<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Equalizing multihop OFDM relay channel under unknown channel orders and doppler frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Zhou, J; Min, R; Qin, J; Wu, YC</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>The 2014 IEEE International Conference on Communications (IEEE ICC 2014), Sydney, Australia, 10-14 June 2014. In IEEE International Conference on Communications, 2014, p. 4512-4517</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2014</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/199395">http://hdl.handle.net/10722/199395</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>IEEE International Conference on Communications. Copyright © IEEE.; ©2014 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Equalizing Multihop OFDM Relay Channel under Unknown Channel Orders and Doppler Frequencies

Jingrong Zhou†‡, Rui Min†, Jiayin Qin†, and Yik-Chung Wu†
†Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong
‡Department of Electronics and Communications Engineering, Sun Yat-Sen University, Guangzhou, China
Email: {jrzhou, minrui, ycwu}@eee.hku.hk
Email: issqjy@mail.sysu.edu.cn

Abstract—In this paper, equalization of multihop relaying orthogonal frequency division multiplexing (OFDM) signal is investigated under time-varying channel with unknown noise powers, channel orders and Doppler frequencies. An iterative algorithm is developed under variational expectation maximization (EM) framework. The proposed algorithm iteratively estimates the channel, learns the channel and noise statistical information, and recovers the unknown data, using only limited number of pilot subcarrier in one OFDM symbol. Simulation results show that, without any statistical information, the performance of the proposed algorithm is very close to that of the optimal channel estimation and data detection algorithm, which requires specific information on system structure, channel tap positions, channel lengths, Doppler shifts as well as noise powers.

I. INTRODUCTION

Next generation broadband systems aim to support higher levels of mobility, connectivity and efficiency. Multihop relaying systems are a perfect suit for such requirement due to their benefits in easy deployment, enhanced connectivity, flexible adaptability, and increased capacity. On the other hand, orthogonal frequency division multiplexing (OFDM) has been adopted as the transmission scheme for many next generation broadband standards, such as WiMAX, LTE and IEEE 802.16. These result in the need to develop receiver algorithms for multihop OFDM system under high mobility. With high mobility, the broadband wireless channel is both frequency-selective and time-varying, a.k.a. doubly-selective. The channel responses vary sample by sample, which destroy the orthogonal property among subcarriers and causes intercarrier interference (ICI). Besides, the relaying system structure and channel statistical information are generally unknown to the receiver, due to flexible configuration of relaying paths. These pose strong challenges to channel estimation and data detection of OFDM relaying system under high mobility.

For point-to-point OFDM systems operating on doubly-selective channels, due to the ICI introduced in frequency domain, pilots and data would interfere with each other, degrading the accuracy of channel estimation via pilots only. One way to handle it is to iterate between data detection and channel estimation, e.g., through expectation-maximization (EM) algorithm [1], assuming knowledge of channel statistics and noise variance. When the noise variance is unknown, a solution based on variational inference was proposed in [2]. Recently, in [3], an iterative algorithm for data detection and channel estimation was proposed for dual-hop amplify-and-forward (AF) OFDM system. With complete information of channel and noise in each hop, the data detection results were very close to the ideal case. Unfortunately, all the above works require the destination receiver to have full statistical information of channels, which might not be readily available in practice.

Traditionally, unknown channel length can be handled by model order selection [4]. However, time synchronization of OFDM symbol is usually not perfect, leaving an unknown number of preceding zeros in the equivalent channel [5]. Together with the unknown Doppler width, the search space of this joint channel order/channel starting position/Doppler spread would be 3 dimensional, as opposed to the 1-D search in basic channel order selection. If the channel taps and Doppler shifts of the channel occur at non-consecutive positions, exhaustive search of all unknown parameters is impossible.

Different from previous works which highly rely on information of system structure, channel tap positions, channel lengths and Doppler frequencies of all channels, as well as noise powers at all receivers, we propose to solve the problems with none of the above information. The composite multihop channel is expanded using generalized complex exponential basis expansion model (GCE-BEM) [6], and Gaussian distributions with Gamma hyperpriors are adopted for the BEM coefficients to facilitate automatic model selection. An iterative algorithm is proposed based on variational EM framework to iteratively estimate the channel, learn the channel and noise statistical information, and recover the unknown data, using only limited number of pilot subcarrier in one OFDM symbol. Simulation results show that the performance of the proposed algorithm is very close to that of an optimal algorithm, which requires detailed statistical information on channels and noises.

