Bayesian Beamforming for Mobile Millimeter Wave Channel Tracking in the Presence of DOA Uncertainty

Yan Yang, *Member, IEEE*, Shuping Dang, *Member, IEEE*, Miaowen Wen, *Senior Member, IEEE*, Shahid Mumtaz, *Senior Member, IEEE*, and Mohsen Guizani, *Fellow, IEEE*

***Abstract*— This paper proposes a Bayesian approach for angle-based hybrid beamforming and tracking that is robust to uncertain or erroneous direction-of-arrival (DOA) estimation in millimeter wave (mmWave) multiple input multiple output (MIMO) systems. Because the resolution of the phase shifters is finite and typically adjustable through a digital control, the DOA can be modeled as a discrete random variable with a prior distribution defined over a discrete set of candidate DOAs, and the variance of this distribution can be introduced to describe the level of uncertainty. The estimation problem of DOA is thereby formulated as a weighted sum of previously observed DOA values, where the weights are chosen according to *a posteriori* probability density function (pdf) of the DOA. To alleviate the computational complexity and cost, we present a motion trajectory-constrained *a priori* probability approximation method. It suggests that within a specific spatial region, a directional estimate can be close to true DOA with a high probability and sufficient to ensure trustworthiness. We show that the proposed approach has the advantage of robustness to uncertain DOA, and the beam tracking problem can be solved by incorporating the Bayesian approach with an expectation-maximization (EM) algorithm. Simulation results validate the theoretical analysis and demonstrate that the proposed solution outperforms a number of state-of-the-art benchmarks.**

Manuscript received August 10, 2019; revised December 29, 2019, March 15, 2020, July 27, 2020, and September 16, 2020; accepted September 17, 2020. Date of publication September 24, 2020; date of current version December 16, 2020. This work was in part supported by the State Key Laboratory of Rail Traffic Control and Safety (Contract No. RCS2020ZT012), Beijing Jiaotong University and China Railway Corporation (Contract No. N2019G028). This article was presented in part at the 2019 IEEE GLOBECOM'19. The associate editor coordinating the review of this article and approving it for publication was O. Oyman. *(Corresponding author: Yan Yang.)*

Yan Yang is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China (e-mail: yyang@bjtu.edu.cn).

Shuping Dang is with the Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia (e-mail: shuping.dang@kaust.edu.sa).

Miaowen Wen is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: eemwwen@scut.edu.cn).

Shahid Mumtaz is with the Instituto de Telecomunicações, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal, and also with the ARIES Research Center, Universidad Antonio de, E-28040 Madrid, Spain (e-mail: smumtaz@av.it.pt).

Mohsen Guizani is with the Department of Computer Science and Engineering, Qatar University, Doha 2713, Qatar (e-mail: mguizani@ieee.org).

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

Digital Object Identifier 10.1109/TCOMM.2020.3026377

***Index Terms*— Bayesian beamforming, direction of arrival (DOA) estimation, millimeter wave, channel tracking, expectation-maximization (EM) algorithm.**

I. Introduction

MILLIMETER wave (mmWave) systems offer a promising solution, through which high data rates (Giga-bit) can be achieved in next-generation mobile communication networks [1]–[5]. To significantly alleviate the destructive effects of path loss of mmWave links, highly directional beamforming at both the transmitter and receiver are required. Meanwhile, the shorter wavelength associated with the higher frequency also enables massive antenna arrays to contain more elements within the same physical dimension, and large-scale antenna arrays, as well as various multiple input multiple output (MIMO) configurations, provide a high antenna array gain and spectral efficiency for mmWave communications [4]. Mobile mmWave communication systems will likely operate over wideband channels with frequency selectivity. Nevertheless, orthogonal frequency-division multiplexing (OFDM) modulation can be employed to effectively combat the channel's frequency selectivity. It divides the wideband signal into many slowly modulated narrowband subcarriers, and the subchannel can be approximated as flat fading within each subcarrier's passband. It is worth noting that in the mmWave bands such as above 28 GHz, the main lobe created by directional beamforming has a narrower beamwidth, and high-gain narrow-beam, boosting the strength of certain paths, is essential to mitigate the high path loss. This is very different from the bands below 3 GHz and/or the bands between 3 GHz and 6 GHz (Sub-6) with a beam of broadening, which is prone to providing more flexible beam tracking. However, narrow mmWave beams must be precisely aligned to afford a feasible transmission link margin. To maintain high-quality transmission links, an efficient approach for beam training and tracking is crucial to determine suitable directions of transmission and reception [6], [7].

To mitigate severe propagation loss in the mmWave band, many recent studies have demonstrated that the hybrid analog/digital beamformer is a cost-effective alternative. Typically, mmWave communication systems that incorporate a hybrid beamforming architecture, is constructed by concatenation of a digital baseband (BB) precoder and an analog beamformer.

The baseband digital part of the hybrid architecture is designed for modifying the data streams, multiplexing as well as reducing the inter-stream interference, and the RF analog part is designed for the spatial focusing gain to compensate/cover up for the higher propagation losses. As a result, it is normally able to achieve significant beamforming gains and synthesize a highly directional beam, as well as to boost spectral efficiency and transmission rates [8]–[15]. Since the emergence of large-scale antenna arrays and analog beamforming, methods for designing a hybrid precoder and combiner have been a research focus with problems in the physical layer or with signal processing aspects, as reported in [3], [15].

In [3], channel estimation and hybrid analog/digital precoding algorithms were proposed to enable multi-stream multiplexing in mmWave systems, with which the authors designed an adaptive compressed sensing (CS)-based channel estimation algorithm as well as a multi-resolution codebook for the training precoder. It is also shown in [15]–[19] that the design of a hierarchical multi-resolution codebook is critical when seeking to improve data rate, including tracking the direction of a mmWave path. Some studies focus on the use of low-complexity hybrid beamforming schemes to achieve the optimal channel capacity, assuming that perfect channel knowledge is available at the receiver. However, such an idealistic assumption rarely holds in practice [4]. In particular, for very high bandwidth channels, high-precision analog-to-digital converters (ADCs) are too costly and power-hungry, whereas low-resolution (e.g., 1-bit) ADCs offer a realistic solution due to its low power consumption and low hardware complexity. However, because of the nonlinearity of quantization, channel estimation with these low-resolution ADCs becomes challenging [20].

Most of the hybrid beamforming schemes require full channel state information (CSI), which includes the estimation of the directions of dominant propagation paths. These directions can be further used to identify important paths for transmission and reception [4], [14], [21]. Due to the spatial sparsity and high directional property of mmWave channels [22], [23], a typical mmWave channel is dominated by a few paths. Implementation of large-scale antenna arrays will lead to a much better resolvability of multiple paths [10], [23]. Accordingly, the problem of mmWave channel estimation and beamforming can be naturally formulated by the angular and delay representation of the channel and can thus be simplified as a straightforward approach to estimate the dominant direction of arrival (DOA) and the corresponding path gains of larger power [10], [11], [20], [24]–[26]. In these studies, the spatial channel is generally decomposed according to the DOA information and the corresponding complex path coefficients. As a result,[1] the angles-of-arrival (AOAs) $\boldsymbol{\theta}$ and angles-of-departure (AODs) $\boldsymbol{\phi}$ can be estimated using a traditional beam training procedure [3], [25].

[1]Strictly speaking, DOA and AOA are the same measurement, which can be used for various purposes. The Tx-Rx beam pairs are generally specified as AOD and AOA of the desired path in the angular space, while DOA can be a better choice to emphasize the direction of the incident wave with respect to the receiver.

In the mobile mmWave channel, as a result of user mobility—especially in high-mobility applications, such as in connected vehicles, uncertain DOAs are directly reflected into the received signal. Accordingly, when using mmWave systems for highly directional communications, even a slight beam misalignment between two communicating devices can lead to a significant signal drop [7], [8], [11], [12]. To address this issue, many Bayesian frameworks for adaptive and blind beamforming algorithms have been developed to combat the DOA uncertainty, where the DOA is modeled as a random variable with a prior distribution defined over a candidate set that describes the level of uncertainty, in accordance with a quantitative or parametric description of uncertainty [16], [17], [26]–[32]. Most recently, Tiwari *et al.* introduced the concept of beam entropy that is used to quantify sparse MIMO channel randomness in beamspace, and investigated the quantitative description of channel randomness as well as the typical values of beam entropy [33]. Classical direction learning techniques, such as MUSIC [34] and ESPRIT [35], diagonal loading, linearly constrained minimum variance (LCMV), and covariance matrix taper methods have been used to enhance the DOA robustness [28], [30]. When using directional beamforming in the context of mobile communications, an efficient channel tracking scheme for mmWave systems relies on DOA accuracy and steering vector estimation to identify the direction of the dominant paths and capture the angle variations [36]. It is noticeable that there has been also considerable research interest in static mmWave scenarios, where beamforming also plays a key role in order to enable long-time sustained high-capacity links [37]–[39].

In this paper, we investigate the robust beamforming for mobile mmWave channel tracking in a downlink MIMO-OFDM system with low-resolution ADCs. We propose a Bayesian-based parameter estimation method to mitigate the performance degradation subject to the DOA uncertainty. Since the resolution of the phase shifter is limited, the phase shifts corresponding to DOA actually change the steering direction by a discrete number of steps [8], [16]. Taking this factor into consideration, the DOA parameter at each time instant can be approximately modeled as a discrete random variable. With our proposed Bayesian framework, a statistical inference approach is formulated to address the uncertainty of DOA estimation. The main contributions of this paper are summarized as follows:

- We formulate the Bayesian framework to provide a statistical inference method that mitigates the DOA mismatch problem. The indeterminate DOA is modeled as a set of random variables with a known *a priori* probability density function (pdf). Accordingly, the final DOA estimate can be determined by a weighted sum of DOA values, where the weights are obtained by means of a posteriori estimator, that is, a maximum *a posteriori* probability (MAP) estimator. Compared to the methods in the existing literature, our method is robust to uncertain DOA and has a higher estimation accuracy.
- By exploiting the characteristics of the motion trajectory of a reference receiver, we propose an approximation method using *a priori* probability. The advantage of this

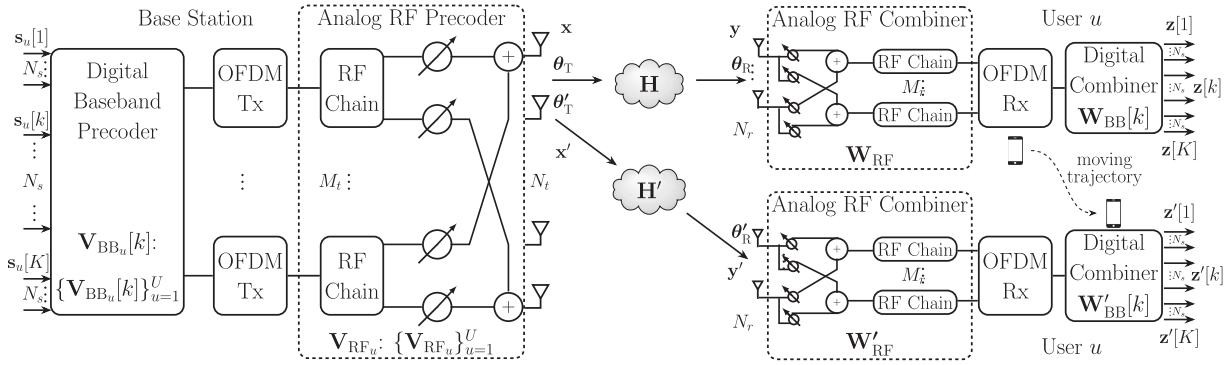


Fig. 1. A downlink mmWave MIMO-OFDM system with hybrid beamforming architecture.

method is that the calculation of *a posteriori* probability only needs to be performed within a range where the actual DOA occurrence probability is high. Therefore, it is not necessary to calculate the posterior probability of all DOA candidate combinations in the DOA parameter space. This results in a significant reduction in degree of freedom and computational complexity. Simulation results show that the proposed method demonstrates a higher parameter estimation accuracy.

- We propose and implement an effective beam tracking algorithm with limited feedback that uses a Bayesian model with the MAP estimation criterion. Considering the need for hyperparameter estimation in the Bayesian framework, we develop an iterative hyperparameter estimation approach, where *a posteriori* DOA mean and covariance can be estimated to provide high-quality CSI by incorporating the Bayesian approach with the estimation-maximization (EM) algorithm.
- From the view of implementation in the frequency domain, we establish a wideband DOA signal detection model corresponding to the beamforming processing and channel tracking. The proposed model introduced herein can be easily extended to multi-user MIMO-OFDM systems.

