HYBRID VARIATIONAL BAYESIAN CHANNEL ESTIMATION, DEMODULATION AND DECODING FOR OFDM UNDER SPARSE MULTIPATH CHANNELS

Chulong Chen, Michael D. Zoltowski

Department of Electrical and Computer Engineering
Purdue University
West Lafayette, Indiana 47907

ABSTRACT

In this paper, a novel hybrid OFDM receiver based on sparse variational Bayesian (VB) learning and soft-input soft-output decoding is proposed. By noticing that a key part of the inference problem approximated by VB (message-passing) methods may be inferred exactly, an genetic interfacing structure is proposed allowing the use of virtually all existing soft-input soft-output decoding schemes. Therefore the tasks of joint channel state estimation, demodulation and decoding are iteratively solved under the proposed hybrid variational Bayesian framework. The bit-interleaved coded modulation with Turbo coding is used to demonstrate the potential performance of the proposed structure. Very promising results in performance are observed in computer simulated experiments.

Index Terms— OFDM, channel estimation, demodulation, decoding, variational Bayesian methods

1. INTRODUCTION

The success of the combination of OFDM and Shannon limit approaching codes has been adopted in many state-of-the-art wireless communication systems. In Long Term Evolution (LTE), for example, OFDM and Turbo coding are successfully employed together to provide near-capacity physical link for mobile phones and data terminals. In order to achieve this goal wireless multipath channels, advanced channel estimation and equalization techniques are required. This paper focuses on the design of a variational Bayesian method based algorithm for joint channel estimation (for both multipath channel impulse response and channel noise variance), demodulation and decoding in OFDM systems.

Throughout this paper the following notations shall be used. The expressions $\mathrm{diag}(\cdot)$ and $\mathrm{Diag}(\cdot)$ denote the vector consisted of diagonal elements of a matrix and the diagonal matrix formed by putting the elements of a vector on its main diagonal, respectively. The superscripts $(\cdot)^T$ and $(\cdot)^H$ denote matrix transposition and Hermitian transposition respectively. The operators $tr(\cdot)$ and $det(\cdot)$ designate trace and determinant of a matrix. The expectation with respect to probability distribution function q is expressed as $\langle\cdot\rangle_q$. Curly letters (e.g. $\mathcal{R}$ and $\mathcal{A}$) are used to designate sets. A subvector is denoted by attaching at subscript an index set pointing to desired elements (e.g. $\boldsymbol{v}_{\mathcal{I}}$ with $\mathcal{I} = \{1, 3, 5\}$). Similarly, a matrix with an index set at its subscript denotes the submatrix consists only the rows specified by the set, i.e. $\boldsymbol{M}_{\mathcal{I}}$.

2. SIGNAL MODEL

Throughout this paper, cyclic-prefix orthogonal frequency division multiplexing (CP-OFDM) is considered and its complex baseband representation is used. The system model of CP-OFDM is summarized as follows. Let N_c and N_{cp} denote the number of subcarriers, the length of cyclic prefix in samples, respectively. Let x_l and y_l, with $l = 0, \dots, N_c - 1$, denote the symbols transmitted and received respectively over the l-th subcarrier. The system input-output relation is given by $y_l = H_l x_l + z_l$ where z_l is the overall interference including the AWGN noise and inter-carrier interference (ICI) whose statistical property is discussed later in this section. And H_l is the narrowband channel coefficient expressed as the Fourier transform of a discrete-time time-varying impulse response h_m, i.e., $H_l = \sum_{m=0}^{N_{cp}-1} h_m e^{-j2\pi lm/N_c}$. Slow fading channel is assumed so that the multipath channel can be considered constant during one OFDM symbol period.

It is convenient to express the system model in a matrix form. First define the $N_c \times 1$ vectors $\mathbf{y} = [y_1, y_2, \dots, y_{N_c}]^T$, $\mathbf{z} = [z_1, z_2, \dots, z_{N_c}]^T$, $\mathbf{H}_k = [H_1, H_2, \dots, H_{N_c}]^T$ and $\boldsymbol{x} = [x_1, x_2, \dots x_{N_c}]^T$. Therefore $\mathbf{H}$ and $\mathbf{h}$ are related through the discrete Fourier transform, i.e. $\mathbf{H} = \mathcal{F}\{\mathbf{h}\}$ with $\mathbf{h}$ being the stack of sampled channel impulse response $\{h_m\}$. Using these notations the system input-output relation can be expressed in terms of $\mathbf{h}$