Notations: Superscripts $H$ and $T$ denote Hermitian and transpose, respectively. The symbol $I_N$ represents the $N \times N$ identity matrix. Symbol $e_l$ denotes the vector with structure given as $[0_{1 \times l}, 1, 0_{1 \times (N-l-1)}]^T$, where $0_{1 \times l}$ is the $l$ dimension all-zero row vector. $\text{diag}(x)$ stands for the diagonal matrix with vector $x$ on its diagonal. The notation $[X]_{m_1:m_2,n_1:n_2}$ represents the submatrix of $X$ consists of entries on the $m_1$-to-$m_2^{th}$ rows and $n_1$-to-$n_2^{th}$ columns. $E\{\cdot\}$
denotes the expectation while \( \mathbb{T} \{ X \} \) is the trace of the square matrix \( X \). \( \text{Re}\{ \} \) denotes the real part. And \( \lfloor x \rfloor \) rounds \( x \) to the nearest integer greater than or equal to \( x \). Finally, \( \mathbb{F} \) represents the discrete Fourier transform (DFT) matrix with \( \mathbb{F}_{m,n} = \frac{1}{\sqrt{N}} e^{-j 2\pi m n / N} \).

## II. SYSTEM MODEL

In this paper, we consider a multihop relaying system that consists of a source \( S \), a destination \( D \) and a number of relays scattered in the middle. Each of them is equipped with single antenna. Without loss of generality, we assume the relays work cooperatively to form \( K \) links, each of them consisting of \( \Upsilon + 1 \) hops. Apart from the \( K \) relaying paths, there is no other link between \( S \) and \( D \), and all relays employ the AF scheme. Denoting the relay on the \( k \)th link connecting the \( \rho \)th and the \(( \rho + 1)\)th hop as \( R_{\rho,k} \).

The channel of each hop is assumed to be doubly-selective channel (DSC). Specifically, at the \( \rho \)th hop of the \( k \)th relaying path, the channel consists of \( N_{k,\rho} \) independent nonzero channel taps with maximum delay of \( (L_{k,\rho}^{\text{max}} - 1)T_s \), where \( T_s \) is the sample interval. We consider the general situation that the channel taps are not necessarily consecutive, so that we have \( N_{k,\rho} \leq L_{k,\rho}^{\text{max}} \). Let \( h_{k,\rho}(n,l) \) be the \( l \)th tap of that channel at time \( nT_s \). For a given \( \rho \) and \( k \), the channel taps are independent and each one being a zero-mean complex Gaussian process with bandlimited power spectral density within \([-f_{k,\rho}(l),f_{k,\rho}(l)]\), where \( f_{k,\rho}(l) \) is the maximum Doppler shift of the \( l \)th tap. In general, \( f_{k,\rho}(l) \) may be distinct for different \( l \), since each tap results from signal transmission through a different physical scattering. Furthermore, it is assumed that the channels for different links \( k \) and hops \( \rho \) are independent from each other.

### A. OFDM Signal Transmitted from \( S \)

In an OFDM system, the frequency domain source data \( x = [x(0),\ldots,x(N-1)]^T \) is first transformed to the time domain data \( s = \mathbb{F}^d x \). In order to facilitate channel estimation and data detection, pilots are inserted in the frequency domain as

\[
x(n) = \begin{cases} x_p(n) & \forall \; n \in \mathcal{J}_p \\
x_d(n) & \forall \; n \in \mathcal{J}_d,
\end{cases}
\]

(1)

where \( \mathcal{J}_d \) is the index set of the \( N_d \) unknown data symbols, \( \mathcal{J}_p \) is the index set of the \( N_p \) pilot symbols and we have \( N = N_d + N_p \). In matrix form, \( x \) can be represented as

\[
x = E_d x_d + E_p x_p,
\]

(2)

where \( E_d \) and \( E_p \), with dimensions \( N \times N_d \) and \( N \times N_p \), respectively, are matrices collecting columns of \( I_N \) that map \( x_d \) and \( x_p \) to subcarriers according to \( \mathcal{J}_p \) and \( \mathcal{J}_d \).

Before transmission, a cyclic prefix (CP) is added at the beginning of the OFDM symbol to prevent intersymbol interference (ISI). Since the OFDM signal goes through a number of relays before reaching the destination, the length of CP \( L_{\text{CP}} \) should be larger than the maximum channel length among all the relaying paths, denoted as \( L_{\text{max}} = \max_k (\sum_{\rho = 1}^{\Upsilon + 1} L_{k,\rho}^{\text{max}} - \Upsilon) \).

### B. Received OFDM Signal

In AF relaying system, each relay merely amplifies the received signal before passing the signal to the next relay or destination. For the \( k \)th relaying path, the signal received at \( R_{k,1} \) is given by

\[
r_{k,1}(n) = \sum_{l=0}^{L_{k,1}^{\text{max}} - 1} h_{k,1}(n,l)s(n-l) + w_{k,1}(n),
\]

