This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2010 proceedings.

Partial Data-Dependent Superimposed Training Based Iterative Channel Estimation for OFDM Systems over Doubly Selective Channels

Lanlan He, Shaodan Ma, Yik-Chung Wu, Tung-Sang Ng
Department of Electrical and Electronic Engineering, The University of Hong Kong
Email:{llhe, sdma, ycwu, tsng}@eee.hku.hk

***Abstract*—In this paper, partial data-dependent superimposed training based channel estimation for OFDM systems over doubly selective channels (DSCs) is addressed. Due to the presence of unknown data as interference, we first derive a minimum mean square error (MMSE) channel estimator by treating the effect of unknown data as noise. To further improve the performance, a novel iterative algorithm which jointly estimates channel and suppresses interference from data is proposed via variational inference approach. Simulation results show that the proposed algorithm converges after a few iterations. Furthermore, after convergence, the performance of the proposed channel estimator is very close to that with full training at high SNRs.**

***Index Terms*—Superimposed training, Channel estimation, Orthogonal frequency division multiplexing (OFDM).**

I. Introduction

For broadband orthogonal frequency division multiplexing (OFDM) systems, high speed movement of mobile terminals causes Doppler spread and results in multi-path time-varying channels, i.e., doubly selective channels (DSCs). Due to time variation of the channel, the number of channel parameters in one OFDM symbol significantly increases, which makes the channel estimation a challenge.

So far, semi-blind channel estimation for OFDM systems over DSCs has been discussed in [1]–[3]. However, in these works, data and pilots occupy exclusive sets of subcarriers, which decreases the spectrum efficiency. An alternative is the superimposed training (ST) [4]–[7], where pilots are added on top of the data symbols. ST based channel estimators are proposed for slow fading channels in [4], [5], which are not applicable for DSCs. Over DSCs, least square (LS) and linear minimum mean square error (LMMSE) channel estimators are proposed in [6] by splitting the whole OFDM symbol into subblocks and ignoring the time-variation of the channel within each subblock. However, the assumption that the channel within each subblock is constant would result in significant modeling error when dealing with large Doppler spread. In [7], partial data-dependent superimposed training (PDDST) based channel estimator is proposed for single carrier transmission systems over DSCs. As a general form of superimposed training generated in frequency domain, PDDST offers a tradeoff between data interference and data integrity by controlling a parameter called self-interference factor. However, extension of the works in [7] to OFDM systems is by no means straightforward.

In this paper, PDDST based channel estimation for OFDM systems over DSCs will be addressed via variational inference approach. The variational inference approach is applicable in cases when direct access or maximization of the posterior distribution of parameter to be estimated is difficult if not impossible. In particular, the variational inference approach constructs a lower bound on the posterior distribution [8], and attempts to optimize this bound iteratively. Since it is basically a Bayesian framework, statistical information (such as channel statistics, power of data and noise) is exploited to aid the estimation. Simulation results show that the proposed joint algorithm converges after a few iterations. Furthermore, after convergence, the performance of the proposed channel estimator is very close to that with full training.

Notation: Boldface letters will be used for matrices and vectors. H and T denotes Hermitian and transpose respectively. The symbol $\mathbf{I}_N$ denotes the $N \times N$ identity matrix, with $\mathbf{e}_l$ denoting the l^{th} column of $\mathbf{I}_N$. diag$\{\mathbf{x}\}$ stands for the diagonal matrix with vector $\mathbf{x}$ on its diagonal. The $(m,n)^{th}$ entry of a matrix $\mathbf{X}$ is denoted by $[\mathbf{X}]_{m,n}$. The symbol $\otimes$ denotes the Kronecker product and $\odot$ denotes the Hadamard product. $\mathbb{E}\{\cdot\}$ denotes the expectation and $\Re\{\cdot\}$ is the real part. Tr$\{\mathbf{X}\}$ and $|\mathbf{X}|$ are the trace and the determinant of a square matrix $\mathbf{X}$ respectively. The matrix $\mathbf{F}$ is the FFT matrix with $[\mathbf{F}]_{m,n} = \frac{1}{\sqrt{N}} e^{-j2\pi mn/N}$.

II. System Model

In an OFDM system, the source data in frequency domain is modulated onto N parallel subcarriers to obtain the time domain signal. With partial data-dependent superimposed training (PDDST), the transmitted signals in time-domain is given by

$$\mathbf{s} = \mathbf{F}^H \mathbf{b} + \mathbf{F}_{\mathcal{K}}^H (\mathbf{c} - \lambda \mathbf{b}_{\mathcal{K}}) \quad (1)$$

where $\mathbf{b}$ is the transmitted data in frequency domain, $\mathbf{c}$ is the known training on subcarriers set denoted by $\mathcal{K}$ with N_c elements, $\mathbf{b}_{\mathcal{K}}$ corresponds to the data transmitted on subcarrier set $\mathcal{K}$, $\lambda \in [0,1]$ is the self-interference factor, and $\mathbf{F}_{\mathcal{K}}^H$ denotes the columns of $\mathbf{F}^H$ corresponding to those subcarriers on $\mathcal{K}$.

