

<b>Title</b>	<b>Superimposed training-based channel estimation and data detection for OFDM amplify-and-forward cooperative systems under high mobility</b>
<b>Author(s)</b>	<b>He, L; Wu, YC; Ma, S; Ng, TS; Poor, HV</b>
<b>Citation</b>	<b>IEEE Transactions on Signal Processing, 2012, v. 60 n. 1, p. 274-284</b>
<b>Issued Date</b>	<b>2012</b>
<b>URL</b>	<b><a href="http://hdl.handle.net/10722/155718">http://hdl.handle.net/10722/155718</a></b>
<b>Rights</b>	<b>IEEE Transactions on Signal Processing. Copyright © IEEE</b>

# Superimposed Training-Based Channel Estimation and Data Detection for OFDM Amplify-and-Forward Cooperative Systems Under High Mobility

Lanlan He, Yik-Chung Wu, Shaoan Ma, Tung-Sang Ng, and H. Vincent Poor, *Fellow, IEEE*

**Abstract**—In this paper, joint channel estimation and data detection in orthogonal frequency division multiplexing (OFDM) amplify-and-forward (AF) cooperative systems under high mobility is investigated. Unlike previous works on cooperative systems in which a number of subcarriers are solely occupied by pilots, partial data-dependent superimposed training (PDDST) is considered here, thus preserving the spectral efficiency. First, a closed-form channel estimator is developed based on the least squares (LS) method with Tikhonov regularization and a corresponding data detection algorithm is proposed using the linear minimum mean square error (LMMSE) criterion. In the derived channel estimator, the unknown data is treated as part of the noise and the resulting data detection may not meet the required performance. To address this issue, an iterative method based on the variational inference approach is derived to improve performance. Simulation results show that the data detection performance of the proposed iterative algorithm initialized by the LMMSE data detector is close to the ideal case with perfect channel state information.

**Index Terms**—Amplify-and-forward, orthogonal frequency division multiplexing (OFDM), time-varying channels.

## I. INTRODUCTION

COOPERATIVE communications has attracted much attention recently due to its advantages in enhancing link reliability and increasing channel capacity [1]–[5]. Since orthogonal frequency division multiplexing (OFDM) transmission has been adopted as the transmission technology for next generation broadband wireless standards [such as IEEE 802.16

and long term evolution (LTE)], this results in the need to develop receiver algorithms for OFDM based cooperative communications [3]–[5]. On the other hand, another important goal for the next generation of wireless broadband systems is to support high user mobility. With high mobility, the time-variation of the channel within one OFDM symbol cannot be ignored and channel responses vary sample by sample, which makes channel estimation challenging. Moreover, high mobility destroys the orthogonality among subcarriers and induces intercarrier interference (ICI), which also complicates data detection [6]–[9]. As a result, channel estimation and data detection for OFDM cooperative systems under high mobility is very challenging.

Channel estimation and data detection for cooperative communications with time-invariant channels has been studied in [1]–[5]. Specifically, channel estimation is studied in [1] and [2] under the assumption of time-invariant flat fading channels. Targeting OFDM transmission, training sequence based channel estimation is investigated in [3] and [4], and data detection performance analysis is considered in [5] by assuming perfect channel state information. However, algorithms applicable to time-varying channels cannot be obtained through direct extension of these previous works. Recently, channel estimation for cooperative OFDM systems with time-varying channels has been studied in [10], in which the amplify-and-forward (AF) scheme is adopted due to its low complexity and minimal delay. By exploiting a time-frequency representation of the received signals, channel estimation algorithms are proposed for two different scenarios. In the first scenario, the corresponding channels are individually estimated at the relay and destination, whereas in the second one, the cascaded source-relay-destination channel is jointly estimated at the destination. Semiblind (i.e., only certain subcarriers are occupied by pilots) channel estimation with unknown data estimated by methods proposed for time-invariant channels is considered, which may lead to severe modeling errors and performance degradation. Moreover, only one relay is considered in [10] and the extension to multiple relays using the time-frequency representation framework is by no means straightforward.

In this paper, an OFDM-based AF cooperative system with one source, multiple relays and one destination is considered. In order to reduce the computational load on relays and time delay of the whole system, no channel estimation is performed at the relays, which corresponds to the second scenario in [10]. Unlike previous works in which subcarriers are occupied by either pilots or data [10], [11], here partial data-dependent superimposed training (PDDST) [8] is adopted for channel estimation and data detection. Notice that superimposed training (ST)

Manuscript received October 06, 2010; revised February 18, 2011 and June 24, 2011; accepted August 19, 2011. Date of publication September 22, 2011; date of current version December 16, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ye (Geoffrey) Li. The work was supported in part by the General Research Fund (GRF) from Hong Kong Research Grant Council (Project No.: HKU 7154/08E), by the GRF (Project No. HKU 7191/11E), and by the U.S. National Science Foundation by Grant CNS-09-05398.

L. He was with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong. She is now with the Huawei Tech. Investment Company (e-mail: llhe@eee.hku.hk).

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

S. Ma was with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong. She is now with the Department of Electrical and Computer Engineering, University of Macau, Macau (e-mail: shaodanma@umac.mo).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2011.2169059

based channel estimation and data detection have been widely investigated in point-to-point systems [8], [12]–[14]. This superimposed training is even more practical for systems over time-varying channels. Otherwise, a number of subcarriers in each OFDM symbol need to be assigned to pilots for channel estimation purpose, which would result in low spectral efficiency [10], [11].

The contributions of this paper are as follow. First, based on the generalized complex exponential basis expansion model (GCE-BEM), the system model is reformulated in a nontrivial way to obtain an expression similar to that for a conventional point-to-point OFDM system. After that, a closed-form channel estimator is developed based on the least squares (LS) method with Tikhonov regularization [15] and a corresponding data detection algorithm is proposed using the linear minimum mean square error (LMMSE) criterion. In the LS-based channel estimator, since the unknown data is treated as part of the noise, the resulting system performance may not meet the requirements. To address this issue, an iterative method based on the variational inference approach is derived to improve performance. The variational inference approach is useful in cases when direct access or maximization of the posterior distribution of parameters to be estimated is difficult, if not impossible. In particular, the variational inference approach constructs a distribution similar to that of the true posterior but with a tractable form [16]. Since it is basically a Bayesian framework, statistical information (such as channel statistics, power of data and noise) is exploited to aid the estimation. Finally, computer simulations are performed to demonstrate the effectiveness of the proposed channel estimation and data detection algorithms.

The rest of the paper is organized as follows. The channel and system model for OFDM cooperative systems is introduced in Section II. The system model is then reformulated using the GCE-BEM in Section III. In Section IV, a channel estimator is developed based on the LS method with the Tikhonov regularization and a corresponding data detection algorithm is proposed using the LMMSE criterion. In Section V, based on the variational inference approach, an iterative enhancement algorithm is developed. Simulation results are presented in Section VI to demonstrate the effectiveness of the proposed algorithms. Finally, conclusions are drawn in Section VII.

**Notation:** Boldface uppercase and lowercase letters are used for matrices and vectors, respectively. Superscripts  $T$ ,  $H$  and  $\dagger$  denote transpose, Hermitian, and pseudo-inverse, respectively. The symbol  $\mathbf{I}_N$  denotes the  $N \times N$  identity matrix. The symbol  $\text{diag}\{\mathbf{x}\}$  signifies the diagonal matrix with vector  $\mathbf{x}$  on its diagonal and  $\|\mathbf{x}\|$  represents the  $L_2$  norm of  $\mathbf{x}$ .  $\text{Tr}\{\mathbf{X}\}$  and  $|\mathbf{X}|$  are the trace and the determinant of a square matrix  $\mathbf{X}$ , respectively. Symbols  $\mathbb{E}\{\cdot\}$  and  $\Re\{\cdot\}$  denote the expectation and the real part of the operand in the brackets, respectively. The symbol  $\otimes$  denotes convolution. The matrix  $\mathbf{F}$  denotes the fast Fourier transform (FFT) matrix with  $[\mathbf{F}]_{m,n} = \frac{1}{\sqrt{N}}e^{-j\frac{2\pi mn}{N}}$  and  $\lceil a \rceil$  rounds  $a$  to the nearest integer greater than or equal to  $a$ .

## II. SYSTEM MODEL

In this paper, we consider a cooperative system with a source  $\mathbb{S}$ , a destination  $\mathbb{D}$ , and  $K$  relays  $\mathbb{R}_k$  scattered between  $\mathbb{S}$  and  $\mathbb{D}$ . Each of these elements is equipped with a single antenna. Denote the channels from the source  $\mathbb{S}$  to the  $k$ th relay  $\mathbb{R}_k$  and from

the  $k$ th relay  $\mathbb{R}_k$  to the destination  $\mathbb{D}$  as  $h_k(t, \tau)$  and  $g_k(t, \tau)$ , respectively. There is no direct link between source and destination. It is assumed that the  $K$  relays are stationary but the source and destination are moving at high speed. Thus the propagation channels are modeled as multi-path time-varying channels. Specifically, the source-relay channel  $h_k(t, \tau)$  has  $L_k^h$  independent taps with the average power of the  $l$ th tap denoted by  $\varpi_{l,k}^2$ . The autocorrelation of the  $l$ th tap follows the classical Jakes' model [6] given by  $\mathbb{E}\{h_k(mT_s, l)h_k^*(nT_s, l)\} = \varpi_{l,k}^2 J_0(2\pi f_h(n - m)T_s)$ , where  $h_k(nT_s, l)$  is the sample of the  $l$ th tap at time  $nT_s$  ( $T_s$  is the sample interval),  $J_0(\cdot)$  represents the zero-order Bessel function of the first kind, and  $f_h$  is the maximal Doppler shift between source and the relays. Similarly, the relay-destination channel  $g_k(t, \tau)$  has  $L_k^g$  independent taps with the average power of the  $l$ th tap denoted by  $\varsigma_{l,k}^2$ . The autocorrelation of the  $l$ th tap is  $\mathbb{E}\{g_k(mT_s, l)g_k^*(nT_s, l)\} = \varsigma_{l,k}^2 J_0(2\pi f_g(n - m)T_s)$ , where  $g_k(nT_s, l)$  is the sample of the  $l$ th tap at time  $nT_s$ , and  $f_g$  represents the maximal Doppler shift between the relays and destination.

**Transmitted Signal at the Source:** In an OFDM system, the source data in the frequency domain  $\mathbf{x} = [x(0), \dots, x(N-1)]^T$  is modulated onto  $N$  parallel subcarriers to obtain the time domain signal  $\mathbf{s} = \mathbf{F}^H \mathbf{x}$ . In this paper, we consider partial data-dependent superimposed training [8], which is a general description on the placement of pilots and data. The  $i$ th element of  $\mathbf{x}$  is given by

$$x(i) = \begin{cases} (1 - \xi)x_d(i) + x_p(i) & \forall i \in \varrho \\ x_d(i) & \text{otherwise} \end{cases} \quad (1)$$

where  $\varrho$ , with cardinality  $N_p$ , is the index set of subcarriers on which both data and pilots are transmitted. Therefore the transmitted symbol  $x(i)$  at the  $i$ th subcarrier is a linear combination of a pilot symbol and a data symbol. The total power for the  $i$ th subcarrier ( $i \in \varrho$ ) is  $(1 - \xi)^2 \mathbb{E}\|x_d(i)\|^2 + \mathbb{E}\|x_p(i)\|^2$ . Furthermore, without loss of generality, it is assumed that the average power of  $x_d(i)$  is  $\sigma_x^2$ . From (1), we have

$$\mathbf{x} = \mathbf{A}_\varrho \mathbf{x}_d + \mathbf{E}_\varrho \mathbf{x}_p \quad (2)$$

where  $\mathbf{A}_\varrho$  is a diagonal matrix with

$$[\mathbf{A}_\varrho]_{i,i} = \begin{cases} 1 - \xi & \forall i \in \varrho \\ 1 & \text{otherwise} \end{cases} \quad (3)$$

$\mathbf{E}_\varrho$  is a matrix collecting columns of  $\mathbf{I}_N$  with indices in  $\varrho$ , and  $\mathbf{x}_p$  and  $\mathbf{x}_d$  denote pilot and data vectors, respectively.