(3)

where \( w_{k,1}(n) \) is the additive white Gaussian noise (AWGN) with power \( \sigma^2_{k,1} \). Upon reception, the relay amplifies the incoming signal as \( z_{k,1}(n) = \delta_{k,1} r_{k,1}(n) \) \([3]\), with \( \delta_{k,1} \) being the amplification factor at \( R_{k,1} \), and then transmits \( z_{k,1}(n) \) to the next relay \( R_{k,2} \) and so on. Then, at the \( \Upsilon \)th hop, the amplified signal is transmitted to the destination. Finally at destination \( D \), the received signal is given by

\[
y(n) = \sum_{k=1}^{K} \sum_{l=0}^{L_{\text{max}}^{\Upsilon+1} - 1} h_{k,\Upsilon+1}(n,l)z_{k,\Upsilon}(n-l) + w_d(n),
\]

(4)

where AWGN \( w_d(n) \) has power \( \sigma^2_d \). Upon reception, the CP is removed and the received signal \( \tilde{y} = [\hat{y}(0),\ldots,\hat{y}(N-1)]^T \) can be written in matrix form as

\[
\tilde{y} = \sum_{k=1}^{K} \mathbb{H}_{k,\Upsilon+1} z_{k,\Upsilon} + w_d,
\]

(5)

where \( z_{k,\Upsilon} = [z_{k,\Upsilon}(-1),\ldots,z_{k,\Upsilon}(0),\ldots,z_{k,\Upsilon}(N-1)]^T \), \( w_d \) is the noise vector with elements \( w_d(n) \), and \( \mathbb{H}_{k,\Upsilon+1} \) is an \( (N + L_{\text{max}}^{\Upsilon+1} - 1) \times N \) channel matrix given by

\[
\mathbb{H}_{k,\Upsilon+1} = \begin{bmatrix}
\tilde{h}_{k,\Upsilon+1}(0) & \tilde{h}_{k,\Upsilon+1}(1,0) & \cdots & \tilde{h}_{k,\Upsilon+1}(N-1,0) \\
\tilde{h}_{k,\Upsilon+1}(0,1) & \tilde{h}_{k,\Upsilon+1}(1,1) & \cdots & \tilde{h}_{k,\Upsilon+1}(N-1,1) \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{h}_{k,\Upsilon+1}(N-1,0) & \cdots & \tilde{h}_{k,\Upsilon+1}(N-1,L_{k,1}^{\text{max}} - 1) \\
\end{bmatrix}.
\]

(6)

Furthermore, \( z_{k,\Upsilon-1} \) can be written in terms of \( z_{k,\Upsilon-1} \) as

\[
z_{k,\Upsilon-1} = \bar{s}_{k,\Upsilon} \mathbb{H}_{k,\Upsilon} z_{k,\Upsilon-1} + \bar{s}_{k,\Upsilon} \mathbb{H}_{k,\Upsilon} w_{k,\Upsilon},
\]

(7)

with \( \bar{s}_{k,\Upsilon-1} = [z_{k,\Upsilon-1}(-1),\ldots,z_{k,\Upsilon-1}(0),\ldots,z_{k,\Upsilon-1}(N-1)]^T \), \( w_{k,\Upsilon} \) is the corresponding noise vector, and \( \mathbb{H}_{k,\Upsilon} \) is an \( (N + L_{\text{max}}^{\Upsilon} - 1) \times (N + L_{\text{max}}^{\Upsilon+1} + L_{k,1}^{\text{max}} - 2) \) matrix given by

\[
\mathbb{H}_{k,\Upsilon} = \begin{bmatrix}
\tilde{h}_{k,\Upsilon}(1-L_{k,1}^{\text{max}}) & \tilde{h}_{k,\Upsilon}(1-L_{k,1}^{\text{max}}) & \cdots & \tilde{h}_{k,\Upsilon}(1-L_{k,1}^{\text{max}},0) \\
\tilde{h}_{k,\Upsilon}(2-L_{k,1}^{\text{max}}) & \tilde{h}_{k,\Upsilon}(2-L_{k,1}^{\text{max}}) & \cdots & \tilde{h}_{k,\Upsilon}(2-L_{k,1}^{\text{max}},1) \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{h}_{k,\Upsilon}(N-1) & \cdots & \tilde{h}_{k,\Upsilon}(N-1,0) & \tilde{h}_{k,\Upsilon}(N-1,1) \\
\end{bmatrix}.
\]

(8)

Tracing back to the 1st hop, we have \( z_{k,1} = \bar{s}_{k,1} \mathbb{H}_{k,1} s_k + \bar{s}_{k,1} \mathbb{H}_{k,1} w_{k,1} \), where \( \mathbb{H}_{k,1} \) is an \( (N + \sum_{\rho = 1}^{\Upsilon} L_{k,\rho}^{\text{max}} - \Upsilon) \times (N + \sum_{\rho = 1}^{\Upsilon+1} L_{k,\rho}^{\text{max}} - \Upsilon - 1) \) channel matrix with structure the same as (6) and (8), and \( s_k = E_k s \) with \( E_k = [I_N;1_N,(N-\sum_{\rho = 1}^{\Upsilon} L_{k,\rho}^{\text{max}}+1):N-1_N]^T \) characterizing the effect of the CP. Based on the above derivations, the received
signal vector $\tilde{y}$ is

