

Variational Bayesian Method for Joint Sparse Channel Estimation and Data Detection in Frequency-Selective Rayleigh-Fading Systems

Zhang Kai¹⁺, Yu Hongyi¹, Hu Yunpeng¹ and Shen Zhixiang¹

¹ The National Digital Switching System Engineering and Technological R&C Center of China

Abstract. This paper deals with the problem of data detection in sparse frequency selective Rayleigh-fading channels. An iterative scheme for joint channel estimation and data detection is proposed based on maximum *a posteriori* (MAP) criterion. We begin with a general sampled representation of the nonlinear sparse multipath channel, which is parameterized by a linear counterpart at a resolution commensurate with the sampling frequency. Then, a hierarchical system model is developed, in which priors for all model parameters, i.e., channel coefficients, symbols, and noise precision parameter, are properly chosen. In particular, the channel sparsity is utilized by defining parametric heavy-tailed priors for its coefficients. And then the optimal joint estimation framework based on MAP criterion is developed. Since direct maximization of the posterior probability density function is infeasible, the variational Bayesian method is invoked as a feasible method to simplify the problem. The estimation of the channel and data are dealt with jointly and iteratively. The superiority of the proposed schemes with respect to several typical methods is presented by both theoretical analysis and Monte Carlo simulations.

Keywords: Sparse channel estimation, Rayleigh-Fading channels, maximum *a posteriori* estimation, variational Bayesian inference

1. Introduction

Over the past decades, the signal detection technology has been widely studied. When the channel impulse response (CIR) or its statistical parameters is unknown, channel estimation procedure must be performed before the detection of the transmitted symbols. Joint channel estimation and data detection techniques, which can save the pilot overhead and improve the communication efficiency, have received considerable attention. In [1], the maximum-likelihood sequence detection and parameter estimation (MLSDE) algorithm is developed, which performs coupled parameter estimation and symbol detection in an iterative scheme. In [2], under the Bayesian framework, the transmitted data sequence is detected by averaging over unknown multipath channel parameters in an EM-based iterative algorithm.

Present experiments have shown that wideband wireless communication channels are always sparse [3]. The approaches [1] and [2] do not specifically utilize this channel property. The statistics of the channel is necessary to derive reliable channel estimate. Otherwise, the associated channel estimation error can be high. This will limit the use of these techniques. Compressive sensing techniques have been widely utilized to the estimation of the sparse channels. Generally, the channel coefficients are regularized by a sparsity-inducing prior. In [4], the basis pursuit (BP) algorithm is applied to channel estimation by incorporating a l1-norm sparsity regularization term on the channel coefficients. It is shown that the BP-based method outperforms the traditional non-sparse estimation methods under sparse channel situations due to the usage of the inherent channel sparsity.

Different from the BP scheme, the SBL method [5] employs a hierarchical structure to model the unknown channel coefficients, which are governed by independent scalar parameters. The conventional EM

⁺ Corresponding author. Tel.: + (86-13673370679); fax: +(0371-60135535).
E-mail address: zk_xxgc@163.com.

algorithm and the SBL technique are combined to perform channel parameters estimation and symbol detection in an interactive manner. The SBL approach estimates both the sparse channel vector and its second order statistics, which is different from the deterministic BP-based approach [4]. It is verified that the performance of the SBL scheme is superior than the BP algorithms. However, there are still several limitations within the SBL scheme. Firstly, the SBL procedure is performed only on a symbol-spaced receiver. Fractionally-spaced receiver, which has been proved to have better performance [6][7], is not included. Secondly, the independence of channel coefficients is not retained since the posterior is non-factorial (see Section 3.2 for details). Thirdly, SBL models the channel coefficients as Gaussian, conditioned on uniformly distributed hyperparameters, which is just a limiting special case. Other parameter selection is not considered. In addition, the computational complexity of SBL scheme is moderate high, mainly from the matrix inversion operation in calculating the posterior covariance matrix. To overcome these shortcomings, we employ the sparse variational Bayesian (VB) framework presented in [8] for joint channel estimation and data detection. The main contributions of this paper are listed as follows.

First, a unified framework for joint estimation of data and sparse channels based on MAP criterion is developed. The term unified has two parts of meanings. (a) A unified processing architectures, containing both symbol-spaced and fractionally-spaced, are included in our framework under the virtual channel model. (b) The channel is constrained by a universal hierarchical prior, from which various sparsity-inducing priors, such as Student's t, and Laplace, can be obtained via different selections of hyperparameters in the hierarchical construction. This makes it possible for us to evaluate the system performance for different sparsity inducing channel priors and different sampling factors. Second, since direct maximization of the posterior probability density function is infeasible, the variational Bayesian method [8] is employed as a feasible method to estimate the channel and symbols. The proposed procedure operates in an iterative manner, from which posterior approximations of all model parameters, instead of just point estimates, are obtained. Finally, the superiority of the proposed VB procedure with respect to several typical methods is presented by theoretical analysis and simulation experiments.

