Message-Passing Receiver for Joint Channel Estimation and Decoding in 3D Massive MIMO-OFDM Systems

Published in:
IEEE Transactions on Wireless Communications

DOI:
10.1109/TWC.2016.2612629

Document Version
Peer reviewed version

Link to publication in the UWA Research Repository

Rights statement
(c) 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.

General rights
Copyright owners retain the copyright for their material stored in the UWA Research Repository. The University grants no end-user rights beyond those which are provided by the Australian Copyright Act 1968. Users may make use of the material in the Repository providing due attribution is given and the use is in accordance with the Copyright Act 1968.

Take down policy
If you believe this document infringes copyright, raise a complaint by contacting repository-lib@uwa.edu.au. The document will be immediately withdrawn from public access while the complaint is being investigated.

Download date: 24. Oct. 2017
Message-Passing Receiver for Joint Channel Estimation and Decoding in 3D Massive MIMO-OFDM Systems

Sheng Wu, Member, IEEE, Linling Kuang, Member, IEEE, Zuyao Ni, Defeng (David) Huang, Senior Member, IEEE, Qinghua Guo, Member, IEEE, and Jianhua Lu, Fellow, IEEE

Abstract—In this paper, we address the design of message-passing receiver for massive MIMO-OFDM systems. With the aid of the central limit argument and Taylor-series approximation, a computationally efficient receiver that performs joint channel estimation and decoding is devised by the framework of expectation propagation. Specifically, the local belief defined at the channel transition function is expanded up to the second order with Wirtinger calculus, to transform the messages sent by the channel transition function to a tractable form. As a result, the channel impulse response (CIR) between each pair of antennas is estimated by Gaussian message passing. In addition, a variational expectation-maximization (EM) based method is derived to learn the channel power-delay-profiles (PDPs). The proposed scheme is assessed in 3D massive MIMO-OFDM systems with spatially correlated channels, and the empirical results corroborate its superiority in terms of performance and complexity.

Index Terms—Expectation Propagation, Joint Channel Estimation and Decoding, Massive MIMO, Message Passing, OFDM.

I. INTRODUCTION

Recently, massive multiple-input multiple-output (MIMO) systems with tens to hundreds of antennas at the base-station (BS) have gained significant attention [1]. It has been proved that massive MIMO systems can scale down transmit power as well as increase spectrum efficiency by orders of magnitude [2]. One of the challenges in massive MIMO systems is estimating the channel impulse response (CIR) for each transmit-receive link, since high data rates and energy efficiency can only be achieved when CIR is known [3], [4]. In contrast to the conventional MIMO systems with a small number of antennas, a large number of channel parameters need to be estimated in massive MIMO systems. In the meantime, the available pilot resources for channel estimation are restricted by the channel coherence time [5]. Hence, the pilot overhead may be one of the limiting factors in massive MIMO systems [1], [6]. On the other hand, the energy consumption by baseband processing grows with the number of antennas, which may obliterate the advantage of massive MIMO systems in energy efficiency. Thus, low-complexity receiver with reduced overhead is critical to massive MIMO systems.

In recent efforts to address the channel estimation problem of massive MIMO systems, several pilot-based channel estimation schemes have been proposed. Specifically, a semi-blind channel estimator based on the eigenvalue decomposition (EVD) method was proposed in [7], but may suffer from multi-user interference [8]. By exploiting the channel hardening effect of massive MIMO, Narasimhan et al. in [9] proposed an effective channel estimation for a message-passing receiver. In [10], the transceiver’s hardware impairments were incorporated into the channel model and a minimum mean-square error (MMSE) channel estimator was proposed. Furthermore, to reduce the computational complexity of MMSE channel estimator, a set of low-complexity Bayesian channel estimators using the L-order matrix polynomial expansion was proposed in [3]. Nevertheless, by exploiting pilots only, the pilot-based channel estimators of [3], [8]–[10] are less competitive in terms of pilots overhead and estimate accuracy [11]. Also, there has been another line of works on the detection problem in massive MIMO systems [9], [12]–[18]. More specifically, Chockalingam et al. addressed the detection problem by likelihood ascent search (LAS) and reactive tabu search (RTS) in [12] and [15], respectively. Švač et al. proposed a multistage detector by combining partial maximum likelihood with genetic algorithm [19]. Additionally, a series of belief propagation (BP) detectors were proposed in [9], [14] and [16], and Monte Carlo based detectors were studied in [20] and [21]. Several low-complexity soft-input soft-output detectors were proposed, based on subspace marginalization with interference suppression (SUMIS) [22], expectation propagation (EP) [17], [23], approximate message passing (AMP) [18], [24], MMSE parallel interference cancellation (MMSE-PIC) [25] and so on. For a detailed list and details on the detection of massive MIMO, please refer to [26] and references therein.

Iterative receivers that jointly estimate the channel coefficients and detect the data symbols are able to provide more accurate channel estimation and more accurate detection with lower pilots overhead [27]–[32]. Factor graph and sum-product algorithm (SPA) [33] have been used as a unified frame-
work for iterative joint data detection, channel estimation, interference cancellation, and decoding [34], [35]. However, exact inference, e.g., clique tree algorithm [36], for joint channel estimation and decoding is computationally infeasible. To this end, various approximate inference algorithms based on message passing have been proposed [28], [37]–[44]. In existing approaches, the message passing strategies include loopy belief propagation (LBP) [28], [37], [40]–[42], variational methods [38], [44], [45], and a hybrid of both [39], [43]. More specifically, LBP has a high complexity when directly applied to graphical models that involve both discrete and continuous random variables, and may even lead to poor performance [46]. This has been addressed by merging the LBP with the expectation-maximization (EM) algorithm [40] or approximating the messages of LBP with Gaussian approximations, e.g., clique tree algorithm [36], for joint detection, channel estimation, and decoding [38]. In [39], Riegler et al. derived a generic message-passing algorithm that merges belief propagation (BP) with the mean-field (MF) approximation (BP-MF), and applied it to joint channel estimation and decoding in single-input single-output orthogonal frequency division multiplex (OFDM) systems and MIMO-OFDM systems [39], [43], [48]. The BP-MF has to learn the noise precision to take into account the residual interference from other users even when the noise power is known [49], [50], as the channel transition functions are incorporated into the MF part [39], [43], [48]. Otherwise, the uncertainty of residual interference is completely ignored, and the likelihood function associated with the messages extracted from observations tends to overwhelm the a priori probability. Besides, the BP-MF requires high computational complexity as large matrices need to be inverted to estimate channel frequency response (CFR) [39], and thereby it is only feasible in the case of a few antennas and subcarriers. We note that there is a low-complexity version of the BP-MF algorithm proposed in [51], but its performance is inferior. The degraded performance may be due to the unrealistic assumption that groups of contiguous channel weights in frequency-domain obey a Markov model. Moreover, a very recent work [52] considered the BP-MF combined with generalized AMP [41], [53] to reduce the complexity of channel estimation in OFDM systems.

To achieve joint channel estimation and decoding for massive MIMO systems using OFDM modulation in frequency-selective channels, the receiver needs to complete three tasks: decoupling frequency-domain channel coefficients and data symbols from noisy observations, decoding, and channel estimation. Via central-limit theorem and moment matching, each being highly complicated. In this paper, we use the framework of EP [54] to derive an efficient message-passing algorithm. However, directly applying the EP to joint channel estimation and decoding involves complex integrals and does not get analytical solutions, in contrast with the detection of massive MIMO systems with known channels as in [17] and [18]. To this end, we use the central-limit theorem to efficiently obtain the beliefs of frequency-domain channel coefficients and the beliefs of data symbols at the channel transition functions, and then employ a quadratic approximation to project them onto the Gaussian family. In the meantime, the way of message updating using the EP principle is applied to the symbol-variable nodes. As the beliefs of frequency-domain channel coefficients are now in the Gaussian family, a Gaussian message passing based estimator [55] is employed, which exploits the fact that the CFR is the Fourier transform of the CIR. Furthermore, using the beliefs of time-domain channel taps, the unknown power-delay-profile (PDP) for each user can be learned by variational expectation maximization. We note that Parker et al. used central-limit theorem and Taylor-series approximations to formulate a bilinear generalized approximate message-passing algorithm for the SPA in the high dimensional limit [56], but its scope is different from that of this work.

In practice, to further improve spectral efficiency, it is desirable to use two-dimensional antenna arrays at the BS to form a three-dimensional (3D) massive MIMO system [57], where channel parameters, such as angles of departure, angles of arrival, and path delays, and the direction of mobility are defined in a 3D space [58]. Due to close antenna spacing at the BS, the channels are highly likely to be correlated [59], [60]. Hence, the proposed scheme of joint channel estimation and decoding is mainly assessed in 3D massive MIMO systems with spatially correlated channels. Simulation results show that the proposed scheme can achieve the performance of the approximate BP proposed in [28], [37] and [42], within 1.5 dB of the known-channel bound in a 128 × 8 MIMO system, a 64 × 8 MIMO system and a 16 × 8 MIMO system, and outperforms the BP-MF and its low-complexity variants considerably. For example, for an OFDM system with 128 subcarriers, the proposed scheme outperforms the BP-MF-GAMP [52] and the BP-MF-M [51] by about 1.75 dB and 4.5 dB, respectively, in the 128 × 8 MIMO system with 16QAM and 8 pilot subcarriers. On the other hand, the complexity of the proposed algorithm is only slightly higher than that of the BP-MF-GAMP, and is about $\frac{1}{2}$ of that of the BP-MF-M, about $\frac{1}{2}$ of the scheme in [42], and about $\frac{1}{80}$ of that of the BP-MF.

The remainder of this paper is organized as follows. The system model is described in Section II. In Section III the message passing for joint detection and decoding is presented. Complexity comparisons are shown in Section IV, and numerical results are provided in Section V, followed by conclusions in Section VI.

