Millimeter Wave Channel Estimation via Exploiting Joint Sparse and Low-Rank Structures

Xingjian Li, Jun Fang, *Member, IEEE*, Hongbin Li, *Senior Member, IEEE*, and Pu Wang

***Abstract*— We consider the problem of channel estimation for millimeter wave (mmWave) systems, where, to minimize the hardware complexity and power consumption, an analog transmit beamforming and receive combining structure with only one radio frequency chain at the base station and mobile station is employed. Most existing works for mmWave channel estimation exploit sparse scattering characteristics of the channel. In addition to sparsity, mmWave channels may exhibit angular spreads over the angle of arrival, angle of departure, and elevation domains. In this paper, we show that angular spreads give rise to a useful low-rank structure that, along with the sparsity, can be simultaneously utilized to reduce the sample complexity, i.e., the number of samples needed to successfully recover the mmWave channel. Specifically, to effectively leverage the joint sparse and low-rank structure, we develop a two-stage compressed sensing method for mmWave channel estimation, where the sparse and low-rank properties are respectively utilized in two consecutive stages, namely, a matrix completion stage and a sparse recovery stage. Our theoretical analysis reveals that the proposed two-stage scheme can achieve a lower sample complexity than a conventional compressed sensing method that exploits only the sparse structure of the mmWave channel. Simulation results are provided to corroborate our theoretical results and to show the superiority of the proposed two-stage method.**

***Index Terms*— mmWave channel estimation, angular spread, jointly sparse and low-rank, compressed sensing.**

I. Introduction

MILLIMETER WAVE (mmWave) communication is a promising technology for future 5G cellular networks [1]–[3]. It has the potential to offer gigabits-per-second communication data rates by exploiting the large bandwidth available at mmWave frequencies. However, a key challenge for mmWave communication is that signals incur a much more significant path loss over the mmWave frequency bands as compared with the path attenuation over the lower frequency bands [4]. To compensate for the significant path loss, large antenna arrays should be used at both the base station (BS) and the mobile station (MS) to provide sufficient beamforming gain for mmWave communications [5].

Although directional beamforming helps overcome the path loss issue, it also complicates the mmWave communication system design. Due to the narrow beam of the antenna array, communication between the transmitter and the receiver is possible only when the transmitter's and receiver's beams are well-aligned, i.e. the beam directions are pointing towards each other. Therefore beamforming training is required to search for the best beamformer-combiner pair that gives the highest channel gain. One method is to exhaustively search for all possible beam pairs to identify the best beam alignment. Nevertheless, this exhaustive search may lead to a prohibitively long training process, particularly when the number of antennas at the BS and MS is large. To address this issue, an adaptive beam alignment algorithm was proposed in [6], where a hierarchical multi-resolution beamforming codebook was employed to avoid the costly exhaustive sampling of all pairs of transmit and receiver beams. Nevertheless, this adaptive beam alignment requires feedback from the receiver to the transmitter. In practice, this feedback involves only a few bits' transmission and thus may work at very low signal-to-noise ratios. It can also be operated in a different frequency band such that the information transmission is possible before channel estimation. Recently, a novel beam steering scheme called as "Agile-Link" [7] was proposed to find the correct beam alignment without scanning the space. The main idea of the Agile-Link is to harsh the beam directions using a few carefully chosen hash functions, and steer the antenna array to beam along multiple directions simultaneously.

Unlike beam scanning techniques whose objective is to find the best beam pair, another approach is to directly estimate the mmWave channel or its associated parameters, e.g. angles of arrival/departure, e.g. [8]–[17]. In particular, by exploiting the sparse scattering nature of mmWave channels, mmWave channel estimation can be formulated as a sparse signal recovery problem [10]–[16], and it has been shown that substantial reduction in training overhead can be achieved. In [7], the estimation of directions of the paths (i.e. the angles of arrival) was also expressed as a sparse signal recovery problem. Nevertheless, unlike [10]–[16], the Agile-Link algorithm [7] uses only magnitudes of the measurements, and developed a simple voting-based scheme to recover the directions of

Manuscript received May 9, 2017; revised September 23, 2017; accepted November 15, 2017. Date of publication November 28, 2017; date of current version February 9, 2018. This work was supported in part by the National Science Foundation of China under Grant 61522104 and in part by the National Science Foundation under Grant ECCS-1408182 and Grant ECCS-1609393. The associate editor coordinating the review of this paper and approving it for publication was M. Di Renzo. *(Corresponding author: Jun Fang.)*

X. Li and J. Fang are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: junfang@uestc.edu.cn).

H. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: hongbin.li@stevens.edu).

P. Wang is with Mitsubushi Electric Research Laboratories, Cambridge, MA 02139 USA (e-mail: pwang@merl.com).

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

Digital Object Identifier 10.1109/TWC.2017.2776108

the paths. Besides the compressed sensing techniques, low-rank tensor factorization methods [17], [18] were recently proposed to exploit the low-rank structure of mmWave channels, and have been shown to outperform the compressed sensing-based methods in terms of both estimation accuracy and computational complexity.

In addition to the sparse scattering characteristic, several real-world measurements in dense-urban propagation environments (e.g. [19]–[22]) reveal that mmWave channels spread in the form of clusters of paths over the angular domains including the angle of arrival (AoA), angle of departure (AoD), and elevation. In [21] and [22], real-world channel measurements at 28 and 73 GHz in New York city were reported, in which the angular spread has been explicitly studied in terms of the root mean-squared (rms) beamspread in the different angular (AoA, AoD, and elevation) dimensions. Specifically, the measured AoA spreads (in terms of rms) are 15.5° and 15.4°, respectively, for the two carrier frequencies, while the measured AoD spreads (in terms of rms) are 10.2° and 10.5°, respectively. As demonstrated in [23], the angular spreads give rise to a structured sparsity pattern that can be exploited to improve the mmWave channel estimation performance.

In this paper, we further show that, in the presence of angular spreads, the mmWave channel exhibits a joint sparse and low-rank structure in which the rank is far less than the sparsity level of the channel. To better utilize the joint structure, we propose a two-stage compressed sensing scheme, where a low-rank matrix completion stage is first performed and then followed by a compressed sensing stage to recover the mmWave channel. Our analysis reveals that the number of measurements required for exact channel recovery is about $O(pL^2)$ for the proposed two-stage method, where L represents the number of scattering clusters and p is a quantity that measures the maximum angular spread among all scattering clusters. While a conventional convex relaxation-based compressed sensing method that exploits only the sparsity of mmWave channels requires a number of measurements of $O(p^2L)$. Thus the proposed two-stage compressed sensing method achieves a lower sample complexity than the conventional compressed sensing method when $L < p$, which is very likely to hold in dense-urban propagation environments.

The rest of the paper is organized as follows. The system model and the problem formulation are discussed in Section II. In Section III, we introduce a geometric mmWave channel model with angular spreads and show that the mmWave channel exhibits a joint sparse and low-rank structure. A two-stage compressed sensing method is developed in Section IV, along with a theoretical analysis for the two-stage method. Simulation results are provided in Section V, followed by concluding remarks in Section VI.

II. System Model and Prior Work

Consider a point-to-point uplink mmWave MIMO system consisting of N_{BS} antennas at the BS and N_{MS} antennas at the MS. Since the radio frequency (RF) chains are costly and power-consuming at mmWave frequency bands, to minimize the hardware complexity and power consumption, we focus on an analog transmit beamforming and receive combining

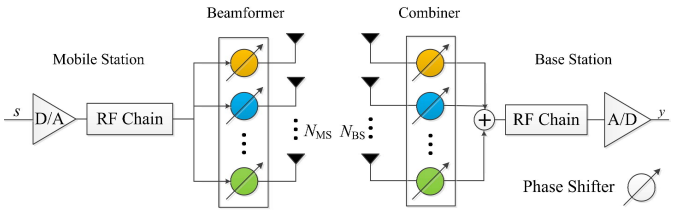


Fig. 1. A block diagram of the analog transmit beamforming and receive combining structure.

structure (see Fig. 1) where only one RF chain is employed at the BS and MS. In this structure, transmit beamforming and receive combining are implemented in the analog domain using digitally controlled phase shifters. On the other hand, although our proposed method may be extended to the hybrid beamforming structure with multiple RF chains, considering the case of one RF chain can simply our exposition and facilitate our algorithmic development.

At time instant t, the transmitter employs a beamforming vector $\boldsymbol{f}(t) \in \mathbb{C}^{N_{\mathrm{MS}}}$ to transmit a symbol $s(t)$, and at the receiver, the received signals on all antennas are combined with a receive combining vector $\boldsymbol{z}(t) \in \mathbb{C}^{N_{\mathrm{BS}}}$. The combined signal at the receiver can therefore be expressed as (see, e.g. [6])

$$y(t) = \boldsymbol{z}^H(t)\boldsymbol{H}\boldsymbol{f}(t)s(t) + w(t) \quad \forall\ t = 1, \ldots, T \tag{1}$$

where $\boldsymbol{H} \in \mathbb{C}^{N_{\mathrm{BS}} \times N_{\mathrm{MS}}}$ is the channel matrix, and $w(t)$ denotes the additive Gaussian noise with zero mean and variance σ^2. Without loss of generality, we set $s(t) = 1$ during the training phase. Since the precoder and combiner are implemented by analog phase shifters, entries of $\boldsymbol{z}(t)$ and $\boldsymbol{f}(t)$ have constant modulus.