The rest of the paper is organized in the following. In Section II, we present the system and signal transmission models, as well as the channel model. Section III expatiates on the problem statement and discusses key technical considerations. In Section IV, we mathematically formulate the Bayesian beamformer and introduce the beam tracking method. Sections V and VI present simulation results and conclusions of this work, respectively.

Notations: Throughout this paper, small normal letters are used for scalars; lower-case boldface $\mathbf{x}$ and upper-case boldface $\mathbf{X}$ denote vectors and matrices, respectively; $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote complex conjugation, transpose, and conjugate transpose, respectively. We use $\mathbb{C}$ to denote the field of complex numbers and $\mathbb{C}^{m\times n}$ denotes an m by n dimensional complex space. Further, we use $\mathrm{Tr}(\cdot)$, $\log(\cdot)$, $\mathbb{E}(\cdot)$, and $||\cdot||$ to denote the trace, the natural logarithm, expected value of the enclosed argument, and the the vector 2-norm (or the matrix Frobenius norm), respectively; $\mathcal{CN}(0;\sigma^2)$ represents the circularly symmetric zero-mean complex Gaussian distribution with variance σ^2.

II. System Model

A. System Framework and Signal Transmission Procedure

Consider a wideband downlink mmWave MIMO-OFDM system with N_t transmit antennas and N_r receive antennas, which is constructed by a two-stage hybrid analog/digital beamforming architecture as shown in Fig. 1. We assume that both the transmitter and receiver have a uniform linear array (ULA) with identical antenna elements spacing. From the view of implementation in the frequency domain, the transmitted data symbols for the uth mobile station (MS) are organized into frequency vectors $\mathbf{s}_u[k] \in \mathbb{C}^{N_s\times 1}, k = 1,\ldots,K$, and the individual data stream is transmitted via N_s K-length data symbols with the property $\mathbb{E}[\mathbf{s}_u[k]\mathbf{s}_u[k]^H] = \mathbf{I}_{N_s}$. Within the hybrid architecture, a base station (BS) using $N_{t,u} < N_t$ antennas and $M_{t,u} < M_t$ RF chains communicates with the uth MS with N_r antennas and M_r RF chains, given $N_s \le M_{t,u} \le N_{t,u}$, $N_s \le M_r \le N_r \le N_{t,u}$ and $M_r \le M_{t,u}$ [4], [6].

As shown in Fig. 1, we use the sets of the matrices $\{\mathbf{V}_{\mathrm{BB}_u}[k]\}_{u=1}^U$ and $\{\mathbf{V}_{\mathrm{RF}_u}\}_{u=1}^U$ to denote the multiple digital precoders and analog precoders, respectively. To perform hybrid beamforming for user u, the BS first applies a digital baseband precoder $\mathbf{V}_{\mathrm{BB}_u}[k] \in \mathbb{C}^{M_{t,u}\times N_s}$ to modify the N_s data symbols $\mathbf{s}_u[k]$ over the kth subcarrier, and the baseband data is fed directly into the OFDM Tx module, which is then transformed to the time domain using $M_{t,u}$ numbers of K-point IFFT. Note that Fig. 1 shows a simplified signal chain in the frequency domain. Typically, the implementation of OFDM at Tx and Rx consists of parallel-to-serial and serial-to-parallel conversions. At Tx, a block of symbols is serial-to-parallel converted onto K subcarriers, and the receiver decodes each frequency bin separately. Additionally, the LNA, mixer, ADC, and variable gain amplifier (VGA) are also referred to as the RF chain. After that, an RF (analog) precoder $\mathbf{V}_{\mathrm{RF}_u} \in \mathbb{C}^{N_{t,u}\times M_{t,u}}$ is implemented using analog phase shifters to support multiple directional beamers in downlink for multiple users. Similarly, the hybrid beamforming architecture at the receiver is constructed by the concatenation of an RF combiner $\mathbf{W}_{\mathrm{RF}} \in \mathbb{C}^{N_r\times M_r}$ and a low-dimensional digital baseband combiner $\mathbf{W}_{\mathrm{BB}}[k] \in \mathbb{C}^{M_r\times N_s}$. Note that the OFDM Rx module is also configured in the streaming mode, and at the digital down-converter, cyclic prefix is removed from the incoming OFDM symbols.

Note that the proposed model can be easily extended to multi-user cases. Assume that a BS with N_t transmit antennas and M_t RF chains simultaneously serves U active users, and each user has been equipped with N_s data streams ($UN_s \leq M_t \leq N_t$). We have $\mathbf{F}_{\text{BB}}[k] \triangleq [\mathbf{V}_{\text{BB}_1}[k], \ldots, \mathbf{V}_{\text{BB}_U}[k]]$ and $\mathbf{F}_{\text{RF}} \triangleq [\mathbf{V}_{\text{RF}_1}, \ldots, \mathbf{V}_{\text{RF}_U}]$, where the hybrid precoders $\mathbf{F}_{\text{BB}}[k] \in \mathbb{C}^{M_t \times UN_s}$ and $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_t \times M_t}$ are represented by a block concatenation of matrices, respectively. In this paper, because we focus on a single-user beam tracking case, we omit the subscript u for simplicity of expression. Parameters $\mathbf{s}[k]$, $\mathbf{V}_{\text{BB}}[k]$ and $\mathbf{V}_{\text{RF}}$ can be directly applied to any single-user case without extra algebraic manipulation.

In the absence of noise, the final transmitted symbol $\mathbf{x}$ associated with the kth subcarrier can be represented as

$$\mathbf{x}[k] = \mathbf{V}[k]\mathbf{s}[k] = \mathbf{V}_{\text{RF}}\mathbf{V}_{\text{BB}}[k]\mathbf{s}[k], \tag{1}$$

where $\mathbf{V}_{\text{RF}}$ is used to process the digitally-precoded transmitted stream and is the same for *all* subcarriers, and $\mathbf{V}_{\text{BB}}[k]$ is designed for each subcarrier since it is applied in the frequency domain. To simplify our exposition, we first adopt a widely-used narrowband block-fading channel model in an ideal setting where a line-of-sight (LOS) path is considered so that the channel can be approximated by the single-tap model. Consequently, the received signal at subcarrier k at the antenna array of MS can be written as

$$\mathbf{y}[k] = \mathbf{H}[k]\mathbf{x}[k] + \mathbf{n}[k], \tag{2}$$

where $\mathbf{H}[k] \in \mathbb{C}^{N_r \times N_{t,u}}$ is the complex channel matrix corresponding to the kth subcarrier and $\mathbf{n}[k] \sim \mathcal{CN}(0, \sigma_k^2\mathbf{I})$ denotes the vector of independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN); σ_k^2 represents the noise variance associated with subcarrier k; $\mathbf{I} \in \mathbb{C}^{N_r \times N_r}$ is the identity matrix. Typically, the channel state matrix $\mathbf{H}$ can be estimated by sending training sequences and/or pilot signals. In stationary scenarios, such as indoor hotspots or backhaul, since the channel is typically slowly varying, an idealistic but acceptable assumption is that $\mathbf{H}$ is perfectly known at the receiver. In contrast, perfect and instantaneous estimation of CSI is a challenging task in mobile networks, due to mobility and the rendered fast fading [5], [12]. Therefore, in this paper, we mainly focus on fast varying channels. Particularly, the problem we are interested in is how to mitigate the effects of uncertain DOA in mobile mmWave systems. In the next section, we will detail the motivation and specify the research problem.

At the receiving end, the overall hybrid combiner can be written as $\mathbf{W}[k] = \mathbf{W}_{\text{RF}}\mathbf{W}_{\text{BB}}[k]$, $\mathbf{W}[k] \in \mathbb{C}^{M_r \times N_s}$, where $\mathbf{W}_{\text{RF}}$ is the RF combiner implemented by phase shifters, assuming that $||\mathbf{W}_{\text{RF}}||^2 = 1$, and $\mathbf{W}_{\text{BB}}[k]$ is the digital combiner in baseband. It means that the estimate of the desired signal vector $\overline{\mathbf{z}}[k] \in \mathbb{C}^{N_s \times 1}$ can be given as a linear combination of the receive signal $\mathbf{y}[k]$ over the current subcarrier. Thus, we mathematically have

$$\mathbf{z}[k] = \mathbf{W}^H[k]\mathbf{H}[k]\mathbf{V}[k]\mathbf{s}[k] + \mathbf{W}^H[k]\mathbf{n}[k]. \tag{3}$$

Essentially, the optimal precoder at the transmitter is the one that maximizes the transmit signal power with a given power budget Ω_{Tx}. Thus, it enables the optimization problem to be decoupled into a series of subproblems, and it is also feasible to apply the same processing method at the receiver. In our work, we adopt these general optimization methods, which decouple and simplify the optimization problem for adjusting the precoding and combining weights [4], [10]. The optimal hybrid precoders at the BS can thereby be determined by the following criterion:

$$\begin{aligned} \mathbf{V}_{\text{opt}}[k] = \arg\max\nolimits_{\mathbf{V}[k]} \ & \mathbf{V}^H[k]\mathbf{s}^H[k]\mathbf{s}[k]\mathbf{V}[k], \\ \text{s.t.} \quad & \sum_{k=1}^{K} \text{Tr}(\mathbf{V}^H[k]\mathbf{V}[k]) < \Omega_{\text{Tx}}. \end{aligned} \tag{4}$$

Similarly, we make an assumption of assigning equal power to all subcarriers as well as all spatial streams, and the average signal-to-noise ratio (SNR) at each receive antenna can be calculated to be $\rho_k/(N_{t,u}\sigma_k^2)$, where $\rho_k = ||\mathbf{W}^H[k]\mathbf{H}[k]\mathbf{V}[k]||^2$, denoting the achieved power gain of the kth subcarrier [4]. The optimal combiner at the receiver which can be extended to multiple RF chains can be expressed as [40], [41]

$$\begin{aligned} \mathbf{W}_{\text{opt}}[k] = \ & \arg\max_{\mathbf{W}[k]} \left(\frac{||\mathbf{W}^H[k]\mathbf{H}[k]\mathbf{V}[k]||^2 \cdot ||\mathbf{s}[k]||^2}{||\mathbf{W}^H[k]\mathbf{n}[k]||^2} \right), \\ \text{s.t.} \quad & \sum_{k=1}^{K} ||\mathbf{W}[k]||^2 = 1. \end{aligned} \tag{5}$$

Assume that the total available power is uniformly allocated over all space-frequency grids, and the data streams transmitted from different antennas are statistically independent. From the basics of information theory, the bound on the capacity of a single-user system is given by [42]

$$\begin{aligned} C_{\text{SU}} &= \sum_{k=1}^{K} \beta_k \log_2 \left(1 + \frac{\rho_k}{N_{t,u}\sigma_k^2} \right) \\ &= \sum_{k=1}^{K} \beta_k \log_2 \left(1 + \frac{||\mathbf{W}^H[k]\mathbf{H}[k]\mathbf{V}[k]||^2}{N_{t,u}\sigma_k^2} \right), \end{aligned} \tag{6}$$

where β_k represents the normalized weight of the rate over the kth subcarrier. Based on (4) and (5), it is clear that a decoupled design of the hybrid precoder and combiner is adopted to solve the joint optimization problem over $\mathbf{V}$ and $\mathbf{W}$. Accordingly, the feedback overhead caused by a training procedure for solving the joint optimization can be subtly circumvented. Although the remaining optimization problem pertaining to $\mathbf{V}$ and $\mathbf{W}$ is still non-convex and thereby difficult to be optimally solved, it enables further decoupling by individually designing RF and baseband. A primary approach is the joint design of digital precoder $\mathbf{V}_{\text{BB}}$ and analog precoder $\mathbf{V}_{\text{RF}}$, in which the RF combiner is first designed based on a locally optimal digital precoder, and the digital combiner can be found afterwards. Note that $\mathbf{V}_{\text{BB}}$ at the BS represents a set of precoders corresponding to all subcarriers, and so on. Similar operations can be applied to the optimization of the digital combiner $\mathbf{W}_{\text{BB}}[k]$ and $\mathbf{W}_{\text{RF}}$. Then, the optimization of each pair can be solved sequentially to achieve the best trade-off between the receiver performance and complexity [10].