$$\mathbf{y} = \sqrt{N_c}\mathrm{Diag}(\boldsymbol{x})\mathbf{D}\mathbf{h} + \mathbf{z} \tag{1}$$

where $\mathbf{D} \in \mathbb{C}^{N_c \times N_{cp}}$ is a partial unitary DFT matrix consisting of the first N_{cp} columns of a $N_c \times N_c$ DFT matrix. The signal model (1) gives the input-output relation of a CP-OFDM system and formulates the time-domain channel estimation problem as an inverse problem. It is also useful to study the pilot and data carrying subcarriers separately through the following equations using the notations defined for submatrices before

$$\mathbf{y}_{\mathcal{P}} = \sqrt{N_c}\mathrm{Diag}(\mathbf{x}_{\mathcal{P}})\mathbf{D}_{\mathcal{P}}\mathbf{h} + \mathbf{z}_{\mathcal{P}} \tag{2}$$

$$\mathbf{y}_{\mathcal{D}} = \sqrt{N_c}\mathrm{Diag}(\mathbf{x}_{\mathcal{D}})\mathbf{D}_{\mathcal{D}}\mathbf{h} + \mathbf{z}_{\mathcal{D}} \tag{3}$$

where $\mathcal{P}$ and $\mathcal{D}$ denotes the index sets pointing to pilot carrying subcarriers and data carrying subcarriers, respectively.

3. PROBABILISTIC GRAPHIC MODEL FOR INFERENCE

A graphical representation of the signal model introduced in previous section will be given here. The joint probabilistic distribution will be factorized according to the graphical model and will be used to develop the inference procedure.

3.1. Probabilistic Characteristics of the Signals

In system model (1) the noise term $z_{k,l}$ is modeled as a AWGN random vector [1]. The probability density function (PDF) of the output vector $\boldsymbol{y}_k$, given the noise precision parameter (inverse of variance) $\tau \triangleq 1/\sigma^2$ and $\boldsymbol{h}_k$, can therefore be written as

$$p(\boldsymbol{y}|\boldsymbol{h},\sigma^2,\boldsymbol{x}) = \mathcal{CN}(\mathbf{y}|\sqrt{N_c}\mathrm{Diag}(\boldsymbol{x})\mathbf{D}\boldsymbol{h},\tau^{-1}\mathbf{I}) \tag{4}$$

$$= \prod_{\ell}\mathcal{CN}(y_\ell|\sqrt{N_c}H_\ell x_\ell,\tau^{-1}). \tag{5}$$

In this work, sparse multipath channels is assumed and each multipath coefficients h_l's are modeled as an independent circularly symmetric Gaussian random variable with many of which has very small amplitude. In order to exploit the sparseness of the channel within the Bayesian framework a sparseness-promoting prior distribution needs to be assigned to the vector $\boldsymbol{h}$. A two-layer hierarchical model [2] is considered to model the CIR vector. First recall that the CIR vector $\boldsymbol{h}$ is modeled as a circularly-symmetric Gaussian vector with PDF

$$p(\boldsymbol{h}|\boldsymbol{\alpha}) = \prod_{l=0}^{N_{cp}-1}(\alpha_l\pi^{-1})\exp\{-\alpha_l|h_l|^2\} \tag{6}$$

where α_l is defined as the precision (inverse of variance) of a Gaussian density function. Notice that the sparseness of $\boldsymbol{h}$ is controlled by the independent α_l's assigned to each h_l, i.e., most α_l's are very large or numerically equals to infinity resulting in the corresponding distribution strongly peaked at zero. To embody this idea, [3] proposed to assign a gamma prior to the parameters $p(\tau)$ and $p(\boldsymbol{\alpha})$:

$$p(\tau|a,b) = Ga(\tau|a,b), \quad p(\boldsymbol{\alpha}|c,d) = \prod_{l=1}^{N_{cp}}Ga(\alpha_l|c,d) \tag{7}$$