In (1), since $\mathbf{b}_{\mathcal{K}}$ is indeed parts of $\mathbf{b}$, we can group those

The work was supported in part by the GRF, Project No. HKU 7154/08E and the HKU Seed Funding Programme, Project No. 200811159094.

unknown data together to obtain

$$\mathbf{s} = \underbrace{\begin{bmatrix}\mathbf{F}_{\bar{\mathcal{K}}}^H & (1-\lambda)\mathbf{F}_{\mathcal{K}}^H\end{bmatrix}}_{\triangleq \tilde{\mathbf{F}}^H} \underbrace{\begin{bmatrix}\mathbf{b}_{\bar{\mathcal{K}}} \\ \mathbf{b}_{\mathcal{K}}\end{bmatrix}}_{\triangleq \tilde{\mathbf{b}}} + \underbrace{\mathbf{F}_{\mathcal{K}}^H \mathbf{c}}_{\triangleq \tilde{\mathbf{c}}} \quad (2)$$

where $\mathbf{F}_{\bar{\mathcal{K}}}$ collects the rows corresponding to the subcarrier set $\bar{\mathcal{K}}$ with $\mathcal{K} \cup \bar{\mathcal{K}} = \{0, \cdots, N-1\}$, and $\tilde{\mathbf{b}}$ comes from re-ordering of $\mathbf{b}$.

From (2), we can find that PDDST include the following three cases:

- When $0 < \lambda < 1$, the data component at each subcarrier $k \in \mathcal{K}$ is reduced to $(1-\lambda)b_k$. Then the interference to training symbol from data on these subcarriers is effectively suppressed while there is no loss of data rate.
- In the case $\lambda = 1$, PDDST reduces to DDST, and the data on subcarriers $\mathcal{K}$ is nulled. Therefore, there is no interference to training from data on these subcarriers. However, DDST results in loss of data rate [5].
- In the case $\lambda = 0$, PDDST reduces to traditional superimposed training [4], and $\mathbf{c}$ becomes the training in frequency domain known to the receiver while the data $\mathbf{b}$ keeps intact. However, the interference to training from data is not suppressed at all at the transmitter side.

With the time domain OFDM symbol $\mathbf{s}$ in (2), a cyclic prefix (CP) with length longer than the delay spread of the channel is inserted at the beginning to prevent intersymbol interference (ISI). The signal is then transmitted through a Rayleigh-distributed doubly selective channel which has L independent taps with the average power of the l^{th} tap denoted by σ_l^2. The auto-correlation of the l^{th} tap follows the classical Jakes' model [1] given by $\mathbb{E}\{h_l(mT_s)h_l(nT_s)\} = \sigma_l^2 J_0(2\pi f_D(n-m)T_s)$, where $J_0(\cdot)$ represents the zero-order Bessel function of the first kind, f_D represents the Doppler spread normalized by the subcarrier spacing, and T_s is the sample interval. At the receiver side, assuming perfect synchronization is achieved, after discarding the CP, the received signal can be written as

$$\mathbf{r} = \mathbf{D}[\mathbf{s}]\mathbf{h} + \mathbf{w} \quad (3)$$

where $\mathbf{r} = [r(0), r(1), \cdots, r(N-1)]^T$ is the received signal, $\mathbf{D}[\mathbf{s}] = [\text{diag}\{\mathbf{\Xi}_0\mathbf{s}\}, ..., \text{diag}\{\mathbf{\Xi}_{L-1}\mathbf{s}\}]$ with $\mathbf{\Xi}_l = [\mathbf{e}_{l+1}, \cdots, \mathbf{e}_N, \mathbf{e}_1, \cdots, \mathbf{e}_l]$, $\mathbf{w} = [w(0), w(1), \cdots, w(N-1)]^T$ is the additive white Gaussian noise vector with zero-mean and covariance $\sigma_w^2\mathbf{I}_N$, and the channel vector $\mathbf{h}$ is defined as

$$\mathbf{h} = [\mathbf{h}_0^T, \cdots, \mathbf{h}_{L-1}^T]^T \quad (4)$$

with $\mathbf{h}_l = [h_l(0), \cdots, h_l(N-1)]^T$ being the channel coefficient of the l^{th} tap during the whole OFDM symbol. For notation simplicity, the sample interval T_s is omitted in $\mathbf{h}_l$. Since the channel taps are independent of each other, the correlation matrix of $\mathbf{h}$ is given by

$$\mathbf{R}_h = \mathbf{R}_L \otimes \mathbf{J} \quad (5)$$

with $\mathbf{R}_L = \text{diag}\{\sigma_0^2, \cdots, \sigma_{L-1}^2\}$ and $[\mathbf{J}]_{m,n} = J_0(2\pi f_D(n-m)T_s)$. Therefore, the probability density function (pdf) of channel follows

$$p(\mathbf{h}) = \frac{1}{\pi^N|\mathbf{R}_h|}\exp\{-\mathbf{h}^H\mathbf{R}_h^{-1}\mathbf{h}\}. \quad (6)$$

Before we proceed, we note a property of the system model (3), which will be used in the derivation of algorithm later.

Lemma 1: If $\mathbf{a}$ is in the structure of (4), we have

$$\mathbf{D}[\mathbf{s}]\mathbf{a} = \mathbf{B}[\mathbf{a}]\mathbf{s} \quad (7)$$

where the matrix operator $\mathbf{B}$ is defined as

$$\mathbf{B}[\mathbf{a}] = \begin{bmatrix} a_0(0) & \mathbf{0} & a_{L-1}(0) & \cdots & a_1(0) \\ a_1(1) & a_0(1) & \mathbf{0} & a_{L-1}(1) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \\ & \mathbf{0} & a_{L-1}(N-1) & \cdots & a_0(N-1) \end{bmatrix}. \quad (8)$$

