Title	Signal Detection for OFDM-Based Virtual MIMO Systems under Unknown Doubly Selective Channels, Multiple Interferences and Phase Noises
Author(s)	Zhong, K; Wu, YC; Li, S
Citation	IEEE Transactions on Wireless Communications, 2013, v. 12 n. 10, p. 5309-5321
Issued Date	2013
URL	http://hdl.handle.net/10722/199096
Rights	IEEE Transactions on Wireless Communications. Copyright © IEEE.

Signal Detection for OFDM-Based Virtual MIMO Systems under Unknown Doubly Selective Channels, Multiple Interferences and Phase Noises

Ke Zhong, Yik-Chung Wu, and Shaoqian Li, *Senior Member, IEEE*

***Abstract*—In this paper, the challenging problem of signal detection under severe communication environment that plagued by unknown doubly selective channels (DSCs), multiple narrowband interferences (NBIs) and phase noises (PNs) is investigated for orthogonal frequency division multiplexing based virtual multiple-input multiple-output (OFDM-V-MIMO) systems. Based on the Variational Bayesian Inference framework, a novel iterative algorithm for joint signal detection, DSC, NBI and PN estimations is proposed. Simulation results demonstrate quick convergence of the proposed algorithm, and after convergence, the bit-error-rate performance of the proposed signal detection algorithm is very close to that of the ideal case which assumes perfect channel state information, no PN, and known positions and powers of NBIs plus additive white Gaussian noise. Furthermore, simulation results show that the proposed signal detection algorithm outperforms other state-of-the-art methods.**

***Index Terms*—Signal detection, virtual multiple-input multiple-output (V-MIMO), orthogonal frequency division multiplexing (OFDM), phase noise (PN), doubly selective channel (DSC), narrowband interference (NBI), Variational Bayesian Inference (VBI).**

I. Introduction

MULTIPLE-input multiple-output (MIMO) communication [1] has emerged as an important technology for boosting spectrum efficiency and system capacity. On the other hand, orthogonal frequency division multiplexing (OFDM) [2] provides high-data-rate transmission capability and is robustness to frequency-selective fading. Therefore, MIMO combined with OFDM (MIMO-OFDM) is a promising access scheme for modern communication systems [3]. However, physical implementation of multiple antennas at a small mobile terminal (MT) may not be feasible due to cost, size, power and complexity constraints. To overcome this problem, virtual MIMO (V-MIMO) [4], [5] is an attractive alternative, where multiple single-antenna MTs independently or cooperatively transmit signals to the multiple-antenna base station (BS). In fact, OFDM-based V-MIMO (OFDM-V-MIMO) has been incorporated in recent wireless standards, such as IEEE 802.11e [6] and 3GPP-LTE [7].

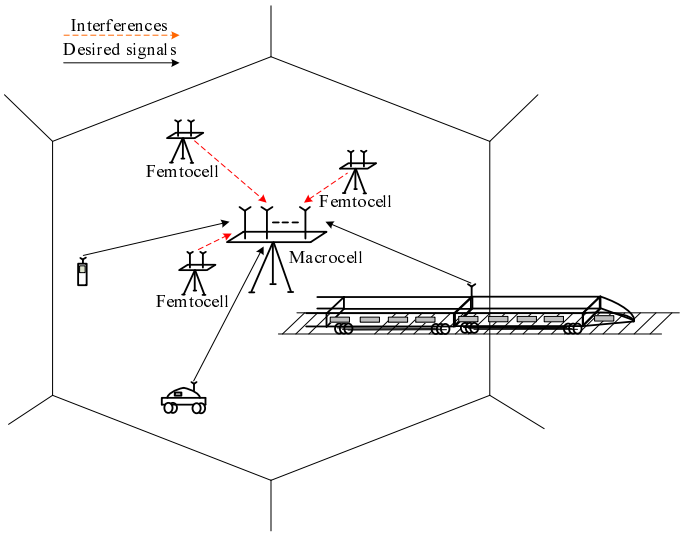


Fig. 1. A V-MIMO system under unknown DSCs, multiple interferences and PNs.

Although with many attractive features, there are a lot of challenges faced by OFDM-V-MIMO systems in practical scenarios (depicted in Fig. 1):

1) Due to cost constraint, low-cost oscillators are usually used in MTs, which results in phase noise (PN). It is known that the performance of OFDM systems is very sensitive to PN due to the induced common phase error (CPE) and inter-carrier interference (ICI) [8], [9]. Making the situation even more challenging is the fact that in V-MIMO, each MT has its own oscillator, leading to received signal at BS corrupted by multiple PNs.
2) In emerging communication scenarios, such as femtocell [10] and cognitive radio systems [11], spectrum sharing is common. In these systems, sophisticated measures for detecting and mitigating multiple narrowband interferences (NBIs) must be employed, otherwise risking significant degradation of system performance [12].
3) Requesting high-data-rate wireless communication while under high-speed movement is one of the key features

Manuscript received February 9, 2013; revised June 6, 2013; accepted July 21, 2013. The associate editor coordinating the review of this paper and approving it for publication was J. Coon.

This work was supported in part by the General Research Fund (GRF) from the Hong Kong Research Grant Council (Project No. HKU 7191/11E), the National Science Foundation of China under grant 61032002, the Chinese Important National Science & Technology Specific Projects under grant 2011ZX03001-007-01, the Program for New Century Excellent Talents in University, NCET-11-0058, and the National High-Tech R&D Program of China ("863" Project under Grant 2011AA01A105).

K. Zhong and S. Li are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu, 611731, P. R. China (e-mail: kezhong@eee.hku.hk; lsq@uestc.edu.cn).

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

Digital Object Identifier 10.1109/TWC.2013.090413.130271

in modern communications. In fact, current and future wireless communication systems, such as 3GPP-LTE and WiMAX, are required to provide support for high-mobility users at speeds up to 350km/h. High mobility together with high data rate result in time and frequency selectivities in channel (i.e., doubly selective channel, DSC), which significantly complicate channel estimation and signal detection in OFDM and multiple-antenna systems [3], [13].

Any one of the above challenges calls for extensive research, let alone all three together. In fact, previous works on related topics only consider subsets of the problems mentioned above. In particular, assuming the channel is perfectly known, signal detection and PN cancelation was investigated under time-invariant frequency-selective channel (TIFSC) in [14], [15] and DSC in [16], respectively. For static channel estimation under PN, it was investigated in [17] using Maximum A Posteriori (MAP) estimator, and in [18] using sequential Monte Carlo (SMC) based algorithm. To deal with NBIs, expectation-maximization (EM) algorithms were developed to jointly estimate TIFSC and NBI in [19], and to detect data under NBI in [20], respectively. On the other hand, joint channel estimation and signal detection was investigated for OFDM systems under static channel [21], [22] and under DSC [23], respectively. For data detection under unknown TIFSC and PN, it was addressed in [24] based on pilot-aided estimator and decision-feedback scheme, and in [25] using Markov Chain Monte Carlo (MCMC). Finally, joint signal detection and channel tracking in the presence of PN was addressed in [26]. However, the channel is assumed to be quasi-static for the duration of an OFDM symbol.

In this paper, we consider all three challenges together. More specifically, signal detection in the uplink of OFDM-V-MIMO systems operating over DSCs in the presence of multiple unknown NBIs and PNs is investigated. In this severe scenario, besides the unknown interferences caused by NBIs and CPEs caused by PNs, the combined ICIs from both DSCs and PNs will make signal detection extremely difficult. Moreover, we consider the worst case where data, DSCs, NBIs and PNs change from one OFDM symbol to another, which means that there is no correlation among adjacent OFDM symbols can be utilized, and only the information from the current OFDM symbol can be used. As is well known, the optimal signal detection in this scenario is to find the maximum value of the marginalized posterior distribution of the data signal. However, finding exact expression for this posterior distribution leads to intractable integrals [27]. In this paper, we employ the Variational Bayesian Inference (VBI) [28], [29] approach to develop an iterative algorithm to solve this challenging problem. VBI is a systematic framework for approximating intractable integrals arising in Bayesian statistics and machine learning, and has recently been successfully applied to tackle subsets of challenges mentioned in this paper [15], [21]-[23], [26]. Simulation results show that after convergence, the proposed signal detection algorithm performs very close to that of the ideal case which assumes perfect channel state information (CSI), no PN, and known positions and powers of NBIs plus additive white Gaussian noise (AWGN). Furthermore, it is shown that the proposed signal detection algorithm outperforms other state-of-the-art methods.

The organization of this paper is as follows. The system model for OFDM-V-MIMO systems operating over DSCs with multiple NBIs and PNs is introduced in Section II. In Section III, based on VBI, an iterative algorithm for joint signal detection, DSC, NBI and PN estimations in such severe communication environment is proposed. Extensive simulation results are given in Section IV to demonstrate the effectiveness of the proposed algorithm. Finally, conclusions are drawn in Section V.

Notation: Matrices and vectors are represented by boldface uppercase and lowercase letters, respectively. A hat over a variable (e.g., $\hat{x}$) indicates an estimate of the variable. $\mathbb{E}\{\cdot\}$ denotes the expectation. Superscripts $[\cdot]^T$, $[\cdot]^{-1}$, $[\cdot]^H$ and $[\cdot]^*$ denote the transpose, the matrix inversion, the Hermitian and the complex conjugate operations, respectively. $\mathbf{I}_N$ is an identity matrix with dimension N. diag$\{\mathbf{x}\}$ stands for the diagonal matrix with vector $\mathbf{x}$ on its diagonal. Bldiag$\{\cdot\}$ denotes the block diagonal concatenation of input arguments. The symbol $\otimes$ denotes the Kronecker product. Tr$\{\mathbf{X}\}$ and det$(\mathbf{X})$ are the trace and the determinant of a square matrix $\mathbf{X}$, respectively. $|\cdot|$ denotes the modulus of a complex number. $\Re\{\cdot\}$ denotes the real part of the quantity in the bracket. The matrix $\mathbf{F}$ is the normalized fast Fourier transform (FFT) matrix with the $(m,n)^{th}$ element given by $\frac{1}{\sqrt{N}}e^{-\jmath 2\pi mn/N}$, where $\jmath \triangleq \sqrt{-1}$. $\propto$ means equal up to a normalizing constant. $\mathbf{0}_m$ and $\mathbf{1}_m$ denote the $m\times 1$ all-zero and all-one column vectors, respectively. $\delta(\cdot)$ denotes the Dirac delta function. The k^{th} element of vector $\mathbf{x}$ is denoted by $\mathbf{x}(k)$.

II. System Model

A. Received OFDM-V-MIMO Signal

We consider an uplink OFDM-V-MIMO system with N_R receive antennas at the BS and N_T ($N_R \geq N_T$) selected MTs with each MT equipped with one transmit antenna (for user selection in V-MIMO, see [30], [31] and references therein). Each MT independently transmits its own OFDM symbols to the BS. For the i^{th} MT, the time domain signal $\mathbf{s}^i = [s^i(0), ..., s^i(N-1)]^T$ is generated by taking the N-point inverse FFT (IFFT) of the source signal in the frequency domain $\mathbf{x}^i = [x^i(0), ..., x^i(N-1)]^T$ as $\mathbf{s}^i = \mathbf{F}^H\mathbf{x}^i$. In general, the elements of $\mathbf{x}^i$ can be categorized into

$$x^i(k) = \begin{cases} x_d^i(k) & \forall \quad k \in \mathfrak{I}_d^i \\ x_p^i(k) & \forall \quad k \in \mathfrak{I}_p^i \end{cases} \tag{1}$$

where $\mathfrak{I}_d^i$ is the index set of subcarriers allocated for data symbols (with N_d elements) and $\mathfrak{I}_p^i$ is the index set of subcarriers allocated for pilot symbols (with N_p elements), respectively. Notice that $N = N_d + N_p$. From (1), we can write $\mathbf{x}^i = \mathbf{E}_d^i\mathbf{x}_d^i + \mathbf{E}_p^i\mathbf{x}_p^i$, where $\mathbf{E}_d^i$ and $\mathbf{E}_p^i$ denote the matrices collecting columns of $\mathbf{I}_N$ corresponding to $\mathfrak{I}_d^i$ and $\mathfrak{I}_p^i$, respectively, and $\mathbf{x}_d^i = [x_d^i(0), ..., x_d^i(N_d-1)]^T$, $\mathbf{x}_p^i = [x_p^i(0), ..., x_p^i(N_p-1)]^T$ denote the data and pilot vectors, respectively. A cyclic prefix (CP) with length N_{cp} longer than the maximum delay spread of channels between MTs and BS is inserted at the beginning of each OFDM symbol to eliminate intersymbol interference (ISI).

At the j^{th} receive antenna of the BS, assuming perfect timing and frequency synchronization are achieved, the received signal for a whole OFDM symbol, after discarding the CP and transforming back into the frequency domain, is given by

$$\mathbf{y}^j = \sum_{i=0}^{N_T-1} \mathbf{F}\mathbf{H}^{ji}\mathbf{P}^i\mathbf{s}^i + \boldsymbol{\varsigma}^j + \mathbf{v}^j, \tag{2}$$

where the symbols in (2) are defined as:

• $\mathbf{H}^{ji}$ represents the DSC matrix between the i^{th} MT and the j^{th} receive antenna at the BS and is given by Eq. (3) at the top of the next page, where $h^{ji}(n,l)$ denotes the corresponding DSC of the l^{th} path at time n, and L^{ji} represents the corresponding channel length. Each channel path is assumed to obey the classical Jakes model [32].

• The PN process for the i^{th} MT is represented by

$$\mathbf{P}^i = \text{diag}\{\mathbf{e}^{\jmath\boldsymbol{\phi}^i}\}, \tag{4}$$

where $\mathbf{e}^{\jmath\boldsymbol{\phi}^i} = [e^{\jmath\phi^i(0)}, ..., e^{\jmath\phi^i(N-1)}]^T$ with $\boldsymbol{\phi}^i = [\phi^i(0), ..., \phi^i(N-1)]^T$ representing the discrete-time PN sequence.