From (1), it is noticed that the PDDST includes the following three cases:

- When  $0 < \xi < 1$ , the data component at each subcarrier  $i \in \varrho$  is  $(1 - \xi)x_d(i)$ . In this case, both data and pilots are transmitted on subcarrier set  $\varrho$ .
- In the case  $\xi = 1$ , the data component at each subcarrier  $i \in \varrho$  is nulled, and the PDDST reduces to the semi-blind case in which subcarriers are uniquely occupied by pilots or data. This case is also referred to as data-dependent superimposed training (DDST) [13].
- In the case  $\xi = 0$ , the PDDST reduces to traditional superimposed training [12], and  $\mathbf{x}_p$  becomes the training in the frequency domain while the data  $\mathbf{x}_d$  remains intact.

Upon transmission, a cyclic prefix (CP) of length  $L_{cp}$  longer than  $\max_{k=1}^K (L_k^h + L_k^g - 1)$  is inserted at the beginning of the

time domain OFDM signal  $\mathbf{s}$  to prevent intersymbol interference (ISI) at the destination D.

*Signal Processing at Relays:* The signal received at the  $k$ th relay  $r_k(n)$  is given by ( $T_s$  is omitted from the time index for notational simplicity)

$$r_k(n) = \sum_{l=0}^{L_k^h-1} h_k(n, l) s(n-l) + v_k(n) \quad (4)$$

where  $v_k(n)$  denotes additive white Gaussian noise (AWGN) with average power  $\sigma_{v_k}^2$ . Upon reception, the  $k$ th relay  $\mathbb{R}_k$  simply amplifies the incoming signal  $r_k(n)$ . The transmitted signal from the  $k$ th relay is thus written as [3]

$$z_k(n) = \alpha_k r_k(n) \quad (5)$$

where

$$\alpha_k = \frac{\sqrt{p_k}}{\sqrt{\sigma_s^2 \sum_{l=0}^{L_k^g-1} \varpi_{l,k}^2 + \sigma_{v_k}^2}} \quad (6)$$

with  $p_k$  being the transmission power at the relay  $\mathbb{R}_k$  and  $\sigma_s^2$  is the average power of  $\mathbf{s}$ .

*Received Signal at the Destination:* The received signal at the destination is a superposition of the signals from  $K$  relays and is given by

$$y(n) = \sum_{k=1}^K \sum_{l=0}^{L_k^g-1} g_k(n, l) z_k(n-l) + w(n). \quad (7)$$

After removing the CP, the received signal vector  $\mathbf{y} = [y(0), \dots, y(N-1)]^T$  is

$$\mathbf{y} = \sum_{k=1}^K \mathbf{G}_k \mathbf{z}_k + \mathbf{w} \quad (8)$$

where  $\mathbf{z}_k = [z_k(-(L_k^g-1)), \dots, z_k(0), \dots, z_k(N-1)]^T$ ,  $\mathbf{G}_k$  is an  $N \times (N + L_k^g - 1)$  channel matrix given by

$$\mathbf{G}_k = \begin{bmatrix} g_k(0, L_k^g-1) \cdots g_k(0, 0) \\ g_k(1, L_k^g-1) \cdots g_k(1, 0) \\ \dots \dots \dots \\ g_k(N-1, L_k^g-1) \cdots g_k(N-1, 0) \end{bmatrix} \quad (9)$$

and  $\mathbf{w}$  denotes an AWGN vector  $[w(0), \dots, w(N-1)]^T$  with average power  $\sigma_w^2$ . Furthermore, according to (4) and (5), the signal vector  $\mathbf{z}_k$  can be compactly expressed as

$$\mathbf{z}_k = \alpha_k \mathbf{H}_k \mathbb{I}_{L_k^c} \mathbf{s} + \alpha_k \mathbf{v}_k \quad (10)$$

where  $\mathbf{H}_k$  is an  $(N + L_k^g - 1) \times (N + L_k^g + L_k^h - 2)$  channel matrix given by

$$\mathbf{H}_k = \begin{bmatrix} h_k(-(L_k^g-1), L_k^h-1) \cdots h_k(-(L_k^g-1), 0) \\ h_k(-(L_k^g-2), L_k^h-1) \cdots h_k(-(L_k^g-2), 0) \\ \dots \dots \dots \\ h_k(N-1, L_k^h-1) \cdots h_k(N-1, 0) \end{bmatrix} \quad (11)$$

$\mathbb{I}_{L_k^c} = [\mathbf{E}_{L_k^c}, \mathbf{I}_N]^T$  characterizes the effect of the CP with  $\mathbf{E}_{L_k^c}$  being the last  $L_k^c \triangleq L_k^g + L_k^h - 2$  columns

of the identity matrix, and  $\mathbf{v}_k$  denotes an AWGN vector  $[v_k(-(L_k^g-1)), \dots, v_k(0), \dots, v_k(N-1)]^T$ . Based on (8), (10) and  $\mathbf{s} = \mathbf{F}^H \mathbf{x}$ , the received signal vector  $\mathbf{y}$  is

$$\mathbf{y} = \sum_{k=1}^K \alpha_k \mathbf{G}_k \mathbf{H}_k \mathbb{I}_{L_k^c} \mathbf{F}^H \mathbf{x} + \underbrace{\sum_{k=1}^K \alpha_k \mathbf{G}_k \mathbf{v}_k}_{\triangleq \mathbf{w}}. \quad (12)$$

### III. REFORMULATION WITH THE BASIS EXPANSION MODEL

Channel state information is generally required for data detection. It is clear from (9) and (11) that the number of unknown channel parameters in channel matrices  $\mathbf{G}_k$  and  $\mathbf{H}_k$  are  $N L_k^g$  and  $(N + L_k^g - 1) L_k^h$ , respectively, which are much larger than the number of received samples  $N$ . This makes direct channel estimation impossible. However, due to the fact that the time-varying channel is time-correlated, the GCE-BEM [6] can be adopted to represent the channels, and the number of unknown channel parameters can be significantly reduced. With GCE-BEM, the matrix  $\mathbf{H}_k$  in (11) can be approximated by

$$\mathbf{H}_k \cong \sum_{m=-\rho_h}^{\rho_h} \Phi(m) \mathbf{H}_k^b(m) \quad (13)$$

where  $\rho_h = \lceil C_h N f_h T_s \rceil$  with  $C_h$  being the oversampling factor in the Doppler domain,

$$\Phi(m) = \text{diag} \left\{ e^{j2\pi m(-(L_k^g-1))/C_h N}, \dots, e^{j2\pi m(N-1)/C_h N} \right\} \quad (14)$$

and

$$\mathbf{H}_k^b(m) = \begin{bmatrix} h_k^b(m, L_k^h-1) \cdots h_k^b(m, 0) \\ h_k^b(m, L_k^h-1) \cdots h_k^b(m, 0) \\ \dots \dots \dots \\ h_k^b(m, L_k^h-1) \cdots h_k^b(m, 0) \end{bmatrix} \quad (15)$$

with  $h_k^b(m, l)$  denoting the BEM coefficient characterizing the source-relay channel  $h_k(t, \tau)$ . Here the oversampling factor  $C_h$  is an integer adjusting the Doppler range sampled by the BEM and the number of basis vectors used to represent the time-varying channel. Similarly, the matrix  $\mathbf{G}_k$  in (9) can be approximated by

$$\mathbf{G}_k \cong \sum_{q=-\rho_g}^{\rho_g} \Theta(q) \mathbf{G}_k^b(q) \quad (16)$$

where  $\rho_g = \lceil C_g N f_g T_s \rceil$  with  $C_g$  (similar to  $C_h$ ) being the corresponding oversampling factor

$$\Theta(q) = \text{diag} \left\{ e^{j2\pi q(0)/C_g N}, \dots, e^{j2\pi q(N-1)/C_g N} \right\} \quad (17)$$

and

$$\mathbf{G}_k^b(q) = \begin{bmatrix} g_k^b(q, L_k^g-1) \cdots g_k^b(q, 0) \\ g_k^b(q, L_k^g-1) \cdots g_k^b(q, 0) \\ \dots \dots \dots \\ g_k^b(q, L_k^g-1) \cdots g_k^b(q, 0) \end{bmatrix} \quad (18)$$

with  $g_k^b(q, l)$  denoting the BEM coefficient characterizing the relay-destination channel  $g_k(t, \tau)$ .

Substituting (13) and (16) into (12), it follows that

$$\mathbf{y} = \sum_{k=1}^K \sum_{q=-\rho_g}^{\rho_g} \sum_{m=-\rho_h}^{\rho_h} \alpha_k \boldsymbol{\Theta}(q) \mathbf{G}_k^b(q) \boldsymbol{\Phi}(m) \mathbf{H}_k^b(m) \mathbb{I}_{L_k^c} \mathbf{F}^H \mathbf{x} + \bar{\mathbf{w}}. \quad (19)$$

Clearly from (19),  $\mathbf{G}_k^b(q)$  and  $\mathbf{H}_k^b(m)$  are unknown and must be estimated. Notice that, from the point of view of data detection, only the combined channels are needed. Therefore, in the following, by exploiting the Toeplitz structure of  $\mathbf{G}_k^b(q)$  and  $\mathbf{H}_k^b(m)$ , it is proved in Appendix A that (19) can be equivalently written as

$$\mathbf{y} = \sum_{q=-\rho_g}^{\rho_g} \sum_{m=-\rho_h}^{\rho_h} \boldsymbol{\Gamma}(q, m) \mathbf{F}^H \text{diag}\{\mathbf{F}_{L_e} \mathbf{u}(q, m)\} \mathbf{x} + \bar{\mathbf{w}} \quad (20)$$

where

$$\boldsymbol{\Gamma}(q, m) = \text{diag} \left\{ e^{\frac{j2\pi(0)(\kappa m + \vartheta q)}{\kappa C_h N}}, \dots, e^{\frac{j2\pi(N-1)(\kappa m + \vartheta q)}{\kappa C_h N}} \right\}$$

with  $\kappa C_h = \vartheta C_g$  being the least common multiple of  $C_h$  and  $C_g$ ,  $\mathbf{F}_{L_e}$  collects the first  $L_e = \max_{k=1}^K L_k^c + 1$  columns of  $\mathbf{F}$ , and  $\mathbf{u}(q, m) = \sum_{k=1}^K \alpha_k \left[ \left( g_k^b(q, 0), \dots, g_k^b(q, L_k^g - 1) e^{-\frac{j2\pi(L_k^g - 1)}{C_h N}} \right) \otimes (h_k^b(m, 0), \dots, h_k^b(m, L_k - 1)) \right]^T$ . From (20), it is clear that all the channel effects are summarized in  $\mathbf{u}(q, m)$ .