$$\tilde{y} = \sum_{k=1}^{K} \left[ \prod_{\rho=1}^{L} \tilde{S}_{k,\rho} \right] \begin{pmatrix} \tilde{H}_{k,1} \times \ldots \times \tilde{H}_{k,\rho} \end{pmatrix} \begin{pmatrix} E_k \end{pmatrix} P^H x + \sum_{k=1}^{K} \sum_{\rho=1}^{L} \left[ \prod_{\mu=1}^{\rho} \bar{S}_{k,\mu} \right] \begin{pmatrix} \tilde{H}_{k,1} \times \ldots \times \tilde{H}_{k,\rho} \end{pmatrix} \begin{pmatrix} w_{k,\rho} \end{pmatrix} + w_d, \quad (9)$$

where $H$ represents the composite channel matrix and $\bar{v}$ represents the composite noise effect.

### III. Reformulation of the Composite Channel

In order to equalize the channel and detect the data, it is important to investigate the structure of the channel matrix $H$. Notice that it can be written as $H = \sum_{k=1}^{K} \prod_{\rho=1}^{L} \tilde{S}_{k,\rho}$, where $H_k = \tilde{H}_{k,1} \times \ldots \times \tilde{H}_{k,\rho} \times E_k$. To find out the structure of $\tilde{H}_{k,1} \times \ldots \tilde{H}_{k,\rho} \times E_k$, we start from $H_k,1$ and $H_k,\rho$ with their expressions given in (6) and (8), respectively. Each matrix represents the linear convolution of a time-varying channel and the matrix multiplication expresses the convolution effect of two time-varying channels. Therefore the resulting matrix $H_k,1 \times \ldots \tilde{H}_{k,\rho} \times E_k$ will also be in the form of (6) and (8), except that the resulting channel length of the new time-varying channel is now being $L_k,1 + L_k,\rho \times \bar{v}$.

Continuing the matrix multiplication, it can be shown that $H_{k,1} \times \ldots \tilde{H}_{k,\rho} \times E_k$ is an $N \times (N + \sum_{\rho=1}^{L} L_k,\rho \times \bar{v} - 1)$ matrix, with equivalent channel length of $\sum_{\rho=1}^{L} L_k,\rho \times \bar{v} - 1$. And eventually $E_k$ moves the $\sum_{\rho=1}^{L_k} L_k,\rho \times \bar{v} - 1$ columns from the left part of $H_k,1 \times \ldots \tilde{H}_{k,\rho} \times E_k$ to the upper right corner. The resulted composite channel matrix $H_k$ is an $N \times N$ circular convolution matrix of a time-varying channel with equivalent channel length of $\sum_{\rho=1}^{L_k} L_k,\rho \times \bar{v} - 1$.

Thus $H$, as the weighted sum of $H_k$‘s, has the same circular convolution matrix structure of a time-varying channel with length $L_{max} = \max(\sum_{\rho=1}^{L} L_k,\rho \times \bar{v} - 1)$, and it can be written as $H = \sum_{l=0}^{L_{max}-1} \text{diag}(\mu_l) \text{P}(l)$, where $\mu_l = [\tilde{\mu}(0, l), \ldots, \tilde{\mu}(N - 1, l)]^T$ consists of all the composite channel coefficients of the $l$th tap and $\text{P}(l) = [e_{l1}, \ldots, e_{lN-1}, e_{l0}, \ldots, e_{l1-1}]$. Thus (9) becomes

$$\tilde{y} = \sum_{l=0}^{L_{max}-1} \text{diag}(\mu_l) \text{P}(l) P^H x + \bar{v}, \quad (10)$$

It should be noticed that, the receiver knows neither the individual channel information of each hop nor the statistical information about the composite channel. This is a natural assumption, as the channels are time-varying and depend on the speed of transceivers and the environment around them.

In order to proceed, we propose to calculate an upper bound on the maximum Doppler shift and the delay for the composite channel. Let $v_{max}$ be the maximum relative velocity between two units in any hop in the relaying system. Since $v_{max} f_e/c \geq f_{k,\rho}(l)$ for all $k, \rho$ and $l$, where $f_e$ and $c$ are the carrier frequency and the speed of light, respectively, we have $f_{max} \leq f_u = (\bar{v} + 1)c_{cp}/c$. And in the delay domain, the best the receiver knows is that $L_{cp}$ is chosen large enough to avoid ISI. Thus $L_{max} \leq L_{cp}$ and all the nonzero taps fall in the range of $\{0, \ldots, L_{cp} - 1\}$. With the ranges of the delay-Doppler domain defined for the composite channel, we can expand the channel with generalized complex exponential basis expansion model (GCE-BEM) as follows

$$\tilde{y}(n,l) = \sum_{q=1}^{Q} \mu_q(l)e^{j2\pi qn/VN} \times \bar{v}, \quad n = 0, \ldots, L_{cp} - 1, \quad (11)$$

where $Q = \lfloor VN f_e T_s \rfloor$ and $V$ is the oversampling factor, and $\mu_q(l)$ is the GCE-BEM coefficient.