• $\boldsymbol{\varsigma}^j = [\varsigma^j(0), ..., \varsigma^j(N-1)]^T$ and $\mathbf{v}^j = [v^j(0), ..., v^j(N-1)]^T$ denote the disturbance terms that account for the NBI and AWGN at the j^{th} receive antenna of the BS, respectively. Elements of $\mathbf{v}^j$ are zero-mean Gaussian distributed with unknown variance σ_v^2. Furthermore, following [19], [20], $\varsigma^j(k)$ is modeled as a complex Gaussian random variable with zero mean and unknown power $\sigma_{\varsigma,j}^2(k)$. The multiple NBIs are assumed to be mutually independent which can be regarded as the worst-case scenario. This is because if correlation exists between NBIs, this correlation can be profitably exploited to improve the system performance against NBI. Notice that the NBI is not the ICI due to PNs and DSCs. ICI has been explicitly modeled in system model (2).

Notice that $\mathbf{H}^{ji}$ in (3) can be decomposed as a sum of L^{ji} terms as

$$\mathbf{H}^{ji} = \sum_{l=0}^{L^{ji}-1} \text{diag}\{\mathbf{h}_l^{ji}\}\mathbf{A}_l, \tag{5}$$

where $\mathbf{h}_l^{ji} = [h^{ji}(0,l), ..., h^{ji}(N-1,l)]^T$ is the l^{th} path channel between the i^{th} MT and the j^{th} receive antenna at the BS, and the $N \times N$ matrix $\mathbf{A}_l$ is a permutation matrix obtained from cyclically shifting the columns of the identity matrix $\mathbf{I}_N$ to the left by l positions. Substituting (4) and (5) into (2), we obtain

$$\mathbf{y}^j = \sum_{i=0}^{N_T-1}\sum_{l=0}^{L^{ji}-1} \mathbf{F}\text{diag}\{\mathbf{h}_l^{ji}\}\mathbf{A}_l\text{diag}\{\mathbf{e}^{\jmath\boldsymbol{\phi}^i}\}\mathbf{s}^i + \mathbf{w}^j, \tag{6}$$

where $\mathbf{w}^j \triangleq \boldsymbol{\varsigma}^j + \mathbf{v}^j$ with the k^{th} element $w^j(k)$ being a complex Gaussian random variable with zero mean and unknown variance $\sigma_j^2(k) \triangleq \sigma_{\varsigma,j}^2(k) + \sigma_v^2$.

B. Reformulation With Basis Expansion Model

From (5) it is observed that the number of unknowns in $\mathbf{H}^{ji}$ is NL^{ji}, which is much larger than the number of received samples. This leads to identifiability problem for direct channel estimation. However, considering the fact that time correlation of channel exists during one OFDM symbol period, the basis expansion model (BEM) [33], [34] can be adopted to represent the channels as

$$h^{ji}(n,l) = \sum_{q=0}^{Q} \beta_{q,l}^{ji} b_{n,q} + \xi^{ji}(n,l), \tag{7}$$

where $\beta_{q,l}^{ji}$ is the coefficient of BEM associated with the channel between the i^{th} MT and the j^{th} receive antenna at the BS; $b_{n,q}$ represents the basis that captures time variations of channel; $Q+1$ is the number of the basis and $\xi^{ji}(n,l)$ represents the BEM modeling error. BEM is motivated by the observation that the temporal (n) variation of $h^{ji}(n,l)$ is usually rather smooth due to the low-pass nature of the channel's band-limited Doppler spread, and therefore, $\{b_{n,q}\}_{q=0}^{Q}$ can be chosen as a small set (i.e., $Q \ll N$) of smooth functions. Motivated by the fact that the BEM modeling error is usually on the order of 10^{-4} [34], [35], which is much smaller than the typical received noise level, using (7) the $\mathbf{h}_l^{ji}$ in (5) can be expressed as

$$\mathbf{h}_l^{ji} = \mathbf{B}\boldsymbol{\beta}_l^{ji}, \tag{8}$$

where $\mathbf{B} = [\mathbf{b}_0, ..., \mathbf{b}_Q]$ with $\mathbf{b}_q = [b_{0,q}, ..., b_{N-1,q}]^T$ and $\boldsymbol{\beta}_l^{ji} = [\beta_{0,l}^{ji}, ..., \beta_{Q,l}^{ji}]^T$. It is observed from (8) that the number of channel parameters to be estimated is significantly reduced.

Substituting (8) into (6), we obtain a BEM-reformulated expression as

$$\mathbf{y}^j = \boldsymbol{\Gamma}[\boldsymbol{\beta}^j, \boldsymbol{\phi}]\mathbf{s} + \mathbf{w}^j, \tag{9}$$

where $\boldsymbol{\Gamma}[\boldsymbol{\beta}^j, \boldsymbol{\phi}] \triangleq [\sum_{l=0}^{L^{j0}-1} \mathbf{F}\text{diag}\{\mathbf{B}\boldsymbol{\beta}_l^{j0}\}\mathbf{A}_l\text{diag}\{\mathbf{e}^{\jmath\boldsymbol{\phi}^0}\}, ..., \sum_{l=0}^{L^{j,N_T-1}} \mathbf{F}\text{diag}\{\mathbf{B}\boldsymbol{\beta}_l^{j,N_T-1}\}\mathbf{A}_l\text{diag}\{\mathbf{e}^{\jmath\boldsymbol{\phi}^{N_T-1}}\}]$ with $\boldsymbol{\beta}^j = [(\boldsymbol{\beta}^{j0})^T, ..., (\boldsymbol{\beta}^{j,N_T-1})^T]^T$ and $\boldsymbol{\beta}^{ji} = [(\boldsymbol{\beta}_0^{ji})^T, ..., (\boldsymbol{\beta}_{L^{ji}-1}^{ji})^T]^T$, $\boldsymbol{\phi} = [(\boldsymbol{\phi}^0)^T, ..., (\boldsymbol{\phi}^{N_T-1})^T]^T$ is the PN vector and $\mathbf{s} = [(\mathbf{s}^0)^T, ..., (\mathbf{s}^{N_T-1})^T]^T$. Stacking all the received signals from N_R receive antennas at the BS, a BEM-reformulated expression that explicitly shows the dependence of the unknown data $\mathbf{x}_d$ can be obtained as

$$\mathbf{y} = \boldsymbol{\Theta}[\boldsymbol{\beta}, \boldsymbol{\phi}]\mathbf{x} + \mathbf{w}, \tag{10}$$

where $\mathbf{y} = [(\mathbf{y}^0)^T, ..., (\mathbf{y}^{N_R-1})^T]^T$, $\boldsymbol{\Theta}[\boldsymbol{\beta}, \boldsymbol{\phi}] \triangleq [\boldsymbol{\Gamma}^H[\boldsymbol{\beta}^0, \boldsymbol{\phi}], ..., \boldsymbol{\Gamma}^H[\boldsymbol{\beta}^{N_R-1}, \boldsymbol{\phi}]]^H$, $\boldsymbol{\beta} = [(\boldsymbol{\beta}^0)^T, ..., (\boldsymbol{\beta}^{N_R-1})^T]^T$, $\mathbf{w} = [(\mathbf{w}^0)^T, ..., (\mathbf{w}^{N_R-1})^T]^T$ and

$$\begin{aligned}\mathbf{x} &= [(\mathbf{x}^0)^T, ..., (\mathbf{x}^{N_T-1})^T]^T \\ &= \text{Bldiag}\{\mathbf{F}^H\mathbf{E}_d^0, ..., \mathbf{F}^H\mathbf{E}_d^{N_T-1}\}\mathbf{x}_d \\ &\quad + \text{Bldiag}\{\mathbf{F}^H\mathbf{E}_p^0, ..., \mathbf{F}^H\mathbf{E}_p^{N_T-1}\}\mathbf{x}_p \\ &\triangleq \mathbf{E}_d\mathbf{x}_d + \mathbf{E}_p\mathbf{x}_p,\end{aligned} \tag{11}$$

where $\mathbf{x}_d = [(\mathbf{x}_d^0)^T, ..., (\mathbf{x}_d^{N_T-1})^T]^T$ and $\mathbf{x}_p = [(\mathbf{x}_p^0)^T, ..., (\mathbf{x}_p^{N_T-1})^T]^T$.

On the other hand, substituting (8) into (6), another BEM-reformulated expression can also be obtained as

$$\mathbf{y}^j = \mathbf{G}^j[\boldsymbol{\phi}, \mathbf{x}]\mathbf{M}^j\boldsymbol{\beta}^j + \mathbf{w}^j, \tag{12}$$

where $\mathbf{G}^j[\boldsymbol{\phi}, \mathbf{x}] \triangleq [\boldsymbol{\varrho}^j[\boldsymbol{\phi}^0, \mathbf{x}^0], ..., [\boldsymbol{\varrho}^j[\boldsymbol{\phi}^{N_T-1}, \mathbf{x}^{N_T-1}]]$ with $\boldsymbol{\varrho}^j[\boldsymbol{\phi}^i, \mathbf{x}^i] \triangleq [\mathbf{F}\text{diag}\{\mathbf{A}_0\text{diag}\{\mathbf{e}^{\jmath\boldsymbol{\phi}^i}\}\mathbf{F}^H\mathbf{x}^i\}, ..., \mathbf{F}\text{diag}\{\mathbf{A}_{L^{ji}-1}\text{diag}\{\mathbf{e}^{\jmath\boldsymbol{\phi}^i}\}\mathbf{F}^H\mathbf{x}^i\}]$, $\mathbf{M}^j \triangleq \text{Bldiag}\{\mathbf{I}_{L^{j0}} \otimes \mathbf{B}, ..., \mathbf{I}_{L^{j,N_T-1}} \otimes \mathbf{B}\}$. Stacking all the received signals from N_R receive antennas at

$$\mathbf{H}^{ji}=\begin{bmatrix} h^{ji}(0,0) & \mathbf{0}\dots & h^{ji}(0,L^{ji}-1)\dots & h^{ji}(0,1) \\ \vdots & \vdots & \ddots & \vdots \\ h^{ji}(L^{ji}-1,L^{ji}-1) & h^{ji}(L^{ji}-1,L^{ji}-2)\dots & h^{ji}(L^{ji}-1,0) & \mathbf{0}\dots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}\dots & h^{ji}(N-1,L^{ji}-1) & h^{ji}(N-1,L^{ji}-2)\dots & h^{ji}(N-1,0) \end{bmatrix}. \quad (3)$$

the BS, an equation that explicitly shows the dependence of the unknown parameter $\boldsymbol{\beta}$ can be obtained as

$$\mathbf{y}=\boldsymbol{\Xi}[\boldsymbol{\phi},\mathbf{x}]\boldsymbol{\beta}+\mathbf{w}, \quad (13)$$

where $\boldsymbol{\Xi}[\boldsymbol{\phi},\mathbf{x}]=\text{Bldiag}\{\mathbf{G}^0[\boldsymbol{\phi},\mathbf{x}]\mathbf{M}^0,...,\mathbf{G}^{N_R-1}[\boldsymbol{\phi},\mathbf{x}]\mathbf{M}^{N_R-1}\}$.

C. PN Modeling

In this paper, the fact that multiple MTs supported by their own oscillators are reflected by the distinct PN process $\mathbf{P}^i$ for each i. On the other hand, the oscillator used at the BS is usually sufficiently stable to disregard its PN. Since PN evolves much slower than the modulation rate, (6) can be expressed alternatively using the small PN approximation [15], [36] $\exp\{\jmath\phi^i(n)\}\approx 1+\jmath\phi^i(n)$ as

$$\mathbf{y}^j=\sum_{i=0}^{N_T-1}\sum_{l=0}^{L^{ji}-1}\mathbf{F}\text{diag}\{\mathbf{h}_l^{ji}\}\mathbf{A}_l\text{diag}\{\mathbf{s}^i\}(\mathbf{1}_N+\jmath\boldsymbol{\phi}^i)+\mathbf{w}^j. \quad (14)$$

Substituting (8) into (14) and stacking all the received signals from N_R receive antennas at the BS, an equation that explicitly shows the dependence of the unknown parameter $\boldsymbol{\phi}$ can be obtained as

$$\mathbf{y}=\boldsymbol{\Omega}[\boldsymbol{\beta},\mathbf{x}](\mathbf{1}_{N_TN}+\jmath\boldsymbol{\phi})+\mathbf{w}, \quad (15)$$

where $\boldsymbol{\Omega}[\boldsymbol{\beta},\mathbf{x}]\triangleq\text{Bldiag}\{\boldsymbol{\Upsilon}[\boldsymbol{\beta}^0,\mathbf{x}],...,\boldsymbol{\Upsilon}[\boldsymbol{\beta}^{N_R-1},\mathbf{x}]\}(\mathbf{1}_{N_R}\otimes\mathbf{I}_{N_TN})$ with $\boldsymbol{\Upsilon}[\boldsymbol{\beta}^j,\mathbf{x}]\triangleq[\sum_{l=0}^{L^{j0}-1}\mathbf{F}\text{diag}\{\mathbf{B}\boldsymbol{\beta}_l^{j0}\}\mathbf{A}_l\text{diag}\{\mathbf{F}^H\mathbf{x}^0\},...,\sum_{l=0}^{L^{j,N_T-1}}\mathbf{F}\text{diag}\{\mathbf{B}\boldsymbol{\beta}_l^{j,N_T-1}\}\mathbf{A}_l\text{diag}\{\mathbf{F}^H\mathbf{x}^{N_T-1}\}]$.

Remark 1: We have developed three equivalent models in (10), (13) and (15) for the received signal $\mathbf{y}$ with dimension N_RN. They all depend on the unknown parameters $\mathbf{x}_d$ with dimension N_TN_d (embedded in $\mathbf{x}$ with dimension N_TN), $\boldsymbol{\beta}$ with dimension $(Q+1)\sum_{j=0}^{N_R-1}\sum_{i=0}^{N_T-1}L^{ji}$, and $\boldsymbol{\phi}$ with dimension N_TN. Their usages will be apparent in later sections.