where $Ga(x|a,b) = \Gamma(a)^{-1}b^a x^{a-1}e^{-bx}$ with $\Gamma(\cdot)$ being the gamma function. And by setting the parameters of these priors to zeros, i.e. $a = b = c = d = 0$, will give uninformative (improper) priors (over a logarithmic scale) to $\boldsymbol{\alpha}$ and τ. The prior distribution of $\boldsymbol{h}$ is revealed by marginalization, i.e. $p(\boldsymbol{h};c,d) = \int_0^\infty p(\boldsymbol{h}|\boldsymbol{\alpha})p(\boldsymbol{\alpha};c,d)\mathrm{d}\boldsymbol{\alpha}$. This integration results in a product of PDFs of Student-t distributions as $p(\boldsymbol{h}|c,d) = \sum_l \frac{1}{\pi}\frac{d^c\Gamma(d+1)}{\Gamma(d)}(d+|h_l|^2)^{-(c+1)}$ whose probability mass heavily concentrates along the coordinate axes in the $\boldsymbol{h}$ space.

Finally, the transmitted data symbols at l-th subcarrier x_l are subject to a discrete uniform distribution, i.e,

$$p(x_l) = \sum_{x\in\mathcal{X}}\frac{1}{M}\delta(x_l - x) \tag{8}$$

where $\mathcal{X}$ denotes the set of modulation symbols, $M = |\mathcal{X}|$ is the modulation order, and $\delta(\cdot)$ denotes the Dirac delta function.

3.2. Graphical Representation

Let $\mathcal{V} = \{\boldsymbol{y}_\mathcal{D},\boldsymbol{y}_\mathcal{P},\boldsymbol{x}_\mathcal{P}\}$ and $\mathcal{U} = \{\boldsymbol{x}_\mathcal{D},\boldsymbol{h},\boldsymbol{\alpha},\tau\}$ denote the set of all observed variables and unknown variables, respectively, in the graphical model. The joint PDF of all variables may then be given by $p(\mathcal{U},\mathcal{V}) = p(\boldsymbol{y}|\tau,\boldsymbol{x},\boldsymbol{h})p(\boldsymbol{x}_\mathcal{D})p(\boldsymbol{h}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\tau)$ and the first term on the right hand side can be further split into $p(\boldsymbol{y}|\tau,\boldsymbol{x},\boldsymbol{h}) = p(\boldsymbol{y}_\mathcal{P}|\tau,\boldsymbol{x}_\mathcal{P},\boldsymbol{h})p(\boldsymbol{y}_\mathcal{D}|\tau,\boldsymbol{x}_\mathcal{D},\boldsymbol{h})$ where the pilot tones and data carrying subcarriers are separated explicitly. This factorization indicates an acyclic directed graph depicted in Figure 1, which lays the foundation to the VB inference procedure. In this figure, shaded nodes represent the observed variables, $\boldsymbol{y}_\mathcal{P}$ and $\boldsymbol{y}_\mathcal{D}$, corresponding to received pilot tones and data carrying symbols respectively; and unshaded nodes represent the hidden nodes, $\boldsymbol{\alpha}$, $1/\sigma^2$, $\boldsymbol{h}$ and $\boldsymbol{x}_\mathcal{D}$, that need to be inferred from the observed data. The box with a positive integer number on the corner around a node denotes a *plate*, which indicates that the contained node and its connected edges are duplicated the number of times specified by this integer number. Note that the nodes corresponding to the message bit stream $\boldsymbol{b}$ does not present in the joint PDF since $\boldsymbol{x}_\mathcal{D}$ is uniquely determined by $\boldsymbol{b}$ through coding and modulation rules.

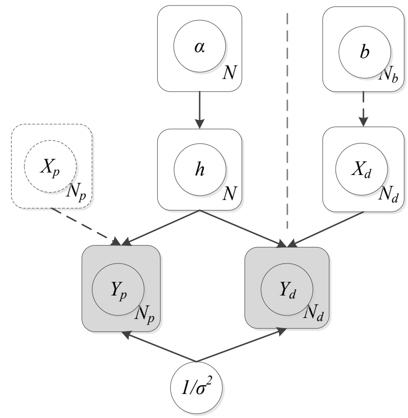

Fig. 1. Bayesian inference model for joint sparse channel estimation, detection and decoding. The dashed nodes represents known variables which is not involved in variational updates

