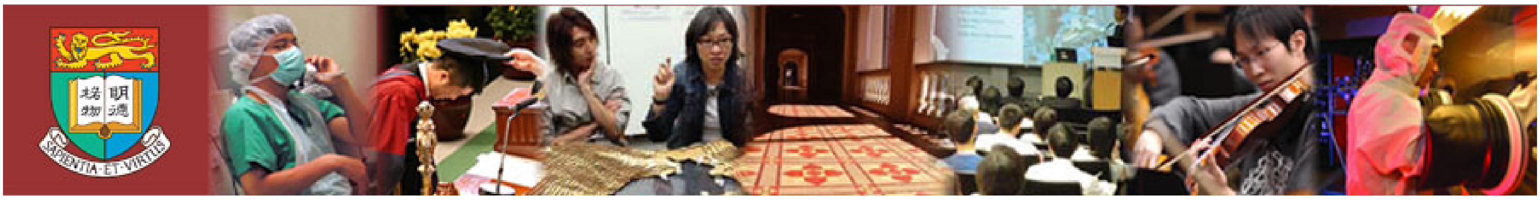

Title	**Variational Inference-based Joint Interference Mitigation and OFDM Equalization Under High Mobility**
Author(s)	**Zhou, J; Qin, J; Wu, YC**
Citation	**IEEE Signal Processing Letters, 2015, v. 22, p. 1970-1974**
Issued Date	**2015**
URL	**http://hdl.handle.net/10722/214165**
Rights	**Creative Commons: Attribution 3.0 Hong Kong License**

Variational Inference-based Joint Interference Mitigation and OFDM Equalization Under High Mobility

Jingrong Zhou, Jiayin Qin, and Yik-Chung Wu

***Abstract*—In OFDM-based spectrum sharing networks, due to inefficient coordination or imperfect spectrum sensing, the signals from femtocells or secondary users appear as interference in a subset of subcarriers of the primary systems. Together with the inter-carrier interference (ICI) introduced by high mobility, equalizing one subcarrier now depends not only on whether interference exists, but also the neighboring subcarrier data. In this letter, we propose a novel approach to iteratively learn the statistics of noise plus interference across different subcarriers, and refine the soft data estimates of each subcarrier based on the variational inference. Simulation results show that the proposed method avoids the error floor effect, which is exhibited by existing algorithms without considering interference mitigation, and performs close to the ideal case with perfect ICI cancelation and knowledge of noise plus interference powers for optimal maximum a posteriori probability (MAP) equalizer.**

***Index Terms*—High mobility, interference mitigation, orthogonal frequency division multiplexing (OFDM), variational inference.**

I. Introduction

HAVING the flexibility in allocating subcarriers to exploit the spatial-temporal spectrum holes, OFDM is a promising transmission scheme used in many spectrum sharing networks to increase spectrum efficiency, e.g., macro-femto networks [1] and cognitive radio [2]. Due to imperfect coordination or spectrum sensing, the signals from femtocells or cognitive users would appear as interference in a subset of subcarriers in the received OFDM signal of the primary systems in a random manner, i.e., with unknown positions and unknown interference powers. If the interference happens to fall on data subcarriers, it will have great impact on the bit error rate (BER) performance. For OFDM systems over static multipath fading channel (i.e., low user mobility), one straightforward way to mitigate the interference is to identify the interference frequencies, and erase the interfered symbols [3], [4]. Since equalization of one subcarrier is performed independently of other subcarriers, the erasure errors do not affect equalizer outputs at other subcarriers and can be corrected by channel decoder. Another approach, which does not explicitly identify interference locations, is to treat the interference as part of the noise, and a joint noise powers estimation and decoding problem is solved using the expectation maximization (EM) algorithm [5].

However, wireless standards such as WiMAX [6] and LTE [7] require support of high mobility users, e.g., high speed vehicles with speed as fast as 350 km/h [6], [7]. Different from low user mobility case, the channel changes sample by sample in the time domain within one OFDM symbol, and becomes both frequency selective and time selective, a.k.a. doubly selective. Due to the inter-carrier interference (ICI) introduced by doubly selective channel, equalization becomes complicated, as equalizing one subcarrier now depends on the accuracy of neighboring subcarrier estimated data. Therefore, OFDM equalization under ICI becomes an important issue [8]–[13]. Unfortunately, they become ineffective in the presence of unknown interference, since the interference will introduce unpredictable errors to data estimates. This results in the need of joint consideration of interference mitigation and OFDM equalization under high mobility.

In this letter, we solve the above problem from Bayesian perspective, and propose a novel approach under the variational inference framework [14], which automatically leads to an iterative algorithm between estimation of noise plus interference powers at different subcarriers and successive soft data estimation at each subcarrier. Simulation results show that while existing algorithms without considering interference mitigation show significant error floors in BER, the proposed algorithm avoids the error floor due to joint interference mitigation and equalization. Furthermore, the proposed method performs close to the match-filtered bound, which assumes perfect ICI cancelation using perfect data at adjacent subcarriers, and perfect knowledge of noise plus interference powers for optimal MAP equalizer.

Notations: Superscripts H, T and $*$ denote Hermitian, transpose and conjugate, respectively. $|\mathcal{S}|$ represents the number of elements in the set $\mathcal{S}$. $\mathbf{e}_l^N$ denotes the length-of-N column vector with the lth element being 1, and 0 otherwise. $\mathrm{diag}\{\mathbf{x}\}$ stands for the diagonal matrix with vector $\mathbf{x}$ on its diagonal. $\mathbf{x}_{\setminus n}$ denotes the vector by deleting the nth element of $\mathbf{x}$, and $\mathbf{X}_{:,\setminus n}$ represent the matrix by deleting the nth column vector of matrix $\mathbf{X}$. $\mathbb{E}_{p(\cdot)}\{\cdot\}$ denotes expectation with respect to distribution $p(\cdot)$, while $\mathrm{Tr}\{\mathbf{X}\}$ and $\det\{\mathbf{X}\}$ are trace and determinant of matrix $\mathbf{X}$.