D. Prior Distributions for the Unknown Quantities

In the Bayesian framework, all the unknown quantities are assumed to be random variables, with each described by a prior distribution. It is assumed that the channel follows Rayleigh fading and each channel tap is independent. The BEM coefficient $\boldsymbol{\beta}$ can be shown to be complex Gaussian variable [33] with zero mean and correlation matrix $\mathbf{R}_{\boldsymbol{\beta}}$:

$$p(\boldsymbol{\beta})=\frac{1}{\pi^{(Q+1)L}\det(\mathbf{R}_{\boldsymbol{\beta}})}\exp\{-\boldsymbol{\beta}^H\mathbf{R}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}\}, \quad (16)$$

where $L\triangleq\sum_{j=0}^{N_R-1}\sum_{i=0}^{N_T-1}L^{ji}$ represents the sum of all the channel lengths. The expression for $\mathbf{R}_{\boldsymbol{\beta}}$ is derived in Appendix A and is given by Eqs. (43) and (45). In case we do not have the statistical information of channel, we can set $\mathbf{R}_{\boldsymbol{\beta}}=\infty\mathbf{I}_{(Q+1)L}$, which is an uninformative prior.

For PN $\boldsymbol{\phi}$, two models are adopted in the literature. 1) When the system is frequency-locked [8], the resulting PN is modeled as a zero-mean, nonstationary and infinite-power Wiener process. 2) When the system is phase-locked [9], the resulting PN is modeled as a zero-mean, stationary and finite-power stochastic process. Since in both cases $\boldsymbol{\phi}$ is Gaussian distributed with zero mean, the prior distribution for $\boldsymbol{\phi}$ is given by

$$p(\boldsymbol{\phi})=\frac{1}{(2\pi)^{\frac{N_TN}{2}}\det(\mathbf{R}_{\boldsymbol{\phi}})^{\frac{1}{2}}}\exp\{-\frac{1}{2}\boldsymbol{\phi}^T\mathbf{R}_{\boldsymbol{\phi}}^{-1}\boldsymbol{\phi}\}, \quad (17)$$

where $\mathbf{R}_{\boldsymbol{\phi}}=\text{Bldiag}\{\mathbf{R}_{\boldsymbol{\phi}^0},...,\mathbf{R}_{\boldsymbol{\phi}^{N_T-1}}\}$ with the specific expression for $\mathbf{R}_{\boldsymbol{\phi}^i}$ determined by the PN model and is documented in [8], [9].

For the prior distribution of the data $\mathbf{x}_d$, we set equal preference to all constellation points since we do not have knowledge on its value before observing the received signal. Furthermore, due to the independent property among data elements, we have

$$p(\mathbf{x}_d)=\frac{1}{(\mathfrak{m})^{N_TN_d}}\prod_{i=0}^{N_T-1}\prod_{k=0}^{N_d-1}\Big\{\sum_{\bar{x}_d^i(k)\in\mathfrak{c}}\delta\big(x_d^i(k)-\bar{x}_d^i(k)\big)\Big\}, \quad (18)$$

where $\mathfrak{m}$ is the modulation order and $\mathfrak{c}$ is the set of constellation points of the modulation.

For the unknown interference plus noise power at the j^{th} receive antenna and k^{th} subcarrier $\sigma_j^2(k)$, an inverse Gamma prior distribution with shape parameter $a_{k,j}$ and scale parameter $b_{k,j}$ is adopted. With the fact that $\sigma_j^2(k)$ are independent for different j and k, we have

$$p(\boldsymbol{\sigma}^2)=\prod_{j=0}^{N_R-1}\prod_{k=0}^{N-1}\frac{b_{k,j}^{a_{k,j}}}{\Gamma(a_{k,j})}\big(\sigma_j^2(k)\big)^{-a_{k,j}-1}\exp\big(-\frac{b_{k,j}}{\sigma_j^2(k)}\big), \quad (19)$$

where $\Gamma(\cdot)$ denotes the gamma function and the $N_RN\times 1$ interference plus noise power vector $\boldsymbol{\sigma}^2=[(\boldsymbol{\sigma}_0^2)^T,...,(\boldsymbol{\sigma}_{N_R-1}^2)^T]^T$ with $\boldsymbol{\sigma}_j^2=[\sigma_j^2(0),...,\sigma_j^2(N-1)]^T$. The inverse Gamma distribution has been adopted to describe the distribution of inter-cell interference in severely fading channels [37] and the power of interference in radar systems [38]. In the absence of prior information, small values of the shape parameter $a_{k,j}$ and the scale parameter $b_{k,j}$ can be chosen (e.g., set $a_{k,j}=b_{k,j}=10^{-6}$) to produce uninformative prior for $\boldsymbol{\sigma}^2$.

III. Iterative Signal Detection with Unknown DSCs, NBIs and PNs

It is observed from (10) and (11) that we aim to detect $\mathbf{x}_d$ in the presence of unknown parameters $\boldsymbol{\beta},\boldsymbol{\phi}$ and unknown

variance $\boldsymbol{\sigma}^2$ of $\mathbf{w}$. The optimal estimate of $\mathbf{x}_d$ is to maximize the marginalized posterior distribution $p(\mathbf{x}_d|\mathbf{y})$ which is given by

$$p(\mathbf{x}_d|\mathbf{y}) = \int_{\boldsymbol{\phi}} \int_{\boldsymbol{\sigma}^2} \int_{\boldsymbol{\beta}} p(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}, \mathbf{x}_d|\mathbf{y}) d\boldsymbol{\beta} d\boldsymbol{\sigma}^2 d\boldsymbol{\phi}. \quad (20)$$

However, this posterior distribution is generally intractable and cannot be calculated analytically due to the high-dimensional integrals associated with it. One way to solve this problem is to approximate the marginalized posterior distribution using particles, such as in MCMC method [39]. However, these Monte-Carlo based statistical methods require high complexity due to the generation and processing of a large number of samples for approximating various distributions. On the other hand, in this paper, we are going to make the following variational approximation [15]

$$p(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}, \mathbf{x}_d|\mathbf{y}) \approx \mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}, \mathbf{x}_d), \quad (21)$$

where $\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}, \mathbf{x}_d)$ is the optimal distributional approximation chosen to minimize the Kullback-Leibler (KL) divergence [27] (also known as relative entropy) from $p(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}, \mathbf{x}_d|\mathbf{y})$, given by Eq. (22) at the top of the next page. To reduce the overall computational complexity of inference and in order to obtain a tractable approximation, conditional independence (also known as the mean field approximation) is enforced as a functional constraint in the variational distribution [40], i.e.,

$$\mathcal{Q}(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}, \mathbf{x}_d) = \mathcal{Q}_1(\boldsymbol{\beta})\mathcal{Q}_2(\boldsymbol{\sigma}^2)\mathcal{Q}_3(\boldsymbol{\phi})\mathcal{Q}_4(\mathbf{x}_d). \quad (23)$$

In essence, the approximation forces $\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}, \mathbf{x}_d$ to be mutually independent conditioned on $\mathbf{y}$. Substituting (21) and (23) into (20), we directly have $p(\mathbf{x}_d|\mathbf{y}) \approx \mathcal{Q}_4(\mathbf{x}_d)$, and therefore the optimal estimate of $\mathbf{x}_d$ over $\mathcal{Q}_4(\mathbf{x}_d)$ is a close approximation to the optimal estimate of $\mathbf{x}_d$ over $p(\mathbf{x}_d|\mathbf{y})$.

A. Computation of the KL Divergence

Using Bayes' rule, and due to the independence among channel, interference plus noise, PN and transmitted data, we have

$$p(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}, \mathbf{x}_d|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}, \mathbf{x}_d)p(\boldsymbol{\beta})p(\boldsymbol{\sigma}^2)p(\boldsymbol{\phi})p(\mathbf{x}_d), \quad (24)$$

where the normalization $p(\mathbf{y})$ does not depend on the unknown variables, and therefore can be dropped. Substituting (24) into (22), and making use of the conditional independence constraint (23), we obtain Eq. (25) at the top of the next page.

It is observed from (25) that for the computation of the KL divergence, one key issue remains to be addressed: the choice of the form of the variational distributions $\mathcal{Q}_1(\boldsymbol{\beta})$,$\mathcal{Q}_2(\boldsymbol{\sigma}^2)$,$\mathcal{Q}_3(\boldsymbol{\phi})$ and $\mathcal{Q}_4(\mathbf{x}_d)$. The key is to choose a distributional form that makes a good approximation to the exact joint posterior distribution and meanwhile provides analytical tractability during the Bayes update. In this paper, $\mathcal{Q}_1(\boldsymbol{\beta})$ and $\mathcal{Q}_3(\boldsymbol{\phi})$ are chosen to be in Gaussian forms. That is,

$$\mathcal{Q}_1(\boldsymbol{\beta}) = \frac{1}{\pi^{(Q+1)L}\det(\boldsymbol{\Psi}_{\boldsymbol{\beta}})} \exp\{-(\boldsymbol{\beta}-\mathbf{m}_{\boldsymbol{\beta}})^H \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta}-\mathbf{m}_{\boldsymbol{\beta}})\}, \quad (26)$$

where $\mathbf{m}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Psi}_{\boldsymbol{\beta}}$ are the unknown mean vector and covariance matrix for $\boldsymbol{\beta}$, respectively; and

$$\mathcal{Q}_3(\boldsymbol{\phi}) = \frac{1}{(2\pi)^{\frac{N_T N}{2}}\det(\boldsymbol{\Psi}_{\boldsymbol{\phi}})^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\boldsymbol{\phi}-\mathbf{m}_{\boldsymbol{\phi}})^T \boldsymbol{\Psi}_{\boldsymbol{\phi}}^{-1}(\boldsymbol{\phi}-\mathbf{m}_{\boldsymbol{\phi}})\}, \quad (27)$$

where $\mathbf{m}_{\boldsymbol{\phi}}$ and $\boldsymbol{\Psi}_{\boldsymbol{\phi}}$ are the unknown mean vector and covariance matrix for $\boldsymbol{\phi}$, respectively. In view of the discrete nature of the data, the variational distribution for $\mathbf{x}_d$ is chosen as

$$\mathcal{Q}_4(\mathbf{x}_d) = \delta(\mathbf{x}_d - \tilde{\mathbf{x}}_d), \quad (28)$$

where the vector Dirac delta function $\delta(\cdot)$ has the properties $\int \delta(\mathbf{x}_d - \tilde{\mathbf{x}}_d)d\mathbf{x}_d = 1$ and $\int \delta(\mathbf{x}_d - \tilde{\mathbf{x}}_d)f(\mathbf{x}_d)d\mathbf{x}_d = f(\tilde{\mathbf{x}}_d)$ for any smooth function $f(\cdot)$. Furthermore, the entries of $\boldsymbol{\sigma}^2$ are chosen to be distributed according to an inverse-gamma distribution with the unknown shape parameter $\tilde{a}_{k,j}$ and scale parameter $\tilde{b}_{k,j}$ as

$$\mathcal{Q}_2(\boldsymbol{\sigma}^2) = \prod_{j=0}^{N_R-1}\prod_{k=0}^{N-1} \frac{\tilde{b}_{k,j}^{\tilde{a}_{k,j}}}{\Gamma(\tilde{a}_{k,j})} \big(\sigma_j^2(k)\big)^{-\tilde{a}_{k,j}-1} \exp\big(-\frac{\tilde{b}_{k,j}}{\sigma_j^2(k)}\big). \quad (29)$$

Substituting the prior distributions (16)-(19) and the variational distributions (26)-(29) into (25), it is shown in Appendix B that after dropping those irrelevant terms, the KL divergence can be expressed as Eq. (30) in the middle of the next page, where $\psi(\tilde{a}_{k,j}) = \frac{\partial \log\Gamma(\tilde{a}_{k,j})}{\partial \tilde{a}_{k,j}}$, $\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2} = \text{Bldiag}\{\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}^0, ..., \boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}^{N_R-1}\}$ with $\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}^j = \text{diag}\{\frac{\tilde{a}_{0,j}}{\tilde{b}_{0,j}}, ..., \frac{\tilde{a}_{N-1,j}}{\tilde{b}_{N-1,j}}\}$, $\boldsymbol{\rho}_{\boldsymbol{\phi}} = (\mathbf{1}_{N_T N} + \jmath\mathbf{m}_{\boldsymbol{\phi}})$, η_z and $\boldsymbol{\mu}_z$ are the z^{th} eigenvalue and corresponding eigenvector of $\boldsymbol{\Psi}_{\boldsymbol{\phi}}$, i.e., $\boldsymbol{\Psi}_{\boldsymbol{\phi}} = \sum_{z=0}^{N_T N-1} \eta_z \boldsymbol{\mu}_z \boldsymbol{\mu}_z^T$. Accordingly, the goal of minimizing $\text{KL}(\mathcal{Q}_1(\boldsymbol{\beta})\mathcal{Q}_2(\boldsymbol{\sigma}^2)\mathcal{Q}_3(\boldsymbol{\phi})\mathcal{Q}_4(\mathbf{x}_d)||p(\boldsymbol{\beta}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}, \mathbf{x}_d|\mathbf{y}))$ reduces to seeking the values for the unknown parameters of the variational distributions $\mathbf{m}_{\boldsymbol{\beta}}, \boldsymbol{\Psi}_{\boldsymbol{\beta}}, \tilde{a}_{k,j}, \tilde{b}_{k,j}$, $\mathbf{m}_{\boldsymbol{\phi}}$, $\boldsymbol{\Psi}_{\boldsymbol{\phi}}$, $\tilde{\mathbf{x}}_d$. After we obtain an estimate of all these parameters, and due to $p(\mathbf{x}_d|\mathbf{y}) \approx \mathcal{Q}_4(\mathbf{x}_d)$, the data estimate directly equals $\tilde{\mathbf{x}}_d$.