4. HYBRID VB INFERENCE ON GRAPH

Again let $\mathcal{V} = \{\boldsymbol{y}_\mathcal{D},\boldsymbol{y}_\mathcal{P},\boldsymbol{x}_\mathcal{P}\}$ and $\mathcal{U} = \{\boldsymbol{x}_\mathcal{D},\boldsymbol{h},\boldsymbol{\alpha},\tau\}$. For the approximation purpose, choose a structured factorial approximation to the posterior distribution $p(\mathcal{U}|\mathcal{V})$ given by $Q(\mathcal{U}) = Q_{\boldsymbol{h}}(\boldsymbol{h})Q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})Q_\tau(\tau)Q_{\boldsymbol{x}_\mathcal{D}}(\boldsymbol{x}_\mathcal{D})$. For convenience, $Q_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ will be denoted by $Q(\boldsymbol{\theta})$ when there is no ambiguity where $\boldsymbol{\theta} \in \{\boldsymbol{\alpha},\boldsymbol{h},\tau,\boldsymbol{x}_\mathcal{D}\}$ is some collection of variables. Therefore, the variational Bayes method may be used to optimize this factorized distribution to approximate the true posterior distribution of desired parameters.

4.1. Optimizing distributions of $\boldsymbol{h}$, $\boldsymbol{\alpha}$ and τ

The maximizing distribution of $Q^*(\boldsymbol{h})$, $Q^*(\boldsymbol{\alpha})$ and $Q^*(\tau)$ is again evaluated using the standard variational procedure due to [4] given by

$$Q^*(\boldsymbol{\theta}) = \frac{1}{Z}\exp\langle\ln P(\mathcal{V},\mathcal{U})\rangle_{\mathcal{U}\backslash\boldsymbol{\theta}} \tag{9}$$

where Z is a normalization constant that ensures the distribution $Q^*(\boldsymbol{\theta})$ integrating to one and $\langle\cdot\rangle_{\mathcal{U}\backslash\boldsymbol{\theta}}$ denotes the expectation with respect to the distributions $Q^*(\mathcal{U}\backslash\boldsymbol{\theta})$. Notice that the likelihood distributions of $\boldsymbol{h}$, $\boldsymbol{\alpha}$ and τ belong to the exponential family. Evaluation of (9) will give raise to the following optimizing distributions,

$$Q^*(\mathbf{h}) = \mathcal{CN}(\mathbf{h}|\langle\mathbf{h}\rangle,\boldsymbol{\Sigma}_\mathbf{h}), \tag{10}$$

$$Q^*(\alpha) = \prod_{l=0}^{N_{cp}}\Gamma(\alpha_l|a_l^*,b_l^*), \tag{11}$$

$$Q^*(\tau) = \Gamma(\tau|c^*,d^*), \tag{12}$$

where N_{cp} is the normalized length of the multipath channel. And setting the hyperparameters $a = b = c = d = 1$ the optimizing parameters are given by

$$\mathbf{\Sigma_h} = \left[\mathrm{Diag}\langle\boldsymbol{\alpha}\rangle + \langle\tau\rangle\left(\mathbf{\Phi}_p^H\mathbf{\Phi}_p + \langle\mathbf{\Phi}_d^H\mathbf{\Phi}_d\rangle\right)\right]^{-1}, \tag{13}$$

$$\langle\mathbf{h}\rangle = \langle\tau\rangle\mathbf{\Sigma_h}\tilde{\mathbf{\Phi}}^H\mathbf{y}, \tag{14}$$

$$\tilde{a}_l = 1, \qquad \tilde{b}_l = \langle|h_l|^2\rangle, \tag{15}$$

$$\tilde{c} = (N_p + N_d), \tag{16}$$

$$\tilde{d} = \|\mathbf{y} - \tilde{\mathbf{\Phi}}\langle\mathbf{h}\rangle\|_2^2 + \mathrm{tr}\left(\mathbf{\Sigma}_\mathbf{h}^T\left(\mathbf{\Phi}_p^H\mathbf{\Phi}_p + \langle\mathbf{\Phi}_d^H\mathbf{\Phi}_d\rangle\right)\right) \tag{17}$$

where $\langle\mathbf{\Phi}_d^H\mathbf{\Phi}_d\rangle = N_c\mathbf{D}_\mathcal{D}^H\mathrm{Diag}(\mathrm{diag}(\langle\boldsymbol{x}_\mathcal{D}\boldsymbol{x}_\mathcal{D}^H\rangle))\mathbf{D}_\mathcal{D}$, $\langle\alpha_l\rangle = \tilde{a}_l/\tilde{b}_l$ and $\langle\tau\rangle = \tilde{c}/\tilde{d}$. It is noted that $Q(\boldsymbol{x}_\mathcal{D})$ will be obtained in Section 4.2. Optimizing distributions of the channel (hyper-)parameters has heretofore been derived. Excluding the terms involving $\boldsymbol{x}_\mathcal{D}$ and computing with only pilot tones, (10) (11) and (12) may serve as a stand-alone VB based channel estimator scheme for OFDM communication systems, similar to the graphical Bayesian learning based channel estimator described in [5].