We see that in mmWave systems, the receiver cannot directly observe $\boldsymbol{H}$, rather it observes a noisy version of $\boldsymbol{z}^H\boldsymbol{H}\boldsymbol{f}$. This is also referred to as the channel subspace sampling limitation [6], [13], which makes channel estimation a challenging problem. By exploiting the sparse scattering nature of mmWave channels, the channel estimation problem can be formulated as a sparse signal recovery problem (e.g. [6], [13]). Specifically, the mmWave channel is usually characterized by a geometric channel model [11]

$$\boldsymbol{H} = \sum_{l=1}^{\tilde{L}} \alpha_l \boldsymbol{a}_{\mathrm{BS}}(\theta_l)\boldsymbol{a}_{\mathrm{MS}}^H(\phi_l) \tag{2}$$

where $\tilde{L}$ is the number of paths, α_l is the complex gain associated with the lth path, $\theta_l \in [0, 2\pi]$ and $\phi_l \in [0, 2\pi]$ are the associated azimuth AoA and azimuth AoD respectively, and $\boldsymbol{a}_{\mathrm{BS}} \in \mathbb{C}^{N_{\mathrm{BS}}}$ ($\boldsymbol{a}_{\mathrm{MS}} \in \mathbb{C}^{N_{\mathrm{MS}}}$) is the array response vector associated with the BS (MS). Suppose a uniform linear array (ULA) antenna array is used. Then the steering vectors at the BS and the MS can be written as

$$\boldsymbol{a}_{\mathrm{BS}}(\theta_l) = \frac{1}{\sqrt{N_{\mathrm{BS}}}}\left[1, e^{j\frac{2\pi}{\lambda}d\sin(\theta_l)}, \ldots, e^{j(N_{\mathrm{BS}}-1)\frac{2\pi}{\lambda}d\sin(\theta_l)}\right]^T$$

$$\boldsymbol{a}_{\mathrm{MS}}(\phi_l) = \frac{1}{\sqrt{N_{\mathrm{MS}}}}\left[1, e^{j\frac{2\pi}{\lambda}d\sin(\phi_l)}, \ldots, e^{j(N_{\mathrm{MS}}-1)\frac{2\pi}{\lambda}d\sin(\phi_l)}\right]^T$$

where λ is the signal wavelength, and d is the distance between neighboring antenna elements. To formulate the channel estimation as a sparse signal recovery problem, we first express the channel as a beam space MIMO representation as follows [13]

$$\boldsymbol{H} = \boldsymbol{A}_{\mathrm{BS}} \boldsymbol{H}_v \boldsymbol{A}_{\mathrm{MS}}^H \tag{3}$$

where $\boldsymbol{A}_{\mathrm{BS}} \triangleq [\boldsymbol{a}_{\mathrm{BS}}(\psi_1), \ldots, \boldsymbol{a}_{\mathrm{BS}}(\psi_{N_1})]$ is an overcomplete matrix ($N_1 \geq N_{\mathrm{BS}}$) with each column a steering vector parameterized by a pre-discretized AoA, $\boldsymbol{A}_{\mathrm{MS}} \triangleq [\boldsymbol{a}_{\mathrm{MS}}(\omega_1), \ldots, \boldsymbol{a}_{\mathrm{MS}}(\omega_{N_2})]$ is an overcomplete matrix (i.e. $N_2 \geq N_{\mathrm{MS}}$) with each column a steering vector parameterized by a pre-discretized AoD, and $\boldsymbol{H}_v \in \mathbb{C}^{N_1 \times N_2}$ is a sparse matrix with $\tilde{L}$ non-zero entries corresponding to the channel path gains $\{\alpha_l\}$. Here we assume that the true AoA and AoD parameters lie on the discretized grids. In practice, the true parameters do not necessarily lie on the discretized grid, which is referred to as grid or basis mismatch. In the presence of grid mismatch, the number of nonzero entries in the beam space channel $\boldsymbol{H}_v$ will not exactly equal $\tilde{L}$, instead, the number of nonzero entries will become larger due to the power leakage caused by grid mismatch. Note that although a variety of super-resolution or off-grid compressed sensing methods, e.g. [24]–[26], were proposed to overcome the grid mismatch issue, these methods incur a higher computational complexity as they usually involve an iterative procedure which alternatively refines the dictionary parameters and the sparse signal.

Substituting (3) into (1), we have

$$\begin{aligned} y(t) &= \boldsymbol{z}^H(t) \boldsymbol{A}_{\mathrm{BS}} \boldsymbol{H}_v \boldsymbol{A}_{\mathrm{MS}}^H \boldsymbol{f}(t) + w(t) \\ &= \left[(\boldsymbol{A}_{\mathrm{MS}}^H \boldsymbol{f}(t))^T \otimes (\boldsymbol{z}(t)^H \boldsymbol{A}_{\mathrm{BS}}) \right] \boldsymbol{h} + w(t) \\ &= (\boldsymbol{f}(t)^T \otimes \boldsymbol{z}(t)^H)(\boldsymbol{A}_{\mathrm{MS}}^* \otimes \boldsymbol{A}_{\mathrm{BS}}) \boldsymbol{h} + w(t) \end{aligned} \tag{4}$$

where $\otimes$ denotes the Kronecker product, $()^*$ represents the complex conjugate, and $\boldsymbol{h} \triangleq \mathrm{vec}(\boldsymbol{H}_v)$. Collecting all measurements $\{y(t)\}$ and stacking them into a vector $\boldsymbol{y} \triangleq [y_1 \ \ldots \ y_T]^T$, we arrive at

$$\begin{aligned} \boldsymbol{y} &= \begin{bmatrix} (\boldsymbol{f}(1)^T \otimes \boldsymbol{z}(1)^H) \\ \vdots \\ (\boldsymbol{f}(T)^T \otimes \boldsymbol{z}(T)^H) \end{bmatrix} (\boldsymbol{A}_{\mathrm{MS}}^* \otimes \boldsymbol{A}_{\mathrm{BS}}) \boldsymbol{h} + \boldsymbol{w} \\ &\triangleq \boldsymbol{\Psi} \boldsymbol{h} + \boldsymbol{w} \end{aligned} \tag{5}$$

Since $\boldsymbol{h}$ is sparse, channel estimation now becomes a sparse signal recovery problem. To estimate $\boldsymbol{h}$, we can resort to the fast iterative shrinkage-thresholding algorithm (FISTA) [27] which involves solving

$$\begin{aligned} &\min \ \|\boldsymbol{h}\|_1 \\ &\text{s.t.} \ \|\boldsymbol{y} - \boldsymbol{\Psi} \boldsymbol{h}\|_2 \leq \varepsilon \end{aligned} \tag{6}$$

where ε is an error tolerance parameter related to noise statistics. Compressed sensing theory tells that, for the noiseless case, we can perfectly recover a high-dimensional sparse signal $\boldsymbol{h}$ from a much lower dimensional linear measurement vector $\boldsymbol{y}$. Thus the compressed sensing-based method has the potential to achieve a substantial training overhead reduction.

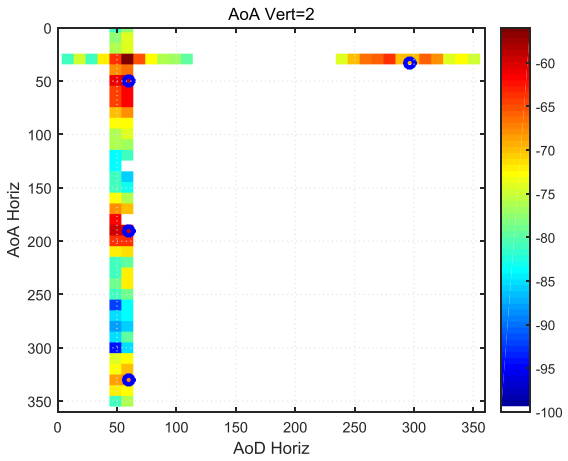


Fig. 2. Rx power angular profile measured at a typical TX-RX location pair at 28 GHz [22].

III. Channel Model With Angular Spreads

In addition to sparsity, mmWave channels also take a form of angular spreads over the AoA, AoD, and elevation domains [21], [22]. The angular spreads are a result of scattering clusters, where each cluster may contribute with multiple rays/paths. It was shown a structured sparsity pattern arises in the presence of angular spreads [23]. In this paper, we further show that, due to the spatial correlation and the unsymmetric angular spreads over different domains, mmWave channels may exhibit a meaningful low-rank structure that can be utilized to improve the sample complexity.

Such a low-rank structure can be observed from recent real-world mmWave channel measurements, e.g. [22]. Specifically, Fig. 2 plots the RX power angular profile measured for a typical TX-RX location pair at 28GHz considered in [22, Fig. 3]. We see that the RX power angular profile consists of several correlated horizontal and vertical strips, and the angular spread over the AoA (or AoD) domain is caused by rays from a common AoD (AoA) or several closely-spaced AoDs (AoAs). This highly structured pattern gives rise to a low-rank (or approximately low-rank) structure in which the rank of the channel matrix is far less than the sparsity level of the channel matrix, i.e. the number of dominant entries in the beam space (angular) domain. To illustrate this, consider a simple case where the angular spread over the AoA domain is a result of rays from a common AoD, in which case the channel can be expressed as

$$\boldsymbol{H} = \left(\sum_{i=1}^{I} \alpha_i \boldsymbol{a}_{\mathrm{BS}}(\theta - \vartheta_i) \right) \boldsymbol{a}_{\mathrm{MS}}^H(\phi) \tag{7}$$

Clearly, the channel matrix $\boldsymbol{H}$ has rank one, while its beam space (i.e. angular) representation has more than one dominant coefficients. If the angular spread over the AoA domain is caused by two closely-spaced AoDs, the channel can be expressed as

$$\begin{aligned} \boldsymbol{H} = &\left(\sum_{i=1}^{I} \alpha_i \boldsymbol{a}_{\mathrm{BS}}(\theta - \vartheta_i) \right) \boldsymbol{a}_{\mathrm{MS}}^H(\phi - \varphi_1) \\ &+ \left(\sum_{i=1}^{I} \alpha_i' \boldsymbol{a}_{\mathrm{BS}}(\theta - \vartheta_i) \right) \boldsymbol{a}_{\mathrm{MS}}^H(\phi - \varphi_2) \end{aligned} \tag{8}$$

which can be further simplified as

$$\boldsymbol{H} = \left(\sum_{i=1}^{I} \alpha_i \boldsymbol{a}_{\mathrm{BS}}(\theta - \vartheta_i)\right)\left(\sum_{j=1}^{2} \boldsymbol{a}_{\mathrm{MS}}^H(\phi - \varphi_j)\right) \tag{9}$$

by considering the fact that the terms $\sum_{i=1}^{I} \alpha_i \boldsymbol{a}_{\mathrm{BS}}(\theta - \vartheta_i)$ and $\sum_{i=1}^{I} \alpha_i' \boldsymbol{a}_{\mathrm{BS}}(\theta - \vartheta_i)$ are highly correlated. Such a spatial correlation can be observed from Fig. 2, in which the two parallel vertical strips associated with closely-spaced AoDs have similar power angular patterns.

