational is an approach of approximating complex probability distribution. It is called variational inference when used to deduce the likelihood function of unknown parameters in statistical problems. It has been widely used in statistical physics, and the details of its application in physics could be found in [6] and [7]. When it is implemented in wireless communication systems, we always need to infer the unknown inputs based on the observed output, which results from the unknown inputs going through the wireless channel. This paper concentrates on the problem of joint estimation of data and channels for OFDM and explains it specifically. We need to infer the unknown inputs $\mathbf{s}$ and $\mathbf{h}$ based on the observed output $\mathbf{r}$. The optimal solution of the estimation of data $\mathbf{s}$ is the maximization of the following:

$$p(\mathbf{s}|\mathbf{r}) = \int p(\mathbf{s}, \mathbf{h}|\mathbf{r})\, d\mathbf{h}. \tag{2}$$

In fact, (2) is obtained by regarding $\mathbf{h}$ as a nuisance parameter and integrating it out based on the estimation theory of [14]. The difficulty of solving (2) lies in two aspects. First, the information symbol $\mathbf{s}$ is always chosen from a discrete set, e.g., binary phase-shift keying or quadrature phase-shift keying (QPSK), leading to the problem of (2) as an NP-complete one. Second, the integrand distribution of interest in (2) $p(\mathbf{s}, \mathbf{h}|\mathbf{r})$ always has a very complex form, making the optimization of (2) hard to be resolved. In regard to these two difficulties, first, although $\mathbf{s}$ is chosen from a discrete set, the theory of variational inference allows us to assume that it follows a continuous distribution, e.g., Gaussian distribution. Second, when straight treatment of $p(\mathbf{s}, \mathbf{h}|\mathbf{r})$ is intractable, a variance inference technique relaxes the problem by allowing us to fall back on a traceable distribution, which is written as $f(\mathbf{s}, \mathbf{h})$, where the constant $\mathbf{r}$ omitted, to approximate $p(\mathbf{s}, \mathbf{h}|\mathbf{r})$.

We want $f(\mathbf{s}, \mathbf{h})$ to approximate $p(\mathbf{s}, \mathbf{h}|\mathbf{r})$ as similar as possible. *Kullback–Leibler divergence* between the two distributions is a good measure of the degree of similarity. However, since $p(\mathbf{s}, \mathbf{h}|\mathbf{r})$ has a very complicated form, we calculate an equivalent option $p(\mathbf{s}, \mathbf{h}, \mathbf{r}) = p(\mathbf{s}, \mathbf{h}|\mathbf{r})p(\mathbf{r})$ instead. As a matter of fact, since $p(\mathbf{r})$ is a constant, $p(\mathbf{s}, \mathbf{h}, \mathbf{r})$ is in direct proportion to $p(\mathbf{s}, \mathbf{h}|\mathbf{r})$ and is called the complete likelihood function. Therefore, the *variational free energy* is defined as (3) and is equivalent to the Kullback–Leibler divergence between $f(\mathbf{s}, \mathbf{h})$ and $p(\mathbf{s}, \mathbf{h}|\mathbf{r})$ up to an additive constant.

$$\omega = \int f(\mathbf{s}, \mathbf{h}) \log \frac{f(\mathbf{s}, \mathbf{h})}{p(\mathbf{r}|\mathbf{s}, \mathbf{h})\, p(\mathbf{s}) p(\mathbf{h})} d\mathbf{s} d\mathbf{h}. \tag{3}$$

In fact, (3) is based on the assumption that $p(\mathbf{s})$, which is the *a priori* distribution of $\mathbf{s}$, follows a continuous distribution. By minimizing (3), we can obtain $f(\mathbf{s}, \mathbf{h})$ that most closely resembles $p(\mathbf{s}, \mathbf{h}, \mathbf{r})$. However, without any restriction, we will get $f(\mathbf{s}, \mathbf{h}) = p(\mathbf{s}, \mathbf{h}, \mathbf{r})$ by optimizing (3), which means nothing. Consequently, we must confine $f(\mathbf{s}, \mathbf{h})$ to some specific form, e.g., Gaussian distribution; then, we could obtain the closed form of variational free energy ω of (3). After that, through minimizing the variational free energy over $f(\mathbf{s},\mathbf{h})$, we can get $f(\mathbf{s},\mathbf{h})$ that mostly approximates $p(\mathbf{s},\mathbf{h},\mathbf{r})$. The technique of computing approximating inference is called variational inference.

The variational technique always further simplifies the problem by assuming that $f(\mathbf{s},\mathbf{h})$ can be factored as (4). The assumption is equivalent to assume that $\mathbf{s}$ and $\mathbf{h}$ are independent conditioned on $\mathbf{r}$.

$$f(\mathbf{s},\mathbf{h}) = f_s(\mathbf{s})f_h(\mathbf{h}). \tag{4}$$

By inserting (4) into (2), we will get

$$\begin{aligned} p(\mathbf{s}|\mathbf{r}) &= \int f(\mathbf{s},\mathbf{h})d\mathbf{h} \\ &= \int f_s(\mathbf{s})f_h(\mathbf{h})d\mathbf{h} \\ &= f_s(\mathbf{s}). \end{aligned} \tag{5}$$

From (5), we can find that $\mathbf{s}^*$, which maximizes $f(\mathbf{s},\mathbf{h})$ over $\mathbf{s}$ and $\mathbf{h}$, also maximizes (2).