Notice that, from the definition of  $\boldsymbol{\Gamma}(q, m)$ , different combinations of  $(q, m)$  may result in the same diagonal matrix  $\boldsymbol{\Gamma}(q, m)$ , i.e.,  $\boldsymbol{\Gamma}(q_1, m_1) = \boldsymbol{\Gamma}(q_2, m_2)$  when  $\kappa q_1 + \vartheta m_1 = \kappa q_2 + \vartheta m_2$ . Combining those terms in (20) with the same  $\boldsymbol{\Gamma}(q, m)$  and denoting the number of distinct matrices  $\boldsymbol{\Gamma}(q, m)$  as  $N_b$ , the received signal vector (20) can be rewritten as

$$\mathbf{y} = \sum_{j=1}^{N_b} \boldsymbol{\Gamma}(\gamma_j) \mathbf{F}^H \text{diag}\{\mathbf{F}_{L_e} \mathbf{u}(\gamma_j)\} \mathbf{x} + \bar{\mathbf{w}} \quad (21)$$

where  $\boldsymbol{\Gamma}(\gamma_j)$  denotes a distinct diagonal matrix  $\boldsymbol{\Gamma}(q, m)$  with  $\kappa q + \vartheta m = \gamma_j$  and  $\mathbf{u}(\gamma_j) = \sum_{\kappa q + \vartheta m = \gamma_j} \mathbf{u}(q, m)$ . Our aim is to estimate the data  $\mathbf{x}_d$  (contained in  $\mathbf{x}$ ) based on (21), while  $\mathbf{u}(\gamma_j)$  is also unknown and required to be estimated.

In particular, substituting  $\mathbf{x} = \mathbf{A}_\varrho \mathbf{x}_d + \mathbf{E}_\varrho \mathbf{x}_p$  into (21), we have

$$\begin{aligned} \mathbf{y} &= \underbrace{\sum_{j=1}^{N_b} \boldsymbol{\Gamma}(\gamma_j) \mathbf{F}^H \text{diag}\{\mathbf{F}_{L_e} \mathbf{u}(\gamma_j)\} \mathbf{A}_\varrho \mathbf{x}_d}_{\triangleq \boldsymbol{\Xi}[\mathbf{u}]} \\ &\quad + \underbrace{\sum_{j=1}^{N_b} \boldsymbol{\Gamma}(\gamma_j) \mathbf{F}^H \text{diag}\{\mathbf{F}_{L_e} \mathbf{u}(\gamma_j)\} \mathbf{E}_\varrho \mathbf{x}_p + \bar{\mathbf{w}}}_{\triangleq \boldsymbol{\eta}[\mathbf{u}]} \end{aligned} \quad (22)$$

with  $\mathbf{u} = [\mathbf{u}^T(\gamma_1), \dots, \mathbf{u}^T(\gamma_{N_b})]^T$ . This equation can be used to estimate the unknown data  $\mathbf{x}_d$  if  $\mathbf{u}(\gamma_j)$  has been estimated. On the other hand, for the convenience of estimating  $\mathbf{u}(\gamma_j)$ , it

will be more efficient to put  $\mathbf{u}(\gamma_j)$  in a linear relationship with  $\mathbf{y}$ . Based on (21), reversing the positions of  $\mathbf{u}(\gamma_j)$  and  $\mathbf{x}$  and writing the summation into matrix form gives

$$\mathbf{y} = \mathbf{B}[\mathbf{x}] \mathbf{u} + \bar{\mathbf{w}} \quad (23)$$

with

$\mathbf{B}[\mathbf{x}] = [\boldsymbol{\Gamma}(\gamma_1) \mathbf{F}^H \text{diag}\{\mathbf{x}\} \mathbf{F}_{L_e}, \dots, \boldsymbol{\Gamma}(\gamma_{N_b}) \mathbf{F}^H \text{diag}\{\mathbf{x}\} \mathbf{F}_{L_e}]$ . The unknown  $\mathbf{u}$  can be estimated by using only the subcarriers containing pilots, which is detailed in the next section.

#### IV. CHANNEL ESTIMATION AND DATA DETECTION

Based on (23) and  $\mathbf{x} = \mathbf{A}_\varrho \mathbf{x}_d + \mathbf{E}_\varrho \mathbf{x}_p$ , taking the FFT and stacking all the received samples corresponding to the subcarrier set  $\varrho$  (where PDDST is located) gives

$$\begin{aligned} \bar{\mathbf{y}}_p &= \mathbf{F}_\varrho \mathbf{B}[\mathbf{A}_\varrho \mathbf{x}_d + \mathbf{E}_\varrho \mathbf{x}_p] \mathbf{u} + \mathbf{F}_\varrho \bar{\mathbf{w}} \\ &= \mathbf{F}_\varrho \mathbf{B}[\mathbf{E}_\varrho \mathbf{x}_p] \mathbf{u} + \underbrace{\mathbf{F}_\varrho \mathbf{B}[\mathbf{A}_\varrho \mathbf{x}_d] \mathbf{u} + \mathbf{F}_\varrho \bar{\mathbf{w}}}_{\triangleq \boldsymbol{\delta}} \end{aligned} \quad (24)$$

where  $\mathbf{F}_\varrho$  collects rows of  $\mathbf{F}$  corresponding to the subcarrier set  $\varrho$ .

According to the definition of the matrix operator  $\mathbf{B}[\cdot]$  in (23), it is composed of a number of submatrices  $\boldsymbol{\Gamma}(\gamma_j) \mathbf{F}^H \text{diag}\{\mathbf{x}\} \mathbf{F}_{L_e}$  which differ only in  $\boldsymbol{\Gamma}(\gamma_j)$ . Since  $\boldsymbol{\Gamma}(\gamma_j) = \text{diag} \left\{ e^{\frac{j2\pi(0)\gamma_j}{\kappa C_h N}}, \dots, e^{\frac{j2\pi(N-1)\gamma_j}{\kappa C_h N}} \right\}$  and  $\gamma_j \ll \kappa C_h N$ ,  $\boldsymbol{\Gamma}(\gamma_j)$  for different values of  $j$  turn out to be quite similar to each other, which would result in columns of  $\mathbf{B}[\mathbf{E}_\varrho \mathbf{x}_p]$  being similar and lead to the problem being ill conditioned. Thus if LS were directly applied to estimate  $\mathbf{u}$ , the estimate would be far away from the true value. Moreover, unlike a rank deficient problem which can be solved via truncated singular value decomposition with a determined numerical rank, for an ill-conditioned matrix, the singular values decay gradually to zero with no significant gap, and therefore, there is no notion of a numerical rank [15]. To deal with this ill-conditioning problem, we hereby employ the Tikhonov regularization method [15] to estimate  $\mathbf{u}$  in (24). By treating  $\boldsymbol{\delta}$  as an effective noise, the estimation problem using LS with Tikhonov regularization can be stated as

$$\min_{\mathbf{u}} \left[ (\mathbf{F}_\varrho \mathbf{B}[\mathbf{E}_\varrho \mathbf{x}_p] \mathbf{u} - \bar{\mathbf{y}}_p)^H (\mathbf{F}_\varrho \mathbf{B}[\mathbf{E}_\varrho \mathbf{x}_p] \mathbf{u} - \bar{\mathbf{y}}_p) + \lambda^2 \mathbf{u}^H \mathbf{L}^H \mathbf{L} \mathbf{u} \right] \quad (25)$$

where  $\mathbf{L}$  denotes a regularization matrix and is chosen according to different criteria (e.g., an identity matrix for minimum energy, a banded matrix for maximal flatness), the variable  $\lambda$  signifies the regularization parameter that balances the minimization of the two terms in (25). For a given  $\lambda$ , based on the LS criterion, the solution for (25) is readily obtained as

$$\mathbf{u}_\lambda = (\mathbf{B}^H [\mathbf{E}_\varrho \mathbf{x}_p] \mathbf{F}_\varrho^H \mathbf{F}_\varrho \mathbf{B} [\mathbf{E}_\varrho \mathbf{x}_p] + \lambda^2 \mathbf{L}^H \mathbf{L})^{-1} \times \mathbf{B}^H [\mathbf{E}_\varrho \mathbf{x}_p] \mathbf{F}_\varrho^H \bar{\mathbf{y}}_p. \quad (26)$$

Notice that in (26), the value of  $\lambda$  could significantly affect the estimation performance by either over-regularization or under-regularization. It is therefore crucial to choose an appropriate  $\lambda$ . An intuitive approach to design  $\lambda$  has been developed

based on the concept of the L-curve [15], which are curves of the smoothing norm  $\|\mathbf{L}\mathbf{u}_\lambda\|^2$  versus the corresponding residual norm  $\|\mathbf{F}_\theta \mathbf{B}[\mathbf{E}_\theta \mathbf{x}_p] \mathbf{u}_\lambda - \bar{\mathbf{y}}_p\|^2$  for different  $\lambda$ . Following this approach, the regularization parameter  $\lambda$  is chosen as the value corresponding to the point with the maximal curvature on the curve  $(\log \|\mathbf{F}_\theta \mathbf{B}[\mathbf{E}_\theta \mathbf{x}_p] \mathbf{u}_\lambda - \bar{\mathbf{y}}_p\|, \log \|\mathbf{L}\mathbf{u}_\lambda\|)$  [15].

After the estimate of  $\mathbf{u}$  is obtained, data can be detected based on the system model (22). Using the assumptions that the noise is AWGN and independent from the channels, and the channel responses are independent from tap to tap, it is easy to verify that

$$\begin{aligned} \mathbb{E}\{\bar{\mathbf{w}}^H \bar{\mathbf{w}}\} &= \mathbb{E}\left\{\left(\sum_{k=1}^K \alpha_k \mathbf{G}_k \mathbf{v}_k + \mathbf{w}\right)^H \left(\sum_{k=1}^K \alpha_k \mathbf{G}_k \mathbf{v}_k + \mathbf{w}\right)\right\} \\ &= \mathbb{E}\left\{\sum_{k=1}^K \alpha_k^2 \sigma_{v_k}^2 \mathbf{G}_k \mathbf{G}_k^H + \sigma_w^2 \mathbf{I}_N\right\} \\ &= \left(\sum_{k=1}^K \alpha_k^2 \sigma_{v_k}^2 \sum_{l=0}^{L_k^g-1} \varsigma_{l,k}^2 + \sigma_w^2\right) \mathbf{I}_N \\ &\triangleq \sigma^2 \mathbf{I}_N. \end{aligned} \quad (27)$$

Based on (22) and (26) and treating  $\bar{\mathbf{w}}$  as the effective noise [1], [3], the approximate LMMSE solution of  $\mathbf{x}_d$  is then given by

$$\hat{\mathbf{x}}_d = \left(\Xi^H[\mathbf{u}_\lambda] \Xi[\mathbf{u}_\lambda] + \frac{\sigma^2}{\sigma_x^2} \mathbf{I}_N\right)^{-1} \Xi^H[\mathbf{u}_\lambda] (\mathbf{y} - \boldsymbol{\eta}[\mathbf{u}_\lambda]). \quad (28)$$

Finally,  $\hat{\mathbf{x}}_d = \text{Qant}[\hat{\mathbf{x}}_d]$  is a data estimator, where  $\text{Qant}[\cdot]$  denotes the hard decision of the operand in the brackets.

*Remark 1:* The complexity of the LMMSE scheme is dominated by one  $N_b L_e \times N_b L_e$  matrix inversion on  $\mathbf{B}^H[\mathbf{E}_\theta \mathbf{x}_p] \mathbf{F}_\theta^H \mathbf{F}_\theta \mathbf{B}[\mathbf{E}_\theta \mathbf{x}_p] + \lambda^2 \mathbf{L}^H \mathbf{L}$  shown in (26) and one  $N_d \times N_d$  matrix inversion on  $\Xi^H[\mathbf{u}_\lambda] \Xi[\mathbf{u}_\lambda] + \frac{\sigma^2}{\sigma_x^2} \mathbf{I}_{N_d}$  shown in (28). For an  $M \times M$  matrix, the complexity of inversion is  $O(M^3)$ . Therefore, the complexity of the LMMSE scheme is  $O((N_b L_e)^3 + N_d^3)$ .

