Received 5 December 2022; accepted 19 December 2022. Date of publication 22 December 2022; date of current version 4 January 2023.
*Digital Object Identifier 10.1109/OJCOMS.2022.3231385*

# Auxiliary Variable-Aided Hybrid Message Passing for Joint Channel, Phase Noise Estimation and Detection for Multi-User MIMO Systems

**XIAOCHEN HE [1], WEI WANG [2] (Member, IEEE), QINGHUA GUO [1] (Senior Member, IEEE), JUN TONG [1], JIANGTAO XI [1] (Senior Member, IEEE), AND YANGUANG YU [1] (Senior Member, IEEE)**

[1] School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW 2522, Australia
[2] School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China
CORRESPONDING AUTHOR: X. HE (e-mail: xh994@uowmail.edu.au)

**ABSTRACT** This paper considers a multi-user multiple-input multiple-output (MU-MIMO) system with independent phase noise (PN) at user terminals (UTs) due to non-synchronous noisy local oscillators. The conventional factor graph of MU-MIMO involves an observation factor with multiple high-dimensional variables in the form of multiplication and summation. In this work, the MU-MIMO factor graph is represented by introducing auxiliary variables, which enables the decomposition of the observation factor into several simpler sub-factors and then use different message passing techniques to obtain approximate marginals. Specifically, belief propagation and expectation propagation (BP-EP) are used for the linear factors for random walk process, modulation and coding, while mean field (MF) is used for non-linear factor in exponential form. To reduce the computational complexity, the non-Gaussian MF messages are approximated to be Gaussian by using the second order Taylor expansion at its belief obtained in the previous iteration. The proposed receiver has a quadratic computational complexity per symbol per iteration. Simulation results show that the performance of the proposed receiver is better than existing methods and is close to the matched filter bound (MFB) in terms of the bit error rate (BER).

**INDEX TERMS** MU-MIMO, phase noise, hybrid message passing, joint channel estimation and detection, iterative receiver.

## I. INTRODUCTION

PHASE noise (PN) is introduced by the unstable circuitry of local oscillators during up-conversion of baseband signals to passband and vice versa. PN is one of main limiting factors in wireless communications as it may severely degrade system performance. The effect of PN has been continuously studied [1], [2]. Most uplink MU-MIMO detection methods are based on the linear observation model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, which assumes ideal hardware components. When PN is modelled as a Random Walk (Wiener) process as in [3], [4], [5], the presence of PN results in a non-linear observation model. Conventionally, the non-linear observation model with PN can be linearized by using the first-order Taylor expansion and resolved by the soft-input extended Kalman smoothing (Soft-input EKS) algorithm to track time-varying PN over a frame. Soft-input EKS has been studied for MIMO systems [6], [7], [8], [9]. Alternatively, it is straightforward to aggregate the PN and channel as an equivalent time-varying channel and then use the existing joint channel estimation and detection approaches [10], [11], [12], [13], [14].

In this work, we consider the uplink MU-MIMO with non-synchronous PN at UTs. Particularly, we focus on developing an iterative receiver with an efficient algorithm for joint PN estimation, channel estimation and detection. The optimum maximum a posteriori (MAP) detection requires to explicitly compute the posteriori distribution of the transmitted symbols of interest, which involves marginalizing out continuous parameters (i.e., PN and channel) and discrete variables (i.e., transmitted symbols of other UTs) from a joint distribution, given the observations, which is computationally intractable. In recent years, message passing techniques have been widely applied to receiver designs. Belief propagation (BP) is the exact inference when factor graphs have tree structures. BP works well when the graphs have large loops. BP is commonly used when all the involved variables are discrete or Gaussian in linear systems [15], [16]. However, BP becomes intractable when both continuous and discrete variables are involved. In this case, we may need to use approximate inference techniques such as expectation propagation (EP) [20] and variational message passing (VMP) sometimes called mean field (MF) [18]. The convergence of VMP is guaranteed and it has simple message updating rules for conjugate-exponential models. EP can be regarded as an approximation to BP, which may have faster convergence speed. To deal with complex problems, combinations of these inference techniques can be used, so that their advantages can be exploited while drawbacks be avoided. For example, BP, EP and MF have been jointly used to design receivers, such as the work based on BP-MF [12], [21], BP-EP [22] and BP-MF-EP [5], [19]. Inspired by these hybrid message passing schemes and with the aid of auxiliary variables, we extend the work in [10] to MU-MIMO systems. By introducing auxiliary variables, the time-varying variables (PN and transmitted symbols) and quasi-static channel are separated in different factors, of which the incoming/outgoing messages are computed by different message passing rules. Specifically, BP is used to deal with the PN process, modulation and coding, EP is used to approximate Gaussian mixture and MF is used to deal with the factors in exponential form. To reduce the computational complexity, the MF messages of PNs are approximated as Gaussian by using the second-order Taylor expansion at its belief obtained in the previous iteration.

*Notation*: The lower and upper case of bold letters represents column vectors $\mathbf{a}$ and matrices $\mathbf{A}$, respectively. The boldness of letters distinguishes vectors/matrices $\mathbf{a}/\mathbf{A}$ with scalars $a$. The superscripts $(\cdot)^\star$, $(\cdot)^T$ and $(\cdot)^H$ represent conjugate, transpose and conjugate transpose, respectively. $\mathbf{A} \odot \mathbf{B}$ represents the Hadamard product of matrices $\mathbf{A}$ and $\mathbf{B}$. The symbol $\propto$ denote the equality of functions up to a scale factor. The real/imaginary part of a complex number $x$ is denoted by $\mathrm{Re}[x]/\mathrm{Im}[x]$. A complex Gaussian random variable $x$ with mean $\mu_x$ and variance $v_x$ is represented by $x \sim \mathcal{CN}(x; \mu_x, v_x)$. $\langle f(x)\rangle_{p(x)}$ denotes the expectation of a function $f(x)$ with respect to a density $p(x)$, i.e., $\langle f(x)\rangle_{p(x)} = \int_x f(x)p(x)dx$. The incoming message from a factor node $f$ to a variable node $x$ as $m_{f\to x}(x)$ and the message out of a variable node $x$ to a factor node $f$ as $n_{x\to f}(x)$. The arrow above messages such as $\vec{m}_{f\to x}(x)$ indicates the direction of message. The belief of a variable $x$ is the product of the incoming and outgoing messages, i.e., $b(x) = m_{f\to x}(x) \cdot n_{x\to f}(x)$.

## II. SYSTEM MODEL

Consider a coded $M \times N$ uplink MU-MIMO system with $M$ denoting the number of BS antennas and $N$ the number of single-antenna UTs, as shown in Fig. 1. At the $n$th UT,

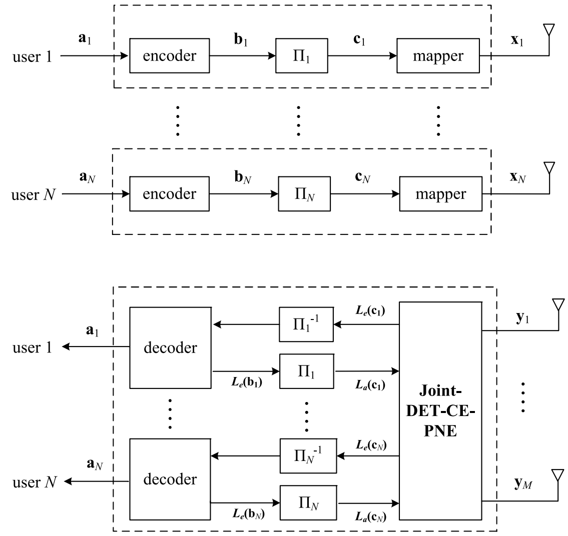



**FIGURE 1.** Block diagram of a coded MU-MIMO system. $\Pi$ and $\Pi^{-1}$ denote random interleaver and de-interleaver, respectively.