From (11), $\tilde{y}(n,l)$ can be put into a vector $\bar{\mu}_l$ as $\bar{\mu}_l = [\mu_q(0), \ldots, \mu_q(L_{cp} - 1)]^T$, then (12) can be written as (13), shown at the top of next page. Further define $\mu = [\mu_1^2, \ldots, \mu_Q^2]^T$ and let $G[x] = [G_1[x], \ldots, G_Q[x]]$, thus we have

$$y = G[x] \mu + v. \quad (14)$$

On the other hand, from (9), let $D[\mu] = FHF^H$, the system model can also be written as

$$y = D[\mu] x + v. \quad (15)$$

It is clear that $D[\mu] x = G[x] \mu$.

Now, using (14), we can write the likelihood function of $y$. In general, the elements of composite noise $v$ are correlated, and the likelihood function would be a function of an unknown correlation matrix. But in order to facilitate the subsequent derivations, we approximate the noise as white but with unknown variance $\sigma_v^2$, and the likelihood function is

$$p(y|\mu, \sigma_v^2, x_d) \sim \frac{1}{(2\pi\sigma_v^2)^N} \exp\left(-\frac{1}{2\sigma_v^2} \|y - G[x] \mu\|^2\right). \quad (16)$$

Notice that the approximation in the likelihood function is used for algorithm derivation only. Correlated noise $v$ will be used in simulation for performance assessment.

### IV. Equalization via Variational EM

From the system models (14) and (15), the problem is to estimate the data $x_d$, under unknown composite channel BEM coefficients $\mu$ and the variance of composite noise $\sigma_v^2$. It is noticed from (14) and (15) that, estimation of channel requires knowledge of data while data detection depends on the accuracy of channel estimate, thus leads to challenges in joint channel estimation and data detection. In this paper,
\[
\sum_{q=-Q}^{Q} \frac{F(\text{diag}\{\varphi(q)\})P(0)F^H x, \ldots, \text{diag}\{\varphi(q)\}P(L_{cp}-1)F^H x)\mu_q + v.}{G_q(x)} \tag{13}
\]

variational EM framework is adopted to iteratively improve the channel estimation and data detection results. Below, we first assign prior distributions to the unknown parameters.

A. Prior Distributions of the Unknown Parameters

First, the prior distribution of \(\mu\) is assumed to be Gaussian
\[
p(\mu|\alpha) \sim \mathcal{CN}(0, \text{diag}\{\alpha\}^{-1}), \tag{17}\]
where \(\alpha = [\alpha_1, \ldots, \alpha_M]^T\) is a vector containing the inverse variance of the elements of \(\mu\), and \(M = (2Q + 1)L_{cp}\). Furthermore, a hyperprior for \(\alpha\) is specified as \([7]\)
\[
p(\alpha_j) = \frac{\Gamma(a_j)}{\Gamma(a_j)} \left(\frac{\alpha_j}{a_j}\right)^{a_j-1} \exp\left(-\frac{\alpha_j}{a_j}\right) \tag{18}\]
with \(a_j, b_j\) being the parameters of the Gamma distribution.

Besides, the unknown noise power is assumed to obey a Gamma prior, such that it can be learned under the variational framework. For ease of expression, let \(\beta = 1/\varpi_e^2\) and then
\[
p(\beta) = \frac{\beta^{c-1}}{\Gamma(c)} \exp(-\beta/d), \tag{19}\]
where \(c, d\) are the parameters of the Gamma distribution. In the absence of prior information, small values for hyperparameters are chosen, i.e., \(a_j = b_j = c = d = 10^{-6}\), so as to produce uninformative priors for the channel and noise power \([7]\).

B. Variational EM

Given the prior distributions and the likelihood function, the unknown data \(x_d\) is estimated by maximizing the marginal log-likelihood function
\[
\log p(y|x_d) = \log \int p(\mu, \alpha, \beta, y|x_d) d\mu d\alpha d\beta, \tag{20}\]
which is difficult to obtain analytically due to the multivariate integration over \(\mu, \alpha\) and \(\beta\).

On the other hand, we can write \(\log p(y|x_d)\) into
\[
\log p(y|x_d) = F(q, x_d) + KL(q||p), \tag{21}\]
with
\[
F(q, x_d) = E_{q(\theta)} \left\{ \log \frac{p(y, \theta|x_d)}{q(\theta)} \right\}, \tag{22}\]
and
\[
KL(q||p) = -E_{q(\theta)} \left\{ \log \frac{p(\theta|x_d)}{q(\theta)} \right\}, \tag{23}\]
where \(q(\theta)\) represents any probability density function (pdf) of \(\theta = \{\mu, \alpha, \beta\}\), and \(KL(q||p)\) denotes the Kullback-Leibler divergence between \(p(\theta|y, x_d)\) and \(q(\theta)\). Due to the non-negativeness of \(KL(q||p)\), it always satisfies \(\log p(y|x_d) \geq F(q, x_d)\).