4.2. Demodulation and Decoding

In the graphical model, the prior of $\boldsymbol{x}_\mathcal{D}$ is naturally set to a discrete distribution. In last section this discrete assumption over $\boldsymbol{x}_\mathcal{D}$ does not cause any difficulty to updating other hidden variables in the model since only the first and second moments of $Q(\boldsymbol{x}_\mathcal{D})$ are needed (17). And these moments can be easily calculated for a discrete distribution. However, direct application of variational optimizing formula will lead to an awkward optimizing distribution $Q^*(\boldsymbol{x}_\mathcal{D})$ since the discrete distribution is conjugate to the the Gaussian likelihood of $\boldsymbol{y}_\mathcal{D}$. By studying the graphical structure it is found that exact method may be used on $\boldsymbol{x}_\mathcal{D}$ on the right hand side of in Figure 1. The combination of variational and exact method thus indicates a hybrid architechure for an OFDM receiver.

Using the optimizing distributions $Q(\boldsymbol{h})Q(\tau)$ as an approximate to the posteriori distribution, the exact marginalization on $\boldsymbol{x}_\mathcal{D}$ gives

$$\tilde{P}(\boldsymbol{x}_\mathcal{D}|\boldsymbol{y}) = \int P(\boldsymbol{x}_\mathcal{D}|\boldsymbol{h},\tau,\boldsymbol{y}_\mathcal{D})Q(\boldsymbol{h})Q(\tau)\mathrm{d}\boldsymbol{h}\mathrm{d}\tau \tag{18}$$

where $Q(\boldsymbol{h})Q(\tau)$ is used in the place of $p(\boldsymbol{h},\tau|\boldsymbol{y}_\mathcal{D},\boldsymbol{y}_\mathcal{P})$ due to VB assumption. The likelihood of $\boldsymbol{x}_\mathcal{D}$ in the integral by Bayes' theorem can be expressed as $P(\boldsymbol{x}|\boldsymbol{h},\tau,\boldsymbol{y}_\mathcal{D}) = p(\boldsymbol{y}_\mathcal{D}|\boldsymbol{h},\tau,\boldsymbol{x})P(\boldsymbol{x}_\mathcal{D} = \boldsymbol{x})/p(\boldsymbol{y}_\mathcal{D})$. Substituting this result into (18), the marginal distribution of each transmitted codeword bit c_j is given by

$$\tilde{P}(c_j = c|\boldsymbol{y}) = \sum_{\forall\boldsymbol{x}:c_j=c}\tilde{p}(\boldsymbol{y}_\mathcal{D}|\boldsymbol{x})P(\boldsymbol{x})/\tilde{p}(\boldsymbol{y}_\mathcal{D}) \tag{19}$$

where $\tilde{p}(\boldsymbol{y}_\mathcal{D}|\boldsymbol{x}) = \int p(\boldsymbol{y}_\mathcal{D}|\boldsymbol{h},\tau,\boldsymbol{x})Q(\boldsymbol{h})Q(\tau)\mathrm{d}\boldsymbol{h}\mathrm{d}\tau$ and $\tilde{p}(\boldsymbol{y}_\mathcal{D}) = \sum_{\boldsymbol{x}}\tilde{p}(\boldsymbol{y}_\mathcal{D}|\boldsymbol{x})P(\boldsymbol{x})$. It is noteworthy that there is no constraints on how does the prior $P(\boldsymbol{x})$ is obtained.