information bits $\mathbf{a}_n$ are encoded and interleaved into a code sequence $\mathbf{c}_n$ (we assume individual encoders, interleavers and modulators at UTs). Then $\mathbf{c}_n$ is divided into a number of length-Q sub-sequences $\mathbf{c}_n^t = [c_{n,1}^t, c_{n,2}^t, \ldots, c_{n,Q}^t]$ and each $\mathbf{c}_n^t$ is mapped into a symbol $x_n^t \in \mathcal{A}$. The symbol alphabet $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_{2^Q}\}$ represents a $2^Q$-ary constellation point sets with $\sum_{i=1}^{2^Q} \alpha_i = 0$ and $2^{-Q}\sum_{i=1}^{2^Q} |\alpha_i|^2 = 1$. Each constellation point $\alpha_i$ corresponds to a binary vector (label) $\mathbf{b}_i = [b_{i,1}, b_{i,2}, \ldots, b_{i,Q}]$. Assume independent PNs at each UT $\theta^t = [\theta_1^t, \theta_2^t, \ldots, \theta_N^t]^T$, perfect timing and frequency synchronization, UTs send a $N \times 1$ symbol vector $\mathbf{x}^t = [x_1^t, x_2^t, \ldots, x_N^t]^T$ for each channel use, the received signals at the $t$th time interval at the $m$th BS antenna can be modelled as

$$y_m^t = \mathbf{h}_m\left(e^{j\theta^t} \odot \mathbf{x}^t\right) + w_m^t, \forall t \in \left[T_p + 1, T_c\right] \qquad (1)$$

where $\mathbf{h}_m = [h_{m1}, h_{m2}, \ldots, h_{mN}]$ represents the channel gains between $K$ UTs and the $m$th BS antenna and is static within a length-$T_c$ coherence block, $T_d$ denotes the length of data ($T_p = T_c - T_d$ denotes the length of pilots for initial channel estimation), $w_m^t \sim \mathcal{N}(0, \sigma_w^2)$ denotes the additive white Gaussian noise (AWGN) at the $m$th BS antenna, and the PN $\theta_n^t$ is modelled as a random-walk (Wiener) process,

$$\theta_n^t = \theta_n^{t-1} + \Delta_n^t, \qquad (2)$$

where the PN increment is real-valued white Gaussian, i.e., $\Delta_n^t \sim \mathcal{N}(0, \sigma_{\Delta_n}^2)$. The variance of the PN increment $\sigma_{\Delta_n}^2$ determines the quality of the oscillators referred to [28]. In practice, $\sigma_\Delta^2 \approx 10^{-3} - 10^{-2}$ in rad$^2$ suggests strong PN whereas $\sigma_\Delta^2 \approx 10^{-5} - 10^{-4}$ indicates moderate-to-weak PN [23]. In the rest of paper, we consider the same quality of oscillators at UTs, i.e., $\sigma_\Delta^2 = \sigma_{\Delta_n}^2, \forall n$.

Based on the received signal model in (1) and the phase noise model in (2), we aim to estimate the transmitted symbol vector $\mathbf{x}^t$ at a low complexity under unknown PN and

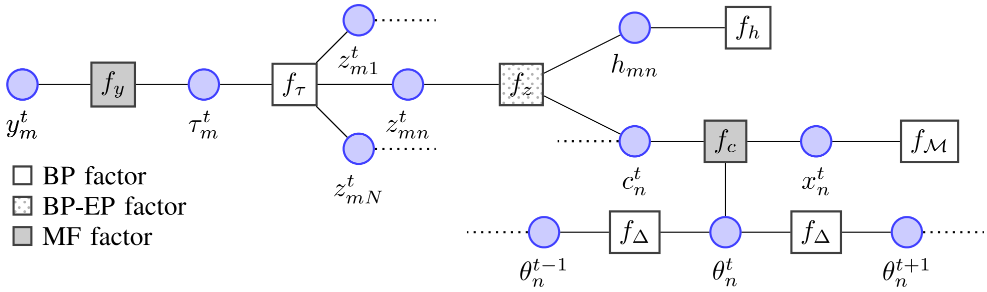



**FIGURE 2.** Factor graph of (3) with $f_o$ in (7).

imperfect estimation of the channel. We achieve this by developing a message passing-based algorithm that iteratively refines the estimates of the transmitted symbols $\mathbf{x}^t$, the PN $\theta^t$ and the channel matrix $\mathbf{H}$. Note that the AWGN variance $\sigma_w^2$ is unknown in practice but can be accurately estimated using preambles [27]. An approach based on BP-VMP was proposed for joint decoding, phase noise estimation and AWGN precision (the reciprocal of the variance) estimation over the AWGN channel [23]. In this paper, we assume $\sigma_w^2$ is known to focus on the treatment of PN and channel estimation and refer interested readers to [23] for the approach to estimate $\sigma_w^2$.

### A. PROBABILISTIC MODEL

Based on the received signal model in (1), the joint probability of $\mathbf{X}, \Theta, \mathbf{H}, \mathbf{Y}$ can be factorized as

$$\begin{aligned} &p(\mathbf{X}, \Theta, \mathbf{H}, \mathbf{Y}) \\ &\quad = p(\mathbf{X}, \Theta, \mathbf{H}|\mathbf{Y})p(\mathbf{Y}) \\ &\quad \propto \prod_{t=1}^{T_d} p\left(\mathbf{x}^t, \theta^t, \mathbf{H}|\mathbf{y}^t\right) \\ &\quad \propto \prod_{t=1}^{T_d} p\left(\mathbf{y}^t|\mathbf{x}^t, \theta^t, \mathbf{H}\right)p\left(\mathbf{x}^t, \theta^t, \mathbf{H}\right) \\ &\quad \propto \prod_{t=1}^{T_d} p\left(\mathbf{y}^t|\mathbf{x}^t, \theta^t, \mathbf{H}\right)\prod_{n=1}^{N}\prod_{t=1}^{T_d} p\left(\theta_n^t|\theta_n^{t-1}\right)\prod_{t=1}^{T_d} P\left(\mathbf{x}^t\right)p(\mathbf{H}) \\ &\quad \propto \prod_{t=1}^{T_d}\prod_{m=1}^{M} \underbrace{p\left(y_m^t|\mathbf{x}^t, \theta^t, \mathbf{h}_m\right)}_{f_o}\prod_{t=1}^{T_d} \underbrace{p\left(\theta_n^t|\theta_n^{t-1}\right)}_{f_\Delta}\prod_{t=1}^{T_d} P\left(\mathbf{x}^t\right)p(\mathbf{H}), \end{aligned} \tag{3}$$

where $\mathbf{H} \triangleq [\mathbf{h}_1, \ldots, \mathbf{h}_M]$, $\mathbf{X} \triangleq [\mathbf{x}^{T_p+1}, \ldots, \mathbf{x}^{T_c}]$, $\Theta \triangleq [\theta^{T_p+1}, \ldots, \theta^{T_c}]$, $\mathbf{Y} \triangleq [\mathbf{y}^{T_p+1}, \ldots, \mathbf{y}^{T_c}]$, $f_o \triangleq \mathcal{CN}(y_m^t; \mathbf{h}_m(e^{j\theta^t} \odot \mathbf{x}^t), \sigma_w^2)$, $f_\Delta \triangleq \mathcal{N}(\theta_n^t; \theta_n^{t-1}, \sigma_\Delta^2)$.

### B. FACTOR GRAPH REPRESENTATION WITH AUXILIARY VARIABLES

Define the following auxiliary variables

$$c_n^t = e^{j\theta_n^t}x_n^t, \tag{4}$$

$$z_{mn}^t = h_{mn}c_n^t, \tag{5}$$

$$\tau_m^t = \Sigma_n z_{mn}^t, \tag{6}$$

the observation factor $f_o$ can be further factorized as

$$f_o = f_y f_\tau f_z f_c, \tag{7}$$

where

$$f_c = p\left(c_n^t|x_n^t, \theta_n^t\right) = \delta\left(c_n^t - e^{j\theta_n^t}x_n^t\right), \tag{8}$$

$$f_z = p\left(z_{mn}^t|h_{mn}, c_n^t\right) = \delta\left(z_{mn}^t - h_{mn}c_n^t\right), \tag{9}$$

$$f_\tau \triangleq p\left(\tau_m^t|\mathbf{z}_m^t\right) = \delta\left(\tau_m^t - \Sigma_n z_{mn}^t\right), \tag{10}$$

$$f_y \triangleq p\left(y_m^t|\tau_m^t, \sigma_w^2\right) = \mathcal{CN}\left(y_m^t; \tau_m^t, \sigma_w^2\right). \tag{11}$$

Fig. 2 shows the factor graph of (3) with $f_o$ defined in (7). Note that in this factor graph, the circles represent variable nodes and the squares represent factor nodes.

## III. ITERATIVE RECEIVER WITH HYBRID MESSAGE PASSING

The nature of message passing facilitates the iterative receiver design, which generally consists of a soft MIMO detector and $N$ soft decoders. The detector and the decoders work iteratively and exchange extrinsic information in terms of log-likelihood ratios (LLRs) of coded bits. In this work, the proposed algorithm enables the soft detection scheme to compute the posteriori LLRs of coded bits given the estimate of channel and PN. Considering the discrete nature of the coded symbols, Gaussian assumption for the channel and the Gaussian approximation to PN, it is expected to come up with some messages in Gaussian mixture form and some other messages in exponential form that is computational intractable. We adopt the approximate inference combining BP-EP to update the former as in [5], [22] and the MF to update the latter as in [5], [21]. We firstly denote the prior LLR of the coded bits at the $n$th soft decoder as $L_a(c_{n,q}^t)$, the soft mapping operation below is needed to carry out to obtain the prior p.m.f of transmitted symbols $P_a(x_n^t)$ as below ([24], [25]):

$$\overleftarrow{P}_a\left(x_n^t = \alpha_i\right) = \prod_{q=1}^{Q}\frac{1}{2}\left(1 + \tilde{b}_{i,q}\tanh\left(L_a\left(c_{n,q}^t\right)/2\right)\right), \tag{12}$$

where

$$\tilde{b}_{i,q} = \begin{cases} +1, & b_{i,q} = 0 \\ -1, & b_{i,q} = 1 \end{cases}$$

Then the message of $x_n^t$ from $f_{\mathcal{M}}$ can be written as

$$m_{f_{\mathcal{M}}\to x}\left(x_n^t\right) = \sum_{\alpha_i\in\mathcal{A}} \overleftarrow{P}_a\left(x_n^t = \alpha_i\right)\delta\left(x_n^t - \alpha_i\right). \tag{13}$$

Assume the message of $x_n^t$ from $f_c$ is Gaussian, i.e.,

$$m_{f_c\to x}\left(x_n^t\right) = \mathcal{CN}\left(x_n^t; \vec{\mu}_{x_n^t}, \vec{v}_{x_n^t}\right), \tag{14}$$

the belief p.m.f of $x_n^t$ is approximated as follows:

$$P_x\left(x_n^t = \alpha_i\right) = \frac{1}{\mathcal{K}}\overleftarrow{P}_a\left(x_n^t = \alpha_i\right)\exp\left\{\frac{-|\alpha_i - \vec{\mu}_{x_n^t}|^2}{\vec{v}_{x_n^t}}\right\}, \tag{15}$$

and $\mathcal{K}$ is a normalization constant.

