

# Channel Estimation for Millimeter-Wave Massive MIMO With Hybrid Precoding Over Frequency-Selective Fading Channels

Zhen Gao, Chen Hu, Linglong Dai, and Zhaocheng Wang

**Abstract**—This letter proposes a multi-user uplink channel estimation scheme for mmWave massive MIMO over frequency selective fading (FSF) channels. Specifically, by exploiting the angle-domain structured sparsity of mmWave FSF channels, a distributed compressive sensing-based channel estimation scheme is proposed. Moreover, by using the grid matching pursuit strategy with adaptive measurement matrix, the proposed algorithm can solve the power leakage problem caused by the continuous angles of arrival or departure. Simulation results verify the good performance of the proposed solution.

**Index Terms**—Millimeter-wave massive MIMO, frequency-selective fading, channel estimation, compressive sensing.

## I. INTRODUCTION

MILLIMETER-WAVE (mmWave) massive MIMO has been considered as a promising key technique for 5G [1]. For mmWave massive MIMO with the phase shifter network based hybrid precoding [1]–[3] or electromagnetic lens based beamspace MIMO [4], [5], the number of radio frequency (RF) chains is usually much smaller than that of antennas for reduced hardware cost and power consumption. However, such architectures will lead to the challenging channel estimation [6].

To date, several channel estimation schemes [6]–[8] have been proposed for mmWave massive MIMO with hybrid precoding. Specifically, the reference signal was designed in [6] for the estimation of angles of arrival or departure (AoA/AoD), but it assumes the discrete AoA/AoD. [8] can directly estimate the singular subspace with Krylov subspace method, but multiple amplify-and-forward operations between the transmitter and receiver will introduce much noise. More importantly, existing schemes [6]–[8] only consider the narrow-band flat fading channel model, while the practical mmWave channels exhibit the broad-band frequency-selective fading (FSF) due to the very large bandwidth and different delays of multipath [9].

To this end, this letter proposes a multi-user uplink channel estimation scheme for mmWave massive MIMO systems, where the broad-band FSF channel is converted to multiple parallel narrow-band flat fading channels when OFDM is considered. Specifically, the mmWave channels exhibit the obviously angle-domain sparsity due to the much higher path loss

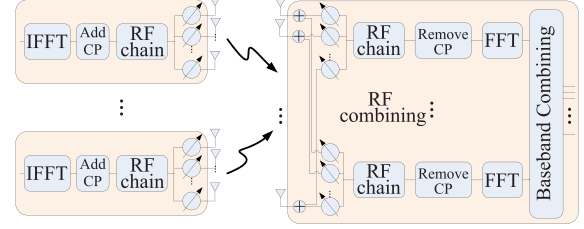


Fig. 1. Illustration of a broad-band mmWave massive MIMO system.

for non-line-of-sight (NLOS) paths than that for line-of-sight (LOS) paths [7]. Moreover, this sparsity is almost unchanged within the system bandwidth according to our derivation. By exploiting such angle-domain structured sparsity of mmWave FSF channels, we propose a distributed compressive sensing (DCS)-based channel estimation scheme, where both the transmit pilot signal and receive channel estimation algorithm are elaborated under the DCS theory for improved performance. Moreover, by using the grid matching pursuit strategy with adaptive measurement matrix, the proposed algorithm can solve the power leakage problem caused by the continuous AoA/AoD. To the best of our knowledge, this is the first paper to investigate the FSF channel estimation for mmWave massive MIMO systems.

**Notation:** The boldface lower and upper-case symbols denote column vectors and matrices, respectively. The Moore-Penrose inversion, transpose, conjugate transpose, integer ceiling, and expectation operators are given by  $(\cdot)^\dagger$ ,  $(\cdot)^T$ ,  $(\cdot)^*$ ,  $\lceil \cdot \rceil$ , and  $E\{\cdot\}$ , respectively.  $|\Gamma|$  is the cardinality of the set  $\Gamma$ . The support set of the vector  $\mathbf{a}$  is denoted by  $\text{supp}\{\mathbf{a}\}$ .  $\otimes$  is the Kronecker product, and  $\text{vect}(\cdot)$  is the vectorization operation according to the columns of the matrix.  $[\mathbf{a}]_i$  denotes the  $i$ th entry of the vector  $\mathbf{a}$ , and  $[\mathbf{A}]_{i,j}$  denotes the  $i$ th-row and  $j$ th-column element of the matrix  $\mathbf{A}$ .

## II. SYSTEM MODEL

We consider a typical mmWave massive MIMO-OFDM system over FSF channels as shown in Fig. 1 [9], where the base station (BS) employs  $N_a^{\text{BS}}$  antennas but only  $N_{\text{RF}}^{\text{BS}}$  RF chains with  $N_a^{\text{BS}} \gg N_{\text{RF}}^{\text{BS}} = K$  to support  $K$  user equipments (UEs), and each UE has  $N_a^{\text{UE}}$  antennas but only  $N_{\text{RF}}^{\text{UE}}$  RF chain with  $N_a^{\text{UE}} \gg N_{\text{RF}}^{\text{UE}} = 1$ . The hybrid analog-digital precoding at the BS can be used to realize the spatial multiplexing with low hardware cost and energy consumption [9]. Particularly, the uplink FSF channel associated with the  $k$ th user in the delay domain can be modeled as [9]

$$\mathbf{H}_k^d(\tau) = \sum_{l=0}^{L_k-1} \mathbf{H}_{l,k}^d \delta(\tau - \tau_{l,k}), \quad (1)$$

where  $L_k$  is the number of multipath,  $\tau_{l,k}$  is the delay of the  $l$ th path,  $\mathbf{H}_{l,k}^d \in \mathbb{C}^{N_a^{\text{BS}} \times N_a^{\text{UE}}}$  is given by

$$\mathbf{H}_{l,k}^d = \alpha_{l,k} \mathbf{a}_{\text{BS}}(d \sin(\theta_{l,k})/\lambda) \mathbf{a}_{\text{UE}}^*(d \sin(\phi_{l,k})/\lambda), \quad (2)$$