Based on (1), the probability of $\mathbf{r}$ conditioned on $\mathbf{s}$ and $\mathbf{h}$ can be written as

$$p(\mathbf{r}|\mathbf{s},\mathbf{h}) = \mathcal{CN}\left(\mathbf{F}^H \mathrm{diag}(\mathbf{F}_L\mathbf{h})\mathbf{s}, \sigma^2\mathbf{I}\right). \tag{6}$$

Inserting (4) into (3), we will get

$$\omega = \int f_s(\mathbf{s})f_h(\mathbf{h}) \log \frac{f_s(\mathbf{s})f_h(\mathbf{h})}{p\left(\mathbf{r}|\mathbf{s},\mathbf{h}\right)p(\mathbf{s})p(\mathbf{h})} d\mathbf{s}d\mathbf{h}. \tag{7}$$

Summing up the latter, the procedure of the algorithm for deducing the approximation distribution is shown as follows:

Step 1: Assume the parametric distribution of $p(\mathbf{s})$, $f_s(\mathbf{s})$, and $f_h(\mathbf{h})$.
Step 2: Compute the closed form of ω in (7).
Step 3: Optimize (7) over the key parameters of the distribution assumed in step 1.

The closed-form solution in step 3 is always hard to obtain so that we often resolve it using iterative methods. The parameters are updated one after another while keeping the others constant.

B. Joint Detection Via Variational Inference

Instead of discussing the discrete distribution of $\mathbf{s}$ over the constellation space, we approximate the prior distribution of it to be Gaussian. The prior distribution of $\mathbf{h}$ is also Gaussian, i.e.,

$$\begin{aligned} p(\mathbf{s}) &= \mathcal{CN}\left(\mathbf{0}, A^2\mathbf{I}_N\right) \\ p(\mathbf{h}) &= \mathcal{CN}(\mathbf{0}, \mathbf{\Phi}) \end{aligned} \tag{8}$$

where A^2 is the average power of the modulated signal, and $\mathbf{\Phi} = \mathrm{diag}(\sigma_0^2, \ldots, \sigma_{L-1}^2)$, with σ_l^2 being the power of the lth path.

Reference [11] points out that the EM algorithm can be viewed as minimizing a free energy problem. In the EM algorithm, the unknowns are divided into parameters that need hard estimate and hidden variables that need soft estimate, respectively. Obviously, we will get a family of algorithms by treating $\mathbf{s}$ and $\mathbf{h}$ as parameters or hidden data, respectively. When both of them are regarded as hidden data, we will get the scheme in [10]. The other three algorithms are shown as follows.

When the unknowns are viewed as parameters that need hard estimates, delta function will be implemented. Let $\delta(\mathbf{x},\widehat{\mathbf{x}})$ denote delta function; then for arbitrary $g(\mathbf{x})$, we have $\int \delta(\mathbf{x},\widehat{\mathbf{x}})g(\mathbf{x})d\mathbf{x} = g(\widehat{\mathbf{x}})$.

Algorithm I: When $\mathbf{s}$ and $\mathbf{h}$ are both regarded as parameters that need hard estimate, let

$$\begin{aligned} f_s(\mathbf{s}) &= \delta\left(\mathbf{s},\widehat{\mathbf{s}}\right) \\ f_h(\mathbf{h}) &= \delta(\mathbf{h},\widehat{\mathbf{h}}). \end{aligned} \tag{9}$$

Insert (6), (8) and (9) into (7), and we have

$$\begin{aligned} \omega &= -\log p(\mathbf{r},\widehat{\mathbf{s}},\widehat{\mathbf{h}}) \\ &= -\log\left(p(\widehat{\mathbf{s}})p(\widehat{\mathbf{h}})p(\mathbf{r}|\widehat{\mathbf{s}},\widehat{\mathbf{h}})\right) \\ &= \frac{1}{A^2}\widehat{\mathbf{s}}^H\widehat{\mathbf{s}} + \widehat{\mathbf{h}}^H\mathbf{\Phi}^{-1}\widehat{\mathbf{h}} + \frac{1}{\sigma^2}(\mathbf{r}-\mathbf{F}^H\widehat{\mathbf{H}}\widehat{\mathbf{s}})^H(\mathbf{r}-\mathbf{F}^H\widehat{\mathbf{H}}\widehat{\mathbf{s}}) \end{aligned} \tag{10}$$

where $\widehat{\mathbf{H}} = \mathrm{diag}(\mathbf{F}_L\widehat{\mathbf{h}})$. Obviously, the closed-form solution of (10) over $\widehat{\mathbf{s}}$ and $\widehat{\mathbf{h}}$ is difficult to obtain. The general practice of the solution is to obtain the unknowns iteratively while keeping the others constant [15]. Reference [11] points out that the schemes converge to local optimal at least. Take the differential over $\widehat{\mathbf{s}}$ and $\widehat{\mathbf{h}}$, and let them be equal to zero, as derived in the Appendix, we will get

$$\widehat{\mathbf{s}} = \left(\frac{\sigma^2}{A^2}\mathbf{I}_N + \widehat{\mathbf{H}}^H\widehat{\mathbf{H}}\right)^{-1}\widehat{\mathbf{H}}\mathbf{F}\mathbf{r} \tag{11}$$

$$\widehat{\mathbf{h}} = \left(\sigma^2\mathbf{\Phi}^{-1} + \mathbf{F}_L^H\mathrm{diag}(\widehat{\mathbf{s}}^*)\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\right)^{-1}\mathbf{F}_L^H\mathrm{diag}(\widehat{\mathbf{s}}^*)\mathbf{F}\mathbf{r}. \tag{12}$$