**Notation:** Lowercase letters (e.g., $x$) denote scalars, bold lowercase letters (e.g., $X$) denote column vectors, and bold uppercase letters (e.g., $X$) denote matrices. The superscripts $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^*$ denote the transpose operation, Hermitian transpose operation, and complex conjugate operation, respectively. Also, $\text{diag}(x)$ denotes a square diagonal matrix with the elements of vector $x$ on the main diagonal; $X \otimes Y$ denotes Kronecker product of $X$ and $Y$; $I$ denotes an identity matrix; and $[X]_{mn}$ denote the element of $X$ at the $m$th row and the $n$th column. Furthermore, $\mathcal{N}_c(x; \hat{x}, \nu_x) = \frac{1}{\sqrt{2\pi\nu_x}} \exp \left(-\frac{|x - \hat{x}|^2}{2\nu_x}\right)$ denotes the Gaussian probability density function (PDF) of $x$ with mean $\hat{x}$ and variance $\nu_x$. 


and \( \text{Gam}(\gamma; \alpha, \beta) = \beta^\alpha \gamma^{\alpha - 1} \exp(-\beta \gamma) / \Gamma(\alpha) \) denotes the Gamma PDF of \( \gamma \) with shape parameter \( \alpha \) and rate parameter \( \beta \), where \( \Gamma(\cdot) \) is the gamma function. Finally, \( \infty \) denotes equality up to a constant scale factor; \( x' \backslash x_{nk} \) denotes all elements in \( x \) but \( x_{nk} \); \( \ln(\cdot) \) denotes the natural logarithm; and \( \mathbb{E}_p(x) \{ \cdot \} \) denotes expectation with respect to distribution \( p(x) \).

II. SYSTEM MODEL

We consider the uplink of 3D massive MIMO systems where \( N \) single-antenna users communicate with the BS simultaneously. The BS employs a uniform planar array (UPA) consisting of \( M = (W \times D) \gg N \) antennas distributed across \( W \) columns and \( D \) rows\(^1\). Frequency-selective block-fading channels are assumed, and OFDM is employed to combat the multipath interference.

A. Channel Model

The CIR between the \( n \)th user and the \( m \)th receive antenna is denoted by \( h_{mn} = [h_{mn1} \cdots h_{mnL}]^T \), where \( h_{mnl} \) is the \( l \)th tap gain and \( L \) is the maximum number of taps. Let \( h_{nl} = [h_{nl1} \cdots h_{nML}]^T \) denote gain vector of all the \( L \) taps between user \( n \) and the \( M \) receive antennas at the BS. Due to close antenna spacing at the BS, we assume that all the \( M \) CIRs between the user \( n \) and the \( M \) receive antennas at the BS follow an identical PDP \( \{p_{nl} \pm \mathbb{E} \{h_{mnl}^2\}, \forall m \} \). We also assume that the transmit antennas from different users are spatially uncorrelated. Accordingly, the Kronecker spatial fading correlation model for the gain vector \( h_{nl} \) is given by [61]

\[
R_{nl} = R_{az}^nl R_{ei}^nl
\]

(1)

where \( R_{az}^nl \in \mathbb{C}^{M \times M} \) denotes the receive correlation matrix, and \( R_{ei}^nl \in \mathbb{C}^{M \times 1} \) denotes independent complex Gaussian vector with zero mean and covariance matrix \( p_{nl}I \). A ray-based channel model in [60] is employed, and the receive correlation matrix \( R_{nl} \) is approximately modeled by

\[
R_{nl} \approx R_{az}^nl \otimes R_{ei}^nl
\]

(2)

where \( R_{az}^nl \in \mathbb{R}^{W \times W} \) and \( R_{ei}^nl \in \mathbb{R}^{D \times D} \) are the correlation matrices in azimuth and elevation directions, respectively, and are defined by [60]

\[
\left[R_{az}^nl\right]_{ww'} = \frac{1}{\sqrt{d}} \exp\left(-\frac{1}{2d^2} \left(\nu^2 \cos^2(\nu_{nl}) - 2j\nu \cos(\nu_{nl})\right) \right]
\]

\[
+ \nu_{nl}^2 e^{2\nu_{nl}^2} \left(\nu_{nl}^2\right)\right)
\]

(3)

\[
\left[R_{ei}^nl\right]_{dd'} = \exp\left(\frac{2\pi}{\lambda^2} \left|\lambda d_{ei} (d' - d) \cos(\nu_{nl})\right)\right.
\]

\[
- \nu_{nl}^2 \pi \left|d_{ei} (d' - d)\right|^2 \sin^2(\nu_{nl})\right)\right)
\]

(4)

in terms of

\[
a = \frac{2\pi d_{az}}{\lambda} \sqrt{\nu_{nl}^2} \cos(\nu_{nl}),
\]

(5)

where \( \lambda \) is the carrier wavelength, \( \nu_{az}^nl \) and \( \nu_{ei}^nl \) are the mean of horizontal angle-of-departure (AoD) and the mean of vertical AoD, respectively; \( \nu_{az}^nl \) and \( \nu_{ei}^nl \) are the variance of horizontal AoD and the variance of vertical AoD, respectively; \( d_{az} \) and \( d_{ei} \) are the horizontal antenna spacing and the vertical antenna spacing, respectively.

B. Signal Model

For the \( n \)th user, the information bits \( b_n \) are encoded and interleaved, yielding a sequence of coded bits \( c_n \). Then each \( Q \) bits in \( c_n \) are mapped to one modulation symbol \( x_n^d \), which is chosen from a \( 2^Q \)-ary constellation set \( \mathcal{A} \), i.e., \( |\mathcal{A}| = 2^Q \). The data symbols \( x_n^d \) are then multiplied with pilot symbols \( x_n^p \), forming the transmitted symbol sequence \( x_n \). Pilot and data symbols are arranged in an OFDM frame of \( T \) OFDM symbols, each consisting of \( K \) subcarriers. Specifically, the frequency-domain symbols in the \( n \)th OFDM symbol transmitted by the \( n \)th user are denoted by \( x_{tn} = [x_{tn1} \cdots x_{tnK}]^T \), where \( x_{tnk} \in \mathcal{A} \) denotes the symbol transmitted at the \( k \)th subcarrier. In each OFDM frame, there are totally \( K_p \leq K \) pilot subcarriers exclusively allocated to one user, uniformly placed in one selected OFDM symbol. The set of pilot-subcarriers of user \( n \) is denoted by \( \mathcal{P}_n = \{t, k : x_{tnk} is a pilot symbol\} \) and the set of data-subcarriers is denoted by \( \mathcal{D} = \bigcup_{n=1}^{N} \mathcal{P}_n \). Note that all the users use the same time-frequency resources \( \mathcal{D} \) to transmit data symbols [62]. On the other hand, to maintain the orthogonality between the pilot sequences sent from different users, the sets of pilot-subcarriers belong to different users are set to be mutually exclusive, i.e., \( \bigcap_{n=1}^{N} \mathcal{P}_n \neq 0 \), and only one user actually transmits a pilot symbol at a given pilot-subcarrier, whereas the other users transmit zero-symbol at this pilot-subcarrier [63], i.e., \( x_{tnk} = 0, \forall n' \neq n, (t, k) \in \mathcal{P}_n \). Hence, multi-user interference at pilot-subcarriers is avoided, which simplifies the initial channel estimation [42]. The symbol sequence \( x_{tn} \) is further modulated by a \( K \)-point inverse discrete Fourier transform (IDFT), and then a cyclic prefix (CP) is added before transmission.

At the receiver, the CP is removed first and then the received signal on each receive antenna is converted into frequency domain by a \( K \)-point discrete Fourier transform (DFT). As it is assumed that the \( N \) transmitters and the receiver are synchronized and the cyclic prefix is longer than each of the CIRs, the received signal corresponding to the \( n \)th OFDM symbol can be written as

\[
y_{t,nk} = \sum_{n=1}^{N} w_{mnk} x_{tnk} + \sigma_{tnk},
\]

(8)

where \( y_{t,nk} \) denotes the received signal at the \( k \)th subcarrier on the \( m \)th receive antenna, \( \sigma_{tnk} \) denotes a circularly symmetric complex noise with zero mean and variance \( \sigma^2 \), and \( w_{mnk} \)

\(^1\)When \( D = 1 \), the UPA degenerates to a uniform linear array (ULA).
denotes the CFR at the kth subcarrier between the nth user and the mth receive antenna, which is given by
\[
W_{mnk} = \sum_{l=1}^{L} h_{mnl} \exp \left( -j2\pi lk \right).
\]  
(9)

The received signal for a frame of T OFDM symbols can be recast in a matrix-vector form as
\[
y = \sum_{n=1}^{N} W_n x_n + \sigma = W x + \sigma,
\]  
(10)

where \( y = \left[ y^T \cdots y^M \right]^T \) with \( y_m = \left[ y_{1m} \cdots y_{Tm} \right]^T \) denoting the received signal at the mth receive antenna for T OFDM symbols, \( W_n = \left[ I_T \otimes \text{diag} \{ w_{1n} \} \cdots I_T \otimes \text{diag} \{ w_{Mn} \} \right]^T \) with \( w_{mn} = \left[ w_{mn1} \cdots w_{mnK} \right]^T \) denoting the CFR from the nth user to the mth antenna, \( W = \left[ W_1 \cdots W_N \right] \), \( x = \left[ x^T \cdots x^N \right]^T \) with \( x_n = \left[ x_{1n} \cdots x_{Tn} \right]^T \) denoting the symbols transmitted by the nth user, and \( \sigma = \left[ \sigma^T \cdots \sigma^M \right]^T \) with \( \sigma_m = \left[ \sigma_{1m} \cdots \sigma_{Tm} \right]^T \) denoting the noise signal at the mth receive antenna.