As a generalization, we adopt the following geometric channel model to characterize the mmWave channel

$$\boldsymbol{H} = \sum_{l=1}^{L}\left(\sum_{i=1}^{I} \alpha_{l,i} \boldsymbol{a}_{\mathrm{BS}}(\theta_l - \vartheta_{l,i})\right)\left(\sum_{j=1}^{J} \beta_{l,j} \boldsymbol{a}_{\mathrm{MS}}^H(\phi_l - \varphi_{l,j})\right) \tag{10}$$

where L denotes the number of clusters, θ_l and ϕ_l represent the mean AoA/AoD associated with each cluster, and $\vartheta_{l,i}$ and $\varphi_{l,j}$ denote the relative AoA and AoD shift from the mean angle. Note that a different clustered channel model was introduced in [28], in which the channel is expressed as

$$\boldsymbol{H} = \sum_{l=1}^{L}\sum_{r_l=1}^{R_l} \alpha_{r_l} \boldsymbol{a}_{\mathrm{BS}}(\theta_l - \vartheta_{r_l}) \boldsymbol{a}_{\mathrm{MS}}^H(\phi_l - \varphi_{r_l}) \tag{11}$$

This clustered channel model [28] exploits the fact that each cluster consists of multiple paths. Nevertheless, it neglects the low-rank structure arising from path clusters. It can be easily checked that the channel matrix given in (11) has a rank identical to its sparsity level (if we ignore the grid mismatch issue). In this case, the channel matrix may still possess a low-rank structure. But, compared with the sparsity, the low-rank structure does not provide extra structural information, and hence cannot be utilized to improve the sample complexity.

Similar to (3), we express the channel (10) as a beam space MIMO representation

$$\begin{aligned}\boldsymbol{H} &= \sum_{l=1}^{L} \boldsymbol{A}_{\mathrm{BS}} \boldsymbol{\alpha}_l \boldsymbol{\beta}_l^T \boldsymbol{A}_{\mathrm{MS}}^H = \boldsymbol{A}_{\mathrm{BS}}\left(\sum_{l=1}^{L} \boldsymbol{\alpha}_l \boldsymbol{\beta}_l^T\right)\boldsymbol{A}_{\mathrm{MS}}^H \\ &\triangleq \boldsymbol{A}_{\mathrm{BS}} \boldsymbol{H}_v \boldsymbol{A}_{\mathrm{MS}}^H \end{aligned} \tag{12}$$

where $\boldsymbol{\alpha}_l \in \mathbb{C}^{N_1}$ and $\boldsymbol{\beta}_l \in \mathbb{C}^{N_2}$ represent the virtual representation over the AoA and AoD domain, respectively. Since the angular spread occupies only a small portion of the whole angular domain, both $\boldsymbol{\alpha}_l$ and $\boldsymbol{\beta}_l$ are sparse vectors with only a few nonzero entries concentrated around the mean AoA and AoD associated with the lth cluster. Hence the virtual beam space channel $\boldsymbol{H}_v$ is a sum of L sparse matrices. Suppose any sparse vector in $\{\boldsymbol{\alpha}_l, \boldsymbol{\beta}_l\}_l$ contains at most p nonzero entries. As a result, $\boldsymbol{H}_v$ is a sparse matrix with at most p^2L nonzero entries. Also, $\boldsymbol{H}_v$ has at most pL nonzero columns and at most pL nonzero rows. Due to the limited scattering nature and small angular spreads, we usually have $pL \ll \min\{N_1, N_2\}$. Meanwhile, $\boldsymbol{H}_v$ has a low rank structure with $\mathrm{rank}(\boldsymbol{H}_v) = L$. Thus the virtual beam space channel has a simultaneously sparse and low-rank structure.

Our objective is to estimate/recover the joint sparse and low-rank virtual channel $\boldsymbol{H}_v$ using as few measurements as possible. Estimation of low-rank matrices or sparse matrices from compressed linear measurements has been studied extensively in various settings, e.g. [29]–[33]. However, there is much less research for cases where the matrix of interest is characterized by two structures simultaneously. In particular, how to simultaneously exploit both structures to improve the sample complexity is of most concern. In [34], an efficient two-stage scheme was developed for recovering a sparse, rank-one and positive semi-definite matrix in the context of compressive phase retrieval, and it was shown that the proposed two-stage scheme can achieve a near-optimal sample complexity and enjoys nice robustness guarantees. In the following section, the two-stage scheme is extended to a more general scenario where the mmWave channel to be estimated is not necessarily a rank-one positive semi-definite matrix. We show that an reduced sample complexity can be obtained as compared with simply exploiting the sparsity of the mmWave channel.

IV. Two-Stage Compressed Sensing Scheme

Before proceeding, we revisit the measurement collection model (1) and reformulate this measurement process as a low-rank matrix sampling process. Assume $\boldsymbol{z}(t)$ and $\boldsymbol{f}(t)$ are randomly chosen from pre-determined beamforming/combining codebooks $\mathcal{Z}$ and $\mathcal{F}$, respectively, where the cardinality of the two sets are $|\mathcal{Z}| = N_Z$ and $|\mathcal{F}| = N_F$ and no beam pair $\{\boldsymbol{z}(t), \boldsymbol{f}(t)\}$ is reused during the sampling process. Let $\boldsymbol{Z} \in \mathbb{C}^{N_{\mathrm{BS}} \times N_Z}$ and $\boldsymbol{F} \in \mathbb{C}^{N_{\mathrm{MS}} \times N_F}$ be matrices constructed by all vectors in $\mathcal{Z}$ and $\mathcal{F}$, respectively. Then the observation model (1) can be expressed as sampling from a low-rank matrix:

$$\boldsymbol{Y}_{ij} = (\boldsymbol{Z}^H \boldsymbol{H} \boldsymbol{F})_{ij} \quad (i, j) \in \Omega \tag{13}$$

where $\boldsymbol{Y} \triangleq \boldsymbol{Z}^H \boldsymbol{H} \boldsymbol{F}$ is a low rank matrix with $\mathrm{rank}(\boldsymbol{Y}) = L$, $\boldsymbol{Y}_{ij}$ denotes the (i, j)th entry of $\boldsymbol{Y}$, and Ω denotes a set indicating which entries of $\boldsymbol{Y}$ are observed. We have $|\Omega| = T$. Also, here the observation noise is temporarily ignored to simplify our subsequent analysis.

Suppose $\boldsymbol{Z}$ and $\boldsymbol{F}$ are full-rank square matrices, i.e. $N_Z = N_{\mathrm{BS}}$ and $N_F = N_{\mathrm{MS}}$. Then the problem of estimating $\boldsymbol{H}$ is equivalent to a low-rank matrix completion problem. Specifically, we first recover the low-rank matrix $\boldsymbol{Y}$ via a nuclear-norm minimization [31]:

$$\begin{aligned} &\min_{\hat{\boldsymbol{Y}}} \ \|\hat{\boldsymbol{Y}}\|_* \\ &\text{s.t. } \hat{\boldsymbol{Y}}_{ij} = \boldsymbol{Y}_{ij} \quad \forall\ (i, j) \in \Omega \end{aligned} \tag{14}$$

After recovering $\boldsymbol{Y}$, the channel $\boldsymbol{H}$ can be estimated as

$$\hat{\boldsymbol{H}} = (\boldsymbol{Z}^H)^{-1} \hat{\boldsymbol{Y}} \boldsymbol{F}^{-1} \tag{15}$$

Nevertheless, according to the matrix completion theory [31], the number of measurements has to satisfy

$$T \geq C\tilde{n}^{5/4} L \log(\tilde{n}) \tag{16}$$

in order to stably reconstruct $\boldsymbol{Y}$ of rank at most L with probability at least $1 - c\tilde{n}^{-3}$, where $\tilde{n} \triangleq \max\{N_{\mathrm{BS}}, N_{\mathrm{MS}}\}$,

and the constants $C, c > 0$ are universal. Hence for the low-rank matrix completion approach, the required number of measurements is of order $O(L\max\{N_{\rm BS}, N_{\rm MS}\}^{5/4})$, which increases approximately linearly with the number of antennas employed at the BS or MS, whichever is greater. We see that the low-rank matrix completion scheme ignores the sparse structure inherent in mmWave channels, and thus can only achieve a sub-optimal sample complexity. To obtain a lower sample complexity, we introduce the following two-stage compressed sensing scheme.

A. Proposed Scheme

The idea of the proposed scheme is to exploit the low-rank and sparse structures in two separate stages. In the first stage, we utilize the low-rank structure to recover $\boldsymbol{Y}$ from observations $\{\boldsymbol{Y}_{i,j}, (i,j) \in \Omega\}$. Note that $\boldsymbol{Z}$ and $\boldsymbol{F}$ do not need to be full-rank; instead, in order to achieve a lower sample complexity, they should have reduced dimensions, i.e. $N_Z < N_{\rm BS}$ and $N_F < N_{\rm MS}$. In other words, the size of $\boldsymbol{Y}$ is much smaller than the size of $\boldsymbol{H}$. In the second stage, based on the reconstructed $\boldsymbol{Y}$, we estimate the virtual beam space channel $\boldsymbol{H}_v$ by exploiting the sparse structure of $\boldsymbol{H}_v$. Through this two-stage scheme, the low-rank and sparse structures of the channel matrix $\boldsymbol{H}_v$ can be effectively decoupled and thus better utilized. For clarity, we summarize the two-stage scheme in Algorithm 1.