Manuscript received March 17, 2015; revised May 20, 2015; accepted June 19, 2015. Date of publication June 24, 2015; date of current version July 07, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61472458, Grant 61202498, and Grant 61173148, in part by Guangdong Natural Science Foundation under Grant 2014A030311032, 2014A030313111, and 2014A030310374, and in part by the Fundamental Research Funds for the Central Universities under Grant 15lgzd10 and Grant 15lgpy15. The work of J. Zhou was performed while he was a visitor at The University of Hong Kong. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sanghoon Lee.

J. Zhou and J. Qin are with the School of Information Science and Technology, Sun Yat-Sen University, Guangzhou, China (e-mail: jrzhou@eee.hku.hk; issqjy@mail.sysu.edu.cn).

Y.-C. Wu is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: ycwu@eee.hku.hk).

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/LSP.2015.2449658

Finally, $\mathbf{F}$ is the discrete Fourier transform (DFT) matrix with $[\mathbf{F}]_{m,n} = \frac{1}{\sqrt{N}} e^{-j2\pi mn/N}$.

II. System Model

In this paper, we consider a single-input single-output OFDM system. In the tth OFDM symbol, the frequency-domain source data is given by $\mathbf{x}^t = [x_0^t, \ldots, x_{N-1}^t]^T$, where N is the number of subcarriers. Before transmission, $\mathbf{x}^t$ is transformed to the time-domain data $\mathbf{F}^H \mathbf{x}^t$, and a cyclic prefix (CP) is added at the beginning of the OFDM symbol to prevent intersymbol interference (ISI). The channel is assumed to be doubly selective with channel length L_h. Let $\boldsymbol{h}_l^t = [h^t(0,l), \ldots, h^t(N-1,l)]^T$ be the lth tap channel corresponding to the tth OFDM symbol, where $h^t(n,l)$ is the channel coefficient at the nth sampling time index.

At the receiver side, after CP removal and DFT transformation, the frequency-domain received signal vector $\mathbf{y}^t = [y_0^t, \ldots, y_{N-1}^t]^T$ of the tth OFDM symbol can be written in matrix form as

$$\mathbf{y}^t = \mathbf{D}^t \mathbf{x}^t + \mathbf{w}^t, \tag{1}$$

where $\mathbf{w}^t = [w_0^t, \ldots, w_{N-1}^t]^T$ represents the frequency-domain noise vector with w_n^t being thermal noise plus any possible interference at the nth subcarrier of the tth OFDM symbol, and $\mathbf{D}^t = \mathbf{F}\mathbf{H}^t\mathbf{F}^H$ is the frequency-domain channel matrix. $\mathbf{H}^t = \sum_{l=0}^{L_h-1} \mathrm{diag}\{\boldsymbol{h}_l^t\}\mathbf{P}(l)$ is an $N \times N$ time-domain channel matrix with circular convolution structure of a time-varying channel with length L_h [15], where $\mathbf{P}(l) = [\mathbf{e}_l^N, \ldots, \mathbf{e}_{N-1}^N, \mathbf{e}_0^N, \ldots, \mathbf{e}_{l-1}^N]$. With a time-varying channel, $\mathbf{D}^t$ is no longer diagonal, thus ICI is inevitable [15].

For the noise component, conditioned on the presence of interference or not, w_n^t is either zero-mean complex Gaussian with variance $\sigma^2 + (\varsigma_n^t)^2$ or σ^2, where σ^2 is the thermal noise power and $(\varsigma_n^t)^2$ represents the interference power at the nth subcarrier [4], [16]. Since υ^2 and $(\varsigma_n^t)^2$ are both unknown, w_n^t can be modeled as complex Gaussian with zero mean and unknown variance $(\varrho_n^t)^2$, which might vary with each subcarrier. This model has been widely used in channel estimation [17] and frame detection [18] in OFDM systems with unknown interference.

In addition, w_n^t is assumed to be statistically independent across subcarriers. For ease of expression, let $\beta_n^t = 1/(\varrho_n^t)^2$ denoting the inverse variance and $\boldsymbol{\beta}^t = [\beta_0^t, \ldots, \beta_{N-1}^t]^T$, then the likelihood function of $\mathbf{y}^t$ is given by

$$p(\mathbf{y}^t|\mathbf{x}^t, \boldsymbol{\beta}^t) = \frac{\exp\{-\mathrm{Tr}\{\mathrm{diag}\{\boldsymbol{\beta}^t\}(\mathbf{y}^t - \mathbf{D}^t\mathbf{x}^t)(\mathbf{y}^t - \mathbf{D}^t\mathbf{x}^t)^H\}\}}{\pi^N \det\{\mathrm{diag}\{\boldsymbol{\beta}^t\}\}^{-1}}, \tag{2}$$

where the trace property $\mathbf{a}^H\mathbf{A}\mathbf{a} = \mathrm{Tr}\{\mathbf{A}\mathbf{a}\mathbf{a}^H\}$ is applied.

III. Variational Inference-based Joint Interference Mitigation and OFDM Equalization

It can be observed from (2) that in addition to the discrete data $\mathbf{x}^t$, the noise statistics $\boldsymbol{\beta}^t$ is also unknown. From Bayesian theory, estimating $\mathbf{x}^t$ requires marginalizing the posterior probability density function (pdf) $p(\mathbf{x}^t, \boldsymbol{\beta}^t|\mathbf{y}^t)$ to obtain marginal posterior pdf of each component of $\mathbf{x}^t$, i.e., $p(x_n^t|\mathbf{y}^t) = \sum_{\mathbf{x}^t_{\setminus n}} \int_{\boldsymbol{\beta}^t} p(\mathbf{x}^t, \boldsymbol{\beta}^t|\mathbf{y}^t) d\boldsymbol{\beta}^t$. However, this requires integration and high-dimensional summation of discrete symbols, which is analytically intractable.