C. Factor Graph Representation of the Massive MIMO-OFDM Systems

Our goal is to infer the information bits \( \{ b_n \} \) from the observations \( y \) with the known pilot symbols \( \{ x_n^p \} \). In particular, we aim to achieve the minimum bit error rate (BER) utilizing the maximum a posteriori marginal criterion, i.e.,
\[
b_{nt} = \text{arg} \max p(b_{nt} \mid y),
\]
where \( b_{m} \) denotes the nth information bit in \( b_n \), and the a posteriori probability \( p(b_{nt} \mid y) \) is given by
\[
p(b_{nt} \mid y) \propto \sum_{b_{nt},x,c} \int_{W} p(b, c, x, y, W, h),
\]  
(11)

where \( h = \left[ h_{11}^T \cdots h_{KN}^T \right] \), \( b = \left[ b_1^T \cdots b_N^T \right] \), and \( c = \left[ c_1^T \cdots c_T^T \right] \). Since \( b \rightarrow c \rightarrow x \rightarrow y \) is a Markov chain and the CFR matrix \( W \) only depends on the CIR vectors \( h \), the joint probability \( p(b, c, x, y, W, h) \) can be factorized into
\[
p(b, c, x, y, W, h) = p(b)p(c \mid b)p(x \mid c)p(y \mid W, x)p(W \mid h)p(h).
\]  
(12)

The conditional probability \( p(x \mid c) \) in (12) can be factorized into
\[
p(x \mid c) = \prod_{t=1}^{T} p(x_t \mid c_t) = \prod_{n=1}^{N} \prod_{(t,k) \in E} p(x_{tnk} \mid c_{tnk}),
\]  
(13)

where \( x_t = \left[ x_{t11} \cdots x_{t1K} \cdots x_{tN1} \cdots x_{tNK} \right]^T \), \( c_t \) is comprised of \( \{ c_{tnk}, \forall n, \forall (t,k) \in E \} \), \( p(x_{tnk} \mid c_{tnk}) = \delta (\varphi (c_{tnk}) - x_{tnk}) \) denotes the deterministic mapping \( x_{tnk} = \varphi (c_{tnk}) \), \( \varphi (c_{tnk}) \) is the mapping function, and \( \delta (\cdot) \) is the Kronecker delta function. In practice, the receive correlation matrices \( \{ R_{nt} \} \) are unknown, so we impose a conditional independent structure on the a priori probability of \( h \), i.e.,
\[
p(h \mid y) = \prod_{n=1}^{N} \prod_{t=1}^{T} p(h_{nt} \mid y_{nt}),
\]  
(14)
\[
p(h_{nt} \mid y_{nt}) = \prod_{m=1}^{M} p(h_{mnt} \mid y_{nt}),
\]  
(15)
\[
p(h_{mnt} \mid y_{nt}) \propto \text{NC} \left( h_{mnt}; 0, \gamma_{nt} \right),
\]  
(16)
\[
p(h_{nt}) \propto \text{Gam}(\gamma_{nt}; 0, 0),
\]  
(17)

where \( y = \left[ y_{11} \cdots y_{1L} \cdots y_{N1} \cdots y_{NL} \right]^T \) denotes all the hyper-parameters, and \( \gamma_{nt} \) denotes the inversion of PDP. As the CFR \( w_{mn} \) is the Fourier transform of the CIR \( h_{mn} \), i.e., \( w_{mn} = \Phi h_{mn} \), \( \forall m, \forall n \), the conditional probability \( p(W \mid h) \) reads
\[
p(W \mid h) = \prod_{n=1}^{N} \prod_{m=1}^{M} \prod_{k=1}^{K} \delta \left( w_{mnk} - \sum_{l=1}^{L} \phi_{kal} h_{mnt} \right),
\]  
(18)

where \( \Phi \in \mathbb{C}^{K \times L} \) denotes the DFT weighting matrix, and \( \phi_{kal} \) denotes the entry in the kth row and /th column of \( \Phi \). The channel transition function \( p(y \mid W, x) \) is factorized into
\[
p(y \mid W, x) = \prod_{t=1}^{T} \prod_{k=1}^{K} f_{tmk} \left( y_{tnk} \mid x_{t-k, km-k} \right),
\]  
(19)

where \( x_{t-k} = \left[ x_{t-k1} \cdots x_{t-kN} \right]^T \), \( w_{m-k} = \left[ w_{m1-k} \cdots w_{mN-k} \right]^T \), and
\[
f_{tmk} (y_{tnk} \mid x_{t-k, km-k}) \propto \text{NC} \left( y_{tnk}; \sum_{n=1}^{N} w_{tnk} x_{tnk}, \sigma_{y}^2 \right).
\]  
(20)

The probabilistic structure defined by the factorization (12)-(19) can be represented by a factor graph, as depicted in Fig. 1. In this factor graph, the mapping constraint \( \delta (\varphi (c_{tnk}) - x_{tnk}) \)
appears as a function node $M_{tnk}$, $(i,k) \in O$, mixing constraint $\delta \left( w_{mnk} - \sum_{n'=n}^N w_{mnk'} \right)$ appears as a function node $g_{mnk}$, and the conditional probability $\psi(h_{mn}, \gamma_{nl}) \approx p(h_{mn} | \gamma_{nl})$ appears as a function node $\psi_{mn}$. There exist two groups of loops, the detection-decoding-loops on the left and the channel-estimation-loops on the right. For the message passing schedules, we choose to start passing messages at the channel transition function nodes $f_{tmk}$, then pass messages concurrently in both the detection-decoding-loops and the channel-estimation-loops. Each of these full cycles of message passing will be referred to as a “turbo iteration”.

III. MESSAGE PASSING FOR JOINT CHANNEL ESTIMATION AND DECODING

The presentation of message passing follows closely the convention in [33], to which we refer the reader for an in-depth review. Due to high-dimensional integration, directly computing the marginal probabilities $p(b_n | y)$ by (11) is computationally prohibitive. Applying the SPA to the factor graph in Fig. 1 leads to a LBP algorithm, however, its complexity is as high as that of directly computing (11). Hence, we will resort to approximate inference based on the LBP to find efficient solutions.

A. BP with Gaussian Approximation in Detection-Decoding-Loops

Note that, to update the outgoing messages from the channel transition node $f_{tmk}$, the received signal shown in (8) can be rewritten as

$$y_{tmk} = w_{mnk}x_{mnk} + \sum_{n'=n}^N w_{mnk'n'}x_{mnk'} + \sigma_{tmk}.$$  

(21)

The interference plus noise term $\sum_{n'=n}^N w_{mnk'n'}x_{mnk'} + \sigma_{tmk}$ in (21) is considered as a Gaussian variable [16], [28], [42], [56], and then $y_{tmk} - (\sigma_{tmk} + \sum_{n'=n}^N w_{mnk'n'}x_{mnk'})$ is also a Gaussian random variable with the mean $z_{tmk}^{(i)}(x_{mnk})$ and variance $\tau_{tmk}^{(i)}(x_{mnk})$ given by

$$z_{tmk}^{(i)}(x_{mnk}) = y_{tmk} - \sum_{n'=n}^N w_{mnk'n'}x_{mnk'} + \sigma_{tmk},$$  

(22)

$$\tau_{tmk}^{(i)}(x_{mnk}) = \sigma_{tmk}^2 + \sum_{n'=n}^N \left( w_{mnk'n'} + \left( \sum_{n''=n}^N w_{mnk'n''}x_{mnk''} \right)^2 \right) \gamma_{mnk'n''}.$$  

(23)

where $z_{tmk}^{(i)}(x_{mnk})$ and $\gamma_{mnk'n''}$ denote the mean and variance of variable $x_{mnk}$ with respect to the message $\mu_{x_{mnk'}^{(i)\rightarrow x_{mnk}}}(x_{mnk})$ at the $(i-1)$th turbo iteration; $w_{mnk'n'}$ and $\gamma_{mnk'n''}$ denote the mean and variance of variable $w_{mnk}$ with respect to the message $\mu_{w_{mnk}^{(i)\rightarrow w_{mnk}}}(w_{mnk})$. From the model shown by (21)-(23), the channel transition function $f_{tmk}$ at the $i$th turbo iteration can be viewed as

$$f_{tmk}^{(i)}(w_{mnk}, x_{mnk}) = \mathcal{N}\left(w_{mnk}x_{mnk}, z_{tmk}^{(i)}(x_{mnk}), \tau_{tmk}^{(i)}(x_{mnk})\right), \forall n.$$  

(24)

Consequently, the message $\mu_{f_{tmk}^{(i)\rightarrow x_{mnk}}}(x_{mnk})$ is calculated by [16], [28], [42]

$$\mu_{f_{tmk}^{(i)\rightarrow x_{mnk}}}(x_{mnk}) = \int_{w_{mnk}} p_{f_{tmk}^{(i)}(w_{mnk}, x_{mnk}) \mu_{w_{mnk}^{(i)\rightarrow f_{tmk}}}(w_{mnk})} \exp \left( -\Delta_{f_{tmk}^{(i)\rightarrow x_{mnk}}}(x_{mnk}) \right),$$  

(25)

where

$$\Delta_{f_{tmk}^{(i)\rightarrow x_{mnk}}}(x_{mnk}) = \frac{\left| z_{tmk}^{(i)}(x_{mnk}) - w_{mnk}^2 \right|^2}{\gamma_{mnk}} + \frac{\left( \gamma_{mnk} - \tau_{tmk}^{(i)}(x_{mnk}) \right)^2}{\gamma_{mnk}}.$$  

(26)

After the updated messages $\mu_{f_{tmk}^{(i)\rightarrow x_{mnk}}}(x_{mnk}), \forall m$ are available, the message $\mu_{x_{mnk'}^{(i)\rightarrow x_{mnk}}}(x_{mnk})$ from the variable node $x_{mnk}$ to the mapper node $M_{tnk}$ is updated by

$$\mu_{x_{mnk'}^{(i)\rightarrow M_{tnk}}}(x_{mnk}) = \prod_{m=1}^M \mu_{f_{tmk}^{(i)\rightarrow x_{mnk}}}(x_{mnk}) \exp \left( -\sum_{m=1}^M \Delta_{f_{tmk}^{(i)\rightarrow x_{mnk}}}(x_{mnk}) \right).$$  

(27)

