# Downlink Channel Estimation in Multiuser Massive MIMO With Hidden Markovian Sparsity

An Liu, *Senior Member, IEEE*, Lixiang Lian, *Student Member, IEEE*, Vincent K. N. Lau, *Fellow, IEEE*, and Xiaojun Yuan, *Senior Member, IEEE*

***Abstract*—Recently, compressive sensing based massive multiple input multiple output (MIMO) channel estimation (CE) has attracted intensive research interest. By exploiting the structured sparsity of massive MIMO channels, it is possible to design structured sparse channel estimation (SSCE) algorithms to significantly reduce the pilot/feedback overheads in massive MIMO downlink CE. However, most existing SSCE algorithms are designed based on oversimplified channel models with restrictive assumptions, and thus perform poorly under realistic channels. In this paper, we propose a Hidden-Markov-Model (HMM) to capture the structured sparsity of multiuser massive MIMO channels. The HMM has the flexibility to model different propagation environments that occur in practice. We derive an efficient approximate message passing algorithm called Turbo-orthogonal approximate message passing (OAMP) to solve the resulting sparse CE problem with an *HMM prior*. The proposed Turbo-OAMP does not require knowledge of the HMM channel parameters, which is automatically learned based on the expectation maximization framework. As a result, the proposed Turbo-OAMP algorithm derived from the HMM works well for realistic channels with complicated propagation environments and unknown channel parameters. Extensive simulations verify that the proposed Turbo-OAMP can achieve significant gains over the existing SSCE algorithms under realistic channels.**

***Index Terms*—Massive MIMO, channel estimation, structured sparsity, message passing.**

Manuscript received January 24, 2018; revised June 24, 2018; accepted July 19, 2018. Date of publication August 6, 2018; date of current version August 16, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xavier Mestre. This work was supported in part by the Science and Technology Program of Shenzhen, China under Grant JCYJ20170818113908577, in part by the National Natural Science Foundation of China under Grant 61571383, and in part by RGC 16209916. The work of A. Liu was supported by the China Recruitment Program of Global Young Experts. The work of X. Yuan was supported in part by National Natural Science Foundation of China under Grant 61471241 and in part by the China Recruitment Program of Global Young Experts. *(Corresponding authors: An Liu and Vincent K. N. Lau.)*

A. Liu is with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China, and also with the HKUST Shenzhen Research Institute, Shenzhen 518057, China (e-mail: anliu@zju.edu.cn).

L. Lian is with the Department of ECE, The Hong Kong University of Science and Technology, Hong Kong (e-mail: llianab@connect.ust.hk).

V. Lau is with the HKUST Shenzhen Research Institute, Shenzhen 518057, China, and also with the Department of ECE, The Hong Kong University of Science and Technology, Hong Kong (e-mail: eeknlau@ust.hk).

X. Yuan is with the Center for Intelligent Networking and Communications, the National Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 610051, China (e-mail: xjyuan@uestc.edu.cn).

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

Digital Object Identifier 10.1109/TSP.2018.2862420

## I. Introduction

MASSIVE multiple input multiple output (MIMO) will be employed in future wireless systems for both mmWave and low frequency bands (less than 6 GHz). In both cases, accurate channel estimation (CE) is recognized as a key challenge of massive MIMO. In time-division duplex (TDD) systems, channel reciprocity can be exploited to estimate the downlink channel via uplink pilot training. However, when the user device is also equipped with a massive MIMO array, exploiting channel reciprocity can no longer reduce the pilot training overhead in TDD systems. Moreover, in practical massive MIMO systems, the number of radio frequency (RF) chains is usually limited (much smaller than the number of antennas) due to the consideration of hardware cost, which further complicates the downlink CE problem.

A lot of research effort has been devoted to address this challenge [1]–[3]. One major approach is to exploit the sparsity of massive MIMO channels via compressive sensing (CS). In practice, there are limited scatterers in the propagation environment and thus the *angular domain massive MIMO channel* [4] can be quite sparse [1], [5]. Several structured sparse channel estimation (SSCE) algorithms have been developed to exploit the structured sparsity of massive MIMO channels to reduce the pilot/feedback overheads in downlink CE for massive MIMO systems, as summarized below.

**Algorithms exploiting joint sparsity:** When several user antennas share some common scatterers, their angular domain channel vectors tend to have some common *channel support*. Such joint sparsity has been exploited to design more efficient compressive CE algorithms in [3] and [5].

**Algorithms exploiting burst sparsity:** In [6], it is shown that due to the physical scattering structure, the significant elements in the angular domain massive MIMO channel will appear in bursts (clusters), where each burst corresponds to a scattering cluster in the propagation environment. Such a burst-sparse structure has been exploited in [6] to design a burst least absolute shrinkage and selection operator (LASSO) CE algorithm.

In [1], [2], [6], the temporal correlation of the channel support has been exploited to reduce the channel state information (CSI) signaling overhead in massive MIMO systems. There are also algorithms exploiting more than one of the above structures to further improve the CE performance, e.g., see [7], [8].

However, the above SSCE algorithms have the following drawbacks. **1)** The algorithms that exploit joint sparsity and/or

burst sparsity rely on some restrictive assumptions. For example, the joint orthogonal matching pursuit (OMP) algorithm in [5] requires that all users share a large portion of common channel support. In [6], the burst LASSO algorithm only works when all bursts of the channel vectors have similar sizes. **2)** Each existing SSCE algorithm is designed under specific assumptions on the sparse structures based on different CS recovery approaches, and there is no systematic framework for the design of SSCE algorithms. **3)** Most SSCE algorithms do not consider massive MIMO at the user side and/or massive MIMO with limited RF chains.

To overcome the above drawbacks of the existing SSCE algorithms, we propose a *Hidden-Markov-Model* (HMM) to capture the structured sparsity in multi-user massive MIMO channels, and develop the associated *Turbo-OAMP* framework for the design and analysis of SSCE algorithms. The main contributions are summarized below.

- **A Hidden Markov Channel Model:** We propose a new statistical channel model called the HMM for a general massive MIMO downlink CE problem in which both base station (BS) and users may have massive MIMO arrays and limited RF chains. The HMM has the flexibility to model different propagation environments that occur in practice. Moreover, we verify the validity of the HMM using realistic channel models. To the best of our knowledge, this is the first work that uses a HMM prior to model the structured sparsity of massive MIMO channels.
- **Design and Analysis of Turbo-OAMP:** By combining the turbo approach and the orthogonal approximate message passing (OAMP) [9]–[11], we propose an efficient message passing algorithm to solve the downlink CE problem with an HMM prior. The OAMP in [10], [11] and the associated state evolution (SE) analysis only works for independent and identically distributed (i.i.d.) priors. We extend the OAMP to the Turbo-OAMP, which works for the HMM prior. Moreover, we analyze the performance of the proposed Turbo-OAMP based on SE.

Recently, a structured Turbo CS algorithm with a Markov channel prior is proposed in [12] for a single-user massive MIMO channel with a single antenna at the user. The Turbo-OAMP with HMM channel prior is an extension of [12] from a single-user to a multi-user massive MIMO with (possibly) limited RF chains at both BSs and users. The rest of the paper is organized as follows. In Section II, we describe the system model. In Section III, we give a brief overview of the main techniques employed throughout the paper. In Section IV, we present the proposed HMM for the multi-user massive MIMO channel. In Sections V and VI, we present the Turbo-OAMP algorithm and the SE-based performance analysis, respectively. The simulation results and conclusion are given in Sections VII and VIII, respectively.

*Notations*: $\mathbf{X}^{-1}$, $\mathbf{X}^T$, $\mathbf{X}^*$and $\mathbf{X}^H$ denote the inverse, transpose, conjugate and conjugate transpose of matrix $\mathbf{X}$, respectively. $|\Omega|$ denotes the cardinality of a set $\Omega$. $\otimes$ means the Kronecker product. $\vee$ represents the logical or operator.

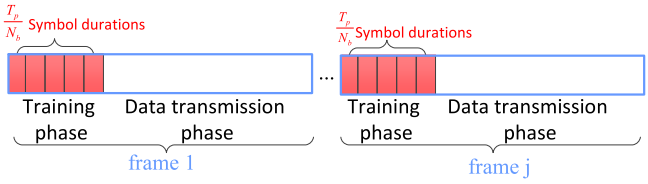



Fig. 1. Frame structure for channel estimation in a massive MIMO system.

## II. System Model

### A. Downlink Training

Consider a massive MIMO system with one BS serving $K$ users indexed by $k \in \{1, ..., K\}$. The BS is equipped with $M \gg 1$ antennas and $M_b \leq M$ transmit RF chains. Each user is equipped with $N$ antennas and $N_b \leq N$ receive RF chains. The time is divided into frames and each frame (channel coherence time) contains a training phase and a data transmission phase, as illustrated in Fig. 1.

The training phase contains $T_p/N_b$ pilot symbol durations, where $T_p$ is the total number of channel measurements which can be divided by $N_b$. At the $i$-th pilot symbol duration, the BS transmits a training vector $\mathbf{v}_i = \mathbf{W}_0^i \mathbf{g}_0^i \in \mathbb{C}^M$ for downlink CE and user $k$ employs $\mathbf{U}_{k,i} = \mathbf{W}_k^i \mathbf{G}_k^i \in \mathbb{C}^{N \times N_b}$ as a combining matrix to combine the received signal into $N_b$ baseband channel measurements, where $\mathbf{W}_0^i \in \mathbb{C}^{M \times M_b}$ and $\mathbf{g}_0^i \in \mathbb{C}^{M_b}$ are the RF training matrix and baseband training vector at the BS in the $i$-th pilot symbol duration, respectively, and $\mathbf{W}_k^i \in \mathbb{C}^{N \times N_b}$ and $\mathbf{G}_k^i \in \mathbb{C}^{N_b \times N_b}$ are the RF and baseband measurement matrices at user $k$ in the $i$-th pilot symbol duration, respectively. In this paper, we focus on the case when the RF combining matrix $\mathbf{W}_k^i$ and RF training matrix $\mathbf{W}_0^i$ are implemented using phase shifters and adders in RF domain [13]. As such, all the elements of $\mathbf{W}_k^i$ and $\mathbf{W}_0^i$ must have equal magnitude. The exact choice of $\mathbf{W}_0^i, \mathbf{g}_0^i, \mathbf{W}_k^i, \mathbf{G}_k^i$ is postponed to Section V-A. Then the received baseband signal in the $i$-th pilot symbol duration for user $k$ is

$$\mathbf{y}_{k,i} = \mathbf{U}_{k,i}^H \mathbf{H}_k \mathbf{v}_i + \mathbf{U}_{k,i}^H \mathbf{n}_{k,i}, \tag{1}$$

where $\mathbf{H}_k \in \mathbb{C}^{N \times M}$ is the channel matrix of user $k$, and $\mathbf{n}_{k,i}$ is the additive complex Gaussian noise (AWGN) at user $k$ with each element being zero mean and variance $\sigma^2$.

The aggregate received pilot signal (channel measurements) of all the $T_p/N_b$ pilot symbols at user $k$ can be expressed in a compact form as

$$\mathbf{y}_k = \left[\mathbf{v}_1^H \mathbf{H}_k^H \mathbf{U}_{k,1}, \cdots, \mathbf{v}_{T_p/N_b}^H \mathbf{H}_k^H \mathbf{U}_{k,T_p/N_b}\right]^H + \mathbf{n}_k, \tag{2}$$

where $\mathbf{n}_k = [\mathbf{n}_{k,1}^H \mathbf{U}_{k,1}, ..., \mathbf{n}_{k,T_p/N_b}^H \mathbf{U}_{k,T_p/N_b}]^H \in \mathbb{C}^{T_p \times 1}$ is the noise vector at user $k$.

### B. Angular Domain Channel Representation

We first describe the angular domain channel representation for the channel between the BS and an individual user. The user index $k$ is omitted for conciseness. For clarity, we focus on the case when both the BS and users are equipped with a

*half-wavelength space* uniform linear array (ULA) and the channel is flat fading.

Consider the widely used discrete multipath channel model in [4], where the azimuth angle of departure (AoD) of the BS and angle of arrival (AoA) of user $k$ are assumed to take values from the following two discrete sets, respectively.

$$\{\theta_{T,m}:$$

$$\sin(\theta_{T,m}) = \frac{2}{\tilde{M}}\left(m - \left\lfloor \frac{\tilde{M}-1}{2} \right\rfloor\right), m = 0, \cdots, \tilde{M}-1\Big\},$$

$$\{\theta_{R,n}:$$

$$\sin(\theta_{R,n}) = \frac{2}{\tilde{N}}\left(n - \left\lfloor \frac{\tilde{N}-1}{2} \right\rfloor\right), n = 0, \cdots, \tilde{N}-1\Big\},$$

where $\tilde{M}$ and $\tilde{N}$ are the total number of discrete AoDs and AoAs, respectively. In this case, the downlink channel matrix $\mathbf{H} \in \mathbb{C}^{N\times M}$ from the BS to a user can be modeled as

$$\mathbf{H} = \sum_{n=1}^{\tilde{N}} \sum_{m=1}^{\tilde{M}} x_{n,m} \mathbf{a}_R(\theta_{R,n}) \mathbf{a}_T^H(\theta_{T,m}), \tag{3}$$

where $\mathbf{a}_T(\theta_{T,m}) \in \mathbb{C}^{M\times 1}$ and $\mathbf{a}_R(\theta_{R,n}) \in \mathbb{C}^{N\times 1}$ are the array response vector for the BS antenna array (associated with the $m$-th AoD $\theta_{T,m}$) and user antenna array (associated with the $n$-th AoA $\theta_{R,n}$), respectively, $x_{n,m}$ is the complex gain of the path from the $m$-th transmit direction (AoD) at the BS to the $n$-th receive direction (AoA) at user $k$ [4]. For half-wavelength space ULAs, the array response vector is

$$\mathbf{a}_T(\theta) = \frac{1}{\sqrt{M}}\left[1, e^{-j\pi\sin(\theta)}, e^{-j2\pi\sin(\theta)}, \cdots, e^{-j(M-1)\pi\sin(\theta)}\right]^T, \tag{4}$$

$$\mathbf{a}_R(\theta) = \frac{1}{\sqrt{N}}\left[1, e^{-j\pi\sin(\theta)}, e^{-j2\pi\sin(\theta)}, \cdots, e^{-j(N-1)\pi\sin(\theta)}\right]^T. \tag{5}$$