### *A. MESSAGE PASSING FOR DETECTION*

This subsection provides the details of the soft detection scheme for computing the posteriori LLRs of the coded bits given the assumptions below about the estimates of the channel and PN. The update of channel and PN estimates will be explained in the latter subsections. Based on the factor graph illustrated in Fig. 2, to compute the message of $x_n^t$ from $f_c$, we start with making the following assumptions for initial messages:

$$b\left(\theta_n^t\right) = \mathcal{CN}\left(\theta_n^t; \hat{\theta}_n^t, v_{\theta_n^t}\right), \tag{16}$$

$$n_{h\to f_z}(h_{mn}) = \mathcal{CN}\left(h_{mn}; \overleftarrow{\mu}_{h_{mn}^t}, \overleftarrow{v}_{h_{mn}^t}\right), \tag{17}$$

$$n_{z\to f_z}\left(z_{mn}^t\right) = \mathcal{CN}\left(z_{mn}^t; \vec{\mu}_{z_{mn}^t}, \vec{v}_{z_{mn}^t}\right). \tag{18}$$

Given (16)(15) and $f_c$ defined by (8), the message of $c_n^t$ to $f_z$ can be obtained by

$$n_{c\to f_z}\left(c_n^t\right) = \sum_{x_n^t\in\mathcal{A}} P_x\left(x_n^t\right)\delta\left(c_n^t - e^{j\hat{\theta}_n^t}x_n^t\right), \tag{19}$$

the BP message of $z_{mn}^t$ from $f_z$ can be computed as follows:

$$\begin{aligned} m_{f_z\to z}^{\mathrm{BP}}\left(z_{mn}^t\right) &= \left\langle f_z\left(z_{mn}^t\right)\right\rangle_{n_{c\to f_z}(c_n^t)n_{h\to f_z}(h_{mn})} \\ &= \sum_{x_n^t\in\mathcal{A}} P_x\left(x_n^t\right)|x_n^t|^{-1}\mathcal{CN}\left(z_{mn}^t; \overleftarrow{\mu}_{h_{mn}}e^{j\hat{\theta}_n^t}x_n^t, |e^{j\hat{\theta}_n^t}x_n^t|^2\overleftarrow{v}_{h_{mn}^t}\right). \end{aligned} \tag{20}$$

This BP message is in a form of Gaussian mixture. For the convenience of computing the backward message of $z_{mn}^t$ to $f_z$, i.e., $n_{z\to f_z}(z_{mn}^t) = m_{f_\tau\to z}(z_{mn}^t)$ with low complexity, we project this Gaussian mixture-form BP message into Gaussian via EP as below (we note that such Gaussian approximation has been widely shown effective in the message passing inference, e.g., [10]):

$$\begin{aligned} b\left(z_{mn}^t\right) &= \mathrm{Proj}_g\left\{m_{f_z\to z}^{\mathrm{BP}}\left(z_{mn}^t\right)n_{z\to f_z}\left(z_{mn}^t\right)\right\} \\ &\triangleq \mathcal{CN}\left(z_{mn}^t; \hat{z}_{mn}^t, v_{z_{mn}^t}\right), \end{aligned} \tag{21}$$

where

$$\hat{z}_{mn}^t = \sum_{x_n^t\in\mathcal{A}} \psi_{z_{mn}^t}\left(x_n^t\right)\mu_{z_{mn}^t}\left(x_n^t\right), \tag{22}$$

$$v_{z_{mn}^t} = \sum_{x_n^t\in\mathcal{A}} \psi_{z_{mn}^t}\left(x_n^t\right)\left(|\mu_{z_{mn}^t}\left(x_n^t\right)|^2 + \tau_{z_{mn}^t}\left(x_n^t\right)\right) - |\hat{z}_{mn}^t|^2, \tag{23}$$

and

$$\mu_{z_{mn}^t}\left(x_n^t\right) = \frac{|e^{j\hat{\theta}_n^t}x_n^t|^2\overleftarrow{v}_{h_{mn}}\vec{\mu}_{z_{mn}^t} + \vec{v}_{z_{mn}^t}\overleftarrow{\mu}_{h_{mn}}e^{j\hat{\theta}_n^t}x_n^t}{|e^{j\hat{\theta}_n^t}x_n^t|^2 v_{h_{mn}} + \vec{v}_{z_{mn}^t}}, \tag{24}$$

$$\tau_{z_{mn}^t}\left(x_n^t\right) = \frac{|e^{j\hat{\theta}_n^t}x_n^t|^2\overleftarrow{v}_{h_{mn}}\vec{v}_{z_{mn}^t}}{|e^{j\hat{\theta}_n^t}x_n^t|^2\overleftarrow{v}_{h_{mn}} + \vec{v}_{z_{mn}^t}}, \tag{25}$$

$$\psi_{z_{mn}^t}\left(x_n^t\right) = \frac{1}{\mathcal{K}}P_x\left(x_n^t\right)|e^{j\hat{\theta}_n^t}|^2|x_n^t|^{-1}. \tag{26}$$

and $\mathcal{K}$ is a normalization constant. Note that $|e^{j\hat{\theta}_n^t}x_n^t|^2 = 1$ due to the modulation constraint $|x_n^t|^2 = 1$. In addition, the EP approximation may result in negative $v_{z_{mn}^t}$, which will significantly degrades the algorithm performance. To avoid the negative variance problem, we replace $v_{z_{mn}^t}$ with its absolute value $|v_{z_{mn}^t}|$, then the mean and variance of the EP message

$$m_{f_z\to z}^{\mathrm{EP}}\left(z_{mn}^t\right) = \mathcal{CN}\left(z_{mn}^t; \overleftarrow{\mu}_{z_{mn}^t}, \overleftarrow{v}_{z_{mn}^t}\right) \tag{27}$$

are given by

$$\overleftarrow{v}_{z_{mn}^t} = 1/\left(1/v_{z_{mn}^t} - 1/\vec{v}_{z_{mn}^t}\right); \tag{28}$$

$$\overleftarrow{\mu}_{z_{mn}^t} = \overleftarrow{v}_{z_{mn}^t}\left(\hat{z}_{mn}^t/v_{z_{mn}^t} - \vec{\mu}_{z_{mn}^t}/\vec{v}_{z_{mn}^t}\right), \tag{29}$$

which is the message of $z_{mn}^t$ passing to the factor $f_\tau$, i.e.,

$$n_{z\to f_\tau}\left(z_{mn}^t\right) = m_{f_z\to z}^{\mathrm{EP}}\left(z_{mn}^t\right). \tag{30}$$

Given the observation $y_m^t$ and assume the AWGN variance $\sigma_w^2$ is known, the BP message of $\tau_m$ to the factor $f_\tau$ is

$$n_{\tau\to f_\tau}\left(\tau_m^t\right) = \mathcal{CN}\left(\tau_m^t; y_m^t, \sigma_w^2\right). \tag{31}$$

Then, the message of $z_{mn}^t$ to $f_z$ can be updated as follows:

$$\begin{aligned} n_{z\to f_z}\left(z_{mn}^t\right) &= m_{f_\tau\to z}^{\mathrm{BP}}\left(z_{mn}^t\right) \\ &= \left\langle \delta\left(\tau_m^t - \sum_n z_{mn}^t\right)\right\rangle_{n_{\tau\to f_\tau}(\tau_m^t)\prod_{n'\neq n} n_{z\to f_\tau}\left(z_{mn'}^t\right)} \\ &= \mathcal{CN}\left(z_{mn}^t; y_m^t - \sum_{n'\neq n}\overleftarrow{\mu}_{z_{mn'}^t}, \sigma_w^2 + \sum_{n'\neq n}\overleftarrow{v}_{z_{mn'}^t}\right) \\ &\triangleq \mathcal{CN}\left(z_{mn}^t; \vec{\mu}_{z_{mn}^t}, \vec{v}_{z_{mn}^t}\right). \end{aligned} \tag{32}$$