2. System Model

The block diagram communication system is presented in Fig. 1, where $s(n) \in S$ is the transmitted information symbol with S being the modulation constellation set, and $p_T(t)$ and $p_R(t)$ represent the impulse responses of the transmit filter and receive filter, respectively. $v(t)$ is the additive white Gaussian noise (AWGN) with variance N_0 . $c(t)$ represents the impulse response of a multipath channel, and can be expressed as

$$c(t) = \sum_{p=0}^{P-1} c_p \delta(t - \tau_p) \quad (1)$$

where $\delta(\cdot)$ is the Dirac delta function, P denotes the number of physical propagation paths, c_p and $\tau_p \in [0, \tau_{max}]$ represent the complex path gain and propagation delay of the p -th path, respectively, with τ_{max} being the maximum delay. We assume that the channel is time-invariant during the observing interval.

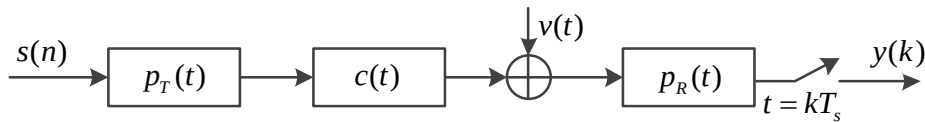


Fig. 1: Block diagram of the baseband communication system.

Thus, the output of the receive filter can be expressed as

$$y(t) = \sum_{p=0}^{P-1} c_p x(t - \tau_p) + z(t) \quad (2)$$

where $x(t) = \sum_{n=0}^{N-1} s(n) p(t - nT)$, N is the symbol length, $p(t) = p_T(t) * p_R(t)$ is the convolution of the transmit filter and the receive filter, with $*$ being the convolution operation, and $z(t) = v(t) * p_R(t)$.

In the sequel, the signal $y(t)$ is sampled at the rate of $F_s = 1/T_s$. Here, we assume that $u/T = 1/T_s$ for some integer $u \geq 1$. The sampled output is

$$y(kT_s) = \sum_{p=0}^{P-1} c_p x(kT_s - \tau_p) + z(kT_s), k = 0, \dots, uN - 1 \quad (3)$$

While the physical discrete path model in (3) is realistic, it is difficult to analyze due to its *nonlinear* dependence on the parameters $\{c_p, \tau_p\}$. However, because of the finite signal bandwidth, the physical model can be well approximated by a linear counterpart with the aid of Fourier series expansion [9]. That is,

$$h_l = \sum_{p \in S_{\tau,l}} c_p \text{sinc}(l - F_s \tau_p), l = 0, \dots, L-1 \quad (3)$$

with $\text{sinc}(a) = \sin(\pi a) / \pi a$, $L = \lceil F_s \tau_{\max} \rceil + 1$ representing the maximum number of virtual delays, and $S_{\tau,l} = \{i : \tau_i \in [l/F_s - 1/2F_s, l/F_s + 1/2F_s]\}$ denoting the l -th virtual delay bin. Let $\mathbf{h} = [h_0, \dots, h_{L-1}]^T \in \mathbb{C}^{L \times 1}$ denote the vector containing all the virtual channel coefficients. When $\|\mathbf{h}\|_0 < L$, i.e., not all delay bins have physical paths, we refer to such channels as sparse channels. In practice, the channel sparsity is the result of sparse distribution of reflectors in space [9]. Under Rayleigh-fading environment, the path gains, c_p , can be typically modeled as independent zero-mean complex Gaussian variables. Then, the virtual channel coefficient vector $\mathbf{h}$ is a zero-mean random vector. From (3), it is clear that the channel covariance matrix $\mathbf{\Sigma}_h = \text{diag}(\sigma_0^2, \dots, \sigma_{L-1}^2)$ is diagonal. We assume that there are sufficiently many unresolvable paths contributing to each non-zero channel coefficient. Invoking the central limit theorem, the channel coefficient vector $\mathbf{h}$ is Gaussian, i.e., $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_h)$. Then, the system (3) can be approximately expressed as

$$y(k) = \sum_{l=0}^{L-1} h_l x(k-l) + z(k), k = 0, \dots, uN-1 \quad (4)$$

where $y(k) = y(kT_s)$, $x(k-l) = x(kT_s - lT_s)$, and $z(k) = z(kT_s)$. The matched filter output $y(k)$ provide a sufficient statistic for extracting the transmitted symbols and channel coefficients. In matrix form, equation (4) can be represented as