Algorithm 1 Two-Stage Compressed Sensing Algorithm

Given the measurements $\boldsymbol{Y}_\Omega$, and the matrices $\boldsymbol{A} \triangleq \boldsymbol{Z}^H \boldsymbol{A}_{\rm BS}$, $\boldsymbol{B} \triangleq \boldsymbol{F}^H \boldsymbol{A}_{\rm MS}$.

1 Recover $\hat{\boldsymbol{Y}}$ by solving

$$\min_{\hat{\boldsymbol{Y}}} \ \|\hat{\boldsymbol{Y}}\|_* \quad \text{s.t. } \hat{\boldsymbol{Y}}_{ij} = \boldsymbol{Y}_{ij} \ \ \forall\ (i,j) \in \Omega \tag{17}$$

2 Estimate $\hat{\boldsymbol{H}}_v$ via

$$\min_{\boldsymbol{H}_v} \ \|\boldsymbol{H}_v\|_1 \quad \text{s.t. } \hat{\boldsymbol{Y}} = \boldsymbol{A}\boldsymbol{H}_v\boldsymbol{B}^H \tag{18}$$

The second stage of our proposed method has a formulation slightly different from the conventional compressed sensing method discussed in Section II. The conventional compressed sensing method estimates the channel based only on those directly observed data, i.e. $\{\boldsymbol{Y}_{i,j} | (i,j) \in \Omega\}$. While our proposed method, in the second stage, not only uses those directly observed data, but also relies on those data recovered from the first stage. In other words, a whole matrix $\hat{\boldsymbol{Y}}$, including those directly observed and those recovered from the first stage, is used to estimate the channel in the second stage.

B. Theoretical Results

We now provide theoretical guarantees for our proposed two-stage compressed sensing scheme. Before proceeding, we first introduce the following lemma.

Lemma 1: Let $\boldsymbol{X} \in \mathbb{C}^{M_1 \times M_2}$ denote a sparse matrix with at most k nonzero columns and rows. $\boldsymbol{\Phi} \in \mathbb{C}^{N_\Phi \times M_1}$ and $\boldsymbol{\Psi} \in \mathbb{C}^{N_\Psi \times M_2}$ satisfy the $2k$-restricted isometry property with δ_{2k}, namely,

$$(1-\delta_{2k})\|\boldsymbol{x}\|_2^2 \le \|\boldsymbol{\Phi}\boldsymbol{x}\|_2^2 \le (1+\delta_{2k})\|\boldsymbol{x}\|_2^2$$
$$(1-\delta_{2k})\|\boldsymbol{x}\|_2^2 \le \|\boldsymbol{\Psi}\boldsymbol{x}\|_2^2 \le (1+\delta_{2k})\|\boldsymbol{x}\|_2^2$$

for all $2k$-sparse vectors $\boldsymbol{x}$, where $\delta_{2k} \triangleq \max\{\delta_{2k}(\boldsymbol{\Phi}), \delta_{2k}(\boldsymbol{\Psi})\}$, with $\delta_{2k}(\boldsymbol{\Phi})$ and $\delta_{2k}(\boldsymbol{\Psi})$ denoting the restricted isometry constants (RIC) of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ respectively. Let

$$\boldsymbol{G} = \boldsymbol{\Phi}\boldsymbol{X}\boldsymbol{\Psi}^H \tag{19}$$

and suppose the following condition holds

$$\delta_{2k} < 1 + \sqrt{2}\left(k - \sqrt{k(k+\sqrt{2})}\right) \tag{20}$$

Then $\boldsymbol{X}$ can be exactly recovered via

$$\min_{\hat{\boldsymbol{X}}} \ \|\hat{\boldsymbol{X}}\|_1 \quad \text{s.t. } \boldsymbol{G} = \boldsymbol{\Phi}\hat{\boldsymbol{X}}\boldsymbol{\Psi}^H \tag{21}$$

It can be easily verified that the term on the right-hand side of (20) is within the interval $(0, 1)$.

Proof: See Appendix A. ∎

Based on Lemma 1, our main results are summarized as follows.

Theorem 1: Consider the channel estimation problem described in (13), where the indexes in Ω are uniformly chosen at random with $|\Omega| = T$. The channel matrix $\boldsymbol{H}$ can be represented in a form of (12). Let L denote the rank of $\boldsymbol{H}$, and p denote the maximum number of nonzero entries in $\{\boldsymbol{\alpha}_l, \boldsymbol{\beta}_l\}_l$. Suppose $\boldsymbol{A} \in \mathbb{C}^{N_Z \times N_1}$ and $\boldsymbol{B} \in \mathbb{C}^{N_F \times N_2}$ are random matrices with i.i.d. Gaussian random entries $a_{i,j} \sim \mathcal{N}(0, \frac{1}{N_Z})$ and $b_{i,j} \sim \mathcal{N}(0, \frac{1}{N_F})$.[1] Define $n \triangleq \max\{N_F, N_Z\}$. There exist positive absolute constants c_1, c_2, c_3, c_4, c_5 and c_6 such that if

$$N_Z \ge c_1 pL\log(N_1/pL) \tag{22}$$
$$N_F \ge c_2 pL\log(N_2/pL) \tag{23}$$
$$T \ge c_3 n^{5/4} L\log(n) \tag{24}$$

then the channel $\boldsymbol{H}$ can be perfectly recovered from Algorithm 1 with probability exceeding $(1 - c_4 n^{-3})(1 - 2e^{-c_5 N_Z})(1 - 2e^{-c_6 N_F})$.

Proof: Please see Appendix B. ∎

C. Discussions

From Theorem 1, we see that the number of measurements T required for exact channel recovery is of order

$$O(p^{5/4} L^{9/4}\log(n)) \tag{25}$$

which scales approximately linearly with p and quadratically with the rank L. Since p and L are usually much smaller than $\max\{N_{\rm BS}, N_{\rm MS}\}$, our proposed two-stage scheme can achieve substantial overhead reduction as compared with the low rank matrix completion scheme whose required number of measurements scales linearly with $\max\{N_{\rm BS}, N_{\rm MS}\}$.

[1] See discussions in Section IV.C regarding this assumption.

It is also interesting to compare our proposed two-stage scheme with a convex relaxation-based compressed sensing method which solves (13) by directly formulating (13) into a sparse recovery problem (6). Note that $\boldsymbol{h} = \text{vec}(\boldsymbol{H}_v)$ has at most p^2L nonzero entries. According to the compressed sensing theory [29], we know that the probability of successful recovery of $\boldsymbol{h}$ via (6) exceeds $1-\delta$ if

$$T \geq Cp^2L\log(N_1N_2/\delta) \tag{26}$$

in which C is a positive constant. Thus the number of measurements required for exact channel recovery is of order

$$O(p^2L) \tag{27}$$

for the conventional convex relaxation-based compressed sensing method. Comparing (25) with (27), we can see that our proposed two-stage scheme achieves a lower sample complexity than the conventional compressed sensing method if $L^{5/4} < p^{3/4}$, or approximately $L < p$. Note that L represents the number of scattering clusters, and p, the largest number of nonzero entries in $\{\boldsymbol{\alpha}_l, \boldsymbol{\beta}_l\}$, is a quantity that measures the maximum angular spread among all scattering clusters. Recent dense-urban mmWave channel measurements indeed show that the condition $L < p$ may be satisfied in practice. In [22, Table I], it was shown that the average number of clusters is 1.8 for 28 GHz and 1.9 for 73 GHz. While the average base station root mean-squared (rms) angular spread is 10.2° for 28 GHz and 10.5° for 73 GHz, in which case we have $p \approx 17$ if we discretize the AoA domain into 64 grid points. Measurements for indoor environments, e.g. [35], also revealed that the average angular spread could be much larger than the number of average clusters. Hence the condition $L < p$ may still hold for indoor environments.

In Theorem 1, we assume that $\boldsymbol{A} \triangleq \boldsymbol{Z}^H\boldsymbol{A}_{\text{BS}}$ and $\boldsymbol{B} \triangleq \boldsymbol{F}^H\boldsymbol{A}_{\text{MS}}$ are random matrices with i.i.d. Gaussian random entries. Nevertheless, noticing that $\boldsymbol{A}_{\text{BS}}$ and $\boldsymbol{A}_{\text{MS}}$ are structured matrices consisting of array response vectors, it may not be possible to devise beamforming and combining matrices $\{\boldsymbol{Z}, \boldsymbol{F}\}$ such that the resulting $\boldsymbol{A}$ and $\boldsymbol{B}$ satisfy the i.i.d. Gaussian assumption. We, however, still make such an assumption in order to facilitate our theoretical analysis. On the other hand, recent theoretical and empirical studies [36] show that structured matrices also enjoy nice restricted isometry properties. Note that the same problem exists for the conventional compressed sensing method, where the sensing matrix is highly structured but a random sensing matrix assumption is evoked in order to obtain its sample complexity.

D. Extension to the Noisy Case

In the previous subsections, we ignore the observation noise in order to simplify our theoretical analysis. Nevertheless, the two-stage compressed sensing scheme can be easily adapted to the noisy case. For clarity, the two-stage algorithm for the noisy case is summarized as follows.

In Algorithm 2, ε and ϵ are error tolerance parameters. Also, the constrained optimizations (28) and (29) can be converted to unconstrained optimization problems by introducing an appropriate choice of the regularization parameter λ. For example, (28) can be replaced by

$$\min_{\hat{\boldsymbol{Y}}} \ \|\hat{\boldsymbol{Y}}_\Omega - \boldsymbol{Y}_\Omega\|_F^2 + \lambda\|\hat{\boldsymbol{Y}}\|_* \tag{30}$$

which can be efficiently solved by the fixed point continuation algorithm [37].