The message from $f_z$ to $c_n^t$ can be updated as follows:

$$\begin{aligned} m_{f_z\to c}^{\mathrm{BP}}\left(c_n^t\right) &= \left\langle \delta\left(z_{mn}^t - h_{mn}c_n^t\right)\right\rangle_{n_{z\to f_z}(z_{mn}^t)n_{h\to f_z}(h_{mn})} \\ &= \frac{1}{|c_n^t|^2 v_{h_{mn}} + \vec{v}_{z_{mn}^t}}\exp\left\{-\frac{|\vec{\mu}_{z_{mn}^t} - c_n^t\overleftarrow{\mu}_{h_{mn}}|^2}{|c_n^t|^2\overleftarrow{v}_{h_{mn}} + \vec{v}_{z_{mn}^t}}\right\} \\ &\propto \mathcal{CN}\left(c_n^t; \frac{\vec{\mu}_{z_{mn}^t}\overleftarrow{\mu}_{h_{mn}}^*}{|\overleftarrow{\mu}_{h_{mn}}|^2}, \frac{\vec{v}_{z_{mn}^t} + \overleftarrow{v}_{h_{mn}}}{|\overleftarrow{\mu}_{h_{mn}}|^2}\right) \\ &\triangleq \mathcal{CN}\left(c_n^t; \vec{\mu}_{c_{mn}^t}, \vec{v}_{c_{mn}^t}\right). \end{aligned} \tag{33}$$

Note that this message in (33) is Gaussian only when equal symbol energy modulation (e.g., QPSK) is employed so that $|c_n^t|^2 = |e^{j\theta_n^t} x_n^t|^2 = 1$. The computation of (33) only counts on the messages from the $m$th observation $y_m^t$. In fact, the message from every BS observation can be utilized because the PN-impaired data symbol from each UT can be seen by all BS antennas with different levels of distortion. Correspondingly, the variable node $c_n^t$ should be connected with $M$ $f_z$ factors that are associated with all observations (Fig. 2 only shows the $m$th observation for simplicity) and message of $c_n^t$ from all observed signals are

$$n_{c\to f_c}(c_n^t) = \mathcal{CN}\left(c_n^t; \frac{1}{M}\sum_m \vec{\mu}_{c_{mn}^t}, \frac{1}{M}\sum_m \vec{v}_{v_{mn}^t}\right), \triangleq \mathcal{CN}\left(c_n^t; \vec{\mu}_{c_n^t}, \vec{v}_{c_n^t}\right). \tag{34}$$

where $\vec{\mu}_{c_n^t} \triangleq \frac{1}{M}\sum_m \vec{\mu}_{c_{mn}^t}$ and $\vec{v}_{c_n^t} = \frac{1}{M}\sum_m \vec{c_{mn}^t}$. Then, the BP message of $x_n^t$ from $f_c$ defined in (8) is

$$\begin{aligned} m_{f_c\to x}^{\mathrm{BP}}(x_n^t) &= \left\langle \delta\left(c_n^t - e^{j\theta_n^t} x_n^t\right)\right\rangle_{n_{c\to f_c}(c_n^t) n_{\theta\to f_c}(\theta_n^t)} \\ &= \int_{\theta_n^t} \mathcal{CN}\left(e^{j\theta_n^t} x_n^t; \vec{\mu}_{c_n^t}, \vec{v}_{c_n^t}\right)\mathcal{N}\left(\theta_n^t; \mu_{\theta_n^t}^{\uparrow}, v_{\theta_n^t}^{\uparrow}\right) d_{\theta_n^t} \end{aligned} \tag{35}$$

Here we assume that the message about PN from random walk to $f_c$ is Gaussian, i.e., $n_{\theta\to f_c} = \mathcal{N}(\theta_n^t; \mu_{\theta_n^t}^{\uparrow}, v_{\theta_n^t}^{\uparrow})$ as we have assumed the belief of PN is Gaussian at the beginning of this section. It is computationally intractable to compute the BP message in (35) as it involves integral of $\theta_n^t$ over a complicated exponential expression. Alternatively, we redefine the factor $f_c$ based on (34) as

$$f_c' \triangleq p\left(\theta_n^t, x_n^t | \vec{\mu}_{c_n^t}, \vec{v}_{c_n^t}\right) \propto \exp\left\{-\frac{|e^{j\theta_n^t} x_n^t - \vec{\mu}_{c_n^t}|^2}{\vec{v}_{c_n^t}}\right\}, \tag{36}$$

and apply the MF rule to compute the message of $x_n^t$ from the factor $f_c'$ as

$$\begin{aligned} n_{x\to f_{\mathcal{M}}} &= m_{f_c\to x}^{\mathrm{MF}}(x_n^t) \\ &= \exp\left\{\langle \log(f_c')\rangle_{b(\theta_n^t)}\right\} \\ &\propto \exp\left\{\int_{\theta_n^t} b(\theta_n^t)\left(-\frac{|e^{j\theta_n^t} x_n^t - \vec{\mu}_{c_n^t}|^2}{\vec{v}_{c_n^t}}\right) d_{\theta_n^t}\right\} \\ &\approx \mathcal{CN}\left(x_n^t; \vec{\mu}_{x_n^t}, \vec{v}_{x_n^t}\right), \end{aligned} \tag{37}$$

where

$$\vec{v}_{x_n^t} = \vec{v}_{c_n^t}, \tag{38}$$

$$\vec{\mu}_{x_n^t} = e^{j\hat{\theta}_n^t}\left(1 - \frac{1}{2} v_{\theta_n^t}\right)\vec{\mu}_{c_n^t}^*. \tag{39}$$

See Appendix A for the proof. Recall we assumed the belief of PN in (16). In this approximation, we use the 2nd order Taylor expansion to approximate $e^{j\theta}$ at the PN belief mean $\hat{\theta}$. Note that higher-order Taylor approximations can also be utilized, which may lead to smaller residual errors. However, we found that the improvement in the converged performance is marginal. We therefore adopt the approximation with the 2nd order Taylor expansion for its simplicity. We also note that the message of $\theta_n^t$ to the factor $f_c'$ is the belief message $n_{\theta\to f_c} m_{f_c\to\theta}$ in MF inference rather than $n_{\theta\to f_c}$ in BP inference. In the iterative receiver, (38), (39) are the extrinsic mean and variance of transmitted symbols, respectively. To collaborate with a soft decoder, they are de-mapped into coded bits in LLR form as follows

$$L_e^{\det}\left(c_{n,q}^t\right) = \mathrm{Ln}\frac{\sum_{\alpha_i\in\mathcal{A}_q^0} \exp\left\{-\frac{|\alpha_i - \vec{\mu}_{x_n^t}|^2}{\vec{v}_{x_n^t}}\right\}\prod_{q'\neq q} \overleftarrow{P}_a\left(c_{n,q'}^t = b_{i,q}\right)}{\sum_{\alpha_i\in\mathcal{A}_q^1} \exp\left\{-\frac{|\alpha_i - \vec{\mu}_{x_n^t}|^2}{\vec{v}_{x_n^t}}\right\}\prod_{q'\neq q} \overleftarrow{P}_a\left(c_{n,q'}^t = b_{i,q}\right)}, \tag{40}$$

where $\mathcal{A}_q^0$ and $\mathcal{A}_q^1$ are the subsets of $\mathcal{A}$ whose label in position $q$, i.e.,$(s_{i,q})$, takes the value of 0 and 1, respectively. $\overleftarrow{P}_a(c_{n,q}^t)$ represents *a priori* probability of $c_{n,q}^t = b_{i,q}$ obtained from a soft decoder in the last iteration. According to [24], (40) can be computed as below if quadrature phase shift keying (QPSK) with Gray mapping is employed

$$L_e^{\det}\left(c_{n,1}^t\right) = 2\sqrt{2}\mathrm{Re}\left[\vec{\mu}_{x_n^t}/\vec{v}_{x_n^t}\right] \tag{41}$$

$$L_e^{\det}\left(c_{n,2}^t\right) = 2\sqrt{2}\mathrm{Im}\left[\vec{\mu}_{x_n^t}/\vec{v}_{x_n^t}\right]. \tag{42}$$

Assume a soft decoder performs MAP decoding of a convolutional code and outputs extrinsic LLR of coded bits $L_e^{\mathrm{dec}}(c_{n,q}^t)$ as the *a priori* input of soft detection $L_a(c_{n,q}^t)$ (demapped as in (12)). The belief p.m.f of transmitted symbol $P_x(x_n^t)$ is then obtained by (15).