The two parameters are updated in turn. The estimate of $\mathbf{s}$ obtained in the last iteration can be used to make a hard decision. In fact, since (11) and (12) have the form of MMSE estimator, Algorithm I can be viewed as an iterative MMSE estimator.

Algorithm II: Let $\mathbf{s}$ be hidden data and $\mathbf{h}$ be parameter with

$$\begin{aligned} f_s(\mathbf{s}) &= \mathcal{CN}(\mathbf{E}_s, \varepsilon_s) \\ f_h(\mathbf{h}) &= \delta(\mathbf{h},\widehat{\mathbf{h}}). \end{aligned} \tag{13}$$

Insert (6), (8) and (13) into (7), and we have

$$\begin{aligned} \omega = &\frac{1}{\sigma^2}\left[\mathrm{tr}(\mathbf{F}^H\widehat{\mathbf{H}}\varepsilon_s\widehat{\mathbf{H}}^H\mathbf{F}) + (\mathbf{F}^H\widehat{\mathbf{H}}\mathbf{E}_s-\mathbf{r})^H(\mathbf{F}^H\widehat{\mathbf{H}}\mathbf{E}_s-\mathbf{r})\right] \\ &-\log|\varepsilon_s| + \frac{1}{A^2}\left(\mathrm{tr}(\varepsilon_s) + \mathbf{E}_s^H\mathbf{E}_s\right) + \widehat{\mathbf{h}}^H\mathbf{\Phi}^{-1}\widehat{\mathbf{h}}. \end{aligned} \tag{14}$$

Take the differential over $\boldsymbol{\varepsilon}_s$, $\mathbf{E}_s$ and $\widehat{\mathbf{h}}$, and let them be zero, as derived in the Appendix. Then, we will obtain

$$\varepsilon_s = \left(\frac{1}{A^2}\mathbf{I}_N + \frac{1}{\sigma^2}\widehat{\mathbf{H}}^T\widehat{\mathbf{H}}^*\right)^{-T} \tag{15}$$

$$\mathbf{E}_s = \frac{1}{\sigma^2}\left(\frac{1}{A^2}\mathbf{I}_N + \frac{1}{\sigma^2}\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}\right)^{-1}\widehat{\mathbf{H}}^H\mathbf{F}\mathbf{r} \tag{16}$$

$$\begin{aligned} \widehat{\mathbf{h}} = &\left(\sigma^2\mathbf{\Phi}^{-1} + \mathbf{F}_L^H\mathrm{diag}(\varepsilon_s)\mathbf{F}_L\right. \\ &\left.+\mathbf{F}_L^H\mathrm{diag}^H(\mathbf{E}_s)\mathrm{diag}(\mathbf{E}_s)\mathbf{F}_L\right)^{-1}\mathbf{F}_L^H\mathrm{diag}^H(\mathbf{E}_s)\,\mathbf{F}\mathbf{r}. \end{aligned} \tag{17}$$

Algorithm III: Let $\mathbf{h}$ be hidden data and $\mathbf{s}$ be a parameter, so that

$$\begin{aligned} f_s(\mathbf{s}) &= \delta(\mathbf{s},\widehat{\mathbf{s}}) \\ f_h(\mathbf{h}) &= \mathcal{CN}(\mathbf{E}_h, \varepsilon_h). \end{aligned} \tag{18}$$

Insert (6), (8) and (18) into (7), and we will obtain

$$\begin{aligned} \omega = &\frac{1}{\sigma^2}\left[\mathrm{tr}\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\varepsilon_h\mathbf{F}_L^H\mathrm{diag}^H(\widehat{\mathbf{s}})\mathbf{F}\right)\right. \\ &\left.+\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\mathbf{E}_h-\mathbf{r})\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}}\right)\mathbf{F}_L\mathbf{E}_h-\mathbf{r})\right]^H \\ &+\mathbf{E}_h^H\mathbf{\Phi}^{-1}\mathbf{E}_h - \log|\varepsilon_h| + \mathrm{tr}(\Phi^{-1}\varepsilon_h). \end{aligned} \tag{19}$$