Algorithm 2 Robust Two-Stage Compressed Sensing Algorithm

Given the measurements $\boldsymbol{Y}_\Omega$, the matrices $\boldsymbol{A} \triangleq \boldsymbol{Z}^H\boldsymbol{A}_{\text{BS}}$, $\boldsymbol{B} \triangleq \boldsymbol{F}^H\boldsymbol{A}_{\text{MS}}$.

1 Recover $\hat{\boldsymbol{Y}}$ by solving

$$\begin{aligned} &\min \ \|\hat{\boldsymbol{Y}}\|_* \\ &\text{s.t. } \|\hat{\boldsymbol{Y}}_\Omega - \boldsymbol{Y}_\Omega\|_F < \varepsilon \end{aligned} \tag{28}$$

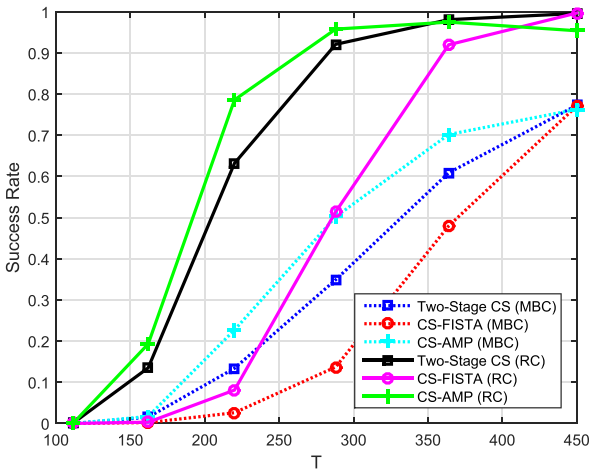
2 Estimate $\hat{\boldsymbol{H}}_v$ via

$$\begin{aligned} &\min \ \|\boldsymbol{H}_v\|_1 \\ &\text{s.t. } \|\hat{\boldsymbol{Y}} - \boldsymbol{A}\boldsymbol{H}_v\boldsymbol{B}^H\|_F < \epsilon \end{aligned} \tag{29}$$

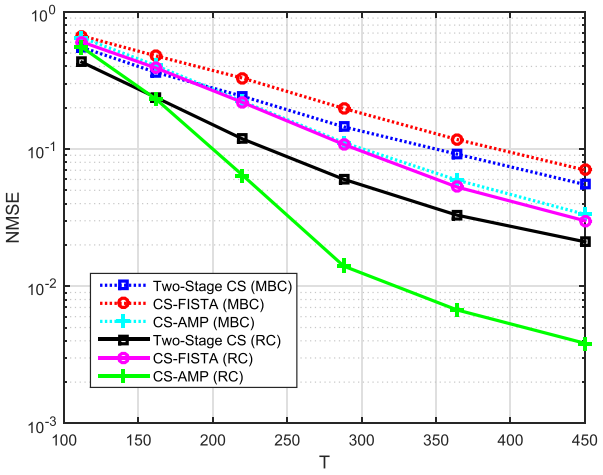
V. Simulation Results

We now carry out simulation results to illustrate the performance of our proposed two-stage compressed sensing (referred to as two-stage CS) method. We compare our method with the conventional compressed sensing approach which formulates mmWave channel estimation into a sparse signal recovery problem (5). Note that different compressed sensing algorithms can be employed to solve (5). Here we consider a fast iterative shrinkage-thresholding algorithm (FISTA) [27] which is a convex relaxation-based compressed sensing method involving solving (6), and an approximate message passing (AMP)-based Bayesian method developed in [13] and [38]. These two conventional compressed sensing methods are respectively referred to as CS-FISTA and CS-AMP. For our proposed method, we use the singular value thresholding (SVT) algorithm [39] and the fixed point continuation (FPC) algorithm [37] to solve the matrix completion problem for the noiseless and noisy case, respectively. The FISTA is employed to perform the sparse recovery stage.

We consider a scenario where both the BS and the MS employ a uniform linear array with $N_{\text{BS}} = N_{\text{MS}} = 64$ antennas. The distance between neighboring antenna elements is assumed to be half the wavelength of the signal. The mmWave channel is assumed to follow the geometric channel model (10) with $L = 2$ clusters. The mean AoAs/AoDs for these two clusters are set to $\theta_1 = \phi_1 = \pi/6$, $\theta_2 = \phi_2 = -\pi/6$, respectively. Unless otherwise specified, the AoA and AoD angular spreads for each cluster are set to $\delta_\theta = 15°$ and $\delta_\phi = 10°$. The relative AoA/AoD shifts are uniformly generated within the angular spreads, i.e. $\vartheta_{l,i} \in (\theta_l - \delta_\theta/2, \theta_l + \delta_\theta/2)$, $\varphi_{l,i} \in (\phi_l - \delta_\phi/2, \phi_l + \delta_\phi/2)$. In our experiments, we discretize the AoA/AoD domains into 64 grid points, i.e. $N_1 = N_{\text{BS}}$, and $N_2 = N_{\text{MS}}$, in which case δ_θ and δ_ϕ span across about 7 and 5 grid points, respectively. The complex gains $\{\alpha_{l,i}\beta_{l,j}\}$

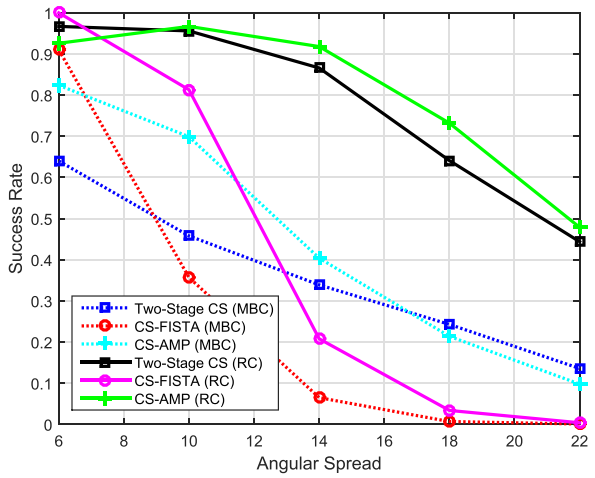


(a) Success rates vs. T.

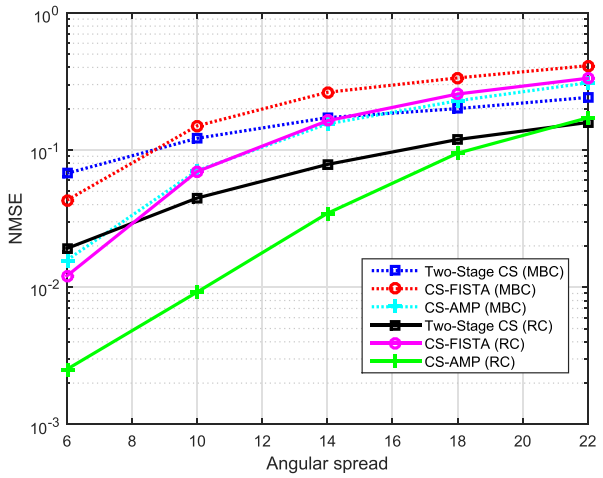


(b) NMSEs vs. T.

Fig. 3. Success rates and NMSEs of respective algorithms vs. T.



(a) Success rates vs. the angular spread.



(b) NMSEs vs. the angular spread.

Fig. 4. Success rates and NMSEs of respective algorithms vs. the angular spread.

are assumed to be random variables following a circularly symmetric complex Gaussian distribution $\mathcal{CN}(0, 1/\rho)$, where ρ is given by $\rho = (4\pi Df_c/c)^2$. Here c represents the speed of light, D denotes the distance between the BS and the MS, f_c is the carrier frequency, and we set $D = 30$m and $f_c = 28$GHz.

The performance is evaluated via two metrics, namely, the normalized mean squared error (NMSE) and the success rate. The NMSE is calculated as

$$\text{NMSE} = E\left[\frac{\left\|\hat{\boldsymbol{H}} - \boldsymbol{H}\right\|_F^2}{\|\boldsymbol{H}\|_F^2}\right] \tag{31}$$

where $\hat{\boldsymbol{H}}$ denotes the estimate of the true channel $\boldsymbol{H}$. The success rate is computed as the ratio of the number of successful trials to the total number of independent runs. A trial is considered successful if the normalized reconstruction error is no greater than 10^{-2}.

In our experiments, the beamforming/combining codebooks, i.e. $\boldsymbol{F}$ and $\boldsymbol{Z}$, are generated according to two different ways. The first is to have the entries of $\boldsymbol{F}$ and $\boldsymbol{Z}$ uniformly chosen from a unit circle, in which case the antenna array has a quasi-omnidirectional beam pattern. This scheme is referred to as a random coding (RC) scheme. Another scheme of devising $\boldsymbol{F}$ and $\boldsymbol{Z}$ is to steer the antenna array to beam along multiple directions simultaneously, which is achieved by dividing the antenna array into a number of sub-arrays and making each sub-array beam toward an individual direction [7]. The steering directions are randomized for each measurement. This scheme is named as multiple-beam coding (MBC) scheme. In order to provide a fair comparison, the columns of $\boldsymbol{F}$ and $\boldsymbol{Z}$ are normalized to unit norm for both beam pattern design schemes. We assume that, at each time instant, the beamforming vector $\boldsymbol{f}(t)$ and the combining vector $\boldsymbol{z}(t)$ are randomly chosen from the beamforming/combining codebooks, respectively. Hence the measurement process can be deemed as randomly collecting samples from a low-rank matrix $\boldsymbol{Y} = \boldsymbol{Z}^H\boldsymbol{H}\boldsymbol{F}$ (cf. (13)), where $\boldsymbol{Y}$ is an $N_Z \times N_F$ matrix. For simplicity, we assume $N_Z = N_F$. Also, in our experiments, the value of N_Z (N_F) is adaptively adjusted such that the ratio of the number of observed entries T to the total number of entries in $\boldsymbol{Y}$ is fixed to be $1/2$, i.e. $T = (1/2)N_Z N_F$. Such a setup can

provide a reliable matrix completion result, which in turn helps achieve an accurate channel estimate for our proposed two-stage method. The adaptive adjustment of the dimensions of the codebooks can be easily implemented in practice. We can first generate augmented beamforming/combining codebooks and then choose $\boldsymbol{Z}$ and $\boldsymbol{F}$ as subsets (with variable dimensions) of the augmented codebooks.