The variational inference approach finds a distribution $Q(\mathbf{x}^t, \boldsymbol{\beta}^t)$, which is in tractable form and closely represents $p(\mathbf{x}^t, \boldsymbol{\beta}^t|\mathbf{y}^t)$, by minimizing the free energy function defined as $\mathcal{F} = \mathbb{E}_{Q(\mathbf{x}^t,\boldsymbol{\beta}^t)}\{\log \frac{Q(\mathbf{x}^t,\boldsymbol{\beta}^t)}{p(\mathbf{x}^t,\boldsymbol{\beta}^t|\mathbf{y}^t)}\}$ [19]. If there is no constraint on $Q(\mathbf{x}^t, \boldsymbol{\beta}^t)$, then $\mathcal{F}$ is minimized when $Q(\mathbf{x}^t, \boldsymbol{\beta}^t) = p(\mathbf{x}^t, \boldsymbol{\beta}^t|\mathbf{y}^t)$, which, however, leads us back to the intractable problem as shown above. Therefore, the mean-field approximation [20], which factorizes $Q(\mathbf{x}^t, \boldsymbol{\beta}^t)$ into a product form, i.e., $Q(\mathbf{x}^t, \boldsymbol{\beta}^t) = Q(\mathbf{x}^t)Q(\boldsymbol{\beta}^t)$, is widely applied. This is equivalent to assuming that $\mathbf{x}^t$ and $\boldsymbol{\beta}^t$ are independent conditioned on $\mathbf{y}^t$, and will greatly simplify the marginalization of $p(x_n^t|\mathbf{y}^t) \approx \sum_{\mathbf{x}^t_{\setminus n}} \int_{\boldsymbol{\beta}^t} Q(\mathbf{x}^t)Q(\boldsymbol{\beta}^t) d\boldsymbol{\beta}^t$.

A. Prior Distributions

First, for the discrete symbol, we consider $\mathbf{x} = [(\mathbf{x}^0)^T, \ldots, (\mathbf{x}^{\mathfrak{T}-1})^T]^T$ to be the signal vector of $\mathfrak{T}$ consecutive OFDM symbols. Let $s_j \in \mathcal{C}$ be the jth element of the constellation points $\mathcal{C}$, for each modulated symbol x_n^t, due to the independence property among symbol elements, we have

$$p(\mathbf{x}) = \prod_{t=0}^{\mathfrak{T}-1} \prod_{n=0}^{N-1} \left(\prod_{j=0}^{|\mathcal{C}|-1} \left(\frac{1}{|\mathcal{C}|} \right)^{\mathbb{I}(x_n^t = s_j)} \right), \tag{3}$$

where $\mathbb{I}(x_n^t = s_j)$ is the indication function that equals 1 if $x_n^t = s_j$ is satisfied and 0 otherwise, and the prior probability that x_n^t takes on $s_j \in \mathcal{C}$ equals $\frac{1}{|\mathcal{C}|}$.

For the unknown additive noise, $\boldsymbol{\beta}^t$ is assumed to be independent of the index $t \in \{0, \ldots, \mathfrak{T}-1\}$, which amounts to saying that the interferers exist in the interfered subcarriers over the duration of $\mathfrak{T}$ consecutive OFDM symbols with fixed interference power. Dropping the OFDM sybmol index t and defining $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_{N-1}]^T$, the prior of $\boldsymbol{\beta}$ takes the form

$$p(\boldsymbol{\beta}) = \prod_{n=0}^{N-1} \mathrm{Ga}(\beta_n|c_n, d_n) = \prod_{n=0}^{N-1} \frac{{d_n}^{c_n}}{\Gamma(c_n)} \beta_n^{c_n-1} \exp(-d_n\beta_n), \tag{4}$$

where c_n and d_n are the parameters of the Gamma distribution. In the absence of prior information, small values for hyperparameters are chosen, i.e., $c_n = d_n = 10^{-6}$, so as to produce uninformative priors [21].

B. Derivation of Proposed Algorithm

With the likelihood function in (2) and prior distributions from (3)-(4), the variational inference approach now finds a distribution $Q(\mathbf{x}, \boldsymbol{\beta})$ to represent $p(\mathbf{x}, \boldsymbol{\beta}|\mathbf{y}) = \prod_{t=0}^{\mathfrak{T}-1} p(\mathbf{x}^t, \boldsymbol{\beta}|\mathbf{y}^t)$, where $\mathbf{y} = [(\mathbf{y}^0)^T, \ldots, (\mathbf{y}^{\mathfrak{T}-1})^T]^T$. According to the mean-field approximation [22], we factorize $Q(\mathbf{x}, \boldsymbol{\beta})$ into a product form as $Q(\boldsymbol{\beta}) \prod_{t=0}^{\mathfrak{T}-1} \prod_{n=0}^{N-1} Q(x_n^t)$. Then the variational free energy is given by [21]

$$\mathcal{F} = \mathbb{E}_{Q(\boldsymbol{\beta}) \prod_{t=0}^{\mathfrak{T}-1} \prod_{n=0}^{N-1} Q(x_n^t)} \left\{ \log \frac{Q(\boldsymbol{\beta}) \prod_{t=0}^{\mathfrak{T}-1} \prod_{n=0}^{N-1} Q(x_n^t)}{p(\mathbf{x}, \boldsymbol{\beta}|\mathbf{y})} \right\}. \tag{5}$$

It turns out that the variational distribution $Q(\cdot)$ that minimizes (5) can be computed from [21]

$$Q(\boldsymbol{z}_i) \propto \exp\{\mathbb{E}_{\prod_{j\neq i} Q(\boldsymbol{z}_j)}\{\log p(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{\mathfrak{T}\mathfrak{N}+1}, \mathbf{y})\}\}, \tag{6}$$

where $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{\mathfrak{T}\mathfrak{N}+1}\} \triangleq \{\boldsymbol{\beta}, x_0^0, \ldots, x_{N-1}^{\mathfrak{T}-1}\}$.