## V. THE VARIATIONAL INFERENCE APPROACH TO ITERATIVE CHANNEL ESTIMATION AND DATA DETECTION

In the previous section, the unknown data is treated as noise (refer to (24)) and it becomes a bottleneck for channel estimation, which in turn affects the data detection performance. In this section,  $\mathbf{u}$  and  $\mathbf{x}_d$  will be jointly estimated to improve performance. Moreover, unlike the Tikhonov regularization method discussed above, we solve the ill-posed problem in a Bayesian framework. Specifically, our aim is to estimate  $\mathbf{u}$  and  $\mathbf{x}_d$ , which maximize the posterior probability density function (pdf)  $p(\mathbf{u}, \mathbf{x}_d | \mathbf{y})$ . However, the computation of  $p(\mathbf{u}, \mathbf{x}_d | \mathbf{y})$  is complicated due to the discrete nature of the data, not to mention the maximization of  $p(\mathbf{u}, \mathbf{x}_d | \mathbf{y})$  with respect to  $\mathbf{x}_d$ . To overcome this problem, we consider the variational inference approach which looks for a parameterized distribution  $Q(\mathbf{u}, \mathbf{x}_d)$  to closely represent the posterior pdf  $p(\mathbf{u}, \mathbf{x}_d | \mathbf{y})$  [17]. Once  $Q(\mathbf{u}, \mathbf{x}_d)$  is found, estimates of  $\mathbf{u}$  and  $\mathbf{x}_d$  are simply obtained by maximizing  $Q(\mathbf{u}, \mathbf{x}_d)$ .

To derive  $Q(\mathbf{u}, \mathbf{x}_d)$  closest to  $p(\mathbf{u}, \mathbf{x}_d | \mathbf{y})$ , we minimize the following *free energy* function defined as in [16]:

$$\mathbb{F} = \int_{\mathbf{u}, \mathbf{x}_d} Q(\mathbf{u}, \mathbf{x}_d) \log \frac{Q(\mathbf{u}, \mathbf{x}_d)}{p(\mathbf{u}, \mathbf{x}_d | \mathbf{y})} d\mathbf{u} d\mathbf{x}_d. \quad (29)$$

Minimizing the free energy function is equivalent to minimizing the difference between  $Q(\mathbf{u}, \mathbf{x}_d)$  and  $p(\mathbf{u}, \mathbf{x}_d | \mathbf{y})$ . A simplification can be made by factorizing  $Q(\mathbf{u}, \mathbf{x}_d)$  into a product form (also known as a mean-field approximation) [18], i.e.,  $Q(\mathbf{u}, \mathbf{x}_d) = Q_1(\mathbf{u}) Q_2(\mathbf{x}_d)$ , which is equivalent to assuming that  $\mathbf{u}$  and  $\mathbf{x}_d$  are independent conditioned on  $\mathbf{y}$ . Then a simple expression for the variational free energy in (29) is given by

$$\begin{aligned} \mathbb{F} &= \int_{\mathbf{u}, \mathbf{x}_d} Q(\mathbf{u}, \mathbf{x}_d) \log \frac{Q(\mathbf{u}, \mathbf{x}_d)}{p(\mathbf{y} | \mathbf{u}, \mathbf{x}_d) p(\mathbf{u}) p(\mathbf{x}_d)} d\mathbf{u} d\mathbf{x}_d \\ &= \int_{\mathbf{u}} Q_1(\mathbf{u}) \log Q_1(\mathbf{u}) d\mathbf{u} + \int_{\mathbf{x}_d} Q_2(\mathbf{x}_d) \log Q_2(\mathbf{x}_d) d\mathbf{x}_d \\ &\quad - \int_{\mathbf{u}} Q_1(\mathbf{u}) \log p(\mathbf{u}) d\mathbf{u} - \int_{\mathbf{x}_d} Q_2(\mathbf{x}_d) \log p(\mathbf{x}_d) d\mathbf{x}_d \\ &\quad - \int_{\mathbf{u}, \mathbf{x}_d} Q_1(\mathbf{u}) Q_2(\mathbf{x}_d) \log p(\mathbf{y} | \mathbf{u}, \mathbf{x}_d) d\mathbf{u} d\mathbf{x}_d. \end{aligned} \quad (30)$$

### A. Free Energy Function

According to (30), the computation of the free energy function requires the likelihood function  $p(\mathbf{y} | \mathbf{u}, \mathbf{x}_d)$  and the prior statistics  $p(\mathbf{u})$  and  $p(\mathbf{x}_d)$ . With  $\mathbf{x} = \mathbf{A}_\theta \mathbf{x}_d + \mathbf{E}_\theta \mathbf{x}_p$  and based on (23), the likelihood function  $p(\mathbf{y} | \mathbf{u}, \mathbf{x}_d)$  is then given by

$$p(\mathbf{y} | \mathbf{u}, \mathbf{x}_d) = \frac{1}{(\pi \sigma^2)^N} \exp \left\{ -\frac{1}{\sigma^2} \|\mathbf{y} - \mathbf{B}[\mathbf{A}_\theta \mathbf{x}_d + \mathbf{E}_\theta \mathbf{x}_p] \mathbf{u}\|^2 \right\}. \quad (31)$$

With respect to the prior statistics, we let  $p(\mathbf{x}_d)$  be complex Gaussian with zero mean and covariance matrix  $\mathbf{\Lambda}_{x_d}$ , i.e.,  $p(\mathbf{x}_d) = \mathcal{N}(0, \mathbf{\Lambda}_{x_d})$  [17], where  $\mathbf{\Lambda}_{x_d}$  is a diagonal matrix with diagonal elements depending on the average power of  $\mathbf{x}_d$ . Note that instead of defining a discrete distribution over the signal constellation, we have made a Gaussian approximation of  $p(\mathbf{x}_d)$ , which will lead to a linear detector. On the other hand, the statistics of  $\mathbf{u}$  are difficult if not impossible to derive. However, it can be shown that  $\mathbb{E}\{\mathbf{u}\} = \mathbf{0}$  and the covariance matrix  $\mathbf{R}_u$  can be obtained in closed-form (shown in Appendix B). Therefore,  $\mathbf{u}$  can be approximated as being Gaussian distributed with pdf

$$p(\mathbf{u}) = \frac{1}{\pi^{N_b L_e} |\mathbf{R}_u|} \exp \left\{ -\mathbf{u}^H \mathbf{R}_u^{-1} \mathbf{u} \right\}. \quad (32)$$

Besides specifying the likelihood function and prior statistics above, the forms of  $Q_2(\mathbf{x}_d)$  and  $Q_1(\mathbf{u})$  need to be fixed. In view of the discrete nature of the data, a close approximation is [17]

$$Q_2(\mathbf{x}_d) = \delta(\bar{\mathbf{x}}_d - \mathbf{x}_d) \quad (33)$$

where  $\delta(\cdot)$  denotes a vector Dirac delta function with the properties  $\int \delta(\bar{\mathbf{x}}_d - \mathbf{x}_d) d\mathbf{x}_d = 1$  and  $\int \delta(\bar{\mathbf{x}}_d - \mathbf{x}_d) f(\mathbf{x}_d) d\mathbf{x}_d = f(\bar{\mathbf{x}}_d)$  for any smooth function  $f(\cdot)$ . For convenience in maximization,  $Q_1(\mathbf{u})$  is chosen to be a Gaussian pdf

$$Q_1(\mathbf{u}) = \frac{1}{\pi^{N_b L_e} |\boldsymbol{\Psi}_u|} \exp \left\{ -(\mathbf{u} - \mathbf{m}_u)^H \boldsymbol{\Psi}_u^{-1} (\mathbf{u} - \mathbf{m}_u) \right\} \quad (34)$$

with  $\mathbf{m}_u$  and  $\Psi_u$  being the posterior mean and covariance matrix of  $\mathbf{u}$ , respectively.

With (31), (32), (33), (34), and  $p(\mathbf{x}_d) = \mathcal{N}(0, \mathbf{\Lambda}_{x_d})$ , the five terms in (30) can be, respectively, computed as

$$\begin{aligned} & \int_{\mathbf{u}} Q_1(\mathbf{u}) \log Q_1(\mathbf{u}) d\mathbf{h} \\ &= -N_b L_e \log \pi - \log |\Psi_u| - \mathbf{m}_u^H \Psi_u^{-1} \mathbf{m}_u \\ & \quad + 2\Re \left\{ \mathbf{m}_u^H \Psi_u^{-1} \mathbf{m}_u \right\} - \text{Tr} \left\{ \Psi_u^{-1} (\mathbf{m}_u \mathbf{m}_u^H + \Psi_u) \right\} \\ &= -N_b L_e \log \pi - \log |\Psi_u| - N_b L_e \end{aligned} \quad (35)$$

$$\int_{\mathbf{x}_d} Q_2(\mathbf{x}_d) \log Q_2(\mathbf{x}_d) d\mathbf{x}_d = 0 \quad (36)$$

$$\begin{aligned} & \int_{\mathbf{u}} Q_1(\mathbf{u}) \log p(\mathbf{u}) d\mathbf{u} \\ &= -N_b L_e \log \pi - \log |\mathbf{R}_u| - \text{Tr} \left\{ \mathbf{R}_u^{-1} (\mathbf{m}_u \mathbf{m}_u^H + \Psi_u) \right\} \end{aligned} \quad (37)$$

$$\begin{aligned} & \int_{\mathbf{x}_d} Q_2(\mathbf{x}_d) \log p(\mathbf{x}_d) d\mathbf{x}_d \\ &= -N_d \log \pi - \log |\mathbf{\Lambda}_{x_d}| - \bar{\mathbf{x}}_d^H \mathbf{\Lambda}_{x_d}^{-1} \bar{\mathbf{x}}_d \\ & \quad \int_{\mathbf{u}, \mathbf{x}_d} Q_1(\mathbf{u}) Q_2(\mathbf{x}_d) \log p(\mathbf{y} | \mathbf{u}, \mathbf{x}_d) d\mathbf{u} d\mathbf{x}_d \\ &= -N \log(\pi \sigma^2) \\ & \quad - \frac{1}{\sigma^2} (\mathbf{y}^H \mathbf{y} - 2\Re \{ \mathbf{y}^H \mathbf{B} [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \mathbf{m}_u \} \\ & \quad + \text{Tr} \{ \mathbf{B}^H [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \mathbf{B} [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \\ & \quad \times (\mathbf{m}_u \mathbf{m}_u^H + \Psi_u) \} ). \end{aligned} \quad (38)$$

Substituting (35)–(39) into (30) and dropping constant terms, we have

$$\begin{aligned} & \mathbb{F}(\mathbf{m}_u, \Psi_u, \bar{\mathbf{x}}_d) \\ & \propto -\log |\Psi_u| + \text{Tr} \left\{ \mathbf{R}_u^{-1} (\mathbf{m}_u \mathbf{m}_u^H + \Psi_u) \right\} + \bar{\mathbf{x}}_d^H \mathbf{\Lambda}_{x_d}^{-1} \bar{\mathbf{x}}_d \\ & \quad + \frac{1}{\sigma^2} (\mathbf{y}^H \mathbf{y} - 2\Re \{ \mathbf{y}^H \mathbf{B} [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \mathbf{m}_u \} \\ & \quad + \text{Tr} \{ \mathbf{B}^H [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \mathbf{B} [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \\ & \quad \times (\mathbf{m}_u \mathbf{m}_u^H + \Psi_u) \} ). \end{aligned} \quad (40)$$

Notice that, after integration, the free energy function depends only on  $\mathbf{m}_u$ ,  $\Psi_u$  and  $\bar{\mathbf{x}}_d$ .