B. Iterative Minimization of the KL Divergence

Jointly optimizing all parameters in the KL divergence is a challenging task, as it is a complicated function with coupled dependences among unknown variables. Fortunately, the estimates of the unknown parameters can be obtained iteratively by minimizing the KL divergence with respect to one set of the parameters at a time and fixing all the others to their last estimated values. The update at the q^{th} iteration follows as:

1) Updating $\mathbf{m}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Psi}_{\boldsymbol{\beta}}$ given $\hat{a}_{k,j}^{q-1}$,$\hat{\tilde{b}}_{k,j}^{q-1}$,$\hat{\mathbf{m}}_{\boldsymbol{\phi}}^{q-1}$,$\hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^{q-1}$,$\hat{\mathbf{x}}_d^{q-1}$:

Setting the first derivative of $\mathbb{F}$ with respect to $\boldsymbol{\Psi}_{\boldsymbol{\beta}}$ to zero, we obtain $\hat{\boldsymbol{\Psi}}_{\boldsymbol{\beta}}^q$, given by Eq. (31) in the middle of the next page, where $\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q-1} = \text{Bldiag}\{\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{0,q-1}, ..., \hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{N_R-1,q-1}\}$ with $\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{j,q-1} = \text{diag}\{\frac{\hat{\tilde{a}}_{0,j}^{q-1}}{\hat{\tilde{b}}_{0,j}^{q-1}}, ..., \frac{\hat{\tilde{a}}_{N-1,j}^{q-1}}{\hat{\tilde{b}}_{N-1,j}^{q-1}}\}$, $\hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q-1} = \mathbf{1}_{N_T N} + \jmath\hat{\mathbf{m}}_{\boldsymbol{\phi}}^{q-1}$, $\hat{\eta}_z^{q-1}$ and $\hat{\boldsymbol{\mu}}_z^{q-1}$ are the z^{th} eigenvalue and corresponding eigenvector of $\hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^{q-1}$, respectively. For $\hat{\mathbf{m}}_{\boldsymbol{\beta}}^q$, it can be obtained by setting the first derivative of $\mathbb{F}$ with respect to $\mathbf{m}_{\boldsymbol{\beta}}$ to zero:

$$\hat{\mathbf{m}}_{\boldsymbol{\beta}}^q = \hat{\boldsymbol{\Psi}}_{\boldsymbol{\beta}}^q \boldsymbol{\Xi}^H [\hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q-1}, \mathbf{E}_d \hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p \mathbf{x}_p] \hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q-1} \mathbf{y}. \quad (32)$$

$$\mathrm{KL}\big(\mathcal{Q}(\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d)||p(\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d|\mathbf{y})\big)=\int_{\mathbf{x}_d}\int_{\boldsymbol{\phi}}\int_{\boldsymbol{\sigma}^2}\int_{\boldsymbol{\beta}}\mathcal{Q}(\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d)\log\frac{\mathcal{Q}(\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d)}{p(\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d|\mathbf{y})}d\boldsymbol{\beta}d\boldsymbol{\sigma}^2d\boldsymbol{\phi}d\mathbf{x}_d. \tag{22}$$

$$\begin{aligned}
&\mathrm{KL}\big(\mathcal{Q}_1(\boldsymbol{\beta})\mathcal{Q}_2(\boldsymbol{\sigma}^2)\mathcal{Q}_3(\boldsymbol{\phi})\mathcal{Q}_4(\mathbf{x}_d)||p(\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d|\mathbf{y})\big)\\
&\propto\int_{\mathbf{x}_d}\int_{\boldsymbol{\phi}}\int_{\boldsymbol{\sigma}^2}\int_{\boldsymbol{\beta}}\mathcal{Q}_1(\boldsymbol{\beta})\mathcal{Q}_2(\boldsymbol{\sigma}^2)\mathcal{Q}_3(\boldsymbol{\phi})\mathcal{Q}_4(\mathbf{x}_d)\log\frac{\mathcal{Q}_1(\boldsymbol{\beta})\mathcal{Q}_2(\boldsymbol{\sigma}^2)\mathcal{Q}_3(\boldsymbol{\phi})\mathcal{Q}_4(\mathbf{x}_d)}{p(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d)p(\boldsymbol{\beta})p(\boldsymbol{\sigma}^2)p(\boldsymbol{\phi})p(\mathbf{x}_d)}d\boldsymbol{\beta}d\boldsymbol{\sigma}^2d\boldsymbol{\phi}d\mathbf{x}_d\\
&=\int_{\boldsymbol{\beta}}\mathcal{Q}_1(\boldsymbol{\beta})\log\mathcal{Q}_1(\boldsymbol{\beta})d\boldsymbol{\beta}+\int_{\boldsymbol{\sigma}^2}\mathcal{Q}_2(\boldsymbol{\sigma}^2)\log\mathcal{Q}_2(\boldsymbol{\sigma}^2)d\boldsymbol{\sigma}^2+\int_{\boldsymbol{\phi}}\mathcal{Q}_3(\boldsymbol{\phi})\log\mathcal{Q}_3(\boldsymbol{\phi})d\boldsymbol{\phi}+\int_{\mathbf{x}_d}\mathcal{Q}_4(\mathbf{x}_d)\log\mathcal{Q}_4(\mathbf{x}_d)d\mathbf{x}_d\\
&\quad-\int_{\boldsymbol{\beta}}\mathcal{Q}_1(\boldsymbol{\beta})\log p(\boldsymbol{\beta})d\boldsymbol{\beta}-\int_{\boldsymbol{\sigma}^2}\mathcal{Q}_2(\boldsymbol{\sigma}^2)\log p(\boldsymbol{\sigma}^2)d\boldsymbol{\sigma}^2-\int_{\boldsymbol{\phi}}\mathcal{Q}_3(\boldsymbol{\phi})\log p(\boldsymbol{\phi})d\boldsymbol{\phi}-\int_{\mathbf{x}_d}\mathcal{Q}_4(\mathbf{x}_d)\log p(\mathbf{x}_d)d\mathbf{x}_d\\
&\quad-\int_{\mathbf{x}_d}\int_{\boldsymbol{\phi}}\int_{\boldsymbol{\sigma}^2}\int_{\boldsymbol{\beta}}\mathcal{Q}_1(\boldsymbol{\beta})\mathcal{Q}_2(\boldsymbol{\sigma}^2)\mathcal{Q}_3(\boldsymbol{\phi})\mathcal{Q}_4(\mathbf{x}_d)\log p(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d)d\boldsymbol{\beta}d\boldsymbol{\sigma}^2d\boldsymbol{\phi}d\mathbf{x}_d.
\end{aligned} \tag{25}$$

$$\begin{aligned}
&\mathrm{KL}\big(\mathcal{Q}_1(\boldsymbol{\beta})\mathcal{Q}_2(\boldsymbol{\sigma}^2)\mathcal{Q}_3(\boldsymbol{\phi})\mathcal{Q}_4(\mathbf{x}_d)||p(\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d|\mathbf{y})\big)\\
&\triangleq\mathbb{F}(\mathbf{m}_{\boldsymbol{\beta}},\boldsymbol{\Psi}_{\boldsymbol{\beta}},\tilde{a}_{k,j},\tilde{b}_{k,j},\mathbf{m}_{\boldsymbol{\phi}},\boldsymbol{\Psi}_{\boldsymbol{\phi}},\tilde{\mathbf{x}}_d)\\
&\propto-\log\det(\boldsymbol{\Psi}_{\boldsymbol{\beta}})+\sum_{j=0}^{N_R-1}\sum_{k=0}^{N-1}\big[(\tilde{a}_{k,j}+1)\psi(\tilde{a}_{k,j})-\log\tilde{b}_{k,j}-\tilde{a}_{k,j}-\log\Gamma(\tilde{a}_{k,j})\big]-\frac{1}{2}\log\det(\boldsymbol{\Psi}_{\boldsymbol{\phi}})+\mathrm{Tr}\big\{\mathbf{R}_{\boldsymbol{\beta}}^{-1}(\mathbf{m}_{\boldsymbol{\beta}}\mathbf{m}_{\boldsymbol{\beta}}^H+\boldsymbol{\Psi}_{\boldsymbol{\beta}})\big\}\\
&\quad-\sum_{j=0}^{N_R-1}\sum_{k=0}^{N-1}\big[(-a_{k,j}-1)(\log\tilde{b}_{k,j}-\psi(\tilde{a}_{k,j}))-\frac{\tilde{a}_{k,j}}{\tilde{b}_{k,j}}b_{k,j}\big]-\sum_{i=0}^{N_T-1}\sum_{k=0}^{N_d-1}\log\Big\{\sum_{\bar{x}_d^i(k)\in\mathfrak{c}}\delta\big(\tilde{x}_d^i(k)-\bar{x}_d^i(k)\big)\Big\}\\
&\quad+\sum_{j=0}^{N_R-1}\sum_{k=0}^{N-1}\big[\log\tilde{b}_{k,j}-\psi(\tilde{a}_{k,j})\big]+\mathbf{y}^H\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\mathbf{y}-2\Re\big\{\mathbf{y}^H\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}\big\}+\frac{1}{2}\big(\mathrm{Tr}\{\mathbf{R}_{\boldsymbol{\phi}}^{-1}\boldsymbol{\Psi}_{\boldsymbol{\phi}}\}+\mathbf{m}_{\boldsymbol{\phi}}^T\mathbf{R}_{\boldsymbol{\phi}}^{-1}\mathbf{m}_{\boldsymbol{\phi}}\big)\\
&\quad+\mathbf{m}_{\boldsymbol{\beta}}^H\boldsymbol{\Xi}^H[\boldsymbol{\rho}_{\boldsymbol{\phi}},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}+\sum_{z=0}^{N_TN-1}\eta_z\mathbf{m}_{\boldsymbol{\beta}}^H\boldsymbol{\Xi}^H[\boldsymbol{\mu}_z,\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\mu}_z,\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}\\
&\quad+\mathrm{Tr}\{\boldsymbol{\Xi}^H[\boldsymbol{\rho}_{\boldsymbol{\phi}},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Psi}_{\boldsymbol{\beta}}\}+\mathrm{Tr}\{\sum_{z=0}^{N_TN-1}\eta_z\boldsymbol{\Xi}^H[\boldsymbol{\mu}_z,\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\mu}_z,\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Psi}_{\boldsymbol{\beta}}\}.
\end{aligned} \tag{30}$$

$$\begin{aligned}
\hat{\boldsymbol{\Psi}}_{\boldsymbol{\beta}}^q=\Big(\mathbf{R}_{\boldsymbol{\beta}}^{-1}+\boldsymbol{\Xi}^H[\hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q-1},\mathbf{E}_d\hat{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q-1}\boldsymbol{\Xi}[\hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q-1},\mathbf{E}_d\hat{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\\
+\sum_{z=0}^{N_TN-1}\hat{\eta}_z^{q-1}\boldsymbol{\Xi}^H[\hat{\boldsymbol{\mu}}_z^{q-1},\mathbf{E}_d\hat{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q-1}\boldsymbol{\Xi}[\hat{\boldsymbol{\mu}}_z^{q-1},\mathbf{E}_d\hat{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\Big)^{-1}.
\end{aligned} \tag{31}$$

2) Updating $\tilde{a}_{k,j}$ and $\tilde{b}_{k,j}$ given $\hat{\mathbf{m}}_{\boldsymbol{\beta}}^q,\hat{\boldsymbol{\Psi}}_{\boldsymbol{\beta}}^q,\hat{\mathbf{m}}_{\boldsymbol{\phi}}^{q-1},\hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^{q-1},\hat{\mathbf{x}}_d^{q-1}$:

Setting the first derivative of $\mathbb{F}$ with respect to $\tilde{a}_{k,j}$ and $\tilde{b}_{k,j}$ to zero and solving them simultaneously, it is shown in Appendix C that

$$\hat{\tilde{a}}_{k,j}^q=a_{k,j}+1, \tag{33}$$

and $\hat{\tilde{b}}_{k,j}^q$ given by Eq. (34) at the top of the next page, where $\hat{\lambda}_r^q$ and $\hat{\boldsymbol{\nu}}_r^q$ are defined from the eigen-decomposition $\hat{\boldsymbol{\Psi}}_{\boldsymbol{\beta}}^q=\sum_{r=0}^{(Q+1)L-1}\hat{\lambda}_r^q\hat{\boldsymbol{\nu}}_r^q(\hat{\boldsymbol{\nu}}_r^q)^H$.

3) Updating $\mathbf{m}_{\boldsymbol{\phi}}$ and $\boldsymbol{\Psi}_{\boldsymbol{\phi}}$ given $\hat{\mathbf{m}}_{\boldsymbol{\beta}}^q,\hat{\boldsymbol{\Psi}}_{\boldsymbol{\beta}}^q,\hat{\tilde{a}}_{k,j}^q,\hat{\tilde{b}}_{k,j}^q,\hat{\mathbf{x}}_d^{q-1}$:

Setting the first derivative of $\mathbb{F}$ with respect to $\boldsymbol{\Psi}_{\boldsymbol{\phi}}$ to zero and using the equivalent expressions between (13) and (15), we obtain $\hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^q$, given by Eq. (35) at the top of the next page. On the other hand, setting the first derivative of $\mathbb{F}$ with respect to $\mathbf{m}_{\boldsymbol{\phi}}$ to zero and using the equivalent expressions between (13) and (15), we obtain $\hat{\mathbf{m}}_{\boldsymbol{\phi}}^q$, given by Eq. (36) at the top of the next page.