Proof: Proved from the straightforward computation. ∎

III. Difficulties in Channel Estimation

The problem we need to address is estimation of $\mathbf{h}$ based on (3). However, notice that, both $\mathbf{h}$ and $\tilde{\mathbf{b}}$ are unknowns. One direct way to proceed is to treat both $\mathbf{h}$ and $\tilde{\mathbf{b}}$ as deterministic unknowns. From (2) and (3), the likelihood function is

$$p(\mathbf{r}|\mathbf{h}, \tilde{\mathbf{b}}) = \frac{1}{(\pi\sigma_w^2)^N}\exp\{-\frac{1}{\sigma_w^2}\|\mathbf{r} - \mathbf{D}[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}} + \tilde{\mathbf{c}}]\mathbf{h}\|^2\}. \quad (9)$$

Notice that $\mathbf{h}$ contains NL unknown parameters, which is larger than the number of received data in $\mathbf{r}$. This makes direct maximum likelihood (ML) solution not available. Even though basis expansion models (BEMs) can be adopted to reduce the number of channel parameters, multi-dimensional search is still required to detect the data, which would result in complexity $O(\mathcal{M}^N)$ with $\mathcal{M}$ being the constellation size. To avoid this, another way is to consider $\tilde{\mathbf{b}}$ as a nuisance parameter and employ the marginal ML method, which maximizes the marginal pdf $p(\mathbf{r}|\mathbf{h}) = \int_{\tilde{\mathbf{b}}} p(\mathbf{r}|\mathbf{h}, \tilde{\mathbf{b}})p(\tilde{\mathbf{b}})d\tilde{\mathbf{b}}$. Unfortunately, the integration with respect to $\tilde{\mathbf{b}}$ is hard to perform because $\tilde{\mathbf{b}}$ is drawn from discrete constellation.

On the other hand, if the statistics of channel and data are known, we can develop the MMSE channel estimator by treating data as noise. With *Lemma 1* and $\mathbf{s} = \tilde{\mathbf{F}}^H\tilde{\mathbf{b}} + \tilde{\mathbf{c}}$, we can rewritten (3) as

$$\mathbf{r} = \mathbf{D}[\tilde{\mathbf{c}}]\mathbf{h} + \mathbf{B}[\mathbf{h}]\tilde{\mathbf{F}}^H\tilde{\mathbf{b}} + \mathbf{w}. \quad (10)$$

Since $\tilde{\mathbf{b}}$ is unknown, the term with $\tilde{\mathbf{b}}$ will be treated as noise. Now we first compute the correlation matrix of $\mathbf{h}$ and $\mathbf{r}$ as

$$\mathbf{R}_{hr^H} = \mathbb{E}_{\mathbf{h},\tilde{\mathbf{b}},\mathbf{w}}\{\mathbf{h}\mathbf{r}^H\} = \mathbf{R}_h\mathbf{D}^H[\tilde{\mathbf{c}}]. \quad (11)$$

Moreover, through tedious but straightforward computations, the auto-correlation matrix of $\mathbf{r}$ is

$$\begin{aligned}\mathbf{R}_{rr^H} &= \mathbf{D}[\tilde{\mathbf{c}}]\mathbf{R}_h\mathbf{D}^H[\tilde{\mathbf{c}}] + \sigma_w^2\mathbf{I}_N \\ &+ \underbrace{\left(\mathbf{J} \odot (\mathbf{F}_{\bar{\mathcal{K}}}^H\mathbf{\Lambda}_{\bar{\mathcal{K}}}\mathbf{F}_{\bar{\mathcal{K}}}) + (1-\lambda)^2(\mathbf{J} \odot (\mathbf{F}_{\mathcal{K}}^H\mathbf{\Lambda}_{\mathcal{K}}\mathbf{F}_{\mathcal{K}}))\right)}_{\triangleq \ \mathbf{\Pi}}\end{aligned} \quad (12)$$

where $\mathbf{\Lambda}_{\bar{\mathcal{K}}} = \mathbb{E}\{\mathbf{b}_{\bar{\mathcal{K}}}\mathbf{b}_{\bar{\mathcal{K}}}^H\}$ and $\mathbf{\Lambda}_{\mathcal{K}} = \mathbb{E}\{\mathbf{b}_{\mathcal{K}}\mathbf{b}_{\mathcal{K}}^H\}$ are diagonal matrices. From (11) and (12), and based on MMSE criterion [6], the channel estimator is then given by

$$\begin{aligned}\hat{\mathbf{h}} &= \mathbf{R}_{hr^H}\mathbf{R}_{rr^H}^{-1}\mathbf{r} \\ &= \mathbf{R}_h\mathbf{D}^H[\tilde{\mathbf{c}}](\mathbf{D}[\tilde{\mathbf{c}}]\mathbf{R}_h\mathbf{D}^H[\tilde{\mathbf{c}}] + \sigma_w^2\mathbf{I}_N + \mathbf{\Pi})^{-1}\mathbf{r}.\end{aligned} \quad (13)$$

Since channel statistics is incorporated and the time-correlation among channel responses is taken into account, the number of effective channel parameters required to be estimated is largely reduced, and a closed-form channel estimator exists. However, notice that, over DSCs, one data subcarrier induces ICI on all the other subcarriers. Even if unknown data on subcarrier set $\mathcal{K}$ are completely nulled with $\lambda = 1$, other data on subcarrier set $\bar{\mathcal{K}}$ still introduce severe interference to the training pilots. If these interference is not mitigated, it will severely degrade the channel estimation performance. More details will be provided in Section V.