### B. Iterative Minimization of the Free Energy Function

The remaining task is to obtain  $(\hat{\mathbf{m}}_u, \hat{\Psi}_u, \hat{\mathbf{x}}_d)$  by minimizing  $\mathbb{F}(\mathbf{m}_u, \Psi_u, \bar{\mathbf{x}}_d)$ . After that, a data estimate can be obtained by maximizing  $Q_2(\mathbf{x}_d)$  given  $\hat{\mathbf{x}}_d$ . Accordingly,  $\hat{\mathbf{x}}_d$  is an estimate of  $\mathbf{x}_d$ . Similarly, a channel estimate can be acquired by maximizing  $Q_1(\mathbf{u})$  given  $\hat{\mathbf{m}}_u$  and  $\hat{\Psi}_u$ . Notice that, since  $Q_1(\mathbf{u})$  is designed to be a complex Gaussian pdf, it is maximized at  $\mathbf{u} = \hat{\mathbf{m}}_u$ , which can be considered as a maximum *a posteriori* probability (MAP) channel estimator. For minimization of the free energy given in (40) with respect to  $(\mathbf{m}_u, \Psi_u, \bar{\mathbf{x}}_d)$ , it is found that, given  $\bar{\mathbf{x}}_d$ , there exist closed-form solutions for  $\mathbf{m}_u$  and  $\Psi_u$ . On the other hand, given  $\mathbf{m}_u$  and  $\Psi_u$ , we can derive a

closed-form solution for  $\bar{\mathbf{x}}_d$ . Therefore,  $\mathbb{F}(\mathbf{m}_u, \Psi_u, \bar{\mathbf{x}}_d)$  is minimized iteratively, starting with an initial value of  $(\mathbf{m}_u, \Psi_u, \bar{\mathbf{x}}_d)$ . The update at the  $i^{\text{th}}$  iteration follows as:

*Updating  $\mathbf{m}_u$  and  $\Psi_u$  Given  $\hat{\mathbf{x}}_d^{i-1}$ :* By setting the first-order derivative of the free energy (40) with respect to  $\mathbf{m}_u$  to zero, we have the estimate of  $\mathbf{m}_u$  as

$$\hat{\mathbf{m}}_u^i = (\mathbf{B}^H [\mathbf{A}_e \hat{\mathbf{x}}_d^{i-1} + \mathbf{E}_e \mathbf{x}_p] \mathbf{B} [\mathbf{A}_e \hat{\mathbf{x}}_d^{i-1} + \mathbf{E}_e \mathbf{x}_p] + \sigma^2 \mathbf{R}_u^{-1})^{-1} \mathbf{B}^H [\mathbf{A}_e \hat{\mathbf{x}}_d^{i-1} + \mathbf{E}_e \mathbf{x}_p] \mathbf{y}. \quad (41)$$

Similarly, by setting the first-order derivative of the free energy with respect to  $\Psi_u$  to zero, we have the estimate of  $\Psi_u$  as

$$\hat{\Psi}_u^i = \sigma^2 (\mathbf{B}^H [\mathbf{A}_e \hat{\mathbf{x}}_d^{i-1} + \mathbf{E}_e \mathbf{x}_p] \times \mathbf{B} [\mathbf{A}_e \hat{\mathbf{x}}_d^{i-1} + \mathbf{E}_e \mathbf{x}_p] + \sigma^2 \mathbf{R}_u^{-1})^{-1}. \quad (42)$$

*Updating  $\bar{\mathbf{x}}_d$  Given  $\hat{\mathbf{m}}_u^i$  and  $\hat{\Psi}_u^i$ :* Notice that the free energy given by (40) depends on  $\bar{\mathbf{x}}_d$  in a non-linear way. To obtain a closed-form solution for  $\bar{\mathbf{x}}_d$ , we first transform (40) into a linear function of  $\bar{\mathbf{x}}_d$ . Given the eigen-decomposition of  $\hat{\Psi}_u^i = \sum_{j=1}^{N_b} \beta_j^i \psi_j^i (\psi_j^i)^H$ , we have

$$\begin{aligned} & \text{Tr} \left\{ \mathbf{B}^H [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \mathbf{B} [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \hat{\Psi}_u^i \right\} \\ &= \sum_{j=1}^{N_b} \beta_j^i (\psi_j^i)^H \mathbf{B}^H [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \mathbf{B} [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \psi_j^i. \end{aligned} \quad (43)$$

Putting (43) into (40), minimizing the free energy given by (40) with respect to  $\bar{\mathbf{x}}_d$  is equivalent to minimizing

$$\begin{aligned} \tilde{\mathbb{F}}(\bar{\mathbf{x}}_d) & \propto \bar{\mathbf{x}}_d^H \mathbf{\Lambda}_{x_d}^{-1} \bar{\mathbf{x}}_d \\ & \quad + \frac{1}{\sigma^2} (-2\Re \{ \mathbf{y}^H \mathbf{B} [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \hat{\mathbf{m}}_u^i \} \\ & \quad + (\hat{\mathbf{m}}_u^i)^H \mathbf{B}^H [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \\ & \quad \times \mathbf{B} [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \hat{\mathbf{m}}_u^i + \sum_{j=1}^{N_b} \beta_j^i (\psi_j^i)^H \\ & \quad \times \mathbf{B}^H [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \\ & \quad \times \mathbf{B} [\mathbf{A}_e \bar{\mathbf{x}}_d + \mathbf{E}_e \mathbf{x}_p] \psi_j^i ). \end{aligned} \quad (44)$$

Due to the fact that the right-hand side (RHS) of (22) is equivalent to that of (23), we have  $\Xi[\mathbf{u}] \mathbf{x}_d + \boldsymbol{\eta}[\mathbf{u}] = \mathbf{B} [\mathbf{A}_e \mathbf{x}_d + \mathbf{E}_e \mathbf{x}_p] \mathbf{u}$ , which still holds when  $\mathbf{x}_d$  is replaced by  $\bar{\mathbf{x}}_d$  and  $\mathbf{u}$  is replaced by any vector having compatible dimensions with  $\Xi[\cdot]$ ,  $\boldsymbol{\eta}[\cdot]$  and  $\mathbf{B}[\cdot]$ . Therefore, (44) can be rewritten as

$$\begin{aligned} \tilde{\mathbb{F}}(\bar{\mathbf{x}}_d) & \propto \bar{\mathbf{x}}_d^H \mathbf{\Lambda}_{x_d}^{-1} \bar{\mathbf{x}}_d \\ & \quad - \frac{1}{\sigma^2} \left( 2\Re \{ \mathbf{y}^H (\Xi [\hat{\mathbf{m}}_u^i] \bar{\mathbf{x}}_d + \boldsymbol{\eta} [\hat{\mathbf{m}}_u^i]) \} \right. \\ & \quad - (\Xi [\hat{\mathbf{m}}_u^i] \bar{\mathbf{x}}_d + \boldsymbol{\eta} [\hat{\mathbf{m}}_u^i])^H \\ & \quad \times (\Xi [\hat{\mathbf{m}}_u^i] \bar{\mathbf{x}}_d + \boldsymbol{\eta} [\hat{\mathbf{m}}_u^i]) \\ & \quad - \sum_{j=1}^{N_b} \beta_j^i \left( \Xi [\psi_j^i] \bar{\mathbf{x}}_d + \boldsymbol{\eta} [\psi_j^i] \right)^H \\ & \quad \left. \times (\Xi [\psi_j^i] \bar{\mathbf{x}}_d + \boldsymbol{\eta} [\psi_j^i]) \right) \end{aligned}$$

$$\begin{aligned}
& \propto -2\Re \left\{ \left( \mathbf{y}^H \Xi [\hat{\mathbf{m}}_u^i] - \boldsymbol{\eta}^H [\hat{\mathbf{m}}_u^i] \Xi [\hat{\mathbf{m}}_u^i] \right. \right. \\
& \quad \left. \left. - \sum_{j=1}^{N_b} \beta_j^i \boldsymbol{\eta}^H [\boldsymbol{\psi}_j^i] \Xi [\boldsymbol{\psi}_j^i] \right) \bar{\mathbf{x}}_d \right\} \\
& + \bar{\mathbf{x}}_d^H \left( \Xi^H [\hat{\mathbf{m}}_u^i] \Xi [\hat{\mathbf{m}}_u^i] + \sum_{j=1}^{N_b} \beta_j^i \Xi^H [\boldsymbol{\psi}_j^i] \Xi [\boldsymbol{\psi}_j^i] \right. \\
& \quad \left. + \sigma^2 \mathbf{\Lambda}_{x_d}^{-1} \right) \bar{\mathbf{x}}_d. \tag{45}
\end{aligned}$$

Although  $\tilde{\mathbb{F}}(\bar{\mathbf{x}}_d)$  in (45) is a quadratic form of  $\bar{\mathbf{x}}_d$ , strictly speaking, minimizing  $\tilde{\mathbb{F}}(\bar{\mathbf{x}}_d)$  with respect to  $\bar{\mathbf{x}}_d$  is still a multi-dimensional search problem due to the discrete nature of  $\bar{\mathbf{x}}_d$ . To overcome this problem, we first relax  $\bar{\mathbf{x}}_d$  to be continuous, which leads to a low-complexity linear solution. By setting the first-order derivative of  $\tilde{\mathbb{F}}(\bar{\mathbf{x}}_d)$  with respect to  $\bar{\mathbf{x}}_d$  to zero, we have

$$\begin{aligned}
\tilde{\mathbf{x}}_d^i = & \left( \Xi^H [\hat{\mathbf{m}}_u^i] \Xi [\hat{\mathbf{m}}_u^i] \right. \\
& + \sum_{j=1}^{N_b} \beta_j^i \Xi^H [\boldsymbol{\psi}_j^i] \Xi [\boldsymbol{\psi}_j^i] + \sigma^2 \mathbf{\Lambda}_{x_d}^{-1} \Big)^{-1} \\
& \times \left( \Xi^H [\hat{\mathbf{m}}_u^i] \mathbf{y} - \Xi^H [\hat{\mathbf{m}}_u^i] \boldsymbol{\eta} [\hat{\mathbf{m}}_u^i] \right. \\
& \left. \left. - \sum_{j=1}^{N_b} \beta_j^i \Xi^H [\boldsymbol{\psi}_j^i] \boldsymbol{\eta} [\boldsymbol{\psi}_j^i] \right) \right). \tag{46}
\end{aligned}$$

Then constellation mapping is carried out to obtain the estimate of  $\bar{\mathbf{x}}_d$  as  $\hat{\mathbf{x}}_d^i = \text{Quant}[\tilde{\mathbf{x}}_d^i]$ .

In summary, the proposed iterative algorithm alternates among (41), (42), and (46) until  $\frac{\mathbb{F}_i - \mathbb{F}_{i-1}}{\mathbb{F}_i}$  is smaller than a threshold  $\epsilon$ . Within the framework of the variational inference approach, the proposed algorithm will approach an approximation of the desired posterior pdf. In the case that the mean field algorithm is adopted to approximate  $Q(\mathbf{u}, \mathbf{x}_d)$  as the product of  $Q_1(\mathbf{u})$  and  $Q_2(\mathbf{x}_d)$ , the convergence guarantee is retained [18]. The iterative minimization of the free energy function is then guaranteed to converge to at least a local minimum [17]. Note that an initialization is required for this iterative algorithm, and the algorithm developed in Section IV can provide one. Therefore, we set  $\hat{\mathbf{x}}_d^0 = \hat{\mathbf{x}}_d$  given in (28).