1) Estimation of $Q(\boldsymbol{\beta})$: Using (6), by substituting the likelihood function from (2) and prior distributions from (3)–(4), $Q(\boldsymbol{\beta})$ can be simplified with straightforward manipulation

$$Q(\boldsymbol{\beta}) \propto \prod_{n=0}^{N-1} \beta_n^{c_n+\mathfrak{T}-1} \exp(-d_n\beta_n) \prod_{t=0}^{\mathfrak{T}-1} \exp\left\{-\mathrm{Tr}\{\mathrm{diag}\{\boldsymbol{\beta}\}\boldsymbol{\Psi}^t\}\right\},$$

where $\boldsymbol{\Psi}^t \triangleq \mathbb{E}_{\prod_{n=0}^{N-1} Q(x_n^t)}\{(\mathbf{y}^t - \mathbf{D}^t\mathbf{x}^t)(\mathbf{y}^t - \mathbf{D}^t\mathbf{x}^t)^H\}$. From the Appendix, it can be further obtained

$$\boldsymbol{\Psi}^t[\mathbf{m}_{\mathbf{x}^t}, \boldsymbol{\Sigma}_{\mathbf{x}^t}] = (\mathbf{y}^t - \mathbf{D}^t\mathbf{m}_{\mathbf{x}^t})(\mathbf{y}^t - \mathbf{D}^t\mathbf{m}_{\mathbf{x}^t})^H + \mathbf{D}^t\boldsymbol{\Sigma}_{\mathbf{x}^t}(\mathbf{D}^t)^H, \tag{7}$$

with $\mathbf{m}_{\mathbf{x}^t} \triangleq \mathbb{E}_{\prod_{n=0}^{N-1} Q(x_n^t)}\{\mathbf{x}^t\}$ and $\boldsymbol{\Sigma}_{\mathbf{x}^t} \triangleq \mathbb{E}_{\prod_{n=0}^{N-1} Q(x_n^t)}\{(\mathbf{x}^t - \mathbf{m}_{\mathbf{x}^t})(\mathbf{x}^t - \mathbf{m}_{\mathbf{x}^t})^H\}$.

Since $\mathrm{Tr}\{\mathrm{diag}\{\boldsymbol{\beta}\}\boldsymbol{\Psi}^t[\mathbf{m}_{\mathbf{x}^t}, \boldsymbol{\Sigma}_{\mathbf{x}^t}]\} = \sum_{n=0}^{N-1} \beta_n \boldsymbol{\Psi}^t[\mathbf{m}_{\mathbf{x}^t}, \boldsymbol{\Sigma}_{\mathbf{x}^t}]_{n,n}$, $Q(\boldsymbol{\beta})$ is shown to follow product of N Gamma distributions

$$\begin{aligned} Q(\boldsymbol{\beta}) &\propto \prod_{n=0}^{N-1} \beta_n^{(c_n+\mathfrak{T}-1)} \exp\{-\beta_n(d_n + \sum_{t=0}^{\mathfrak{T}-1} \boldsymbol{\Psi}^t[\mathbf{m}_{\mathbf{x}^t}, \boldsymbol{\Sigma}_{\mathbf{x}^t}]_{n,n})\} \\ &= \prod_{n=0}^{N-1} \mathrm{Ga}(\beta_n | c_n + \mathfrak{T}, d_n + \sum_{t=0}^{\mathfrak{T}-1} \boldsymbol{\Psi}^t[\mathbf{m}_{\mathbf{x}^t}, \boldsymbol{\Sigma}_{\mathbf{x}^t}]_{n,n}). \end{aligned} \tag{8}$$

Then $\mathbf{m}_{\boldsymbol{\beta}} \triangleq \mathbb{E}_{Q(\boldsymbol{\beta})}\{\boldsymbol{\beta}\}$ is given by [21]

$$\mathbf{m}_{\boldsymbol{\beta}} = \left[\frac{c_0 + \mathfrak{T}}{d_0 + \sum_{t=0}^{\mathfrak{T}-1} \boldsymbol{\Psi}^t[\mathbf{m}_{\mathbf{x}^t}, \boldsymbol{\Sigma}_{\mathbf{x}^t}]_{0,0}}, \ldots, \frac{c_{N-1} + \mathfrak{T}}{d_{N-1} + \sum_{t=0}^{\mathfrak{T}-1} \boldsymbol{\Psi}^t[\mathbf{m}_{\mathbf{x}^t}, \boldsymbol{\Sigma}_{\mathbf{x}^t}]_{N-1,N-1}}\right]^T. \tag{9}$$

2) Estimation of $Q(x_n^t)$ *:* Similarly, following (6), $Q(x_n^t)$ can be obtained as

$$Q(x_n^t) \propto \left(\prod_{j=0}^{|\mathcal{C}|-1} \left(\frac{1}{|\mathcal{C}|}\right)^{\mathbb{I}(x_n^t = s_j)}\right) \exp\left\{-\mathrm{Tr}\{\mathrm{diag}\{\mathbf{m}_{\boldsymbol{\beta}}\}\boldsymbol{\Upsilon}^t\}\right\}, \tag{10}$$