4) Updating $\hat{\mathbf{x}}_d$ given $\hat{\mathbf{m}}_{\boldsymbol{\beta}}^q,\hat{\boldsymbol{\Psi}}_{\boldsymbol{\beta}}^q,\hat{\tilde{a}}_{k,j}^q,\hat{\tilde{b}}_{k,j}^q,\hat{\mathbf{m}}_{\boldsymbol{\phi}}^q,\hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^q$:

For minimizing $\mathbb{F}$ with respect to $\tilde{\mathbf{x}}_d$, the term $\sum_{i=0}^{N_T-1}\sum_{k=0}^{N_d-1}\log\{\sum_{\bar{x}_d^i(k)\in\mathfrak{c}}\delta(\tilde{x}_d^i(k)-\bar{x}_d^i(k))\}$ accounts for the constellation constraint. But a multidimensional exhaustive search over all possible discrete data combinations is a formidable task. Here, we propose a complexity-reduced linear data estimate by relaxing $\tilde{\mathbf{x}}_d$ to be continuous and setting the

$$\hat{\tilde{b}}_{k,j}^{q} = b_{k,j} + |\mathbf{y}(k+jN)|^2 - 2\Re\Big\{\mathbf{y}^*(k+jN)\big[\boldsymbol{\Xi}[\hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q-1}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}\big](k+jN)\Big\}$$
$$+ \Big|\big[\boldsymbol{\Xi}[\hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q-1}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}\big](k+jN)\Big|^2 + \sum_{z=0}^{N_TN-1}\hat{\eta}_z^{q-1}\Big|\big[\boldsymbol{\Xi}[\hat{\boldsymbol{\mu}}_z^{q-1}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}\big](k+jN)\Big|^2$$
$$+ \sum_{r=0}^{(Q+1)L-1}\hat{\lambda}_r^{q}\Big|\big[\boldsymbol{\Xi}[\hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q-1}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\hat{\boldsymbol{\nu}}_r^{q}\big](k+jN)\Big|^2 + \sum_{z=0}^{N_TN-1}\sum_{r=0}^{(Q+1)L-1}\hat{\eta}_z^{q-1}\hat{\lambda}_r^{q}\Big|\big[\boldsymbol{\Xi}[\hat{\boldsymbol{\mu}}_z^{q-1}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\hat{\boldsymbol{\nu}}_r^{q}\big](k+jN)\Big|^2. \tag{34}$$

$$\hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^{q} = \Bigg(\mathbf{R}_{\boldsymbol{\phi}}^{-1} + 2\Big(\boldsymbol{\Omega}^H[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Omega}[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]$$
$$+ \sum_{r=0}^{(Q+1)L-1}\hat{\lambda}_r^{q}\boldsymbol{\Omega}^H[\hat{\boldsymbol{\nu}}_r^{q}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Omega}[\hat{\boldsymbol{\nu}}_r^{q}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]\Big)\Bigg)^{-1}. \tag{35}$$

$$\hat{\mathbf{m}}_{\boldsymbol{\phi}}^{q} = 2\hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^{q}\times\Big(-\jmath\boldsymbol{\Omega}^H[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\mathbf{y} + \jmath\boldsymbol{\Omega}^H[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Omega}[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]\mathbf{1}_{N_TN}$$
$$+ \sum_{r=0}^{(Q+1)L-1}\hat{\lambda}_r^{q}\jmath\boldsymbol{\Omega}^H[\hat{\boldsymbol{\nu}}_r^{q}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Omega}[\hat{\boldsymbol{\nu}}_r^{q}, \mathbf{E}_d\hat{\mathbf{x}}_d^{q-1} + \mathbf{E}_p\mathbf{x}_p]\mathbf{1}_{N_TN}\Big). \tag{36}$$

$$\acute{\mathbf{x}}_d^{q} = \Big(\mathbf{E}_d^H\boldsymbol{\Theta}^H[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q}]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Theta}[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q}]\mathbf{E}_d + \sum_{r=0}^{(Q+1)L-1}\hat{\lambda}_r^{q}\mathbf{E}_d^H\boldsymbol{\Theta}^H[\hat{\boldsymbol{\nu}}_r^{q}, \hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q}]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Theta}[\hat{\boldsymbol{\nu}}_r^{q}, \hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q}]\mathbf{E}_d$$
$$+ \sum_{z=0}^{N_TN-1}\hat{\eta}_z^{q}\mathbf{E}_d^H\boldsymbol{\Theta}^H[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \hat{\boldsymbol{\mu}}_z^{q}]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Theta}[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \hat{\boldsymbol{\mu}}_z^{q}]\mathbf{E}_d + \sum_{z=0}^{N_TN-1}\sum_{r=0}^{(Q+1)L-1}\hat{\eta}_z^{q}\hat{\lambda}_r^{q}\mathbf{E}_d^H\boldsymbol{\Theta}^H[\hat{\boldsymbol{\nu}}_r^{q}, \hat{\boldsymbol{\mu}}_z^{q}]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Theta}[\hat{\boldsymbol{\nu}}_r^{q}, \hat{\boldsymbol{\mu}}_z^{q}]\mathbf{E}_d\Big)^{-1}$$
$$\times\Big(\mathbf{E}_d^H\boldsymbol{\Theta}^H[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q}]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\mathbf{y} - \mathbf{E}_d^H\boldsymbol{\Theta}^H[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q}]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Theta}[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q}]\mathbf{E}_p\mathbf{x}_p$$
$$- \sum_{r=0}^{(Q+1)L-1}\hat{\lambda}_r^{q}\mathbf{E}_d^H\boldsymbol{\Theta}^H[\hat{\boldsymbol{\nu}}_r^{q}, \hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q}]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Theta}[\hat{\boldsymbol{\nu}}_r^{q}, \hat{\boldsymbol{\rho}}_{\boldsymbol{\phi}}^{q}]\mathbf{E}_p\mathbf{x}_p - \sum_{z=0}^{N_TN-1}\hat{\eta}_z^{q}\mathbf{E}_d^H\boldsymbol{\Theta}^H[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \hat{\boldsymbol{\mu}}_z^{q}]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Theta}[\hat{\mathbf{m}}_{\boldsymbol{\beta}}^{q}, \hat{\boldsymbol{\mu}}_z^{q}]\mathbf{E}_p\mathbf{x}_p$$
$$- \sum_{z=0}^{N_TN-1}\sum_{r=0}^{(Q+1)L-1}\hat{\eta}_z^{q}\hat{\lambda}_r^{q}\mathbf{E}_d^H\boldsymbol{\Theta}^H[\hat{\boldsymbol{\nu}}_r^{q}, \hat{\boldsymbol{\mu}}_z^{q}]\hat{\boldsymbol{\Lambda}}_{\boldsymbol{\sigma}^2}^{q}\boldsymbol{\Theta}[\hat{\boldsymbol{\nu}}_r^{q}, \hat{\boldsymbol{\mu}}_z^{q}]\mathbf{E}_p\mathbf{x}_p\Big). \tag{37}$$

first derivative of $\mathbb{F}$ (without the term corresponding to the constellation constraint) with respect to $\tilde{\mathbf{x}}_d$ to zero. Using the equivalent expressions between (10) and (13), the linear data estimate can be obtained as Eq. (37) in the middle of this page. Then the constellation constraint is enforced by quantizing $\acute{\mathbf{x}}_d^q$ as $\hat{\mathbf{x}}_d^q = \text{Qant}(\acute{\mathbf{x}}_d^q)$.

Remark 2: The computational complexity of each iteration is dominated by (31), (35) and (37). In particular, in (31), the complexity mainly comes from the $(Q+1)L \times (Q+1)L$ matrix inversion. In (35), there are two eigen-decompositions of matrices with size $N_TN \times N_TN$ ($\hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^{q-1}$ and $\hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^{q}$), an eigen-decomposition of $\hat{\boldsymbol{\Psi}}_{\boldsymbol{\beta}}^{q}$ with size $(Q+1)L \times (Q+1)L$ and one $N_TN \times N_TN$ matrix inversion. In (37), the complexity is dominated by the $N_TN_d \times N_TN_d$ matrix inversion. For an $M \times M$ matrix, the complexities of inversion and eigen-decomposition are both $O(M^3)$. Therefore, the computational complexity for each iteration of the proposed VBI-based algorithm is $O(2((Q+1)L)^3 + 3(N_TN)^3 + (N_TN_d)^3)$. Since $Q \ll N$ and $N_d < N$, the computational complexity is dominated by the term $(N_TN)^3$, and the overall computational complexity depends on the number of iterations.

C. Summary and Initialization

In summary, the proposed algorithm proceeds as follows:
1) The priori distributions of the DSC, PN, data and NBI are assumed to be known and are given by (16)-(19), respectively.
2) The forms of variational distributions of the unknown parameters are assumed in (26)-(29), with distributional parameters to be optimized. Conditional independence (i.e., mean-field approximation) among variational distributions is applied when KL divergence is computed in (30).
3) Assuming an initial value for $\frac{\hat{\tilde{a}}_{k,j}^0}{\hat{\tilde{b}}_{k,j}^0}, \hat{\mathbf{m}}_{\boldsymbol{\phi}}^0, \hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^0, \hat{\mathbf{x}}_d^0$, the unknown parameters of variational distributions are iteratively estimated by minimizing the KL divergence as (31)-(37).

4) After convergence, the desired data estimate can be obtained using (37).

This iterative process is guaranteed to converge monotonically to at least a stationary point [28], [29]. From (37) it can be observed that in the proposed signal detection, the ICIs caused by DSCs, the interferences result from NBIs as well as the CPEs and ICIs caused by PNs are all iteratively eliminated.

To provide an initial estimate, recalling (10), (11) and (13), we have

$$\mathbf{y} = \boldsymbol{\Xi}[\boldsymbol{\phi}, \mathbf{E}_p\mathbf{x}_p]\boldsymbol{\beta} + \boldsymbol{\Theta}[\boldsymbol{\beta}, \boldsymbol{\phi}]\mathbf{E}_d\mathbf{x}_d + \mathbf{w}. \tag{38}$$

By collecting the samples in $\mathbf{y}$ corresponding to pilot symbols, and treating the term containing $\mathbf{x}_d$ as interference, the least squares (LS) estimate of $\boldsymbol{\beta}$ is obtained as

$$\hat{\boldsymbol{\beta}}^0 = \left(\boldsymbol{\Xi}_p^H[\hat{\mathbf{m}}_{\boldsymbol{\phi}}^0, \mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Xi}_p[\hat{\mathbf{m}}_{\boldsymbol{\phi}}^0, \mathbf{E}_p\mathbf{x}_p]\right)^{-1}\boldsymbol{\Xi}_p^H[\hat{\mathbf{m}}_{\boldsymbol{\phi}}^0, \mathbf{E}_p\mathbf{x}_p]\mathbf{y}_p, \tag{39}$$

where the elements of $\mathbf{y}_p$ and the rows of $\boldsymbol{\Xi}_p$ are taken from $\mathbf{y}$ and $\boldsymbol{\Xi}$, respectively, with corresponding positions of pilot; and $\hat{\mathbf{m}}_{\boldsymbol{\phi}}^0$ is taken as the prior mean of $\boldsymbol{\phi}$, i.e., $\hat{\mathbf{m}}_{\boldsymbol{\phi}}^0 = \mathbf{0}_{N_TN}$. Substituting (39) into (10), the initial signal detection is thus obtained as

$$\hat{\mathbf{x}}_d^0 = \text{Qant}\Big(\left(\mathbf{E}_d^H\boldsymbol{\Theta}^H[\hat{\boldsymbol{\beta}}^0, \hat{\mathbf{m}}_{\boldsymbol{\phi}}^0]\boldsymbol{\Theta}[\hat{\boldsymbol{\beta}}^0, \hat{\mathbf{m}}_{\boldsymbol{\phi}}^0]\mathbf{E}_d\right)^{-1} \times \mathbf{E}_d^H\boldsymbol{\Theta}^H[\hat{\boldsymbol{\beta}}^0, \hat{\mathbf{m}}_{\boldsymbol{\phi}}^0]\left(\mathbf{y} - \boldsymbol{\Theta}[\hat{\boldsymbol{\beta}}^0, \hat{\mathbf{m}}_{\boldsymbol{\phi}}^0]\mathbf{E}_p\mathbf{x}_p\right)\Big). \tag{40}$$

For the initial value $\hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^0$, we set $\hat{\boldsymbol{\Psi}}_{\boldsymbol{\phi}}^0$ to be an $N_TN \times N_TN$ all-zero matrix. Finally, since $\frac{\tilde{a}_{k,j}}{\tilde{b}_{k,j}} = \mathbb{E}\{\frac{1}{\sigma_j^2(k)}\}$, where $\sigma_j^2(k)$ is the power of the $(k+jN)^{th}$ element in $\mathbf{w}$, the initial value of $\frac{\hat{\tilde{a}}_{k,j}^0}{\hat{\tilde{b}}_{k,j}^0}$ can be set as $\frac{\hat{\tilde{a}}_{k,j}^0}{\hat{\tilde{b}}_{k,j}^0} = \frac{1}{\hat{\sigma}_j^2(k)}$, with $\hat{\sigma}_j^2(k) = |\mathbf{w}(k+jN)|^2 = |[\mathbf{y} - \boldsymbol{\Xi}[\hat{\mathbf{m}}_{\boldsymbol{\phi}}^0, \mathbf{E}_d\hat{\mathbf{x}}_d^0 + \mathbf{E}_p\mathbf{x}_p]\hat{\boldsymbol{\beta}}^0](k+jN)|^2$.

IV. Simulation Results and Discussions

In this section, the performance of the proposed algorithm for signal detection operating over unknown DSCs in the presence of multiple NBIs and PNs is demonstrated by Monte Carlo simulations. There are 4 single-antenna MTs ($N_T = 4$) and a BS with 4 receive antennas ($N_R = 4$). Each OFDM symbol has 64 subcarriers ($N = 64$) and the length of the CP is $N_{cp} = 8$. The signal bandwidth is 20 MHz and the sampling interval T_s is thus 50 ns. The normalized maximal Doppler shift is set as $Nf_dT_s = 0.15$[1], where f_d represents the maximum Doppler frequency. The PN is generated by the true PN model $\mathbf{e}^{j\phi^i}$, and the Wiener PN is adopted with the PN rate [8] $\beta^i NT_s = 10^{-3}$ for each MT, where β^i is the two-side 3-dB linewidth of the Lorentzian-shaped PSD of the oscillator. The received OFDM-V-MIMO signal at each receive antenna of the BS is affected by multiple NBIs. The multiple NBIs add Gaussian disturbances to the received signal and are randomly positioned in the signal bandwidth, affecting 4 contiguous subcarriers for each receive antenna at the BS. Notice that the positions of interferences at different receive antennas are in general different. The signal-to-interference ratio (SIR) over the jammed subcarriers is $10\log(E_s/\sigma_\varsigma^2)$ with

[1]Similar conclusions can be drawn for other normalized maximal Doppler shift (e.g., 0.075 and 0.03), and the corresponding results are not presented here.