**Remark 2:** The complexity of each iteration is dominated by one  $N_b L_e \times N_b L_e$  matrix inversion on  $\mathbf{B}^H [\mathbf{A}_e \hat{\mathbf{x}}_d^{i-1} + \mathbf{E}_e \mathbf{x}_p] \mathbf{B} [\mathbf{A}_e \hat{\mathbf{x}}_d^{i-1} + \mathbf{E}_e \mathbf{x}_p] + \sigma^2 \mathbf{R}_u^{-1}$  shown in (41), an eigen-decomposition of the  $N_b L_e \times N_b L_e$  matrix  $\hat{\mathbf{\Psi}}_u^i$  and one  $N_d \times N_d$  matrix inversion on  $\Xi^H [\hat{\mathbf{m}}_u^i] \Xi [\hat{\mathbf{m}}_u^i] + \sum_{j=1}^{N_b} \beta_j^i \Xi^H [\boldsymbol{\psi}_j^i] \Xi [\boldsymbol{\psi}_j^i] + \sigma^2 \mathbf{\Lambda}_{x_d}^{-1}$  shown in (46). For an  $M \times M$  matrix, the complexity of eigen-decomposition is  $O(M^3)$ . Therefore, the complexity for each iteration of the variational inference approach is  $O(2(N_b L_e)^3 + N_d^3)$ . Each iteration of the variational inference approach has approximately the same complexity as the LMMSE scheme since it is dominated by the term  $N_d^3$ , and the overall complexity depends on the number of iterations.

## VI. SIMULATION RESULTS

In this section, computer simulation results are presented to demonstrate the performance of the proposed channel estimation and data detection algorithms. Unless otherwise specified, the simulation setting is given as follows. An OFDM AF cooperative system with one source, two relays ( $K = 2$ ) and one destination is considered. Each OFDM symbol has 128 subcarriers ( $N = 128$ ) and the length of the CP is  $L_{cp} = 8$ . The carrier frequency is  $f_c = 2$  GHz and the sampling interval is  $T_s = 2 \mu\text{s}$ . The normalized maximal Doppler shifts are set as  $N f_h T_s = 0.05$  (corresponding to a speed of 105.5 km/h) and  $N f_g T_s = 0.15$  (corresponding to a speed of 316.4 km/h), respectively. The source-relay channels have  $L_1^h = 2$  and  $L_2^h = 3$  taps, and the relay-destination channels have  $L_1^g = 3$  and  $L_2^g = 2$  taps, respectively, all following exponential power delay profiles normalized to unity. The pilot subcarriers are located on the frequencies that are multiples of  $\frac{N}{N_p}$  [8]. The pilots are generated as zero-mean complex Gaussian random variables and data are modulated by quadrature phase-shift keying (QPSK) with unit power. The average training-to-signal power ratio is defined as

$$\frac{\sum_{i \in \mathcal{Q}} \mathbb{E} \|x_p(i)\|^2}{\sum_{i=0}^{N-1} \mathbb{E} \|x(i)\|^2} = \frac{N_p}{N + (1 - \xi) N_p}.$$

The corresponding oversampling factors in the Doppler domain are  $C_h = 12$  and  $C_g = 6$ . The regularization matrix  $\mathbf{L}$  is chosen as  $\mathbf{I}_{N_b L_e}$  such that the regularization term  $\|\mathbf{I}_{N_b L_e} \mathbf{u}\|^2$  represents the energy of  $\mathbf{u}$ . The transmission power at the source is normalized to unity across different scenarios without loss of generality. For illustration, the noise powers at the relays and destination are set to be the same, i.e.,  $\sigma_v^2 \triangleq \sigma_{v_1}^2 = \sigma_{v_2}^2 = \sigma_w^2$ . The signal-to-noise ratio (SNR) is defined as  $\text{SNR} = \frac{\sigma_s^2}{\sigma_v^2}$  [1]. The unit power is equally allocated among the relays, namely  $p_1 = p_2 = \frac{1}{2}$  and  $\alpha_1 = \alpha_2 = \frac{1}{\sqrt{2(\sigma_v^2 + \sigma_v^2)}}$ . The threshold  $\epsilon$  is set to 0.0001. Each point in the following figures is obtained by averaging the results over 10 000 runs.

Fig. 1 presents the bit error rate (BER) performance of LMMSE data detection using the regularized LS channel estimator under different choices of  $\xi$  in order to investigate the effects of the PDDST. As can be seen, the data detection performance improves significantly as  $\xi$  increases from 0 to 0.8, since the interference from unknown data to training reduces as  $\xi$  increases, which benefits the channel estimation and accordingly the data detection. However, that is not the case when  $\xi$  reaches 0.9, where data transmitted on the subcarrier set  $\mathcal{Q}$  is severely distorted by training and difficult to recover. To offer a balance between limited interference for channel estimation and data integrity,  $\xi = 0.8$  is taken as an example in the following simulations.

To investigate the effects of regularization on the system performance, the performance of LMMSE data detection using unregularized/regularized LS channel estimators is presented in Fig. 2. Data detection assuming perfect channel state information is also shown for comparison. As can be seen, data detection with the unregularized LS channel estimator performs poorly. The incorporation of regularization improves the performance considerably, although there still exists a significant performance gap compared with the ideal case.

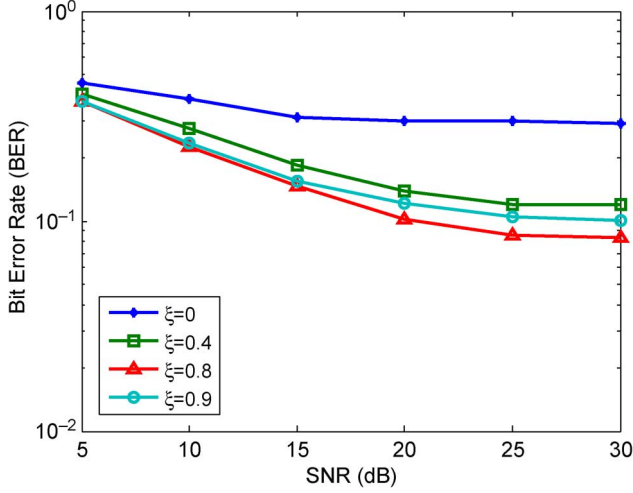


Fig. 1. Performance of LMMSE data detection using the regularized LS channel estimator for different values of  $\xi$ .

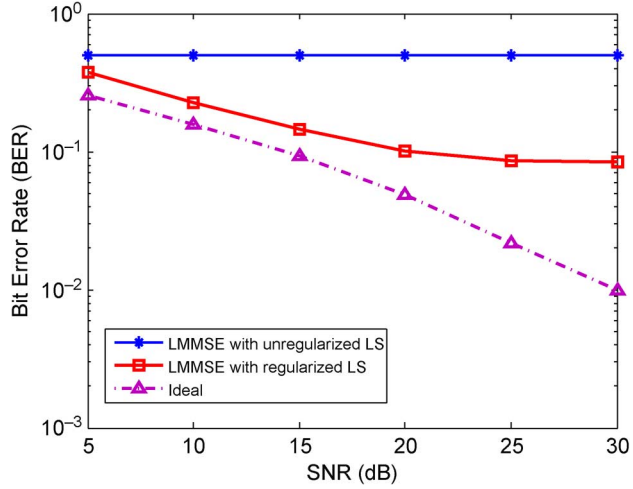


Fig. 2. Performance comparison of LMMSE data detection with unregularized and regularized LS channel estimators.

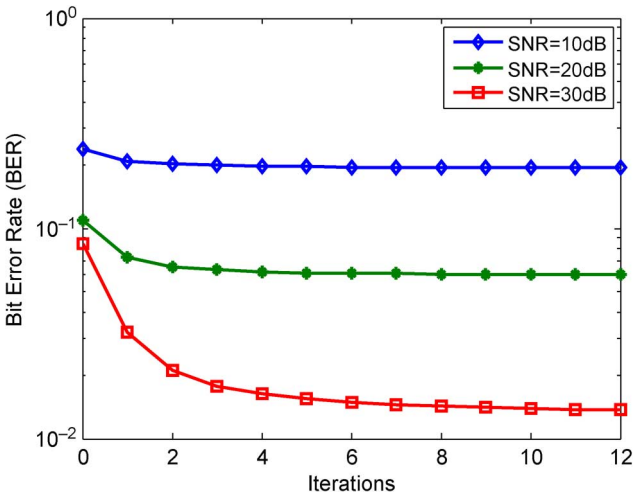


Fig. 3. Convergence of the proposed variational inference approach.

Fig. 3 shows the convergence performance of the proposed variational inference approach with SNR = 10, 20, and 30 dB. It can be seen that the BERs improve significantly in the first iteration and they converge to stable values after about eleven

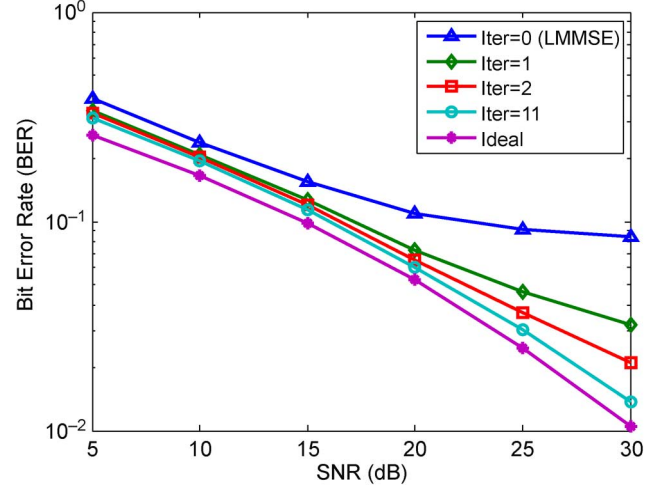


Fig. 4. Performance improvement obtained via the proposed variational inference approach.

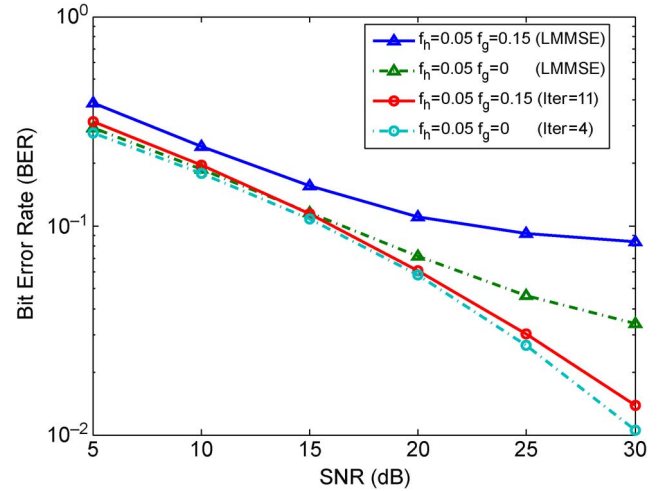


Fig. 5. Performance comparison with different Doppler spreads.

iterations for the considered SNR range. Fig. 4 further shows the performance achieved by the proposed variational inference approach versus SNR. The proposed LMMSE data detection scheme using the regularized LS channel estimator is adopted for initialization. As can be seen, the BER performance improves significantly when the number of iterations increases, and after convergence, it is very close to the ideal case. Simulations using longer channel lengths have been conducted leading to similar conclusions. In fact, the proposed LMMSE scheme with regularized LS employed for initialization is a noniterative method and significant performance gain can be further achieved by the proposed iterative algorithm.