It should be noted that even though the beamforming processing for the data transmission and reception is modeled

in a single-user system from the perspective of a single subcarrier, the proposed method introduced herein can be easily extended to multi-user MIMO-OFDM systems. Taking into account the effect of average signal-to-interference-plus-noise ratio (SINR) of a U-user MIMO-OFDM system, the achievable spectral efficiency in (4) becomes:

$$C_{\mathrm{MU}} = \sum_{k=1}^{K} \beta_k \log_2 \left(1 + \frac{||\mathbf{W}^H[k]\mathbf{H}[k]\mathbf{V}[k]||^2}{N_{t,u}\sigma_k^2 + \sum\limits_{\upsilon=1,\,\upsilon\neq u}^{U} \gamma_\upsilon \rho_{\upsilon_k}} \right), \quad (7)$$

where $\rho_{\upsilon_k} = |\mathbf{W}_\upsilon^H[k]\mathbf{H}_\upsilon[k]\mathbf{V}_\upsilon[k]|^2$ accounts for the effect of inter-user interference; $\mathbf{V}_\upsilon[k]$ and $\mathbf{W}_\upsilon[k]$ represent the hybrid digital/analog precoder of the υth user, assuming that all users share the whole hybrid precoder with M_t RF chains at the BS. Note that, the SINR is calculated from a total number of $U-1$ available users subject to the co-channel and/or adjacent interference, and γ_υ denotes the weight of the interference ratio of the υth user. For a multi-user downlink system, without loss of generality, the desired signal can be expressed as a true signal of the whole hybrid precoder and combiner designs, which considers the interference caused by the other $U-1$ users and can be written as

$$\tilde{\mathbf{z}}[k] = \mathbf{z}[k] + \sum_{\upsilon\neq u}^{U} \mathbf{W}_\upsilon^H[k]\mathbf{H}_\upsilon[k]\mathbf{V}_\upsilon[k]\mathbf{s}_\upsilon[k]. \quad (8)$$

B. Channel Model

From a viewpoint of subcarrier, the narrowband channel model incorporates the propagation characteristics of mmWave channels. Consequently, we first adopt an extended Saleh-Valenzuela geometric model, where the channel matrix $\mathbf{H}$ is assumed to be determined by the sum of L multipath clusters, assuming each cluster to be a complex Gaussian variable. With this clustered channel model, the channel matrix of the kth subcarrier $\mathbf{H}[k]$ can be written as

$$\mathbf{H}[k] = \mathsf{N}_L \sum_{\ell=0}^{L-1} \alpha_{k,l} \mathbf{a}_{\mathrm{R}_\ell} \mathbf{a}_{\mathrm{T}_\ell}^H e^{-j\frac{2\pi k}{K}}, \quad (9)$$

where $\mathsf{N}_L = \sqrt{\frac{N_{t,u}N_r}{L}}$ is a normalized factor; $\alpha_{k,\ell} \sim \mathcal{CN}(0,1)$ is the complex gain of ℓth path over the k subcarrier; $\mathbf{a}_{\mathrm{R}_\ell} \in \mathbb{C}^{N_r\times 1}$ denotes the receive array steering vector and $\mathbf{a}_{\mathrm{T}_\ell} \in \mathbb{C}^{N_{t,u}\times 1}$ denotes the transmit array steering vector. In a certain global reference assuming the azimuth angles are fixed, let θ_{R} and θ_{T} be the AOA and the AOD in elevation, respectively. The array steering vectors $\mathbf{a}_{\mathrm{R}_\ell}$ and $\mathbf{a}_{\mathrm{T}_\ell}$ corresponding to θ_{R} and θ_{T} are given by

$$\mathbf{a}_{\mathrm{R}_\ell} = \frac{1}{\sqrt{N_r}} \left[1, e^{\cos(\theta_{\mathrm{R}_\ell})}, \ldots, e^{(N_r-1)\cos(\theta_{\mathrm{R}_\ell})}\right]^T,$$
$$\mathbf{a}_{\mathrm{T}_\ell} = \frac{1}{\sqrt{N_{t,u}}} \left[1, e^{\cos(\theta_{\mathrm{T}_\ell})}, \ldots, e^{(N_{t,u}-1)\cos(\theta_{\mathrm{T}_\ell})}\right]^T. \quad (10)$$

For wideband and limited scattering mmWave communication systems, the measurement results reveal the nature of the wideband mmWave channels that there are $\{R_\ell\}_{\ell=0}^{L-1}$ rays within each of L multipath clusters. Given a cyclic-prefix length D, the discrete-time wideband channel $\mathbf{H}[k]$ associated with the delay-d can be obtained as [43], [44]

$$\mathbf{H}[d] = \mathsf{N}_{\mathrm{PL}} \sum_{\ell=0}^{L-1} \sum_{r=0}^{R_\ell-1} \xi(dT_s - \tau_{\ell,r}) \mathbf{a}_{\mathrm{R}_{\ell,r}} \mathbf{a}_{\mathrm{T}_{\ell,r}}^H, \quad (11)$$

where $\mathsf{N}_{\mathrm{PL}} = \sqrt{\frac{N_{t,u}N_r}{\mathsf{E}_{\mathrm{PL}}}}$ is a normalized factor dependent on path loss E_{PL}; $\xi(\cdot)$ is the transmit pulse shaping gain; d is the inter-element distance of the ULA; T_s is the duration of symbol; r represents the index of subpath; L is the number of multiple paths; $\tau_{\ell,r}$ denotes the time delay of the rth ray within cluster ℓ. For such a scattering environment, the channel matrix over subcarrier k can be expressed as

$$\mathbf{H}[k] = \sum_{d=0}^{D-1} \mathbf{H}[d] e^{-j\frac{2\pi dk}{K}}. \quad (12)$$

III. Problem Statement

For downlink mobile mmWave communications considered in this paper, a key consideration is that the received wideband signal is quite sensitive to the error of steering direction. A slight beam misalignment may lead to a significant drop of reception quality. Therefore, accurate DOA information is essential for both BS and MS. However, precise knowledge of the steering vector is usually difficult to be determined in a fast-varying channel environment, while the DOA information required by MS and BS is usually inaccurate and even lost due to the effects of the local scatterers, gain and phase errors in calibrated arrays, node mobility, as well as DOA estimation error. In a mobile mmWave network, destructive effect caused by mobility on the DOA accuracy shall be dealt with carefully. Moreover, the directional DOA perturbation correspondingly produces an erroneous steering vector, resulting in a significant drop of reception quality. In practice, a low-complexity mmWave front-end design with a low-resolution ADC or even few-bit (e.g., 1-3 bits) ADCs is a feasible solution, which ensures an acceptable level of power consumption and system costs. As mentioned earlier, channel estimation with few-bit ADCs is challenging, because the DOA estimation has to be achieved from coarsely quantized data.

Motivated by the above challenges, we propose a Bayesian beamforming method to mitigate the effects of uncertain DOA, which is modeled as a random variable with *a priori* pdf. With Bayes' theorem, the parameter estimation problem is converted into a more reliable probability optimization problem that combines *a priori* knowledge and *a posteriori* knowledge of DOA. When an indeterminate DOA occurs, the Bayesian estimator is able to probabilistically infer what the accurate DOA would be. It allows us to reconstruct DOA parameters by using MAP estimates by considering a reasonable first-order Markov model for DOA evolution.

Because our work mainly focuses on urban LOS and non-LOS (NLOS) channels, (12) can be rewritten as

$$\mathbf{H}[k] = \bar{\mathbf{H}}_{\mathrm{LOS}}[k] + \widetilde{\mathbf{H}}[k] = \mathcal{G}_0 \mathbf{a}_{\mathrm{R}_0} \mathbf{a}_{\mathrm{T}_0}^H + \sum_{d=1}^{D-1} \mathbf{H}_{\mathrm{NLOS}}[d] e^{-j\frac{2\pi dk}{K}}, \quad (13)$$

where $\mathcal{G}_0$ is the antenna gain of LOS. The NLOS cluster can be treated as ground-reflected paths within an angular spread and lower power (e.g., 10∼20 dB) than the LOS path, and the LOS path dominates in outdoor mmWave channels. For the sake of simplicity, we temporarily disregard the NLOS components, and treat their resultant effect as a disturbance term.

From (8), a closed-form expression for the desired signal can be expressed as

$$\mathbf{z}[k] = \underbrace{\mathcal{G}_0 \mathbf{W}^H[k]\mathbf{a}_{\mathrm{R}_0}\mathbf{a}_{\mathrm{T}_0}^H \mathbf{V}[k]\mathbf{s}[k]}_{\text{desired signals: LOS}} + \underbrace{\mathbf{W}^H[k]\widetilde{\mathbf{H}}[k]\mathbf{V}[k]\mathbf{s}[k] + \mathbf{W}^H[k]\mathbf{n}[k]}_{\text{effective signals plus noise: } \mathbf{e}[k]}. \quad (14)$$

According to the prior studies in [10], [23], and [47], the mmWave transmission is highly susceptible to blockages and radio propagation attenuation, and thereby relies on the LOS path. For this reason, the NLOS component can be regarded as a disturbance term. Consider that the global optimization problem is decoupled and the optimal beamformer and combiner can be systematically solved. Accordingly, we do not take the global optimization into account because of its overlarge computational complexity. After some simple manipulations, the final received signal can be simplified as

$$\mathbf{z}[k] = \mathbf{W}^H[k]\mathbf{a}[\theta]\mathbf{a}^H[\phi_0]\mathbf{V}_{\text{opt}}[k]\mathbf{s}[k] + \mathbf{e}[k], \quad (15)$$

where $\mathbf{e}[k]$ represents the perturbation term, and ϕ_0 is a temporary steering angle in the selected beam direction. From the viewpoint of the array geometry, the ULA can be regarded as a special subarray of a planar array (e.g., rectangular array), not only oriented on x-axis but also y-axis. As an example of the outdoor mmWave antenna array, a set of multiple connected antenna arrays is created in the x-y plane to yield different radiation patterns. To track temporal variations in the mobile channel, the full DOA information should be launched to select appropriate beam and/or switch from one beam to another. A generalized vector denotation not only can easily express array factors for linear arrays along other cardinal directions, but also is more effective to represent spatial reference, either azimuth or elevation beamforming or in both dimensions. Without loss of generality, we use $\boldsymbol{\theta}$ to represent the DOA information set in the following sections which can be viewed as the combination of any azimuth and elevation angle.

IV. Bayesian Approach for Robust Beamforming

A. Bayesian Beamforming With Uncertain DOA

Let $\mathbf{Z} = \{\mathbf{z}[n,k], \ldots, \mathbf{z}[n+m-1,k]\}$ denote a collection of m snapshots of the received data vectors. From (15), the kth subcarrier of the nth received signal at the subcarrier level, can be rewritten as

$$\mathbf{z}[n,k] = \mathbf{W}^H[n,k]\mathbf{a}[\boldsymbol{\theta}]\mathbf{a}^H[\boldsymbol{\phi}_0]\mathbf{V}_{\text{opt}}[k]\mathbf{s}[n,k] + \mathbf{e}[n,k], \quad (16)$$

where $\boldsymbol{\theta}$ could be viewed as the incident DOA of the received signal at time slot n.

Using the Bayesian approach, the unknown DOA parameter at each sampling moment is generally assumed to be a discrete random variable with *a priori* $p(\boldsymbol{\theta})$, and there exists a discrete set of $\mathcal{J}$ points $\boldsymbol{\theta} = \{\boldsymbol{\theta}_j\}\ \forall j \in \{1, .., \mathcal{J}\}$ over the whole parameter space. From the viewpoint of statistical inference, given observation $\mathbf{Z}$ with a hidden random variable $\boldsymbol{\theta}$, the detection probability for the desired $\mathbf{Z}$ is proportional to the likelihood function $\mathcal{L}(\boldsymbol{\theta}|\mathbf{Z})$, which can be written as

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{Z}) = p(\mathbf{Z}|\boldsymbol{\theta}) = \prod_{i=0}^{m-1} p(\mathbf{z}[n+i,k]|\boldsymbol{\theta}), \quad (17)$$

where $p(\mathbf{Z}|\boldsymbol{\theta})$ is the conditional pdf of observations.