Take the differential over $\boldsymbol{\varepsilon}_h$, $\mathbf{E}_h$, and $\widehat{\mathbf{s}}$ and let them be zero, as derived in the Appendix. Then, we can obtain

$$\boldsymbol{\varepsilon}_h = \left(\boldsymbol{\Phi}^{-1} + \frac{1}{\sigma^2}\mathbf{F}_L^H \mathrm{diag}^H(\widehat{\mathbf{s}})\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\right)^{-T)} \tag{20}$$

$$\mathbf{E}_h = \left(\sigma^2\boldsymbol{\Phi}^{-1} + \mathbf{F}_L^H \mathrm{diag}^H(\widehat{\mathbf{s}})\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\right)^{-1}\mathbf{F}_L^H \mathrm{diag}^H(\widehat{\mathbf{s}})\mathbf{Fr} \tag{21}$$

$$\widehat{\mathbf{s}} = \left(\mathrm{diag}(\mathbf{F}_L\boldsymbol{\varepsilon}_h\mathbf{F}_L^H) + \mathrm{diag}^H(\mathbf{F}_L\mathbf{E}_h)\mathrm{diag}(\mathbf{F}_L\mathbf{E}_h)\right)^{-1} \times \mathrm{diag}^H(\mathbf{F}_L\mathbf{E}_h)\mathbf{Fr}. \tag{22}$$

C. Complexity Analysis

Equation (11) of Algorithm I involves an inversion of an $N \times N$ diagonal matrix $\widehat{\mathbf{s}} = ((\sigma^2/A^2)\mathbf{I}_N + \widehat{\mathbf{H}}^H\widehat{\mathbf{H}})^{-1}$ and multiplication of a diagonal matrix $\widehat{\mathbf{H}}$ and the fast Fourier transform (FFT) of $\mathbf{r}$. Therefore, the complexity of calculating $\widehat{\mathbf{s}}$ is $O(NlogN)$. Based on the property of a circulant matrix, a circulant matrix can be diagonalized by pre- and postmultiplication with the FFT and inverse FFT matrices [16]. Hence, $\mathbf{F}_L^H\mathrm{diag}(\widehat{\mathbf{s}}^*)\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L)^{-1}$ in (12) can be derived efficiently using the FFT, leading to complexity of $O(NlogN)$. Furthermore, we can conclude directly that the total complexity of (12) is O_c, where $O_c = \max(NlogN, L^3)$, so that the complexity of Algorithm I is $(NlogN + O_c)$ in one iteration. Similarly, the complexity of Algorithms II and III in one iteration is $(N + NlogN + O_c)$ and $(3O_c)$, respectively. We can also find that the complexity of the scheme of [10] is $(4O_c)$. On the other hand, the MMSE estimation of data by assuming that the CSI is accurate is $\widehat{\mathbf{s}} = ((\sigma^2/A^2)\mathbf{I}_N + \widehat{\mathbf{H}}^H\widehat{\mathbf{H}})^{-1}\widehat{\mathbf{H}}\mathbf{Fr}$, of which the complexity is $O(NlogN)$. Futhermore, we can say that the algorithm of [10] and the MMSE algorithm have the highest and lowest complexity, respectively. The proposed methods of this paper provide several options with lower complexity compared with [10].

D. Channel Estimation Error Model Used for Initialization of the Algorithms

Initial imperfect channel estimation must be given before executing the proposed algorithms. Many published articles addressed the problem of channel estimation for OFDM systems, e.g., [13]. This paper does not take details of channel estimation algorithms into account. We mainly pay attention to additive channel estimation errors. Other models will be further studied for future work. The channel fade v_i, channel estimate $\widehat{v}_i$, and the additive Gaussian channel estimation error u_i at the ith, $0 \le i \le N-1$, subcarrier are all assumed to be complex Gaussian random variables with zero mean and can be written as $\widehat{v}_i = v_i + u_i$. The real and imaginary parts of v_i are assumed to be uncorrelated and both with variance $\sigma_v^2/2$. u_i is independent of v_i. Let ρ be the correlation coefficient between channel response and its estimation. ρ can be written as $\rho = (\mathbf{E}[v_i\widehat{v}_i^*]/\sqrt{\mathbf{E}[v_iv_i^*]\mathbf{E}[\widehat{v}_i\widehat{v}_i^*]} = (\sigma_v/\sqrt{\sigma_v^2+\sigma_u^2})$. The real and imaginary parts of u_i are assumed to be uncorrelated and both with variance $\sigma_u^2/2$. Obviously, $\sigma_u^2 = \sigma_v^2(1/\rho^2 - 1)$.

IV. Simulations

To illustrate the performance of the algorithms, simulations are performed here. The system parameters are as follows. The number of subcarriers is 256, and the length of the prefix is 32. A six-path Rayleigh fading multipath wireless channel model is used. The power delay profile follows, exponentially decreasing with a decay constant of four taps. The baseband sampling rate is 1 MHz. The modulation is QPSK.