For convenience, define two array response matrices $\mathbf{A}_R = [\mathbf{a}_R(\theta_{R,1}), ..., \mathbf{a}_R(\theta_{R,\tilde{N}})] \in \mathbb{C}^{N\times\tilde{N}}$ and $\mathbf{A}_T = [\mathbf{a}_T(\theta_{T,1}), ..., \mathbf{a}_T(\theta_{T,\tilde{M}})] \in \mathbb{C}^{M\times\tilde{M}}$. Furthermore, define $\mathbf{X} \in \mathbb{C}^{\tilde{N}\times\tilde{M}}$ as the angular domain channel matrix with the $(n,m)$-th element given by $x_{n,m}$. Then $\mathbf{H}$ can be expressed in a compact form as

$$\mathbf{H} = \mathbf{A}_R \mathbf{X} \mathbf{A}_T^H. \tag{6}$$

The same discrete channel model is used in most existing works on massive MIMO CE [5], [14], [15]. Note that we can also define the angular domain representation for more practical 2-dimensional (2D) antenna arrays. In this case, the array response vector $\mathbf{a}_T(\theta,\varphi)$ at the BS (or $\mathbf{a}_R(\theta,\varphi)$ at the user) can be expressed as a function of the azimuth angle $\theta$ and elevation angle $\varphi$. Please refer to [16] for the details. Assume that the azimuth-elevation AoD pair $(\theta,\varphi)$ of the BS takes values from a discrete set of size $\tilde{M}$: $\{(\theta_{T,m},\varphi_{T,m}) : m = 1, ..., \tilde{M}\}$, and the azimuth-elevation AoA pair $(\theta,\varphi)$ of the user takes values from a discrete set of size $\tilde{N}$: $\{(\theta_{R,n},\varphi_{R,n}) : n = 1, ..., \tilde{N}\}$. Then the channel matrix $\mathbf{H}$ can also be expressed using a similar angular domain representation as in (6), except that the array response matrices $\mathbf{A}_R$ and $\mathbf{A}_T$ are redefined as $\mathbf{A}_R = [\mathbf{a}_R(\theta_{R,1},\varphi_{R,1}), ..., \mathbf{a}_R(\theta_{R,\tilde{N}},\varphi_{R,\tilde{N}})]$ and $\mathbf{A}_T = [\mathbf{a}_T(\theta_{T,1},\varphi_{T,1}), ..., \mathbf{a}_T(\theta_{T,\tilde{M}},\varphi_{T,\tilde{M}})]$.

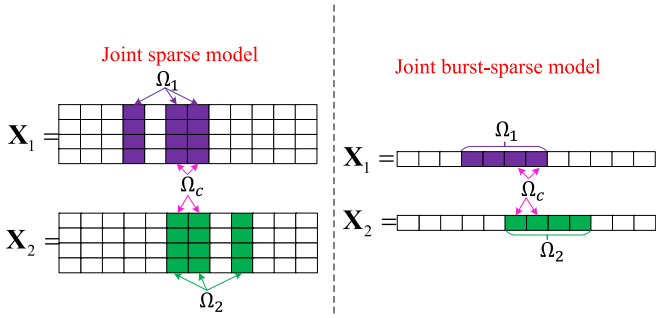



Fig. 2. Structured sparsity of multi-user massive MIMO channels in [5] and [8].

## III. Preliminary Results

In this section, we review some preliminary results and main techniques employed throughout the paper.

### A. Structured Sparsity of Massive MIMO Channel

Multi-user massive MIMO channels have been modeled as joint sparse channels and joint burst-sparse channels in [5] and [8], respectively, as shown in Fig. 2. Specifically, in a joint sparse channel, each row of the channel matrix shares the same sparsity support, and different channel matrices share a partially common support. In a joint burst-sparse channel, the channels of single antenna users only have one single non-zero cluster, and different user channels share a partially common support. These structured sparsity models rely on some restrictive assumptions. For example, they require that all users share a large portion of common channel support, which is usually not satisfied in practice due to the complicated propagation environment. [5] also requires that there are dense clusters at user side, and [8] only works for single antenna users and when there is only one cluster in each channel vector. Furthermore, the algorithms proposed in [5] and [8] cannot be used to deal with more realistic and complicated sparsity structures.

### B. Orthogonal Approximate Message Passing

OAMP proposed in [10], [11], [17] is a variation of the well-known approximate message passing (AMP) [18]. OAMP can handle a wide range of partial unitary sensing matrices[1], and it is shown to achieve a better performance than AMP. Consider the linear observation model

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{n}, \tag{7}$$

[1]For matrix $\mathbf{F} \in \mathbb{C}^{L\times Q}$ with $L < Q$, we say matrix $\mathbf{F}$ is partial unitary if $\mathbf{F}\mathbf{F}^H = \mathbf{I}$, where $\mathbf{I}$ is a $L \times L$ dimensional identity matrix.

where $\mathbf{x} \in \mathbb{C}^{Q\times 1}$ is a sparse signal to be estimated, $\mathbf{y} \in \mathbb{C}^{L\times 1}$ is the received signal, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ is the Gaussian noise, $\mathbf{F} \in \mathbb{C}^{L\times Q}$ is a partial unitary matrix. The entries of the sparse signal $\mathbf{x}$ are assumed to be i.i.d., with the $j$-th entry following the Bernoulli-Gaussian distribution:

$$x_j \sim \begin{cases} 0 & \text{probability} = 1-\lambda, \\ \mathcal{CN}(0,\zeta) & \text{probability} = \lambda. \end{cases} \tag{8}$$

OAMP is designed to recover the i.i.d. sparse signal $\mathbf{x}$ from the linear observation model in (7).

OAMP contains two modules: Module A is a linear minimum mean square error (LMMSE) estimator based on the observation and the messages from Module B, and Module B performs minimum mean square error (MMSE) estimator that combines the i.i.d sparse prior distribution in (8) and the messages from Module A. The extrinsic output [19] of a module is fed to the other module as a prior input. The two modules are executed iteratively until convergence. At the end, the estimation of $\mathbf{x}$ is based on the posterior output of Module B.

Specifically, for Module A, it is based on the assumption that the entries of $\mathbf{x}$ are i.i.d with a prior mean $\mathbf{x}_A^{pri}$ and variance $v_A^{pri}$, where $\mathbf{x}_A^{pri}$ and $v_A^{pri}$ are the messages passed from Module B. Then under this assumption, the LMMSE estimator and the mean-square error (MSE) of $\mathbf{x}$ based on model (7) are respectively given by [11]

$$\mathbf{x}_A^{post} = \mathbf{x}_A^{pri} + \frac{v_A^{pri}}{v_A^{pri}+\sigma^2}\mathbf{F}^H(\mathbf{y} - \mathbf{F}\mathbf{x}_A^{pri}), \tag{9}$$

$$v_A^{post} = v_A^{pri} - \frac{L}{Q}\cdot\frac{(v_A^{pri})^2}{v_A^{pri}+\sigma^2}. \tag{10}$$

The extrinsic LMMSE estimate and the MSE of $\mathbf{x}$ can be computed by [11]

$$\mathbf{x}_B^{pri} = \mathbf{x}_A^{ext} = v_A^{ext}\left(\frac{\mathbf{x}_A^{post}}{v_A^{post}} - \frac{\mathbf{x}_A^{pri}}{v_A^{pri}}\right), \tag{11}$$

$$v_B^{pri} = v_A^{ext} = \left(\frac{1}{v_A^{post}} - \frac{1}{v_A^{pri}}\right)^{-1}, \tag{12}$$

which are passed to Module B as its input messages.

For Module B, it is based on the assumption that $\mathbf{x}_B^{pri}$ is modeled as an AWGN observation of $\mathbf{x}$, i.e.,

$$\mathbf{x}_B^{pri} = \mathbf{x} + \mathbf{w}, \tag{13}$$

where $\mathbf{w} \sim \mathcal{CN}(0, v_B^{pri}\mathbf{I})$, and is independent of $\mathbf{x}$. Based on this assumption, the posterior mean and variance can be respectively calculated as [11]

$$x_{B,j}^{post} = \mathrm{E}\left(x_j|\mathbf{x}_B^{pri}\right) = \mathrm{E}\left(x_j|x_{B,j}^{pri}\right), \tag{14}$$

$$\begin{aligned} v_B^{post} &= \frac{1}{Q}\sum_{j=1}^{Q} v_{B,j}^{post} = \frac{1}{Q}\sum_{j=1}^{Q}\mathrm{Var}\left(x_j|x_{B,j}^{pri}\right) \\ &= \frac{1}{Q}\sum_{j=1}^{Q}\mathrm{E}\left(\left|x_j - \mathrm{E}\left(x_j|x_{B,j}^{pri}\right)\right|^2\right) \end{aligned} \tag{15}$$

where $x_{B,j}^{post}$ and $x_{B,j}^{pri}$ denote the $j$-th entry of $\mathbf{x}_B^{post}$ and $\mathbf{x}_B^{pri}$. $\mathrm{E}(\cdot)$ is with respect to the posterior distribution of $\mathbf{x}$, which is given by $p(x_j|x_{B,j}^{pri}) \propto p(x_{B,j}^{pri}|x_j)p(x_j)$, where $p(x_j)$ is given in (8). Based on (13), $p(x_{B,j}^{pri}|x_j)$ is given by $\mathcal{CN}(x_{B,j}^{pri}; x_j, v_B^{pri})$. Note that as the entries of $\mathbf{x}$ are prior independent (from (8)) and according to the assumption in (13), the entries of $\mathbf{x}$ are also posterior independent. The extrinsic mean and variance of $\mathbf{x}$ can be computed by [11]

$$\mathbf{x}_A^{pri} = \mathbf{x}_B^{ext} = v_B^{ext}\left(\frac{\mathbf{x}_B^{post}}{v_B^{post}} - \frac{\mathbf{x}_B^{pri}}{v_B^{pri}}\right), \tag{16}$$

$$v_A^{pri} = v_B^{ext} = \left(\frac{1}{v_B^{post}} - \frac{1}{v_B^{pri}}\right)^{-1}. \tag{17}$$

### C. State Evolution of OAMP

The performance of OAMP can be characterized by a recursion of two states, $v_A^{pri}$ and $v_B^{pri}$. It is shown in [10], [11], [17] that the SE equations of OAMP are given by

$$v_B^{pri}(t+1) = \frac{Q}{L}\left(v_A^{pri}(t) + \sigma^2\right) - v_A^{pri}(t), \tag{18}$$

$$v_A^{pri}(t+1) = \left(\frac{1}{\mathrm{MMSE}\left(v_B^{pri}(t+1)\right)} - \frac{1}{v_B^{pri}(t+1)}\right)^{-1}, \tag{19}$$

where $\mathrm{MMSE}(v_B^{pri}) = \mathrm{E}(|x - \mathrm{E}(x|x+w)|^2)$, and $x$ is a sparse signal modeled as (8), $w \sim \mathcal{CN}(0, v_B^{pri})$, $t$ indicates the iteration index.

## IV. Hidden Markov Model for the Angular Domain Channel

In the multi-user scenario, the channel of each user $k$ can be expressed as

$$\mathbf{H}_k = \mathbf{A}_R\mathbf{X}_k\mathbf{A}_T^H \tag{20}$$

via the angular domain channel representation in (6), where $\mathbf{X}_k$ is the angular domain channel matrix of user $k$.

The channel model in (20) lacks a probability model for $\mathbf{X}_k$. Such a probability model provides the foundation for exploiting the structured sparsity of massive MIMO channels. As discussed in Section III-A, the existing structured sparsity models rely on restrictive assumptions. As a result, we shall introduce a hidden Markov model to capture a more realistic scattering structure at the Tx and the Rx of the multi-user massive MIMO channel. Fig. 3 illustrates the high level architecture of the HMM for the support of the massive MIMO channel matrices $\mathbf{X}_1, ..., \mathbf{X}_K$.

Let $\boldsymbol{\alpha}_k \in \{0,1\}^{\tilde{N}\times\tilde{M}}$ denote the *channel support matrix* of user $k$, whose $(n,m)$-th element, denoted by $\alpha_{k,n,m}$, indicates whether the channel coefficient $x_{k,n,m}$ is zero ($\alpha_{k,n,m} = 0$) or not ($\alpha_{k,n,m} = 1$). Conditioned on the channel support matrix $\boldsymbol{\alpha}_k$, the conditional prior distribution of the elements of $\mathbf{X}_k$ is

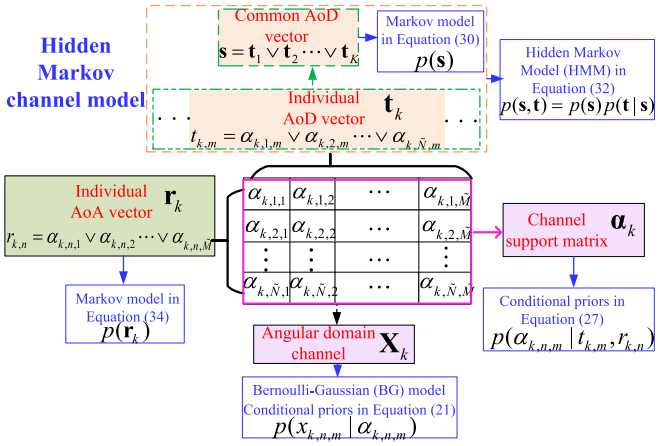



Fig. 3. Hidden Markov channel model.

independent and is given by

$$\underbrace{p\left(x_{k,n,m}|\alpha_{k,n,m}\right)}_{f_{k,n,m}\left(x_{k,n,m},\alpha_{k,n,m}\right)} = \left(1-\alpha_{k,n,m}\right)\delta\left(x_{k,n,m}\right)+\alpha_{k,n,m}\mathcal{CN}\left(x_{k,n,m};0,\sigma^2_{k,n,m}\right), \tag{21}$$

where $\mathcal{CN}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the probability density function of a complex Gaussian random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, given by

$$\mathcal{CN}\left(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}\right)=\frac{e^{-(\mathbf{x}-\boldsymbol{\mu})^H\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}}{\det\left(\pi\Sigma\right)}$$

where $\det(\mathbf{X})$ means the determinant of matrix $\mathbf{X}$.