By adopting a parametric representation of the Bayesian approach, the estimation problem with uncertainty is to infer $\boldsymbol{\theta}$ in terms of $p(\mathbf{Z}\,|\boldsymbol{\theta}\,)$ from the noisy observation $\mathbf{Z}$. It is clear that the azimuth and elevation angles associated with DOA are independent, and $\boldsymbol{\theta}$ can therefore be considered as a single variable to reduce the computational complexity. Again, denoting the joint pdf of $\mathbf{Z}$ and $\boldsymbol{\theta}$ as $f(\mathbf{Z}, \boldsymbol{\theta})$, we have

$$\begin{aligned} f(\mathbf{Z}, \boldsymbol{\theta}) &= p(\mathbf{Z}\,|\boldsymbol{\theta}\,)p(\boldsymbol{\theta}) \\ &= p(\boldsymbol{\theta}\,|\mathbf{Z})m(\mathbf{Z}) = p(\boldsymbol{\theta}\,|\mathbf{Z}\,)\int_{\boldsymbol{\theta}} p(\mathbf{Z}\,|\boldsymbol{\theta}\,)p(\boldsymbol{\theta})d\boldsymbol{\theta}, \end{aligned} \quad (18)$$

where $m(\mathbf{Z})$ is the probability mass function (PMF). In this case, only the *a posteriori* probability can be used for inference, which is calculated by

$$p(\boldsymbol{\theta}\,|\mathbf{Z}\,) = \frac{f(\mathbf{Z}, \boldsymbol{\theta})}{m(\mathbf{Z})} = \frac{p(\mathbf{Z}\,|\boldsymbol{\theta}\,)p(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} p(\mathbf{Z}\,|\boldsymbol{\theta}\,)p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (19)$$

In this way, the solution to the Bayesian parameter inference problem is converted into a more tractable *a posteriori* pdf based statistical inference by incorporating the PMF, the prior belief, and the evidence provided by the observed data. Since $m(\mathbf{z})$ does not depend on $\boldsymbol{\theta}$, it can be treated as a normalized constant. As a result, (19) can be rewritten in a mathematically equivalent form as

$$p(\boldsymbol{\theta}\,|\mathbf{Z}\,) \propto p(\mathbf{Z}\,|\boldsymbol{\theta}\,)p(\boldsymbol{\theta}). \quad (20)$$

According to Bayes' theorem, our task is to realize statistical inferences about the unknown parameter $\boldsymbol{\theta}$. Suppose that a meaningful a priori pdf is a normal distribution with given parameter μ_n and δ_n that can be chosen for the parameter $\boldsymbol{\theta}$, and then the a posteriori pdf of $\boldsymbol{\theta}$ also corresponds to a normal distribution and can thus improve the parameter estimation accuracy.

In the presence of DOA uncertainty, given the noisy observation data set $\mathbf{Z}$, the Bayesian approach enables us to reconstruct DOA parameter value by using its MAP estimate. With the logarithmic representation, the MAP estimate can be obtained by [45]

$$\hat{\boldsymbol{\theta}}_{MAP} = \arg\max_{\boldsymbol{\theta}}\ p(\mathbf{Z}\,|\boldsymbol{\theta}\,)p(\boldsymbol{\theta}). \quad (21)$$

To evaluate the level of uncertainty, a simple approach is to employ the Kullback-Leibler (KL) divergence as a distance measure of the dissimilarity between two distributions of estimates [28], which can be calculated by

$$D(\boldsymbol{\theta}_j||\boldsymbol{\theta}) = \mathbb{E}_j\left\{\ln\frac{p(\mathbf{Z}\,|\boldsymbol{\theta}_j\,)}{p(\mathbf{Z}\,|\boldsymbol{\theta}\,)}\right\}. \quad (22)$$

For analytical purposes, we consider a discrete setting, in which $\boldsymbol{\theta}$ is defined as a discrete set of $\mathcal{J}$ candidates. Consequently, the estimate of the desired signal in (16) can be rewritten in the form of

$$\begin{aligned}&\hat{\mathbf{z}}[n,k]\\&=\sum_{j=1}^{\mathcal{J}} p(\boldsymbol{\theta}_j\,|\mathbf{z}[n,k]\,)\mathbf{W}^H[\boldsymbol{\theta}_j,n,k]\mathbf{a}[\boldsymbol{\theta}_j]\mathbf{b}_{n,k}^H\mathbf{s}[n,k]+\mathbf{e}[n,k].\end{aligned} \tag{23}$$

Accordingly, the optimal estimation of $\boldsymbol{\theta}_j$ is to find one minimizing the KL divergence by

$$\hat{\boldsymbol{\theta}}_j = \underset{\boldsymbol{\theta}_j\in\boldsymbol{\theta}}{\arg\min}\, D(\boldsymbol{\theta}||\boldsymbol{\theta}_j). \tag{24}$$

Recall from (5) that the weight vectors of combined matrix $\mathbf{W}$ can be determined by solving the optimization problem by certain criteria. In this paper, we follow the approach proposed in [4], [10]. Considering $\hat{\mathbf{Z}}$ with $p(\mathbf{Z}|\boldsymbol{\theta}_j)$, the optimization problem of the Bayesian beamformer $\mathbf{W}_{\mathrm{B}}[\boldsymbol{\theta};n,k]$ turns out to minimize the conditional minimum mean squared error (MMSE) [45], namely

$$\mathbf{W}_{\mathrm{B}}[\boldsymbol{\theta};n,k] = \underset{\mathbf{W}_{\mathrm{B}}[\boldsymbol{\theta};n,k]}{\arg\min}\,\mathrm{Tr}\left[(\hat{\mathbf{Z}}-\mathbf{Z})(\hat{\mathbf{Z}}-\mathbf{Z})^H\right]. \tag{25}$$

Given M possible DOA values, the weight vector of $\mathbf{W}_{\mathrm{B}}$ is chosen according to *a posteriori* pdf of the DOA, and can be expressed as

$$\mathbf{w}_{\mathrm{B}}[n,k] = \sum_{m=1}^{M} p(\boldsymbol{\theta}_m|\mathbf{Z}[n,k])\mathbf{w}_{\mathrm{B}}[\boldsymbol{\theta}_m;n,k]. \tag{26}$$

Note that, based directly on Bayes' theorem, the use of Bayesian estimator is equivalent to yielding an estimate of $\hat{\boldsymbol{\theta}}$ as close as possible to the actual value of $\boldsymbol{\theta}$, with respect to the posterior information.

B. Determination of A Priori Probability

Taking into account the angle and delay distribution model in [46] and the most recent measurement results [23], [43], our paper considers the *a priori* DOA distribution as a Bernoulli Gaussian-mixture (GM) with an unknown parameter $\boldsymbol{\theta}$. Based on the channel estimation with few-bit ADCs results in [20], we aim to obtain an approximate fit to the true channel distribution. Meanwhile, motivated by the results presented in [20], [46], the angle and delay of the ℓth path are modeled by the angle and delay channel coefficient matrix $\boldsymbol{\chi}[\ell]\in\mathbb{C}^{N_r\times N_{t,u}}$, in which the (i,j)th entry of the matrix is the channel gain between the jth discrete transmit angle and the ith discrete receive angle, equivalent to the angle and delay channel representation proposed in [20]. Accordingly, the coefficient χ_i of $\boldsymbol{\chi}$ is generated by the following expression:

$$p(\chi_i;\boldsymbol{\theta}) = \lambda_0\delta(\chi_i) + \sum_i \lambda_i\mathcal{CN}(\chi_i;\mu_i,\Sigma_i), \tag{27}$$

where $\lambda_0 = \Pr\{\chi_i = 0\}$ and λ_i $(0\le\lambda_i\le 1)$, μ_i, and Σ_i are the weights, means, and variances of the GM over $\boldsymbol{\theta}$, respectively; $\delta(\cdot)$ is the Dirac delta function. The parameter λ provides an additional degree of freedom characterizing the GM distribution. As the details of the modeling process are out of the scope of this paper, we do not present them here. Note that the underlying distributions will not necessarily be limited to the Gaussian distribution. Most recently, the authors in [47] also proposed that the angular-domain channel coefficients could be modeled by Laplacian distributed random variables.

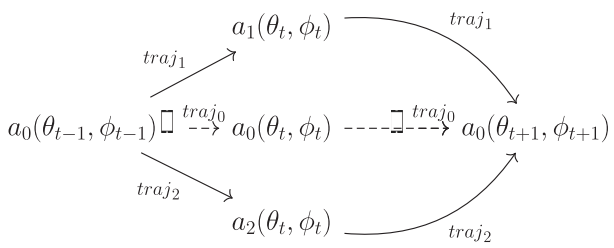


Fig. 2. A priori information of a two-dimensional trajectory.

Generally, the estimates of *a priori* pdf in (27) can be determined by a series of iterative quantization algorithms, e.g., MAP and Bayesian linear regression. However, the estimates of (μ_i,Σ_i) of *a priori* pdf have to consider all possible combinations of discrete DOA values in the whole parameter space as well as all possible combinations of a posterior probability estimation. Therefore, once the optimization procedure falls into a wide search range, it would take a great amount of time to converge to the optimal solution. Such a pitfall will also lead to high computational complexity.

To solve the pitfall above, we propose a moving trajectory based constraint relying on the motion features that are not completely random. Therefore, we can to some extent utilize the regularity to restrict the search space of DOA estimates. It means that $\boldsymbol{\theta}$ must lie in a known interval, and we suppose that the true value of $\boldsymbol{\theta}$ has been chosen from that interval, which is capable of reducing the variance and therefore improving the estimation accuracy [48]. The constraint is pictorially illustrated in Fig. 2. We assume that each moving trajectory corresponds to a specific *a priori* pdf subset, and the channel information, i.e., the spatial directions, can be collected from multiple specific moving tracjectories. With this constraint, the parameter range of the expected *a posteriori* pdf, e.g., variance, can therefore be limited to a smaller region, resulting in a faster convergence rate and a higher estimation accuracy. As the future work, it is worth noting that the beam entropy proposed in [33] can also provide an indication for the beamspace randomness of mmWave channels. It will be thereby helpful to investigate the quantitative relation between the channel randomness and the corresponding beam entropy values, in accordance with the spatial characterization of beam combining in [43].

The motivation behind this constraint is that the moving trajectory contains valuable *a priori* information, since the BS is geographically stationary, and the moving trajectory of a target object is usually regular. Particularly, this regular feature might be more obvious in high-mobility scenarios. Suppose that there exists a $\mathcal{K}$-element overlapping subset $\{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\ldots\boldsymbol{\theta}_{\mathcal{K}}\}$ in the whole DOA parameter search space.

Each element in the overlapping subset could be defined as a subspace and is comprised of *a priori* pdf corresponding to a specific moving trajectory. This feature implies that $\boldsymbol{\theta}$ can be reconstructed by

$$\boldsymbol{\theta}_1 \cup \boldsymbol{\theta}_2 \ldots \cup \boldsymbol{\theta}_{\mathcal{K}} = \boldsymbol{\theta}. \tag{28}$$

At each time instant, it is common that we do not completely know *a priori* distribution of the angle and delay channel over the whole space. These subspace constraints lead to a significant reduction in the degree of freedom and thereby lower the variance of the DOA estimates. In the trajectory-constrained case, it is important to note that the joint Bayesian estimation of AOA/AOD is a very common type of parameter point estimation problem, and the formulation can be similarly derived to form the best beam pair. For this reason, the developed methodology can be directly applied to facilitate the high-resolution estimation of the azimuth/elevation AOD at the BS.

C. Estimation of A Posteriori pdf

According to Bayes' theorem with *a priori* pdf that is modeled as a GM, the *a posteriori* pdf of $\boldsymbol{\theta}$, a discrete setting of (19), can be expressed as

$$p(\chi_i; \boldsymbol{\theta} \,|\mathbf{z}) = \frac{p(\mathbf{Z} \,|\chi_i; \boldsymbol{\theta})p(\chi_i; \boldsymbol{\theta})}{\sum_{i=1}^{\mathcal{J}} p(\mathbf{Z} \,|\chi_i; \boldsymbol{\theta}_i)p(\chi_i; \boldsymbol{\theta}_i)}. \tag{29}$$

The DOA-dependent hyperparameter estimates $\hat{\mu}_i$ and $\hat{\Sigma}_i$ are obtained by maximizing the marginal log-likelihood distribution [20]. Mathematically, the estimates can be written as

$$(\hat{\mu}_i, \hat{\Sigma}_i) = \arg\max_{\mu_i, \Sigma_i} \sum_{i=1}^{\mathcal{J}} (\ln p(\mathbf{Z}|\chi_i; \mu_i, \Sigma_i) + \ln p(\chi_i; \mu_i, \Sigma_i)). \tag{30}$$

Note that the mean and variance of this *a posteriori* approximation are tractable by means of an iterative method, e.g., a gradient-based adaptive estimator with optimal step-size control. In our work, the approximation of mean and variance can be carried out analytically by a data-aided iteration procedure. The significance of this *a posteriori* approximation is that an alternative method, based on the observed signal, is provided by a variational distribution approximation, resulting in a considerable reduction in computational complexity. In this case, the optimization objective becomes

$$\hat{\boldsymbol{\theta}}_{MAP} = \arg\max_{\boldsymbol{\theta}} \, (\ln p(\mathbf{Z} \,|\chi_i; \boldsymbol{\theta}) + \ln p(\chi_i; \boldsymbol{\theta})). \tag{31}$$

Obviously, the instantaneous *a posteriori* pdf is also unavailable in practice. Therefore, what we can do is to apply an iterative EM online processing scheme. Consequently, both mean and variance of the *a posteriori* pdf are estimated in an iterative manner from the received signals across all receive antennas over the kth subcarrier.