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{z} \quad (5)$$

where $\mathbf{y} = [y(0), \dots, y(uN-1)]^T \in \mathbb{C}^{uN \times 1}$, $\mathbf{z} = [z(0), \dots, z(uN-1)]^T \in \mathbb{C}^{uN \times 1}$, and $\mathbf{X}$ is a $uN \times L$ convolution Toeplitz matrix. The noise vector $\mathbf{z}$ is zero-mean complex Gaussian distributed with a variance matrix $\mathbf{R}_z = E[\mathbf{z}\mathbf{z}^H] = N_0 \mathbf{R}_p$, where $\mathbf{R}_p \in \mathbb{R}^{uN \times uN}$ is constructed by the autocorrelation function of the receive filter. It is obviously to see that, if $\mathbf{R}_p$ is a diagonal matrix, the noise component $z(k)$ is still white. Otherwise, $z(k)$ is colored due to the effects of the receive filter. Define the pseudo-inverse of $\mathbf{R}_z$ as $\mathbf{R}_z^\dagger = N_0^{-1} \mathbf{V} \mathbf{\Omega}^{-1} \mathbf{V}^H$ with $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{u_p}] \in \mathbb{C}^{uN \times u_p}$ and $\mathbf{\Omega} = \text{diag}[\omega_1, \omega_2, \dots, \omega_{u_p}] \in \mathbb{C}^{u_p \times u_p}$, where u_p is the number of non-zero eigenvalues of $\mathbf{R}_p$, $\mathbf{\Omega}$ is a diagonal matrix, with the diagonal elements, $\{\omega_i\}_{i=1}^{u_p}$, being the non-zero eigenvalues of $\mathbf{R}_p$, and $\{\mathbf{v}_i\}_{i=1}^{u_p}$ are the corresponding orthonormal eigenvectors.

Let $\mathbf{D} = \mathbf{\Omega}^{-1/2} \mathbf{V}^H \in \mathbb{C}^{u_p \times uN}$, multiplying $\mathbf{D}$ on both sides of (5), we will derive an equivalent system

$$\bar{\mathbf{y}} = \bar{\mathbf{X}}\mathbf{h} + \bar{\mathbf{z}} \quad (6)$$

where $\bar{\mathbf{y}} = \mathbf{D}\mathbf{y}$ is a vector containing the output of the whitened matched filter (WMF), $\bar{\mathbf{X}} = \mathbf{D}\mathbf{X}$ is the whitened convolution matrix, $\bar{\mathbf{z}} = \mathbf{D}\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, N_0 \mathbf{I})$ is the whitened noise vector. The ultimate goal is to estimate the transmitted symbols from the WMF output $\bar{y}(n)$, $n = 0, \dots, u_p - 1$.

3. Joint Sparse Channel Estimation and Data Detection Using Sparse Variational Bayesian Algorithm

3.1. Derivation of the VB scheme

Given the system model in Section 2, the optimal estimate can be derived by maximizing the full posterior $p(\boldsymbol{\theta}|\bar{\mathbf{y}})$, where $\boldsymbol{\theta}$ is a vector containing all parameters to be estimated. Using the Bayes rule, the ideal full posterior can be written as $p(\boldsymbol{\theta}|\bar{\mathbf{y}}) = p(\bar{\mathbf{y}}, \boldsymbol{\theta}) / \int p(\bar{\mathbf{y}}, \boldsymbol{\theta}) d\boldsymbol{\theta}$. It is shown that the joint distribution $p(\bar{\mathbf{y}}, \boldsymbol{\theta})$ is the product of the conditional distribution for the vector $\bar{\mathbf{y}}$ and the prior distributions of unknown variables. Under the system model in (11), the conditional distribution for $\bar{\mathbf{y}}$ is multivariate complex Gaussian, i.e., $p(\bar{\mathbf{y}}|\mathbf{h}, \beta) = \mathcal{CN}(\bar{\mathbf{y}}|\bar{\mathbf{X}}\mathbf{h}, \beta^{-1}\mathbf{I})$, where $\beta = 1/N_0$ is the noise precision parameter of the observation model. A *conjugate* Gamma distribution is placed on the parameter β , i.e., $p(\beta) = G(\rho, \delta)$. As illustrated in Section 2, the channel coefficients h_l are approximately statistically independent, zero-mean complex Gaussian random variables. And, many elements in $\mathbf{h}$ are zero under sparse channel scenarios. To this end, a hierarchical heavy-tailed prior is chosen for the channel coefficient vector $\mathbf{h}$. In the first level, we choose a Gaussian prior distribution over $\mathbf{h}$, i.e., $p(\mathbf{h}|\boldsymbol{\alpha}, \beta) = \prod_{l=0}^{L-1} \mathcal{CN}(h_l|0, (\beta\alpha_l)^{-1})$, with $\beta\alpha_l = 1/\sigma_l^2$ being the precision parameter of h_l , $\boldsymbol{\alpha} = [\alpha_0, \dots, \alpha_{L-1}]^T$, and $\mathbf{A} = \text{diag}(\alpha_0, \dots, \alpha_{L-1})$. Note that $\mathbf{h}$ is conditioned on β , which ensures the unimodality of the posterior joint distribution [10][11].

Then the marginal probability distribution of $\mathbf{h}$ can be deduced by integrating out the latent variables α and β as $p(\mathbf{h}) = \int p(\mathbf{h}|\alpha) p(\alpha) p(\beta) d\alpha d\beta$. Various heavy-tailed (marginal probability) distributions can be obtained for $\mathbf{h}$ by different choice of $p(\alpha)$. In this paper, we will consider three special and widely used cases [8], i.e., the Student's-t distribution, the Laplace distribution, and the multi-parameter Laplace distribution. To provide unification for these three distributions in a single model, the prior for α_l is chosen as the generalized inverse Gaussian (GIG) as given by [8]