For the purpose of demodulation, the a priori probability of the sequence $\boldsymbol{x}$, $P(\boldsymbol{x}_\mathcal{D} = \boldsymbol{x})$, may be assumed fully factorized, i.e., $P(\boldsymbol{x}) = \prod_k P(x_k)$ with $c_j^{(k)}$ being the j-th bit of the codeword labeling the k-th symbol of $\boldsymbol{x}$. Applying this factorization of $P(\boldsymbol{x}_\mathcal{D})$ to the bitwise APP of c_k in (19) and expressing the results in logarithmic likelihood ratio (LLR) form yields the following:

$$L(c_j) = \ln\frac{\sum_{\forall\boldsymbol{x}:c_j=1}\tilde{p}(\boldsymbol{y}_\mathcal{D}|\boldsymbol{x})\prod_k P(x_k)}{\sum_{\forall\boldsymbol{x}:c_j=0}\tilde{p}(\boldsymbol{y}_\mathcal{D}|\boldsymbol{x})\prod_k P(x_k)} \tag{20}$$

This value then may be fed to a SISO decoder if the bits are coded.

The output LLR values from the decoder may be regarded as an updated a priori information on $c_k's$. Using $P(c_k = c) = \frac{\exp(c\cdot L(c_k))}{1+\exp(L(c_k)}$ the LLR values may further convert to a prior mass of $\boldsymbol{x}_\mathcal{D}$. Figure 2 shows the encapsulated block diagram of this architecture. This suggests an iterative algorithm that provides a general framework for VB based joint channel estimation, demodulation and decoding for a digital receiver. It is especially suitable to work with modern random-like codes such as coded modulation (CM), Turbo codes, LDPC codes and etc. since they are born suitable for SISO decoding algorithms.

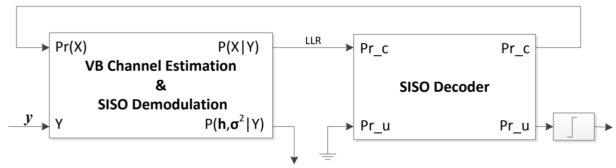

Fig. 2. Joint receiver scheme with VB block and SISO decoder: the VB block takes as input the received symbols and prior distribution of $\boldsymbol{x}$ and gives as output of approximate a posterior distribution of $\boldsymbol{x}_\mathcal{D}$. And the decoder takes as input from VB block the updated distribution over $\boldsymbol{x}_\mathcal{D}$ and produces updated distribution of $\boldsymbol{x}_\mathcal{D}$ (in the form of LLR).

5. IMPLEMENTATION AND SIMULATION RESULTS

In this section aspects of implementation is discussed and the use of the bit-interleaved coded modulation (BICM) with Turbo coding is presented to demonstrate the performance of the proposed scheme.

5.1. Calculation of $\tilde{p}(\boldsymbol{y}_\mathcal{D}|\boldsymbol{x}_\mathcal{D})$ and $L_{ext}(c_k)$

Following the formula for equalization in (20), the a-posteriori distribution $\tilde{p}(\boldsymbol{y}_\mathcal{D}|\boldsymbol{x}_\mathcal{D})$ needs to be evaluated first. In an CP-OFDM system the frequency-selective channel is converted into a set of parallel narrow band channels. Hence, equalization and demodulation can be performed per subcarrier and be arranged in an computationally efficient manner. Let y, H and C_y respectively denote the received symbol, narrow band channel coefficient and noise variance on certain OFDM subcarrier, where subscript for indexing each subcarrier is dropped for convenience. Considering one subcarrier at a time, the approximate a-posteriori probability for a received symbol is given by

$$\tilde{p}(y|x) = \langle p(y|\sqrt{N_c}Hx,\tau)\rangle_{q(H)q(\tau)} \tag{21}$$

where the variational distribution $Q(H)$ can be obtained easily from $Q(\mathbf{h})$. However the marginalization of $p(y|\sqrt{N_c}Hx,\tau)$ with respect to $q(H)$ and $q(\tau)$ cannot be evaluated analytically. We are therefore forced to adopt the following approximations:

$$\tilde{p}(y|x) \approx \mathcal{CN}\left(y|\sqrt{N_c}\langle H\rangle x, C_{y|x}\right), \tag{22}$$

where $C_{y|x} = \langle\tau\rangle^{-1} + N_c|x|^2\Sigma_H$. In this approximation, a Dirac delta function is used to replace the gamma distribution of τ in the expectation. It may be interpreted as a type-II likelihood maximization method used commonly in statistics. It is justified by noticing that 1) the variance of $Q(\tau)$ given by $\tilde{c}/\tilde{d}^2$ in typical SNR region is in the order of 10^{-3}, hence $Q(\tau)$ is heavily concentrated at the vicinity of its mean $\langle\tau\rangle$ and 2) the maximum likelihood approximation is sufficiently accurate for only predictive purposes [3].