D. Complexity Analysis

With the aforementioned trajectory based constraint, we have proposed an approximation method of the estimates of *a priori* pdf to reduce the computational complexity. This parameter point estimation for the approximation of *a posteriori* probability can be performed with lower computational complexity. In other words, it is not necessary to calculate the full posterior $p(\boldsymbol{\theta}|\mathbf{Z})$. Taking into account the computational complexity corresponding to the estimates of *a priori* pdf, one simple but powerful alternative is the calculation of the *a priori* pdf that can usually be generated by Markov chain Monte Carlo (MCMC) simulation [49]. Meanwhile, a first-order autoregressive model can be further employed to capture the temporal variation. In this case, the remaining computations are the estimation of the Gaussian mixture parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ of the *a posteriori* pdf.

To evaluate the a posterior probability model to estimate the unknown $\boldsymbol{\theta}$, we specify a squared error loss function $L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^2$, resulting in an optimization problem with the MMSE criterion for the posterior media. For a given observation $\mathbf{Z}$, it is more common to use the posterior mean as the Bayesian estimate, and the optimal estimate of $\boldsymbol{\theta}$ can be determined in closed form as

$$\hat{\boldsymbol{\theta}} = \mathbb{E}[\boldsymbol{\theta} \,|\mathbf{Z}] = \int_{\boldsymbol{\Theta}_{\mathcal{K}}} \boldsymbol{\theta} p(\boldsymbol{\theta} \,|\mathbf{Z}) d\boldsymbol{\theta}. \tag{32}$$

Generally speaking, the statistics of $\boldsymbol{\theta}$ vary on a continuous scale, whereas it is usually hard to obtain the form of the continuous distribution in practice. For simplicity, it is customarily modeled in the discrete-time domain, i.e., $\boldsymbol{\theta}$ takes discrete values to yield a general discrete probability distribution. For a specific moving trajectory $\mathcal{K}$, given the observed data $\mathbf{Z}_{\mathcal{K}}$, we can manipulate the above expression to cast it in a discrete form, and the estimation of $\boldsymbol{\theta}$ can be given by

$$\hat{\boldsymbol{\theta}} = \mathbb{E}[\boldsymbol{\theta} \,|\mathbf{Z}_{\mathcal{K}}] = \sum_{j=1}^{\mathcal{J}} \boldsymbol{\theta}_j p(\boldsymbol{\theta}_j \,|\mathbf{Z}_{\mathcal{K}}). \tag{33}$$

In the subsequent step of this paper, the parameter estimates associated with the posteriori are iteratively updated using the EM algorithm. As a data-driven online learning method, the EM algorithm has been proven to be an effective tool for simplifying computational problems associated with the maximum posteriori. Generally, the complexity of the estimation algorithm is dominated by the expectation step (E-step), whereas the calculation of maximization step (M-step) is negligible in terms of computational complexity.

Considering the overall transmit data matrix $\mathbf{X} \in \mathbb{C}^{N \times NN_t}$ with N_p pilot subcarriers, we denote ρ_q as the transmit power at the qth OFDM symbol. With sampling at the pilot subcarriers, the MIMO-OFDM system model in (2) can be rewritten as

$$\mathbf{Z}_{p,q} = \sqrt{\rho_q}\mathbf{H}_q \boldsymbol{\Phi}_{p,q}^T + \mathbf{N}_{p,q}, \tag{34}$$

where $\mathbf{Z}_{p,q} \in \mathbb{C}^{N_p \times N_r}$, $\boldsymbol{\Phi}_{p,q} \in \mathbb{C}^{LN_t \times N_p}$ and $\mathbf{N}_{p,q} \in \mathbb{C}^{N_p \times N_r}$.

At the rth received antenna, since the posterior distribution with the hidden random variable $\boldsymbol{\theta}$ in $\mathbf{H}_q$ is iteratively obtained by combining the likelihood and the prior distribution, for the ith iteration of the EM algorithm, we obtain the corresponding mean and covariance by the method developed in [50]

as follows

$$\boldsymbol{\mu}_r^{(i)} = \frac{1}{\sigma^2}\boldsymbol{\Sigma}\boldsymbol{\Phi}_p^* \mathbf{z}_{p,r}^{(i)},\ \boldsymbol{\Sigma} = \left(\frac{\boldsymbol{\Phi}_p^* \boldsymbol{\Phi}_p}{\sigma^2} + \boldsymbol{\Delta}^{(i)^{-1}}\right)^{-1}, \quad (35)$$

where $\boldsymbol{\Delta}^{(i)} = \text{diag}(\boldsymbol{\delta}_\ell)$, and $\boldsymbol{\delta}_\ell = [\delta_1, \ldots, \delta_L]^T$ is the variances of the elements of the *a priori* pdf, i.e., the hyperparameter value in the ith iteration.

From (9) and (10), we can see that the E-step will involve the computation of the posterior associated with the DOA-dependent hyperparameter estimates as given in (31). At each iteration, the EM algorithm increases a lower bound on the logarithmic posterior $\log p(\boldsymbol{\mu}|\mathbf{z})$. Accordingly, the computational cost of the E-step at each iteration can be given by $\mathcal{O}(N_p^2 L)$, while the M-step performs the update of hyperparameter involved in the computational complexity $\mathcal{O}(N_t N_r L)$.

The benchmark results in Section VI-C will show that the proposed algorithm is feasible to reduce the computational complexity to a practical level, so as to provide guarantees for the Bayesian inference in a finite time-window. The proposed Bayesian framework is guaranteed to converge to a stationary point and/or a local maximum of the posterior with a small number of samples (in the order of hundreds of OFDM symbols) of iterations. This result suggests that the proposed algorithm is effective for many practical applications.

V. Robust Channel Tracking

A. DOA Tracking With EM Algorithm

To apply the proposed scheme to the time-varying channels, it is reasonable to assume that the evolution of the DOA is approximately constant for a number of symbol intervals. For notational convenience, we define a discrete time index t to be a normalized symbol block duration. According to a time-variant state-space model, the dynamics of DOA can be described as a first-order Gauss-Markov process by

$$\alpha^t = \nu\alpha^{t-1} + \xi_\alpha, \quad (36)$$

and

$$\boldsymbol{\theta}^t = \boldsymbol{\theta}^{t-1} + \Delta\boldsymbol{\theta} + \xi_{\boldsymbol{\theta}}, \quad (37)$$

where $\nu \in (0, 1]$ is the fading correlation coefficient; $\xi_{\boldsymbol{\theta}} \sim \mathcal{CN}(0, \sigma_{\boldsymbol{\theta}}^2)$ and $\xi_\alpha \sim \mathcal{CN}(0, \sigma_\alpha^2)$ are zero-mean Gaussian random noise terms that allow $\boldsymbol{\theta}$ and α to change over time.

Recalling the signal model proposed in Section II-A, the estimate of DOA can be performed by a subspace-based narrowband DOA estimator, such as the classical MUSIC or ESPRIT algorithm. Considering that the ULA satisfies the time shift requirement without deformation, we choose the ESPRIT method to obtain the direction estimate. More importantly, the ESPRIT method has the ability to obtain the angular delay component estimate for each path. In addition, the lower computational complexity of ESPRIT is one of the features that we are interested in. From the definition given in (37), the *a posteriori* pdf associated with the observations is given by

$$p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=0}^{n} p(x_i|\boldsymbol{\theta}) \prod_{i=0}^{n} p(\boldsymbol{\theta}^{i+1}|\boldsymbol{\theta}^i). \quad (38)$$

Starting from (30), the Bayesian DOA estimation is affected by the a priori probability and the *a posteriori* probability. In the actual tracking process, we need to dynamically estimate the GM distribution parameters associated with a particular motion trajectory according to (31). Furthermore, it is clear that once the *a priori* pdf is determined, the Bayesian estimation only depends on the *a posteriori* probability.

To find the unknown parameters of the *a posteriori* pdf, we adopt a data-aided EM algorithm, where the transmitted symbol $\mathbf{x}$ is supposed to be known. Typically, the EM algorithm is an iterative method, and each iteration cycle consists two steps: the E-step and M-step [51]. We introduce both steps as follows.

E-step: Define $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ as the expected value of the logarithmic a posterior function of $\boldsymbol{\theta}$. The new parameters are estimated from the recent snapshot $\mathbf{Z}$ by

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i-1)}) = \mathbb{E}\left[\log p(\mathbf{z}|\boldsymbol{\theta})|\mathbf{z} = \mathrm{Z}, \boldsymbol{\theta}^{(i-1)}\right]. \quad (39)$$

M-step: Iteratively find the parameters according to the criterion given as follows:

$$\boldsymbol{\theta}^{(i)} = \arg\max_{\boldsymbol{\theta}}\ \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{(i-1)}). \quad (40)$$

Note that the EM algorithm starts with an arbitrary initial guess, due to the GM assumption, and the parameters are characterized in Section IV-B. In the M-step, the latest parameters are updated. The M-step ensures that each iteration cycle increases the *a posteriori* until a local maximum is reached. For hyperparameter estimates of GMs, $\hat{\mu}_i$ and $\hat{\Sigma}_i$ can be obtained by (30) straightforwardly, and there is only a slight difference in the form of the $\mathcal{Q}$ function, as shown in (29) and (30).

B. Procedure of Channel Tracking

To ensure the actual performance of the Bayesian beamformer, we devise an efficient beam tracking strategy based on the iterative EM algorithm to have better robustness and maintain a high tracking accuracy considering node mobility. We focus on tracking the dynamics of channel in mobile mmWave systems and adopt the angular motion model proposed in [52]. We summarize our methodology in Algorithm 1, and the descriptions are given as follows.

We start with the traditional beam searching and training procedure introduced in [21] for initializing access beam training, and obtain the *a priori* pdf. At the transmitter, the BS uses the training precoding matrix of the first level of the codebook. At the receiver, the MS uses the measurement vectors of the first level to combine the received signal. In the iteration step, we assume that the MS is moving along a fixed trajectory or with a mobility pattern. According to Bayes' theorem, the estimation of the latest set $\hat{\boldsymbol{\theta}}$ of suitable directions is used to alleviate the DOA uncertainty. During the iterative EM procedure, we do not need to feed back explicit DOA estimate to the BS, and the training overhead is therefore reduced.