$$p(\alpha_l; a, b_l, c) = \text{GIG}(\alpha_l; a, b_l, c) = \frac{(a/b_l)^{(c/2)}}{2K_c(\sqrt{ab_l})} \alpha_l^{c-1} \exp\left[-\frac{1}{2}\left(a\alpha_l + \frac{b_l}{\alpha_l}\right)\right] \quad (7)$$

where $a, b_l > 0$, $c \in \mathbb{R}$ and $K_c(\cdot)$ is the modified Bessel function of the second kind. In equation (7), the parameters a , c and b_l can be deterministic or random variables. Different selections will result in different marginal probability distributions for $\mathbf{h}$. Finally, since the symbol vector $\mathbf{s}$ is deterministic but unknown parameter that need point estimate, we select $p(\mathbf{s}) = \sum_{i=1}^M M^{-N} \delta(\mathbf{s} - \mathbf{s}_i)$, where M represents the modulation level, and $\mathbf{s}_i \in S^N$ represents a possible symbol sequence. Then, the joint distribution of all unknown and observed quantities can be obtained. Unfortunately, since the integral in the denominator of (12) is intractable, the posterior cannot be analytically calculated, i.e., some approximation must be employed to simplify the problem.

In the following, we utilize the variational Bayesian method to approximate $p(\boldsymbol{\theta}|\bar{\mathbf{y}})$ by a variational distribution $q(\boldsymbol{\theta})$. Within the VB framework, the parameters of $q(\boldsymbol{\theta})$ are selected so as to minimize the following Kullback-Leibler (KL) divergence

$$\text{KL}(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta}|\bar{\mathbf{y}})) = \int q(\boldsymbol{\theta}) \log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\bar{\mathbf{y}})} d\boldsymbol{\theta} \quad (8)$$

Since the parameter vector $\boldsymbol{\theta}$ is the combination of several (approximately) disjoint sets, the variational distribution $q(\boldsymbol{\theta})$ can be approximately factorized as

$$q(\boldsymbol{\theta}) \approx q(\mathbf{s}) q(\beta) \prod_{l=0}^{L-1} q(h_l) q(\alpha_l) q(b_l) \quad (9)$$

Then, the optimal solutions for each factor $q(\boldsymbol{\theta}_k)$ can be calculated as

$$q(\boldsymbol{\theta}_k) \propto \exp\left[\int q(\boldsymbol{\theta}_{-\boldsymbol{\theta}_k}) \log p(\bar{\mathbf{y}}, \boldsymbol{\theta}) d\boldsymbol{\theta}_{-\boldsymbol{\theta}_k}\right] \quad (10)$$

where $\boldsymbol{\theta}_k \in \boldsymbol{\theta}$ denotes a subset of $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_{-\boldsymbol{\theta}_k}$ is constructed by removing $\boldsymbol{\theta}_k$ from $\boldsymbol{\theta}$. It is obvious to see from (10) that each distribution $q(\boldsymbol{\theta}_k)$ is iteratively updated by holding all the other factors fixed. This gives rise to a cyclic optimization scheme. In the following part, we provide the explicit forms of the update rules for all unknown factors.

A. Estimation of the channel coefficients

Applying (18) to the proposed model, the posterior approximation of h_l is found to be Gaussian as

$$q(h_l) \sim \mathcal{CN}(h_l; \langle h_l \rangle, \lambda_l^{-1}) \quad (11)$$

with posterior mean $\langle h_l \rangle$ and posterior precision parameter λ_l given by

$$\langle h_l \rangle = \lambda_l^{-1} \langle \beta \rangle \bar{\mathbf{x}}_l^H (\bar{\mathbf{y}} - \bar{\mathbf{X}}_{-l} \langle \mathbf{h}_{-l} \rangle) \quad (12)$$

$$\lambda_l = \langle \beta \rangle (\langle \alpha_l \rangle + \bar{\mathbf{x}}_l^H \bar{\mathbf{x}}_l) \quad (13)$$

where $\bar{\mathbf{X}}_{-l}$ is constructed from $\bar{\mathbf{X}}$ by removing its l -th column $\bar{\mathbf{x}}_l$, $\mathbf{h}_{-l}$ is constructed from $\mathbf{h}$ by removing its l -th element h_l , and $\langle \mathbf{h}_{-l} \rangle$ is constructed from $\langle \mathbf{h} \rangle = [\langle h_0 \rangle, \dots, \langle h_{L-1} \rangle]^T$ by removing its l -th element. Note that the independence of the channel coefficients h_l is retained due to the *fully* channel factorization scheme in (9).

B. Estimation of the noise precision parameter

Working in a similar manner, we can derive that $q(\beta) = G(\tilde{\rho}, \tilde{\delta})$ where $\tilde{\rho} = u_\rho + L + \rho$, $\tilde{\delta} = \|\bar{\mathbf{y}} - \bar{\mathbf{X}} \langle \mathbf{h} \rangle\|^2 + \text{tr}(\bar{\mathbf{X}}^H \bar{\mathbf{X}} \boldsymbol{\Sigma}) + \sum_{l=0}^{L-1} \langle \alpha_l \rangle \langle h_l^2 \rangle + \delta$ with $\langle h_l^2 \rangle = \langle h_l \rangle^2 + \lambda_l^{-1}$, and $\boldsymbol{\Sigma} = \text{diag}(\lambda_0^{-1}, \dots, \lambda_{L-1}^{-1})$ being the posterior covariance matrix of $\mathbf{h}$. Thus, the mean value of β is $\langle \beta \rangle = \tilde{\rho} / \tilde{\delta}$.