For convenience, let $\boldsymbol{\alpha}=[\boldsymbol{\alpha}_1,...,\boldsymbol{\alpha}_K]$ denote the *aggregate channel support matrix* of all users. The prior distribution of $\boldsymbol{\alpha}$ captures the structured sparsity of multi-user massive MIMO channels $\{\mathbf{X}_k\}$. In this paper, we propose a hierarchical prior, called the *hidden Markov Model (HMM) prior*, for $\boldsymbol{\alpha}$, as illustrated in Fig. 3. Specifically, the HMM introduces a second layer of random vectors $\mathbf{t}_1,...,\mathbf{t}_K,\mathbf{s},\mathbf{r}_1,...,\mathbf{r}_K$. The random vector $\mathbf{t}_k=[t_{k,1},...,t_{k,\tilde{M}}]\in\{0,1\}^{\tilde{M}}$ with

$$t_{k,m}=\alpha_{k,1,m}\vee\alpha_{k,2,m}\vee\cdots\vee\alpha_{k,\tilde{N},m} \tag{22}$$

represents the *AoD vector* of user $k$. In other words, $t_{k,m}=1$ indicates that there is an active path from the $m$-th AoD direction at the BS and thus the $m$-th column of $\mathbf{X}_k$ is a non-zero vector, and $t_{k,m}=0$ indicates the opposite. The random vector $\mathbf{s}=[s_1,...,s_{\tilde{M}}]\in\{0,1\}^{\tilde{M}}$ with

$$s_m=t_{1,m}\vee t_{2,m}\vee\cdots\vee t_{K,m} \tag{23}$$

represents the *common AoD vector*. The random vector $\mathbf{r}_k=[r_{k,1},...,r_{k,\tilde{N}}]\in\{0,1\}^{\tilde{N}}$ with

$$r_{k,n}=\alpha_{k,n,1}\vee\alpha_{k,n,2}\vee\cdots\vee\alpha_{k,n,\tilde{M}} \tag{24}$$

represents the *AoA vector* of user $k$. In other words, $r_{k,n}=1$ indicates that there is an active path to the $n$-th AoA direction at user $k$ and thus the $n$-th row of $\mathbf{X}_k$ is a non-zero vector, and $r_{k,n}=0$ indicates the opposite. For convenience, define $\mathbf{t}\triangleq[\mathbf{t}_1^T,...,\mathbf{t}_K^T]^T$ and $\mathbf{r}=[\mathbf{r}_1^T,...,\mathbf{r}_K^T]^T$. Then the HHM prior distribution (joint distribution of $\boldsymbol{\alpha},\mathbf{s},\mathbf{t},\mathbf{r}$) is given by

$$p\left(\boldsymbol{\alpha},\mathbf{s},\mathbf{t},\mathbf{r}\right)=\underbrace{p\left(\mathbf{s}\right)}_{\text{Common AoD}}\underbrace{p\left(\mathbf{t}|\mathbf{s}\right)}_{\text{AoD}}\underbrace{p\left(\mathbf{r}\right)}_{\text{AoA}}\underbrace{p\left(\boldsymbol{\alpha}|\mathbf{t},\mathbf{r}\right)}_{\text{channel support}}, \tag{25}$$

where the the probability model for channel support $\boldsymbol{\alpha}$ is conditioned on AoD and AoA vectors, the common AoD vector $\mathbf{s}$ forms a Markov model, the joint distribution of $\mathbf{s}$ and individual AoD vectors $\mathbf{t}$ forms a hidden Markov model, and the individual AoA vector $\mathbf{r}$ forms a Markov model, as detailed below.

### A. Probability Model for Channel Support $\boldsymbol{\alpha}$

The conditional prior $p(\boldsymbol{\alpha}|\mathbf{t},\mathbf{r})$ is given by

$$p\left(\boldsymbol{\alpha}|\mathbf{t},\mathbf{r}\right)=\prod_{k,n,m}p\left(\alpha_{k,n,m}|t_{k,m},r_{k,n}\right), \tag{26}$$

i.e., conditioned on $\mathbf{t},\mathbf{r}$, $\alpha_{k,n,m}$'s are independent. By definition, $\alpha_{k,n,m}$ must be zero if either $t_{k,m}$ or $r_{k,n}$ is zero. When both $t_{k,m}$ and $r_{k,n}$ are one, the probability of $\alpha_{k,n,m}=1$ is non-zero. Therefore, we can model the conditional prior distribution of $\alpha_{k,n,m}$ as

$$\underbrace{p\left(\alpha_{k,n,m}|t_{k,m},r_{k,n}\right)}_{\eta_{k,n,m}\left(\alpha_{k,n,m},t_{k,m},r_{k,n}\right)}=\left(1-t_{k,m}r_{k,n}\right)\delta\left(\alpha_{k,n,m}\right)+t_{k,m}r_{k,n}\left(\gamma^k_\alpha\right)^{\alpha_{k,n,m}}\left(1-\gamma^k_\alpha\right)^{1-\alpha_{k,n,m}}, \tag{27}$$

where $\gamma^k_\alpha$ is the probability of $\alpha_{k,n,m}=1$ conditioned on $t_{k,m}=r_{k,n}=1$.

The HHM channel model is motivated by the physical propagation mechanism of radio waves. Due to the different heights of the BS and users, the scatterers in the propagation environment can be naturally classified into Tx scatterers at the BS side and local scatterers at the user side, where the Tx scatterers correspond to the AoDs from the BS, and the local scatterers correspond to the AoAs to each user $k$. In practice, the BS is usually elevated high, and thus the Tx scatterers at the BS side are usually higher/larger, making it easier for a group of users to share a set of common Tx scatterers. On the other hand, the probability that a set of users share a large portion of local scatterers is much smaller. As a result, it is natural to model the channel support matrix $\boldsymbol{\alpha}$ using a conditional prior conditioned on the "product" of the AoD and AoA vector, where a hidden Markov model $p(\mathbf{s})p(\mathbf{t}|\mathbf{s})$ in (32) is used to capture the clustered and common AoD structure, as will be explained in detail in Sections IV-B and IV-C; and a Markov model $p(\mathbf{r})$ in (34) is used to capture the clustered structure of each individual AoA vector, as will be explained in Section IV-D.

### B. Markov Model of Common AoD

Physically, each Tx scatterer at the BS side corresponds to an AoD, as shown in Fig. 4. If user $k$ can see a Tx scatterer corresponding to the $m$-th AoD (we say a user can see a Tx scatterer if the signal reflected/scattered from this Tx scatterer can reach the user), there will be a non-zero element in the $m$-th column of the angular domain channel matrix $\mathbf{X}_k$. Therefore,

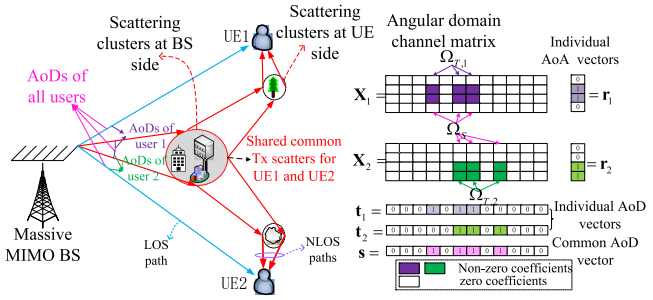



Fig. 4. Illustration of the structured sparsity of a multi-user massive MIMO channel for $K = 2$ users. In this example, the non-line-of-sight (NLOS) paths of the two users share a set of common Tx scatterers, and thus the associated AoDs are also shared by both users. On the other hand, the two users lie in different directions of the BS and thus the line-of-sight (LOS) paths do not share a common AoD. The random common AoD structure can be captured using the common AoD vector $\mathbf{s}$. Due to the limited and clustered scattering at the BS, the common AoD vector $\mathbf{s}$ will be sparse with clustered non-zero elements. Similarly, each AoA vector $\mathbf{r}_k$ is also sparse with clustered non-zero elements.

the *AoD set*

$$\Omega_{T,k} \triangleq \{m:\ t_{k,m} = 1\} \tag{28}$$

indicates the set of AoDs of user $k$ (which corresponds to the Tx scatterers seen by user $k$), and the *common AoD set*

$$\Omega_S \triangleq \{m:\ s_m = 1\} = \cup_{k=1}^{K} \Omega_{T,k} \tag{29}$$

indicates the set of AoDs of all users (which corresponds to all the Tx scatterers at the BS side).

Since there are limited and clustered Tx scatterings at the BS side, the non-zero elements of common AoD vector $\mathbf{s}$ will be sparse and clustered, as shown in Fig. 4. Such clustered sparsity of $\mathbf{s}$ can be modeled using a Markov prior as

$$p(\mathbf{s}) = \underbrace{p(s_1)}_{h_{S,1}(s_1)} \prod_{m=1}^{\tilde{M}-1} \underbrace{p(s_{m+1}|s_m)}_{h_{S,m+1}(s_m, s_{m+1})}, \tag{30}$$

with the transition probability given by $p(s_{m+1} = 0|s_m = 1) = \rho_{1,0}^S$ and $p(s_{m+1} = 1|s_m = 0) = \rho_{0,1}^S$, where $\rho_{1,0}^S$ and $\rho_{0,1}^S$ determine the average cluster size and the average gap between two clusters in $\mathbf{s}$. Intuitively, if the probability $p(s_{m+1} = 1|s_m = 1) = 1 - \rho_{1,0}^S$ is large, the non-zero elements in the common AoD vector $\mathbf{s}$ will tend to be more clustered with random locations and sizes. Hence, a smaller $\rho_{1,0}^S$ leads to a larger average cluster size. Similarly, a smaller $\rho_{0,1}^S$ leads to a larger average gap between two clusters. Therefore, the Markov prior in (30) provides a natural probability model for the random cluster structures of the common AoD vector $\mathbf{s}$. The initial distribution $p(s_1)$ is set to be the steady state distribution of the Markov chain in (30), i.e.,

$$\lambda_S \triangleq p(s_1 = 1) = \frac{\rho_{0,1}^S}{\rho_{0,1}^S + \rho_{1,0}^S}. \tag{31}$$

This ensures that all elements of $\mathbf{s}$ have the same marginal distribution $p(s_m) = \lambda_S^{s_m}(1 - \lambda_S)^{1-s_m}, \forall m$.

### C. Hidden Markov Model of Individual AoD

In practice, a group of users may share some common Tx scatterers, and therefore they share partially common AoD sets [5], [8], as illustrated in Fig. 4. In practice, different user groups may have overlaps. Moreover, the partially common AoDs shared by different user groups have random locations and sizes. However, the existing models in [5] and [8] cannot capture such a random common AoD structure that occurs in practice.

In this paper, the joint prior distribution of $\mathbf{s}, \mathbf{t}$ is modeled as a hidden Markov model (HMM):

$$p(\mathbf{s}, \mathbf{t}) = p(\mathbf{s}) \prod_{k=1}^{K} p(\mathbf{t}_k|\mathbf{s})$$

$$= p(s_1) \prod_{k=1}^{K} \underbrace{p(t_{k,1}|s_1)}_{h_{T,k,1}(s_1, t_{k,1})} \prod_{m=2}^{\tilde{M}} p(s_m|s_{m-1}) \prod_{k=1}^{K} \underbrace{p(t_{k,m}|s_m)}_{h_{T,k,m}(s_m, t_{k,m})}, \tag{32}$$

where $\mathbf{s}$ is the hidden Markov state variable and $\mathbf{t}_1, \cdots, \mathbf{t}_K$ are the symbol sequences that are generated by the hidden state $\mathbf{s}$ in the HMM. Note that $\mathbf{t}_1, ..., \mathbf{t}_K$ are independent conditioned on $\mathbf{s}$, and $p(t_{k,m} = 1|s_m = 0) = 0, p(t_{k,m} = 1|s_m = 1) = q_k^T$, where

$$q_k^T \triangleq \frac{|\Omega_{T,k}|}{|\Omega_S|} \tag{33}$$

measures the the portion of the Tx scatterers that can be seen by user $k$ and $q_k^T$'s characterize the extent of common support sharing among the $K$ users. For example, when $q_k^T = 1, \forall k$, all users will share the same common AoD set, i.e., there is only a single user group and $\Omega_{T,k} = \Omega_S, \forall k$. When $q_k^T$'s are small, different user groups will have limited common AoD sets (with random locations and sizes), meaning that they only have limited common scattering structure at the Tx.

### D. Markov Model of Individual AoA

Similarly, since there are limited and clustered Rx scatterings at the user side, the non-zero elements of $\mathbf{r}_k, \forall k$ will be sparse and clustered, as shown in Fig. 4. Such clustered sparsity of $\mathbf{r}_k, \forall k$ can be modeled using a Markov prior as

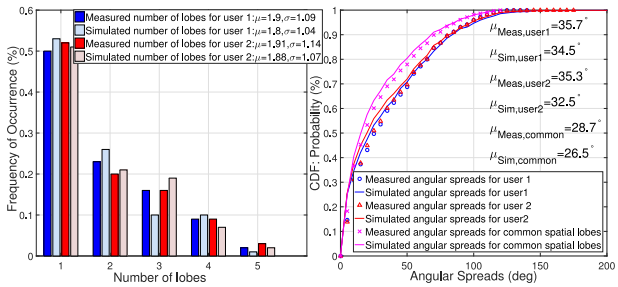
$$p(\mathbf{r}_k) = \underbrace{p(r_{k,1})}_{\triangleq h_{R,k,1}(r_{k,1})} \prod_{n=1}^{\tilde{N}-1} \underbrace{p(r_{k,n+1}|r_{k,n})}_{\triangleq h_{R,k,n+1}(r_{k,n}, r_{k,n+1})}, \tag{34}$$

with the transition probability given by $p(r_{k,n+1} = 1|r_{k,n} = 0) = \rho_{0,1}^k$ and $p(r_{k,n+1} = 0|r_{k,n} = 1) = \rho_{1,0}^k$. The initial distribution $p(r_{k,1})$ is set to be the steady state distribution of the Markov chain in (34), i.e.,

$$\lambda_{R,k} \triangleq p(r_{k,1} = 1) = \frac{\rho_{0,1}^k}{\rho_{0,1}^k + \rho_{1,0}^k}. \tag{35}$$

Finally, the sparsity ratio of $\mathbf{X}_k$ (i.e., the sparsity level over the total number of elements of $\mathbf{X}_k$) is

$$\lambda_k = \frac{q_k^T \rho_{0,1}^S \rho_{0,1}^k \gamma_\alpha^k}{\left(\rho_{0,1}^S + \rho_{1,0}^S\right)\left(\rho_{0,1}^k + \rho_{1,0}^k\right)}. \tag{36}$$




(a) (b)

Fig. 5. Comparison of measured channel property extracted from 28-GHz NLOS mm-SSCM [20] and simulated channel property extracted from the HMM channel. We set $K = 2, M = 256$ and $N = 1$. (a) Number of AoD lobes. (b) AoD global angular spread.