Because the optimal digital precoder $\mathbf{V}_{\text{BB}}$ needs to be jointly designed with the analog beamforming/combining matrix, a large amount of feedback and training overhead between

Algorithm 1 EM-Based mmWave Channel Tracking

Input: $\mathbf{s}(n,k)$, $\mathbf{V}_{\mathrm{BB}}[k]$, $\mathbf{V}_{\mathrm{RF}}$, $\mathbf{W}_{\mathrm{RF}}$, $\mathbf{W}_{\mathrm{BB}}[k]$
Output: estimate of $\boldsymbol{\theta}$, $\mathbf{z}(n,k)$

initialization: Guess $\hat{\boldsymbol{\theta}}_0$, $p(\boldsymbol{\theta})$, Determine $p(\boldsymbol{\theta})$ with $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \ldots \boldsymbol{\theta}_{\mathcal{K}}\}$ ▷ initialization step
Repeat ▷ iterative step
for all $n = 1, 2, ..$ **do**
  E-step: Compute the latest $\boldsymbol{\theta}$ from $\boldsymbol{\theta}_{t-1}$
  $\mathcal{Q} \leftarrow \ln p(\mathbf{Z}\,|\chi_i; \boldsymbol{\theta}) + \ln p(\chi_i; \boldsymbol{\theta})$
  M-step: Find $\boldsymbol{\theta}$, which maximizes $\mathcal{Q}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_{t-1})$
  for all $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{\mathcal{J}}\}$ **do** ▷ candidate $\boldsymbol{\theta}$
  Evaluate *a posteriori* pdf, $\forall \mathcal{J}$
  end for
  Bayesian inference
  Update $\mathbf{W}_B[\boldsymbol{\theta}; n, k] \leftarrow \mathbf{W}_{\mathrm{RF}}$, $\mathbf{W}_{\mathrm{BB}}[k]$
  Detect abrupt changes [53]
  if abrupt changes do not occur **then**
  Feedback: $\boldsymbol{\theta}$ ▷ Feedback step
  Update $\mathbf{V}_{\mathrm{RF}}$, $\mathbf{V}_{\mathrm{BB}}[k]$ ▷ $\mathbf{V}_{\mathrm{opt}}$
  else
  Return $(\boldsymbol{\theta})$
  end if
end for
Until $D(\boldsymbol{\theta}_{t-1}||\boldsymbol{\theta}) < \epsilon$ ▷ KL divergence

BS and MS is required, including CSI [6], [52]. However, it should be noted that instantaneous CSI is generally inaccessible at the transmitter, because the numbers of transmit and receive antennas are very large, which means the overall feedback overhead is very high, resulting in a significant loss of transmission efficiency.

To address this problem, we thereby design an alternative limited feedback strategy. Our idea is that the current channel tracking is performed by the Bayesian estimator, while the tracking error accumulates over time. At the same time, we adopt an abrupt change detection scheme proposed in [53], to determine whether the tracking is no longer reliable or the path does not exist anymore by blockage. Once the abrupt change point is detected and/or the Bayesian approach cannot infer the beam directions, the corresponding countermeasure will be triggered to deal with the abrupt change. In Algorithm 1, the mobile system transmits the channel estimate information back to the BS to calculate the precoding weights needed for the subsequent data transmission. According to the reciprocity principle, the transmitter can update the beamforming matrix according to the feedback from the receiver, and then the highly directional beam matching the current DOA can be generated. This mechanism allows us to facilitate closed-loop transmit processing via a limited feedback link [6], and robustly track the dynamics of the channel between the transmitter and receiver.

VI. Numerical Results

A. Simulation Settings

We consider a 64×16 hybrid beamforming architecture with a ULA, where the arrays at the BS and MS are equipped with $N_t = 64$, $N_r = 16$ antenna elements and 4 RF chains. To investigate the performance of the proposed approach in close to the real scenarios, our simulation is conducted over 28 GHz frequency band in a typical urban macro-cellular (UMa) scenario with LOS and NLOS components, using the angular motion model proposed in [52]. The system has a bandwidth of 100 MHz and the path loss exponent is 3. In this simulation, we adopt an open-source channel simulator, NYUSIM, which supports the parameterization and validation of channel models on a reliable statistical basis [55]. With NYUSIM, channel statistics based on extensive mmWave measurements can be recreated from a variety of antenna beamwidths and environmental conditions. We consider a sparse mmWave channel with $L = 3$ multipath clusters and $R_\ell = 6$ delay spread, which are reconstructed according to the channel model described in (11). Moreover, we assume that the AOD ϕ and AOA θ are continuous and uniformly distributed over $[-\pi/2, \pi/2]$. Note that ϕ and θ are measured from the +x-axis in the x-y plane and the x-z plane, respectively. In the simulations, the MIMO-OFDM system has $k = 512$ subcarriers. Table 1 summarizes the adopted parameters and simulation settings.

TABLE I
System Parameters and Simulation Settings

System parameter	simulation setting
Carrier frequency	28 GHz
Bandwidth	100 MHz
Symbol duration	3.7 μs
Cell radius	50~100 m
Maximum tap delay	20 ns
DOA of clustered waves	6
Receiver noise figure	10 dB
Effective radiated power	+ 24 dBm (peak)
Speed of MS	20 m/s
Channel model	mmWave statistical channel model [54]

B. Performance of Bayesian Beamformer

To evaluate the effects of DOA uncertainty, we first consider the case of the existence of array calibration errors, as well as phase and amplitude errors on the array gain (beamforming). For instance, the array suffers from both geometrical and electrical uncertainties, including the uncertainty of the phased array antenna, i.e., phase fluctuation or error. To simulate the random array errors, we suppose that each element of the steering vector $\mathbf{a}(\phi)$ is subject to a perturbation of a complex Gaussian random variable with zero mean and variance σ_c^2. For simplicity, we suppose that all randomly perturbed arrays have exactly the same statistical properties, and the expected azimuth angle is set to be $\phi = 0^\circ$. Fig. 3 compares the proposed Bayesian method and existing data-independent methods in the high SNR regime. It demonstrates that the significant gains can be achieved by the proposed Bayesian beamforming approach. As can be seen from Fig. 3, the method proposed in [7] may not take into account the effect of array error, e.g., phase fluctuation and/or interference, and does not form the expected gain in a particular direction. In this case, we notice that the method proposed in [20] performs better

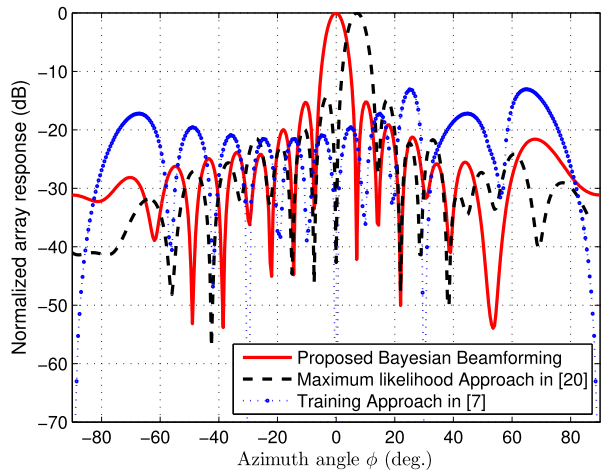


Fig. 3. Synthetic array response patterns of beamformer with perturbation of a complex Gaussian random variable. θ is set to be $\pi/4$; σ_c^2 is set to be 0.01; SNR $= 20$ dB.

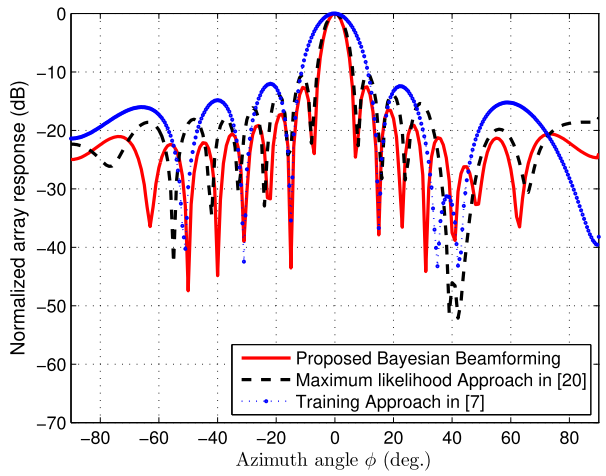


Fig. 4. Synthetic array response patterns of beamformer under inter-user interference, which assume that the same beam is assigned to different users. θ is set to be $\pi/4$; σ_c^2 is set to be 0.01; SNR $= 5$ dB.

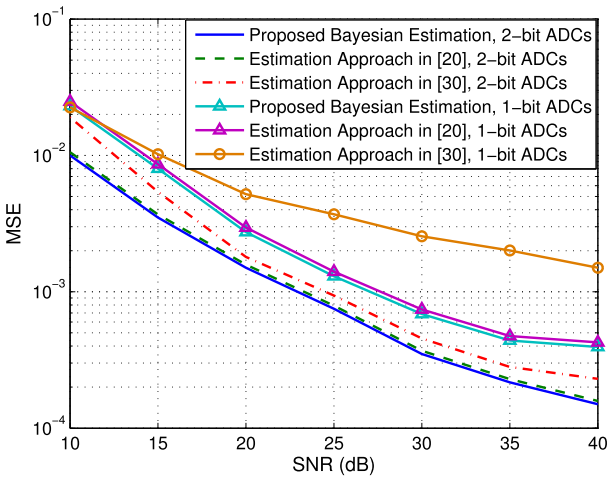


Fig. 5. MSE of DOA estimation vs. SNR: a lower-bit (1-2 bits) quantization in the high SNR regime.

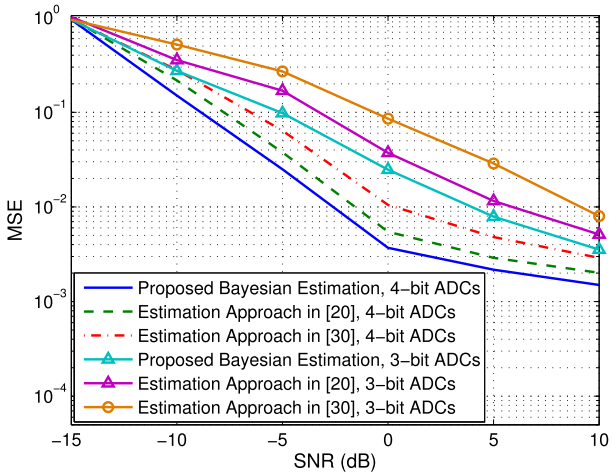


Fig. 6. MSE of DOA estimation vs. SNR: a higher-bit (3-4 bits) quantization in the low SNR regime.

than that of [7] but has a direction error of about 5°. We can also see that the Bayesian approach offers better performance than the aforementioned two methods, and the steering gain obtained is almost the same as the expected direction. This demonstrates that our proposed method is robust to uncertain estimates of θ.

In Fig. 4, we further evaluate the performance of the calibrated radiation pattern in the low SNR regime, where the inter-user interference is taken into account. According to the analytical model constructed in (7), the achievable spectral efficiencies multi-user case can be investigated. In our simulation, the hypothetical interference beams come from two directions, 20° and 10°, and each presents a strong interference signal out of the receive band. For simplicity, the signal strength is uniformly set to be 20 dBm, where we assume that each interference signal has the identical leakage ratio that is set to be $\gamma = 0.1$. Thereafter, we observe that all schemes can provide accurate steering vectors, but the array response pattern obtained in [7], [20] have a wider response range. In contrast, the Bayesian method produces a 5 dB higher steering gain in an expected direction. This observation suggests that the Bayesian method provides a more considerable reduction in interference, and is feasible to minimize the loss of sum rate with low complexity.

In Figs. 5 and 6, the MSE performance of DOA estimation is evaluated at different ADC quantization levels. As described in Section V, the statistical properties of DOA are the product of the GM distribution, where a sufficiently large number of snapshots could be generally used to obtain the correct statistical properties. Correspondingly, the MSE of the DOA estimate can be expressed as $\mathbb{E}\{|\theta - \hat{\theta}|^2\}$. Fig. 5 shows the MSE performance for the estimated parameters θ using 1- and 2-bit quantization in the high SNR regime. It is shown in Fig. 5 that the proposed method first discussed in [20] are significantly better than the method given in [30]. Furthermore, it is noteworthy that the Bayesian estimation is slightly improved compared to the noisy quantized CS algorithm in [20], whereas similar angle and delay distribution models are taken into account in our work. This can be also interpreted as that the method proposed by [20] exploited sparsity in both the

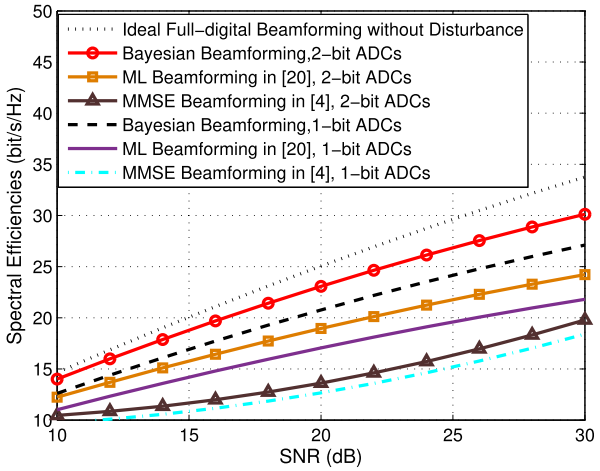


Fig. 7. Spectral efficiencies vs. SNR for different methods comparison with a complex Gaussian perturbation at high SNR. $\gamma = 0.1$; μ_c and σ_c^2 are set to be 1 and 0.01, respectively.