With the message $\mu_{x_{mnk'}^{(i)\rightarrow M_{tnk}}}(x_{mnk})$ in (27) and the a priori log likelihood ratio (LLR) $\lambda_a^{(i)}(q_{r_{tnk}}, q_{\gamma})$ fed back by the decoder of user $n$ at previous turbo iteration, the extrinsic LLRs $\lambda_e^{(i)}(q_{r_{tnk}}, q_{\gamma})$ corresponding to the symbol $x_{mnk}$ are obtained by de-mapping [35] as follows

$$\lambda_e^{(i)}(q_{r_{tnk}}, q_{\gamma}) = \frac{\sum_{x_{mnk} \in \mathcal{A}} \mu_{x_{mnk'}^{(i)\rightarrow M_{tnk}}}(x_{mnk}) \mu_{M_{tnk}^{(i)\rightarrow x_{mnk}}}(x_{mnk})}{\sum_{x_{mnk} \in \mathcal{A}} \mu_{x_{mnk'}^{(i)\rightarrow M_{tnk}}}(x_{mnk}) \mu_{M_{tnk}^{(i)\rightarrow x_{mnk}}}(x_{mnk})} - \lambda_a^{(i)}(q_{r_{tnk}}, q_{\gamma}),$$  

(28)

where the message $\mu_{M_{tnk}^{(i)\rightarrow x_{mnk}}}(x_{mnk})$ at the $(i-1)$th turbo iteration is given in the following by (29). Given that soft-in soft-out decoders are employed at the receiver, once the extrinsic LLRs $\lambda_e^{(i)}(q_{r_{tnk}}, q_{\gamma})$ are available at their inputs, each channel decoder performs decoding and feeds back the a priori LLRs of coded bits $\lambda_a^{(i)}(q_{r_{tnk}}), \forall m$ which are then interleaved and mapped to the following message

$$\mu_{M_{tnk}}^{(i)}(x_{mnk}) \equiv \left\{ \begin{array}{ll} \frac{\exp \left( q_{r_{tnk}} \cdot \lambda_a^{(i)}(q_{r_{tnk}}) \right)}{1 + \exp \left( \lambda_a^{(i)}(q_{r_{tnk}}) \right)} & \text{if } q_{r_{tnk}} > 0 \\ 0 & \text{otherwise} \end{array} \right.,$$  

(29)

Using the SPA rules [33] as well as (25) and (29), the message from the variable $x_{mnk}$ to the channel transition node $f_{tmk}$ is updated by

$$\mu_{x_{mnk}^{(i)\rightarrow f_{tmk}}}(x_{mnk}) = \mu_{M_{tnk}^{(i)\rightarrow x_{mnk}}}(x_{mnk}) \exp \left( -\sum_{m'=m}^M \Delta_{f_{tmk}^{(i)\rightarrow x_{mnk}}}^{(i)}(x_{mnk}) \right).$$  

(30)

To get $z_{tmk}^{(i)}(x_{mnk})$ in (22) and $\tau_{tmk}^{(i)}(x_{mnk})$ in (23), the mean and variance of variable $x_{mnk}$ with respect to the message $\mu_{x_{mnk}^{(i)\rightarrow f_{tmk}}}(x_{mnk})$ are calculated by

$$z_{tmk}^{(i)}(x_{mnk}) = \frac{\Delta_{f_{tmk}^{(i)\rightarrow x_{mnk}}}^{(i)}(x_{mnk})}{\sum_{x_{mnk} \in \mathcal{A}} \mu_{x_{mnk}^{(i)\rightarrow f_{tmk}}}(x_{mnk})},$$  

(31)
To retrieve the information of CFRs from the channel transition functions for channel estimation, the messages $\tilde{\mu}_{\text{frnk}}^{(i)}(w_{mnk})$, $\nu_{tnk}, \vartheta_{tnk}, \forall k$ also need to be calculated in the Detection-Decoding-Loops. Using the Gaussian approximation shown by (21)-(23), again, the message $\mu_{\text{frmk-wmnk}}^{(i)}(w_{mnk})$ is then updated by

$$
\mu_{\text{frmk-wmnk}}^{(i)}(w_{mnk}) = \sum_{x_{tnk} \in \mathcal{A}} \sum_{x_{tnk} \in \mathcal{A}} \frac{\tilde{\mu}_{\text{frnk}}^{(i-1)}(x_{tnk})}{\sum_{x_{tnk} \in \mathcal{A}} \mu_{\text{frnk-fmk}}^{(i-1)}(x_{tnk})} - \frac{(x_{tnk})^{(i-1)}}{\mu_{\text{frnk-fmk}}^{(i-1)}(x_{tnk})}.
$$

(32)

As $\mu_{\text{frmk-wmnk}}^{(i)}(w_{mnk})$ given by (33) is a Gaussian mixture, the number of components in it will increase exponentially in the subsequent message updating. To avoid the increase, the message $\mu_{\text{frmk-wmnk}}^{(i)}(w_{mnk})$ shown by (33) is projected onto a Gaussian function by the criteria of minimum KL divergence as in [28] and [37]. The projection reduces to matching the first two moments of a Gaussian function $\mathcal{N}_2(w_{mnk}; \tilde{\mu}_{\text{frmk-wmnk}}^{(i)}, \tilde{\mu}_{\text{frmk-wmnk}}^{(i)}(w_{mnk}))$ and the message $\mu_{\text{frmk-wmnk}}^{(i)}(w_{mnk})$ [64], leading to

$$
\tilde{w}_{\text{frmk-wmnk}}^{(i)} x_{tnk} = \tilde{\mu}_{\text{frmk-wmnk}}^{(i)}(w_{mnk}) = \sum_{x_{tnk} \in \mathcal{A}} \frac{\tilde{\mu}_{\text{frmk-wmnk}}^{(i)}(x_{tnk})}{x_{tnk}},
$$

(35)

$$
\nu_{\text{frmk-wmnk}}^{(i)} = (\tilde{x}_{\text{frmk-wmnk}}^{(i)} + \nu_{\text{frmk-wmnk}}^{(i)}) \sum_{x_{tnk} \in \mathcal{A}} \frac{\tilde{\mu}_{\text{frmk-wmnk}}^{(i)}(x_{tnk})}{x_{tnk}^2}.
$$

(36)

Note that, the messages at the channel transition node $f_{mk}, \forall (t, k) \in \mathcal{P}_n$ associated with a known pilot symbol $x_{tnk}$ boil down to the following simple forms

$$
\mu_{\text{frmk-wmnk}}^{(i)}(w_{mnk}) \propto \mathcal{N}_2(w_{mnk}; \tilde{y}_{tnk}, \frac{\sigma_{\tilde{y}_{tnk}}^2}{x_{tnk}^2}),
$$

(37)

$$
\mu_{\text{frmk-wmnk}}^{(i)-}(w_{mnk}) \propto \mathcal{N}_2(w_{mnk}; 0, \infty), \forall n', \forall n.
$$

(38)

where we use the fact that other users transmit zero-symbols on the pilot subcarriers $\mathcal{P}_n$, and $\mathcal{N}_2(w_{mnk}; 0, \infty)$ plays the role of a non-informative prior PDF.

Table I: The BP-GA at the $i$th iteration.

<table>
<thead>
<tr>
<th>for $m = 1 \rightarrow M$, $n = 1 \rightarrow N$, $(t, k) \in \mathcal{D}$ do</th>
</tr>
</thead>
<tbody>
<tr>
<td>Go to (22) and (23);</td>
</tr>
<tr>
<td>Update $\mu_{\text{frmk-wmnk}}^{(i)}(x_{tnk})$ by (27) and (28);</td>
</tr>
<tr>
<td>Decode and generate LLRs ${y_{tnk}^{(i)}(e_{tnk}) }, \forall (t, k) \in \mathcal{D}, \forall q$;</td>
</tr>
<tr>
<td>Update $\mu_{\text{frmk-wmnk}}^{(i)}(x_{tnk})$ by (29);</td>
</tr>
<tr>
<td>Update $\mu_{\text{frmk-wmnk}}^{(i)}(x_{tnk})$, $\forall (t, k) \in \mathcal{D}$ by (30);</td>
</tr>
<tr>
<td>Update $\mu_{\text{frmk-wmnk}}^{(i)}(x_{tnk})$, $\forall (t, k) \in \mathcal{D}$ by (31), and update $\nu_{\text{frmk-wmnk}}^{(i)}$ by (32);</td>
</tr>
<tr>
<td>end for</td>
</tr>
</tbody>
</table>

for $n = 1 \rightarrow N$, $(t, k) \in \mathcal{D}$ do

<table>
<thead>
<tr>
<th>Perform de-mapping, decoding and mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Update $\mu_{\text{frmk-wmnk}}^{(i)}(x_{tnk})$, $\forall (t, k) \in \mathcal{D}$ by (27) and (28);</td>
</tr>
<tr>
<td>Decode and generate LLRs ${y_{tnk}^{(i)}(e_{tnk}) }, \forall (t, k) \in \mathcal{D}, \forall q$;</td>
</tr>
<tr>
<td>Update $\mu_{\text{frmk-wmnk}}^{(i)}(x_{tnk})$ by (29);</td>
</tr>
<tr>
<td>Update $\mu_{\text{frmk-wmnk}}^{(i)}(x_{tnk})$, $\forall (t, k) \in \mathcal{D}$ by (30);</td>
</tr>
<tr>
<td>Update $\mu_{\text{frmk-wmnk}}^{(i)}(x_{tnk})$, $\forall (t, k) \in \mathcal{D}$ by (31), and update $\nu_{\text{frmk-wmnk}}^{(i)}$ by (32);</td>
</tr>
<tr>
<td>end for</td>
</tr>
</tbody>
</table>

B. EP with Quadratic Approximation in Detection-Decoding-Loops

In this subsection, we will derive an efficient message-passing algorithm by the framework of EP [54], [65], [66] in the Detection-Decoding-Loops. Let us return to the issue of message updating at the channel transition functions, i.e., $\{f_{mk}, i, t, \forall m, \forall k\}$. As shown by (25), (27), (31) and (32), the form of the messages $\mu_{\text{frnk-fmk}}(x_{tnk})$ in (25) leads to significant computational complexity in the BP-GA. We will reduce the complexity by representing the messages $\{\tilde{x}_{tnk}, \nu_{tnk}, \forall k\}$ and $\{\mu_{\text{frmk-wmnk}}^{(i)}(x_{tnk})\}$ with Gaussian functions of $x_{tnk}$ using the principle of EP. More specifically, the belief of $x_{tnk}$ at each function node $f_{mk}$ and the belief of $x_{tnk}$ at the variable node $x_{tnk}$ are respectively projected onto a Gaussian function, then the messages in the form of Gaussian function are indirectly obtained by a factor-division operation [66], rather than obtained by directly projecting messages themselves onto Gaussian functions.