### E. Verification of Hidden Markov Channel Model

Compared to the existing models in [5], [8], the proposed HMM channel prior provides more flexibility to model the complicated random cluster structures of the common AoD vector and individual AoD/AoA vectors that may occur in practice. Moreover, the proposed Turbo-OAMP does not require knowledge of the HMM channel parameters. These parameters can be automatically learned by the Turbo-OAMP based on the expectation maximization framework. As a result, the proposed HMM channel and Turbo-OAMP algorithm work well for realistic channels including both non-line-of-sight (NLOS) and line-of-sight (LOS) channels.

In this section, we will provide some verification of the proposed HMM channel.

Fig. 5-a shows a typical empirical histogram plot of the number of AoD spatial lobes (SLs) [20] extracted from 28-GHz NLOS millimeter-wave statistical spatial channel model (mm-SSCM) proposed in [20], next to the simulated histograms extracted from the proposed HMM channel. It can be seen that the proposed HMM channel prior yields good agreement with the practical mmWave channels. For example, the mean numbers of AoD SLs are 1.9 (1.91) and 1.8 (1.88) for measured channel and simulated channel, respectively, for user 1 (user 2), which are close to each other.

AoD global angular spread [20] describes the degree of angular dispersion at the BS. The AoD global angular spreads in Fig. 5-b are computed based on a $-10$ dB lobe threshold from the measured data in 28-GHz NLOS mm-SSCM [20], and are compared with simulated values using the HMM channel. It can be seen that the simulated and measured global angular spreads match well, not only for individual users, but also for common spatial lobes between users. For example, the mean angular spread values for measured channel and simulated channel are $35.7^\circ$ ($35.3^\circ$) and $34.5^\circ$ ($32.5^\circ$), respectively, for user 1 (user 2), which are again, close.

## V. Sparse Channel Recovery for Massive MIMO

### A. Sparse Channel Estimation Formulation

Using the angular domain channel representation, (2) can be rewritten as a standard CS model [2], [5], [6] as

$$\mathbf{y}_k = \mathbf{F}_k \mathbf{x}_k + \mathbf{n}_k, \forall k, \tag{37}$$

where the measurement matrix $\mathbf{F}_k$ is given by

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{v}_{k,1}^T \otimes \mathbf{U}_{k,1}^H \\ \vdots \\ \mathbf{v}_{k,T_p/N_b}^T \otimes \mathbf{U}_{k,T_p/N_b}^H \end{bmatrix} (\mathbf{A}_T^* \otimes \mathbf{A}_R) \in \mathbb{C}^{T_p \times \tilde{N}\tilde{M}}, \tag{38}$$

$\mathbf{x}_k = \mathrm{vec}(\mathbf{X}_k) \in \mathbb{C}^{\tilde{N}\tilde{M}}$ is the sparse channel vector.

Note that the CE performance depends heavily on the measurement matrix $\mathbf{F}_k$, which in turn depends on the choice of $\mathbf{v}_{k,i}$ and $\mathbf{U}_{k,i}$. In this paper, the training vector $\mathbf{v}_{k,i}$ and combining matrix $\mathbf{U}_{k,i}$ are chosen to be $\mathbf{v}_{k,i} = \mathbf{E}_M \mathbf{P}_{k,i} \mathbf{E}_M \mathbf{b}_{k,i}$ and $\mathbf{U}_{k,i} = \mathbf{B}_{k,i}$, respectively, where $\mathbf{E}_M$ is the DFT matrix of dimension $M$, $\mathbf{P}_{k.i} \in \{0,1\}^{M\times M}$ is a random permutation matrix generated by a randomly reordered $M \times M$ identity matrix, $\mathbf{b}_{k,i} \in \{0,1\}^{M\times 1}$ and $\mathbf{B}_{k,i} \in \{0,1\}^{N\times N_b}$ are selection matrices consisting of randomly selected and reordered columns of the $M \times M$ and $N \times N$ identity matrices, respectively. Moreover, $\mathbf{b}_{k,i}$'s and $\mathbf{B}_{k,i}$'s should satisfy the constraint $\mathbf{b}_{k,i}^T \mathbf{b}_{k,j} \otimes \mathbf{B}_{k,i}^T \mathbf{B}_{k,j} = \mathbf{0}, \forall i,j \in \{1,\cdots,T_p/N_b\}$, which is used to ensure that $\mathbf{F}_k$ is a partial unitary matrix. A sufficient condition for $\mathbf{b}_{k,i}^T \mathbf{b}_{k,j} \otimes \mathbf{B}_{k,i}^T \mathbf{B}_{k,j} = \mathbf{0}$ is that $\mathbf{b}_{k,i}^T \mathbf{b}_{k,j} = 0$ or $\mathbf{B}_{k,i}^T \mathbf{B}_{k,j} = 0$. Specifically, $\mathbf{b}_{k,i}^T \mathbf{b}_{k,j} = 0$ means $\mathbf{b}_{k,i}$ and $\mathbf{b}_{k,j}$ are selected from different columns of the $M \times M$ identity matrix. $\mathbf{B}_{k,i}^T \mathbf{B}_{k,j} = 0$ means $\mathbf{B}_{k,i}$ and $\mathbf{B}_{k,j}$ do not share any common column. The above choice of training vectors and combining matrices has several advantages. First, the training vector $\mathbf{v}_{k,i} = \mathbf{E}_M \mathbf{P}_{k,i} \mathbf{E}_M \mathbf{b}_{k,i}$ ensures that the transmit power of each antenna at the BS is similar and we can always find a feasible RF precoder $\mathbf{W}_0^i$ with constant envelope elements and a baseband precoder $\mathbf{g}_0^i$ to form the training vector $\mathbf{W}_0^i \mathbf{g}_0^i = \mathbf{E}_M \mathbf{P}_{k,i} \mathbf{E}_M \mathbf{b}_{k,i}$, as long as $M_b \geq 2$ [13]. Second, the combining matrix $\mathbf{U}_{k,i} = \mathbf{B}_{k,i}$ can be easily implemented using a phase-shifting network or a simple RF switch since $\mathbf{U}_{k,i}$ only contains elements of 0 or 1. Finally, the resulting measurement matrix $\mathbf{F}_k$ is a partial unitary matrix, which is shown in [10], [11] to be beneficial for both algorithm design and recovery performance. The superior performance of such a design is verified in the simulations in Section VII.

Our primary goal is to estimate the structured sparse channel vector $\mathbf{x} = [\mathbf{x}_1^T, ..., \mathbf{x}_K^T]^T$, given the observations $\mathbf{y} = [\mathbf{y}_1^T, ..., \mathbf{y}_K^T]^T$ in model (37). In particular, we are interested in computing minimum mean-squared error (MMSE) estimates of $\{x_{k,n,m}\}$, $\hat{x}_{k,n,m} = \mathrm{E}(x_{k,n,m}|\mathbf{y})$ where the expectation is over the marginal posterior $p(x_{k,n,m}|\mathbf{y})$, where

$$\begin{aligned} p(x_{k,n,m}|\mathbf{y}) &\propto \int_{\mathbf{x}_{-(k,n,m)}} \sum_{\mathbf{s},\mathbf{t},\mathbf{r},\boldsymbol{\alpha}} p(\mathbf{y},\mathbf{x},\boldsymbol{\alpha},\mathbf{s},\mathbf{t},\mathbf{r}) \\ &= \int_{\mathbf{x}_{-(k,n,m)}} \sum_{\mathbf{s},\mathbf{t},\mathbf{r},\boldsymbol{\alpha}} p(\mathbf{s})\, p(\mathbf{t}|\mathbf{s})\, p(\mathbf{r}) \prod_{k=1}^{K} \prod_{i=1}^{T_p} p(y_{k,i}|\mathbf{x}_k) \\ &\quad \times \prod_{k,n,m} p(x_{k,n,m}|\alpha_{k,n,m})\, p(\alpha_{k,n,m}|t_{k,m}, r_{k,n}). \end{aligned} \tag{39}$$

$\mathbf{x}_{-(k,n,m)}$ denotes the vector $\mathbf{x}$ excluding the element $x_{k,n,m}$, $p(y_{k,i}|\mathbf{x}_k) = \mathcal{CN}(y_{k,i}; \mathbf{f}_{k,i}^T \mathbf{x}_k, \sigma^2)$, $y_{k,i}$ is the $i$-th element of $\mathbf{y}_k$ and $\mathbf{f}_{k,i}^T$ is the $i$-th row of $\mathbf{F}_k$. We use $\propto$ to denote equality after scaling. One challenge in computing the MMSE estimate is the

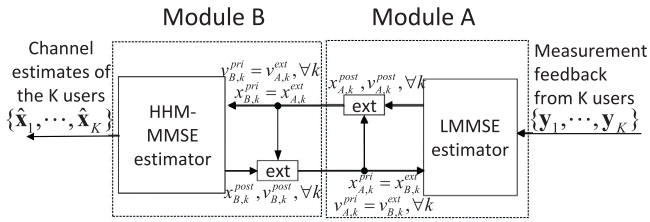



Fig. 6. Modules of the Turbo-OAMP algorithm and message flows between different modules.

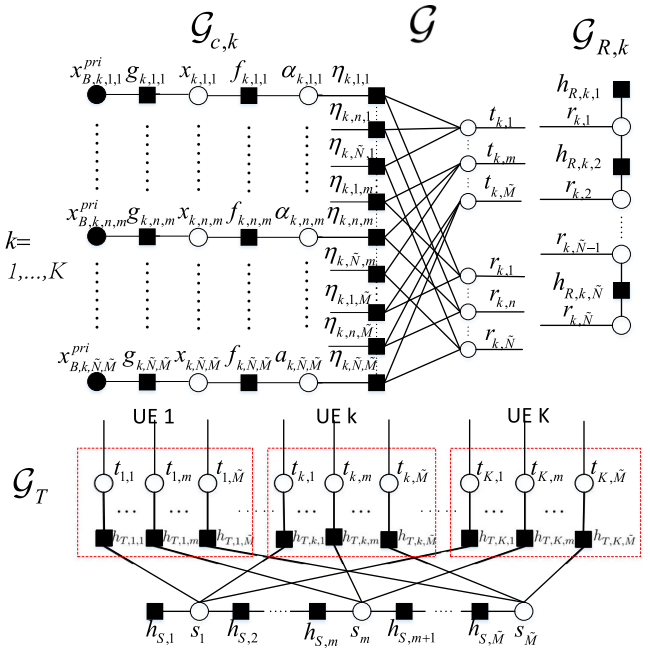



Fig. 7. Factor graph of the joint distribution in (39).

exact posterior calculation in (39) because the factor graph of the underlying model in (39) has loops (as illustrated in Fig. 7). In the next subsection, we propose a centralized Turbo-OAMP algorithm at the BS to approximately calculate the marginal posteriors $\{p(x_{k,n,m}|\mathbf{y})\}$ and the associated MMSE estimates, which will be shown in the simulations to achieve a good performance.

### *B. Turbo-OAMP Algorithm: Centralized Solution at the BS*

In the centralized Turbo-OAMP, user $k$ first feeds back $\mathbf{y}_k$ to the BS. Then BS runs the Turbo-OAMP algorithm to estimate the individual channel vectors $\{\mathbf{x}_k\}$ of the $K$ users based on the compressed aggregate CSI measurements $\{\mathbf{y}_k\}$ fed back from all users. Such a design allows us to exploit the joint sparsity (common AoD structure) among the user channels to enhance the CE performance.

*1) Modules of the Turbo-OAMP Algorithm:* The proposed Turbo-OAMP (Algorithm 1) contains two modules, as shown in Fig. 6. Module A is a LMMSE estimator based on the observation $\mathbf{y}$ and messages from Module B. Module B performs MMSE estimation that combines the HMM channel prior in (25) and the messages from Module A. The two modules are executed iteratively until convergence.

**Algorithm 1:** Turbo-OAMP Algorithm.

**Input:** $\{\mathbf{y}_1, \cdots, \mathbf{y}_K\}$, measurement matrix $\mathbf{F}_k, \forall k$, and addition noise variance $\sigma^2$.
**Output:** $\{\hat{\mathbf{x}}_1, \cdots, \hat{\mathbf{x}}_K\}$.
**Initialize:** $\mathbf{x}_{A,k}^{pri} = \mathbf{0}, \forall k$, $v_{A,k}^{pri} = \frac{\lambda_k}{\tilde{N}\tilde{M}} \sum_{n=1}^{\tilde{N}} \sum_{m=1}^{\tilde{M}} \sigma_{k,n,m}^2, \forall k$, $\nu_{t_{k,m} \to f_{k,n,m}}(s) = \frac{\rho_{0,1}^S q_k^T}{\rho_{0,1}^S + \rho_{1,0}^S}, \forall k, n, m, t = 0$.
**Module A:**
**% LMMSE estimator**
1: Update $\mathbf{x}_{A,k}^{post}$ and $v_{A,k}^{post}$, $\forall k$ using (40) and (41).
**% Update extrinsic**
2: Update $\mathbf{x}_{B,k}^{pri} = \mathbf{x}_{A,k}^{ext}$ and $v_{B,k}^{pri} = v_{A,k}^{ext}, \forall k$, using (43) and (44).
**Module B:**
**% HMM-MMSE estimator (Message Passing over $\mathcal{G}$)**
3: Message passing over the path $x_{k,n,m} \to f_{k,n,m} \to \alpha_{k,n,m} \to \eta_{k,n,m} \to r_{k,n}, \forall k$ using Eqs. (55)–(58), with $\{\mathbf{x}_{B,k}^{pri}\}$ and $\{v_{B,k}^{pri}\}$ as the input.
4: AoA vector estimation for $k = 1, \cdots, K$ using Eqs. (59)–(62).
5: Message passing over the path $r_{k,n} \to \eta_{k,n,m} \to t_{k,m}, \forall k$ using Eqs. (63)–(65).
6: AoD vector estimation for $k = 1, \cdots, K$ using Eqs. (66)–(72).
7: Message passing over the path $t_{k,m} \to \eta_{k,n,m} \to \alpha_{k,n,m} \to f_{k,n,m} \to x_{k,n,m}, \forall k$ using Eqs. (73)–(75).
8: Calculate the approximate posterior distributions $\hat{p}(x_{k,n,m}|\mathbf{x}_B^{pri})$ using (46), and update $\mathbf{x}_{B,k}^{post}$ and $v_{B,k}^{post}$ ,$\forall k$ using (47) and (48).
**% Update extrinsic**
9: Update $\mathbf{x}_{A,k}^{pri} = \mathbf{x}_{B,k}^{ext}$ and $v_{A,k}^{pri} = v_{B,k}^{ext}, \forall k$ using (49) and (50).
Repeat Module A and Module B until convergence or the maximum iteration number is exceeded.