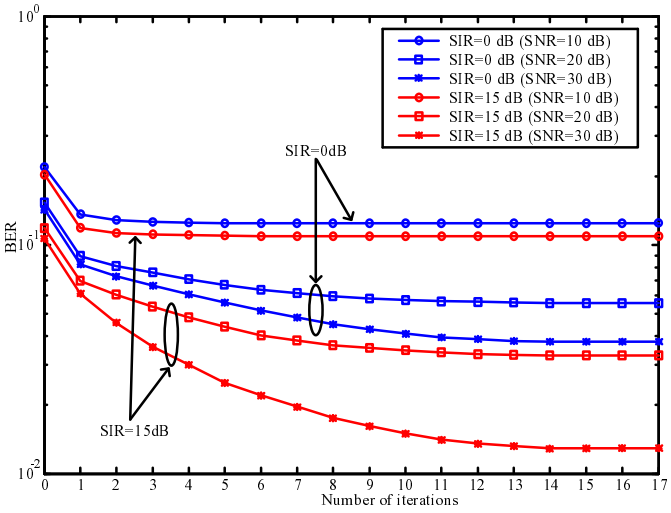


Fig. 2. Convergence performance of the proposed method (4 NBIs, PN rate=10^{-3} and normalized Doppler spread=0.15).

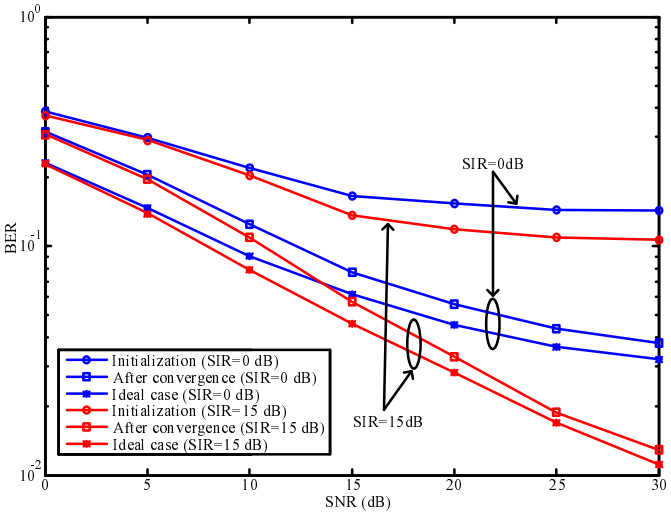


Fig. 3. Performance of signal detection versus SNR (4 NBIs, PN rate=10^{-3} and normalized Doppler spread=0.15).

$E_s = \mathbb{E}\{|x_d^i|^2\}$ denotes the symbol energy and σ_ς^2 is assumed to be constant over the observation period. The signal-to-noise ratio (SNR) is defined as $10\log(E_s/\sigma_v^2)$. Without loss of generality, the channel for each MT has three paths with an exponential power delay profile, namely $\varpi_l^2 = \exp(-\kappa l)((1-\exp(-\kappa))/(1-\exp(-3\kappa))), l = 0, 1, 2$ with $\kappa = 1/3$. Each path coefficient follows a complex Gaussian distribution. The data symbols are modulated by quadrature phase-shift keying (QPSK) with unit power. The pilot cluster follows the structure from [41]. More specifically, seven pilot clusters are used for each MT. The clusters are equally spaced among subcarriers and 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. Furthermore, the generalized complex exponential BEM (GCE-BEM) [34] is adopted and the statistical information of channel is assumed to be known such that the prior covariance derived in Appendix A can be computed.

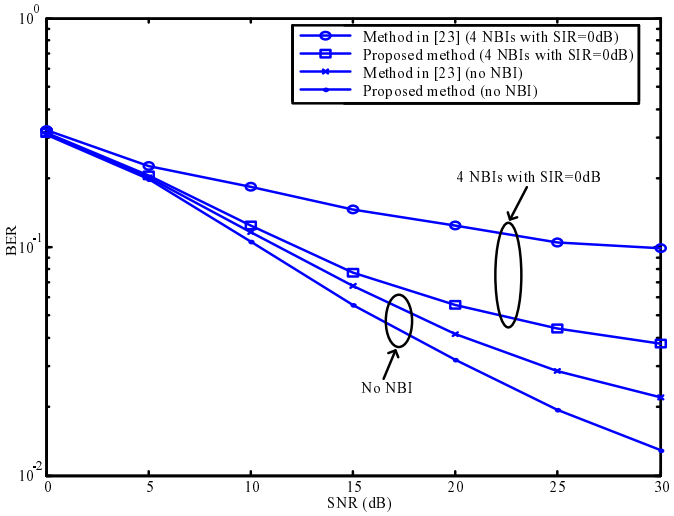


Fig. 4. Performance comparison between the proposed method and method in [23] (PN rate=10^{-3} and normalized Doppler spread=0.15).

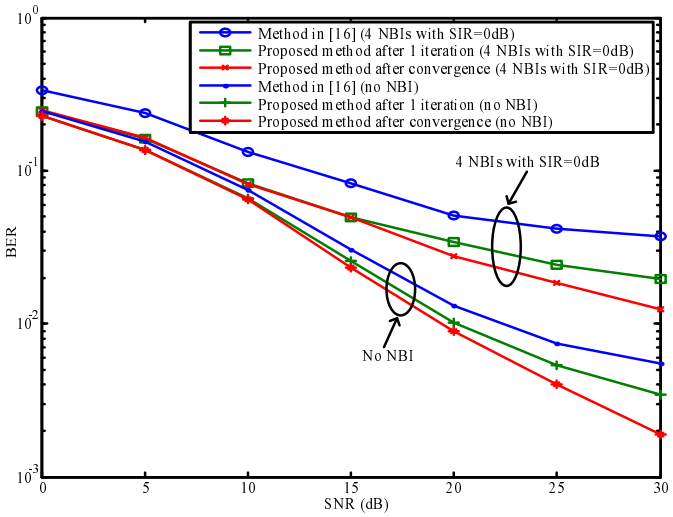


Fig. 5. Performance comparison between the proposed method and method in [16] (PN rate=10^{-3} and normalized Doppler spread=0.15).

Figure 2 presents the convergence performance of the proposed iterative algorithm for the 4 NBIs case with SIR=0dB and 15dB. From Fig. 2, it can be seen that the bit-error-rates (BERs) improve significantly in the first few iterations and converge to stable values within 15 iterations. For the rest of the figures, the proposed iterative algorithm is terminated at 15 iterations.

Figure 3 shows the signal detection performance versus SNRs for the proposed initialization, and the proposed iterative algorithm after convergence. The performance of the ideal case which assumes perfect CSI, no PN, and known positions and powers of NBIs plus AWGN is also shown for comparison. It can be seen that the iterative algorithm improves with respect to the initialization significantly, especially at high SNRs. After convergence, the performance of the proposed iterative signal detection algorithm is very close to that of the ideal case, which requires a lot more information.

Figure 4 compares the performance of the proposed algorithm to that of [23]. Also based on VBI, [23] proposes an iterative algorithm jointly estimate DSC and detect data in single-input single-output (SISO) system, assuming the noise variance is known. For the purpose of comparison, the method in [23] was extended to MIMO system. It can be seen from Fig. 4 that in case of no NBI, the proposed algorithm, which not only jointly estimate channel and detect data but also deals with PNs, performs better than that of [23]. This shows the importance of PN compensation in OFDM systems. Furthermore, in case of 4 NBIs with SIR=0dB, the proposed algorithm also performs better than that of [23], with performance gap more significant than the case with no NBI. This shows the importance of NBI mitigation in OFDM systems.

Figure 5 compares the performance of the proposed algorithm to that of [16], which tackles the PN by estimating and eliminating the CPE, assuming the CSI is perfectly known. For fair comparison, CSI is also assumed to be perfectly known for the proposed algorithm. It can be seen from Fig. 5 that for both cases of no NBI and 4 NBIs, the proposed algorithm performs much better than that of [16]. In fact, with only a single iteration, the proposed algorithm already outperforms the method from [16]. The reason is that the proposed scheme not only compensates the CPEs, but also the ICIs introduced by the PNs. This shows the importance of estimating and compensating the whole realization of PNs.

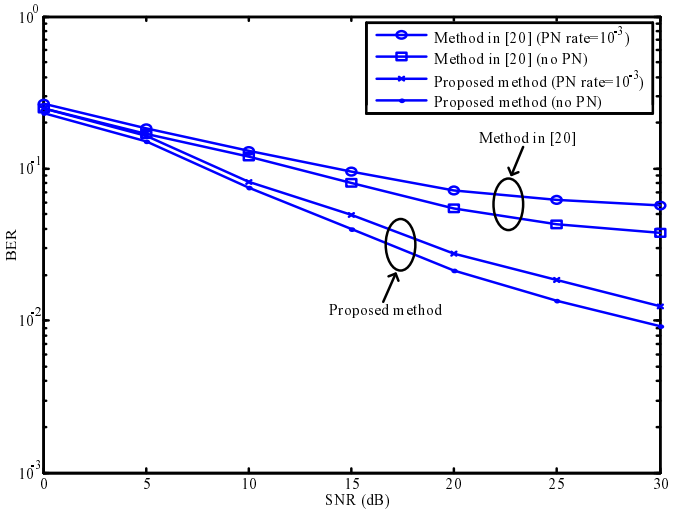


Fig. 6. Performance comparison between the proposed method and method in [20] (4 NBIs with SIR=0dB and normalized Doppler spread=0.15).

Finally, Fig. 6 compares the performance of the proposed algorithm to that of [20]. Based on EM, [20] proposes an iterative algorithm jointly estimate NBI and detect data in single-input single-output (SISO) system, assuming the channel is static and perfectly known. For fair comparison, the channel is also assumed to be perfectly known for the proposed algorithm. For the purpose of comparison, the method in [20] was extended to MIMO system. It can be seen from Fig. 6 that in case of no PN, the proposed algorithm, which not only jointly estimate NBI and detect data but also deals with ICI induced by high mobility, performs better than that of [20]. In case of PN with PN rate=10^{-3}, the proposed algorithm also performs significantly better than that of [20].

V. Conclusion

In this paper, we have addressed the challenging problem of signal detection for OFDM-V-MIMO systems operating over DSCs in the presence of multiple unknown NBIs and PNs. Based on the VBI framework, an iteratively updating algorithm for joint signal detection, DSC, NBI and PN estimations has been developed. Simulation results showed that after convergence, the performance of the proposed signal detection algorithm is very close to that of the ideal case which assumes perfect CSI, no PN, and known positions and powers of NBIs plus AWGN. Moreover, simulation results showed that the performance of the proposed signal detection algorithm outperforms other existing methods.

Appendix A
Derivation of $\mathbf{R}_{\boldsymbol{\beta}}$ in (16)

From (7) we have

$$\mathbf{h} = \mathbf{C}\boldsymbol{\beta} + \boldsymbol{\xi}, \tag{41}$$

where $\mathbf{h} = [(\mathbf{h}^0)^T, ..., (\mathbf{h}^{N_R-1})^T]^T$ with $\mathbf{h}^j = [(\mathbf{h}^{j0})^T, \dots, (\mathbf{h}^{j,N_T-1})^T]^T$ and $\mathbf{h}^{ji} = [(\mathbf{h}_0^{ji})^T, ..., (\mathbf{h}_{L^{ji}-1}^{ji})^T]^T$, the $NL \times (Q+1)L$ matrix $\mathbf{C} = \text{Bldiag}\{\mathbf{M}^0, ..., \mathbf{M}^{N_R-1}\}$ with $\mathbf{M}^j \triangleq \text{Bldiag}\{\mathbf{I}_{L^{j0}} \otimes \mathbf{B}, ..., \mathbf{I}_{L^{j,N_T-1}} \otimes \mathbf{B}\}$, $\boldsymbol{\xi}$ represents the corresponding BEM modeling error, which is to be minimized in the MSE sense. The optimal BEM coefficient is given by the LS solution

$$\boldsymbol{\beta} = (\mathbf{C}^H\mathbf{C})^{-1}\mathbf{C}^H\mathbf{h}. \tag{42}$$

Using (42), we have

$$\begin{aligned}\mathbf{R}_{\boldsymbol{\beta}} &= \mathbb{E}\{\boldsymbol{\beta}\boldsymbol{\beta}^H\} \\ &= (\mathbf{C}^H\mathbf{C})^{-1}\mathbf{C}^H\mathbb{E}\{\mathbf{h}\mathbf{h}^H\}\mathbf{C}(\mathbf{C}^H\mathbf{C})^{-1}.\end{aligned} \tag{43}$$

Since the channel follows Rayleigh fading and the correlation of different channel paths is given by

$$\begin{aligned}\mathbb{E}\Big\{h^{j_1i_1}(n_1,l_1)\big(h^{j_2i_2}(n_2,l_2)\big)^*\Big\} &= \varpi_{l_1}^2\delta(j_1-j_2)\delta(i_1-i_2) \\ &\times \delta(l_1-l_2)J_0\big(2\pi f_d(n_1-n_2)T_s\big),\end{aligned} \tag{44}$$

where ϖ_l^2 denotes the average power of the l^{th} path; $J_0(\cdot)$ represents the zero-order Bessel function of the first kind; f_d represents the maximum Doppler frequency and T_s is the sample interval. Expressing (44) in the form of $\mathbb{E}\{\mathbf{h}\mathbf{h}^H\}$, we obtain

$$\begin{aligned}\mathbb{E}\{\mathbf{h}\mathbf{h}^H\} = \text{Bldiag}\{&\mathbf{R}_{\mathbf{h}^{0,0}}, ..., \mathbf{R}_{\mathbf{h}^{0,N_T-1}}, ..., \\ &\mathbf{R}_{\mathbf{h}^{N_R-1,0}}, ..., \mathbf{R}_{\mathbf{h}^{N_R-1,N_T-1}}\},\end{aligned} \tag{45}$$

where $\mathbf{R}_{\mathbf{h}^{ji}} = \text{diag}\{\varpi_0^2, ..., \varpi_{L^{ji}-1}^2\} \otimes \mathbf{J}$ and the $(m,n)^{th}$ element of the $N \times N$ matrix $\mathbf{J}$ is $J_0\big(2\pi f_d(m-n)T_s\big)$. Finally, substituting (45) into (43), the correlation matrix $\mathbf{R}_{\boldsymbol{\beta}}$ can be obtained.