In Fig. 5, the effects of different Doppler spreads are studied by comparing the case with  $Nf_hT_s = 0.05$  and  $Nf_gT_s = 0$  (which corresponds to the scenario in which the source is moving and the destination is static) to the case with  $Nf_hT_s = 0.05$  and  $Nf_gT_s = 0.15$ . Accordingly, we set the oversampling factors  $C_h = 12$  and  $C_g = 1$  for the former case. As can be seen, for the LMMSE data detector, the case with  $Nf_hT_s = 0.05$  and  $Nf_gT_s = 0$  performs much better than that with  $Nf_hT_s = 0.05$  and  $Nf_gT_s = 0.15$ . However, for the proposed variational inference approach, after convergence, these two cases show only

$$\mathbf{H}_k^b(m) : (N + L_k^g - 1) \times (N + L_k^c) \quad \mathbb{I}_{L_k^c} : (N + L_k^c) \times N \quad \mathbf{H}_k^b(m) \mathbb{I}_{L_k^c} : (N + L_k^g - 1) \times N$$

■ — Non-zero elements      ▨ —  $\tilde{\mathbf{H}}_k^b(m) : N \times N$

Fig. 6.  $\mathbf{H}_k^b(m) \mathbb{I}_{L_k^c} = \mathbb{I}_{L_k^g - 1} \tilde{\mathbf{H}}_k^b(m)$ .

a slight performance gap, which demonstrates the effectiveness and robustness of the proposed variational inference approach to different Doppler spreads.

## VII. CONCLUSION

In this paper, an OFDM-based AF cooperative system with one source, multiple relays and one destination has been considered with partial data-dependent superimposed training. In the first part of the paper, a closed-form channel estimator has been developed based on the LS method with Tikhonov regularization and a corresponding data detection algorithm has been proposed using the LMMSE criterion. In the second part, an iterative method based on the variational inference approach has been derived for performance improvement. Simulation results have demonstrated that, after convergence, the performance of the proposed iterative algorithm initialized by the LMMSE data detector is close to the ideal case with perfect channel state information.

## APPENDIX A PROOF OF (20)

For the matrix  $\mathbf{H}_k^b(m) \mathbb{I}_{L_k^c}$  (shown graphically in Fig. 6), the upper  $L_k^g - 1$  rows are identical to the last  $L_k^g - 1$  rows, and the lower  $N \times N$  matrix denoted by  $\tilde{\mathbf{H}}_k^b(m)$  (the cross-hatched part in Fig. 6) is a circulant matrix. Therefore

$$\mathbf{H}_k^b(m) \mathbb{I}_{L_k^c} = \mathbb{I}_{L_k^g - 1} \tilde{\mathbf{H}}_k^b(m) \quad (47)$$

as the effect of multiplying  $\tilde{\mathbf{H}}_k^b(m)$  by  $\mathbb{I}_{L_k^g - 1}$  from the left is to copy the last  $L_k^g - 1$  rows and stack them on the top of  $\tilde{\mathbf{H}}_k^b(m)$ .

Through direct computation

$$\mathbf{G}_k^b(q) \Phi(m) = \bar{\Phi}(m) \mathbf{G}_k^b(q, m) \quad (48)$$

where  $\bar{\Phi}(m) = \text{diag} \left\{ 1, e^{\frac{j2\pi m}{C_h N}}, \dots, e^{\frac{j2\pi m(N-1)}{C_h N}} \right\}$  and  $\mathbf{G}_k^b(q, m)$  has the same structure as  $\mathbf{G}_k^b(q)$  but with the first row being  $[g_k^b(q, 0), g_k^b(q, 1)e^{-\frac{j2\pi m}{C_h N}}, \dots, g_k(q, L_k^g - 1)e^{-\frac{j2\pi m(L_k^g - 1)}{C_h N}}, 0, \dots, 0]$ . Using (47) and (48), the first term on the RHS of (19) is rewritten as

$$\begin{aligned} & \alpha_k \Theta(q) \mathbf{G}_k^b(q) \Phi(m) \mathbf{H}_k^b(m) \mathbb{I}_{L_k^c} \mathbf{F}^H \mathbf{x} \\ &= \alpha_k \Theta(q) \bar{\Phi}(m) \mathbf{G}_k^b(q, m) \mathbb{I}_{L_k^g - 1} \tilde{\mathbf{H}}_k^b(m) \mathbf{F}^H \mathbf{x} \\ &= \alpha_k \Theta(q) \bar{\Phi}(m) \tilde{\mathbf{G}}_k^b(q, m) \tilde{\mathbf{H}}_k^b(m) \mathbf{F}^H \mathbf{x} \end{aligned} \quad (49)$$

where the second equality is derived similarly to (47) and  $\tilde{\mathbf{G}}_k^b(q, m)$  is a circulant matrix with the first column being  $[g_k^b(q, 0), g_k^b(q, 1)e^{-\frac{j2\pi m}{C_h N}}, \dots, g_k(q, L_k^g - 1)e^{-\frac{j2\pi m(L_k^g - 1)}{C_h N}}, 0, \dots, 0]^T$ . Accordingly, (19) is equivalent to

$$\begin{aligned} \mathbf{y} = & \sum_{k=1}^K \sum_{q=-\rho_g}^{\rho_g} \sum_{m=-\rho_h}^{\rho_h} \alpha_k \Theta(q) \bar{\Phi}(m) \\ & \times \underbrace{\tilde{\mathbf{G}}_k^b(q, m) \tilde{\mathbf{H}}_k^b(m)}_{\triangleq \mathbf{U}_k(q, m)} \mathbf{F}^H \mathbf{x} + \bar{\mathbf{w}}. \end{aligned} \quad (50)$$

Since both  $\tilde{\mathbf{G}}_k^b(q, m)$  and  $\tilde{\mathbf{H}}_k^b(m)$  are circulant matrices, the product  $\mathbf{U}_k(q, m)$  is also a circulant and is characterized by

$$\begin{aligned} \mathbf{u}_k^T(q, m) \triangleq & [g_k^b(q, 0), \dots, g_k^b(q, L_k^g - 1)e^{-j2\pi m(L_k^g - 1)/C_h N}] \\ & \otimes [h_k^b(m, 0), \dots, h_k^b(m, L_k^h - 1)]. \end{aligned} \quad (51)$$

The length of  $\mathbf{u}_k(q, m)$  is  $L_k^c + 1$ . Applying the result  $\mathbf{F} \mathbf{U}_k(q, m) \mathbf{F}^H = \text{diag}\{\mathbf{F}_{L_k^c + 1} \mathbf{u}_k(q, m)\}$  in (50) leads to

$$\begin{aligned} \mathbf{y} = & \sum_{k=1}^K \sum_{q=-\rho_g}^{\rho_g} \sum_{m=-\rho_h}^{\rho_h} \alpha_k \Theta(q) \bar{\Phi}(m) \mathbf{F}^H \\ & \times \text{diag}\{\mathbf{F}_{L_k^c + 1} \mathbf{u}_k(q, m)\} \mathbf{x} + \bar{\mathbf{w}}. \end{aligned} \quad (52)$$

Notice that  $\frac{\Theta(q) \bar{\Phi}(m)}{\text{diag}\left\{e^{\frac{j2\pi(0)(qC_h + mC_g)}{C_h C_g N}}, \dots, e^{\frac{j2\pi(N-1)(qC_h + mC_g)}{C_h C_g N}}\right\}} = 1$ . There must be integers  $\kappa$  and  $\vartheta$  with  $\kappa C_h = \vartheta C_g$  being the least common multiple of  $C_h$  and  $C_g$ . Denoting  $\Gamma(q, m) \triangleq \Theta(q) \bar{\Phi}(m) = \text{diag}\left\{e^{\frac{j2\pi(0)(\kappa m + \vartheta q)}{\kappa C_h N}}, \dots, e^{\frac{j2\pi(N-1)(\kappa m + \vartheta q)}{\kappa C_h N}}\right\}$  and combining the  $K$  relay signals together, (52) can be expressed as

$$\begin{aligned} \mathbf{y} = & \sum_{q=-\rho_g}^{\rho_g} \sum_{m=-\rho_h}^{\rho_h} \Gamma(q, m) \mathbf{F}^H \\ & \times \text{diag}\left\{ \underbrace{\sum_{k=1}^K \mathbf{F}_{L_k^c + 1} \alpha_k \mathbf{u}_k(q, m)}_{\triangleq \mathbf{F}_{L_e} \mathbf{u}(q, m)} \right\} \mathbf{x} + \bar{\mathbf{w}} \end{aligned} \quad (53)$$

where  $\mathbf{F}_{L_e}$  collects the first  $L_e = \max_{k=1}^K L_k^c + 1$  columns of  $\mathbf{F}$  and  $\mathbf{u}(q, m) = \sum_{k=1}^K \alpha_k [\mathbf{u}_k^T(q, m), \mathbf{0}_{1 \times (L_e - L_k^c - 1)}]^T$ .

## APPENDIX B

### DERIVATION OF THE MEAN AND COVARIANCE MATRIX OF $\mathbf{u}$

The received signal at the  $n$ th sampling interval is given by

$$\begin{aligned} y(n) = & \sum_{k=1}^K \sum_{l=0}^{L_k^g - 1} g_k(n, l) z_k(n - l) \\ = & \sum_{k=1}^K \sum_{l=0}^{L_k^g - 1} \sum_{q=0}^{L_k^h - 1} \alpha_k g_k(n, l) h_k(n - l, q) s(n - l - q) \end{aligned} \quad (54)$$

$$\begin{aligned}
[\mathbf{R}_{l'}]_{m,n} &= \mathbb{E} \{ \bar{h}(m, l') \bar{h}^*(n, l') \} \\
&= \mathbb{E} \left\{ \left( \sum_{k=1}^K \sum_{l=\max(0, l'-L_k^h+1)}^{\min(l', L_k^g-1)} \alpha_k g_k(m, l) h_k(m-l, l'-l) \right) \left( \sum_{k=1}^K \sum_{l=\max(0, l'-L_k^h+1)}^{\min(l', L_k^g-1)} \alpha_k g_k(n, l) h_k(n-l, l'-l) \right)^* \right\} \\
&= \sum_{k=1}^K \sum_{l=\max(0, l'-L_k^h+1)}^{\min(l', L_k^g-1)} \alpha_k^2 \mathbb{E} \{ g_k(m, l) h_k(m-l, l'-l) g_k^*(n, l) h_k^*(n-l, l'-l) \} \\
&= \sum_{k=1}^K \sum_{l=\max(0, l'-L_k^h+1)}^{\min(l', L_k^g-1)} \alpha_k^2 \zeta_l^2 \varpi_{l'-l}^2 J_0(2\pi f_g(n-m)T_s) J_0(2\pi f_h(n-m)T_s). \tag{59}
\end{aligned}$$

where the noise is ignored for simplicity, as our interest here is in the statistics of channels but not the received signals. Denoting  $l' = l + q$ , (54) can be rewritten as

$$y(n) = \sum_{k=1}^K \sum_{l'=0}^{L_k^g+L_k^h-2} \sum_{l=\max(0, l'-L_k^h+1)}^{\min(l', L_k^g-1)} \alpha_k g_k(n, l) h_k(n-l, l'-l) s(n-l'). \tag{55}$$

Noticing that  $L_e = \max_k L_k^g + L_k^h - 1$ , (55) is equivalent to

$$y(n) = \underbrace{\sum_{l'=0}^{L_e-1} \sum_{k=1}^K \sum_{l=\max(0, l'-L_k^h+1)}^{\min(l', L_k^g-1)} \alpha_k g_k(n, l) h_k(n-l, l'-l)}_{\triangleq \bar{h}(n, l')} \times s(n-l') \tag{56}$$

where  $\bar{h}(n, l')$  can be considered as the equivalent cascaded channel of multiple relays. The received OFDM symbol  $\mathbf{y} = [y(0), \dots, y(N-1)]^T$  then can be expressed as the convolution of the time domain data  $\mathbf{s}$  and the equivalent channel vector  $\bar{\mathbf{h}} = [\bar{\mathbf{h}}^T(0), \dots, \bar{\mathbf{h}}^T(L_e-1)]^T$  with  $\bar{\mathbf{h}}(l') = [\bar{h}(0, l'), \dots, \bar{h}(N-1, l')]^T$ .