Based on (20), the EM algorithm replaces maximization of \(\log p(y|x_d)\) by a two-stage iterative algorithm that alternatively maximizes its lower-bound \(F(q, x_d)\) with respect to \(q(\theta)\) given previous estimate of \(x_d^{old}\), and with respect to \(x_d\) given previous estimate of \(q^{old}(\theta)\) \([8]\).

In the first stage, maximization of \(F(q, x_d)\) with respect to \(q(\theta)\) given previous estimate of \(x_d^{old}\) produces \(q(\theta) = p(\mu, \alpha, \beta|y, x_d^{old})\). However, calculation of \(p(\mu, \alpha, \beta|y, x_d^{old})\) is in general very hard to obtain in closed-form, and in consequence maximization of \(F(q, x_d)\) with respect to \(x_d\) given previous estimate of \(q^{old}(\theta)\) in the second stage is inconvenient.

In the variational EM framework, a factorized form \(q(\theta) = q(\mu)q(\alpha)q(\beta)\) is adopted. With this factorized form, the maximization of \(F(q, x_d)\) with respect to \(q(\mu)q(\alpha)q(\beta)\) produces the following expressions \([8]\)
\[
q(\mu) = \frac{\exp\{E_{q(\alpha)q(\beta)}(\log p(\mu, \alpha, \beta, y|x_d))\}}{\int \exp\{E_{q(\alpha)q(\beta)}(\log p(\mu, \alpha, \beta, y|x_d))\} d\alpha d\beta}, \tag{24}\]
and
\[
q(\alpha) = \frac{\exp\{E_{q(\mu)q(\beta)}(\log p(\mu, \alpha, \beta, y|x_d))\}}{\int \exp\{E_{q(\mu)q(\beta)}(\log p(\mu, \alpha, \beta, y|x_d))\} d\mu d\beta}, \tag{25}\]
(26) - (29) into (23) - (25), and through similar derivations as in \([8]\), we obtain
\[
q(\mu) = \frac{\exp\{E_{q(\alpha)}(\log p(\mu, \alpha, y|x_d))\}}{\int \exp\{E_{q(\alpha)}(\log p(\mu, \alpha, y|x_d))\} d\alpha}, \tag{26}\]
and
\[
q(\beta) = \frac{\exp\{E_{q(\mu)q(\beta)}(\log p(\mu, \beta, y|x_d))\}}{\int \exp\{E_{q(\mu)q(\beta)}(\log p(\mu, \beta, y|x_d))\} d\mu d\beta}, \tag{27}\]
(28) with parameters given by
\[
\hat{\Sigma}_\mu = \left(\text{diag}\left\{\frac{\tilde{a}_1}{b_1}, \ldots, \frac{\tilde{a}_M}{b_M}\right\}\right)^{-1} \tag{29}\]
\[
\hat{\mu}_\mu = \left(\frac{\tilde{c}}{d}\right) \hat{\Sigma}_\mu G^H \hat{x} G \hat{x} \tag{30}\]
\[
\tilde{a}_j = a_j + 1 \tag{31}\]
\[
\tilde{b}_j = b_j + ||\hat{\mu}_\mu||^2 + ||\hat{\Sigma}_\mu||_j \tag{32}\]
\[
\tilde{c} = c + 3 \tag{33}\]
\[
\tilde{d} = d + y^H y - 2 \text{Re} \left\{ y^H G \hat{x} \hat{\mu}_\mu \right\} + \text{Tr} \left\{ G^H G \hat{x} (\hat{\mu}_\mu \hat{\mu}_\mu^H + \hat{\Sigma}_\mu) \right\}. \tag{34}\]

Given the estimate of \(q^{old}(\mu, \alpha, \beta) = q^{old}(\mu)q^{old}(\alpha)q^{old}(\beta)\) from the first stage, the lower-bound \(F(q^{old}, x_d)\) is now written as
\[
F(q^{old}, x_d) = \mathbb{E}_{q^{old}(\mu)q^{old}(\alpha)q^{old}(\beta)} \left\{ \log \frac{p(y, \mu, \alpha, \beta|x_d)}{q^{old}(\mu)q^{old}(\alpha)q^{old}(\beta)} \right\} \tag{35}\]
Putting (16), (26) and (28) into (35), and dropping those terms independent of \(x_d\), we have

\[
\mathcal{F}(q^{\text{old}}; x_d) \propto \mathbb{E}_{q^{\text{old}}(\mu)^{\text{old}}(\beta)} \{ \log(y | \mu, \beta, x_d) \} \\
\propto -\mathbb{E}_{q^{\text{old}}(\mu)} \{ ||y - G[x]|^2 || \} \\
- \text{Tr} \{ G_H^* [G] G[x] (\tilde{m}_\mu \tilde{m}_\mu^H + \tilde{\Sigma}_\mu) \} \\
+ 2 \text{Re} \{ \tilde{m}_\mu G_H [y] \} - y^H y. \tag{36}
\]