C. Estimation of the symbol vector

The approximating posterior for the symbol vector $\mathbf{s}$ is

$$q(\mathbf{s}) \propto \exp \left[-\|\bar{\mathbf{y}} - \bar{\mathbf{X}}\langle \mathbf{h} \rangle\|^2 - \text{tr}(\bar{\mathbf{X}}^H \bar{\mathbf{X}} \Sigma) \right] \quad (14)$$

By specifying a Dirac delta function to the symbol vector, we restrict ourselves to point estimates of the transmitted symbols. Thus, the approximate optimal symbol vector estimation is given by

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s} \in S^N} \left[\|\bar{\mathbf{y}} - \bar{\mathbf{X}}\langle \mathbf{h} \rangle\|^2 + \text{tr}(\bar{\mathbf{X}}^H \bar{\mathbf{X}} \Sigma) \right] \quad (15)$$

D. Estimation of the precision parameters α_l

For the parameters α_l , we consider three special cases, i.e., Student's-t, Laplace and multi-parameter Laplace.

(1) Student's-t distribution

Let $b_l \rightarrow 0$ and $c > 0$, the GIG prior in (7) reduces to a Gamma distribution with $p(\alpha_l; a, c) = G(\alpha_l; c, a/2)$. Using (13) and (14), it is easily to verify the marginal distribution for the channel coefficients is Student's-t distribution. The variational posterior for α_l is Gamma, i.e., $q(\alpha_l) = G(\alpha_l; c+1, \langle \beta \rangle \langle h_l^2 \rangle + a/2)$. The mean value of α_l is

$$\langle \alpha_l \rangle = \frac{c+1}{\langle \beta \rangle \langle h_l^2 \rangle + a/2} \quad (16)$$

(2) Laplace and multi-parameter Laplace distributions

We further choose a Gamma prior for the parameters b_l 's, i.e., $p(b_l) = G(b_l | \kappa, \nu)$. Let $a \rightarrow 0$ and $c = -3/2$, the prior α_l reduces to

$$p(\alpha_l; b_l) = \frac{2^{c+1}}{\Gamma(-c) b_l^c} \alpha_l^{c-1} \exp \left(-\frac{b_l}{2\alpha_l} \right) \quad (18)$$

In this case, the marginal probability distribution of $\mathbf{h}$ can be obtained as multi-parameter Laplace (mpL), which can be expressed as $p(\mathbf{h} | \mathbf{b}, \beta) \propto \prod_{l=0}^{L-1} \exp(\sqrt{2\beta b_l} |h_l|)$. The variational distributions $q(\alpha_l)$ and $q(b)$ become

$$q(\alpha_l) = \text{GI } G(\alpha_l; 2\langle \beta \rangle \langle h_l^2 \rangle, \langle b_l \rangle, -1/2) \quad (17)$$

$$q(b_l) = G(b; 3/2 + \kappa, \nu + \langle 1/\alpha_l \rangle / 2) \quad (18)$$

where $\langle 1/\alpha_l \rangle = 1/\langle \alpha_l \rangle + 1/\langle b_l \rangle$. Thus, the posterior mean values of α_l and b are

$$\langle \alpha_l \rangle = \sqrt{\langle b_l \rangle / (2\langle \beta \rangle \langle h_l^2 \rangle)} \quad (19)$$

$$\langle b_l \rangle = \frac{3/2 + \kappa}{\nu + \langle 1/\alpha_l \rangle / 2} \quad (20)$$

Particularly, if we select a single variable $p(b) = G(b | \kappa, \nu)$ to replace all b_l in (7), and also set $a \rightarrow 0$ and $c = -3/2$, it is easily to be verified that the marginal probability distribution of $\mathbf{h}$ is Laplace with a single parameter governing all channel coefficients. In this case, the variational distributions $q(\alpha_l)$ and $q(b)$ become

$$q(\alpha_l) = \text{GI } G(\alpha_l; 2\langle \beta \rangle \langle h_l^2 \rangle, \langle b \rangle, -1/2), l = 0, \dots, L-1 \quad (23)$$

$$q(b) = G\left(b; 3L/2 + \kappa, \nu + \frac{1}{2} \sum_{l=0}^{L-1} \langle 1/\alpha_l \rangle\right) \quad (21)$$

Thus, the posterior mean values of α_l and b are

$$\langle \alpha_l \rangle = \sqrt{\langle b \rangle / (2\langle \beta \rangle \langle h_l^2 \rangle)} \quad (22)$$

$$\langle b \rangle = \frac{3L/2 + \kappa}{\nu + 0.5 \sum_{l=0}^{L-1} \langle 1/\alpha_l \rangle} \quad (23)$$

Up to now, we have completed one iteration with approximate posterior for all unknown variables. It is obvious to see that variational distribution for each parameter is the function of the expectations of the other parameters. This gives rise to an iterative scheme for joint channel estimation and symbol detection. Note that the notations VB-S, VB-L, and VB-mpL represent the proposed VB-based procedure with Student's t, Laplace, and multi-parameter Laplace channel priors, respectively. Since the factors $q(\alpha_i)$, $q(\beta)$, and $q(h_i)$ are all convex, the VB algorithm can always converge in a few cycles.