IV. The Variational Inference Approach to Iterative Channel Estimation

In the following, instead of treating data as noise, data is also exploited for channel estimation. Moreover, different from the ML approaches discussed above, we jointly estimate the channel and detect the data by exploiting statistics of channel, data and noise in the Bayesian framework. Specifically, our aim is to estimate $\mathbf{h}$ and $\tilde{\mathbf{b}}$, which maximizes of the posterior pdf $p(\mathbf{h}, \tilde{\mathbf{b}}|\mathbf{r})$. In general, direct computation of $p(\mathbf{h}, \tilde{\mathbf{b}}|\mathbf{r})$ is complicated and not convenient for maximization. To overcome this problem, we consider the variational inference approach. It looks for a parameterized distribution, $Q(\mathbf{h}, \tilde{\mathbf{b}})$, which closely represents the posterior pdf $p(\mathbf{h}, \tilde{\mathbf{b}}|\mathbf{r})$. Once $Q(\mathbf{h}, \tilde{\mathbf{b}})$ has been found, estimates of $\mathbf{h}, \tilde{\mathbf{b}}$ are simply given by maximizing $Q(\mathbf{h}, \tilde{\mathbf{b}})$ with respect to $\mathbf{h}, \tilde{\mathbf{b}}$.

To obtain $Q(\mathbf{h}, \tilde{\mathbf{b}})$ closest to $p(\mathbf{h}, \tilde{\mathbf{b}}|\mathbf{r})$, we minimize the following *free energy* function defined as [8]:

$$\mathbb{F} = \int_{\mathbf{h},\tilde{\mathbf{b}}} Q(\mathbf{h}, \tilde{\mathbf{b}}) \log \frac{Q(\mathbf{h}, \tilde{\mathbf{b}})}{p(\mathbf{h}, \tilde{\mathbf{b}}, \mathbf{r})} \mathrm{d}\mathbf{h}\mathrm{d}\tilde{\mathbf{b}}. \tag{14}$$

A simplification can be made by factorizing $Q(\mathbf{h}, \tilde{\mathbf{b}})$ into a product form (also known as mean-field approximation) [9], i.e., $Q(\mathbf{h}, \tilde{\mathbf{b}}) = Q(\mathbf{h})Q(\tilde{\mathbf{b}})$, which is equivalent to assuming that $\mathbf{h}$ and $\tilde{\mathbf{b}}$ are independent conditioned on $\mathbf{r}$. Then a simple expression of the variational free energy in (14) is given by

$$\begin{aligned}\mathbb{F} &= \int_{\mathbf{h},\tilde{\mathbf{b}}} Q(\mathbf{h}, \tilde{\mathbf{b}}) \log \frac{Q(\mathbf{h}, \tilde{\mathbf{b}})}{p(\mathbf{r}|\mathbf{h}, \tilde{\mathbf{b}})p(\mathbf{h})p(\tilde{\mathbf{b}})} \mathrm{d}\mathbf{h}\mathrm{d}\tilde{\mathbf{b}} \\ &= \int_{\mathbf{h}} Q(\mathbf{h}) \log Q(\mathbf{h})\mathrm{d}\mathbf{h} + \int_{\tilde{\mathbf{b}}} Q(\tilde{\mathbf{b}}) \log Q(\tilde{\mathbf{b}})\mathrm{d}\tilde{\mathbf{b}} \\ &- \int_{\mathbf{h}} Q(\mathbf{h}) \log p(\mathbf{h})\mathrm{d}\mathbf{h} - \int_{\tilde{\mathbf{b}}} Q(\tilde{\mathbf{b}}) \log p(\tilde{\mathbf{b}})\mathrm{d}\tilde{\mathbf{b}} \\ &- \int_{\mathbf{h},\tilde{\mathbf{b}}} Q(\mathbf{h})Q(\tilde{\mathbf{b}}) \log p(\mathbf{r}|\mathbf{h}, \tilde{\mathbf{b}})\mathrm{d}\mathbf{h}\mathrm{d}\tilde{\mathbf{b}}.\end{aligned} \tag{15}$$

For convenience in maximization [10], we assume

$$Q(\tilde{\mathbf{b}}) = \delta(\tilde{\mathbf{b}}) \tag{16}$$

and

$$Q(\mathbf{h}) = \frac{1}{\pi^{NL}|\mathbf{\Phi}_h|} \exp\{-(\mathbf{h} - \mathbf{m}_h)^H \mathbf{\Phi}_h^{-1}(\mathbf{h} - \mathbf{m}_h)\}. \tag{17}$$

Here $\delta(\tilde{\mathbf{b}})$ denotes a vector Diracs delta function with the properties $\int \delta(\tilde{\mathbf{b}})\mathrm{d}\tilde{\mathbf{b}} = 1$ and $\int \delta(\tilde{\mathbf{b}})\mathrm{f}(\tilde{\mathbf{b}})\mathrm{d}\tilde{\mathbf{b}} = \mathrm{f}(\tilde{\mathbf{b}})$ for any function $\mathrm{f}(\cdot)$. On the other hand, $\mathbf{m}_h$ and $\mathbf{\Phi}_h$ are parameters to be designed later.

With (16) and (17), the first term in (15) can be computed as

$$\begin{aligned}&\int_h Q(\mathbf{h}) \log Q(\mathbf{h})\mathrm{d}\mathbf{h} = -NL\log(\pi) - \log|\mathbf{\Phi}_h| - \mathbf{m}_h^H\mathbf{\Phi}_h^{-1}\mathbf{m}_h \\ &+ 2\Re\{\mathbf{m}_h^H\mathbf{\Phi}_h^{-1}\mathbf{m}_h\} - \mathrm{Tr}\{\mathbf{\Phi}_h^{-1}(\mathbf{m}_h\mathbf{m}_h^H + \mathbf{\Phi}_h)\} \\ &= -NL\log(\pi) - \log|\mathbf{\Phi}_h| - NL.\end{aligned} \tag{18}$$