Since both  $g_k(n, l)$  and  $h_k(n-l, l'-l)$  in  $\bar{h}(n, l')$  have zero means and are independent of each other, we have  $\mathbb{E}\{\bar{\mathbf{h}}\} = \mathbf{0}$ . With respect to the covariance of  $\bar{\mathbf{h}}$ , since the taps of  $g_k(n, l)$  and  $h_k(n-l, l'-l)$  are also independent, it follows that:

$$\mathbb{E} \{ \bar{\mathbf{h}}(l'_1) \bar{\mathbf{h}}^H(l'_2) \} = \mathbf{0}, \quad l'_1 \neq l'_2. \tag{57}$$

Furthermore, when  $l'_1 = l'_2 = l'$ , we have

$$\mathbb{E} \{ \bar{\mathbf{h}}(l') \bar{\mathbf{h}}^H(l') \} = \mathbf{R}_{l'} \tag{58}$$

with  $[\mathbf{R}_{l'}]_{m,n}$  given as (59), shown at the top of the page.

On the other hand, according to (23), with GCE-BEM the  $n$ th sample of the received signal is

$$y(n) = \sum_{l'=0}^{L_e-1} \sum_{j=1}^{N_b} e^{j2\pi n \gamma_j / (\kappa C_h N)} u(\gamma_j, l') s(n-l') \tag{60}$$

with  $u(\gamma_j, l')$  being the  $l'$ th element of  $\mathbf{u}(\gamma_j)$ . Comparing (56) and (60), we have  $\bar{h}(n, l') = \sum_{j=1}^{N_b} e^{\frac{j2\pi n \gamma_j}{\kappa C_h N}} u(\gamma_j, l')$  and therefore, stacking  $\bar{h}(n, l')$  into a vector gives

$$\bar{\mathbf{h}}(l') = \mathbf{W} \mathbf{u}(l') \tag{61}$$

where  $\mathbf{W}$  collects the BEM basis vectors with the  $j$ th column being  $\left[ e^{\frac{j2\pi(0)\gamma_j}{\kappa C_h N}}, \dots, e^{\frac{j2\pi(N-1)\gamma_j}{\kappa C_h N}} \right]^T$  and  $\mathbf{u}(l')$  is given by  $[u(\gamma_1, l'), \dots, u(\gamma_{N_b}, l')]^T$ . Similarly to [11], we can easily derive that

$$\mathbb{E} \{ \mathbf{u}(l') \} = \mathbf{W}^\dagger \mathbb{E} \{ \bar{\mathbf{h}}(l') \} = \mathbf{0} \tag{62}$$

and

$$\mathbb{E} \{ \mathbf{u}(l'_1) \mathbf{u}^H(l'_2) \} = \begin{cases} \mathbf{W}^\dagger \mathbf{R}_{l'} \mathbf{W}^{\dagger H} \triangleq \mathbf{R}_u(l'), & l'_1 = l'_2 = l' \\ \mathbf{0}, & l'_1 \neq l'_2. \end{cases} \tag{63}$$

With (62) and (63), the mean and covariance matrix of  $\mathbf{u}$  are, respectively, given by  $\mathbb{E}\{\mathbf{u}\} = \mathbf{0}$  and

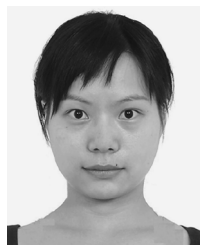
$$\mathbb{E}\{\mathbf{u} \mathbf{u}^H\} \triangleq \mathbf{R}_u \tag{64}$$

with  $[\mathbf{R}_u]_{(l':L_e:N_b L_e, l':L_e:N_b L_e)} = \mathbf{R}_u(l')$  (here Matlab notation is used).

## REFERENCES

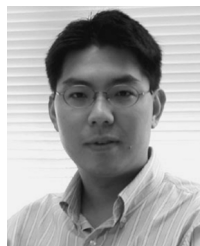
- [1] F. Gao, T. Cui, and A. Nallanathan, "On channel estimation and optimal training design for amplify and forward relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1907–1916, May 2008.
- [2] G. Wang and C. Tellambura, "Super-imposed pilot-aided channel estimation and power allocation for relay systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Budapest, Hungary, May 2009, pp. 1017–1022.
- [3] K. Yan, S. Ding, Y. Qiu, Y. Wang, and H. Liu, "A low-complexity LMMSE channel estimation method for OFDM-based cooperative diversity systems with multiple amplify-and-forward relays," *EURASIP J. Wireless Commun. Netw.*, vol. 2008, no. 149803, pp. 1–9, 2008.
- [4] Z. Zhang, W. Zhang, and C. Tellambura, "Cooperative OFDM channel estimation in the presence of frequency offsets," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3447–3458, Sep. 2009.
- [5] H. Mheidat, M. Uysal, and N. Al-Dhahir, "Equalization techniques for distributed space-time block codes with amplify-and-forward relaying," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1839–1852, May 2007.
- [6] I. Barhum, G. Leus, and M. Moonen, "Equalization for OFDM over doubly selective channels," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1445–1457, Apr. 2006.

- [7] T. Cui and C. Tellambura, "Blind receiver design for OFDM systems over doubly selective channels," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 906–917, May 2007.
- [8] S. He and J. Tugnait, "On doubly selective channel estimation using superimposed training and discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3214–3228, Jul. 2008.
- [9] E. Panayirci, H. Senol, and H. V. Poor, "Joint channel estimation, equalization, and data detection for OFDM systems in the presence of very high mobility," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4225–4238, Aug. 2010.
- [10] E. Önen, N. Odabaşioğlu, and A. Akany, "Time-frequency based channel estimation for high-mobility OFDM systems—Part II: Cooperative relaying case," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 973286, pp. 1–7, May 2010.
- [11] Z. Tang, G. Leus, R. C. Cannizzaro, and P. Banelli, "Pilot-assisted time-varying channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2226–2238, May 2007.
- [12] A. Orozco-Lugo, M. Lara, and D. McLernon, "Channel estimation using implicit training," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 240–254, Jan. 2004.
- [13] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 226–229, Mar. 2005.
- [14] T. Cui and C. Tellambura, "Superimposed pilot symbols for channel estimation in OFDM systems," in *Proc. Global Telecommun. Conf.*, St. Louis, MO, Nov./Dec. 2005, pp. 2229–2233.
- [15] P. C. Hansen, *Rank Deficient and Discrete Ill-Posed Problems—Numerical Aspects of Linear Inversion*. Philadelphia, PA: SIAM, 1998.
- [16] B. Frey, *Graphical Models for Machine Learning and Digital Communication*. Cambridge, MA: MIT Press, 1998.
- [17] D. Lin and T. Lim, "The variational inference approach to joint data detection and phase noise estimation in OFDM," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1862–1873, May 2007.
- [18] S. Haykin, J. C. Principe, T. J. Sejnowski, and J. McWhirter, *New Directions in Statistical Signal Processing: From Systems to Brains*. Cambridge, MA: MIT Press, 2005.



**Lanlan He** received the B.Sc.(Eng.) degree in 2004 and the M.Eng. degree in 2006, both from the Huazhong University of Science and Technology (HUST), Wuhan, China, in automation control and system engineering, respectively, and the Ph.D. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, in 2011.

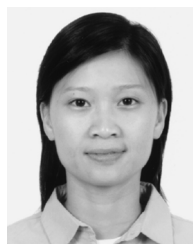
Since April 2011, she has been an ASIC system engineer with Huawei Tech. Investment Company. Her research interests include physical layer algorithms for wireless and wireline communication systems.



**Yik-Chung Wu** received the B.Eng. (EEE) degree in 1998 and the M.Phil. degree in 2001 from the University of Hong Kong (HKU). He received the Ph.D. degree from Texas A&M University, College Station, in 2005.

From August 2005 to August 2006, he was with the Thomson Corporate Research, Princeton, NJ, as a Member of Technical Staff. Since September 2006, he has been with the HKU as an Assistant Professor. He was a visiting scholar at Princeton University, Princeton, in summer 2011. His research interests

are in general area of signal processing and communication systems, and in particular, receiver algorithm design, synchronization techniques, channel estimation, and equalization.



**Shaodan Ma** received the B.Sc. (Eng.) and the M.Eng. degrees from NanKai University, Tianjin, China, in 1999 and 2002, respectively, all in electrical engineering, and the Ph.D. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, in 2006.

She was a Postdoctoral Research Fellow with the University of Hong Kong from 2006 to 2011. From March to August 2010, she was with Princeton University, Princeton, NJ, as a visiting scholar. Since August 2011, she has been with the University of Macau as an Assistant Professor. Her research interests include MIMO systems, OFDM technique, cooperative networks, cognitive radio systems, and related digital signal processing.



**Tung-Sang Ng** received the B.Sc. (Eng.) degree from The University of Hong Kong in 1972, and the M.Eng.Sc. and Ph.D. degrees from the University of Newcastle, Australia, in 1974 and 1977, respectively, all in electrical engineering.

He worked for BHP Steel International and The University of Wollongong, Australia, after graduation for 14 years before returning to The University of Hong Kong in 1991, taking up the position of Professor and Chair of Electronic Engineering. He was Head of Department of Electrical and Electronic Engineering from 2000 to 2003 and Dean of Engineering from 2003 to 2007. His current research interests include wireless communication systems, spread spectrum techniques, CDMA, and digital signal processing. He has published more than 300 international journal and conference papers.

Dr. Ng was the General Chair of ISCAS'97 and the VP-Region 10 of IEEE CAS Society in 1999 and 2000. He was an Executive Committee Member and a Board Member of the IEE Informatics Divisional Board (1999–2001) and was an ordinary member of IEE Council (1999–2001). He was awarded the Honorary Doctor of Engineering degree by the University of Newcastle, Australia, in 1997, the Senior Croucher Foundation Fellowship in 1999, the IEEE Third Millennium medal in 2000, and the Outstanding Researcher Award by The University of Hong Kong in 2003. He is a Fellow of IET and HKIE.



**H. Vincent Poor** (S'72–M'77–SM'82–F'87) received the Ph.D. degree in electrical engineering and computer science from Princeton University, Princeton, NJ, in 1977.

From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton University, where he is the Michael Henry Strater University Professor of Electrical Engineering and Dean of the School of Engineering and Applied Science. His research interests are in the areas of stochastic analysis,

statistical signal processing and their applications in wireless networks and related fields. Among his publications in these areas are the recent books *MIMO Wireless Communications* (Cambridge University Press, 2007) and *Quickest Detection* (Cambridge University Press, 2009).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, a Fellow of the American Academy of Arts and Sciences, and an International Fellow of the Royal Academy of Engineering (U.K.). He is also a Fellow of the Institute of Mathematical Statistics, the Acoustical Society of America, and other organizations. In 1990, he served as President of the IEEE Information Theory Society, and during 2004–2007, he served as the Editor-in-Chief of the IEEE TRANSACTIONS ON INFORMATION THEORY. He received a Guggenheim Fellowship in 2002 and the IEEE Education Medal in 2005. Recent recognition of his work includes the 2010 IET Ambrose Fleming Medal, the 2011 IEEE Eric E. Sumner Award, and an honorary doctorate from the University of Edinburgh, conferred in June 2011.