where $\mathbf{m}_{\boldsymbol{\beta}} = \mathbb{E}_{Q(\boldsymbol{\beta})}\{\boldsymbol{\beta}\}$ is given by (9), and $\boldsymbol{\Upsilon}^t \triangleq \mathbb{E}_{\prod_{m\neq n} Q(x_m^t)}\{(\mathbf{y}^t - \mathbf{D}^t\mathbf{x}^t)(\mathbf{y}^t - \mathbf{D}\mathbf{x}^t)^H\} = (\mathbf{y}^t - \mathbf{D}^t_{:,\backslash n}\mathbf{m}_{\mathbf{x}^t_{\backslash n}} - \mathbf{D}^t_{:,n}x_n^t)(\mathbf{y}^t - \mathbf{D}^t_{:,\backslash n}\mathbf{m}_{\mathbf{x}^t_{\backslash n}} - \mathbf{D}^t_{:,n}x_n^t)^H + \mathrm{const}$ follows from the Appendix, with $\mathbf{m}_{\mathbf{x}^t_{\backslash n}} \triangleq \mathbb{E}_{\prod_{m\neq n} Q(x_m^t)}\{\mathbf{x}^t_{\backslash n}\}$. Further denote

$$\begin{aligned} &\boldsymbol{\Xi}^t[\mathbf{m}_{\mathbf{x}^t_{\backslash n}}, x_n^t] \\ &\triangleq (\mathbf{y}^t - \mathbf{D}^t_{:,\backslash n}\mathbf{m}_{\mathbf{x}^t_{\backslash n}} - \mathbf{D}^t_{:,n}x_n^t)(\mathbf{y}^t - \mathbf{D}^t_{:,\backslash n}\mathbf{m}_{\mathbf{x}^t_{\backslash n}} - \mathbf{D}^t_{:,n}x_n^t)^H, \end{aligned} \tag{11}$$

together with the fact that x_n^t falls on the constellation points $s_j \in \mathcal{C}$, it can be obtained [23] as (12), shown at the bottom of the page, which is a Categorical distribution with the jth parameter

$$r_{n,j}^t = \frac{\frac{1}{|\mathcal{C}|}\exp\{-\mathrm{Tr}\{\mathrm{diag}\{\mathbf{m}_{\boldsymbol{\beta}}\}\boldsymbol{\Xi}^t[\mathbf{m}_{\mathbf{x}^t_{\backslash n}}, s_j]\}\}}{\sum_{j=0}^{|\mathcal{C}|-1} \frac{1}{|\mathcal{C}|}\exp\{-\mathrm{Tr}\{\mathrm{diag}\{\mathbf{m}_{\boldsymbol{\beta}}\}\boldsymbol{\Xi}^t[\mathbf{m}_{\mathbf{x}^t_{\backslash n}}, s_j]\}\}}. \tag{13}$$

It satisfies $r_{n,j}^t \geq 0$ and $\sum_{j=0}^{|\mathcal{C}|-1} r_{n,j}^t = 1$ [23], and the mean and variance of x_n^t are given by

$$m_{x_n^t} \triangleq \sum_{j=0}^{|\mathcal{C}|-1} r_{n,j}^t s_j, \tag{14}$$

$$v_{x_n^t} \triangleq \sum_{j=0}^{|\mathcal{C}|-1} r_{n,j}^t |s_j - m_{x_n^t}|^2. \tag{15}$$

C. Summary of Algorithm

From the expressions of $Q(\boldsymbol{\beta})$ in (8) and $Q(x_n^t)$ in (12), it is noticed that updating $Q(\boldsymbol{\beta})$ depends on $\mathbf{m}_{\mathbf{x}^t}$ and $\boldsymbol{\Sigma}_{\mathbf{x}^t}$, $t = 0, \ldots, \mathfrak{T}-1$, while updating $Q(x_n^t)$ for $t = 0, \ldots, \mathfrak{T}-1$ and $n = 0, \ldots, N-1$ relies on both $\mathbf{m}_{\boldsymbol{\beta}}$ and $\mathbf{m}_{\mathbf{x}^t_{\backslash n}}$. Consequently, they should be be updated iteratively.

In each iteration, given $\mathbf{m}_{\mathbf{x}^t}$ and $\boldsymbol{\Sigma}_{\mathbf{x}^t}$ for $t = 0, \ldots, \mathfrak{T}-1$, the noise statistics $\mathbf{m}_{\boldsymbol{\beta}}$ is obtained through (9). Then, with $\mathbf{m}_{\boldsymbol{\beta}}$ and $\mathbf{m}_{\mathbf{x}^t_{\backslash n}}$, for each discrete symbol x_n^t, $r_{n,j}^t$ is computed via (13), thus $m_{x_n^t}$ via (14) is used to refine $\mathbf{m}_{\mathbf{x}^t_{\backslash(n+1)}} = [m_{x_0^t}, \ldots, m_{x_n^t}, m_{x_{n+2}^t}, \ldots, m_{x_{N-1}^t}]^T$ for the estimation of next symbol x_{n+1}^t. The symbols are updated successively from $n = 0$ to $n = N-1$, and parallelly for $t \in \{1, \ldots, \mathfrak{T}-1\}$. Upon completion, based on (14) and (15), $\mathbf{m}_{\mathbf{x}^t} = [m_{x_0^t}, \ldots, m_{x_{N-1}^t}]^T$ and $\boldsymbol{\Sigma}_{\mathbf{x}^t} = \mathrm{diag}\{[v_{x_0^t}, \ldots, v_{x_{N-1}^t}]^T\}$ for $t = 0, \ldots, \mathfrak{T}-1$ are then refined for the next iteration. Notice that, in estimating each symbol, calculation of $\mathbf{y}^t - \mathbf{D}^t_{:,\backslash n}\mathbf{m}_{\mathbf{x}^t_{\backslash n}}$ in (11) reveals a soft interference cancelation procedure for symbol x_n^t.