Similarly, other four terms in (15) are

$$\int_{\tilde{\mathbf{b}}} Q(\tilde{\mathbf{b}}) \log Q(\tilde{\mathbf{b}})\mathrm{d}\tilde{\mathbf{b}} = 0, \tag{19}$$

$$\begin{aligned}\int_h Q(\mathbf{h}) \log p(\mathbf{h})\mathrm{d}\mathbf{h} = -NL\log(\pi) - \log|\mathbf{R}_h| \\ - \mathrm{Tr}\{\mathbf{R}_h^{-1}(\mathbf{m}_h\mathbf{m}_h^H + \mathbf{\Phi}_h)\},\end{aligned} \tag{20}$$

$$\int_{\tilde{\mathbf{b}}} Q(\tilde{\mathbf{b}}) \log p(\tilde{\mathbf{b}})\mathrm{d}\tilde{\mathbf{b}} = \log p(\tilde{\mathbf{b}}), \tag{21}$$

and

$$\begin{aligned}&\int_{\mathbf{h},\tilde{\mathbf{b}}} Q(\mathbf{h})Q(\tilde{\mathbf{b}}) \log p(\mathbf{r}|\mathbf{h}, \tilde{\mathbf{b}})\mathrm{d}\mathbf{h}\mathrm{d}\tilde{\mathbf{b}} \\ &= -N\log(\pi\sigma_w^2) - \frac{1}{\sigma_w^2}(\mathbf{r}^H\mathbf{r} - 2\Re\{\mathbf{r}^H\mathbf{D}[\tilde{\mathbf{F}}^{\mathbf{H}}\tilde{\mathbf{b}} + \tilde{\mathbf{c}}]\mathbf{m}_h\} \\ &+ \mathrm{Tr}\{\mathbf{D}^H[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}} + \tilde{\mathbf{c}}]\mathbf{D}[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}} + \tilde{\mathbf{c}}](\mathbf{m}_h\mathbf{m}_h^H + \mathbf{\Phi}_h)\}).\end{aligned} \tag{22}$$

Putting (18), (19), (20), (21) and (22) into (15), we have

$$\begin{aligned}&\mathbb{F}(\mathbf{m}_h, \mathbf{\Phi}_h, \tilde{\mathbf{b}}) = -\log|\mathbf{\Phi}_h| + \log|\mathbf{R}_h| \\ &+ \mathrm{Tr}\{\mathbf{R}_h^{-1}(\mathbf{m}_h\mathbf{m}_h^H + \mathbf{\Phi}_h)\} - \log p(\tilde{\mathbf{b}}) \\ &+ \frac{1}{\sigma_w^2}\Big(\mathbf{r}^H\mathbf{r} - 2\Re\{\mathbf{r}^H\mathbf{D}[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}} + \tilde{\mathbf{c}}]\mathbf{m}_h\} \\ &+ \mathrm{Tr}\{\mathbf{D}^H[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}} + \tilde{\mathbf{c}}]\mathbf{D}[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}} + \tilde{\mathbf{c}}](\mathbf{m}_h\mathbf{m}_h^H + \mathbf{\Phi}_h)\}\Big).\end{aligned} \tag{23}$$

The free energy function in (23) depends on $\mathbf{m}_h$, $\mathbf{\Phi}_h$ and $\tilde{\mathbf{b}}$. The remaining task is to obtain $(\hat{\mathbf{m}}_h, \hat{\mathbf{\Phi}}_h, \hat{\tilde{\mathbf{b}}})$ by minimizing $\mathbb{F}(\mathbf{m}_h, \mathbf{\Phi}_h, \tilde{\mathbf{b}})$. After that, channel estimate can be acquired by maximizing $Q(\mathbf{h})$ given $\hat{\mathbf{m}}_h$ and $\hat{\mathbf{\Phi}}_h$. Notice that, since $Q(\mathbf{h})$ is designed to be complex Gaussian distribution, therefore, it is maximized at $\mathbf{h} = \hat{\mathbf{m}}_h$, which can be considered as a MAP channel estimator. Moreover, as a byproduct, $\hat{\tilde{\mathbf{b}}}$ is an estimate of data.

For minimization of the free energy given in (23) with respect to $(\mathbf{m}_h, \mathbf{\Phi}_h, \tilde{\mathbf{b}})$, it is found that, given $\tilde{\mathbf{b}}$, there exist closed-form solutions for $\mathbf{m}_h$ and $\mathbf{\Phi}_h$. On the other hand, given $\mathbf{m}_h$ and $\mathbf{\Phi}_h$, we can derive closed-form solution for $\tilde{\mathbf{b}}$. Therefore, $\mathbb{F}(\mathbf{m}_h, \mathbf{\Phi}_h, \tilde{\mathbf{b}})$ is minimized iteratively, starting with an initial value.

A. Updating $\mathbf{m}_h$ and $\mathbf{\Phi}_h$ given $\tilde{\mathbf{b}}$

By setting the first order derivative of the free energy with respect to $\mathbf{m}_h$ to zero, we have the estimate of $\mathbf{m}_h$ as

$$\begin{aligned}\hat{\mathbf{m}}_h = &\big(\mathbf{D}^H[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}} + \tilde{\mathbf{c}}]\mathbf{D}[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}} + \tilde{\mathbf{c}}] + \sigma_w^2\mathbf{R}_h^{-1}\big)^{-1} \\ &\times \mathbf{D}^H[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}} + \tilde{\mathbf{c}}]\mathbf{r}.\end{aligned} \tag{24}$$