In Module A, channel vector $\mathbf{x}_k$ of user $k$ is estimated based on the observation $\mathbf{y}_k$ with a prior distribution $\mathcal{CN}(\mathbf{x}_k; \mathbf{x}_{A,k}^{pri}, v_{A,k}^{pri}\boldsymbol{I})$, where $\mathbf{x}_{A,k}^{pri}$ and $v_{A,k}^{pri}$ are the extrinsic mean and variance from the HMM-MMSE estimator that will be elaborated in detail later. Then the posterior distribution of $\mathbf{x}_k$ is still complex Gaussian with mean and variance given by

$$\mathbf{x}_{A,k}^{post} = \mathbf{x}_{A,k}^{pri} + \frac{v_{A,k}^{pri}}{v_{A,k}^{pri} + \sigma^2} \mathbf{F}_k^H (\mathbf{y}_k - \mathbf{F}_k \mathbf{x}_{A,k}^{pri}), \tag{40}$$

$$v_{A,k}^{post} = v_{A,k}^{pri} - \frac{T_p}{\tilde{N}\tilde{M}} \cdot \frac{(v_{A,k}^{pri})^2}{v_{A,k}^{pri} + \sigma^2}. \tag{41}$$

After that, we need to calculate the extrinsic message passing [19], which can decorrelate the input and output messages of

TABLE I
FACTORS, DISTRIBUTIONS AND FUNCTIONAL FORMS IN FIG. 7

| Factor | Distribution | Functional form |
|---|---|---|
| $g_{k,n,m}(x^{pri}_{B,k,n,m}, x_{k,n,m})$ | $p\left(x_{k,n,m}\vert x^{pri}_{B,k,n,m}\right)$ | $\mathcal{CN}(x_{k,n,m}; x^{pri}_{B,k,n,m}, v^{pri}_{B,k})$ |
| $f_{k,n,m}\left(x_{k,n,m}, \alpha_{k,n,m}\right)$ | $p\left(x_{k,n,m}\vert\alpha_{k,n,m}\right)$ | $(1-\alpha_{k,n,m})\,\delta\left(x_{k,n,m}\right)+\alpha_{k,n,m}\mathcal{CN}\left(x_{k,n,m}; 0, \sigma^2_{k,n,m}\right).$ |
| $\eta_{k,n,m}\left(\alpha_{k,n,m}, t_{k,m}, r_{k,n}\right)$ | $p\left(\alpha_{k,n,m}\vert t_{k,m}, r_{k,n}\right)$ | $t_{k,m}r_{k,n}\left(\gamma^k_\alpha\right)^{\alpha_{k,n,m}}\left(1-\gamma^k_\alpha\right)^{1-\alpha_{k,n,m}}+(1-t_{k,m}r_{k,n})\,\delta\left(\alpha_{k,n,m}\right)$ |
| $h_{T,k,m}\left(s_m, t_{k,m}\right)$ | $p\left(t_{k,m}\vert s_m\right)$ | $p\left(t_{k,m}=1\vert s_m=0\right)=0,\, p\left(t_{k,m}=1\vert s_m=1\right)=q^T_k$ |
| $h_{S,1}(s_1)$ | $p\left(s_1\right)$ | $\left(\lambda_S\right)^{s_1}\left(1-\lambda_S\right)^{1-s_1}$ |
| $h_{S,m+1}\left(s_m, s_{m+1}\right)$ | $p\left(s_{m+1}\vert s_m\right)$ | $\begin{cases}\left(\rho^S_{01}\right)^{s_{m+1}}\left(1-\rho^S_{01}\right)^{1-s_{m+1}}, & s_m=0\\ \left(1-\rho^S_{10}\right)^{s_{m+1}}\left(\rho^S_{10}\right)^{1-s_{m+1}} & s_m=1\end{cases}$ |
| $h_{R,k,1}\left(r_{k,1}\right)$ | $p\left(r_{k,1}\right)$ | $\left(\lambda_{R,k}\right)^{r_{k,1}}\left(1-\lambda_{R,k}\right)^{1-r_{k,1}}$ |
| $h_{R,k,n+1}\left(r_{k,n}, r_{k,n+1}\right)$ | $p\left(r_{k,n+1}\vert r_{k,n}\right)$ | $\begin{cases}\left(\rho^k_{01}\right)^{r_{k,n+1}}\left(1-\rho^k_{01}\right)^{1-r_{k,n+1}}, & r_{k,n}=0\\ \left(1-\rho^k_{10}\right)^{r_{k,n+1}}\left(\rho^k_{10}\right)^{1-r_{k,n+1}} & r_{k,n}=1\end{cases}$ |

the estimator. The extrinsic distribution of $\mathbf{x}_k$ satisfies

$$\mathcal{CN}(\mathbf{x}_k; \mathbf{x}^{post}_{A,k}, v^{post}_{A,k}\boldsymbol{I}) \propto \mathcal{CN}(\mathbf{x}_k; \mathbf{x}^{pri}_{A,k}, v^{pri}_{A,k}\boldsymbol{I})\mathcal{CN}(\mathbf{x}_k; \mathbf{x}^{ext}_{A,k}, v^{ext}_{A,k}\boldsymbol{I}). \quad (42)$$

Therefore, the extrinsic mean and variance are given by

$$\mathbf{x}^{pri}_{B,k} = \mathbf{x}^{ext}_{A,k} = v^{ext}_{A,k}\left(\frac{\mathbf{x}^{post}_{A,k}}{v^{post}_{A,k}} - \frac{\mathbf{x}^{pri}_{A,k}}{v^{pri}_{A,k}}\right), \quad (43)$$

$$v^{pri}_{B,k} = v^{ext}_{A,k} = \left(\frac{1}{v^{post}_{A,k}} - \frac{1}{v^{pri}_{A,k}}\right)^{-1}, \quad (44)$$

In Module B, the extrinsic calculation is similar to that in Module A, but the HMM-MMSE estimator is more complicated. Compared to the OAMP in [11], the main difference is that the MMSE estimator for the i.i.d. prior in the OAMP is replaced by the HMM-MMSE estimator for the HMM prior in the Turbo-OAMP. Therefore, one major contribution of this paper from the algorithm aspect is the design of the HMM-MMSE estimator, which takes into account the structured sparsity of the multi-user massive MIMO channel as captured by the HMM channel prior in (25). Specifically, a message passing scheme will be designed for the HMM-MMSE estimator to calculate the posterior mean and variance of the channel $\mathbf{x}$, based on the HMM channel prior and the messages from Module A. The details of the HMM-MMSE estimator is presented in Section V-B2.

*2) Message Passing in Module B (HMM-MMSE Estimator):* In this subsection, we explain the details of Module B for HMM channel prior. First of all, a basic assumption is to model $\mathbf{x}^{pri}_{B,k}$, the extrinsic mean from the LMMSE estimator as given in (43), as an AWGN observation [10], [11], i.e.,

$$\mathbf{x}^{pri}_{B,k} = \mathbf{x}_k + \mathbf{z}_k, \forall k, \quad (45)$$

where $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0}, v^{pri}_{B,k}\boldsymbol{I})$ is independent of $\mathbf{x}_k$, and $v^{pri}_{B,k}$ is the extrinsic variance from the LMMSE estimator as given in (44). Similar assumptions have been widely used in other message passing algorithms [10]–[12], [18]. A formal proof of Assumption (45) has been provided in [21] for the AMP with an i.i.d. Gaussian measurement matrix and in [22], [23] for a wider class of sensing matrices. Extensive simulations have also been conducted in [10], [11] and [12] to verify the validity of Assumption (45) for the OAMP. The main advantage of replacing the original observation model in (37) with the approximate AWGN observation model in (45) is that the per iteration complexity of the message passing algorithm can be reduced from $O(T_p\tilde{N}\tilde{M})$ to only $O(\tilde{N}\tilde{M})$.

Let $\mathbf{x}^{pri}_B = [(\mathbf{x}^{pri}_{B,1})^T, ..., (\mathbf{x}^{pri}_{B,K})^T]^T$. Under Assumption (45), the factor graph of joint distribution $p(\mathbf{x}^{pri}_B, \mathbf{x}, \boldsymbol{\alpha}, \mathbf{s}, \mathbf{t}, \mathbf{r})$, denoted by $\mathcal{G}$, is shown in Fig. 7, where the function expression of each factor node is listed in Table I. The factor graph $\mathcal{G}$ contains $2K+1$ subgraphs: the AoD vector subgraph $\mathcal{G}_T$; $K$ AoA vector subgraphs $\mathcal{G}_{R,k}$'s; and $K$ channel coefficient subgraphs $\mathcal{G}_{c,k}$'s. Since $\mathcal{G}$ has loops, it is difficult to calculate the exact posterior distributions $\{p(x_{k,n,m}\vert\mathbf{x}^{pri}_B)\}$. Therefore, Module B aims at calculating the approximate posterior distributions $\{\hat{p}(x_{k,n,m}\vert\mathbf{x}^{pri}_B)\}$ using the sum-product message passing rule.

We now outline the message passing scheme on graph $\mathcal{G}$. The details are elaborated in Appendix A. According to the sum-product rule, the message passing over the path $x_{k,n,m} \rightarrow f_{k,n,m} \rightarrow \alpha_{k,n,m} \rightarrow \eta_{k,n,m} \rightarrow r_{k,n}$ are given by Eqs. (55)–(58), with $\{\mathbf{x}^{pri}_{B,k}\}$ and $\{v^{pri}_{B,k}\}$ as the input to update the messages $\{\nu_{\eta_{k,n,m}\rightarrow r_{k,n}}\}$. Then a forward-backward message passing is performed over the Markov chains $\mathbf{r}_k$ in $\mathcal{G}_{R,k}$, for $k=1,...,K$ respectively, through Eqs. (59)–(62). After this, the message is passed over the path $r_{k,n} \rightarrow \eta_{k,n,m} \rightarrow t_{k,m}$ through Eqs. (63)–(65) to update the messages $\{\nu_{\eta_{k,n,m}\rightarrow t_{k,m}}\}$. Then a forward-backward message passing is performed over the HMM associated with $\mathbf{t}$ in $\mathcal{G}_T$ through Eqs. (66)–(72). After this, the message is passed back over the path $t_{k,m} \rightarrow \eta_{k,n,m} \rightarrow \alpha_{k,n,m} \rightarrow f_{k,n,m} \rightarrow x_{k,n,m}$ using Eqs. (73)–(75) to update the messages $\{\nu_{f_{k,n,m}\rightarrow x_{k,n,m}}\}$.

After calculating the updated messages $\{\nu_{f_{k,n,m}\rightarrow x_{k,n,m}}\}$, the approximate posterior distributions are given by

$$\hat{p}(x_{k,n,m}\vert\mathbf{x}^{pri}_B) \propto \nu_{f_{k,n,m}\rightarrow x_{k,n,m}}\left(x_{k,n,m}\right)\nu_{g_{k,n,m}\rightarrow x_{k,n,m}}\left(x_{k,n,m}\right), \quad (46)$$

where $\nu_{g_{k,n,m}\rightarrow x_{k,n,m}}(x_{k,n,m}) = \mathcal{CN}(x_{k,n,m}; x^{pri}_{B,k,n,m}, v^{pri}_{B,k,n,m})$. Then the posterior mean and variance can be respectively

calculated as

$$x_{B,k,n,m}^{post} = \mathrm{E}(x_{k,n,m}|\mathbf{x}_B^{pri}) = \int_{x_{k,n,m}} x_{k,n,m}\hat{p}(x_{k,n,m}|\mathbf{x}_B^{pri}), \tag{47}$$

$$\begin{aligned} v_{B,k}^{post} &= \frac{1}{\tilde{N}\tilde{M}}\sum_{n=1}^{\tilde{N}}\sum_{m=1}^{\tilde{M}}\mathrm{Var}(x_{k,n,m}|\mathbf{x}_B^{pri}) \\ &= \frac{1}{\tilde{N}\tilde{M}}\sum_{n=1}^{\tilde{N}}\sum_{m=1}^{\tilde{M}}\int_{x_{k,n,m}} |x_{k,n,m} - \mathrm{E}(x_{k,n,m}|\mathbf{x}_B^{pri})|^2 \\ &\quad \times\hat{p}(x_{k,n,m}|\mathbf{x}_B^{pri}). \end{aligned} \tag{48}$$

Based on the derivation in [11], the corresponding extrinsic update can be calculated as

$$\mathbf{x}_{A,k}^{pri} = \mathbf{x}_{B,k}^{ext} = v_{B,k}^{ext}\left(\frac{\mathbf{x}_{B,k}^{post}}{v_{B,k}^{post}} - \frac{\mathbf{x}_{B,k}^{pri}}{v_{B,k}^{pri}}\right), \forall k \tag{49}$$

$$v_{A,k}^{pri} = v_{B,k}^{ext} = \left(\frac{1}{v_{B,k}^{post}} - \frac{1}{v_{B,k}^{pri}}\right)^{-1}, \forall k. \tag{50}$$

Finally, the overall Turbo-OAMP algorithm is summarized in Algorithm 1.