### B. MESSAGE PASSING FOR PHASE NOISE ESTIMATION

Having explained the message passing inference for data detection, we will now move on to discuss how to estimate PN. Recall that the defined PN factor $f_c = \delta(c_n^t - e^{j\theta_n^t} x_n^t)$ in (8) is an exponential function. By using MF rules, the message of $\theta_n^t$ from $f_c'$ based on the belief of transmitted symbol $P_x(x_n^t)$ can be approximated to be Gaussian as below:

$$\begin{aligned} m_{f_c'\to\theta}^{\mathrm{MF}}(\theta_n^t) &= \exp\left\{\sum_{x_n^t\in\mathcal{A}} P_x(x_n^t)\log\left(f_c'\left(\theta_n^t, x_n^t | \vec{\mu}_{c_n^t}, \vec{v}_{c_n^t}\right)\right)\right\} \\ &\approx \mathcal{N}\left(\theta_n^t, \mu_{\theta_n^t}^{\downarrow}, \vec{v}_{\theta_n^t}^{\downarrow}\right), \end{aligned} \tag{43}$$

where

$$1/v_{\theta_n^t}^{\downarrow} = \mathrm{Re}\left[r_n^t e^{j\hat{\theta}_n^t}\right], \tag{44}$$

$$\mu_{\theta_n^t}^{\downarrow}/v_{\theta_n^t}^{\downarrow} = \mathrm{Re}\left[r_n^t e^{j\hat{\theta}_n^t}\left(j + \hat{\theta}_n^t\right)\right], \tag{45}$$

$$r_n^t = \frac{2}{\vec{v}_{c_n^t}}\mu_{x_n^t}\vec{\mu}_{c_n^t}^*. \tag{46}$$

and $\mu_{x_n^t} = \sum_{x_n^t\in\mathcal{A}} P_x(x_n^t) x_n^t$. See Appendix B for the proof. Denote $m_d(\theta_n^t) \triangleq m_{f_c'\to\theta}^{\mathrm{MF}}(\theta_n^t)$ in the message passing inference for random-walk process. Using BP rules, the forward message of $\theta_n^t$ from the factor $f_\Delta(\theta_n^t, \theta_n^{t-1})$ is computed as

$$m_f\left(\theta_n^t\right)=\left\langle f_\Delta\left(\theta_n^t-\theta_n^{t-1}\right)\right\rangle_{m_l\left(\theta_n^{t-1}\right)m_f\left(\theta_n^{t-1}\right)}$$
$$=\int_{\theta_n^{t-1}} m_f\left(\theta_n^{t-1}\right)m_d\left(\theta_n^{t-1}\right)f_\Delta\left(\theta_n^t-\theta_n^{t-1}\right)d_{\theta_n^{t-1}}, \quad (47)$$

which is a recursive integral. Note that if $\sigma_\Delta^2\to 0$, we have $f_\Delta(\theta_n^t-\theta_n^{t-1})\approxeq\delta(\theta_n^t-\theta_n^{t-1})$. In fact, this approximation is reasonable because $\sigma_\Delta^2\approx 10^{-3}-10^{-2}$ rad$^2$ in strong PN scenarios or $\sigma_\Delta^2\approx 10^{-5}-10^{-4}$ rad$^2$ in weak-to-moderate scenarios [23]. With this approximation, the solution of the recursive integral in (47) is given by

$$m_f\left(\theta_n^t\right)\propto\mathcal{N}\left(\theta_n^t;\vec{\mu}_{\theta_n^t},\vec{v}_{\theta_n^t}\right), \quad (48)$$

with the initial mean and variance of the forward PN message assumed to be $\vec{\mu}_{\theta_n^0}=0$ and $\vec{v}_{\theta_n^0}=0$, respectively. In fact, the true initial PN is uniformly distributed within $[0,2\pi]$, but it is absorbed into $\mathbf{H}$ during the pilot-based channel estimation (coarse estimation for the initialize iterative receiver). Similarly, the backward message of $\theta_n^t$ from the factor $f_\Delta(\theta_n^{t+1},\theta_n^t)$ is computed as

$$m_b\left(\theta_n^t\right)\propto\mathcal{N}\left(\theta_n^t;\overleftarrow{\mu}_{\theta_n^t},\overleftarrow{v}_{\theta_n^t}\right). \quad (49)$$

Then, the belief of $\theta_n^t$ is

$$b\left(\theta_n^t\right)=m_d\left(\theta_n^t\right)m_f\left(\theta_n^t\right)m_b\left(\theta_n^t\right)$$
$$\triangleq\mathcal{CN}\left(\theta_n^t;\hat{\theta}_n^t,v_{\theta_n^t}\right) \quad (50)$$

where

$$v_{\theta_n^t}=1/\left(1/\vec{v}_{\theta_n^t}+1/\overleftarrow{v}_{\theta_n^t}+1/v_{\theta_n^t}^{\downarrow}\right); \quad (51)$$
$$\hat{\theta}_n^t=v_{\theta_n^t}\left(\vec{\mu}_{\theta_n^t}/\vec{v}_{\theta_n^t}+\overleftarrow{\mu}_{\theta_n^t}/\overleftarrow{v}_{\theta_n^t}+\vec{\mu}_{\theta_n^t}^{\downarrow}/v_{\theta_n^t}^{\downarrow}\right) \quad (52)$$

### C. MESSAGE PASSING FOR CHANNEL ESTIMATION

We use the factor $f_h$ to represent the initial coarse estimate of $h_{mn}$ obtained by pilot-based channel estimation for initialization of the proposed algorithm, i.e., $m_{f_h\to h}(h_{mn})=\mathcal{CN}(h_{mn};\hat{h}_{mn}^{\text{mmse}},\hat{v}_{mn}^{mmse})$. Recall that EP has been used to approximate a Gaussian mixture-form BP message to be Gaussian when handling the message $m_{f_z\to z}^{\text{EP}}$ in (21)-(27). Given that $n_{c\to f_z}(c_n^t)=\sum_{x_n^t\in\mathcal{A}}P_x(x_n^t)\delta(c_n^t-e^{j\hat{\theta}_n^t}x_n^t)$ and $n_{z\to f_z}=\mathcal{CN}(z_{mn}^t;\vec{\mu}_{z_{mn}^t},\vec{v}_{z_{mn}^t})$, the BP message of $h_{mn}$ from $f_z$ is also a Gaussian mixture,

$$m_{f_z\to h}^{\text{BP}}(h_{mn})=\left\langle\delta\left(z_{mn}^t-h_{mn}c_n^t\right)\right\rangle_{n_{c\to f_z}(c_n^t)n_{z\to f_z}(z_{mn}^t)} \quad (53)$$

Provided that $n_{h\to f_z}^t(h_{mn})=\mathcal{CN}(h_{mn};\overleftarrow{\mu}_{h_{mn}}^t,\overleftarrow{v}_{h_{mn}}^t)$, the Gaussian approximated belief of $h_{mn}$ resembles as (21), i.e.,

$$b^t(h_{mn})=\mathcal{CN}\left(h_{mn};\hat{h}_{mn}^t,v_{h_{mn}}^t\right), \quad (54)$$

where

$$\hat{h}_{mn}^t=\sum_{x_n^t\in\mathcal{A}}\phi_{h_{mn}}\left(x_n^t\right)\mu_{h_{mn}}\left(x_n^t\right), \quad (55)$$

$$v_{h_{mn}}^t=\sum_{x_n^t\in\mathcal{A}}\phi_{h_{mn}}\left(x_n^t\right)\left(|\mu_{h_{mn}}\left(x_n^t\right)|^2+v_{h_{mn}}\left(x_n^t\right)\right)-|\hat{h}_{mn}^t|^2 \quad (56)$$

with

$$\mu_{h_{mn}}\left(x_n^t\right)=\frac{\vec{v}_{z_{mn}^t}\overleftarrow{\mu}_{h_{mn}}^t+\overleftarrow{v}_{h_{mn}}^t\vec{\mu}_{z_{mn}^t}e^{-j\hat{\theta}_n^t}x_n^{t\,*}}{\vec{v}_{z_{mn}^t}+\overleftarrow{v}_{h_{mn}}^t}, \quad (57)$$
$$v_{h_{mn}}\left(x_n^t\right)=\frac{\vec{v}_{z_{mn}^t}\overleftarrow{v}_{h_{mn}}^t}{\vec{v}_{z_{mn}^t}+\overleftarrow{v}_{h_{mn}}^t}, \quad (58)$$
$$\phi_{h_{mn}}\left(x_n^t\right)=\frac{1}{\mathcal{K}}P_x\left(x_n^t\right)\mathcal{CN}\left(\overleftarrow{\mu}_{h_{mn}}^te^{j\hat{\theta}_n^t}x_n^t;\vec{\mu}_{z_{mn}^t},\vec{v}_{z_{mn}^t}+\overleftarrow{v}_{h_{mn}}^t\right). \quad (59)$$

Then the EP message $m_{f_z\to h}^{\text{EP}}(h_{mn})$ is given by

$$m_{f_z\to h}^t(h_{mn})=\mathcal{CN}\left(h_{mn};\vec{\mu}_{h_{mn}}^t,\vec{v}_{h_{mn}}^t\right), \quad (60)$$

where

$$\vec{v}_{h_{mn}}^t=1/\left(1/\hat{v}_{h_{mn}}^t-1/\overleftarrow{v}_{h_{mn}}^t\right), \quad (61)$$
$$\vec{\mu}_{h_{mn}}^t=\vec{v}_{h_{mn}}^t\left(\hat{h}_{mn}^t/\hat{v}_{h_{mn}}^t-\overleftarrow{\mu}_{h_{mn}}^t/\overleftarrow{v}_{h_{mn}}^t\right). \quad (62)$$

It indicates that the EP mean/variance of $h_{mn}$ varies with respect to the belief p.m.f of $x_n^t$, resulting in time-varying channel estimates. For the first iteration, the message of $h_{mn}$ to $f_z$ for all $t\in[T_p+1,T_c]$ can be updated by computing (61), (62) with $n_{h\to f_z}^t(h_{mn})=m_{f_z\to h}^{t-1}(h_{mn})$, i.e.,

$$\overleftarrow{v}_{h_{mn}}^{t+1}=\vec{v}_{h_{mn}}^t, \quad (63)$$
$$\overleftarrow{\mu}_{h_{mn}}^{t+1}=\vec{\mu}_{h_{mn}}^t, \quad (64)$$

and $m_{f_z\to h}^{T_p}(h_{mn})=m_{f_h\to h}(h_{mn})$. For the following iterations, we can ignore the factor $f_h$ and the message updating is performed by exact sum-product algorithm as

$$n_{h\to f_z}^t(h_{mn})=\prod_{t'\neq t}m_{f_z\to h}^{t'}(h_{mn}). \quad (65)$$

Note that the considered MU-MIMO channel is time invariant but the calculated EP message of each channel $h_{mn}$ is Gaussian and its mean/variance change in terms of time. Eq. (65) can be understood that we utilize the messaging from all observations of a coherence block to improve the accuracy of channel estimation.