To start the iterative algorithm, $\mathbf{m}_{\mathbf{x}^t}$ and $\boldsymbol{\Sigma}_{\mathbf{x}^t}$ are required. Since we have no information about symbols, we can set them to be uniformly distributive, i.e., $[r_{n,0}^t, \ldots, r_{n,|\mathcal{C}|-1}^t] = [\frac{1}{|\mathcal{C}|}, \ldots, \frac{1}{|\mathcal{C}|}]_{|\mathcal{C}|\times 1}$. Thus, $\mathbf{m}_{\mathbf{x}^t} = [m_{x_0^t}, \ldots, m_{x_{N-1}^t}]^T$ and $\boldsymbol{\Sigma}_{\mathbf{x}^t} = \mathrm{diag}\{[v_{x_0^t}, \ldots, v_{x_{N-1}^t}]^T\}$ are easily obtained with (14) and (15).

The complexity of the proposed algorithm in each iteration is dominated by (9) and (12). For (9), since only the diagonal elements of $\boldsymbol{\Psi}^t[\mathbf{m}_{\mathbf{x}^t}, \boldsymbol{\Sigma}_{\mathbf{x}^t}]$ are needed, the complexity of computing those diagonal elements is $\mathcal{O}(N^2)$ based on the diagonal structure of $\boldsymbol{\Sigma}_{\mathbf{x}^t}$ in (7). Thus, the complexity of (9) is $\mathcal{O}(\mathfrak{T}\mathfrak{N}^2)$. For (12), $\mathrm{Tr}\{\mathrm{diag}\{\mathbf{m}_{\boldsymbol{\beta}}\}\boldsymbol{\Xi}^t[\mathbf{m}_{\mathbf{x}^t_{\backslash n}}, s_j]\}$ can be performed in a recursive way based on (11) for $n = 0, \ldots, N-1$, whose complexity is $\mathcal{O}(N^2)$ in the initial step and $\mathcal{O}(N|\mathcal{C}|)$ in the recursive process. Therefore, the complexity of (12) for $t = 0, \ldots, \mathfrak{T}-1$ and $n = 0, \ldots, N-1$ is $\mathcal{O}(\mathfrak{T}\mathfrak{N}^2) + \mathcal{O}(\mathfrak{T}\mathfrak{N}|\mathcal{C}|)$. Thus, the total complexity of the proposed algorithm in each iteration is $\mathcal{O}(\mathfrak{T}\mathfrak{N}^2) + \mathcal{O}(\mathfrak{T}\mathfrak{N}|\mathcal{C}|)$.

IV. Simulation Results and Discussions

This section presents numerical results to assess the performance of the proposed algorithm. In the simulation below, we consider $\mathfrak{T} = 6$ consecutive OFDM symbols, each with $N = 128$ subcarriers. QPSK modulation is used with the transmission signal power set as σ_s^2. Both uncoded and coded systems are considered. For coded systems, a rate $r = 1/2$ convolutional code $(171, 133)_8$ with a block interleaver is used. We assume the carrier frequency to be 3 GHz and the channel bandwidth as 1.25 MHz. The channel is assumed to have $L_h = 6$ paths with power delay profile given by $\sigma_h^2(l) = c_0\exp(-0.6l)$, where l represents the path number and c_0 is the normalization term such that $\sum_{l=0}^{L_h-1}\sigma_h^2(l) = 1$. Each path $h(n,l)$ is generated according to a zero-mean complex Gaussian process with autocorrelation of the lth path given by $\mathbb{E}\{h(m,l)h^*(n,l)\} = \sigma_h^2(l)J_0(2\pi f_d(m -$

$$Q(x_n^t) \propto \prod_{j=0}^{|\mathcal{C}|-1}\left(\frac{1}{|\mathcal{C}|}\exp\{-\mathrm{Tr}\{\mathrm{diag}\{\mathbf{m}_{\boldsymbol{\beta}}\}\boldsymbol{\Xi}^t[\mathbf{m}_{\mathbf{x}^t_{\backslash n}}, s_j]\}\}\right)^{\mathbb{I}(x_n^t = s_j)} = \mathrm{Cat}(x_n^t | r_{n,0}^t, \ldots, r_{n,|\mathcal{C}|-1}^t), \tag{12}$$

$$
\begin{aligned}
&\mathbb{E}_{\prod_{m\neq n} Q(x_m^t)}\{(\mathbf{y}^t-\mathbf{D}^t\mathbf{x}^t)(\mathbf{y}^t-\mathbf{D}^t\mathbf{x}^t)^H\} = (\mathbf{y}^t-\mathbf{D}^t_{:,n}x_n^t)(\mathbf{y}^t-\mathbf{D}^t_{:,n}x_n^t)^H\\
&\quad-(\mathbf{y}^t-\mathbf{D}^t_{:,n}x_n^t)(\mathbf{D}^t_{:,\backslash n}\mathbf{m}_{\mathbf{x}^t_{\backslash n}})^H-(\mathbf{D}^t_{:,\backslash n}\mathbf{m}_{\mathbf{x}^t_{\backslash n}})(\mathbf{y}^t-\mathbf{D}^t_{:,n}x_n^t)^H+\mathbf{D}^t_{:,\backslash n}(\mathbf{m}_{\mathbf{x}^t_{\backslash n}}(\mathbf{m}_{\mathbf{x}^t_{\backslash n}})^H+\mathbf{\Sigma}_{\mathbf{x}^t_{\backslash n}})(\mathbf{D}^t_{:,\backslash n})^H\\
&=(\mathbf{y}^t-\mathbf{D}^t_{:,\backslash n}\mathbf{m}_{\mathbf{x}^t_{\backslash n}}-\mathbf{D}^t_{:,n}x_n^t)(\mathbf{y}^t-\mathbf{D}^t_{:,\backslash n}\mathbf{m}_{\mathbf{x}^t_{\backslash n}}-\mathbf{D}^t_{:,n}x_n^t)^H+\text{const},
\end{aligned}
$$