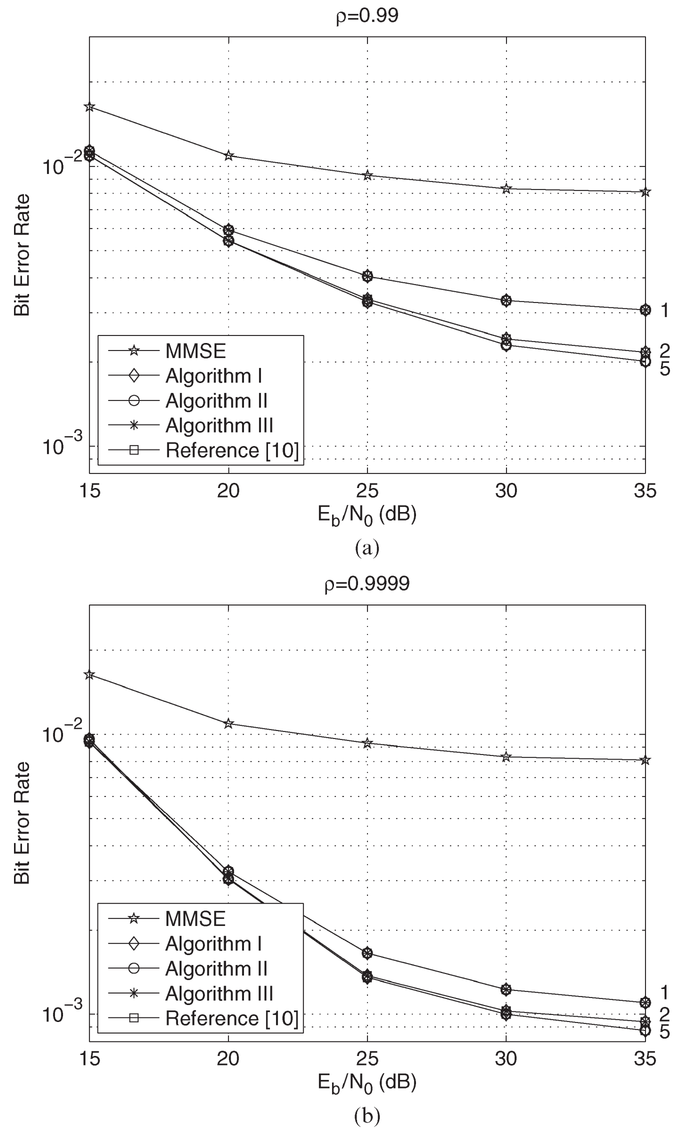

Fig. 1. BER performance comparison of the proposed method with existing methods.

At the initial step, we ignore the channel estimation error and obtain the estimate of $\mathbf{s}$ using MMSE, which is the initial value of $\mathbf{s}$. The initial value of $\boldsymbol{\varepsilon}_s$ and $\boldsymbol{\varepsilon}_h$ is set to be $\mathbf{0}$.

The bit error rate (BER) and SNR performance of the three algorithms of this paper, [10] and MMSE [17], [18] are demonstrated in Fig. 1. Fig. 1(a) and (b) is obtained with initial correlation coefficients $\rho = 0.99$ and $\rho = 0.9999$, respectively. The BER curves with one, two, and five iterations are shown. We find that, when the number of iterations is larger than 5, there is no significant improvement of system performance when the number of iterations increases. From the simulation result, It can be seen that the performance of the four variational-inference-based methods outperforms the traditional MMSE method by 1-dB SNR gain at least. It can be seen that the performance of Algorithms I, II, and III is identical to that in [10]. The reason is that we only consider uncoded system here, and the variance is not fully used. When the unknown parameter is viewed as hidden data, we can obtain the variance of its estimate, which can be viewed as the reliability of the estimate when Gaussian approximation is implemented in coded systems.

V. Conclusion

We have studied the detection problem for OFDM with additive channel estimate errors. We propose three low-complex detection methods based on variational inference. The data and CSI are estimated jointly and iteratively. For uncoded systems, the proposed algorithms have the same performance as [10] with lower complexity. The algorithms with the unknown parameters being viewed as hidden data are more attractive in coded systems since soft information provided by the variance of the estimate can be used in a turbo equalizer. Turbo detection for coded OFDM systems will be studied in future work.

Appendix
Derivation of the Algorithms

Algorithm I: Based on (10), we have two equations as follows.

For $\widehat{\mathbf{s}}$, we have

$$\begin{aligned}\partial\omega/\partial\widehat{\mathbf{s}}^* &= \frac{\partial}{\partial\widehat{\mathbf{s}}^*}\left(\frac{1}{A^2}\widehat{\mathbf{s}}^H\widehat{\mathbf{s}} + \frac{1}{\sigma^2}(\mathbf{F}^H\widehat{\mathbf{H}}\widehat{\mathbf{s}}-\mathbf{r})^H(\mathbf{F}^H\widehat{\mathbf{H}}\widehat{\mathbf{s}}-\mathbf{r})\right)\\ &= \frac{1}{A^2}\widehat{\mathbf{s}} + \frac{1}{\sigma^2}\widehat{\mathbf{H}}^H\mathbf{F}(\mathbf{F}^H\widehat{\mathbf{H}}\widehat{\mathbf{s}}-\mathbf{r}).\end{aligned}$$

Therefore, $\partial\omega/\partial\widehat{\mathbf{s}}^* = 0$ leads to (11).