### D. MESSAGE PASSING SCHEDULING

As the observation carries new information, but all other messages in the initial states do not carry new information, we start from the left-most factor $f_\tau$ in Fig. 2 and update the messages to the right in a forward-backward manner until convergence, which is assured by proper initialization of the messages from the right according to the priori assumptions about the channel, PN and coded bits. The scheduling and messages updating of the proposed algorithm is summarized in Algorithm 1. As mentioned, within every length-$T_c$ coherence block, $T_p$ pilots are inserted at the beginning followed

**Algorithm 1** Proposed Joint PNE-CE-DET Algorithm

**INITIALIZATION**:
$L_a(c_{n,q}^t) = 0$, $\forall n, q$ and $t \in [T_p + 1, T_c]$
$\vec{\mu}_{z_{mn}^t} = 0$, $\vec{v}_{z_{mn}^t} = 1$, $\forall m, n$ and $t \in [T_p + 1, T_c]$
$\hat{\theta}_n^t = 0$, $v_{\theta_n^t} = 1$, $\forall n$ and $t \in [T_p + 1, Tc]$
$\overleftarrow{\mu}_{h_{mn}}^t = (\hat{\mathbf{H}}^{\mathrm{MMSE}})_{mn}$, $\overleftarrow{v}_{h_{mn}}^t = (\hat{\mathbf{V}}_{\mathbf{H}}^{\mathrm{MMSE}})_{mn}$, $\forall m, n$ and $t \in [1, T_p]$

**INPUT:** observations $\mathbf{Y}$
**OUTPUT:** $L_e^{\mathrm{det}}(c_n^t)$, $\forall n, t$.
  **for** $i = 1 \to N_{\mathrm{iter}}$ **do**
    $\forall n, t$: soft map $L_a(c_{n,q}^t)$ to $\overleftarrow{P}_a(x_n^t)$ by (12)
    $\forall n, t$: compute data belief $P_x(x_n^t)$ by (15)
    **for** $t = T_p + 1 \to T_c$ **do**
      $\forall m, n$: compute $\hat{z}_{mn}^t$, $v_{z_{mn}^t}$ by (22), (23), (24), (25), (26)
      $\forall m, n$: compute $\overleftarrow{\mu}_{z_{mn}^t}$, $\overleftarrow{v}_{z_{mn}^t}$ by (28)(29)
      $\forall m, n$: update $\vec{\mu}_{z_{mn}^t}$, $\vec{v}_{z_{mn}^t}$ by (32)
      $\forall n$: compute $\vec{\mu}_{c_n^t}$ and $\vec{v}_{c_n^t}$ by (33)(34)
      $\forall n$: compute $\vec{\mu}_{x_n^t}$ and $\vec{v}_{x_n^t}$ by (38)(39)
      $\forall n$: compute $L_e(c_{n,q}^t)$ by (40)
      $\forall n$: compute $\mu_{\theta_n^t}^{\downarrow}$ and $v_{\theta_n^t}^{\downarrow}$ by (44)(45)
      $\forall n$: compute $\vec{\mu}_{\theta_n^t}$ and $\vec{v}_{\theta_n^t}$ by (48)
      $\forall n$: compute $\overleftarrow{\mu}_{\theta_n^t}$ and $\overleftarrow{v}_{\theta_n^t}$ by (49)
      $\forall n$: update $\hat{\theta}_n^t$ and $v_{\theta_n^t}$ by (51)(52)
      $\forall m, n$: compute $\hat{h}_{mn}^t$, $v_{h_{mn}^t}$ by (55)(56)(57)(58)(59)
      $\forall m, n$: update $\overleftarrow{\mu}_{h_{mn}}^{t+1}$ and $\overleftarrow{v}_{h_{mn}}^{t+1}$ by (61)(62)
    **end for** $t$
    $\forall n, t$, update $L_a(\mathbf{c}_n)$ by MAP decoder
  **end for** $i$

by $T_d = T_c - T_p$ data symbols. We perform the MMSE approach [26] that takes all $T_p$ pilots as a pilot matrix and the MMSE channel estimate denoted as $\hat{\mathbf{H}}^{\mathrm{MMSE}}$ with covariance $\hat{\mathbf{V}}_{\mathbf{H}}^{\mathrm{MMSE}}$ are used to initialize the prior Gaussian distribution of the channel for the bootstrap of the proposed message passing. During the pilot-based channel estimation, PN is aggregated with channel as an equivalent time-varying channel, making the assumption $\vec{\mu}_{\theta_n^{T_p}} = 0, \vec{v}_{\theta_n^{T_p}} = 1$ reasonable. The soft-input soft-output (SISO) MAP decoder provides the update of the prior LLR of coded bits $L_a(c_n^t)$. It is initialized to 0 indicating that the p.m.f of transmitted symbols is a discrete uniform distribution according to the constellation of the modulation scheme employed and then iteratively updated by the updates of the prior LLR of coded bits from the SISO decoder. The computational complexity of the proposed algorithm is $O(MN)$ per symbol per iteration.

## IV. SIMULATIONS

In the simulations, we evaluate the BER performance of the proposed algorithm and compare it with that of the soft-input EKS [9], the auxiliary variable-aided hybrid message passing algorithm (AVA-HMP) [17] and the matched filter bound

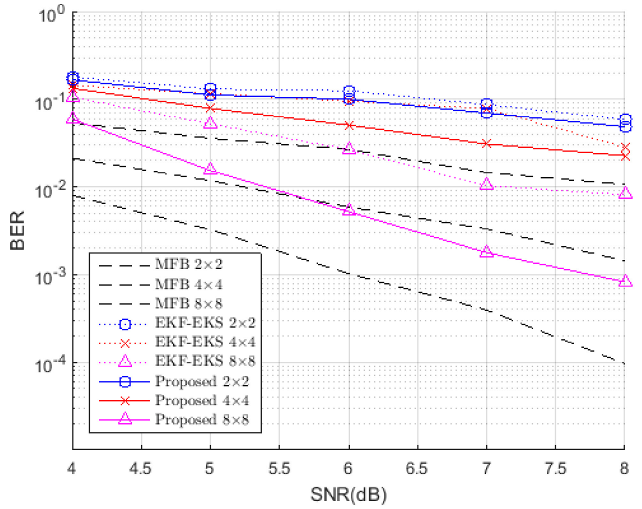



**FIGURE 3.** BER performance versus SNR.

(MFB). Note that BP-MF is used for detection as in [5] when using soft-input EKS. The PN and channel are aggregated as an equivalent time-varying channel when using the AVA-HMP that was designed for joint channel estimation and decoding. The MFB is achieved when the multi-user interference is completely removed and the channel and PN are perfectly estimated. In this case, the performance is dominated by the AWGN, and matched filtering yields the optimal detection performance. The MFB serves as a lower bound for the error rates for the detectors that aim to cancel the interference with estimated channel and PN.

The system setup are as follows: medium level of PN at UTs $\sigma_\Delta^2 = 10^{-4}\mathrm{rad}^2$, a rate-1/2 $[5, 7]_8$ convolutional code, Random interleaver and QPSK with Gray mapping. $T_c$ is 1000. $N_{\mathrm{iter}}$ is 10. Channel is following Rayleigh fading. $T_p = \max[M, N]$ pilots are inserted at the beginning of every coherence block. In Fig. 3, we compare the BER performance of the proposed algorithm with that of the soft-input EKS and the MFB with respect to $E_b/N_0$ in $2 \times 2$, $4 \times 4$ and $8 \times 8$ MU-MIMO systems. It can be seen that the proposed algorithm outperforms EKS and its performance is more significant in larger MU-MIMO. We note that the proposed algorithm exhibits a 2dB gap to the MFB, which is due to the residual multi-user interference and errors in the channel and PN estimates. Fig. 4 shows that moderate PN degrades the algorithm by approximately 2dB, as compared to its limit performance with perfect knowledge of channel and PN. Our proposed algorithm can provide a gain of about 1dB. This indicates the necessity of modelling and tracking PN. The BER with different iteration numbers are shown in Fig. 5. It indicates that the AVA-HMP reaches performance floor more quickly than the proposed algorithm.