Since \(x_d\) (contained in \(x\) via (2)) is nonlinear in (36), maximization of (36) is cumbersome. In order to proceed, we perform the eigen-decomposition \(\tilde{\Sigma}_\mu = \sum_{j=1}^M \lambda_j \xi_j \xi_j^H\), and we have

\[
\text{Tr} \{ G_H^* [G] G[x] \tilde{\Sigma}_\mu \} = \sum_{j=1}^M \lambda_j G_H^* [G] G[x] \xi_j. \tag{37}
\]

Putting (16), (26) and (28) into (35), and dropping those terms independent of \(x_d\), we have

\[
\mathcal{F}(q^{\text{old}}; x_d) \propto -x_d^H \left( \sum_{j=1}^M \lambda_j D_H^* \xi_j D_{\xi_j} + D_H^* \tilde{m}_\mu D_{\tilde{m}_\mu} \right) x_d \\
+ 2 \text{Re} \left\{ x_d^H \left( D_H^* \tilde{m}_\mu [y] - \sum_{j=1}^M \lambda_j D_H^* \xi_j D_{\xi_j} x_p \right) \right\}. \tag{38}
\]

Then constellation mapping is carried out to obtain the estimate of \(x_d\) as \(\tilde{x}_d = \text{Qant}[\tilde{x}_d]\).

In summary, the variational EM algorithm is performed by iterating among (29)-(34) and (38) until it converges. It is worth noting that, along with each iteration, when \(||\tilde{m}_{\mu,j}||_2^2 + ||\tilde{\Sigma}_{\mu,j}||_2^2\) gets close to zero, meaning both mean and variance of the corresponding BEM coefficients \(\mu_j\) are close to zero, then \(\mu_j\) can be treated as null entry and pruned from further iteration. In practice, a threshold with the order of 10^{-10} is used to compare with \(||\tilde{m}_{\mu,j}||_2^2 + ||\tilde{\Sigma}_{\mu,j}||_2^2\) to determine which \(\mu_j\) is being pruned [9].

V. INITIALIZATION OF THE ITERATIVE ALGORITHM

According to (2), \(y = \mathbf{E}_d x_d + \mathbf{E}_p x_p\). Together with the fact that \(\mathbb{G} = \mathbb{E}_p x_p + \mathbb{E}_d x_d\) and \(\mathbb{G} = \mathbb{G}_p \mathbb{E}_p x_p + \mathbb{G}_d \mathbb{E}_d x_d\), (14) can be written as \(y = \mathbb{G} \mathbb{E}_p x_p + \mu + v\). Collecting the output samples corresponding to pilot positions \(\gamma_p\), the equation for initial channel estimation can be written as

\[
y_p = \mathbb{G}_p \mathbb{E}_p x_p + \mu + \mathbb{G}_d \mathbb{E}_d x_d + v_p, \tag{32}
\]

where \(\mathbb{G}_p[\cdot]\) is constructed from the rows of \(\mathbb{G}[\cdot]\) corresponding to \(\gamma_p\). The initial channel estimation can be obtained by treating the second term of (32) as noise and performing LS algorithm, that is \(\hat{\mu} = (\mathbb{G}_p^H \mathbb{G}_p)^{-1} \mathbb{G}_p^H y_p\).

With the estimated channel \(\hat{\mu}\), we rewrite (15) as \(y = D \hat{\mu} \mathbb{E}_d x_d + D \hat{\mu} \mathbb{E}_p x_p + v\). Applying LS estimation again, we have \(\hat{x}_d^0 = (\mathbb{E}_d^H D \hat{\mu} \mathbb{E}_d)^{-1} \mathbb{E}_d^H D \hat{\mu} \mathbb{E}_p x_p + v\). The obtained \(\hat{x}_d^0\) may not reside on the constellation map, thus quantization is performed on \(\hat{x}_d^0\) and the initial channel estimation can be obtained by \(\hat{x}_d^0 = \text{Qant}[\hat{x}_d^0]\). Notice that in DSC, theICI is not negligible, and \(\mathbb{G}_p \mathbb{E}_p x_d + \mu \neq 0\), which decreases the accuracy of initial channel estimation, and in turn affects the accuracy of initial data detection. This is the reason why an iterative algorithm is necessary.