We now examine the estimation performance of our proposed two-stage CS method and the conventional CS methods. Fig. 3 plots the success rates for the noiseless case and NMSEs for the noisy case as a function of the number of measurements T, where for the noisy case, the SNR, defined as $10\log(\|\boldsymbol{H}\|_F^2/(N_{\mathrm{BS}}N_{\mathrm{MS}}\sigma^2))$, is set equal to 20dB. From Fig. 3, we see that better performance can be obtained by using the beamforming/combining codebooks that are generated according to the RC scheme. Also, our proposed two-stage CS method presents a clear performance advantage over the CS-FISTA algorithm, whichever beamforming/combining codebooks are used. This result corroborates our claim that the proposed two-stage CS method can achieve a lower sample complexity than the CS-FISTA method. We also observe that the CS-AMP method offers performance better than our proposed two-stage CS method. The CS-AMP is a non-convex method developed in a Bayesian framework to solve the sparse signal recovery problem (5). Since the sparsity-promoting prior in [38] behaves more like the ℓ_0-norm than the ℓ_1-norm, the CS-AMP method is able to achieve superior performance. This result, however, does not contradict the claim of superiority of the two-stage scheme over the conventional compressed sensing method because our conclusion is reached based on a sample complexity analysis of the convex relaxation-based (i.e. the CS-FISTA) method. It should be noted that, the CS-AMP, as a nonconvex method, may converge to an undesirable local minimum. In addition, since an exact reconstruction theoretical guarantee is unavailable for the CS-AMP, the CS-AMP method is not guaranteed to yield the true solution even if a global minimum is reached.

Next, in Fig. 4, we examine the performance of respective algorithms as a function of the angular spread, where the AoA and AoD angular spreads are assumed to be the same and vary from 6° to 22°, i.e. $\delta_\theta = \delta_\phi \in [6°, 22°]$. Also, we set $N_Z = N_F = 24$, $T = 0.5N_ZN_F$, and the SNR is set to 20dB for the noisy case. From Fig. 4, we see that the CS-FISTA method outperforms our proposed two-stage scheme when the angular spread is small, say, $\delta_\theta = \delta_\phi = 6°$, whereas our proposed method achieves a performance improvement over the CS-FISTA (even over the CS-AMP in some cases) as the angular spread becomes large. This result, again, substantiates our theoretical analysis. In Fig. 5, we depict the NMSEs of respective algorithms vs. the SNR, where we set $N_Z = N_F = 24$, $T = 0.5N_ZN_F$, $\delta_\theta = 15°$ and $\delta_\phi = 10°$. We see that the proposed two-stage CS method outperforms the CS-FISTA method in moderate and high SNR regimes.

Lastly, we examine the effect of channel estimation accuracy on the bit error rate (BER) performance. For each method, after the channel is estimated, the beamforming/combining vectors are calculated via the technique developed in [40]. Fig. 6 plots the BERs of respective algorithms vs. the number of observed samples T, where we set SNR = 30dB, $\delta_\theta = 15°$ and $\delta_\phi = 10°$. We see that our proposed method achieves consistently lower BERs than the CS-FISTA method.

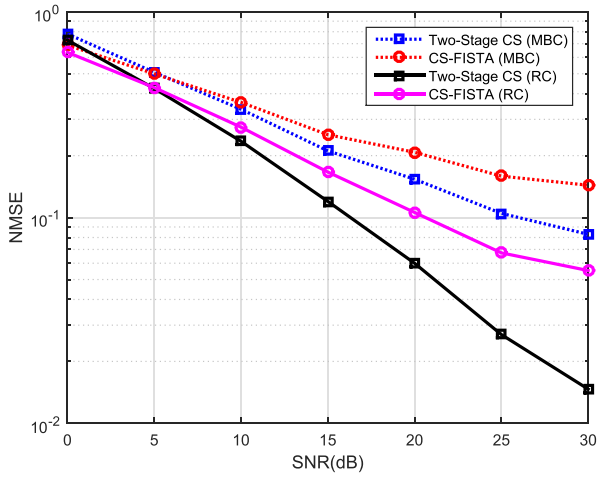


Fig. 5. NMSEs of respective algorithms vs. SNR.

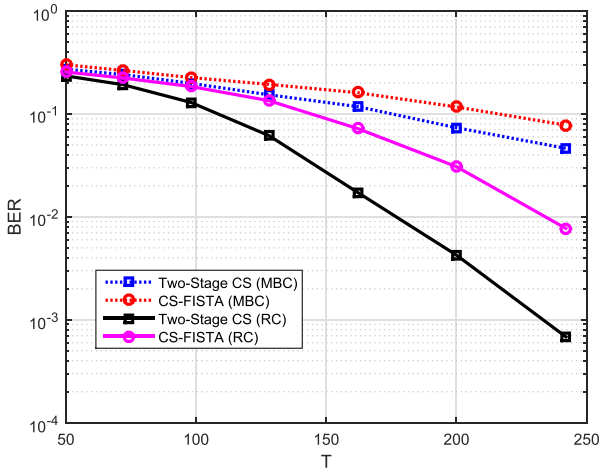


Fig. 6. BERs of respective algorithms vs. T.

VI. Conclusions

We studied the problem of channel estimation for mmWave systems with only one RF chain used at the BS and MS. Besides the sparse scattering characteristics, we also considered the effect of angular spreads in channel modeling and algorithm development. We showed that, in the presence of angular spreads, mmWave channels exhibit a jointly sparse and low-rank structure. A two-stage compressed sensing method was developed, in which a matrix completion stage is first performed, and then followed by a sparse recovery stage to estimate the mmWave channel. Theoretical analysis was also conducted. It reveals that the proposed two-stage method requires fewer measurements than a convex relaxation-based compressed sensing method which only exploits the sparsity of mmWave channels. Simulation results were provided to corroborate our theoretical analysis and demonstrate the superiority of the proposed two-stage compressed sensing method.

Appendix A
Proof of Lemma 1

Before proving $\hat{\boldsymbol{X}} = \boldsymbol{X}$, we first show that for any sparse matrix $\boldsymbol{S} \in \mathbb{C}^{M_1 \times M_2}$ with at most $2k$ nonzero columns

and rows, we have

$$(1-\delta_{2k})^2\left\|\boldsymbol{S}\right\|_F^2 \le \left\|\boldsymbol{\Phi}\boldsymbol{S}\boldsymbol{\Psi}^H\right\|_F^2 \le (1+\delta_{2k})^2\left\|\boldsymbol{S}\right\|_F^2 \quad (32)$$

Since $\boldsymbol{\Phi}$ satisfies the $2k$-RIP and each column of $\boldsymbol{S}$ is a $2k$-sparse vector, adding all the inequalities together leads to

$$(1-\delta_{2k})\left\|\boldsymbol{S}\right\|_F^2 \le \left\|\boldsymbol{\Phi}\boldsymbol{S}\right\|_F^2 \le (1+\delta_{2k})\left\|\boldsymbol{S}\right\|_F^2 \quad (33)$$

Meanwhile, note that $\boldsymbol{S}^H\boldsymbol{\Phi}^H$ has at most $2k$ non-zero rows, i.e. each column of $\boldsymbol{S}^H\boldsymbol{\Phi}^H$ is also a $2k$-sparse vector. Using the RIP associated with $\boldsymbol{\Psi}$, we have

$$\left\|\boldsymbol{\Psi}\boldsymbol{S}^H\boldsymbol{\Phi}^H\right\|_F^2 \le (1+\delta_{2k})\left\|\boldsymbol{S}^H\boldsymbol{\Phi}^H\right\|_F^2 \le (1+\delta_{2k})^2\left\|\boldsymbol{S}\right\|_F^2 \quad (34)$$

$$\left\|\boldsymbol{\Psi}\boldsymbol{S}^H\boldsymbol{\Phi}^H\right\|_F^2 \ge (1-\delta_{2k})\left\|\boldsymbol{S}^H\boldsymbol{\Phi}^H\right\|_F^2 \ge (1-\delta_{2k})^2\left\|\boldsymbol{S}\right\|_F^2 \quad (35)$$

Combining (34)–(35), we arrive at (32).

Using (32), we now prove that $\boldsymbol{E} \triangleq \hat{\boldsymbol{X}} - \boldsymbol{X}$ equals zero, i.e. $\|\boldsymbol{E}\|_F = 0$. Let Ω denotes the support set (i.e. the set of indices of non-zeros entries) of $\boldsymbol{X}$. $\boldsymbol{E}$ can be decomposed as

$$\boldsymbol{E} = \sum_{i=0}^{N}\boldsymbol{E}_i \quad (36)$$

where $\boldsymbol{E}_0$ is a matrix whose entries in the set Ω are equivalent to those of $\boldsymbol{E}$, while the rest of entries are equal to zero, $\boldsymbol{E}_i$ ($i \neq 0$) are matrices with at most k nonzero columns and rows, and they have disjoint support sets such that $(1/k)\|\boldsymbol{E}_i\|_1 \ge \|\boldsymbol{E}_{i+1}\|_\infty$ for $i = 1, \ldots, N-1$. Note that this inequality can be automatically satisfied if we arrange the entries of $\boldsymbol{E} - \boldsymbol{E}_0$ in descending order in terms of magnitude and choose entries from the $((i-1)k+1)$th to the (ik)th (may be smaller if $i = N$) as the entries of $\boldsymbol{E}_i$, while the rest of entries of $\boldsymbol{E}_i$ are equal to zero.