For $\widehat{\mathbf{h}}$, we have

$$\begin{aligned}\partial\omega/\partial\widehat{\mathbf{h}}^* &= \frac{\partial}{\partial\widehat{\mathbf{h}}^*}\left(\widehat{\mathbf{h}}^H\mathbf{\Phi}^{-1}\widehat{\mathbf{h}} + \frac{1}{\sigma^2}(\mathbf{F}^H\widehat{\mathbf{H}}\widehat{\mathbf{s}}-\mathbf{r})^H(\mathbf{F}^H\widehat{\mathbf{H}}\widehat{\mathbf{s}}-\mathbf{r})\right)\\ &= \frac{\partial}{\partial\widehat{\mathbf{h}}^*}\left(\widehat{\mathbf{h}}^H\mathbf{\Phi}^{-1}\widehat{\mathbf{h}} + \frac{1}{\sigma^2}\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\widehat{\mathbf{h}}-\mathbf{r}\right)^H\right.\\ &\qquad\left.\times\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\widehat{\mathbf{h}}-\mathbf{r}\right)\right)\\ &= \mathbf{\Phi}^{-1}\widehat{\mathbf{h}} + \frac{1}{\sigma^2}\mathbf{F}_L^H\mathrm{diag}(\widehat{\mathbf{s}}^*)\mathbf{F}\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\widehat{\mathbf{h}}-\mathbf{r}\right).\end{aligned} \tag{23}$$

Thus, let $\partial\omega/\partial\widehat{\mathbf{h}}^* = 0$, we will get (12).

Algorithm II: Based on (14), we have three equations as follows.

For $\boldsymbol{\varepsilon}_s$. we have

$$\begin{aligned}\partial\omega/\partial\varepsilon_s^{-1} &= \frac{\partial}{\partial\varepsilon_s^{-1}}\left\{-\log|\varepsilon_s| + \frac{1}{A^2}\left(\mathrm{tr}(\varepsilon_s) + \mathbf{E}_s^H\mathbf{E}_s\right)\right.\\ &\qquad\left.-\frac{1}{\sigma^2}\mathrm{tr}(\mathbf{F}^H\widehat{\mathbf{H}}\varepsilon_s\widehat{\mathbf{H}}^H\mathbf{F})\right\}\\ &= \varepsilon_s^T - \frac{1}{A^2}\varepsilon_s^T\varepsilon_s^T - \frac{1}{\sigma^2}(\varepsilon_s\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}\varepsilon_s)^T.\end{aligned}$$

Therefore, let $\partial\omega/\partial\varepsilon_s^{-1} = 0$; then, we will obtain (15).

For $\mathbf{E}_s$, we have

$$\begin{aligned}\partial\omega/\partial\mathbf{E}_s^* &= \frac{\partial}{\partial\mathbf{E}_s^*}\left\{\frac{1}{A^2}\mathbf{E}_s^H\mathbf{E}_s + \frac{1}{\sigma^2}(\mathbf{F}^H\widehat{\mathbf{H}}\mathbf{E}_s-\mathbf{r})^H(\mathbf{F}^H\widehat{\mathbf{H}}\mathbf{E}_s-\mathbf{r})\right\}\\ &= \frac{1}{A^2}\mathbf{E}_s + \frac{1}{\sigma^2}\widehat{\mathbf{H}}^H\mathbf{F}(\mathbf{F}^H\widehat{\mathbf{H}}\mathbf{E}_s-\mathbf{r}).\end{aligned}$$

Therefore, let $\partial\omega/\partial\mathbf{E}_s^* = 0$, we will obtain (16).

For $\widehat{\mathbf{h}}$, we have

$$\begin{aligned}\partial\omega/\partial\widehat{\mathbf{h}}^* &= \frac{\partial}{\partial\widehat{\mathbf{h}}^*}\left\{\widehat{\mathbf{h}}^H\mathbf{\Phi}^{-1}\widehat{\mathbf{h}} + \frac{1}{\sigma^2}\right.\\ &\qquad\times\left[\mathrm{tr}(\mathbf{F}^H\widehat{\mathbf{H}}\boldsymbol{\varepsilon}_s\widehat{\mathbf{H}}^H\mathbf{F})\right.\\ &\qquad\left.\left.+(\mathbf{F}^H\widehat{\mathbf{H}}\mathbf{E}_s-\mathbf{r})(\mathbf{F}^H\widehat{\mathbf{H}}\mathbf{E}_s-\mathbf{r})^H\right]\right\}\\ &= \mathbf{\Phi}^{-1}\widehat{\mathbf{h}} + \frac{1}{\sigma^2}\left[\mathbf{F}_L^H\mathrm{diag}(\boldsymbol{\varepsilon}_s)\mathbf{F}_L\widehat{\mathbf{h}}\right.\\ &\qquad\left.+\mathbf{F}_L^H\mathrm{diag}^H(\mathbf{E}_s)\mathbf{F}(\mathbf{F}^H\widehat{\mathbf{H}}\mathbf{E}_s-\mathbf{r})\right].\end{aligned}$$

Therefore, let $\partial\omega/\partial\widehat{\mathbf{h}}^* = 0$; then, we will obtain (17).

Algorithm III: Based on (19), we have three equations as follows.