$n)T_s)$ [24], where $J_0(\cdot)$ represents the zero-order Bessel function of the first kind and T_s denotes inverse of the channel bandwidth. The received OFDM signal is affected by interference. The interference is randomly positioned in the signal bandwidth affecting 20 subcarriers. At each interfered subcarrier, the interference is a superposition of multiple (uniformly generated from 1 to 10) unknown interferers, each with randomly modulated QPSK symbols going through Rayleigh flat fading channel. The average signal-to-interference ratio (SIR) is defined as $\text{SIR} = \frac{N\sigma_s^2}{\sum_{n\in S_I}\varsigma_n^2}$, which is the ratio of total transmit signal power over total interference power in one OFDM symbol with S_I denoting the set of subcarriers affected by interference. The thermal noise power σ^2 is set to be the same for all subcarriers. The signal-to-noise ratio (SNR) is defined as $\text{SNR} = \sigma_s^2/\sigma^2$. The proposed algorithm is executed for 6 iterations in both uncoded and coded systems. For interfacing with channel decoder, upon completion of iterations the proposed algorithm outputs bit probabilities by marginalization for channel decoding. Each point in the following figures is obtained by averaging the results over 1000 runs, each with different channel, noise and interference realizations.

For comparison, the algorithms from [10] and [11] are also simulated. In particular, the BCJR-based MAP equalizer from [10] making use of the approximately banded ICI structure, with one-side bandwidth of two is considered. On the other hand, the iterate nonbanded block linear equalizer from [11] for OFDM systems under doubly selective channels based on linear minimum mean square error (MMSE) [25], which outperforms low-complexity linear equalizers from [8] and [12], is compared. Both of them assumed perfect knowledge of noise statistics, and did not consider external interference. The complexity of the applied methods from [10] and [11] are $\mathcal{O}(\mathfrak{T}\mathfrak{N}|\mathcal{C}|^4)$ and $\mathcal{O}(\mathfrak{T}\mathfrak{N}^3|\mathcal{C}|)$, respectively. As a benchmark, the matched-filter bound is also provided, which assumes perfect ICI cancelation with perfect data at adjacent subcarriers, and also knowledge of noise plus interference powers for optimal MAP equalization. BCJR decoder is used for channel decoding if coding is applied.

The interference power is set at $\text{SIR} = 8$ dB for coded systems and $\text{SIR} = 24$ dB for uncoded systems, while the normalized Doppler spread is set as 0.1, which corresponds to transceiver speed of about 350 kilometers per hour. As can be observed from Fig. 1, both schemes from [10] and [11] show BER error floors in both coded and uncoded systems. This is due to the presence of interference, rendering the system to be interference-limited if the interference is ignored at the receiver. On the other hand, the proposed approach, which explicitly considers interference, does not have error floor effect, and is only about 1 dB away from the matched-filter bound at $\text{BER} = 10^{-4}$ in coded systems, and 3 dB away at $\text{BER} = 10^{-3}$ in uncoded systems. As can be observed from Fig. 1, both schemes from [10] and [11] show BER error floors in both coded and uncoded systems. This is due to the presence of interference, rendering the system to be interference-limited if the interference is ignored at the receiver. On the other hand, the proposed approach, which explicitly considers interference, does not have error floor effect, and is only about 1 dB away.

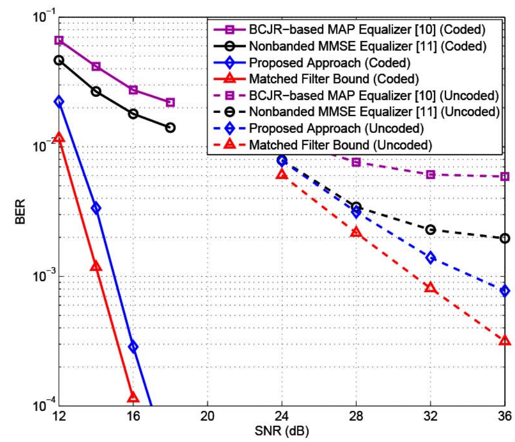

Fig. 1. BER versus SNR under Unknown Interference.

V. Conclusions

In this letter, we proposed a variational inference-based approach to joint interference mitigation and OFDM equalization with high mobility. It iteratively learns the noise plus interference powers at different subcarriers and refines the soft data estimate at each subcarrier successively. Simulation results demonstrated the robustness of the proposed algorithm to unknown interference, and its superiority with BER performance closed to the matched-filter bound.

Appendix

For $\mathbf{\Psi}^t[\mathbf{m}_{\mathbf{x}^t},\mathbf{\Sigma}_{\mathbf{x}^t}]$ in (7), denote $\mathbf{m}_{\mathbf{x}^t} \triangleq \mathbb{E}_{\prod_{n=0}^{N-1} Q(x_n^t)}\{\mathbf{x}^t\}$, and $\mathbf{\Sigma}_{\mathbf{x}^t} \triangleq \mathbb{E}_{\prod_{n=0}^{N-1} Q(x_n^t)}\{(\mathbf{x}^t-\mathbf{m}_{\mathbf{x}^t})(\mathbf{x}^t-\mathbf{m}_{\mathbf{x}^t})^H\}$, then

$$
\begin{aligned}
&\mathbb{E}_{\prod_{n=0}^{N-1} Q(x_n^t)}\{(\mathbf{y}^t-\mathbf{D}^t\mathbf{x}^t)(\mathbf{y}^t-\mathbf{D}^t\mathbf{x}^t)^H\}\\
&=\mathbf{y}^t(\mathbf{y}^t)^H-\mathbf{y}^t(\mathbf{m}_{\mathbf{x}^t})^H(\mathbf{D}^t)^H-\mathbf{D}^t\mathbf{m}_{\mathbf{x}^t}(\mathbf{y}^t)^H\\
&\quad+\mathbf{D}^t(\mathbf{m}_{\mathbf{x}^t}(\mathbf{m}_{\mathbf{x}^t})^H+\mathbf{\Sigma}_{\mathbf{x}^t})(\mathbf{D}^t)^H\\
&=(\mathbf{y}^t-\mathbf{D}^t\mathbf{m}_{\mathbf{x}^t})(\mathbf{y}^t-\mathbf{D}^t\mathbf{m}_{\mathbf{x}^t})^H+\mathbf{D}^t\mathbf{\Sigma}_{\mathbf{x}^t}(\mathbf{D}^t)^H.
\end{aligned}
$$

Thus, $\mathbf{\Psi}^t[\mathbf{m}_{\mathbf{x}^t},\mathbf{\Sigma}_{\mathbf{x}^t}]$ in (7) is obtained.