### C. Statistical Parameters Learning

In practice, the statistical parameters $\boldsymbol{\rho} \triangleq \{\rho_{0,1}^S, \rho_{1,0}^S, q_k^T, \rho_{0,1}^k, \rho_{1,0}^k, \gamma_\alpha^k, \sigma_{k,n,m}^2\}$ can be learned using the expectation maximization (EM) algorithm [24]. Specifically, the statistical parameters $\boldsymbol{\rho}$ are initialized using any available prior knowledge. Then in each iteration of the EM algorithm, we first perform one iteration of the Turbo-OAMP algorithm for given $\boldsymbol{\rho}$. After that, estimates of the marginal posterior distributions are available for $\mathbf{x}, \mathbf{t}, \mathbf{r}, \boldsymbol{\alpha}$. With these marginals, we can obtain closed-form update equations for $\boldsymbol{\rho}$, as shown in [25]. Since the focus of the paper is how to estimate the marginal posterior distributions of $\mathbf{x}, \mathbf{t}, \mathbf{r}, \boldsymbol{\alpha}$ for given $\boldsymbol{\rho}$, we omit the detailed update equations for $\boldsymbol{\rho}$ for conciseness. Interested readers please refer to [25] for the detailed derivation of the EM algorithm.

## VI. State Evolution Analysis

It is very difficult to directly analyze the MSE performance of Algorithm 1 because $\mathcal{G}$ has loops and the message passing over $\mathcal{G}$ is not exact. Therefore, in the following, we aim at deriving a lower bound for the MSE of Algorithm 1 based on the SE of a *gene-aided algorithm* which has perfect knowledge of the true AoD vector $\mathbf{t}$.

### A. SE of a Gene-Aided Algorithm

Compared to Algorithm 1, the only difference of the gene-aided algorithm is that the joint distribution $p(\mathbf{x}_B^{pri}, \mathbf{x}, \boldsymbol{\alpha}, \mathbf{t}, \mathbf{r})$ associated with the AWGN observation model (45) is replaced with the conditional joint distribution $p(\overline{\mathbf{x}}_B^{pri}, \mathbf{x}, \boldsymbol{\alpha}, \mathbf{r}|\mathbf{t})$, where

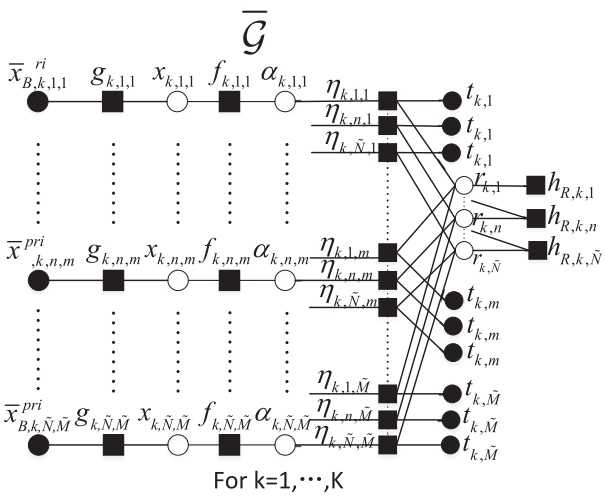



Fig. 8. Factor graph $\overline{\mathcal{G}}$ of the modified Module B in the gene-aided algorithm when the AoD vector $\mathbf{t}$ is perfectly known.

$\mathbf{t}$ is the true AoD vector. Note that for clarity, we use

$$\left\{\mathbf{x}_{A,k}^{pri}, \mathbf{x}_{B,k}^{pri}, \mathbf{x}_{A,k}^{post}, \mathbf{x}_{B,k}^{post}, v_{A,k}^{pri}, v_{B,k}^{pri}, v_{A,k}^{post}, v_{B,k}^{post}\right\}$$

to denote the key variables in Algorithm 1, and

$$\left\{\overline{\mathbf{x}}_{A,k}^{pri}, \overline{\mathbf{x}}_{B,k}^{pri}, \overline{\mathbf{x}}_{A,k}^{post}, \overline{\mathbf{x}}_{B,k}^{post}, \overline{v}_{A,k}^{pri}, \overline{v}_{B,k}^{pri}, \overline{v}_{A,k}^{post}, \overline{v}_{B,k}^{post}\right\}$$

to denote the variables in the gene-aided algorithm. The factor graph of $p(\overline{\mathbf{x}}_B^{pri}, \mathbf{x}, \boldsymbol{\alpha}, \mathbf{r}|\mathbf{t})$, denoted by $\overline{\mathcal{G}}$, is shown in Fig. 8. Consequently, the gene-aided algorithm can be obtained by replacing Steps 3–9 in Algorithm 1 (message passing algorithm over $\mathcal{G}$) with the message passing algorithm over $\overline{\mathcal{G}}$. Since $\overline{\mathcal{G}}$ has no loop, the modified Steps 3–9 (*modified Module B*) in the gene-aided algorithm can find the true posterior expectation conditioned on $\mathbf{t}_k$, i.e., $\overline{\mathbf{x}}_{B,k}^{post} = \mathrm{E}(x_{k,n,m}|\overline{\mathbf{x}}_B^{pri}, \mathbf{t}_k)$. Clearly, the MSE of the gene-aided algorithm provides a lower bound for the MSE of Algorithm 1.

When the resulting factor graph under the AWGN observation model in (45) has no loop, the performance of the OAMP can be characterized by a simple scalar recursion called SE [11]. A similar technique can be applied to the gene-aided algorithm by tracking the input variances $\overline{v}_{A,k}^{pri}$ and $\overline{v}_{B,k}^{pri}$ of Module A and the modified Module B as

$$\begin{aligned} \overline{v}_{B,k}^{pri}(t) &\triangleq \frac{1}{\tilde{N}\tilde{M}}\mathrm{E}\|\mathbf{x}_k - \overline{\mathbf{x}}_{B,k}^{pri}(t)\|_2^2 \\ &= \frac{\tilde{N}\tilde{M}}{T_p}(\overline{v}_{A,k}^{pri}(t-1) + \sigma^2) - \overline{v}_{A,k}^{pri}(t-1), \forall k, \end{aligned} \tag{51}$$

$$\begin{aligned} \overline{v}_{A,k}^{pri}(t) &\triangleq \frac{1}{\tilde{N}\tilde{M}}\mathrm{E}\|\mathbf{x}_k - \overline{\mathbf{x}}_{A,k}^{pri}(t)\|_2^2 \\ &= \left(\frac{1}{\overline{\mathrm{MMSE}}_k(\overline{v}_{B,k}^{pri}(t))} - \frac{1}{\overline{v}_{B,k}^{pri}(t)}\right)^{-1}, \forall k, \end{aligned} \tag{52}$$

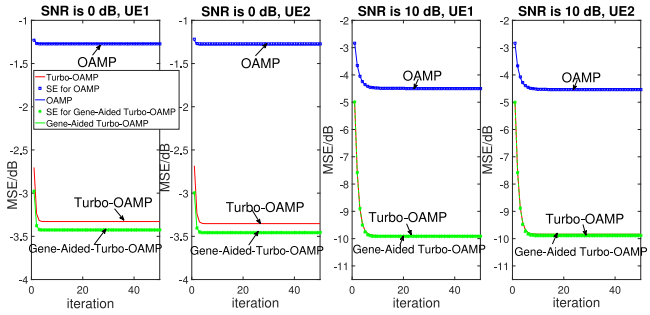



Fig. 9. MSE performance of the Turbo-OAMP under SNR $= 0$ dB and 10 dB. The system parameters are set as $K = 2$, $\tilde{M} = 64$, $\tilde{N} = 4$, $T_p = 80$, $N_b = 1$, $q_1^T = q_2^T = 0.9375$, $\rho_{0,1}^S = \rho_{1,0}^S = 0.25$, $\rho_{0,1}^1 = \rho_{1,0}^1 = \rho_{0,1}^2 = \rho_{1,0}^2 = 0.25$, $\lambda_S = 0.5$, $\lambda_{R,1} = \lambda_{R,2} = 0.5$ and $\gamma_\alpha^1 = \gamma_\alpha^2 = 1$.

where $t$ denotes the $t$-th iteration and

$$\overline{\text{MMSE}}_k(\bar{v}_{B,k}^{pri}) = \frac{1}{\tilde{N}\tilde{M}} \sum_{n,m} \text{E}\left[|x_{k,n,m} - \text{E}(x_{k,n,m}|\mathbf{x}_k + \mathbf{z}_k, \mathbf{t}_k)|^2\right] \quad (53)$$

is the MMSE achieved by the posterior mean estimation for the AWGN observation model $\mathbf{x}_{B,k}^{pri} = \mathbf{x}_k + \mathbf{z}_k$ conditioned on the true AoD vector $\mathbf{t}$. Note that (51) [10, (21a)] can be obtained by substituting $v_{A,k}^{post}$ in (41) into $v_{B,k}^{pri}$ in (44) in Algorithm 1, and (52) is obtained from [10, (21b)].

### B. MSE Lower Bound of Turbo-OAMP

The SE in (51) and (52) for the gene-aided algorithm provides a lower bound for the MSE performance of the Turbo-OAMP. To be more specific, define the MSE performance of the Turbo-OAMP at iteration $t$ as

$$\text{MSE}_k(t) = \frac{1}{\tilde{N}\tilde{M}} \text{E}\|\mathbf{x}_k - \mathbf{x}_{B,k}^{post}(t)\|_2^2, \forall k.$$

Here, we use $\text{MSE}_k(t)$ instead of $\text{MMSE}_k(t)$ to denote the MSE of the Turbo-OAMP at iteration $t$ because the message passing over $\mathcal{G}$ in the Turbo-OAMP is not exact and thus $\mathbf{x}_{B,k}^{post}(t)$ is not the true MMSE estimate. Then we have

$$\text{MSE}_k(t) \geq \overline{\text{MMSE}}_k(\bar{v}_{B,k}^{pri}(t)), \quad (54)$$

since the gene-aided algorithm can achieve a better MSE performance at each iteration due to the additional information on the true AoD vector $\mathbf{t}$. Extensive simulations show that the MSE lower bound in (54) obtained from the SE analysis of the gene-aided algorithm is quite tight at medium or high SNR. This is because by exploiting the structured sparsity in multi-user massive MIMO channels, the Turbo-OAMP can almost perfectly recover the AoD vector $\mathbf{t}$ at a medium or high SNR. As a result, it can approach the performance of the gene-aided algorithm. Similar results have also been observed in [5], [8].

### C. Comparison of the SEs of Different Algorithms

In Fig. 9, we compare the SE of the OAMP [11] and gene-aided algorithms, as well as the simulated MSE, under different SNRs. It can be seen that the performance of the Turbo-OAMP is close to both the SE prediction and MSE performance of the gene-aided algorithm, especially when the SNR is higher. Moreover, compared to the OAMP, the Turbo-OAMP also achieves faster convergence and smaller MSE (more than 5 dB MSE gain when the SNR is 10 dB).

## VII. Simulation Results Under Realistic Channel Models

In this section, we evaluate the performance of the proposed Turbo-OAMP algorithm under two realistic channel models: the spatial channel model (SCM) [26] developed in 3GPP/3GPP2 for low frequency bands (less than 6 GHz), and the millimeter-wave statistical spatial channel model (mm-SSCM) proposed in [20] for high frequency bands (28–73 GHz). In the following, we will use extensive simulations to verify that the proposed Turbo-OAMP can achieve superior performance over the following state-of-the-art baseline algorithms in these two realistic channel models.

**J-BurstLASSO** [8]**:** This algorithm is based on LASSO (convex optimization) and it can exploit the joint burst sparsity of multi-user massive MIMO channels.

**OAMP** [11]**:** The OAMP assumes i.i.d. sparse channel prior.

**J-OMP** [5]**:** The J-OMP is based on OMP and it can exploit the joint sparsity of multi-user channel matrices.

To apply the message-passing-based algorithms (OAMP and Turbo-OAMP) in practical channel estimation, we update the channel statistical parameters $\boldsymbol{\rho}$ by the EM framework [25]. Note that the OAMP only needs parameters $\{\lambda_k, \sigma_{k,n,m}^2\}$. On the other hand, for burst LASSO and J-OMP, the number of non-zero clusters and the average cluster size are learned from the channel data generated by the SCM/mm-SSCM in an offline way. Both low-frequency and high-frequency massive MIMO systems will be considered. In the low-frequency massive MIMO system, the BS has $M = 128$ antennas and $M_b = 128$ RF chains, each user has a single antenna, and SCM will be used to generate channels. In the high-frequency massive MIMO systems, the BS has $M = 256$ antennas and $M_b = 16$ RF chains, each user has $N = 8$ or $N = 2$ antennas and $N_b = 2$ or $N_b = 1$ RF chains, and mm-SSCM will be used to generate channels. We let $\tilde{M} = M$ and $\tilde{N} = N$ in all simulations.

### A. Impact of Common Sparsity Ratio

The common sparsity ratio is defined as $\rho_c = \min_k q_k^T$. When $\rho_c = 1$, all users share the same common AoD set $\Omega_S$. Therefore, $\rho_c$ reflects the correlation among the AoD sets of different users. In Fig. 10, we compare the MSE performance versus $\rho_c$ for low-frequency and high-frequency massive MIMO systems, respectively. It can be seen that as the common sparsity ratio $\rho_c$ increases, and the MSEs of Turbo-OAMP, J-OMP and J-BurstLASSO decrease. This shows that except for the OAMP, other schemes can exploit the joint sparsity among users to enhance the channel estimation performance. However, the proposed Turbo-OAMP can achieve a better performance for any given $\rho_c$, which shows that it can exploit the joint sparsity more efficiently than the baselines. Comparing Fig. 10b and Fig. 10c,

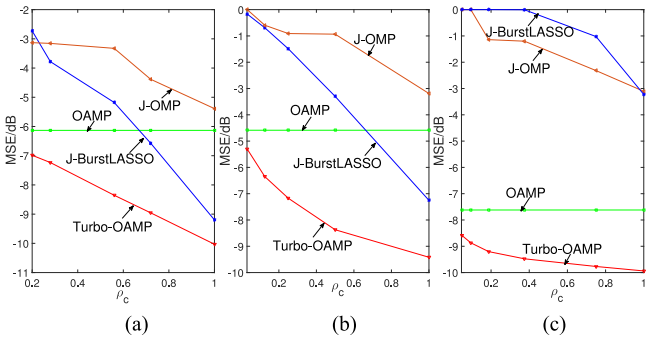



Fig. 10. MSE performance of the Turbo-OAMP versus the common sparsity ratio $\rho_c$. Set SNR= 10 dB, $K = 20$. (a) $N = 1$, $M = 128$, $T_p = 36$, $N_b = 1$ under SCM. (b) $N = 2$, $M = 256$, $T_p = 32$, $N_b = 1$ under mm-SSCM. (c) $N = 8$, $M = 256$, $T_p = 160$, $N_b = 2$ under mm-SSCM.