For other initial values \(\{\hat{\alpha}_1^0, \ldots, \hat{\alpha}_M^0\}, \{\hat{\beta}_1^0, \ldots, \hat{\beta}_N^0\}\) in the iterative algorithm, it is should be noticed that only the ratios \(\hat{\alpha}_j^0 / \hat{\beta}_j^0\) and \(\hat{c}/\hat{d}\) are required, thus we only need to specify the initial values of the ratios to start the iteration. From (27) and the property of Gamma distribution, \(\hat{\alpha}_j^0 / \hat{\beta}_j^0\) represents the mean value of \(\alpha_j\), which is the inverse variance of channel GCE-BEM coefficients. Since we have no information about their relative values, we can set them to be equal. That is, let \(\hat{\alpha}_j^0 / \hat{\beta}_j^0 = 1/\hat{M}\) for all \(j\). Furthermore, from (28) and the property of Gamma distribution, \(\hat{c}/\hat{d} = \mathbb{E}\{\hat{\beta}\} = \mathbb{E}\{1/\hat{\alpha}_k^0\}\). Therefore the initial value can be set as \(\hat{c}^0 / \hat{d}^0 = 1/\hat{\alpha}_k^0\), where \(\hat{\alpha}_k^0\) is an estimate of noise power \(\hat{\alpha}_k^0 = ||y - \mathbb{G} \mathbb{E}_p x_p + \mathbb{G}_d \mathbb{E}_d x_d||_2^2 / N\).

VI. SIMULATION RESULTS AND DISCUSSIONS

In this section, simulation results of a three-hop cooperative OFDM system are provided. The system has the following settings. Each OFDM symbol has 128 subcarriers and the length of CP is 8. Carrier frequency is \(f_c = 2\) GHz and the sample interval is \(T_s = 2\) µs. Fourteen pilot clusters are used. The clusters are equal-spaced and interleaved with data subcarriers. In each cluster, one nonzero pilot is guarded by one zero pilot on each side. The nonzero pilots are generated as zero-mean complex Gaussian random variables with power three times that of data symbols. The data is modulated with QPSK of unit power. The maximal normalized Doppler shifts for the first and third hop are set as 0.05 while that of the second hop is set as 0.15. Two relaying paths are considered \((K = 2)\). For the first relaying path, the number of channel taps in the three hops are \(\{2, 3, 2\}\), respectively; while that for the second relaying path is \(\{3, 2, 2\}\), respectively. The channel taps in each hop are consecutive but with unknown zeros on both sides.

The normalized channel mean-square error (MSE) and data detection bit error rate (BER) are plotted to demonstrate the performance. The MSE of the channel estimate is defined as \(\text{MSE} = ||H - \tilde{H}||_2^2 / ||H||_2^2\), where \(H\) is the channel matrix recovered from the GCE-BEM estimate. The noise power at the relays and destination are set to be the same \(\bar{\alpha}_k^2 = \bar{\alpha}_k \rho^2\), for all \(k\) and \(\rho\). The signal-to-noise ratio (SNR) in the figures is

\(^1\)Normalized Doppler shift is defined as \(f_d T_s\), with \(f_d\) being the Doppler frequency.
defined as $\text{SNR} = \frac{\sigma^2_d}{\omega^2_n}$. The oversampling factor of GCE-BEM is chosen as $V = 20$ for iterative algorithm, and $V = 1$ is chosen for initialization. Each point is obtained by averaging the results over 1,000 runs.

Figure 1 and Figure 2 show the MSE of channel estimation and BER performance achieved by the proposed iterative algorithm versus SNRs, with results taken after 10 iterations. In both figures, performance curves of KLEM, which is the EM algorithm with channel expanded on Karhunen-Loève (KL) bases, and with detailed information of relaying system structure, channel and noise statistics, are depicted as a reference for optimal channel estimation and data detection. From Figure 1, it is seen that, the proposed iterative algorithm greatly improve the performance from the initial channel estimation, indicating the ability of the proposed algorithm to cancel interference between unknown data and pilots through iterations. Furthermore, after convergence, only a small performance gap exists between the proposed algorithm and KLEM. This exhibits the strong ability of our proposed algorithm in learning the statistics of both channel and noise. From Figure 2, the BER performance of our proposed method is also shown to improve significantly compared to the initial data detection and is very close to that of KLEM. Furthermore, the proposed algorithm and KLEM are close to the ideal data detection, which assumes perfect knowledge of all the time-varying channels.

VII. CONCLUSIONS

In this paper, equalization of multihop OFDM relaying channel under high mobility has been investigated with focus on unknown channel orders and Doppler frequencies. By exploring the matrix structure of channels in different hops, we first simplified the multihop multilink channel matrix into a composite channel matrix. A pilot-aided iterative algorithm was developed under the variational EM framework, using only limited number of pilot subcarrier in one OFDM symbol. The proposed algorithm iteratively estimates the channel, learns the channel and noise statistical information, and recovers the unknown data. Simulation results showed that, even without any specific information on system structure, channel tap positions, channel lengths, Doppler shifts and noise powers, the proposed algorithm exhibited performance very close to that of an optimal channel estimation and data detection algorithm, which requires all of the above information.

ACKNOWLEDGMENTS

This work was supported by the Hong Kong Research Grants Council (Grant No.: 719111E).

REFERENCES