Hence LLR values of c_j in (20) can be rewritten as

$$L_j^e = \underbrace{\ln \frac{\sum_{x \in \mathcal{X}_{j,1}} \exp\left(-C_{y|x}^{-1} |y - \sqrt{N_c}\langle H \rangle x|^2 + \sum_{\substack{i=0\\ i\neq j}}^{n-1} b_i^x L_i^a\right)}{\sum_{x \in \mathcal{X}_{j,0}} \exp\left(-C_{y|x}^{-1} |y - \sqrt{N_c}\langle H \rangle x|^2 + \sum_{\substack{i=0\\ i\neq j}}^{n-1} b_i^x L_i^a\right)}}_{\tilde{L}_{ext}(c_k|\boldsymbol{y})} + L_j^x \tag{23}$$

where $\mathcal{X}_{j,b}$ is the subset of symbols from $\mathcal{X}$ having the k-th bit equal to $b \in \{0,1\}$, with $k = 0 \ldots n-1$, and b_j^x is the j-th bit labeling the symbol x and L_j^a denotes a-priori LLR values of the j-th bit. The quantity $L_j^e \triangleq \tilde{L}_{ext}(c_k|\boldsymbol{y})$ is referred to as the *extrinsic information* in literatures. However $\tilde{L}_{ext}(c_k|\boldsymbol{y})$ differs from the conventional extrinsic information in the way it is obtained: it is calculated using the approximate a posterior probability $\tilde{p}(\boldsymbol{y}_{\mathcal{D}}|\boldsymbol{x}_{\mathcal{D}})$ obtained from the current iteration of VB process.

5.2. Simulation Results

An CP-OFDM system operating at 1800Mhz is simulated to verify the performance of the proposed algorithm. Each OFDM symbol consists of $N_c = 1024$ subcarriers, cyclic-prefix of length $N_{cp} = 144$ which gives a total OFDM symbol duration of $N = N_c + N_c = 1168$ time samples. The transmitted bits are coded using 1/3-rate Turbo code specified in the 3GPP TS 36.212 technical specification. The coded bits are then scrambled and mapped to QAM-4,16 or 64 complex valued symbols. For simplicity, it is also assumed that the maximum channel length is same as the length of the CP for the worst case scenario. The multipath channel have $N_{path} = 6$ significant multipath component whose location τ_i in delay is chosen from a uniform distribution and the complex delay gain h_l is a complex circularly symmetric Gaussian random variable. The number of significant multipath components and their location are unknown to the receiver. The performance of proposed algorithm is compared side-by-side with a receiver working under decision directed scheme with MMSE channel estimator. For all cases, the performance is measured by bit-error rate (BER).

In Figure 3, the BER performance of the proposed algorithm is compared side by side with conventional MMSE receiver for different modulation orders. The 64-QAM, 16-QAM and 4-QAM modulation along with 1/3-rate Turbo code give overall bandwidth efficiency $\eta = 2, 1\frac{1}{3}$ and $\frac{2}{3}$ bits/symbol, respectively. With the proposed algorithm, about 2dB gain in SNR at $\text{BER} = 10^{-6}$ is observed over the conventional receiver working on MMSE principle for channel estimation and equalization. Also the BER performance of the proposed receiver given perfect CSI is also given to show the adverse effect of channel mismatch to the performance. In this experiment, updating steps involving variable $\mathbf{h}$, τ and $\boldsymbol{\alpha}$ is bypassed and true CSI is passed to the equalizer. And iterations are performed between the equalizer and the decoder. In this case, only 1dB or less of loss in SNR is observed at 10^{-6}.

In Figure 4 the behavior of the algorithms in terms of convergence is revealed by carrying out simulated experiment with 64-QAM modulation under AWGN channel with SNR $= 8dB$. This empirical experiments show that for decoding purpose only the algorithm converges after seven iterations. In higher SNR situations, the algorithm converges under five iterations. In practice, adjusting the maximum number of iterations provides a trade-off in computation complexity and error performance. Interestingly, after the BER converges at the seventh iteration, the quality of the channel estimate continues to improve significantly before approximately the 10th iteration. This indicates a mismatch in information extracting power among all updating steps. Since the VB principle allows updating variables in an arbitrary order this phenomenon indicates better updating sequences may exist besides the round-robin scheme.