Similarly, by setting the first order derivative of the free energy with respect to $\mathbf{\Phi}_h$ to zero, we have the estimate of $\mathbf{\Phi}_h$ as

$$\hat{\mathbf{\Phi}}_h = \sigma_w^2\big(\mathbf{D}^H[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}}]\mathbf{D}[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}}]+\sigma_w^2\mathbf{R}_h^{-1}\big)^{-1}. \quad (25)$$

B. Updating $\tilde{\mathbf{b}}$ *given* $\mathbf{m}_h$ *and* $\mathbf{\Phi}_h$

Notice that the free energy given by (23) depends on $\tilde{\mathbf{b}}$ in a non-linear way. To obtain a close-form solution for $\tilde{\mathbf{b}}$, we first transform (23) into the linear form of $\tilde{\mathbf{b}}$. Given the eigen-decomposition of $\mathbf{\Phi}_h = \sum_{m=1}^{r_{\Phi_h}}\alpha_m\mathbf{u}_m\mathbf{u}_m^H$ with r_{Φ_h} being the rank of $\mathbf{\Phi}_h$ [1], we have

$$\begin{aligned}&\mathrm{Tr}\{\mathbf{D}^H[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}}]\mathbf{D}[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}}]\mathbf{\Phi}_h\}\\&=\sum_{m=1}^{r_{\Phi_h}}\alpha_m\mathbf{u}_m^H\mathbf{D}^H[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}}]\mathbf{D}[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}}]\mathbf{u}_m.\end{aligned} \quad (26)$$

Putting (26) into (23), and dropping those terms irrelevant to $\tilde{\mathbf{b}}$, it follows that

$$\begin{aligned}\mathbb{F}(\mathbf{m}_h,\mathbf{\Phi}_h,\tilde{\mathbf{b}}) &= -\log p(\tilde{\mathbf{b}}) + \frac{1}{\sigma_w^2}(-2\Re\{\mathbf{r}^H\mathbf{D}[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}}]\mathbf{m}_h\}\\&+\mathbf{m}_h^H\mathbf{D}^H[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}}]\mathbf{D}[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}}]\mathbf{m}_h\\&+\sum_{m=1}^{r_{\Phi_h}}\alpha_m\mathbf{u}_m^H\mathbf{D}^H[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}}]\mathbf{D}[\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}}]\mathbf{u}_m).\end{aligned} \quad (27)$$

By *Lemma 1*, (27) can be written as

$$\begin{aligned}\mathbb{F}(\mathbf{m}_h,\mathbf{\Phi}_h,\tilde{\mathbf{b}}) &= -\log p(\tilde{\mathbf{b}}) - \frac{1}{\sigma_w^2}\Big(2\Re\{\mathbf{r}^H\mathbf{B}[\mathbf{m}_h](\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}})\}\\&-(\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}})^H\underbrace{\big(\mathbf{B}^H[\mathbf{m}_h]\mathbf{B}[\mathbf{m}_h]+\sum_{m=1}^{r_{\Phi_h}}\alpha_m\mathbf{B}^H[\mathbf{u}_m]\mathbf{B}[\mathbf{u}_m]\big)}_{\triangleq\mathbf{M}}\\&\times(\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}+\tilde{\mathbf{c}})\Big).\end{aligned} \quad (28)$$

Again dropping the terms independent of $\tilde{\mathbf{b}}$, we have

$$\begin{aligned}&\mathbb{F}(\mathbf{m}_h,\mathbf{\Phi}_h,\tilde{\mathbf{b}}) = -\tilde{\mathbf{b}}\mathbf{\Lambda}_{\tilde{\mathbf{b}}}^{-1}\tilde{\mathbf{b}}^H\\&-\frac{1}{\sigma_w^2}\big(2\Re\{\mathbf{r}^H\mathbf{B}[\mathbf{m}_h]\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}-\tilde{\mathbf{c}}^H\mathbf{M}\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}\}-\tilde{\mathbf{b}}^H\tilde{\mathbf{F}}\mathbf{M}\tilde{\mathbf{F}}^H\tilde{\mathbf{b}}\big)\end{aligned} \quad (29)$$

where we let $p(\tilde{\mathbf{b}})$ be a complex Gaussian with zero mean and covariance matrix $\mathbf{\Lambda}_{\tilde{\mathbf{b}}}$, a diagonal matrix whose elements depend on the average power of $\tilde{\mathbf{b}}$ [10]. Note that instead of defining a discrete distribution over the signal constellation, we have made a Gaussian approximation, which leads to a linear detector.

By setting the first order derivative of the free energy in (29) with respect to $\tilde{\mathbf{b}}$ to zero, we have the estimate of $\check{\mathbf{b}}$ as

$$\check{\mathbf{b}} = (\tilde{\mathbf{F}}\mathbf{M}\tilde{\mathbf{F}}^H+\sigma_w^2\mathbf{\Lambda}_{\tilde{\mathbf{b}}}^{-1})^{-1}(\tilde{\mathbf{F}}\mathbf{B}^H[\mathbf{m}_h]\mathbf{r}-\tilde{\mathbf{F}}\mathbf{M}\tilde{\mathbf{c}}). \quad (30)$$

Accordingly, $\hat{\tilde{\mathbf{b}}} = \mathrm{Qant}[\check{\mathbf{b}}]$ is a data estimator, where $\mathrm{Qant}[\check{\mathbf{b}}]$ denotes making hard decision on $\check{\mathbf{b}}$.