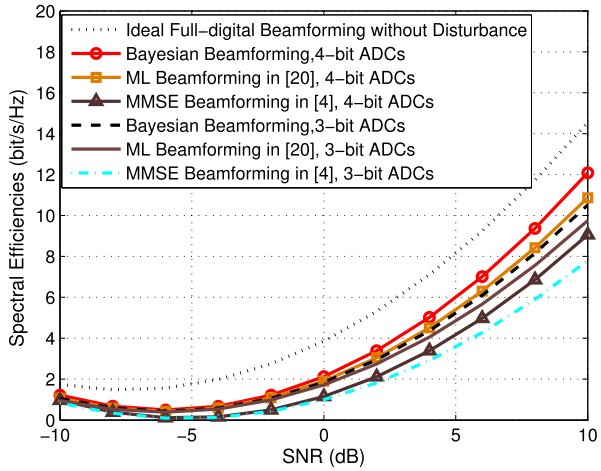


Fig. 8. Spectral efficiencies vs. SNR for different methods comparison under inter-user interference at low SNR. Here, the direction of hypothesis interference beams are 20° and 10° with transmit power of 10 dBm.

angle and delay domains and used an efficient approximate message passing algorithm to provide estimates with nearly minimum MSE. Another significant difference is that [20] would introduce an additional 10-20% tracking overhead, resulting in a loss of transmission efficiency. In light of this, the data-driven Bayesian method provides better overall performance than that expounded in [20]. In the low SNR regime, according to the experimental results in [4], [20], all schemes use a higher-bit (3-4 bits) quantization scheme. The performance gap between the proposed approach and prior literature is significant. In light of this, we confirm that the Bayesian scheme offers better estimation performance.

Fig. 6 illustrates the MSE performance of the DOA estimation in the low SNR regime. Interestingly, we find that the proposed method and the approach in [20] outperform that of [30]. At a higher resolution ADC level (3-4-bit), the Bayesian method is significantly better than that advanced by [20], and we observe an increase of channel gain in the antenna arrays by 35% compared to conventional methods. In return, it is capable of providing SNR boost via beamforming gain to invoke higher order modulations, which leads to a significantly improvement of system throughout without any additional SNR margin. This has been interpreted in [20] and [30] by using *a priori* maximum likelihood parameter estimation methods, while the Bayesian method proposed herein implements a combination of *a priori* and *a posteriori* information, which is more appropriate for the GM parameter estimation.

Next, according to the same configurations of simulation, we adopt Monte Carlo simulations to evaluate the spectral efficiency of the system with noisy few-bit quantization. These spectral efficiency metrics are illustrated as a function of SNR, which are approximated by averaging over 1000 independent channel realizations. Fig. 7 shows a comparison of Bayesian beamforming with other advanced hybrid beamforming schemes in the high SNR regime, where the achievable spectral efficiency is also shown to serve as a performance benchmark. In the presence of random DOA perturbation with a complex Gaussian random variable, where the mean μ_c is normalized, and the variance is set to be 0.01. Compared with the method of [20], the gain of this method is about 2 dB, which is almost 1.5 dB higher than that proposed in [4]. In Fig. 7, it is shown that when random DOA perturbation is small, the spectral efficiency achieved by the Bayesian method is very close to the approximate optimal case of full-digital beamforming without DOA perturbation, which confirms that the proposed method is near-optimal. In the low SNR regime, we still focus on the situation of multi-user interference. In Fig. 8, we compare the total system capacity obtained by different methods. In particular, we investigated a strong inter-user interference presenting in a specific direction, and compare spectral efficiencies yielded by different methods. This situation would be understood as a higher level of DOA uncertainty. The numerical results show that using the proposed technique can achieve better spectral efficiency than the methods of [20] and [4] even if a strong inter-user interference presents in a specific direction.

C. Performance of Robust Channel Tracking

To simulate the variations of the steering angle of the data beam, in this section we utilize the angular motion model proposed in [52] to simulate the DOA angle variations of impinge on the antenna arrays, associated with a specific motion trajectory. From the perspective of the frequency domain, the narrowband system method proposed by [52] is obviously applicable to the case where the LOS path dominates in the wideband mmWave channel environment. Therefore, the angular motion model can be used to characterize the angular variations with respect to a particular moving trajectory.

Fig. 9 shows the simulation of the temporal evolution of the mobile mmWave channel by using the above angle motion model. Because of the mobility of the terminal in the downlink transmission, the AOD of the BS is to a great extent dependent on the CSI information fed back by the terminal. As such, we mainly simulated the variants of the impinging angle at the receiver. According to the angular motion model, the angle variants of $\Delta\phi$ and $\Delta\theta$ can be obtained according to the MS's

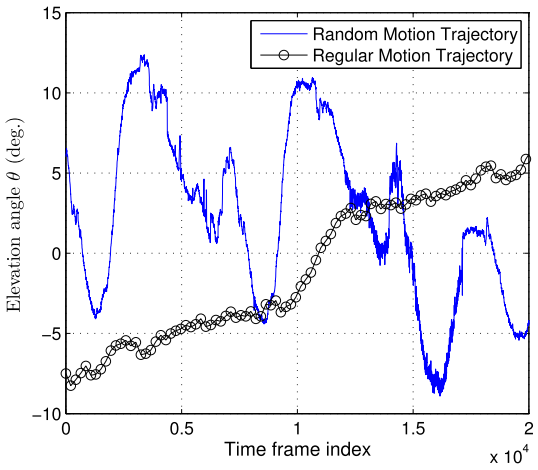


Fig. 9. Examples of actual angle variations using the proposed motion tracjectory model in sparse mmWave channels.

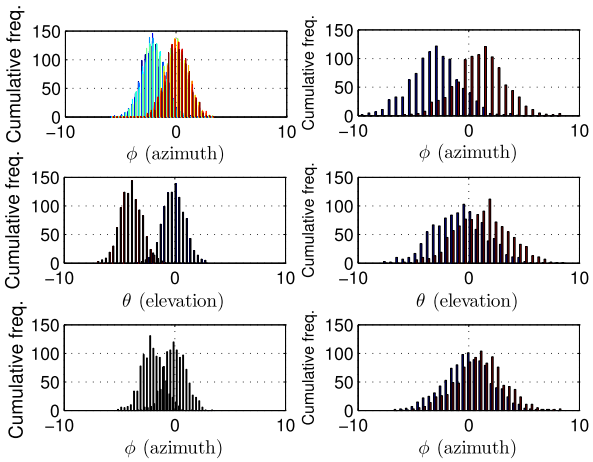


Fig. 10. Histogram of *a posteriori* pdf of the discrete Bayesian beamformer (left part) with different motion trajectory restraints and histogram of *a posteriori* pdf without a priori constraint (right part), given SNR = 15 dB.

azimuth and elevation velocities v_{az} and v_{el}. In this study, we assume that the angle-correlated noise statistics are fixed, and the random angular variants can be modeled as GM distributed random variables, according to (19). Fig. 9 illustrates a simulation of three angular motion models associated with the motion trajectory of the MS, where the BS is set at the center of the cell [52]. We consider three typical motion trajectories: (1) the MS either moves away the BS or close to it with a fixed direction. That is, the main change occurs at the elevation angle θ, whereas the azimuth ϕ is relatively stable or varies in a small range; (2) the MS is moving along a ring that the BS is located at the origin of the ring; the elevation angle θ would keep unchanged at this time, whereas the main change occurs at the azimuth angle ϕ; (3) the MS randomly moves, resulting in random variant of ϕ and θ simultaneously.

Based on the aforementioned motion model, we next use the *a priori* information to evaluate the estimated performance of the *a posteriori* pdf. Without loss of generality, we assume that all motion trajectories start at the same position. The starting distance between MS and BS is 50 m, and the initial values of θ and ϕ are set to be 0° and 90°, respectively. The set motion trajectory moves back and forth over a range of motion of ± 5 m with certain absolute speed v. We assume that all MS have the same speed of motion. Because the spatial information θ and ϕ are mutually uncorrelated, we can independently evaluate the estimation performance of θ and ϕ. Fig. 10 shows the *a posteriori* pdf estimation performance for the unknown parameters θ and ϕ, considering an approximation of a GM distribution with two-component mixed model. The GM model is thereby considered to be given by the product: $p(\boldsymbol{\theta}) = (1-\lambda)g_1(\boldsymbol{\theta}) + \lambda g_2(\boldsymbol{\theta})$, where $\mu_1 = 0$, $\mu_2 = 3$, and Σ is set to be 0.3. The estimates of Gaussian parameters are obtained by using an iterative EM algorithm. Note that the estimation of the GM distribution is performed at present time frame n and future time intervals, where the estimated *a posteriori* pdf is approximately calculated by 2,000 snapshots. During the data transmission epoch, it implies that 256 OFDM symbols are sufficient to provide guarantee for the Bayesian interference in a finite statistical period. Compared to the *a posteriori* pdf estimation without *a priori* knowledge on the

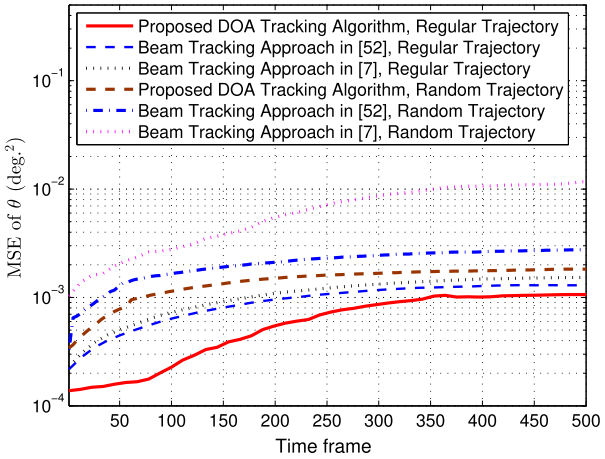


Fig. 11. Average tracking errors yielded by different methods.

right part, it is also noticeable that the proposed approach has better estimation performance. Obviously, Fig. 10 confirms that the trajectory-constrained approach, i.e., the empirical Bayesian estimation, has a high concentration of estimates of *a posteriori* pdf at the actual DOA region. When the *a priori* information is known, the estimation of *a posteriori* pdf in terms of the moving trajectory constraint is closer to the true value when the variance is smaller, and it can yield a higher accuracy and with lower complexity than the estimate in the full probability space.

Finally, we validate the performance of beam tracking algorithms in more realistic situations, while sufficient training is assumed at the initial stage. For a fair comparison, we set an identical feedback cycle. As shown in Fig. 11, the proposed method evidently yields better tracking performance than the angle tracking strategies developed in [52] and [7], which apply the extended Kalman filter (EKF) to track the channel with the aid of beamswitching at the anology beamformers at the transmitter and receiver. It should be noted that beam tracking algorithm provides a higher degree of DOA error suppression. An intuitive explanation for this phenomenon is that, only when the nonlinearity of the solution problem is mild, the Kalman filter can perform well. However, the actual

dynamics of DOA and the property of ADC quantization are nonlinear, as described in [23], [43]. From Fig. 11, we observe that, as the tracking time increases, the DOA estimation becomes less accurate, which, in turn, leads to a larger estimation bias. It should further note that although the accumulated DOA estimation error will severely degrade system performance, the performance penalty is negligible via a limited feedback link, by which the detected abrupt channel variation is fed back [53]. By doing so, the tracking procedure remains robust to enable a continuous beam tracking and to update the steering directions directly within each feedback cycle.

VII. Conclusion

In this paper, we proposed a Bayesian beamforming approach for dealing with the DOA uncertainty in a mobile mmWave system. Because *a priori* information of the motion trajectory dependent DOA is well exploited, the proposed Bayesian beamformer is robust to uncertain DOA estimation and is suitable even when the DOA is completely unknown. To alleviate the computational complexity, we presented a motion trajectory-constrained priori probability approximation method, which implies that a directional estimate within a specific spatial region can be achieved as close as to the true DOA with high probability. The performance of the proposed method was evaluated by simulation, which showed that the proposed data-driven method improves over prior works and enables directional beam tracking with significantly less amount of feedback overhead, and the proposed Bayesian beamforming approach exhibits the robustness against the DOA uncertainty. The future work will be an empirical study based on the promising results achieved in this paper and investigating different probability distributions combined with other criteria. From the information-theoretic perspective, it would also be important to apply the concept of beam entropy to the investigation of the beamspace randomness in mobile mmWave communication scenarios.