4. Discussion on the Proposed VB procedure

Thus far, we have developed an iterative sparse variational Bayesian scheme for joint channel estimation and data detection. In this section, we will make a brief discussion of the proposed procedure as compared with BP [4] and SBL [5] methods. First, both VB and SBL belong to *Type II* Bayesian method, while BP belongs *Type I* Bayesian method [12]. Since the samples $x(k)$ in equation (4) are correlated with each other due to the effects of the receive filter, the matrix $\bar{\mathbf{X}}$ will exhibit correlation property between its columns and/or rows. The mutual coherence of $\bar{\mathbf{X}}$, which can be calculated as

$$\mu(\bar{\mathbf{X}}) = \max_{i \neq j} \frac{|\bar{\mathbf{x}}_i^H \bar{\mathbf{x}}_j|}{\|\bar{\mathbf{x}}_i\|_2 \|\bar{\mathbf{x}}_j\|_2} \quad (24)$$

is shown in Table 2 for QPSK modulated signals. Both the transmit and receive filters are the square root raised cosine (SRRC) filter. The roll-off factor is set to $\alpha = 0.1/0.5/1$. And, the oversampling factor is set to $u=1$ and 2, respectively. It is clear to see that when $u=1$, the mutual coherence values of $\bar{\mathbf{X}}$ are approximately equal to zero, which means that the columns of $\bar{\mathbf{X}}$ are approximately independent. However, when $u=2$, the matrix $\bar{\mathbf{X}}$ is characterized by highly correlated columns. And the smaller the α value, the larger the value of $\mu(\bar{\mathbf{X}})$, which indicates a stronger correlation property. In this case, *Type II* methods, which handle dictionary correlations using an implicit dictionary-dependent sparse penalty, are expected to maintain desirable advantages over *Type I* methods, especially in oversampling cases. This will be verified by simulations in Section 5. Second, the channel coefficients in SVB procedures are posterior independent, meaning the posterior approximation of $\mathbf{h}$ is factorial, i.e., $q(\mathbf{h}) = \prod_{l=0}^{L-1} q(h_l)$. However, in SBL scheme, the channel coefficients are posterior dependent since the posterior covariance matrix is always not a diagonal matrix. The only exception occurs when $\bar{\mathbf{X}}^H \bar{\mathbf{X}} = \mathbf{I}$. Third, proposed VB approaches derive variational distribution for all parameters. However, the SBL method derives posterior distribution for the channel coefficients, and point estimates for the hyperparameters. BP only derive point estimates for all parameters. SBL and BP can be viewed as a special limiting case of VB. Finally, due to the factorization mechanism in (9), the computational complexity in the calculation of the posterior covariance matrix will be significantly reduced as compared to SBL.

Table 1: Mutual Coherence of $\bar{\mathbf{X}}$ (QPSK)

	$\alpha = 1$	$\alpha = 0.5$	$\alpha = 0.1$
$u = 1$	0.05	0.05	0.05
$u = 2$	0.50	0.60	0.64

5. Simulation Results

In this section, we present experimental results to evaluate the performance of the proposed procedure. During these simulations, the channels were generated based on the model used in [13]. We assume the path delays τ_p uniformly distributed in $[0, \tau_{\max}]$. The number of physical paths P is Poisson distributed with mean $\langle P \rangle = 5$. The complex path gains c_p are independent zero-mean complex circular symmetric Gaussian random variables with variance $\text{var}(c_p) = C \cdot \exp(-\tau_p/\nu)$, and $C, \nu > 0$. The parameter C is computed such that $\sum_{p=1}^P E(|c_p|^2) = 1$. Here, we set $\tau_{\max} = 30T$, and $\nu = 8T$. It is assumed that the channel is quasistatic and remains constant in the observing interval. The input sequence consisted of uncoded QPSK symbols. Both the transmit and receive filters are SRRC filter with a roll-off factor $\alpha = 1$. The sampling factor was set to

$u=1$ and 2, which corresponds to symbol- and fractionally-spaced receivers, respectively. The hyperparameters ρ , δ , κ and ν are set close to zero so as to create non-informative prior for β and b_i . Besides, in VB-S scheme, hyperparameters a and c are also set close to zero, thus creating non-informative prior for α_i . Initial values of the channel precision parameters α_i are generally set to $\alpha_i = 1/L, i = 0, \dots, L-1$. In all the following simulations, $K = 30$ symbols are utilized to derive an initial channel coefficients values for all methods. The block size of symbols used in the iterative procedure is set to $N = 50$. Moreover, since the optimum symbol detection scheme in (26) is often prohibitive in practice due to its huge computational complexity, the block decision feedback equalizer (BDFE) [14] is employed in our simulations to detect the symbols in each iteration. The system performance is measured in terms of the symbol error rate (SER) and the normalized root-mean-square-error (NRMSE).