Using $\mu_{\text{frmk-wmnk}}^{(i)}(x_{tnk})$ shown in (25), the local belief of $x_{tnk}$ at the function node $f_{mk}$ is defined by [18], [23], [67]

$$
\hat{\beta}_{\text{frmk}}^{(i)}(x_{tnk}) = \mu_{\text{frmk-wmnk}}^{(i)}(x_{tnk}) \mu_{\text{frmk-wmnk}}^{(i-1)}(x_{tnk}) \exp \left(-\Delta_{\text{frmk}}^{(i)}(x_{tnk}) \right),
$$

(39)

where

$$
\Delta_{\text{frmk}}^{(i)}(x_{tnk}) = \Delta_{\text{frmk}}^{(i)}(x_{tnk}) + \frac{(x_{tnk} - \hat{x}_{tnk})^2}{\nu_{tnk}^{(i)} - \nu_{tnk}^{(i-1)}}.
$$

(40)

We will impose a continuous complex Gaussian distribution constraint on the belief of $x_{tnk}$, i.e., project $\beta_{\text{frmk}}^{(i)}(x_{tnk})$ onto a Gaussian distribution. The projection reduces to a moment matching: however, the mean and variance of $\beta_{\text{frmk}}^{(i)}(x_{tnk})$
involve complex integrals and do not have analytical solutions. Therefore, we resort to a quadratic approximation for analytically calculating the first two moments of $\beta_{f_{mk}}^{(i)} (x_{tnk})$. As shown in the Appendix, the exponent $\Delta_{f_{mk}}^{(i)} (x_{tnk})$ in (39) can be approximated into $\tilde{\Delta}_{f_{mk}}^{(i)} (x_{tnk})$ shown by (76). Then $\tilde{\beta}_{f_{mk}}^{(i)} (x_{tnk}) \approx \exp \{-\tilde{\Delta}_{f_{mk}}^{(i)} (x_{tnk})\}$ is essentially a Gaussian approximation of $\beta_{f_{mk}}^{(i)} (x_{tnk})$, i.e.,

$$\tilde{\beta}_{f_{mk}}^{(i)} (x_{tnk}) = \mathcal{N}_x \left( x_{tnk}; \tilde{\zeta}_{f_{mk}}^{(i)}, \tilde{\gamma}_{f_{mk}}^{(i)} \right),$$

(41)

where the mean and variance of $\tilde{\beta}_{f_{mk}}^{(i)} (x_{tnk})$ are given by

$$\tilde{\zeta}_{f_{mk}}^{(i)} = \frac{\tilde{\gamma}_{f_{mk}}^{(i)} - \Delta_{f_{mk}}^{(i)} (x_{tnk})}{\Delta_{f_{mk}}^{(i)} (x_{tnk})},$$

(42)

and $\tilde{\zeta}_{f_{mk}}^{(i)}$ and $\tilde{\gamma}_{f_{mk}}^{(i)}$ are given by (72) and (73) in the Appendix, respectively. According to (39), we have $\beta_{f_{mk}}^{(i)} (x_{tnk}) = \mu_{f_{mk} \rightarrow x_{tnk}}^{(i)} (x_{tnk}) \mu_{x_{tnk} \rightarrow f_{mk}}^{(i)} (x_{tnk})$. Then the message $\mu_{f_{mk} \rightarrow x_{tnk}}^{(i)} (x_{tnk})$ can be approximately obtained by the factor-division operation [66] as follows

$$\mu_{f_{mk} \rightarrow x_{tnk}}^{(i)} (x_{tnk}) = \frac{\beta_{f_{mk}}^{(i)} (x_{tnk})}{\mu_{x_{tnk} \rightarrow f_{mk}}^{(i)} (x_{tnk})} \propto \mathcal{N}_x \left( x_{tnk}; \tilde{\zeta}_{f_{mk} \rightarrow x_{tnk}}^{(i)}, \tilde{\gamma}_{f_{mk} \rightarrow x_{tnk}}^{(i)} \right),$$

(44)

where the approximate belief $\beta_{f_{mk}}^{(i)} (x_{tnk})$ shown by (41) replaces the true belief $\beta_{f_{mk}}^{(i)} (x_{tnk})$, $\mu_{f_{mk} \rightarrow x_{tnk}}^{(i)} (x_{tnk})$ is shown in the following by (56), and the parameter $\tilde{\zeta}_{f_{mk} \rightarrow x_{tnk}}^{(i)}$ and $\tilde{\gamma}_{f_{mk} \rightarrow x_{tnk}}^{(i)}$ are given by

$$\tilde{\zeta}_{f_{mk} \rightarrow x_{tnk}}^{(i)} = \frac{\tilde{\gamma}_{f_{mk} \rightarrow x_{tnk}}^{(i)} - \Delta_{f_{mk} \rightarrow x_{tnk}}^{(i)} (x_{tnk})}{\Delta_{f_{mk} \rightarrow x_{tnk}}^{(i)} (x_{tnk})},$$

(45)

$$\tilde{\gamma}_{f_{mk} \rightarrow x_{tnk}}^{(i)} = \frac{\tilde{\gamma}_{f_{mk} \rightarrow x_{tnk}}^{(i)} - \Delta_{f_{mk} \rightarrow x_{tnk}}^{(i)} (x_{tnk})}{\Delta_{f_{mk} \rightarrow x_{tnk}}^{(i)} (x_{tnk})}. $$

(46)

Here, $\tilde{\zeta}_{f_{mk} \rightarrow x_{tnk}}^{(i)}$ and $\tilde{\gamma}_{f_{mk} \rightarrow x_{tnk}}^{(i)}$ are conveniently obtained by using the canonical form of Gaussian function [66].

Also, the local belief of $w_{mk}$ at the function node $f_{mk}$ is defined by [18], [23], [67]

$$\beta_{f_{mk}}^{(i)} (w_{mk}) = \mu_{w_{mk} \rightarrow f_{mk}}^{(i-1)} (w_{mk}) \int_{x_{tnk}} f_{tnk}^{(i)} (w_{mk}, x_{tnk}) \mu_{x_{tnk} \rightarrow f_{mk}}^{(i-1)} (x_{tnk}) \propto \exp \{-\Delta_{f_{mk}}^{(i)} (w_{mk})\},$$

(47)

where $\mu_{w_{mk} \rightarrow f_{mk}}^{(i-1)} (w_{mk})$ is shown in Tab. III as the output of channel-estimation-loops, and

$$\Delta_{f_{mk}}^{(i)} (w_{mk}) = \frac{1}{\tau_{f_{mk} \rightarrow x_{tnk}}} \left[ \frac{\Delta_{f_{mk}}^{(i-1)} (x_{tnk}) - \Delta_{f_{mk} \rightarrow x_{tnk}}^{(i-1)} (w_{mk}) + \Delta_{f_{mk}}^{(i)} (x_{tnk}) - \Delta_{f_{mk} \rightarrow x_{tnk}}^{(i)} (w_{mk})}{\mu_{w_{mk} \rightarrow f_{mk}}^{(i)} (w_{mk})} \right] \left[ \mu_{w_{mk} \rightarrow f_{mk}}^{(i-1)} (w_{mk}) \right]^2 + \left[ \Delta_{f_{mk}}^{(i)} (x_{tnk}) - \Delta_{f_{mk} \rightarrow x_{tnk}}^{(i-1)} (w_{mk}) \right] \left[ \mu_{w_{mk} \rightarrow f_{mk}}^{(i)} (w_{mk}) \right]^2$$

(48)

As shown by (80) in the Appendix, $\Delta_{f_{mk}}^{(i)} (w_{mk})$ can be approximated by $\tilde{\Delta}_{f_{mk}}^{(i)} (w_{mk})$ using Taylor expansion. Accordingly, $\beta_{f_{mk}}^{(i)} (w_{mk})$ is approximated by

$$\beta_{f_{mk}}^{(i)} (w_{mk}) \approx \exp \{-\tilde{\Delta}_{f_{mk}}^{(i)} (w_{mk})\}.$$ Similar to (44), the message $\mu_{f_{mk} \rightarrow w_{mk}}^{(i)} (w_{mk})$ can be approximately obtained by the factor-division operation, $\mu_{f_{mk} \rightarrow w_{mk}}^{(i)} (w_{mk}) \approx \beta_{f_{mk}}^{(i)} (w_{mk}) / \mu_{w_{mk} \rightarrow f_{mk}}^{(i)} (w_{mk})$, and we have

$$\mu_{f_{mk} \rightarrow w_{mk}}^{(i)} (w_{mk}) \propto \mathcal{N}_w \left( w_{mk}; \tilde{\zeta}_{f_{mk} \rightarrow w_{mk}}^{(i)}, \tilde{\gamma}_{f_{mk} \rightarrow w_{mk}}^{(i)} \right),$$

(49)

where $\tilde{\zeta}_{f_{mk} \rightarrow w_{mk}}^{(i)}$ and $\tilde{\gamma}_{f_{mk} \rightarrow w_{mk}}^{(i)}$ are given by

$$\tilde{\zeta}_{f_{mk} \rightarrow w_{mk}}^{(i)} = \frac{\tilde{\gamma}_{f_{mk} \rightarrow w_{mk}}^{(i)} - \Delta_{f_{mk} \rightarrow w_{mk}}^{(i)} (w_{mk})}{\Delta_{f_{mk} \rightarrow w_{mk}}^{(i)} (w_{mk})},$$

(50)

$$\tilde{\gamma}_{f_{mk} \rightarrow w_{mk}}^{(i)} = \frac{\tilde{\gamma}_{f_{mk} \rightarrow w_{mk}}^{(i)} - \Delta_{f_{mk} \rightarrow w_{mk}}^{(i)} (w_{mk})}{\Delta_{f_{mk} \rightarrow w_{mk}}^{(i)} (w_{mk})}.$$ (51)

Here, $\tilde{\zeta}_{f_{mk} \rightarrow w_{mk}}^{(i)}$ and $\tilde{\gamma}_{f_{mk} \rightarrow w_{mk}}^{(i)}$ are conveniently obtained by using the canonical form of Gaussian function [66].