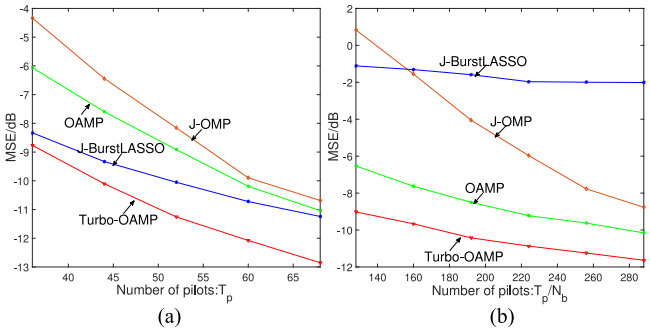



Fig. 11. MSE performance of the Turbo-OAMP versus the pilot number $T_p/N_b$. Set SNR= 10 dB, $K = 5$ and $\rho_c = 0.75$. (a) $N = 1$, $M = 128$ and $N_b = 1$ under SCM. (b) $N = 8$, $M = 256$ and $N_b = 2$ under mm-SSCM.

we see that as the Rx antenna number decreases, the performance gain of exploiting the common sparsity among users improves. For example, for the Turbo-OAMP with $\rho_c = 0.75$, the MSE gain over the OAMP is about 2 dB when $N = 8$ and $M = 256$. If $N$ decreases to 2, the MSE gain is improved to 4.5 dB.

### *B. Impact of Number of Pilots*

The impact of pilot number $T_p/N_b$ is shown in Fig. 11. It can be observed that the MSE performance improves with the number of pilots for all schemes. The proposed Turbo-OAMP achieves a large performance gain over baselines, especially when the pilot number is smaller. This demonstrates the advantage of fully exploiting both the burst sparsity of individual channels and the joint burst sparsity structure among multi-user channels.

### *C. Impact of SNR and Number of Rx RF Chains*

In Fig. 12, we compare the MSE performance versus SNR. When $N_b$ increases from 1 to 2, the performance improves for all schemes. It can be seen that the proposed Turbo-OAMP achieves a significant gain over all baseline algorithms. This demonstrates that the proposed algorithm is a powerful method for accurate channel estimation in practical massive MIMO systems when

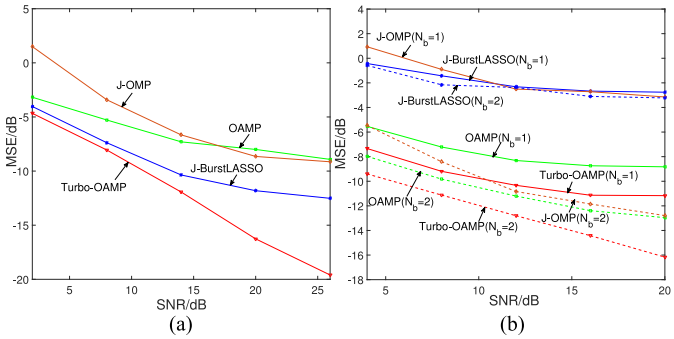



Fig. 12. MSE performance of the Turbo-OAMP versus the SNR. Set $K = 5$ and $\rho_c = 0.8$. (a) $N = 1$, $M = 128$, $T_P = 36$ and $N_b = 1$ under SCM. (b) $N = 8$, $M = 256$, $T_P/N_b = 160$, $N_b = 1$ and $N_b = 2$ under mm-SSCM.

the number of Rx RF chains at the user side is very limited in order to maintain a low cost user device.

## VIII. Conclusion

We consider the downlink channel estimation problem for general multi-user massive MIMO systems with (possibly) massive MIMO arrays and limited RF chains at both the BS and users. We first propose a statistical channel model called the HMM, in which the channel support with structured sparsity is represented as a hidden Markov process. Then we generalize the OAMP algorithm for recovering sparse signals with i.i.d. priors [11] to derive a Turbo-OAMP algorithm that can be used to recover sparse massive MIMO channels with HMM priors. In addition, we analyze the lower bound of the MSE that can be achieved by the Turbo-OAMP via the SE of a gene-aided algorithm. Simulations verify that the proposed Turbo-OAMP derived from the HMM channel can better exploit the structured sparsity of more realistic massive MIMO channels to achieve significant gain over various baseline algorithms.

For clarify, we focus on frequency-flat fading channels in this paper. In frequency-selective fading channels, there is also structured sparsity in the delay/frequency domain [27]. An interesting future work is to propose a proper probability model to capture the joint structured sparsity in the spatial/delay/frequency domain. The proposed Turbo-OAMP framework can also be extended to handle more practical frequency-selective fading channels, by modifying the MMSE estimator (Module B) according to the new probability model associated with the joint structured sparsity.

## Appendix

### *A. Message Update Equations for Module B*

*1) Message Passing Over the Path $x_{k,n,m} \to f_{k,n,m} \to \alpha_{k,n,m} \to \eta_{k,n,m} \to r_{k,n}$:* The message from variable node $x_{k,n,m}$ to factor node $f_{k,n,m}$ is

$$\nu_{x_{k,n,m}\to f_{k,n,m}}(x_{k,n,m}) = \mathcal{CN}(x_{k,n,m}; x^{pri}_{B,k,n,m}, v^{pri}_{B,k}), \tag{55}$$

and the message from $f_{k,n,m}$ to variable node $\alpha_{k,n,m}$ is

$$\begin{aligned}
&\nu_{f_{k,n,m}\to\alpha_{k,n,m}}(\alpha_{k,n,m})\\
&\propto \int f_{k,n,m}(x_{k,n,m},\alpha_{k,n,m})\,\nu_{x_{k,n,m}\to f_{k,n,m}}(x_{k,n,m})\,dx_{k,n,m}\\
&= \int (1-\alpha_{k,n,m})\,\delta(x_{k,n,m}) + \alpha_{k,n,m}\mathcal{CN}\big(x_{k,n,m};0,\sigma^2_{k,n,m}\big)\\
&\quad\times \mathcal{CN}\left(x_{k,n,m};x^{pri}_{B,k,n,m},v^{pri}_{B,k}\right)dx_{k,n,m}\\
&= (1-\alpha_{k,n,m})\,\mathcal{CN}\left(0;x^{pri}_{B,k,n,m},v^{pri}_{B,k}\right)\\
&\quad+\alpha_{k,n,m}\mathcal{CN}\left(0;x^{pri}_{B,k,n,m},v^{pri}_{B,k}+\sigma^2_{k,n,m}\right)\\
&\propto \pi^{in}_{\alpha,k,n,m}\delta(\alpha_{k,n,m}-1)+\left(1-\pi^{in}_{\alpha,k,n,m}\right)\delta(\alpha_{k,n,m}),
\end{aligned}\tag{56}$$

where $\pi^{in}_{\alpha,k,n,m}=(1+\frac{\mathcal{CN}(0;x^{pri}_{B,k,n,m},v^{pri}_{B,k})}{\mathcal{CN}(0;x^{pri}_{B,k,n,m},v^{pri}_{B,k}+\sigma^2_{k,n,m})})^{-1}$. The message from variable node $\alpha_{k,n,m}$ to factor node $\eta_{k,n,m}$ is the same as $\nu_{f_{k,n,m}\to\alpha_{k,n,m}}(\alpha_{k,n,m})$. The message from factor node $\eta_{k,n,m}$ to variable node $r_{k,n}$ is

$$\begin{aligned}
&\nu_{\eta_{k,n,m}\to r_{k,n}}(r_{k,n})\\
&\propto \sum_{\alpha,t}\eta_{k,n,m}(\alpha,t,r_{k,n})\,\nu_{t_{k,m}\to\eta_{k,n,m}}(t)\,\nu_{\alpha_{k,n,m}\to\eta_{k,n,m}}(\alpha)\\
&=\pi^{in}_{R,k,n,m}\delta(r_{k,n}-1)+(1-\pi^{in}_{R,k,n,m})\delta(r_{k,n}),
\end{aligned}\tag{57}$$

where

$$\pi^{in}_{R,k,n,m}=\frac{\pi_R}{\pi_R+1-\pi^{in}_{\alpha,k,n,m}},\tag{58}$$

with

$$\begin{aligned}
\pi_R=\gamma^k_\alpha\pi^{in}_{\alpha,k,n,m}\pi^{out}_{T,k,n,m}+\left(1-\gamma^k_\alpha\right)\left(1-\pi^{in}_{\alpha,k,n,m}\right)\pi^{out}_{T,k,n,m}\\
+\left(1-\pi^{in}_{\alpha,k,n,m}\right)\left(1-\pi^{out}_{T,k,n,m}\right)
\end{aligned}$$

and $\pi^{out}_{T,k,n,m}=\nu_{t_{k,m}\to\eta_{k,n,m}}(t_{k,m}=1)$.

*2) Message Passing Over the Markov Chains* $\mathbf{r}_k$*:* The forward-backward message passing is performed over the Markov chains $\mathbf{r}_k,\forall k$ for given input messages $\{\nu_{\eta_{k,n,m}\to r_{k,n}}\}$. The forward and backward messages are

$$\nu_{h_{R,k,n}\to r_{k,n}}\propto\gamma^f_{k,n}r_{k,n}+(1-\gamma^f_{k,n})(1-r_{k,n}),\tag{59}$$

$$\nu_{h_{R,k,n+1}\to r_{k,n}}\propto\gamma^b_{k,n}r_{k,n}+(1-\gamma^b_{k,n})(1-r_{k,n}),\tag{60}$$

where

$$\gamma^f_{k,n}=\frac{\rho^k_{0,1}(1-\pi^{in}_{n-1})(1-\gamma^f_{k,n-1})+\rho^k_{1,1}\pi^{in}_{n-1}\gamma^f_{k,n-1}}{(1-\pi^{in}_{n-1})(1-\gamma^f_{k,n-1})+\pi^{in}_{n-1}\gamma^f_{k,n-1}},\tag{61}$$

$$\gamma^b_{k,n}=\frac{\rho^k_{1,0}\left((\pi^{in}_{n+1})^{-1}-1\right)\left((\gamma^b_{k,n+1})^{-1}-1\right)+1-\rho^k_{1,0}}{(\rho^k_{0,0}+\rho^k_{1,0})\left((\pi^{in}_{n+1})^{-1}-1\right)\left((\gamma^b_{k,n+1})^{-1}-1\right)+\rho^k_{1,1}+\rho^k_{0,1}}\tag{62}$$

with $\pi^{in}_n=\frac{\prod_m\pi^{in}_{R,k,n,m}}{\prod_m\pi^{in}_{R,k,n,m}+\prod_m(1-\pi^{in}_{R,k,n,m})}$ and $\gamma^f_{k,1}=\frac{\rho^k_{0,1}}{\rho^k_{0,1}+\rho^k_{1,0}}$, $\gamma^b_{k,N}=\frac{1}{2}$.

*3) Message Passing Over the Path* $r_{k,n}\to\eta_{k,n,m}\to t_{k,m}$*:* The message from $r_{k,n}$ to factor node $\eta_{k,n,m}$ is

$$\begin{aligned}
&\nu_{r_{k,n}\to\eta_{k,n,m}}(r_{k,n})\\
&\propto \nu_{h_{R,k,n}\to r_{k,n}}\nu_{h_{R,k,n+1}\to r_{k,n}}\prod_{m'\neq m}\nu_{\eta_{k,n,m'}\to r_{k,n}}\\
&=\pi^{out}_{R,k,n,m}\delta(r_{k,n}-1)+(1-\pi^{out}_{R,k,n,m})\delta(r_{k,n}),
\end{aligned}\tag{63}$$

where $\pi^{out}_{R,k,n,m}=\frac{\pi^{out}_{R,k,n}}{\pi^{out}_{R,k,n}+(1-\pi^{out}_{R,k,n})\prod_{m'\neq m}((\pi^{in}_{R,k,n,m'})^{-1}-1)}$, $\pi^{out}_{R,k,n}=\frac{\gamma^f_{k,n}\gamma^b_{k,n}}{(1-\gamma^f_{k,n})(1-\gamma^b_{k,n})+\gamma^f_{k,n}\gamma^b_{k,n}}$. The message from factor node $\eta_{k,n,m}$ to variable node $t_{k,m}$ is

$$\begin{aligned}
&\nu_{\eta_{k,n,m}\to t_{k,m}}(t_{k,m})\\
&\propto\sum_{r,\alpha}\eta_{k,n,m}(\alpha,r,t_{k,m})\,\nu_{r_{k,n}\to\eta_{k,n,m}}(r)\,\nu_{\alpha_{k,n,m}\to\eta_{k,n,m}}(\alpha)\\
&=\pi^{in}_{T,k,n,m}\delta(t_{k,m}-1)+(1-\pi^{in}_{T,k,n,m})\delta(t_{k,m}),
\end{aligned}\tag{64}$$

where $\pi^{in}_{T,k,n,m}$ is given in equation (65), shown at the bottom of this page.