Acknowledgment

The authors would like to thank the anonymous reviewers for providing valuable comments that helped in improving the quality of this paper.

References

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[2] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[3] A. Alkhateeb, J. Mo, N. González-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[4] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[5] M. Xiao *et al.*, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[6] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[7] D. De Donno, J. Palacios, and J. Widmer, "Millimeter-wave beam training acceleration through low-complexity hybrid transceivers," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3646–3660, Jun. 2017.

[8] X. Zhang, A. F. 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.

[9] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.

[10] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[11] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[12] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[13] F. Gutierrez, S. Agarwal, K. Parrish, and T. S. Rappaport, "On-chip integrated antenna structures in CMOS for 60 GHz WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1367–1378, Oct. 2009.

[14] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 949–973, 2nd Quart., 2016.

[15] F. Khalid and J. Speidel, "Robust hybrid precoding for multiuser MIMO wireless communication systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3353–3363, Jun. 2014.

[16] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[17] P. Kuo, J. Ahn, and A. Mourad, "Adaptive digital precoder codebook resolution for millimeter wave hybrid beamforming," in *Proc. IEEE PIMRC*, Montreal, QC, Canada, Oct. 2017, pp. 1–6.

[18] Y. Liu, X. Fang, M. Xiao, and S. Mumtaz, "Decentralized beam pair selection in multi-beam millimeter-wave networks," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2722–2737, Jun. 2018.

[19] N. Garcia, H. Wymeersch, and D. T. M. Slock, "Optimal precoders for tracking the AoD and AoA of a mmWave path," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5718–5729, Nov. 2018.

[20] J. Mo, P. Schniter, and R. W. Heath, Jr., "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[21] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath, and O. Koymen, "Beamforming tradeoffs for initial UE discovery in millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 543–559, Apr. 2016.

[22] T. Rappaport *et al.*, "Cellular broadband millimeter wave propagation and angle of arrival for adaptive beam steering systems," in *Proc. Radio Wireless Symp. (RWS)*, Santa Clara, CA, USA, Jan. 2012, pp. 151–154.

[23] H. Zhao *et al.*, "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in new york city," in *Proc. IEEE ICC*, Budapest, Hungary, Jun. 2013, pp. 5163–5167.

[24] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin, "Angle domain hybrid precoding and channel tracking for millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6868–6880, Oct. 2017.

[25] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *Proc. IEEE SPAWC*, Sydney, NSW, Australia, Jun. 2015, pp. 146–150.

[26] C. Chen and W. Wu, "Joint AoD, AoA, and channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5806–5820, Jul. 2018.

[27] K. L. Bell, Y. Ephraim, and H. L. Van Trees, "A Bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 386–398, Feb. 2000.

[28] C. J. Lam and A. C. Singer, "Bayesian beamforming for DOA uncertainty: Theory and implementation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4435–4445, Nov. 2006.

[29] M. U. Aminu, M. Codreanu, and M. Juntti, "Variational Bayesian learning for channel estimation and transceiver determination," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2018, pp. 1–9.

[30] S. Malik, J. Benesty, and J. Chen, "A Bayesian framework for blind adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2370–2384, May 2014.

[31] M. U. Aminu, M. Codreanu, and M. Juntti, "Bayesian learning based millimeter-wave sparse channel estimation with hybrid antenna array," in *Proc. IEEE SPAWC*, Kalamata, Greece, Jun. 2018, pp. 1–5.

[32] P. Gerstoft, C. F. Mecklenbräuker, A. Xenaki, and S. Nannuru, "Multisnapshot sparse Bayesian learning for DOA," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1469–1473, Oct. 2016.

[33] K. K. Tiwari, E. Grass, J. S. Thompson, and R. Kraemer, "Beam entropy of 5G cellular millimetre-wave channels," in *Proc. IEEE VTC Fall*, Honolulu, HI, USA, Sep. 2019, pp. 1–5.

[34] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[35] R. Roy and T. Kailath, "ESPRIT: Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, Jul. 1989.

[36] Y. Yang, S. Dang, M. Wen, S. Mumtaz, and M. Guizani, "Mobile millimeter wave channel tracking: A Bayesian beamforming framework against DOA uncertainty," in *Proc. IEEE GLOBECOM*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[37] Y. Niu, C. Gao, Y. Li, L. Su, D. Jin, and A. V. Vasilakos, "Exploiting device-to-device communications in joint scheduling of access and backhaul for mmWave small cells," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2052–2069, May 2015.

[38] Y. Niu *et al.*, "Energy-efficient scheduling for mmWave backhauling of small cells in heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2674–2687, Jun. 2017.

[39] E. Arribas *et al.*, "Optimizing mmWave wireless backhaul scheduling," *IEEE Trans. Mobile Comput.*, vol. 19, no. 10, pp. 2409–2428, Oct. 2020.

[40] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, Mar. 1998.

[41] H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 225–234, Feb. 2002.

[42] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[43] *Study on Channel Model for Frequency Spectrum Above 6 GHz (Release 15)*, document 3GPP TR 36.901 V15.0.0, Jul. 2018.

[44] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[45] H. L. V. Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, 1st ed. Hoboken, NJ, USA: Wiley, 2002.

[46] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[47] F. Bellili, F. Sohrabi, and W. Yu, "Generalized approximate message passing for massive MIMO mmWave channel estimation with Laplacian prior," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3205–3219, May 2019.

[48] S. M. Kay, *Fundamentals of Statistical Processing: Estimation Theory*, vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[49] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.

[50] R. Prasad, C. R. Murthy, and B. Rao, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," *IEEE Trans. Signal Process*, vol. 63, no. 20, pp. 3704–3716, Oct. 2015.

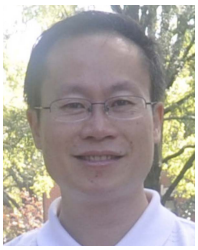
[51] J. A. Bilmes, "A gentle tutorial EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Comput. Sci. Division, UC Berkeley, Berkeley, CA, USA, 1998.

[52] D. Zhu, J. Choi, Q. Cheng, W. Xiao, and R. W. Heath, Jr., "High-resolution angle tracking for mobile wideband millimeter-wave systems with antenna array calibration," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7173–7189, Nov. 2018.

[53] C. Zhang, D. Guo, and P. Fan, "Tracking angles of departure and arrival in a mobile millimeter wave channel," in *Proc. IEEE ICC*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[54] 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.

[55] *NYUSIM Channel Simulator v2.01*. Accessed: Nov. 2019. [Online]. Available: https://wireless.engineering.nyu.edu/nyusim/



Yan Yang (Member, IEEE) received the B.Sc. degree in electronics engineering, and the M.Sc. and D.Sc. degrees in signal processing from the University of Electronic Science and Technology of China, Sichuan University, Institute of Acoustic, Chinese Academy of Science, China, in 1990, 1997, and 2004, respectively. From 2014 to 2015, he was a Visiting Scholar with the Georgia Institute of Technology, Atlanta, USA. He is currently an Associate Professor with the State Key Lab. of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China. His current research interests include wireless communications, signal processing, and artificial intelligence for cognitive wireless communications. He has served as a Reviewer for various journals, including IEEE TRANSACTIONS ON NETWORK, IEEE INTERNET OF THINGS JOURNAL, IEEE WIRELESS COMMUNICATIONS LETTERS, IEEE ACCESS, *IEEE Network Magazine*, *IEEE Wireless Communications Magazine* and several more. He was a recipient of the Best Paper Awards from the IEEE ComComAp'2019. He received the First Research Award from the Science and Technology of China Railways Society in 2007 and 2014, respectively. He is an active participant in the Working Party 5A (WP 5A), the International Telecommunication Union (ITU), and is a Technical Specialist for the research item 1.11 of the Resolution 236 World Radiocommunication Conference (WRC-15).

Shuping Dang (Member, IEEE) received the B.Eng. degree (Hons.) in electrical and electronic engineering from the University of Manchester, the B.Eng. degree in electrical engineering and automation from Beijing Jiaotong University, in 2014, via a joint '2+2' dual-degree program, and the D.Phil. degree in engineering science from the University of Oxford, in 2018. He joined in the R&D Center, Huanan Communication Company, Ltd., after graduating from the University of Oxford and is currently working as a Postdoctoral Fellow with the Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST). He was a co-recipient of the 'best paper' award for work presented at 2019 19th IEEE International Conference on Communication Technology. He serves as a reviewer for a number of key journals in communications and information science, including IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS LETTERS, IEEE COMMUNICATIONS LETTERS, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is recognized as the Exemplary Reviewer of IEEE COMMUNICATIONS LETTERS in 2019. His current research interests include novel modulation schemes, cooperative communications, terahertz communications, and 6G wireless network design.

Miaowen Wen (Senior Member, IEEE) received the Ph.D. degree from Peking University, Beijing, China, in 2014. From 2012 to 2013, he was a Visiting Student Research Collaborator with Princeton University, Princeton, NJ, USA. He is currently an Associate Professor with the South China University of Technology, Guangzhou, China, and a Hong Kong Scholar with The University of Hong Kong, Hong Kong. He has published a Springer book entitled *Index Modulation for 5G Wireless Communications* and more than 100 journal articles. His research interests include a variety of topics in the areas of wireless and molecular communications.

Dr. Wen was a recipient of four Best Paper Awards from the IEEE ITST'2012, the IEEE ITSC'2014, the IEEE ICNC'2016, and IEEE ICCT'2019. He has served on the Editorial Boards of IEEE ACCESS, and *EURASIP Journal on Wireless Communications and Networking*, and a Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Special Issue on Spatial Modulation for Emerging Wireless Systems) and for IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (Special Issue on Index Modulation for Future Wireless Networks: A Signal Processing Perspective). He is currently serving as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE COMMUNICATIONS LETTERS, and *Physical Communication* (Elsevier).

Shahid Mumtaz (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical and electronic engineering from the Blekinge Institute of Technology (BTH) Karlskrona, Sweden, and University of Aveiro, Portugal, in 2006 and 2011, respectively. He has more than ten years of wireless industry experience and is currently working as a Senior Research Scientist with the Instituto de Telecomunicações, Aveiro, Portugal, ARIES Research Center, Universidad Antonio de Nebrija, Madrid, Spain. Prior to his current position, he worked as a Research Intern at Ericsson and Huawei Research Labs in 2005 at Karlskrona, Sweden. He has more than 150 publications in international conferences, journal papers, and book chapters. He was awarded an 'Alain Bensoussan' fellowship by ERCIM to pursue research in communication networks for one year at the VTT Technical Research Centre of Finland in 2012. He was nominated as a Vice Chair for IEEE new standardization on P1932.1: Standard for Licensed/Unlicensed Spectrum Interoperability in Wireless Mobile Networks. He is also actively involved in 3GPP standardization on LTE release 12 onwards, along with major manufacturers. He is an ACM distinguished speaker.

Mohsen Guizani (Fellow, IEEE) received the B.S. (Hons.) and M.S. degrees in electrical engineering, and the M.S. and Ph.D. degrees in computer engineering from Syracuse University, Syracuse, NY, USA, in 1984, 1986, 1987, and 1990, respectively. He is currently a Professor and the ECE Department Chair with the University of Idaho, USA. Previously, he served as the Associate Vice President for Graduate Studies, Qatar University, a Chair for the Computer Science Department, Western Michigan University, and a Chair for the Computer Science Department, University of West Florida. He also served in academic positions at the University of Missouri-Kansas City, University of Colorado-Boulder, and Syracuse University. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He is currently the Editor-in-Chief of the IEEE Network Magazine, serves on the editorial boards for several international technical journals and the Founder and the Editor-in-Chief of *Wireless Communications and Mobile Computing* journal (Wiley). He is the author of nine books and more than 500 publications in refereed journals and conferences. He guest edited a number of special issues in IEEE journals and magazines. He also served as a member, Chair, and General Chair for a number of international conferences. He received three teaching awards and four research awards throughput his career. He received the 2017 IEEE Communications Society Recognition Award for his contribution to outstanding research in Wireless Communications. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the Chair of the TAOS Technical Committee. He served as the IEEE Computer Society Distinguished Speaker from 2003 to 2005. He is a Senior Member of ACM.