For $\boldsymbol{\varepsilon}_h$, we have

$$\begin{aligned}&\partial\omega/\partial\boldsymbol{\varepsilon}_h^{-1}\\ &= \frac{\partial}{\partial\boldsymbol{\varepsilon}_h^{-1}}\left[\frac{1}{\sigma^2}\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\mathbf{E}_h-\mathbf{r}\right)^H\right.\\ &\qquad\times\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\mathbf{E}_h-\mathbf{r}\right)+\mathrm{tr}(\mathbf{\Phi}^{-1}\boldsymbol{\varepsilon}_h)-\log|\boldsymbol{\varepsilon}_h|\\ &\qquad\left.+\frac{1}{\sigma^2}\mathrm{tr}\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\boldsymbol{\varepsilon}_h\mathbf{F}_L^H\mathrm{diag}^H(\widehat{\mathbf{s}})\mathbf{F}\right)\right]\\ &= \boldsymbol{\varepsilon}_h^T - \boldsymbol{\varepsilon}_h^T\mathbf{\Phi}^{-1}\boldsymbol{\varepsilon}_h^T - \frac{1}{\sigma^2}\boldsymbol{\varepsilon}_h\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\right)^H\\ &\quad\times\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\right)\boldsymbol{\varepsilon}_h.\end{aligned}$$

Therefore, let $\partial\omega/\partial\boldsymbol{\varepsilon}_h^{-1} = 0$; then, we will obtain (20).

For $\mathbf{E}_h$, we have

$$\begin{aligned}\partial\omega/\partial\mathbf{E}_h^* &= \frac{\partial}{\partial\mathbf{E}_h^*}\left(\mathbf{E}_h^H\mathbf{\Phi}^{-1}\mathbf{E}_h + \frac{1}{\sigma^2}\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\mathbf{E}_h-\mathbf{r}\right)^H\right.\\ &\qquad\left.\times\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\mathbf{E}_h-\mathbf{r}\right)\right)\\ &= \mathbf{\Phi}^{-1}\mathbf{E}_h + \frac{1}{\sigma^2}\mathbf{F}_L^H\mathrm{diag}^H(\widehat{\mathbf{s}})\mathbf{F}\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\mathbf{E}_h-\mathbf{r}\right)\end{aligned}$$

Therefore, let $\partial\omega/ = \partial\mathbf{E}_h^* = 0$; then, we will obtain (21). Then, we have

$$\begin{aligned}\partial\omega/\partial\widehat{\mathbf{s}}^* &= \frac{\partial}{\partial\widehat{\mathbf{s}}^*}\left\{\frac{1}{\sigma^2}\mathrm{tr}\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\boldsymbol{\varepsilon}_h\mathbf{F}_L^H\mathrm{diag}^H(\widehat{\mathbf{s}})\mathbf{F}\right)\right.\\ &\qquad+\frac{1}{\sigma^2}\times\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\mathbf{E}_h-\mathbf{r}\right)^H\\ &\qquad\left.\times\left(\mathbf{F}^H\mathrm{diag}(\widehat{\mathbf{s}})\mathbf{F}_L\mathbf{E}_h-\mathbf{r}\right)\right\}\\ &= \frac{\partial}{\partial\widehat{\mathbf{s}}^*}\left\{\frac{1}{\sigma^2}\widehat{\mathbf{s}}^H\mathrm{diag}\left(\mathbf{F}_L\boldsymbol{\varepsilon}_h\mathbf{F}_L^H\right)\widehat{\mathbf{s}} + \frac{1}{\sigma^2}\right.\\ &\qquad\times\left(\mathbf{F}^H\mathrm{diag}(\mathbf{F}_L\mathbf{E}_h)\widehat{\mathbf{s}}-\mathbf{r}\right)^H\\ &\qquad\left.\times\left(\mathbf{F}^H\mathrm{diag}(\mathbf{F}_L\mathbf{E}_h)\widehat{\mathbf{s}}-\mathbf{r}\right)\right\}\\ &= \frac{1}{\sigma^2}\mathrm{diag}^H(\mathbf{F}_L\mathbf{E}_h)\mathbf{F}\left(\mathbf{F}^H\mathrm{diag}(\mathbf{F}_L\mathbf{E}_h)\widehat{\mathbf{s}}-\mathbf{r}\right)\\ &\quad+\frac{1}{\sigma^2}\mathrm{diag}\left(\mathbf{F}_L\boldsymbol{\varepsilon}_h\mathbf{F}_L^H\right)\widehat{\mathbf{s}}.\end{aligned}$$

Therefore, let $\partial\omega/\partial\widehat{\mathbf{s}}^* = 0$, we will obtain (22).

Acknowledgment

The authors would like to thank the editor and the anonymous reviewers for their valuable suggestions.

References

[1] T. Hwang, C. Yang, G. Wu, S. Li, and G. Li, "OFDM and its wireless applications: A survey," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1673–1694, May 2009.

[2] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York: Wiley, 2003, ser. Probability and Statistics.

[3] R. Annavajjala, P. C. Cosman, and L. B. Milstein, "Performance analysis of linear modulation schemes with generalized diversity combining on rayleigh fading channels with noisy channel estimates," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4701–4727, Dec. 2007.