Since $\hat{\boldsymbol{X}}$ is an optimal solution to (21), we have

$$\begin{aligned}\|\boldsymbol{X}\|_1 \ge \|\hat{\boldsymbol{X}}\|_1 &= \|\boldsymbol{E}+\boldsymbol{X}-\boldsymbol{E}_0+\boldsymbol{E}_0\|_1 \\ &\ge \|\boldsymbol{E}+\boldsymbol{X}-\boldsymbol{E}_0\|_1 - \|\boldsymbol{E}_0\|_1 \\ &= \|\boldsymbol{X}\|_1 + \|\boldsymbol{E}-\boldsymbol{E}_0\|_1 - \|\boldsymbol{E}_0\|_1 \end{aligned} \quad (37)$$

Thus we obtain

$$\|\boldsymbol{E}-\boldsymbol{E}_0\|_1 \le \|\boldsymbol{E}_0\|_1 \overset{(a)}{\le} k\|\boldsymbol{E}_0\|_F \quad (38)$$

where (a) comes from the Cauchy-Schwarz inequality. Also, we have

$$\begin{aligned}\|\boldsymbol{E}-(\boldsymbol{E}_0+\boldsymbol{E}_1)\|_F &\le \sum_{i=2}^{N}\|\boldsymbol{E}_i\|_F \overset{(a)}{\le} \sum_{i=1}^{N-1}\|\boldsymbol{E}_i\|_1 \overset{(b)}{\le} \|\boldsymbol{E}_0\|_1 \\ &\overset{(c)}{\le} k\|\boldsymbol{E}_0\|_F \le k\|\boldsymbol{E}_0+\boldsymbol{E}_1\|_F \end{aligned} \quad (39)$$

where (a) comes from the fact that

$$\|\boldsymbol{E}_i\|_1 \ge k\|\boldsymbol{E}_{i+1}\|_\infty \ge \|\boldsymbol{E}_{i+1}\|_F \quad (40)$$

and the inequalities (b) and (c) follow from (38). The result (39) implies that

$$\|\boldsymbol{E}\|_F \le (k+1)\|\boldsymbol{E}_0+\boldsymbol{E}_1\|_F \quad (41)$$

We now prove $\|\boldsymbol{E}_0+\boldsymbol{E}_1\|_F = 0$. Note that $\boldsymbol{E}_0+\boldsymbol{E}_1$ is a sparse matrix with at most $2k$ nonzero columns and rows. Using (32), we have

$$\begin{aligned}&(1-\delta_{2k})^2\|\boldsymbol{E}_0+\boldsymbol{E}_1\|_F^2 \le \left\|\boldsymbol{\Phi}(\boldsymbol{E}_0+\boldsymbol{E}_1)\boldsymbol{\Psi}^H\right\|_F^2 \\ &= \mathrm{tr}[(\boldsymbol{\Phi}(\boldsymbol{E}_0+\boldsymbol{E}_1)\boldsymbol{\Psi}^H)^H\boldsymbol{\Phi}(\boldsymbol{E}-\sum_{i=2}^{N}\boldsymbol{E}_i)\boldsymbol{\Psi}^H] \\ &= \Re\{\mathrm{tr}[(\boldsymbol{\Phi}(\boldsymbol{E}_0+\boldsymbol{E}_1)\boldsymbol{\Psi}^H)^H\boldsymbol{\Phi}\boldsymbol{E}\boldsymbol{\Psi}^H]\} \\ &\quad -\Re\{\mathrm{tr}[(\boldsymbol{\Phi}(\boldsymbol{E}_0+\boldsymbol{E}_1)\boldsymbol{\Psi}^H)^H\boldsymbol{\Phi}\sum_{i=2}^{N}\boldsymbol{E}_i\boldsymbol{\Psi}^H]\} \\ &\le \Re\{\mathrm{tr}[(\boldsymbol{\Phi}(\boldsymbol{E}_0+\boldsymbol{E}_1)\boldsymbol{\Psi}^H)^H\boldsymbol{\Phi}\boldsymbol{E}\boldsymbol{\Psi}^H]\} \\ &\quad +\left|\Re\{\mathrm{tr}[(\boldsymbol{\Phi}(\boldsymbol{E}_0+\boldsymbol{E}_1)\boldsymbol{\Psi}^H)^H\boldsymbol{\Phi}\sum_{i=2}^{N}\boldsymbol{E}_i\boldsymbol{\Psi}^H]\}\right| \\ &\overset{(a)}{\le} \sum_{i=0}^{1}\sum_{j=2}^{N}\left|\Re\{\mathrm{tr}((\boldsymbol{\Phi}\boldsymbol{E}_i\boldsymbol{\Psi}^H)^H\boldsymbol{\Phi}\boldsymbol{E}_j\boldsymbol{\Psi}^H)\}\right| \\ &= \sum_{i=0}^{1}\sum_{j=2}^{N}\left|\Re\{\mathrm{tr}((\boldsymbol{\Phi}\frac{\boldsymbol{E}_i}{\|\boldsymbol{E}_i\|_F}\boldsymbol{\Psi}^H)^H\boldsymbol{\Phi}\frac{\boldsymbol{E}_j}{\|\boldsymbol{E}_j\|_F}\boldsymbol{\Psi}^H)\}\right|\cdot\|\boldsymbol{E}_i\|_F\|\boldsymbol{E}_j\|_F \\ &\overset{(b)}{=} \sum_{i=0}^{1}\sum_{j=2}^{N}\frac{1}{4}\Big|\|\boldsymbol{\Phi}(\frac{\boldsymbol{E}_i}{\|\boldsymbol{E}_i\|_F}+\frac{\boldsymbol{E}_j}{\|\boldsymbol{E}_j\|_F})\boldsymbol{\Psi}^H\|_F^2 \\ &\quad -\|\boldsymbol{\Phi}(\frac{\boldsymbol{E}_i}{\|\boldsymbol{E}_i\|_F}-\frac{\boldsymbol{E}_j}{\|\boldsymbol{E}_j\|_F})\boldsymbol{\Psi}^H\|_F^2\Big|\cdot\|\boldsymbol{E}_i\|_F\|\boldsymbol{E}_j\|_F \\ &\le \sum_{i=0}^{1}\sum_{j=2}^{N}\frac{1}{4}\Big((1+\delta_{2k})^2\|\frac{\boldsymbol{E}_i}{\|\boldsymbol{E}_i\|_F}+\frac{\boldsymbol{E}_j}{\|\boldsymbol{E}_j\|_F}\|_F^2 \\ &\quad -(1-\delta_{2k})^2\|\frac{\boldsymbol{E}_i}{\|\boldsymbol{E}_i\|_F}-\frac{\boldsymbol{E}_j}{\|\boldsymbol{E}_j\|_F}\|_F^2\Big)\cdot\|\boldsymbol{E}_i\|_F\|\boldsymbol{E}_j\|_F \\ &\overset{(c)}{=} \sum_{i=0}^{1}\sum_{j=2}^{N}\frac{1}{4}((1+\delta_{2k})^2\left(\|\frac{\boldsymbol{E}_i}{\|\boldsymbol{E}_i\|_F}\|_F^2+\|\frac{\boldsymbol{E}_j}{\|\boldsymbol{E}_j\|_F}\|_F^2\right) \\ &\quad -(1-\delta_{2k})^2\left(\|\frac{\boldsymbol{E}_i}{\|\boldsymbol{E}_i\|_F}\|_F^2+\|\frac{\boldsymbol{E}_j}{\|\boldsymbol{E}_j\|_F}\|_F^2)\right)\cdot\|\boldsymbol{E}_i\|_F\|\boldsymbol{E}_j\|_F \\ &= \sum_{i=0}^{1}\sum_{j=2}^{N}\frac{1}{2}((1+\delta_{2k})^2-(1-\delta_{2k})^2)\cdot\|\boldsymbol{E}_i\|_F\|\boldsymbol{E}_j\|_F \\ &= 2\delta_{2k}\sum_{i=0}^{1}\sum_{j=2}^{N}\|\boldsymbol{E}_i\|_F\left\|\boldsymbol{E}_j\right\|_F \\ &= 2\delta_{2k}(\|\boldsymbol{E}_0\|_F+\|\boldsymbol{E}_1\|_F)\sum_{j=2}^{N}\left\|\boldsymbol{E}_j\right\|_F \\ &\overset{(d)}{\le} 2k\delta_{2k}(\|\boldsymbol{E}_0\|_F+\|\boldsymbol{E}_1\|_F)\|\boldsymbol{E}_0+\boldsymbol{E}_1\|_F \\ &\overset{(e)}{\le} 2\sqrt{2}k\delta_{2k}\|\boldsymbol{E}_0+\boldsymbol{E}_1\|_F^2 \end{aligned} \quad (42)$$

where (a) comes from the fact that

$$\boldsymbol{\Phi}\boldsymbol{E}\boldsymbol{\Psi}^H = \boldsymbol{\Phi}\boldsymbol{X}\boldsymbol{\Psi}^H - \boldsymbol{\Phi}\hat{\boldsymbol{X}}\boldsymbol{\Psi}^H = \boldsymbol{0} \quad (43)$$

(b) follows from the equality

$$4\Re\{\mathrm{tr}(\boldsymbol{P}\boldsymbol{Q}^H)\} = \|\boldsymbol{P}+\boldsymbol{Q}\|_F^2 - \|\boldsymbol{P}-\boldsymbol{Q}\|_F^2 \quad (44)$$

for any complex matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$, (c) is due to the reason that $\boldsymbol{E}_i$ and $\boldsymbol{E}_j$ have disjoint supports, (d) follows from (39), and (e) can be easily verified by noting that

$$\|\boldsymbol{E}_0 + \boldsymbol{E}_1\|_F = (\|\boldsymbol{E}_0\|_F^2 + \|\boldsymbol{E}_1\|_F^2)^{1/2} \tag{45}$$

If $2\sqrt{2}k\delta_{2k} - (1-\delta_{2k})^2 < 0$, i.e.

$$\delta_{2k} < 1 + \sqrt{2}\left(k - \sqrt{k(k+\sqrt{2})}\right) \tag{46}$$

then we have $\|\boldsymbol{E}_0 + \boldsymbol{E}_1\|_F = 0$ from (42), which implies that $\|\boldsymbol{E}\|_F = 0$, i.e. $\boldsymbol{X} = \hat{\boldsymbol{X}}$. The proof is completed here.