A. Comparison of the proposed VB algorithm under different priors

Firstly, we evaluate the performance of the proposed VB algorithm with different priors and different sampling factors. Fig. 2 shows the SER curves of the VB-S, VB-L, and VB-mpL versus the E_s/N_0 values. Moreover, the LS method that operates on the *a priori* known support set (locations of nonzero elements) of the channel coefficients, termed as the genie aided LS (GALS), is also included to serve as a benchmark. Fig. 3 shows the NRMSE performance curves versus E_s/N_0 values. The performance curves shown in Fig. 2 and Fig. 3 indicate that fractionally-spaced receivers ($u = 2$) have better performance than the symbol-spaced receivers ($u = 1$), especially for the high E_s/N_0 operating range. The performance improvement mainly comes from the multipath diversity contributed by fractionally-spaced processing architecture. Besides, we can derive that the performance of the variational Bayesian algorithms with Student's-t (VB-S) and multi-parameter Laplace (VB-mpL) is almost identical, both of which are significantly better than the VB-L method for both symbol- and fractionally-spaced cases.

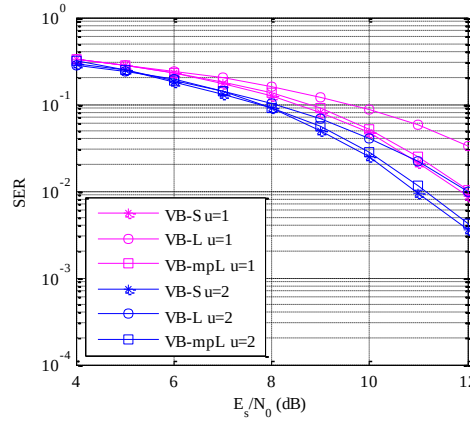


Fig. 2: SER performance of the proposed VB schemes with different channel priors and different oversampling factors.

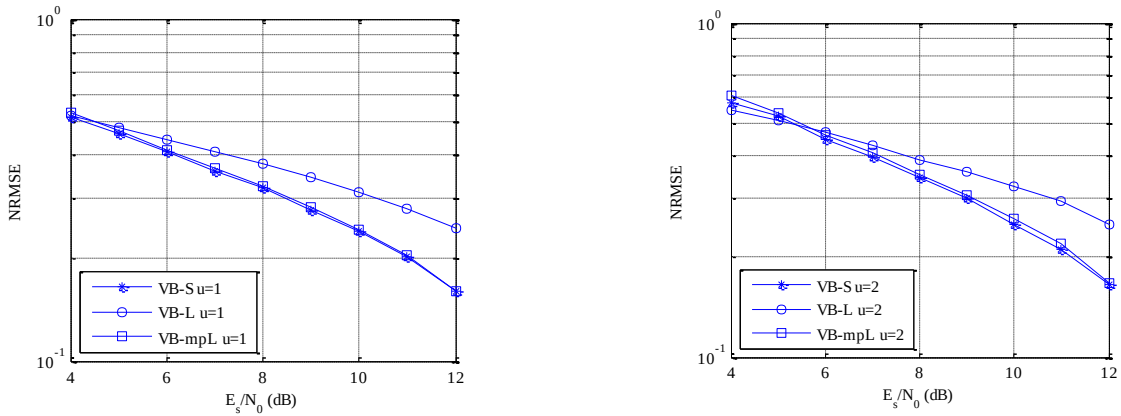


Fig. 3: NRMSE performance of the proposed VB schemes with different channel priors and different sampling factors.

B. Comparison of VB-S and SBL methods

Then, we compare SER performance of the proposed VB-S algorithm with SBL-based algorithm [5]. Within the VB-S scheme, by setting ρ , δ , a and c close to zero, we obtain uniform distributed hyperpriors which is the same as SBL. Then we can make a fair comparison between SBL and VB-S schemes. Note that the original SBL scheme in [5] is only applicable to symbol-spaced receiver. Here, we have extended it to our system model. It is shown in Fig. 4 that VB-S outperforms SBL for both symbol-spaced and fractionally-spaced cases. This is mainly due to the channel vector factorization mechanism in VB scheme depicted in Eq.(17), which constrain the channel posterior independence.

C. Compared to BP method

Finally, the performance comparison of the proposed VB-L and VB-mpL algorithms over BP-based [4] algorithm is shown in Fig. 5. Since the BP-based scheme is inherently sensitive to the regularization parameter. The BP curve presented in Fig. 5 is derived after fine-tuning the regularization parameter values of BP through extensive experiments. On the other hand, the proposed VB-L and VB-mpL are fully automatic. When $u=1$, the proposed variational Bayesian algorithm with multi-parameter Laplace channel prior (VB-mpL) achieves the best SER performance in moderate and high E_s/N_0 values, for both symbol- and fractionally-spaced cases. And, the performance difference between VB-L and BP is negligible. However, in low E_s/N_0 values, the parameter estimation performance within VB is poor. Thus, the SER performance of VB-L and VB-mpL is poorer than the BP approach as shown in Fig. 5. However, when $u = 2$, since the matrix $\bar{\mathbf{X}}$ is characterized by highly correlated columns as illustrated in Section 3.2, the performance of BP, which belongs to *Type I* Bayesian estimator, is poorer than both VB-L and VB-mpL, especially in low E_s/N_0 range.