Appendix B
Derivation of (30)

We derive term-by-term in (25). First, using the variational distributions (26)-(29), the first four terms of (25) can be computed as

$$\begin{aligned}&\int_{\boldsymbol{\beta}} \mathcal{Q}_1(\boldsymbol{\beta})\log\mathcal{Q}_1(\boldsymbol{\beta})d\boldsymbol{\beta} \\ &= -(Q+1)L\log\pi - \log\det(\boldsymbol{\Psi}_{\boldsymbol{\beta}}) - \mathbf{m}_{\boldsymbol{\beta}}^H\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}\mathbf{m}_{\boldsymbol{\beta}} \\ &\quad + 2\Re\Big\{\mathbb{E}_{\boldsymbol{\beta}}\{\boldsymbol{\beta}^H\}\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}\mathbf{m}_{\boldsymbol{\beta}}\Big\} - \text{Tr}\Big\{\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}\mathbb{E}_{\boldsymbol{\beta}}\{\boldsymbol{\beta}\boldsymbol{\beta}^H\}\Big\} \\ &= -\log\det(\boldsymbol{\Psi}_{\boldsymbol{\beta}}) - (Q+1)L\log\pi - (Q+1)L;\end{aligned} \tag{46}$$

$$\begin{aligned}&\int_{\boldsymbol{\sigma}^2} \mathcal{Q}_2(\boldsymbol{\sigma}^2)\log\mathcal{Q}_2(\boldsymbol{\sigma}^2)d\boldsymbol{\sigma}^2 \\ &= \sum_{j=0}^{N_R-1}\sum_{k=0}^{N-1}\Big[\tilde{a}_{k,j}\log\tilde{b}_{k,j} + (-\tilde{a}_{k,j}-1)\mathbb{E}_{\sigma_j^2(k)}\big\{\log\sigma_j^2(k)\big\} \\ &\qquad - \mathbb{E}_{\sigma_j^2(k)}\big\{\frac{1}{\sigma_j^2(k)}\big\}\tilde{b}_{k,j} - \log\Gamma(\tilde{a}_{k,j})\Big] \\ &= \sum_{j=0}^{N_R-1}\sum_{k=0}^{N-1}\Big[(\tilde{a}_{k,j}+1)\psi(\tilde{a}_{k,j}) - \log\tilde{b}_{k,j} - \tilde{a}_{k,j} - \log\Gamma(\tilde{a}_{k,j})\Big],\end{aligned} \tag{47}$$

where $\psi(\tilde{a}_{k,j}) = \frac{\partial\log\Gamma(\tilde{a}_{k,j})}{\partial\tilde{a}_{k,j}}$;

$$\begin{aligned}&\int_{\boldsymbol{\phi}} \mathcal{Q}_3(\boldsymbol{\phi})\log\mathcal{Q}_3(\boldsymbol{\phi})d\boldsymbol{\phi} \\ &= -\frac{1}{2}\log\det(\boldsymbol{\Psi}_{\boldsymbol{\phi}}) - \frac{N_TN}{2}\log 2\pi - \frac{1}{2}\Big(\text{Tr}\big\{\boldsymbol{\Psi}_{\boldsymbol{\phi}}^{-1}\mathbb{E}_{\boldsymbol{\phi}}\{\boldsymbol{\phi}\boldsymbol{\phi}^T\}\big\} \\ &\quad - \mathbf{m}_{\boldsymbol{\phi}}^T\boldsymbol{\Psi}_{\boldsymbol{\phi}}^{-1}\mathbb{E}_{\boldsymbol{\phi}}\{\boldsymbol{\phi}\} - \mathbb{E}_{\boldsymbol{\phi}}\{\boldsymbol{\phi}\}\boldsymbol{\Psi}_{\boldsymbol{\phi}}^{-1}\mathbf{m}_{\boldsymbol{\phi}} + \mathbf{m}_{\boldsymbol{\phi}}^T\boldsymbol{\Psi}_{\boldsymbol{\phi}}^{-1}\mathbf{m}_{\boldsymbol{\phi}}\Big) \\ &= -\frac{1}{2}\log\det(\boldsymbol{\Psi}_{\boldsymbol{\phi}}) - \frac{N_TN}{2}\log 2\pi - \frac{N_TN}{2};\end{aligned} \tag{48}$$

$$\int_{\mathbf{x}_d} \mathcal{Q}_4(\mathbf{x}_d)\log\mathcal{Q}_4(\mathbf{x}_d)d\mathbf{x}_d = 0. \tag{49}$$

For the next four terms in (25), they make use of the variational distributions (26)-(29) and the prior distributions (16)-(19). Straightforward computations give

$$\begin{aligned}&\int_{\boldsymbol{\beta}} \mathcal{Q}_1(\boldsymbol{\beta})\log p(\boldsymbol{\beta})d\boldsymbol{\beta} \\ &= -\text{Tr}\big\{\mathbf{R}_{\boldsymbol{\beta}}^{-1}(\mathbf{m}_{\boldsymbol{\beta}}\mathbf{m}_{\boldsymbol{\beta}}^H + \boldsymbol{\Psi}_{\boldsymbol{\beta}})\big\} - \log\det(\mathbf{R}_{\boldsymbol{\beta}}) - (Q+1)L\log\pi;\end{aligned} \tag{50}$$

$$\begin{aligned}
&\int_{\boldsymbol{\sigma}^2} \mathcal{Q}_2(\boldsymbol{\sigma}^2)\log p(\boldsymbol{\sigma}^2)d\boldsymbol{\sigma}^2 \\
&= \sum_{j=0}^{N_R-1}\sum_{k=0}^{N-1}\Big[a_{k,j}\log b_{k,j} + (-a_{k,j}-1)\mathbb{E}_{\sigma_j^2(k)}\big\{\log\sigma_j^2(k)\big\} \\
&\qquad - \mathbb{E}_{\sigma_j^2(k)}\Big\{\frac{1}{\sigma_j^2(k)}\Big\}b_{k,j} - \log\Gamma(a_{k,j})\Big] \\
&= \sum_{j=0}^{N_R-1}\sum_{k=0}^{N-1}\Big[a_{k,j}\log b_{k,j} + (-a_{k,j}-1)\big(\log\tilde{b}_{k,j} - \psi(\tilde{a}_{k,j})\big) \\
&\qquad - \frac{\tilde{a}_{k,j}}{\tilde{b}_{k,j}}b_{k,j} - \log\Gamma(a_{k,j})\Big]; \qquad (51)
\end{aligned}$$

$$\begin{aligned}
&\int_{\boldsymbol{\phi}} \mathcal{Q}_3(\boldsymbol{\phi})\log p(\boldsymbol{\phi})d\boldsymbol{\phi} \\
&= -\frac{1}{2}\mathrm{Tr}\big\{\mathbf{R}_{\boldsymbol{\phi}}^{-1}\mathbb{E}_{\boldsymbol{\phi}}\{\boldsymbol{\phi}\boldsymbol{\phi}^T\}\big\} - \frac{N_T N}{2}\log 2\pi - \frac{1}{2}\log\det(\mathbf{R}_{\boldsymbol{\phi}}) \\
&= -\frac{1}{2}\Big(\mathrm{Tr}\big\{\mathbf{R}_{\boldsymbol{\phi}}^{-1}\boldsymbol{\Psi}_{\boldsymbol{\phi}}\big\} + \mathbf{m}_{\boldsymbol{\phi}}^T\mathbf{R}_{\boldsymbol{\phi}}^{-1}\mathbf{m}_{\boldsymbol{\phi}}\Big) \\
&\quad - \frac{N_T N}{2}\log 2\pi - \frac{1}{2}\log\det(\mathbf{R}_{\boldsymbol{\phi}}); \qquad (52)
\end{aligned}$$

$$\begin{aligned}
&\int_{\mathbf{x}_d} \mathcal{Q}_4(\mathbf{x}_d)\log p(\mathbf{x}_d)d\mathbf{x}_d \\
&= \sum_{i=0}^{N_T-1}\sum_{k=0}^{N_d-1}\log\Big\{\sum_{\bar{x}_d^i(k)\in\mathfrak{c}}\delta\big(\tilde{x}_d^i(k)-\bar{x}_d^i(k)\big)\Big\} - N_T N_d\log\mathfrak{m}. \qquad (53)
\end{aligned}$$

For the last term of (25), we first compute the likelihood function. With the observation model defined in (13) and the Gaussian property of the NBI plus AWGN term, the log-likelihood function of $\mathbf{y}$ can be expressed as

$$\begin{aligned}
&\log p(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d) \\
&= -\sum_{j=0}^{N_R-1}\sum_{k=0}^{N-1}\log\sigma_j^2(k) - N_R N\log\pi - \mathbf{y}^H\mathbf{R}_{\mathbf{w}}^{-1}\mathbf{y} \\
&\quad + 2\Re\big\{\mathbf{y}^H\mathbf{R}_{\mathbf{w}}^{-1}\boldsymbol{\Xi}[\boldsymbol{\phi},\mathbf{x}]\boldsymbol{\beta}\big\} \\
&\quad - \mathrm{Tr}\big\{\boldsymbol{\Xi}^H[\boldsymbol{\phi},\mathbf{x}]\mathbf{R}_{\mathbf{w}}^{-1}\boldsymbol{\Xi}[\boldsymbol{\phi},\mathbf{x}]\boldsymbol{\beta}\boldsymbol{\beta}^H\big\}, \qquad (54)
\end{aligned}$$

where $\mathbf{R}_{\mathbf{w}} = \mathrm{Bldiag}\{\mathbf{R}_{\mathbf{w}}^0,...,\mathbf{R}_{\mathbf{w}}^{N_R-1}\}$ with $\mathbf{R}_{\mathbf{w}}^j = \mathrm{diag}\{\sigma_j^2(0),...,\sigma_j^2(N-1)\}$. Using (54), we have Eq. (55) at the top of the next page. Define $\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2} = \mathrm{Bldiag}\{\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}^0,...,\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}^{N_R-1}\}$ with $\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}^j = \mathrm{diag}\{\frac{\tilde{a}_{0,j}}{\tilde{b}_{0,j}},...,\frac{\tilde{a}_{N-1,j}}{\tilde{b}_{N-1,j}}\}$ and using the equivalent expressions between (13) and (15), we obtain Eq. (56) at the top of the next page, where $\boldsymbol{\rho}_{\boldsymbol{\phi}} = (\mathbf{1}_{N_T N} + \jmath\mathbf{m}_{\boldsymbol{\phi}})$. Since the covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{\phi}}$ is Hermitian, based on eigen-decomposition, we have $\boldsymbol{\Psi}_{\boldsymbol{\phi}} = \sum_{z=0}^{N_T N-1}\eta_z\boldsymbol{\mu}_z\boldsymbol{\mu}_z^T$ with η_z being the z^{th} eigenvalue of $\boldsymbol{\Psi}_{\boldsymbol{\phi}}$, and $\boldsymbol{\mu}_z$ being the z^{th} eigenvector, associated with η_z. By using the equivalent expressions between (13) and (15), the last term of (56) can be derived as Eq. (57) in the middle of the next page. Substituting (57) into (56), the last term in the KL divergence (25) can be obtained. Finally, substituting (46)-(53) and (56) into (25) and dropping those irrelevant terms leads to (30).

APPENDIX C
DERIVATION OF (33) AND (34)

Gathering those terms related to $\tilde{a}_{k,j}$ in (30) and computing the first derivative with respect to $\tilde{a}_{k,j}$, we obtain Eq. (58) in the middle of the next page, where $\psi'(\tilde{a}_{k,j}) = \frac{\partial\psi(\tilde{a}_{k,j})}{\partial\tilde{a}_{k,j}}$ with $\psi(\tilde{a}_{k,j}) = \frac{\partial\log\Gamma(\tilde{a}_{k,j})}{\partial\tilde{a}_{k,j}}$. Gathering those terms related to $\tilde{b}_{k,j}$ in (30) and computing the first derivative with respect to $\tilde{b}_{k,j}$, we obtain Eq. (59) at the bottom of the next page. Setting both (58) and (59) to zero and solving them simultaneously, after some straightforward computations we obtain (33) and (34).

REFERENCES

[1] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, and H. V. Poor, *MIMO Wireless Communications*. Cambridge University Press, 2007.

[2] R. V. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. Artech House Publishers, 2000.

[3] L. Hanzo, J. Akhtman, M. Jiang, and L. Wang, *MIMO-OFDM for LTE, WiFi and WiMAX: Coherent versus Non-coherent and Cooperative Turbo Transceivers*. Wiley, 2010.

[4] S. Cui, A. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, Aug. 2004.

[5] A. del Coso, U. Spagnolini, and C. Ibars, "Cooperative distributed MIMO channels in wireless sensor network," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 402–414, Feb. 2007.

[6] Q. Li and X. E. Lin, "Advancement of MIMO technology in WiMAX: from IEEE 802.16d/e/j to 802.16m," *IEEE Commun. Mag.*, vol. 47, no. 6, pp. 100–107, June 2009.

[7] R. Irmer *et al.*, "Multisite field trial for LTE and advanced concepts," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 92–98, Feb. 2009.

[8] L. Tomb, "On the effect of Wiener phase noise in OFDM systems," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 580–583, May 1998.

[9] D. Petrovic, W. Rave, and G. Fettweis, "Effects of phase noise on OFDM systems with and without PLL: characterization and compensation," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1607–1616, Aug. 2007.

[10] S. P. Yeh, S. Talwar, S. C. Lee, and H. Kim, "WiMAX femtocells: a perspective on network architecture, capacity, and coverage," *IEEE Commun. Mag.*, vol. 46, no. 10, pp. 58–65, Oct. 2008.

[11] D. Cabric, I. D. O'Donnel, M. S. W. Chen, and R. W. Brodersen, "Spectrum sharing radios," *IEEE Circuits Syst. Mag.*, vol. 6, pp. 30–45, July 2006.

[12] A. J. Coulson, "Bit error rate performance of OFDM in narrowband interference with excision filtering," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2484–2492, Sept. 2006.