For $\mathbf{\Upsilon}^t$ in (10), we factorize $\mathbf{D}^t\mathbf{x}^t = \mathbf{D}^t_{:,n}x_n^t + \mathbf{D}^t_{:,\backslash n}\mathbf{x}^t_{\backslash n}$, and denote $\mathbf{m}_{\mathbf{x}^t_{\backslash n}} \triangleq \mathbb{E}_{\prod_{m\neq n} Q(x_m^t)}\{\mathbf{x}^t_{\backslash n}\}$ and $\mathbf{\Sigma}_{\mathbf{x}^t_{\backslash n}} \triangleq \mathbb{E}_{\prod_{m\neq n} Q(x_m^t)}\{(\mathbf{x}^t_{\backslash n}-\mathbf{m}_{\mathbf{x}^t_{\backslash n}})(\mathbf{x}^t_{\backslash n}-\mathbf{m}_{\mathbf{x}^t_{\backslash n}})^H\}$. Then, it yields the equation shown at the top of the page, where $\text{const} = \mathbf{D}^t_{:,\backslash n}\mathbf{\Sigma}_{\mathbf{x}^t_{\backslash n}}(\mathbf{D}^t_{:,\backslash n})^H$. Thus, $\mathbf{\Upsilon}^t$ in (10) is derived.

REFERENCES

[1] D. López-Pérez, A. Valcarce, G. De La Roche, and J. Zhang, "OFDMA femtocells: A roadmap on interference avoidance," *IEEE Commun. Mag.*, vol. 47, no. 9, pp. 41–48, Sep. 2009.

[2] H. Mahmoud, T. Yucek, and H. Arslan, "OFDM for cognitive radio: Merits and challenges," *IEEE Wireless Commun.*, vol. 16, no. 2, pp. 6–15, Apr. 2009.

[3] A. J. Coulson, "Bit error rate performance of OFDM in narrowband interference with excision filtering," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2484–2492, Sep. 2006.

[4] T. Li, W. H. Mow, V. K. Lau, M. Siu, R. S. Cheng, and R. D. Murch, "Robust joint interference detection and decoding for OFDM-based cognitive radio systems with unknown interference," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 566–575, Apr. 2007.

[5] L. Sanguinetti, M. Morelli, and H. V. Poor, "BICM decoding of jammed OFDM transmissions using the EM algorithm," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 2800–2806, Sep. 2011.

[6] S. Ahmadi, *Mobile WiMAX: A Systems Approach to Understanding IEEE 802.16 m Radio Cccess Technology*. New York, NY, USA: Academic, 2010.

[7] S. Sesia, I. Toufik, and M. Baker, *LTE- The UMTS Long Term Evolution: From Theory to Practice*, 2nd ed. Hoboken, NJ, USA: Wiley, 2011.

[8] P. Schniter, "Low-complexity equalization of OFDM in doubly selective channels," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1002–1011, 2004.

[9] I. Barhumi, G. Leus, and M. Moonen, "Equalization for OFDM over doubly selective channels," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1445–1458, 2006.

[10] D. Liu and M. Fitz, "Iterative MAP equalization and decoding in wireless mobile coded OFDM," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 2042–2051, Jul. 2009.

[11] K. Fang, L. Rugini, and G. Leus, "Low-complexity block turbo equalization for OFDM systems in time-varying channels," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5555–5566, Nov. 2008.

[12] T. Hrycak, S. Das, G. Matz, and H. Feichtinger, "Low complexity equalization for doubly selective channels modeled by a basis expansion," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5706–5719, Nov. 2010.

[13] P. Schniter, S.-J. Hwang, S. Das, and A. P. Kannu, , F. Hlawatsch and G. Matz, Eds., *Equalization of Time-varying Channels, in Wireless Communications over Rapidly Time-varying Channels*, 1st ed. Amsterdam, The Netherlands: Elsevier, 2011.

[14] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, no. 1–2, pp. 1–305, Jan. 2008.

[15] T. Y. Al-Naffouri, K. Z. Islam, N. Al-Dhahir, and S. Lu, "A model reduction approach for OFDM channel estimation under high mobility conditions," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2181–2193, Apr. 2010.

[16] S. Kalyani, V. Raj, and K. Giridhar, "Narrowband interference mitigation in turbo-coded OFDM systems," in *IEEE Int. Conf. Communications (ICC)*, 2007, pp. 1059–1064, IEEE.

[17] M. Morelli and M. Moretti, "Channel estimation in OFDM systems with unknown interference," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5338–5347, Oct. 2009.

[18] L. Sanguinetti, M. Morelli, and H. V. Poor, "Frame detection and timing acquisition for OFDM transmissions with unknown interference," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 1226–1236, Mar. 2010.

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

[20] 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.

[21] C. Bishop and M. Tipping, "Variational relevance vector machines," in *The 16th Conf. Uncertainty in Artifcial Intelligence*, 2000, pp. 46–53.

[22] S. Haykin *et al.*, *New Directions in Statistical Signal Processing: From Systems to Brain*. Cambridge, MA, USA: MIT Press, 2005.

[23] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. Berlin, Germany: Springer, 2006.

[24] W. C. Jakes and D. C. Cox, *Microwave Mobile Communications*. Piscataway, NJ, USA: Wiley-IEEE Press, 1994.

[25] 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, 1999.