So far, the message updating based on EP with quadratic approximation at the function node $f_{mk}$ has been derived. With (44), the message $\mu_{x_{tnk} \rightarrow M_{tnk}}^{(i)} (x_{tnk})$ from the variable node $x_{tnk}$ to the function node $M_{tnk}$ is updated by

$$\mu_{x_{tnk} \rightarrow M_{tnk}}^{(i)} (x_{tnk}) = \prod_{m=1}^{M} \mu_{f_{mk} \rightarrow x_{tnk}}^{(i)} (x_{tnk}) \propto \mathcal{N}_x \left( x_{tnk}; \tilde{\zeta}_{x_{tnk} \rightarrow M_{tnk}}^{(i)}, \tilde{\gamma}_{x_{tnk} \rightarrow M_{tnk}}^{(i)} \right),$$

(52)

where $\tilde{\zeta}_{x_{tnk} \rightarrow M_{tnk}}^{(i)} = \sum_{m=1}^{M} \left( \tilde{\zeta}_{f_{mk} \rightarrow x_{tnk}}^{(i)} \right) / \mu_{x_{tnk} \rightarrow f_{mk}}^{(i)} (x_{tnk})$ and $\tilde{\gamma}_{x_{tnk} \rightarrow M_{tnk}}^{(i)} = \sum_{m=1}^{M} \left( \tilde{\gamma}_{f_{mk} \rightarrow x_{tnk}}^{(i)} \right) / \mu_{x_{tnk} \rightarrow f_{mk}}^{(i)} (x_{tnk})$. After the same process of de-mapping, decoding and mapping as described in subsection III-A, we now need to update $\{x_{tnk}, y_{f_{mk} \rightarrow x_{tnk}}^{(i)}\}$ at variable nodes $\{x_{tnk}\}$. One strategy to update them is using moment matching like (31) and (32), but the complexity is high, and to make it worse, the number of message parameters $\{x_{tnk}, y_{f_{mk} \rightarrow x_{tnk}}^{(i)}\}$ is up to $2TMNK$ per iteration. Following the way of EP based message updating as in [18], [23] and [67], we can reduce the computational complexity of $\{x_{tnk}, y_{f_{mk} \rightarrow x_{tnk}}^{(i)}\}$. First, the local belief of $x_{tnk}$ at the variable node $x_{tnk}$ is defined by

$$\beta_{x_{tnk}}^{(i)} (x_{tnk})$$
Table II: The EP-QA at the $i$th turbo iteration.

for $m = 1 \rightarrow M$, $n = 1 \rightarrow N$, $(t, k) \in \mathcal{D}$ do

- Extract information of symbols and channel coefficients from receive signals
  
  $\begin{align*}
  \text{Update } \xi_{mk}^i & \rightarrow x_{mk}^i, \text{ and } \eta_{mk}^i \rightarrow x_{mk}^i \text{ by (22) and (23);} \\
  \hat{x}_{mk}^i & = \frac{\eta_{mk}^i}{\xi_{mk}^i x_{mk}^i}, \text{ and } \tilde{x}_{mk}^i = \frac{\eta_{mk}^i}{\xi_{mk}^i x_{mk}^i} x_{mk}^i; \\
  \hat{y}_{mk}^i & = \frac{\eta_{mk}^i}{\xi_{mk}^i x_{mk}^i} + \frac{\eta_{mk}^i}{\xi_{mk}^i x_{mk}^i} \eta_{mk}^i x_{mk}^i; \\
  \hat{y}_{mk}^i & = \frac{\eta_{mk}^i}{\xi_{mk}^i x_{mk}^i} + \frac{\eta_{mk}^i}{\xi_{mk}^i x_{mk}^i} x_{mk}^i. \\
  \text{Update } \hat{v}_{mk}^i & \text{ by (45), and update } \xi_{mk}^i \rightarrow x_{mk}^i \text{ by (46);} \\
  \text{Update } \hat{v}_{mk}^i & \text{ by (50), and update } \hat{v}_{mk}^i \rightarrow x_{mk}^i \text{ by (51).}
  \end{align*}$

end for

for $n = 1 \rightarrow N$, $(t, k) \in \mathcal{D}$ do

- De-mapping, decoding and mapping
  
  $\gamma_{xnk}^i = \left\langle \sum_{x_{nk} \in \mathcal{A}} [\xi_{mk}^i x_{mk}^i x_{mk}^i], \gamma_{xnk}^i \right\rangle$; x_{mk}^i \right\rangle; \\
  \mu_{xnk}^i \rightarrow M_{xnk}^i (x_{nk}) = \mathcal{N} (x_{nk}, \hat{x}_{nk}^i, \gamma_{xnk}^i), \forall x_{nk} \in \mathcal{A}; \\
  \text{Update } \hat{x}_{nk}^i \rightarrow x_{nk}^i \text{ by (28)}; \\
  \text{Decode and generate LLRs } [\hat{y}_{nk}^i, q_{nk}^i], \forall (t, k) \in \mathcal{D}, \forall q; \\
  \text{Update } \hat{v}_{nk}^i \rightarrow x_{nk}^i \text{ by (29), and update } \xi_{nk}^i \rightarrow x_{nk}^i \text{ by (53);} \\
  \text{Update } \hat{v}_{nk}^i \rightarrow x_{nk}^i \text{ by (54), and update } \hat{v}_{nk}^i \rightarrow x_{nk}^i \text{ by (55);} \\
  \text{Update } \hat{v}_{nk}^i \rightarrow x_{nk}^i \text{ by (57), and update } \hat{v}_{nk}^i \rightarrow x_{nk}^i \text{ by (58).}
  \end{align*}$

end for

\begin{align*}
\hat{y}_{nk}^i & = \frac{\hat{x}_{nk}^i}{\gamma_{xnk}^i} \xi_{nk}^i, \\
\hat{y}_{nk}^i & = \frac{\hat{x}_{nk}^i}{\gamma_{xnk}^i} + \frac{\hat{x}_{nk}^i}{\gamma_{xnk}^i} \hat{x}_{nk}^i.
\end{align*}

Then, the local belief $p_{xnk}^i (x_{nk})$ is projected onto a Gaussian PDF, $\mathcal{N} (x_{nk}, \hat{x}_{nk}^i, \gamma_{xnk}^i)$, by moment matching, where

\begin{align*}
\hat{x}_{nk}^i & = \sum_{x_{nk} \in \mathcal{A}} x_{nk} p_{xnk}^i (x_{nk}), \\
\gamma_{xnk}^i & = \sum_{x_{nk} \in \mathcal{A}} |x_{nk}|^2 p_{xnk}^i (x_{nk}) - |\hat{x}_{nk}^i|^2.
\end{align*}

Similar to (44), the message $\mu_{xnk}^i \rightarrow f_{mk}^i (x_{nk})$ is approximated by

\begin{align*}
\mu_{xnk}^i \rightarrow f_{mk}^i & \propto \mathcal{N} (x_{nk}, \hat{x}_{nk}^i, \gamma_{xnk}^i, \hat{y}_{nk}^i, \gamma_{xnk}^i),
\end{align*}

where

\begin{align*}
\hat{y}_{nk}^i & = \frac{x_{nk}^i}{\gamma_{xnk}^i}, \\
\hat{y}_{nk}^i & = \frac{x_{nk}^i}{\gamma_{xnk}^i} + \frac{x_{nk}^i}{\gamma_{xnk}^i} \hat{x}_{nk}^i.
\end{align*}

C. Message Updating in Channel-Estimation-Loops

In this section, we focus on the message passing in the Channel-Estimation-Loops. With the outputs of the Detection-Decoding-Loops, i.e., $\mu_{f_{mk}^i \rightarrow w_{mnk}^i} (w_{mnk}^i) = \mathcal{N} (w_{mnk}^i, \hat{w}_{f_{mk}^i \rightarrow w_{mnk}^i} (w_{mnk}^i), \gamma_{f_{mk}^i \rightarrow w_{mnk}^i})$, we can obtain

\begin{align*}
\mu_{w_{mnk}^i \rightarrow g_{mnk}^i} & = \sum_{t=1}^{T} \mu_{f_{mk}^i \rightarrow w_{mnk}^i} (w_{mnk}^i) \gamma_{f_{mk}^i \rightarrow w_{mnk}^i}^{-1}, \\
\hat{w}_{w_{mnk}^i \rightarrow g_{mnk}^i} & = \mathcal{N} (w_{mnk}^i, \hat{w}_{w_{mnk}^i \rightarrow g_{mnk}^i} (w_{mnk}^i), \gamma_{w_{mnk}^i \rightarrow g_{mnk}^i})
\end{align*}

Following the derivation in [55], the Gaussian message passing for channel estimation, i.e., calculating $\{w_{mnk}^i \rightarrow f_{mk}^i, \hat{w}_{w_{mnk}^i \rightarrow f_{mk}^i}\}$, is summarized in Tab. III, which will be referred to as “GMP”. In particular, we employ variational expectation-maximization to tune the
Table IV: Receiver schemes and their component algorithms.