[4] H. Song, S. Kang, M. Kim, and Y. You, "Error performance analysis of STBC-OFDM systems with parameter imbalances," *IEEE Trans. Broadcast.*, vol. 50, no. 1, pp. 76–82, Mar. 2004.

[5] M. Krondorf and G. Fettweis, "Numerical performance evaluation for alamouti space time coded OFDM under receiver impairments," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1446–1455, Mar. 2009.

[6] B. J. Frey, *Graphical Models for Machine Learning and Digital Communication*. Cambridge, MA: MIT Press, 1998.

[7] D. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[8] D. D. Lin and T. J. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1862–1874, May 2007.

[9] S. Stefanatos and A. K. Katsaggelos, "Joint data detection and channel tracking for OFDM systems with phase noise," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4230–4243, Sep. 2008.

[10] F. Li, S. Zhu, and M. Rong, "Detection for OFDM systems with channel estimation errors using variational inference," *IEEE Signal Process. Lett.*, vol. 16, no. 5, pp. 434–437, May 2009.

[11] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in Graphical Models*, M. I. Jordan, Ed. Norwell, MA: Kluwer, 1998.

[12] M. Tüchler, "Turbo Equalization: An Overview," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 920–952, Feb. 2011.

[13] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *Commun. Surveys Tuts.*, vol. 9, no. 2, pp. 18–48, Second Quarter, 2007.

[14] S. M. Kay, *Fundamentals of Statistical Signal Processing—Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993, ch. 10.7, pp. 328–330.

[15] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999.

[16] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[17] M. A. Gharabally and P. Das, "On the performance of OFDM systems in time varying channels with channel estimation error," in *Proc. IEEE ICC*, Jun. 2006, vol. 11, pp. 5180–5185.

[18] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3065–3073, Dec. 2001.

Cooperative Relaying in Next-Generation Mobile WiMAX Networks

Chun Nie, Pei Liu, Thanasis Korakis, Elza Erkip, and Shivendra S. Panwar

***Abstract*—Distributed space–time coding (DSTC) is a key physical (PHY) layer technique to enable cooperative relaying in wireless networks. Compared with direct transmission and single-relay cooperation, DSTC exploits spatial diversity gain and achieves a higher end-to-end throughput performance. In particular, a novel variation of DSTC, which is called randomized DSTC (R-DSTC), further enhances the performance of conventional DSTC in a mobile environment due to its decentralized relay recruitment. This paper presents a medium access control (MAC) protocol, which is called *CoopMAX*, for Worldwide Interoperability for Microwave Access (WiMAX) systems to provide cooperative relaying with DSTC and R-DSTC techniques by the use of fixed and mobile relays. *CoopMAX* explores a joint PHY layer and MAC layer optimization for rate adaptation. The efficiency of *CoopMAX* is investigated, showing pronounced advantages, such as a doubling of throughput as compared with conventional WiMAX protocols in some typical scenarios.**

***Index Terms*—Cooperative relaying, IEEE 802.16, medium access control (MAC), mobility, throughput, Worldwide Interoperability for Microwave Access (WiMAX).**

I. Introduction

The Worldwide Interoperability for Microwave Access (WiMAX) system has attracted great research attention and been standardized in [2]. The latest IEEE 802.16m WiMAX standard aims at developing advanced techniques to support higher system capacity and extended cell coverage. To provision high throughput for subscribers at the cell edge, the concept of mobile multihop relaying (MMR) [3] defines a framework for multihop transmission. In an MMR network, the conventional low-speed single-hop transmission is replaced by relay-assisted high-speed multihop transmission between the source and the destination. At each hop, intermediate fixed nodes, which are called relay stations (RSs), can forward the signal. The functionalities of an RS in a WiMAX system include relaying the data packets and signaling messages [3]. While the use of a single RS can increase the end-to-end throughput, it suffers in a fading environment, where the RS may fail to decode the source signal and, thus, cannot deliver it to the destination. WiMAX allows multiple RSs [3], coordinated using a distributed space–time code (DSTC) [4], to provide a higher spatial diversity gain. By selecting an optimal set of relays in a centralized manner, DSTC can outperform the direct transmission and single-RS cooperation schemes. However, DSTC performance is degraded in a mobile environment, where coordinating distributed relays becomes

Manuscript received December 31, 2011; revised May 20, 2012 and September 17, 2012; accepted October 23, 2012. Date of publication November 29, 2012; date of current version March 13, 2013. This work was supported in part by the National Science Foundation (NSF) under Grant CNS-0520054 and Grant CNS-0905446, by the New York State Center for Advanced Technology in Telecommunications (CATT), and by the Wireless Internet Center for Advanced Technology (WICAT), an NSF Industry/University Cooperative Research Center. This paper was presented in part at the IEEE International Conference on Communications (ICC 2009), Dresden, Germany, June 14–18, 2009. The review of this paper was coordinated by Prof. W. A. Hamouda.

The authors are with the Polytechnic Institute of New York University, Brooklyn, NY 11201 USA (e-mail: cnie01@students.poly.edu; peiliu@gmail.com; nasoskor@gmail.com; elza@poly.edu; panwar@catt.poly.edu).

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

Digital Object Identifier 10.1109/TVT.2012.2230281