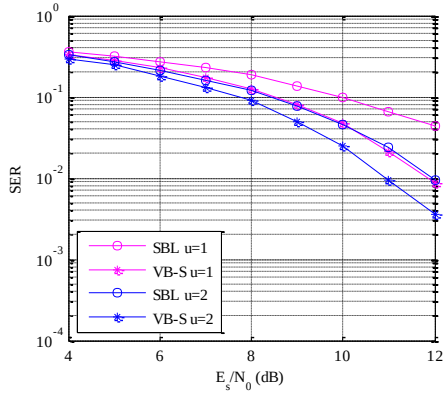


Fig. 4 : SER performance comparisons between SBL and VBS algorithms.

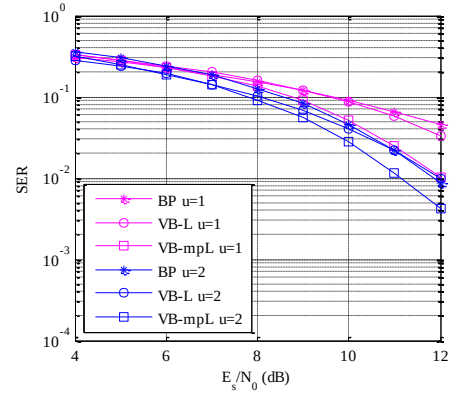


Fig. 5: SER performance comparisons between SBL and VBS algorithms.

6. Conclusions

In the proposed VB-based receiver, both fractionally- and symbol-spaced architectures are included. Using the variational Bayesian inference scheme, an approximately joint posterior for all model parameters is derived, which operates in an iterative scheme. Three special and widely used sparsity-promoting channel priors, i.e., Student's t, Laplace, and multi-parameter Laplace, are employed as examples. The proposed algorithms make good use of prior information of the channel. The superiority of the proposed method is verified by experiments. Note that the channel model considered is a sparse, time-invariant, Rayleigh-fading multipath channel with independent scattering physical paths. In practice, the channels may be time-varying, and/or exhibit correlation property among different paths. This will be considered in our follow-up studies.

7. Acknowledgements

The authors would like to thank their colleagues for useful discussions and suggestions, which helped to substantially improve the quality of this manuscript. This work was supported by a grant from the National Natural Science Foundation of China (No.61501517).

8. References

- [1] Zamiri-Jafarian H, Pasupathy S. Adaptive MLSDE using the EM algorithm[J]. IEEE Transactions on Communications, 1999, 47(8): 1181-1193.
- [2] Ma X, Kobayashi H, Schwartz S C. An EM-based estimation of OFDM signals[C]//Wireless Communications and Networking Conference, 2002. WCNC2002. 2002 IEEE. IEEE, 2002, 1: 228-232.
- [3] Vuokko L, Kolmonen V M, Salo J, et al. Measurement of large-scale cluster power characteristics for geometric channel models[J]. IEEE Transactions on Antennas and Propagation, 2007, 55(11): 3361-3365.
- [4] Mileounis G, Kalouptsidis N, Babadi B, et al. Blind identification of sparse channels and symbol detection via the EM algorithm[C]//2011 17th International Conference on Digital Signal Processing (DSP). 2011.
- [5] Prasad R, Murthy C R. Bayesian learning for joint sparse OFDM channel estimation and data detection[C]//Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE. IEEE, 2010: 1-6.
- [6] Chugg K M, Polydoros A. MLSE for an unknown channel. I. Optimality considerations[J]. IEEE Transactions on Communications, 1996, 44(7): 836-846.
- [7] Tepedelenlioglu C, Challagulla R. Low-complexity multipath diversity through fractional sampling in OFDM[J]. IEEE Transactions on Signal Processing, 2004, 52(11): 3104-3116.
- [8] Themelis K E, Rontogiannis A A, Koutroumbas K D. A variational Bayes framework for sparse adaptive estimation[J]. IEEE Transactions on Signal Processing, 2014, 62(18): 4723-4736.
- [9] Bello P. Characterization of randomly time-variant linear channels[J]. IEEE transactions on Communications Systems, 1963, 11(4): 360-393.
- [10] Park T, Casella G. The bayesian lasso[J]. Journal of the American Statistical Association, 2008, 103(482): 681-686.
- [11] Wipf D P, Nagarajan S S. A new view of automatic relevance determination[C]//Advances in neural information processing systems. 2008: 1625-1632.
- [12] Wipf D P. Sparse estimation with structured dictionaries[C]//Advances in Neural Information Processing Systems. 2011: 2016-2024.
- [13] Jakobsen M L, Laugesen K, Manchón C N, et al. Parametric modeling and pilot-aided estimation of the wireless multipath channel in OFDM systems[C]//Communications (ICC), 2010 IEEE International Conference on. IEEE, 2010: 1-6.
- [14] Stamoulis A, Giannakis G B, Scaglione A. Block FIR decision-feedback equalizers for filterbank precoded transmissions with blind channel estimation capabilities[J]. IEEE Transactions on Communications, 2001, 49(1): 69-83.