<table>
<thead>
<tr>
<th>Receiver Scheme</th>
<th>Channel Estimation Algorithm</th>
<th>Detection and Decoding Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP-GA</td>
<td>GMP with oracle PDPs</td>
<td>BP-GA in Tab. I</td>
</tr>
<tr>
<td>EP-QA</td>
<td>GMP with oracle PDPs</td>
<td>EP-QA in Tab. II</td>
</tr>
<tr>
<td>BP-MF</td>
<td>Algorithm in [39] using disjoint channel model</td>
<td>BP-MF [39], [43]</td>
</tr>
<tr>
<td>BP-MF-M</td>
<td>Algorithm in [51] using Markov channel model</td>
<td>BP-MF [39], [43]</td>
</tr>
<tr>
<td>BP-MF-GAMP</td>
<td>GAMP with oracle PDPs [52]</td>
<td>BP-MF [39], [43]</td>
</tr>
<tr>
<td>MMSE</td>
<td>Frequency-domain MMSE channel estimation [29]</td>
<td>MMSE detection, followed by soft decoding [26]</td>
</tr>
<tr>
<td>MFB-PCSI</td>
<td>Perfect CSI</td>
<td>MAP decoding under perfect interference cancellation</td>
</tr>
</tbody>
</table>

Table V: Complexity of detection and decoding per turbo iteration in terms of FLOPs.

<table>
<thead>
<tr>
<th>Receiver Scheme</th>
<th>FLOPs per Turbo Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>EP-QA-L / EP-QA</td>
<td>$47TNK + (11N + 4)M (K - K_p) + (23</td>
</tr>
<tr>
<td>BP-GA [28], [37]</td>
<td>$(28</td>
</tr>
<tr>
<td>Scheme in [42]</td>
<td>$(17</td>
</tr>
<tr>
<td>BP-MF [39]</td>
<td>$19TNK + (11N + 4)M (K - K_p) + (23</td>
</tr>
<tr>
<td>BP-MF-M [51] / BP-MF-GAMP [52]</td>
<td>$33TNK + (11N + 4)M (K - K_p) + (23</td>
</tr>
</tbody>
</table>

Table VI: Complexity of channel estimation per turbo iteration in terms of FLOPs.

<table>
<thead>
<tr>
<th>Receiver Scheme</th>
<th>FLOPs per Turbo Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scheme in [42]</td>
<td>$MN(39KL + 30TK - 6K - 26TKp + 3Kp)$</td>
</tr>
</tbody>
</table>

hyper-parameters \(\{\gamma_{nl}\}\) automatically, as the PDPs may be unknown to the receiver. According to the variational message-passing rule [70], the message from the function node \(\psi_{mnl}\) to the variable \(\gamma_{nl}\) is given by

\[
\mu^{(i)}_{\gamma_{mnl} \to \gamma_{nl}}(\gamma_{nl}) = \exp \left( \mathbb{E}_{\hat{h}^{(i-1)}_{h_{mnl}}} \ln |\psi_{mnl}(h_{mnl}, \gamma_{nl})| \right) \\
\propto \frac{\Gamma(\gamma_{nl}, 0, \hat{h}^{(i-1)}_{h_{mnl}} + \gamma_{nl})}{\Gamma(\gamma_{nl}, 0, \gamma_{nl})}, \quad (60)
\]

where \(\beta^{(i-1)}_{h_{mnl}}\) is the belief of \(h_{mnl}\) at the \((i - 1)\)th iteration, as shown in the following by (63). Then the belief of precision variable \(\gamma_{nl}\) is updated by

\[
\beta^{(i)}_{\gamma_{mnl}}(\gamma_{nl}) = p(\gamma_{nl}) \prod_{m=1}^{M} \mu^{(i)}_{\gamma_{mnl} \to \gamma_{nl}}(\gamma_{nl}) \\
\propto \text{Gam} \left(\gamma_{nl}; M, \sum_{m=1}^{M} \left( \hat{h}^{(i-1)}_{h_{mnl}} + \gamma_{nl} \right) \right), \quad (61)
\]

Using the variational message-passing rule again, the message from the function node \(\psi_{mnl}(h_{mnl}, \gamma_{nl})\) to variable node \(h_{mnl}\) reads

\[
\mu^{(i)}_{\gamma_{mnl} \to h_{mnl}}(h_{mnl}) = \exp \left( \mathbb{E}_{\hat{h}^{(i-1)}_{h_{mnl}}} \ln |\psi_{mnl}(h_{mnl}, \gamma_{nl})| \right) \\
\propto \mathcal{N}_C \left( h_{mnl}; 0, \frac{1}{M} \sum_{m'=1}^{M} \hat{h}^{(i-1)}_{h_{mnl}} + \gamma_{nl} \right), \quad (62)
\]

and the belief of \(h_{mnl}\) is updated by

\[
\beta^{(i)}_{h_{mnl}}(h_{mnl}) = \mu^{(i)}_{\gamma_{mnl} \to h_{mnl}}(h_{mnl}) \prod_{k=1}^{K} \mu^{(i)}_{\gamma_{mnk} \to h_{mnl}}(h_{mnl}) \\
= \mathcal{N}_C \left( h_{mnl}; \hat{h}^{(i)}_{h_{mnl}}, \nu_{h_{mnl}} \right), \quad (63)
\]

where the parameter \(\hat{h}^{(i)}_{h_{mnl}}\) and \(\nu_{h_{mnl}}\) are shown in Tab. III. Note that, at the first turbo iteration, i.e., \(i = 1\), we set \(h^{(0)}_{mnl} = \nu^{(0)}_{mnl} = L^{-1}, \forall m, \forall n, \forall l\) and \(\xi^{(0)}_{mnl} = 0, \forall m, \forall n, \forall l\).
Figure 2: Normalized complexity of receiver schemes versus the number of antennas $M$ in the $M \times 8$ MIMO-OFDM system with 16QAM, where $K = 128$, $K_p = L = 16$, and $T = 8$. The complexity is normalized over the complexity of the EP-QA-L.

V. SIMULATION RESULTS

The proposed EP-QA and the EP-QA-L are compared with the BP-GA, the variants of BP-MF, the MMSE, and the MFB-PCSI in terms of normalized mean square error (NMSE) of channel estimation and BER. We consider the uplink of multiuser systems with $N = 8$ independent users, each of which is equipped with single transmit antenna. For each user, the transmission is based on OFDM with $K = 128$ subcarriers and 16-QAM (i.e., $Q = 4$). Information bits are encoded by a $R = 1/2$ terminated recursive systematic convolutional (RSC) code with generator polynomial $[G_1, G_2] = [117, 155]_{86}$, yielding $(TK - NK_p)Q$ coded bits, and then interleaved randomly. For bit-to-symbol mapping, a multilevel Gray-mapping is used [41]. The maximum multipath delay $L = 16$ is assumed and the PDPs are modeled identically as exponentially decaying$^2$, i.e., $p_{nl} = e^{-l/10} / \sum_{l=1}^{L} e^{-l/10}, \forall n$. Accordingly, the CP length is set to be $L_{CP} = L$. Realizations of the CIRs are generated from the MIMO channel model (1) as described in II-A. Specifically, a $16(W) \times 8(D)$ UPA, $16(W) \times 4(D)$ UPA and $16(W) \times 1(D)$ uniform linear array (ULA) at the BS are considered, corresponding to a $128 \times 8, 64 \times 8$ and $16 \times 8$ MIMO system, respectively. We set the antenna spacing to be $d_{ax} = d_{ol} = \lambda$, and uniformly generate following random variables independently for each channel realization of each user: the mean of horizontal AoD $\theta_{nl}^{ax}$ in $[\pi/6,5\pi/6]$, the mean of vertical AoD $\theta_{nl}^{ol}$ in $[\pi/12,\pi/3]$, and the standard deviations of horizontal AoD $\sqrt{d_{ax}^{nl}}$ and vertical AoD $\sqrt{d_{ol}^{nl}}$ both in $[\pi/12,\pi/6]$. At the receiver, the BCJR algorithm is used to decode the convolutional codes. The channels are assumed to be block-static for the selected $T = 8$ transmitted OFDM symbols.

Taking into account of the overhead incurred by the CP and the frequency-domain pilots, the spectral efficiency $\eta$ of the MIMO-OFDM scheme normalized by the ideal case without any overhead is expressed as $\eta = \left(TN/K - N^2K_p\right) / \left(TN(L_{CP} + K)\right)$ [63]. For $K_p = 16$, $K_p = 8$ and $K_p = 4$, the spectral efficiency is $\eta = 77.8\%$, $\eta = 83.3\%$, and $\eta = 86.1\%$, respectively. The energy per bit to noise power spectral density ratio $E_b/N_0$ is defined as [71]

$$\frac{E_b}{N_0} = \frac{E_s}{N_0} + 10\log_{10} \frac{M}{\eta RNQ},$$

where $E_s/N_0$ is the average energy per transmitted symbol. Obviously, when $E_b/N_0$ is fixed, $E_s/N_0$ is scaled down by the number of receive antennas $M$.

A. Channel-Tap NMSE Versus $E_b/N_0$

At the initial turbo iteration of these message-passing receivers, only the pilots are available for channel estimation. The EP-QA-L, the EP-QA and the BP-MF-GAMP all perform 5 inner iterations in the channel-estimation-loops at the initial turbo iteration and perform only 1 inner iteration at each of the following turbo iterations. The channel estimator in the BP-MF is equivalent to a pilot-based LMMSE estimator at the initial turbo iteration, and becomes a data-aided LMMSE estimator at the following turbo iterations. The channel estimation in the BP-MF-M is performed by a Kalman smoother proposed in [51], where the group-size of contiguous channel weights is set to be $G = 4$.

Fig. 3 shows the NMSE of the channel estimation versus $E_b/N_0$ in the $128\times8$ MIMO system, the $64\times8$ MIMO system, $2$The receiver does not know the information that the PDPs are identical.
and the 16×8 MIMO system, respectively. A maximum of 50 turbo iterations are used in all message-passing receivers, and the NMSE at the 4th turbo iteration is calculated by

\[
\text{NMSE} = \frac{1}{\Theta} \sum_{\theta=1}^{\Theta} \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{\sum_{k=1}^{L} |h_{mmn} - f_{mmn}(t)|^2}{\sum_{k=1}^{L} |h_{mmn}|^2},
\]

where Θ is the number of Monte Carlo runs. It is shown that the NMSE of the proposed EP-QA and EP-QA-L always outperform the MMSE, the BP-MF-M, the BP-MF-GAMP and the BP-MF (which is evaluated only in the 16×8 MIMO system due to complexity issue) in all cases. Meanwhile, in the low Eb/N0 region, e.g., Eb/N0 < 6.5 dB, only in the case where the number of pilots is as low as Kp = 4, as shown by Fig. 3 (b), the EP-QA is slightly inferior to the BP-GA. Using more pilots, the EP-QA can even outperform the BP-GA, as shown by Fig. 3 (a), Fig. 3 (c), and Fig. 3 (d). Compared with the EP-QA using oracle channel PDPs, the EP-QA-L is only slightly degraded in the low Eb/N0 region. Remarkably, in the high Eb/N0 region e.g., Eb/N0 > 8.5 dB, the EP-QA-L, EP-QA and BP-GA achieve almost the same NMSE. On the other hand, the BP-MF-M and the BP-MF-GAMP are more sensitive to the number of pilots. Comparing Fig. 3 (b) with Fig. 3 (a), we can see that the BP-MF-GAMP degrades about 5 dB when target NMSE is -9 dB, and the BP-MF-M degrades much more. A similar phenomenon can be observed by comparing Fig. 3 (d) with Fig. 3 (c).