Since initialization is very essential to the performance of the proposed iterative algorithm, the MMSE channel estimator given in (13) and its corresponding theoretical MSE (whose expression can be readily obtained, e.g., according to [3]) are chosen to be $\hat{\mathbf{m}}_h^0$ and $\hat{\mathbf{\Phi}}_h^0$ respectively. In summary, the proposed iterative algorithm is:

Initialization:

Choose $\hat{\mathbf{m}}_h^0 = \hat{\mathbf{h}}$ given in (13) and $\hat{\mathbf{\Phi}}_h^0 = \mathbf{R}_h^{-1}+\mathbf{D}^H[\tilde{\mathbf{c}}](\mathbf{\Pi}+\sigma_w^2\mathbf{I}_N)^{-1}\mathbf{D}[\tilde{\mathbf{c}}]$.

Iteration (for $i = 1, 2, 3\cdots$):

$$\hat{\tilde{\mathbf{b}}}^i = \mathrm{Qant}\Big[(\tilde{\mathbf{F}}\mathbf{M}^i\tilde{\mathbf{F}}^H+\sigma_w^2\mathbf{\Lambda}_{\tilde{\mathbf{b}}}^{-1})^{-1}(\tilde{\mathbf{F}}\mathbf{B}^H[\hat{\mathbf{m}}_h^{i-1}]\mathbf{r}-\tilde{\mathbf{F}}\mathbf{M}\tilde{\mathbf{c}})\Big] \quad (31)$$

where

$$\mathbf{M}^i = \mathbf{B}^H[\hat{\mathbf{m}}_h^{i-1}]\mathbf{B}[\hat{\mathbf{m}}_h^{i-1}]+\sum_{m=1}^{r_{\Phi_h^{i-1}}}\alpha_m^{i-1}\mathbf{B}^H[\mathbf{u}_m^{i-1}]\mathbf{B}[\mathbf{u}_m^{i-1}] \quad (32)$$

with α_m^{i-1} being the m^{th} eigenvalue of $\hat{\mathbf{\Phi}}_h^{i-1}$ and $\mathbf{u}_m^{i-1}$ being the corresponding eigenvector;

$$\begin{aligned}\hat{\mathbf{m}}_h^i &= (\mathbf{D}^H[\tilde{\mathbf{F}}^H\hat{\tilde{\mathbf{b}}}^i+\tilde{\mathbf{c}}]\mathbf{D}[\tilde{\mathbf{F}}^H\hat{\tilde{\mathbf{b}}}^i+\tilde{\mathbf{c}}]+\sigma_w^2\mathbf{R}_h^{-1})^{-1}\\&\times\ \mathbf{D}^H[\tilde{\mathbf{F}}^H\hat{\tilde{\mathbf{b}}}^i+\tilde{\mathbf{c}}]\mathbf{r}; \quad (33)\\\hat{\mathbf{\Phi}}_h^i &= \sigma_w^2(\mathbf{D}^H[\tilde{\mathbf{F}}^H\hat{\tilde{\mathbf{b}}}^i+\tilde{\mathbf{c}}]\mathbf{D}[\tilde{\mathbf{F}}^H\hat{\tilde{\mathbf{b}}}^i+\tilde{\mathbf{c}}]+\sigma_w^2\mathbf{R}_h^{-1})^{-1} \quad (34)\end{aligned}$$

Until $\|\hat{\mathbf{m}}_h^i-\hat{\mathbf{m}}_h^{i-1}\|^2$ is smaller than a threshold ϵ. Assuming the algorithm converges at the $\mathcal{C}^{th}$ iteration, we finally obtain $\hat{\mathbf{m}}_h^{\mathcal{C}}$ as the channel estimate.

V. Simulation Results

In this section, the performances of both the MMSE channel estimator given in (13) and the proposed iterative channel estimator are demonstrated by Monte Carlo simulations, where each point is obtained by averaging over $\mathcal{R} = 10000$ runs. Each OFDM symbol has 128 subcarriers (N=128) and the length of CP is 8. Carrier frequency is $f_c = 2$ GHz, the sample interval $T_s = 2\mu s$ and the speed of vehicle is $v = 219$ km/hr, which results the normalized Doppler spread $NT_sf_D = NT_s\frac{vf_c}{c} = 0.1$ with c being the speed of light. The channel has five taps ($L = 5$) with an exponential power delay profile, namely $\sigma_l^2 = \exp(-\beta l)(\frac{1-\exp(-\beta)}{1-\exp(-\beta L)}), l = 0,\cdots,L-1$ with $\beta = 0.2$. Each tap is Rayleigh distributed and is assumed to experience the same f_D, and the time correlation of each tap follows the Jakes' model. The pilots in frequency domain is equally spaced with $N_c = 16$, and each pilot has equal power α. The training-to-signal power ratio in time domain is fixed to $\frac{\alpha N_c}{N+\big((1-\lambda)^2+\alpha-1\big)N_c} = 0.124$ by controlling α once λ is fixed. The pilots are generated following complex Gaussian distribution and the data are chosen from QPSK constellation. We set thresholds $\epsilon = 10^{-4}$ to terminate the iterative algorithm. The simulated NMSE of channel estimation at the i^{th} iteration is defined as

$$NMSE_h = \frac{\sum_{n=0}^{\mathcal{R}-1}||\hat{\mathbf{m}}_h^i-\mathbf{h}||^2}{\sum_{n=0}^{\mathcal{R}-1}||\mathbf{h}||^2}. \quad (35)$$

First, we focus on the performance of the MMSE channel estimator given in (13), which is used for initialization. In Fig. 1, comparison of the MMSE channel estimator under different λ is illustrated. As can be seen, the channel estimation performance improves significantly as λ increases, since the interference from data to training reduces as λ increases. To offer a balance between limited interference for channel estimation and data integrity, $\lambda = 0.8$ is taken as an example in the following simulations.