Appendix B
Proof of Theorem 1

Our proof proceeds in two steps. We first investigate the condition under which $\boldsymbol{Y}$ can be perfectly recovered from (17), and then examine the exact recovery condition for (18). By combining the results of the two stages, we arrive at results in Theorem 1.

Since $\boldsymbol{Y}$ has a low rank structure, the first stage is essentially a matrix completion stage. Invoking the matrix completion theory [31], we know that for some positive constants c_3 and c_4, if (24) holds, then $\boldsymbol{Y}$ can be perfectly recovered with probability exceeding $1 - c_4 n^{-3}$.

The second stage is a sparse matrix recovery stage. Note that $\boldsymbol{H}_v$ is a sparse matrix with at most pL nonzero columns and rows. Meanwhile, it is well-known that for a random matrix $\boldsymbol{\Psi} \in \mathbb{R}^{m_1 \times m_2}$ whose i.i.d. entries follow a Gaussian distribution with zero mean and variance $1/m_1$, if the following condition

$$m_1 \geq \eta k \log(m_2/k) \tag{47}$$

holds for a sufficiently large constant $\eta > 0$, then $\boldsymbol{\Psi}$ satisfies the $2k$-restricted isometry property for a sufficiently small restricted isometry constant $\delta_{2k}(\boldsymbol{\Psi})$ with probability exceeding $1 - 2e^{-cm_1}$ for some constant $c > 0$ that depends only on $\delta_{2k}(\boldsymbol{\Psi})$ [41]. Recalling Lemma 1, we therefore can naturally arrive at the following: for some positive constants c_1, c_2, c_5 and c_6, if (22) and (23) hold valid, then $\boldsymbol{H}_v$ can be perfectly recovered via (18) with probability exceeding $(1 - 2e^{-c_5 N_Z})(1 - 2e^{-c_6 N_F})$.

By combining the results from both stages, we now reach that there exist positive absolute constants c_1, c_2, c_3, c_4, c_5 and c_6 such that if (22)–(24) are satisfied, then the channel $\boldsymbol{H}$ can be perfectly recovered from Algorithm 1 with probability exceeding $(1 - c_4 n^{-3})(1 - 2e^{-c_5 N_Z})(1 - 2e^{-c_6 N_F})$. The proof is completed here.

References

[1] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, Jr., "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proc. IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

[2] 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.

[3] A. Ghosh *et al.*, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1152–1163, Jun. 2014.

[4] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[5] A. Alkhateeb, J. Mo, N. Gonzalez-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.

[6] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[7] O. Abari, H. Hassanieh, M. Rodriguez, and D. Katabi, "Millimeter wave communications: From point-to-point links to agile network connections," in *Proc. 15th ACM Workshop Hot Topics Netw.*, Atlanta, GA, USA, Nov. 2016, pp. 169–175.

[8] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2012, pp. 234–239.

[9] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive tracking with 1000-element arrays: A framework for multi-Gbps mm wave cellular downlinks," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput.*, Oct. 2012, pp. 690–697.

[10] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. 40th IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brisbane, QLD, Australia, Apr. 2015, pp. 2909–2913.

[11] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[12] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2014, pp. 273–277.

[13] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *Proc. 16th IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, Jun./Jul. 2015, pp. 146–150.

[14] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-wave picocells," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 514–527, Apr. 2016.

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

[16] X. Gao, L. Dai, and A. M. Sayeed. (2016). "Low RF-complexity technologies to enable millimeter-wave MIMO with large antenna array for 5G wireless communications." [Online]. Available: https://arxiv.org/abs/1607.04559

[17] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and S. Li, "Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7501–7516, Nov. 2016.

[18] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, Jul. 2017.

[19] M. Samimi *et al.*, "28 GHz angle of arrival and angle of departure analysis for outdoor cellular communications using steerable beam antennas in New York City," in *Proc. IEEE 77th Veh. Technol. Conf. (VTC Spring)*, Dresden, Germany, Jun. 2013, pp. 1–6.

[20] 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 Int. Conf. Commun. (ICC)*, Budapest, Hungary, Jun. 2013, pp. 5163–5167.

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

[22] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[23] P. Wang, M. Pajovic, P. V. Orlik, T. Koike-Akino, K. J. Kim, and J. Fang, "Sparse channel estimation in millimeter wave communications: Exploiting joint AoD-AoA angular spread," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[24] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.

[25] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, Jun. 2014.
[26] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, "Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4649–4662, Sep. 2016.
[27] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, Mar. 2009.
[28] 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.
[29] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[30] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[31] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009.
[32] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, Aug. 2010.
[33] V. Koltchinskii, K. Lounici, and A. Tsybakov, "Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion," *Ann. Statist.*, vol. 39, no. 5, pp. 2302–2329, Oct. 2011.
[34] S. Bahmani and J. Romberg, "Efficient compressive phase retrieval with constrained sensing vectors," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 28. Montreal, QC, Canada, Dec. 2015, pp. 523–531.
[35] X. Wu *et al.*, "60-GHz millimeter-wave channel measurements and modeling for indoor office environments," *IEEE Trans. Antennas Propag.*, vol. 65, no. 4, pp. 1912–1924, Apr. 2017.
[36] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.
[37] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Math. Program.*, vol. 128, nos. 1–2, pp. 321–353, Jun. 2011.
[38] J. Vila and P. Schniter, "Expectation-maximization Bernoulli–Gaussian approximate message passing," in *Proc. 45th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2011, pp. 799–803.
[39] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Mar. 2010.
[40] O. El 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.
[41] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approx.*, vol. 28, no. 3, pp. 253–263, Jan. 2008.

Xingjian Li received the B.Sc. degree from the University of Electronic Science and Technology, China, in 2015, where he is currently pursuing the Ph.D. degree. His current research interests include compressed sensing, millimeter wave, and massive MIMO communications.

Jun Fang (M'08) received the B.S. and M.S. degrees from Xidian University, Xi'an, China, in 1998 and 2001, respectively, and the Ph.D. degree from the National University of Singapore, Singapore, in 2006, all in electrical engineering.

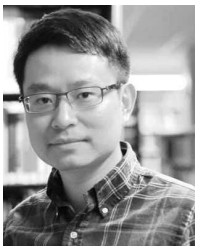
In 2006, he was a Post-Doctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University. From 2007 to 2010, he was a Research Associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology. Since 2011, he has been with the University of Electronic of Science and Technology of China. His research interests include compressed sensing and sparse theory, massive MIMO/mmWave communications, and statistical inference.

Dr. Fang received the IEEE Jack Neubauer Memorial Award in 2013 for the best systems paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He serves as an Associate Technical Editor for the *IEEE Communications Magazine*, and an Associate Editor for the IEEE *Signal Processing Letters*.

Hongbin Li (M'99–SM'08) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, in 1999, all in electrical engineering.

From 1996 to 1999, he was a Research Assistant with the Department of Electrical and Computer Engineering, University of Florida. Since 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, where he is currently a Professor. He was a Summer Visiting Faculty Member with the Air Force Research Laboratory in 2003, 2004, and 2009. His research interests include statistical signal processing, wireless communications, and radars.

Dr. Li has been a member of the IEEE SPS Signal Processing Theory and Methods since 2011, Technical Committee (TC), and the IEEE SPS Sensor Array and Multichannel TC from 2006 to 2012. He is a member of Tau Beta Pi and Phi Kappa Phi. He received the IEEE Jack Neubauer Memorial Award in 2013 for the best systems paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the Outstanding Paper Award from the IEEE AFICON Conference in 2011, the Harvey N. Davis Teaching Award in 2003, and the Jess H. Davis Memorial Award for excellence in research in 2001 from the Stevens Institute of Technology, and the Sigma Xi Graduate Research Award from the University of Florida in 1999. He has been involved in various conference organization activities, including serving as the General Co-Chair for the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop, Hoboken, NJ, in 2012. He has been an Associate Editor for the *Signal Processing* (Elsevier) since 2013, the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2006 to 2009 and since 2014, the IEEE SIGNAL PROCESSING LETTERS from 2005 to 2006, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2003 to 2006, and a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and the *EURASIP Journal on Applied Signal Processing*.



Pu Wang (S'05–M'12) received the Ph.D. degree in electrical engineering from the Stevens Institute of Technology, Hoboken, NJ, USA, in 2011.

He was an intern with the Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA, in 2010. He was a Research Scientist with Schlumberger-Doll Research, Cambridge, MA, contributing to developments of logging-while-drilling Acoustics/NMR products, before returning to the MERL in 2016, where he is a Principle Research Scientist. His current research interests include statistical signal processing, Bayesian inference, sparse signal recovery, and their applications to sensing, wireless communications, and networks.

Dr. Wang was selected as a 2016–2017 Distinguished Speaker of the Society of Petrophysicists and Well Log Analysts (SPWLA) for "Dipole Shear Anisotropy Using Logging-While-Drilling Sonic Tools," one of the top papers presented at the 57th SPWLA Annual Symposium, Reykjavik, Iceland, in 2016. He was a recipient of the Head of the Cooperate Research and Development Award of Mitsubishi Electric Cooperation in 2017, the Conrad Schlumberger Award for Technical Depth from the Schlumberger Reservoir Symposium in 2015 and 2017, the IEEE Jack Neubauer Memorial Award for the best systems paper published by the IEEE Vehicular Technology Society in 2013, and the Outstanding Paper Award from the IEEE AFRICON Conference in 2011. He received the Outstanding Doctoral Thesis in the EE Award in 2011, the Edward Peskin Award in 2011, the Francis T. Boesch Award in 2008, and the Outstanding Research Assistant Award in 2007, all from the Stevens Institute of Technology.