## V. CONCLUSION

In this paper, we develop an iterative receiver with an efficient algorithm based on hybrid message passing for joint phase noise estimation, channel estimation and detection

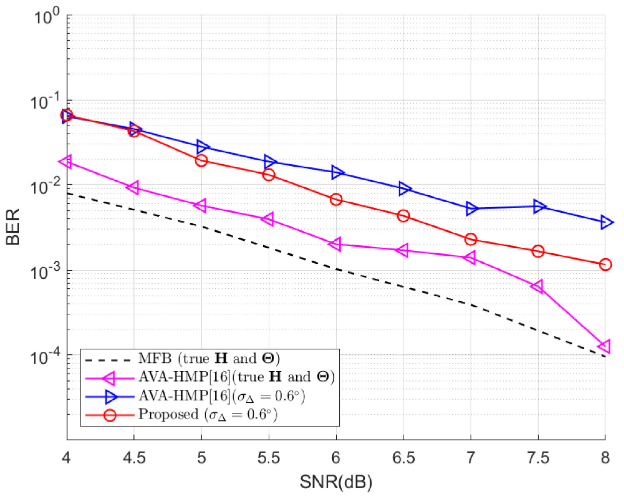



**FIGURE 4.** BER performance versus SNR in 8 × 8 MU-MIMO.

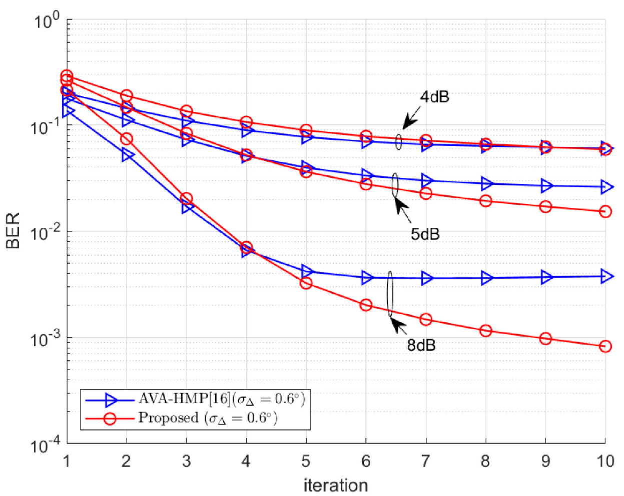



**FIGURE 5.** BER performance versus iteration in 8 × 8 MU-MIMO.

in uplink MU-MIMO on quasi-static Rayleigh-fading channel. Simulation results show that our proposed algorithm outperform Soft-input EKS and the AVA-HMP algorithm.

## APPENDIX A
## DERIVATION OF EQUATION (37)

For simplicity, we neglect the time and user indices $n, t$ of variables in the derivation. Assume the PN belief in current iteration is Gaussian, i.e., $b(\theta) \sim \mathcal{N}(\hat{\theta}, v_\theta)$,

$$\begin{aligned}
m^{\mathrm{MF}}_{f_c\to x}(x) &= \exp\left\{\int_\theta b(\theta)\log(f_c(\vec{\mu}_c, \vec{v}_c|\theta, x))d_\theta\right\}\\
&= \exp\left\{\int_\theta b(\theta)\log\left(-\frac{|e^{j\theta}x - \vec{\mu}_c|^2}{\vec{v}_c}\right)d_\theta\right\}\\
&= \exp\left\{\int_\theta b(\theta)\left[\frac{\mathrm{Re}\left[e^{j\theta}x\vec{\mu}_c^*\right]}{\vec{v}_c} - \frac{|x|^2}{\vec{v}_c}\right]d_\theta\right\}\\
&= \exp\left\{-\frac{|x|^2}{\vec{v}_c} + \frac{2\mathrm{Re}\left[\left\langle e^{j\theta}\right\rangle_{b(\theta)}x\vec{\mu}_c^*\right]}{\vec{v}_c}\right\}
\end{aligned}$$

$$\approx \mathcal{CN}(\vec{\mu}_x, \vec{v}_x), \tag{66}$$

where

$$1/\vec{v}_x = 1/\vec{v}_c, \tag{67}$$

$$\vec{\mu}_x/\vec{v}_x = \frac{\left\langle e^{j\theta}\right\rangle_{b(\theta)}\vec{\mu}_c^*}{\vec{v}_c}, \tag{68}$$

and the expectation of $e^{j\theta}$ w.r.t the PN belief $b(\theta) \sim \mathcal{N}(\hat{\theta}, v_\theta)$ is approximated by the 2nd order Taylor expansion at the point of the belief mean $\hat{\theta}$ as follow, i.e.,

$$\begin{aligned}
\left\langle e^{j\theta}\right\rangle_{b(\theta)} &\approx \int_\theta b(\theta)e^{j\hat{\theta}}\left[1 + j\left(\theta - \hat{\theta}\right) - \frac{1}{2}\left(\theta - \hat{\theta}\right)^2\right]d_\theta\\
&= e^{j\hat{\theta}}\left[1 + j\left(\hat{\theta} - \hat{\theta}\right) - \frac{1}{2}\left(\hat{\theta}^2 + v_\theta - 2\hat{\theta}^2 + \hat{\theta}^2\right)\right]
\end{aligned} \tag{69}$$

$$= e^{j\hat{\theta}}\left(1 - \frac{1}{2}v_\theta\right). \tag{70}$$

Substituting (70) into (68),

$$\vec{\mu}_x/\vec{v}_x \approx \frac{e^{j\hat{\theta}}(1 - v_\theta)\vec{\mu}_c^*}{\vec{v}_c}. \tag{71}$$

## APPENDIX B
## DERIVATION OF EQUATION (43)

For simplicity, we neglect the time and user indices $n, t$ of variables in the derivation. Assume the PN belief obtained in previous iteration is Gaussian, i.e., $b(\theta) \sim \mathcal{N}(\hat{\theta}, v_\theta)$,

$$\begin{aligned}
m^{\mathrm{MF}}_{f_c\to\theta}(\theta) &= \exp\left\{\sum_{x\in\mathcal{A}} P_x(x)\log(f_c(\vec{\mu}_c, \vec{v}_c|\theta, x))\right\}\\
&= \exp\left\{\sum_{x\in\mathcal{A}} P_x(x)\log\left(-\frac{|e^{j\theta}x - \vec{\mu}_c|^2}{\vec{v}_c}\right)\right\}\\
&= \exp\left\{\sum_{x\in\mathcal{A}} P_x(x)\left[\frac{\mathrm{Re}\left[e^{j\theta}x\vec{\mu}_c^*\right]}{\vec{v}_c} - \frac{|x|^2}{\vec{v}_c}\right]\right\}\\
&= \exp\left\{\frac{2}{\vec{v}_c}\mathrm{Re}\left[e^{j\theta}\underbrace{\sum_{x\in\mathcal{A}} P_x(x)x}_{\hat{\mu}_x}\vec{\mu}_c^*\right] - \frac{1}{\vec{v}_c}\underbrace{\sum_{x\in\mathcal{A}} P_x(x)|x|^2}_{v_x}\right\}\\
&= \exp\left\{\underbrace{\frac{2}{\vec{v}_c}\hat{\mu}_x\vec{\mu}_c^*}_{r}e^{j\theta} - \frac{v_x}{\vec{v}_c}\right\}\\
&\approx \exp\left\{\mathrm{Re}\left[re^{j\theta}\right]\right\}
\end{aligned} \tag{72}$$

To facilitate the low-complexity PN estimation, the second-order Taylor expansion of $e^{j\theta}$ at its belief mean $\hat{\theta}$ obtained in previous iteration is exploited, i.e., $e^{j\theta} \approx e^{j\hat{\theta}} + je^{j\hat{\theta}}(\theta - \hat{\theta}) - \frac{1}{2}e^{j\hat{\theta}}(\theta - \hat{\theta})^2$. Then, (72) can be further approximated as

$$m^{\mathrm{MF}}_{f_c\to\theta}(\theta) \approx \exp\left\{\mathrm{Re}\left[-\frac{1}{2}re^{j\hat{\theta}}\theta^2 + re^{j\hat{\theta}}\left(j + \hat{\theta}\right)\theta\right]\right\}$$

$$\triangleq \mathcal{N}(\theta; \vec{\mu}_\theta, \vec{v}_\theta), \tag{73}$$

where

$$1/\vec{v}_\theta = \mathrm{Re}\Big[re^{j\hat{\theta}}\Big], \tag{74}$$

$$\vec{\mu}_\theta/\vec{v}_\theta = \mathrm{Re}\Big[re^{j\hat{\theta}}\Big(j+\hat{\theta}\Big)\Big]. \tag{75}$$

## REFERENCES


[1] A. Piemontese, G. Colavolpe, and T. Eriksson, “A new analytical model of phase noise in communication systems,” in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2022, pp. 926–931, doi: 10.1109/WCNC51071.2022.9771968.

[2] A. Piemontese, G. Colavolpe, and T. Eriksson, “A new discrete-time model for channels impaired by phase noise,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2021, pp. 1–6, doi: 10.1109/GLOBECOM46510.2021.9685737.

[3] H. Ghozlan and G. Kramer, “Models and information rates for Wiener phase noise channels,” *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2376–2393, Apr. 2017, doi: 10.1109/TIT.2017.2662698.

[4] N. Hajiabdolrahim, S. R. Aghdam, and T. Eriksson, “An extended Kalman filter framework for joint phase noise, CFO and sampling time error estimation,” in *Proc. IEEE 31st Annu. Int. Symp. Pers. Indoor Mobile Radio Commun.*, 2020, pp. 1–6, doi: 10.1109/PIMRC48278.2020.9217229.