*4) Message Passing Over the HMM of* $\mathbf{t}$*:* The forward-backward message passing is performed over the HMM $\mathbf{t}$ for given input messages $\{\nu_{\eta_{k,n,m}\to t_{k,m}}\}$. The message from factor node $h_{T,k,m}$ to variable node $s_m$ is

$$\begin{aligned}
&\nu_{h_{T,k,m}\to s_m}(s_m)\\
&\propto\sum_{t_{k,m}}\prod_n\nu_{\eta_{k,n,m}\to t_{k,m}}(t_{k,m})\,h_{T,k,m}(t_{k,m},s_m)\\
&=\pi^{in}_{S,k,m}\delta(s_m-1)+(1-\pi^{in}_{S,k,m})\delta(s_m),
\end{aligned}\tag{66}$$

where $\pi^{in}_{S,k,m}=\frac{q^T_k\pi^{in}_{k,m}+(1-q^T_k)(1-\pi^{in}_{k,m})}{q^T_k\pi^{in}_{k,m}+(1-q^T_k)(1-\pi^{in}_{k,m})+(1-\pi^{in}_{k,m})}$, $\pi^{in}_{k,m}=\frac{\prod_n\pi^{in}_{T,k,n,m}}{\prod_n\pi^{in}_{T,k,n,m}+\prod_n(1-\pi^{in}_{T,k,n,m})}$. The forward and backward

$$\pi^{in}_{T,k,n,m}=\frac{\gamma^k_\alpha\pi^{in}_{\alpha,k,n,m}\pi^{out}_{R,k,n,m}+\left(1-\gamma^k_\alpha\right)\left(1-\pi^{in}_{\alpha,k,n,m}\right)\pi^{out}_{R,k,n,m}+\left(1-\pi^{in}_{\alpha,k,n,m}\right)\left(1-\pi^{out}_{R,k,n,m}\right)}{\gamma^k_\alpha\pi^{in}_{\alpha,k,n,m}\pi^{out}_{R,k,n,m}+\left(1-\gamma^k_\alpha\right)\left(1-\pi^{in}_{\alpha,k,n,m}\right)\pi^{out}_{R,k,n,m}+\left(1-\pi^{in}_{\alpha,k,n,m}\right)\left(1-\pi^{out}_{R,k,n,m}\right)+1-\pi^{in}_{\alpha,k,n,m}}.\tag{65}$$

$$\gamma_{S,m}^{b} = \frac{\rho_{1,0}^{S}\left(1-\pi_{S,m+1}^{in}\right)\left(1-\gamma_{S,m+1}^{b}\right)+\left(1-\rho_{1,0}^{S}\right)\pi_{S,m+1}^{in}\gamma_{S,m+1}^{b}}{\left(\rho_{0,0}^{S}+\rho_{1,0}^{S}\right)\left(1-\pi_{S,m+1}^{in}\right)\left(1-\gamma_{S,m+1}^{b}\right)+\left(\rho_{1,1}^{S}+\rho_{0,1}^{S}\right)\pi_{S,m+1}^{in}\gamma_{S,m+1}^{b}} \tag{70}$$

messages over the Markov chain $\mathbf{s}$ are given by

$$\nu_{h_{s,m}\to s_m} \propto \gamma_{S,m}^{f} s_m + (1-\gamma_{S,m}^{f})(1-s_m), \tag{67}$$

and

$$\nu_{h_{s,m+1}\to s_m} \propto \gamma_{S,m}^{b} s_m + (1-\gamma_{S,m}^{b})(1-s_m), \tag{68}$$

where

$$\begin{aligned}&\gamma_{S,m}^{f}\\ &= \frac{\rho_{0,1}^{S}(1-\pi_{S,m-1}^{in})(1-\gamma_{S,m-1}^{f})+\rho_{1,1}^{S}\pi_{S,m-1}^{in}\gamma_{S,m-1}^{f}}{(1-\pi_{S,m-1}^{in})(1-\gamma_{S,m-1}^{f})+\pi_{S,m-1}^{in}\gamma_{S,m-1}^{f}},\end{aligned} \tag{69}$$

and $\gamma_{S,m}^{b}$ are given by equation (70) shown at the top of this page, with $\pi_{S,m}^{in} = \frac{\prod_k \pi_{S,k,m}^{in}}{\prod_k \pi_{S,k,m}^{in}+\prod_k(1-\pi_{S,k,m}^{in})}$ and $\gamma_{S,1}^{f} = \frac{\rho_{0,1}^{S}}{\rho_{0,1}^{S}+\rho_{1,0}^{S}}$, $\gamma_{S,M}^{b} = \frac{1}{2}$. Then the message from the variable node $s_m$ to factor node $h_{T,k,m}$ is

$$\begin{aligned}&\nu_{s_m\to h_{T,k,m}}(s_m)\\ &\propto \nu_{h_{S,m}\to s_m}(s_m)\,\nu_{h_{S,m+1}\to s_m}(s_m)\\ &= \pi_{S,k,m}^{out}\delta(s_m-1)+(1-\pi_{S,k,m}^{out})\delta(s_m),\end{aligned} \tag{71}$$

with $\pi_{S,k,m}^{out} = \frac{\gamma_{S,m}^{f}\gamma_{S,m}^{b}}{\gamma_{S,m}^{f}\gamma_{S,m}^{b}+\prod_{k'\neq k}((\pi_{S,k',m}^{in})^{-1}-1)(1-\gamma_{S,m}^{f})(1-\gamma_{S,m}^{b})}$. The message from factor node $h_{T,k,m}$ to variable node $t_{k,m}$ is

$$\begin{aligned}&\nu_{h_{T,k,m}\to t_{k,m}}(t_{k,m})\\ &\propto \sum_{s_m}\nu_{s_m\to h_{T,k,m}}(s_m)\,h_{T,k,m}(t_{k,m},s_m)\\ &= \pi_{T,k,m}^{out}\delta(t_{k,m}-1)+(1-\pi_{T,k,m}^{out})\delta(t_{k,m}),\end{aligned} \tag{72}$$

where $\pi_{T,k,m}^{out} = \pi_{S,k,m}^{out} q_k^{T}$.

*5) Message Passing Over the Path $t_{k,m}\to\eta_{k,n,m}\to\alpha_{k,n,m}\to f_{k,n,m}\to x_{k,n,m}$:* According to the sum-product rule, the message from variable node $t_{k,m}$ to factor node $\eta_{k,n,m}$ is

$$\begin{aligned}&\nu_{t_{k,m}\to\eta_{k,n,m}}(t_{k,m})\\ &\propto \nu_{h_{T,k,m}\to t_{k,m}}\prod_{n'\neq n}\nu_{\eta_{k,n',m}\to t_{k,m}}\\ &= \pi_{T,k,n,m}^{out}\delta(t_{k,m}-1)+(1-\pi_{T,k,n,m}^{out})\delta(t_{k,m}),\end{aligned} \tag{73}$$

where $\pi_{T,k,n,m}^{out} = \frac{\pi_{T,k,m}^{out}}{\pi_{T,k,m}^{out}+(1-\pi_{T,k,m}^{out})\prod_{n'\neq n}((\pi_{T,k,n',m}^{in})^{-1}-1)}$. The message from factor node $\eta_{k,n,m}$ to variable node $\alpha_{k,n,m}$ is

$$\begin{aligned}&\nu_{\eta_{k,n,m}\to\alpha_{k,n,m}}(\alpha_{k,n,m})\\ &\propto \sum_{t,r}\eta_{k,n,m}(\alpha_{k,n,m},r,t)\,\nu_{t_{k,m}\to\eta_{k,n,m}}(t)\,\nu_{r_{k,n}\to\eta_{k,n,m}}(r)\\ &= \pi_{\alpha,k,n,m}^{out}\delta(\alpha_{k,n,m}-1)+\left(1-\pi_{\alpha,k,n,m}^{out}\right)\delta(\alpha_{k,n,m}),\end{aligned} \tag{74}$$

where $\pi_{\alpha,k,n,m}^{out} = \pi_{T,k,n,m}^{out}\pi_{R,k,n,m}^{out}\gamma_{\alpha}^{k}$. And the message from variable node $\alpha_{k,n,m}$ to factor node $f_{k,n,m}$ is the same as $\nu_{\eta_{k,n,m}\to\alpha_{k,n,m}}$. The message from factor node $f_{k,n,m}$ to variable node $x_{k,n,m}$ is

$$\begin{aligned}&\nu_{f_{k,n,m}\to x_{k,n,m}}(x_{k,n,m})\\ &\propto \int f_{k,n,m}(x_{k,n,m},\alpha_{k,n,m})\,\nu_{\alpha_{k,n,m}\to f_{k,n,m}}(\alpha_{k,n,m})\,d\alpha_{k,n,m}\\ &= \pi_{\alpha,k,n,m}^{out}g\left(x_{k,n,m};0,\sigma_{k,n,m}^{2}\right)+\left(1-\pi_{\alpha,k,n,m}^{out}\right)\delta(x_{k,n,m}).\end{aligned} \tag{75}$$

## REFERENCES


[1] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[2] X. Rao and V. Lau, "Compressive sensing with prior support quality information and application to massive MIMO channel estimation with temporal correlation," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4914–4924, Sep. 2015.

[3] M. Masood, L. Afify, and T. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 104–118, Jan. 2015.

[4] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[5] X. Rao and V. K. Lau, "Distributed compressive CSIT estimation and feedback for FDD multiuser massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[6] A. Liu, V. K. N. Lau, and W. Dai, "Exploiting burst-sparsity in massive MIMO with partial channel support information," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7820–7830, Nov. 2016.

[7] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 601–617, Feb. 2016.

[8] A. Liu, V. Lau, and W. Dai, "Joint burst LASSO for sparse channel estimation in multiuser massive MIMO," in *Proc. IEEE Int. Conf. Commun.*, May 2016, pp. 1–6.

[9] M. Opper and O. Winther, "Expectation consistent approximate inference," *J. Mach. Learn. Res.*, vol. 6, pp. 2177–2204, 2005.

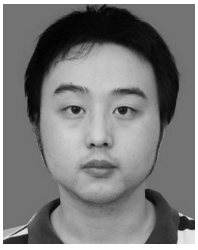
[10] J. Ma, X. Yuan, and L. Ping, "On the performance of turbo signal recovery with partial DFT sensing matrices," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1580–1584, Oct. 2015.

[11] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[12] L. Chen, A. Liu, and X. Yuan, "Structured turbo compressed sensing for massive MIMO channel estimation using a Markov prior," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4635–4639, May 2018.

[13] X. Zhang, A. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.

[14] A. Alkhateeb, O. El Ayach, G. Leusz, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
[15] A. Alkhateeb, G. Leusz, and R. W. Heath, "Compressed sensing based multiuser millimeter wave systems: How many measurements are needed?" in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2015, pp. 2909–2913.
[16] C. B. Dietrich Jr., "Adaptive arrays and diversity antenna configurations for handheld wireless communication terminals. Chapter 3: Antenna arrays and beamforming," Ph.D. dissertation, Dept. Elect. Comput. Eng., Virginia Tech, Blacksburg, VA, USA, 2005.
[17] J. Ma, X. Yuan, and L. Ping, "Turbo compressed sensing with partial DFT sensing matrix," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 158–161, Feb. 2015.
[18] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, pp. 18914–18919, 2009.
[19] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.
[20] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2207–2225, Jul. 2016.
[21] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.
[22] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," in *Proc. IEEE Int. Symp. Inf. Theory*, 2017, pp. 501–505.
[23] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2017, pp. 1588–1592.
[24] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statistical Soc. Ser. B (Methodological)*, vol. 39, pp. 1–38, 1977.
[25] J. Vila and P. Schniter, "Expectation-maximization Bernoulli–Gaussian approximate message passing," in *Proc. Conf. Record 45th Asilomar Conf. Signals, Syst. Comput.*, 2011, pp. 799–803.
[26] J. Salo *et al.*, "MATLAB implementation of the 3GPP spatial channel model (3GPP TR 25.996)," Jan. 2005. [Online]. Available: http://www.tkk.fi/Units/Radio/scm/
[27] Z. Gao, L. Dai, S. Han, C.-L. I, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, Jun. 2018.



**An Liu** (S'07–M'09–SM'17) received the Ph.D. and B.S. degrees in electrical engineering from Peking University, Beijing, China, in 2011 and 2004, respectively. From 2008 to 2010, he was a Visiting Scholar with the Department of ECEE, University of Colorado at Boulder. He was a Postdoctoral Research Fellow from 2011–2013, a Visiting Assistant Professor in 2014, and a Research Assistant Professor from 2015–2017, with the Department of ECE, Hong Kong University of Science and Technology. He is currently a Distinguished Research Fellow with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. His research interests include wireless communications, stochastic optimization, and compressive sensing.

**Lixiang Lian** (S'16) received the B.Eng. degree in information and communication engineering from Zhejiang University, Hangzhou, China, in 2014. She is currently working toward the Ph.D. degree with the Department of ECE, Hong Kong University of Science and Technology, Hong Kong. Her research interests include wireless communication and compressive sensing.

**Vincent K. N. Lau** (SM'04–F'12) received the B.Eng. degree (Distinction 1st Hons.) from the University of Hong Kong, Hong Kong, in 1992, and the Ph.D. degree from the Cambridge University, Cambridge, U.K., in 1997. He joined Bell Labs from 1997–2004 and the Department of ECE, Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2004. He is currently a Chair Professor and the Founding Director of Huawei-HKUST Joint Innovation Laboratory, HKUST. His current research focus includes robust and delay-optimal cross layer optimization for MIMO/OFDM wireless systems, interference mitigation techniques for wireless networks, massive MIMO, M2M, and network control systems.

**Xiaojun Yuan** (S'04–M'09–SM'15) received the Ph.D. degree in electrical engineering from the City University of Hong Kong, Hong Kong, in 2008.

From 2009 to 2011, he was a Research Fellow with the Department of Electronic Engineering, the City University of Hong Kong. He was also a Visiting Scholar with the Department of Electrical Engineering, the University of Hawaii at Manoa, in spring and summer 2009, as well as in the same period of 2010. From 2011 to 2014, he was a Research Assistant Professor with the Institute of Network Coding, The Chinese University of Hong Kong. From 2014 to 2017, he was an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University. He is now a Professor with the National Key Laboratory of Science and Technology on Communications, the University of Electronic Science and Technology of China, Chengdu, China, supported by the Thousand Youth Talents Plan in China. He has authored or co-authored more than 110 peer-reviewed research papers in the leading international journals and conferences in the related areas. His research interests cover a broad range of signal processing, machine learning, and information theory, including but not limited to multiantenna and cooperative communications, sparse and structured signal recovery, Bayesian approximate inference, network coding, etc. He was on a number of technical programs for international conferences. He has been an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS since 2017. He was a co-recipient of the Best Paper Award of IEEE International Conference on Communications 2014, and also a co-recipient of the Best Journal Paper Award of IEEE Technical Committee on Green Communications and Computing 2017.