[13] 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.

[14] Y. Gong and X. Hong, "OFDM joint data detection and phase noise cancellation for constant modulus modulations," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2864–2868, July 2009.

[15] 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.

[16] Y. Zhang and H. Li, "MIMO-OFDM systems in the presence of phase noise and doubly selective fading," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 2277–2285, July 2007.

[17] J. Tao, J. Wu, and C. Xiao, "Estimation of channel transfer function and carrier frequency offset for OFDM systems with phase noise," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4380–4387, Oct. 2009.

[18] F. Septier, Y. Delignon, A. Menhaj-Rivenq, and C. Garnier, "Monte Carlo methods for channel, phase noise, and frequency offset estimation with unknown noise variances in OFDM systems," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3613–3626, Aug. 2008.

[19] M. Morelli and M. Moretti, "Channel estimation in OFDM systems with unknown interference," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 5338–5347, Oct. 2009.

[20] L. Sanguinetti, M. Morelli, and H. V. Poor, "BICM decoding of jammed OFDM transmissions using the EM algorithm," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 2800–2806, Sept. 2011.

$$\int_{\mathbf{x}_d}\int_{\boldsymbol{\phi}}\int_{\boldsymbol{\sigma}^2}\int_{\boldsymbol{\beta}} \mathcal{Q}_1(\boldsymbol{\beta})\mathcal{Q}_2(\boldsymbol{\sigma}^2)\mathcal{Q}_3(\boldsymbol{\phi})\mathcal{Q}_4(\mathbf{x}_d)\log p(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d)d\boldsymbol{\beta}d\boldsymbol{\sigma}^2 d\boldsymbol{\phi}d\mathbf{x}_d$$
$$= -\sum_{j=0}^{N_R-1}\sum_{k=0}^{N-1}\mathbb{E}_{\sigma_j^2(k)}\big\{\log\sigma_j^2(k)\big\} - N_R N\log\pi - \mathbf{y}^H\mathbb{E}_{\boldsymbol{\sigma}^2}\{\mathbf{R}_{\mathbf{w}}^{-1}\}\mathbf{y} + 2\Re\big\{\mathbf{y}^H\mathbb{E}_{\boldsymbol{\sigma}^2}\{\mathbf{R}_{\mathbf{w}}^{-1}\}\int_{\boldsymbol{\phi}}\mathcal{Q}_3(\boldsymbol{\phi})\boldsymbol{\Xi}[\boldsymbol{\phi},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\mathbb{E}_{\boldsymbol{\beta}}\{\boldsymbol{\beta}\}d\boldsymbol{\phi}\big\}$$
$$- \mathrm{Tr}\big\{\int_{\boldsymbol{\phi}}\mathcal{Q}_3(\boldsymbol{\phi})\boldsymbol{\Xi}^H[\boldsymbol{\phi},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\mathbb{E}_{\boldsymbol{\sigma}^2}\{\mathbf{R}_{\mathbf{w}}^{-1}\}\boldsymbol{\Xi}[\boldsymbol{\phi},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\mathbb{E}_{\boldsymbol{\beta}}\{\boldsymbol{\beta}\boldsymbol{\beta}^H\}d\boldsymbol{\phi}\big\}. \quad (55)$$

$$\int_{\mathbf{x}_d}\int_{\boldsymbol{\phi}}\int_{\boldsymbol{\sigma}^2}\int_{\boldsymbol{\beta}} \mathcal{Q}_1(\boldsymbol{\beta})\mathcal{Q}_2(\boldsymbol{\sigma}^2)\mathcal{Q}_3(\boldsymbol{\phi})\mathcal{Q}_4(\mathbf{x}_d)\log p(\mathbf{y}|\boldsymbol{\beta},\boldsymbol{\sigma}^2,\boldsymbol{\phi},\mathbf{x}_d)d\boldsymbol{\beta}d\boldsymbol{\sigma}^2 d\boldsymbol{\phi}d\mathbf{x}_d$$
$$= -\sum_{j=0}^{N_R-1}\sum_{k=0}^{N-1}\big[\log\tilde{b}_{k,j} - \psi(\tilde{a}_{k,j})\big] - N_R N\log\pi - \mathbf{y}^H\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\mathbf{y} + 2\Re\big\{\mathbf{y}^H\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}\big\}$$
$$- \mathrm{Tr}\big\{\int_{\boldsymbol{\phi}}\mathcal{Q}_3(\boldsymbol{\phi})\boldsymbol{\Xi}^H[\boldsymbol{\phi},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\phi},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p](\mathbf{m}_{\boldsymbol{\beta}}\mathbf{m}_{\boldsymbol{\beta}}^H+\boldsymbol{\Psi}_{\boldsymbol{\beta}})d\boldsymbol{\phi}\big\}. \quad (56)$$

$$- \mathrm{Tr}\big\{\int_{\boldsymbol{\phi}}\mathcal{Q}_3(\boldsymbol{\phi})\boldsymbol{\Xi}^H[\boldsymbol{\phi},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\phi},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p](\mathbf{m}_{\boldsymbol{\beta}}\mathbf{m}_{\boldsymbol{\beta}}^H+\boldsymbol{\Psi}_{\boldsymbol{\beta}})\big\}$$
$$= -\mathbf{m}_{\boldsymbol{\beta}}^H\boldsymbol{\Xi}^H[\boldsymbol{\rho}_{\boldsymbol{\phi}},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}$$
$$- \sum_{z=0}^{N_T N-1}\eta_z\mathbf{m}_{\boldsymbol{\beta}}^H\boldsymbol{\Xi}^H[\boldsymbol{\mu}_z,\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\mu}_z,\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}$$
$$- \mathrm{Tr}\big\{\boldsymbol{\Xi}^H[\boldsymbol{\rho}_{\boldsymbol{\phi}},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}},\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Psi}_{\boldsymbol{\beta}}\big\}$$
$$- \mathrm{Tr}\big\{\sum_{z=0}^{N_T N-1}\eta_z\boldsymbol{\Xi}^H[\boldsymbol{\mu}_z,\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Lambda}_{\boldsymbol{\sigma}^2}\boldsymbol{\Xi}[\boldsymbol{\mu}_z,\mathbf{E}_d\tilde{\mathbf{x}}_d+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\Psi}_{\boldsymbol{\beta}}\big\}. \quad (57)$$

$$(\tilde{a}_{k,j}-a_{k,j}-1)\psi'(\tilde{a}_{k,j}) - 1 + \frac{1}{\tilde{b}_{k,j}}\Big[b_{k,j} + \big|\mathbf{y}(k+jN)\big|^2 - 2\Re\Big\{\mathbf{y}^*(k+jN)\big[\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}}^{q-1},\mathbf{E}_d\tilde{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}^q\big](k+jN)\Big\}$$
$$+ \Big|\big[\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}}^{q-1},\mathbf{E}_d\tilde{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}^q\big](k+jN)\Big|^2 + \sum_{z=0}^{N_T N-1}\eta_z^{q-1}\Big|\big[\boldsymbol{\Xi}[\boldsymbol{\mu}_z^{q-1},\mathbf{E}_d\tilde{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}^q\big](k+jN)\Big|^2$$
$$+ \sum_{r=0}^{(Q+1)L-1}\lambda_r^q\Big|\big[\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}}^{q-1},\mathbf{E}_d\tilde{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\nu}_r^q\big](k+jN)\Big|^2 + \sum_{z=0}^{N_T N-1}\sum_{r=0}^{(Q+1)L-1}\eta_z^{q-1}\lambda_r^q\Big|\big[\boldsymbol{\Xi}[\boldsymbol{\mu}_z^{q-1},\mathbf{E}_d\tilde{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\nu}_r^q\big](k+jN)\Big|^2\Big]. \quad (58)$$

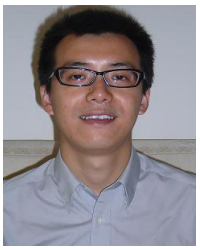
$$\frac{1}{\tilde{b}_{k,j}}(a_{k,j}+1) - \frac{\tilde{a}_{k,j}}{\tilde{b}_{k,j}^2}\Big[b_{k,j} + \big|\mathbf{y}(k+jN)\big|^2 - 2\Re\Big\{\mathbf{y}^*(k+jN)\big[\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}}^{q-1},\mathbf{E}_d\tilde{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}^q\big](k+jN)\Big\}$$
$$+ \Big|\big[\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}}^{q-1},\mathbf{E}_d\tilde{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}^q\big](k+jN)\Big|^2 + \sum_{z=0}^{N_T N-1}\eta_z^{q-1}\Big|\big[\boldsymbol{\Xi}[\boldsymbol{\mu}_z^{q-1},\mathbf{E}_d\tilde{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\mathbf{m}_{\boldsymbol{\beta}}^q\big](k+jN)\Big|^2$$
$$+ \sum_{r=0}^{(Q+1)L-1}\lambda_r^q\Big|\big[\boldsymbol{\Xi}[\boldsymbol{\rho}_{\boldsymbol{\phi}}^{q-1},\mathbf{E}_d\tilde{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\nu}_r^q\big](k+jN)\Big|^2 + \sum_{z=0}^{N_T N-1}\sum_{r=0}^{(Q+1)L-1}\eta_z^{q-1}\lambda_r^q\Big|\big[\boldsymbol{\Xi}[\boldsymbol{\mu}_z^{q-1},\mathbf{E}_d\tilde{\mathbf{x}}_d^{q-1}+\mathbf{E}_p\mathbf{x}_p]\boldsymbol{\nu}_r^q\big](k+jN)\Big|^2\Big]. \quad (59)$$

[21] F. Li, S. Zhu, and M. Rong, "Detection for OFDM systems with channel estimation errors using variational inference," *IEEE Signal Process. Lett.*, vol. 16, no. 5, pp. 434–437, May 2009.

[22] X. Y. Zhang, D. G. Wang, and J. B. Wei, "Joint symbol detection and channel estimation for MIMO-OFDM systems via the variational bayesian EM algorithm," in *Proc. 2008 IEEE Wireless Commun. Netw. Conf.*, pp. 13–17.

[23] L. He, Y. Wu, S. Ma, T. Ng, and H. V. Poor, "Superimposed training-based channel estimation and data detection for OFDM amplify-and-forward cooperative systems under high mobility," *IEEE Trans. Signal*

Process., vol. 60, no. 1, pp. 274–284, Jan. 2012.
[24] R. Corvaja and A. G. Armad, "Joint channel and phase noise compensation for OFDM in fast-fading multipath applications," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 636–643, Feb. 2009.
[25] F. Merli, X. Wang, and G. Vitetta, "A Bayesian multiuser detection algorithm for MIMO-OFDM systems affected by multipath fading, carrier frequency offset, and phase noise," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 506–516, Apr. 2008.
[26] S. Stefanatos and A. K. Katsaggelos, "Joint data detection and channel tracking for OFDM systems with phase noise," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4230–4243, Sept. 2008.
[27] B. Frey, *Graphical Models for Machine Learing and Digital Communication*.: MIT Press, 1998.
[28] V. Smidl and A. Quinn, *The Variational Bayes Method in Signal Processing*. Springer-Verlag, 2006.
[29] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
[30] X. Chen, H. Hu, H. Wang, H.-H. Chen, and M. Guizani, "Double proportional fair user pairing algorithm for uplink virtual MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2425–2429, July 2008.
[31] M. Ruder, U. Dang, and W. Gerstacker, "User pairing for multiuser SC-FDMA transmission over virtual MIMO ISI channels," in *Proc. 2009 IEEE Global Telecommun. Conf.*, pp. 1–7.
[32] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
[33] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, May 2003.
[34] 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.
[35] T. Zemen and C. F. Mecklenbrauker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3597–3607, Sept. 2005.
[36] P. Rabiei, W. Namgoong, and N. Al-Dhahir, "A non-iterative technique for phase noise ICI mitigation in packet-based OFDM systems," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5945–5950, Nov. 2010.
[37] S. H. Choi, P. Smith, B. Allen, W. Q. Malik, and M. Shafi, "Severely fading MIMO channels: models and mutual information," in *Proc. 2007 IEEE International Conf. Commun.*, pp. 4628–4633.
[38] J. Wang, A. Dogandžić, and A. Nehorai, "Maximum likelihood estimation of compound-Gaussian clutter and target parameters," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3884–3898, Oct. 2006.
[39] A. Doucet and X. Wang, "Monte Carlo methods for signal processing: a review in the statistical signal processing context," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 152–170, Nov. 2005.
[40] S. Haykin, J. C. Principe, T. J. Sejnowski, and J. McWhirter, *New Directions in Statistical Signal Processing: From Systems to Brains.* MIT Press, 2005.
[41] A. Kannu and P. Schniter, "Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1148–1160, Mar. 2008.



Ke Zhong received the B.Eng. degree in 2008, from Chengdu University of Technology, Chengdu, China.

Since September 2008, he has been in successive postgraduate and doctoral programs of study at the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China (UESTC), Chengdu, China.

He was a visiting researcher at the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, from 2011 to 2012. His research interests include physical layer algorithms for wireless 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 Croucher Foundation scholarship in 2002 to study Ph.D. degree at Texas A&M University, College Station, and graduated 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 HKU, currently as an Associate Professor. He was a visiting scholar at Princeton University, in summer 2011. His research interests are in general area of signal processing and communication systems, and in particular distributed signal processing and communications; optimization theories for communication systems; estimation and detection theories in transceiver designs; and smart grid.

Dr. Wu served as an Editor for IEEE COMMUNICATIONS LETTERS, is currently an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS and *Journal of Communications and Networks*.

Shaoqian Li received the B.S.E. degree in communication technology from Northwest Institute of Telecommunication (renamed Xidian University) in 1982 and the M.S.E. degree from the University of Electronic Science and Technology of China (UESTC) in 1984.

He is now a Professor and the Director of the National Key Laboratory of Science and Technology on Communications in UESTC. He is also a member of National High Technology R&D Program, which is also known as 863 Program Communications Group. His research interests include wireless information and communication theory, mobile and personal communications, anti-interference technology in wireless communications, spread spectrum, and frequency-hopping techniques.