[5] W. Wang, Z. Wang, C. Zhang, Q. Guo, P. Sun, and X. Wang, “A BP–MF–EP based iterative receiver for joint phase noise estimation, equalization, and decoding,” *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1349–1353, Oct. 2016, doi: 10.1109/LSP.2016.2593917.

[6] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. Karagiannidis, and T. Svensson, “Joint estimation of channel and oscillator phase noise in MIMO systems,” *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4790–4807, Sep. 2012, doi: 10.1109/TSP.2012.2202652.

[7] A. A. Nasir, H. Mehrpouyan, R. Schober, and Y. Hua, “Phase noise in MIMO systems: Bayesian Cramér–Rao bounds and soft-input estimation,” *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2675–2692, May 2013, doi: 10.1109/TSP.2013.2243444.

[8] R. Krishnan, G. Colavolpe, A. G. I. Amat, and T. Eriksson, “Algorithms for joint phase estimation and decoding for MIMO systems in the presence of phase noise and quasi-static fading channels,” *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3360–3375, Jul. 2015, doi: 10.1109/TSP.2015.2420533.

[9] M. Nissila and S. Pasupathy, “Adaptive iterative detectors for phase-uncertain channels via variational bounding,” *IEEE Trans. Commun.*, vol. 57, no. 3, pp. 716–725, Mar. 2009, doi: 10.1109/TCOMM.2009.03.070068.

[10] Z. Yuan, C. Zhang, Z. Wang, Q. Guo, and J. Xi, “An auxiliary variable-aided hybrid message passing approach to joint channel estimation and decoding for MIMO-OFDM,” *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 12–16, Jan. 2017, doi: 10.1109/LSP.2016.2632180.

[11] X. Shao, X. Chen, and R. Jia, “A dimension reduction-based joint activity detection and channel estimation algorithm for massive access,” *IEEE Trans. Signal Process.*, vol. 68, pp. 420–435, 2020, doi: 10.1109/TSP.2019.2961299.

[12] M.-A. Badiu, G. E. Kirkelund, C. N. Manchón, E. Riegler, and B. H. Fleury, “Message-passing algorithms for channel estimation and decoding using approximate inference,” in *Proc. IEEE Int. Symp. Inf. Theory*, 2012, pp. 2376–2380, doi: 10.1109/ISIT.2012.6283939.

[13] R. Prasad, C. R. Murthy, and B. D. Rao, “Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach,” *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5369–5382, Oct. 2015, doi: 10.1109/TSP.2015.2451071.

[14] Q. Guo and D. Huang, “EM-based joint channel estimation and detection for frequency selective channels using Gaussian message passing,” *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 4030–4035, Aug. 2011, doi: 10.1109/TSP.2011.2153201.

[15] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001, doi: 10.1109/18.910572.

[16] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, “The factor graph approach to model-based signal processing,” *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, Jun. 2007, doi: 10.1109/JPROC.2007.896497.

[17] Q. Shi, N. Wu, H. Wang, and W. Yuan, “Joint channel estimation and decoding in the presence of phase noise over time-selective flat-fading channels,” *IET Communn.*, vol. 10, no. 5, pp. 577–585, 2016.

[18] J. Winn, C. Bishop, and T. Jaakkola, “Variational message passing,” *J. Mach. Learn. Res.*, vol. 6, pp. 661–694, Apr. 2005.

[19] D. J. Jakubisin, R. M. Buehrer, and C. R. C. M. Silva, “Probabilistic receiver architecture combining BP, MF, and EP for multi-signal detection,” 2016, *arXiv:1604.04834*.

[20] J. Hu, H.-A. Loeliger, J. Dauwels, and F. Kschischang, “A general computation rule for lossy summaries/messages with examples from equalization,” 2006, *arXiv:cs/0607027*.

[21] E. Riegler, G. E. Kirkelund, C. N. Manchon, M.-A. Badiu, and B. H. Fleury, “Merging belief propagation and the mean field approximation: A free energy approach,” *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 588–602, Jan. 2013, doi: 10.1109/TIT.2012.2218573.

[22] P. Sun, C. Zhang, Z. Wang, C. Manchón, and B. Fleury, “Iterative receiver design for ISI channels using combined belief-and expectation-propagation,” *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1733–1737, Oct. 2015, doi: 10.1109/LSP.2015.2404822.

[23] W. Wang, Y. Zhang, Z. Wang, Q. Guo, and J. Xi, “Low complexity message passing-based receiver design for Wiener phase-noise channels,” *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 88–91, Jan. 2017, doi: 10.1109/LCOMM.2016.2615901.

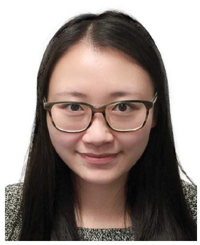
[24] Q. Guo and D. Huang, “A concise representation for the soft-in soft-out LMMSE detector,” *IEEE Commun. Lett.*, vol. 15, no. 5, pp. 566–568, May 2011, doi: 10.1109/LCOMM.2011.032811.102073.

[25] X. He, Q. Guo, J. Tong, J. Xi, and Y. Yu, “Low-complexity approximate iterative LMMSE detection for large-scale MIMO systems,” *Digit. Signal Process.*, vol. 60, pp. 134–139, Jan. 2017.

[26] M. Biguesh and A. B. Gershman, “Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006, doi: 10.1109/TSP.2005.863008.

[27] A. Barbieri, A. Cero, and G. Colavolpe, “Iterative per-frame gain and SNR estimation for DVB-S2 receivers,” in *Proc. 9th Int. Workshop Signal Process. Space Commun.*, 2006, pp. 1–8.

[28] G. Colavolpe, A. Barbieri, and G. Caire, “Algorithms for iterative decoding in the presence of strong phase noise,” *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1748–1757, Sep. 2005, doi: 10.1109/JSAC.2005.853813.

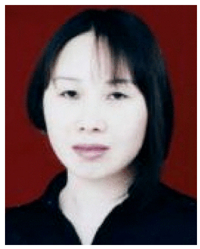
**XIAOCHEN HE** received the B.E. degree and the M.E.(Res) degree in telecommunications engineering from the University of Wollongong, Wollongong, Australia, in 2010 and 2011, respectively, where she is currently pursuing the Ph.D. degree. Her research interests include massive MIMO and signal processing for wireless communications.

**WEI WANG** (Member, IEEE) received the M.S. and Ph.D. degrees from the School of Information Engineering, Zhengzhou University, Zhengzhou, China, in 2003 and 2017, respectively. From 2004 to 2011, she was a Lecturer with the School of Information Engineering, Zhengzhou University, where she is currently with the School of Electrical and Information Engineering as an Associate Professor. Her research interests include signal processing and communication theory.

**QINGHUA GUO** (Senior Member, IEEE) received the B.E. degree in electronic engineering and the M.E. degree in signal and information processing from Xidian University, Xi'an, China, in 2001 and 2004, respectively, and the Ph.D. degree in electronic engineering from the City University of Hong Kong, Hong Kong, SAR, China, in 2008. He is currently an Associate Professor with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW, Australia, and an Adjunct Associate Professor with the School of Engineering, The University of Western Australia, Perth, WA, Australia. His research interests include signal processing, telecommunications, radar, and optical sensing. He was a recipient of the Australian Research Council's Discovery Early Career Researcher Award.

**JUN TONG** received the B.E. and M.E. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 2001 and 2004, respectively, and the Ph.D. degree in electronic engineering from the City University of Hong Kong, Hong Kong, in 2009. He is currently a Senior Lecturer with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW, Australia. His research interests are in signal processing and its applications.

**JIANGTAO XI** (Senior Member, IEEE) received the B.E. degree in electrical engineering from the Beijing Institute of Technology, Beijing, China, in 1982, the M.E. degree in electrical engineering from Tsinghua University, Beijing, in 1985, and the Ph.D. degree in electrical engineering from the University of Wollongong, Wollongong, NSW, Australia, in 1996. He was a Postdoctoral Fellow with the Communications Research Laboratory, McMaster University, Hamilton, ON, Canada, from 1995 to 1996, a member of Technical Staff with Bell Laboratories, Lucent Technologies Inc., Murray Hill, NJ, USA, from 1996 to 1998, and the Chief Technical Officer of TCL IT Group Company, Huizhou, China, from 2000 to 2002. In 2003, he rejoined the University of Wollongong as a Senior Lecturer, where he is currently a Professor and also the Head of the School of Electrical, Computer, and Telecommunications Engineering. His research interests include signal processing and its applications in various areas, such as optoelectronics, instrumentation and measurement, and telecommunications.



**YANGUANG YU** (Senior Member, IEEE) received the B.E. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1986, and the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2000. In 2007, she joined the University of Wollongong, Wollongong, NSW, Australia, where she is currently an Associate Professor with the School of Electrical, Computer and Telecommunications Engineering. She was with the College of Information Engineering, Zhengzhou University, Zhengzhou, China, on various appointments, including a Lecturer from 1986 to 1999, an Associate Professor from 2000 to 2004, and a Professor from 2005 to 2007. Her research interests include semiconductor lasers with optical feedback and signal processing and their applications to various areas.