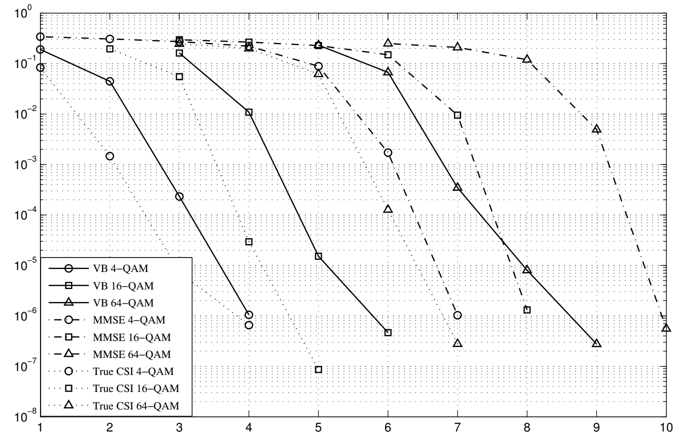

Fig. 3. Bit Error Rate vs. SNR Performance of Hybrid Variational Bayesian Receiver with 1/3-Rate Turbo Coding

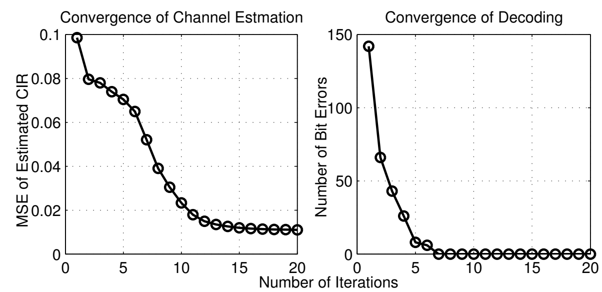

Fig. 4. Rate of Convergence

6. CONCLUSION

In this paper, by the use of graphical model and Bayesian variational method, the proposed algorithms for OFDM receiver jointly performs channel estimation, detection and decoding in a unified framework. Comparing to related prior works, the proposed algorithm improves in various aspects both theoretically and practically. In the seminal work on Turbo equalization [6], the CSI that the equalizer operates on is assumed accurate, which is generally unrealistic in practice. The proposed algorithm includes the channel estimation stage into the iterative loop and carries the soft information of the CSI onto the demodulation and decoding stages naturally. It therefore provides more robust performance under practical scenarios. Moreover, in [7] a variational message passing (VMP) based receiver is proposed. However, since the VMP is only valid for model with conjugate distributions, an Gaussian prior is assumed for the transmitted symbols, which has been emphasized as untrue in Section 4. This work also provides a more robust channel estimation result by the use of a hierarchical graphical model which promotes a succinct or sparse and hence robust solution on the CIR.

7. REFERENCES

[1] G.L. Stuber and M. Russell, "Terrestrial digital video broadcasting for mobile reception using ofdm," in *Global Telecommunications Conference, 1995. GLOBECOM '95., IEEE*, nov 1995, vol. 3, pp. 2049 –2053 vol.3.

[2] Judea Pearl, *Causality: models, reasoning, and inference*, Cambridge University Press, New York, NY, USA, 2000.

[3] Michael E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, September 2001.

[4] Steve Waterhouse, David Mackay, and Tony Robinson, "Bayesian methods for mixtures of experts," in *In*. 1996, pp. 351–357, MIT Press.

[5] Chulong Chen and M.D. Zoltowski, "Bayesian sparse channel estimation and tracking," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, aug. 2012, pp. 472 –475.

[6] R. Koetter, A.C. Singer, and M. Tuchler, "Turbo equalization," *Signal Processing Magazine, IEEE*, vol. 21, no. 1, pp. 67 – 80, jan. 2004.

[7] G.E. Kirkelund, Manchon C.N., L.P.B. Christensen, E. Riegler, and B.H. Fleury, "Variational message-passing for joint channel estimation and decoding in mimo-ofdm," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, dec. 2010, pp. 1 –6.