Next, we look at the convergence behavior of the proposed algorithm, which is shown through the NMSEs of channel estimates versus the number of iterations in Fig. 2. It can be seen that the performance of channel estimation improves significantly in the first iteration. At SNR=10dB, the NMSEs converge to stable values quickly. At SNR=30dB, the channel estimation performance continually improve until about seven iterations.

Finally, the performance of the proposed iterative algorithm versus signal-to-noise ratios (SNRs) is illustrated. Fig. 3 depicts the NMSEs of channel estimates. The channel estimation with full training is also shown for comparison. As can be seen, after convergence, the NMSEs of channel estimation almost touches that of full training at high SNRs. Simulation results also show that, after convergence, the data detection performance of the proposed iterative algorithm is very close to the ideal case with perfect channel state information. The figure is not presented here due to space limitation.

VI. Conclusions

In this paper, PDDST based channel estimation for OFDM systems over DSCs was addressed via variational inference approach. The proposed channel estimation method works with limited superimposed pilot, and iteratively estimate channel and detect the unknown data. Simulation results show that the proposed algorithm converges after a few iterations. Furthermore, after convergence, the performance of the proposed channel estimator is very close to that with full training.

References

[1] G. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *IEEE Proceedings*, vol. 86, no. 10, pp. 1969-1986, Oct. 1998.

[2] Z. Tang, G. Leus, R. Cannizzaro and P. Banelli, "Low-complexity pilot-assisted time-varying channel estimation for OFDM systems," *IEEE Trans. Signal Processing*, vol. 55, no. 5, pp. 2226-2238, May 2007.

[3] A. Kannu and P. Schniter, "Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. Signal Processing*, vol. 56, pp. 1148-1160, Mar. 2008.

[4] A. Orozco-Lugo, M. Lara and D. McLernon, "Channel estimation using implicit training," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp.240-254, Jan. 2004.

[5] M. Ghogho, D. McLernon, E. Alameda-Hernandez and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training" *IEEE Trans. Signal Processing Letters*, vol.12, no.3, pp. 226-229, Mar. 2005.

[6] Q. Yang and K. Kwak, "Superimposed training for estimating of doubly-selective OFDM channels," in *Proc. IEEE Advance Commun. Tech.*, vol. 3, pp. 1652-1655, Feb. 2007.

[7] S. He and J. Tugnait, "On doubly selective channel estimation using superimposed training and discrete prolate spheroidal sequences" *IEEE Trans. Signal Processing*, vol.56, no.7, pp. 3214-3228, July 2008.

[8] B. Frey, *Graphical Models for Machine learing and digital communication,* Cambridge, MA: MIT Press, 1998.

[9] S. Haykin, J. Prı́ncipe, T. Sejnowski and J. McWhirter, *New Directions in Statistical Signal Processing: From Systems to Brains,* MIT Press, 2005.

[10] D. Lin and T. Lim, "The variatinal inference approach to joint data detection and phase noise estimation in OFDM " *IEEE Trans. Signal Processing*, vol.55, no.5, pp. 1862-1873, May 2007.

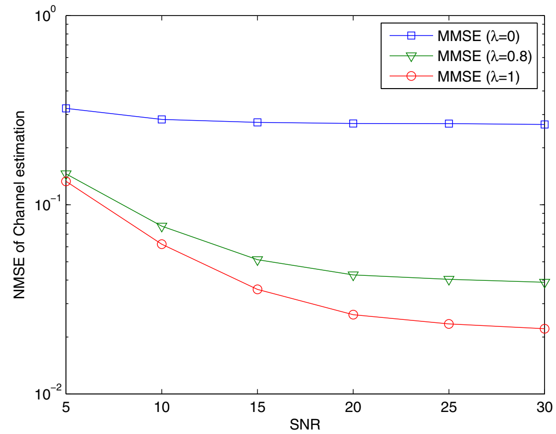

Fig. 1. Performance comparison of the MMSE channel estimator under different λ

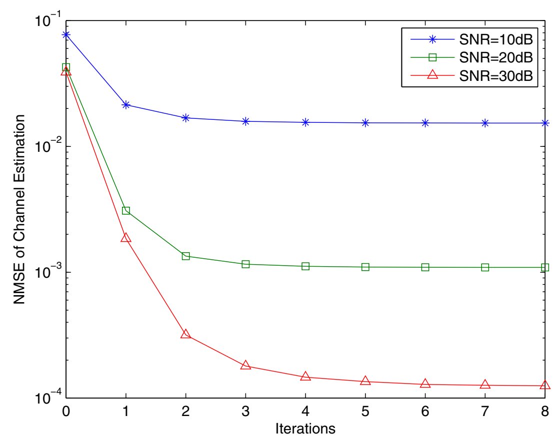

Fig. 2. Convergence of the proposed iterative algorithm

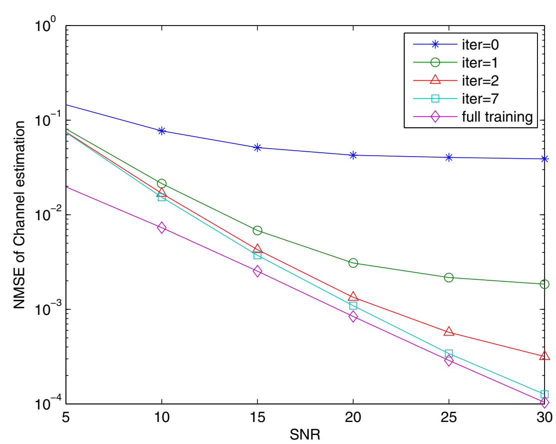

Fig. 3. NMSE of channel estimation for the proposed iterative algorithm