Fig. 4 presents the NMSE performance versus the number of turbo iterations. Results indicate that the more the pilots are used, the less number of turbo iterations is needed to achieve convergence. In all cases, the EP-QA and BP-GA demonstrate almost the same convergence, and need less than 15 iterations to converge. In cases with fewer pilots, the EP-QA-L needs more iterations to reach convergence, as shown by Fig. 4 (b) and Fig. 4 (d).

B. BER Versus Eb/N0

Fig. 5 shows the BER performance versus Eb/N0 in the 128×8 MIMO system, the 64×8 MIMO system, and the 16×8 MIMO system, respectively. It is shown that the BP-GA, the EP-QA and the EP-QA-L achieve the same performance, being away from the MFB-PCSI within 1.5 dB in all cases, when the target BER is 10^{-5}. Meanwhile, the BP-GA, the EP-QA, and the EP-QA-L all outperform the MMSE, the BP-MF and its variants considerably. For example, the EP-QA
outperforms the BP-MF-GAMP and the BP-MF-M about 1.75 dB and 4.5 dB, respectively, in the 128 × 8 MIMO system with 8 pilots (Fig. 5 (a)). Remarkably, when the number of pilots reduces to $K_p = 4$ (Fig. 5 (b)), the gap between the EP-QA and the BP-MF-GAMP is enlarged to about 7.5 dB. A similar phenomenon can be observed in the 64 × 8 MIMO system when comparing Fig. 5 (d) with Fig. 5 (c). In the 16 × 8 MIMO system (Fig. 5 (e)), we can see that even the BP-MF with much higher complexity is still inferior to the EP-QA-L about 0.5 dB at BER $10^{-5}$. The EP-QA under perfect CSI are also evaluated in all scenarios, namely the EP-QA-PCSI. Performance gap between the EP-QA-PCSI and the MFB-PCSI is almost negligible. From this point of view, we may postulate that the gap between the EP-QA and MFB-PCSI is incurred by the channel estimation error, leading to a loss in SNR.

Fig. 6 shows the BER performance versus the number of iterations in the 128 × 8 MIMO system, the 64 × 8 MIMO system, and the 16 × 8 MIMO system, respectively. Similar to Fig. 4, the EP-QA and BP-GA demonstrate almost the same convergence in all cases. In cases with more pilots, the BP-GA, the EP-QA, and the EP-QA-L need less than 10 iterations to achieve convergence under $E_b/N_0 = 8.75$ dB, as shown by Fig. 6 (a), Fig. 6 (c) and Fig. 6 (e). On the other hand, in the cases with fewer pilots, all the algorithms need more iterations to reach convergence, as shown by Fig. 6 (b) and Fig. 6 (d). Compared with cases under $E_b/N_0 = 8.75$ dB, all the algorithms under $E_b/N_0 = 7.25$ dB also need more iterations to converge.

VI. CONCLUSION

In this paper, we presented an efficient message-passing receiver for joint channel-estimation and decoding in 3D massive MIMO systems transmitting over frequency-selective block fading channels. EP with quadratic approximation was derived to deal with the decoupling of channel coefficients and data symbols, and a low-complexity Gaussian message-passing algorithm was applied for the channel estimation. It was verified through simulations that our proposed algorithm could approach to the MFB with limited loss in 3D massive MIMO systems.

ACKNOWLEDGMENT

The authors would like to thank Prof. P. Schniter for valuable suggestions, especially on the issue of negative variances.

Figure 5: BER versus $E_b/N_0$.

Figure 6: BER versus number of iterations, under $E_b/N_0 = 7.25$ dB (dashed lines) and $E_b/N_0 = 8.75$ dB (solid lines).
occurred in the EP algorithm. Also, the authors thank the Editor, Prof. Lars K. Rasmussen and the anonymous reviewers for their valuable comments and suggestions that improve the quality of this paper.

**APPENDIX**

A real function of \( z \in \mathbb{C}, \tau \in \mathbb{C} \) and \( w \in \mathbb{C} \) is defined by

\[
\mathcal{H}(u) = \frac{2 |z|^2}{\tau + \tau^*} + \ln \left( \frac{\tau + \tau^*}{2} \right) + |w|^2,
\]

where \( u = [z, \tau, w]^T \), and \( \nu > 0 \) is a constant. Using the Wirtinger calculus [72], the power series of \( \mathcal{H}(u) \) is expanded at the point \( u_0 = [z_0, \tau_0, w_0]^T \) up to the second order, i.e.,

\[
\mathcal{H}(u) \approx \mathcal{H}(u_0) + \Re \left\{ \frac{\partial \mathcal{H}}{\partial u} \right|_{u_0} \Delta u + \Re \left\{ \frac{\partial^2 \mathcal{H}}{\partial u \partial \bar{u}} \right|_{u_0} \Delta u \right\},
\]

where \( \Delta u \equiv u - u_0 \). For the function \( \mathcal{H}(u) \) in (67), some of its partial derivatives are given by

\[
\frac{\partial \mathcal{H}}{\partial u} \equiv \left[ \begin{array}{c}
\frac{\partial \mathcal{H}}{\partial z} \\
\frac{\partial \mathcal{H}}{\partial \tau} \\
\frac{\partial \mathcal{H}}{\partial w}
\end{array} \right] = \left[ \begin{array}{c}
2z \frac{\tau^*}{\tau + \tau^*} + \frac{2 |z|^2}{\tau + \tau^*} w \\
0 \\
0
\end{array} \right].
\]

In the form of (66), where \( u = [z_0 + \Delta z, \tau_0 + \Delta \tau, w_0 + \Delta w]^T \) and \( \nu = \frac{1}{\Delta \tau} \), (71) will be omitted for tractability, i.e.,

\[
\mathcal{H}(u) \approx \mathcal{H}(u_0) + 2 \Re \left\{ \frac{z_0}{\tau_0} \Delta z + \frac{w_0}{\tau_0} \Delta w + \frac{1}{\tau_0} \Delta \tau + \frac{|z_0|^2}{\tau_0^2} \Delta z + \frac{|w|^2}{|\tau|^2} \right\}.
\]

By (75), it is not hard to get an approximate version of \( \mathcal{H}(x_{i\text{nk}}) \), namely,

\[
\mathcal{H}(x_{i\text{nk}}) = \text{const} + |x_{i\text{nk}}|^2 \left[ \frac{1}{\mathcal{H}(x_{i\text{nk}})} \right] + \Re \left\{ \frac{\mathcal{H}(x_{i\text{nk}})}{\mathcal{H}(x_{i\text{nk}})} \right\} - 2 \Re \left\{ \mathcal{H}(x_{i\text{nk}}) \right\},
\]

where the invariant terms with respect to \( x_{i\text{nk}} \) are absorbed into the constant term \( \text{const} \).

Similarly, the term \( \mathcal{H}(x_{i\text{nk}}) \) can be rewritten as

\[
\mathcal{H}(x_{i\text{nk}}) = \mathcal{H}(x_{i\text{nk}}) \left[ \frac{1}{x_{i\text{nk}}} \right] + \Re \left\{ \frac{\mathcal{H}(x_{i\text{nk}})}{x_{i\text{nk}}} \right\} - 2 \Re \left\{ \mathcal{H}(x_{i\text{nk}}) \right\},
\]

where the invariant terms with respect to \( x_{i\text{nk}} \) are absorbed into the constant term \( \text{const} \).
\[
\tau_{f_{mk} \rightarrow x_{mk}} \delta_{f_{mk} \rightarrow f_{mk}} W_{mnk} = \tau_{f_{mk} \rightarrow x_{mk}} W_{mnk}^2 + \tau_{x_{mk} \rightarrow f_{mk}} W_{mnk}^2 = 2 \tau_{f_{mk} \rightarrow x_{mk}} W_{mnk}^2 + \tau_{x_{mk} \rightarrow f_{mk}} W_{mnk}^2
\]

where the invariant terms with respect to \( \tau = u \) are

\[
\tau_{f_{mk} \rightarrow x_{mk}} W_{mnk}^2 = \tau_{f_{mk} \rightarrow x_{mk}} W_{mnk}^2 + \tau_{x_{mk} \rightarrow f_{mk}} W_{mnk}^2
\]

and we get an approximate version of \( \Delta_{f_{mk}}^i (\cdot) \), i.e.,

\[
\tau_{f_{mk} \rightarrow x_{mk}} \delta_{f_{mk} \rightarrow f_{mk}} W_{mnk} = \text{const} + \tau_{x_{mk} \rightarrow f_{mk}} W_{mnk}^2 \tau_{f_{mk} \rightarrow x_{mk}} W_{mnk}^2 \tau_{x_{mk} \rightarrow f_{mk}} W_{mnk}^2
\]

where the invariant terms with respect to \( W_{mnk} \) are absorbed into the constant term const.

References:


Y. Kim, H. Ji, J. Lee, Y. H. Nam, B. L. Ng, I. Tzanidis, Y. Li, and J. Zhang, “Full dimension MIMO (FD-MIMO): the next evolution of MIMO in LTE systems,” *IEEE Wireless Commun. Mag.*, vol. 21, no. 2, pp. 26–33, Apr. 2014.