Manuscript received March 20, 2016; accepted April 11, 2016. Date of publication April 20, 2016; date of current version June 8, 2016. This work was supported by the Beijing Natural Science Foundation (Grant No. 4142027), the National Natural Science Foundation of China (Grant Nos. 61271266 and 61571270), the International Science & Technology Cooperation Program of China (Grant No. 2015DFG12760), and the Foundation of Shenzhen government. The associate editor coordinating the review of this letter and approving it for publication was K. J. Kim. (Corresponding author: Linglong Dai.)

The authors are with the Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: gaozhen010375@foxmail.com; daill@tsinghua.edu.cn; zcwang@tsinghua.edu.cn).

Digital Object Identifier 10.1109/LCOMM.2016.2555299

$\alpha_{l,k}$  is the complex gain of the  $l$ th path, and  $\theta_{l,k} \in [0, 2\pi]$  and  $\phi_{l,k} \in [0, 2\pi]$  are azimuth AoA/AoD if we consider the typical uniform linear array (ULA). For path gains, we consider Rician fading channels consisting of one LOS path (the 0th path) and  $L_k - 1$  NLOS paths (the  $l$ th path for  $1 \leq l \leq L_k - 1$ ), where path gains follow the mutually independent complex Gaussian distribution with zero means, and  $K_{\text{factor}}$  denotes the ratio between the power of LOS path and the power of NLOS paths. In addition,

$$\begin{aligned} \mathbf{a}_{\text{BS}} \left( \frac{d \sin(\theta_{l,k})}{\lambda} \right) &= \left[ e^{j2\pi n_{\text{BS}} d \sin(\theta_{l,k})/\lambda} \right]_{n_{\text{BS}} \in [0, \dots, N_a^{\text{BS}}-1]}^T \\ \mathbf{a}_{\text{UE}} \left( \frac{d \sin(\phi_{l,k})}{\lambda} \right) &= \left[ e^{j2\pi n_{\text{UE}} d \sin(\phi_{l,k})/\lambda} \right]_{n_{\text{UE}} \in [0, \dots, N_a^{\text{UE}}-1]}^T \end{aligned} \quad (3)$$

are steering vectors at the BS and the  $k$ th user, respectively, where  $\lambda$  is wavelength and  $d$  is the antenna spacing.

### III. DCS-BASED CHANNEL ESTIMATION SCHEME

#### A. Uplink Pilot Training

We consider that the training sequences used for channel estimation adopt OFDM to combat the FSF channels, where lengths of cyclic prefix (CP) and discrete Fourier transform (DFT) are  $L_{\text{CP}} > (\max\{\tau_{l,k}\}_{l=0, k=1}^{L_k-1, K} - \min\{\tau_{l,k}\}_{l=0, k=1}^{L_k-1, K}) f_s$  and  $P > L_{\text{CP}}$ , respectively, where  $f_s$  is the sampling rate. At the BS, after the CP removal and DFT operation, the received signal at the  $p$ th ( $1 \leq p \leq P$ ) subcarrier of the  $t$ th OFDM symbol in the frequency domain can be expressed as

$$\mathbf{r}_p^{(t)} = (\mathbf{Z}_{\text{RF}}^{(t)} \mathbf{Z}_{\text{BB},p}^{(t)})^* \sum_{k=1}^K \mathbf{H}_{p,k}^f \mathbf{F}_{\text{RF},k}^{(t)} \mathbf{F}_{\text{BB},p,k}^{(t)} \mathbf{s}_{p,k}^{(t)} + \mathbf{v}_p^{(t)}, \quad (4)$$

where  $\mathbf{r}_p^{(t)} \in \mathbb{C}^{N_{\text{RF}}^{\text{BS}} \times 1}$  is the received signal dedicated to the  $p$ th pilot subcarrier in the  $t$ th OFDM symbol,  $\mathbf{Z}_{\text{BB},p}^{(t)} \in \mathbb{C}^{N_a^{\text{BS}} \times N_{\text{RF}}^{\text{BS}}}$  is the digital combining matrix,  $\mathbf{Z}_{\text{RF}}^{(t)} \in \mathbb{C}^{N_a^{\text{BS}} \times N_{\text{RF}}^{\text{BS}}}$  is the RF combining matrix,  $\mathbf{Z}_p^{(t)} = \mathbf{Z}_{\text{RF}}^{(t)} \mathbf{Z}_{\text{BB},p}^{(t)} \in \mathbb{C}^{N_a^{\text{BS}} \times N_{\text{RF}}^{\text{BS}}}$  is the composite combining matrix at the BS,

$$\begin{aligned} \mathbf{H}_{p,k}^f &= \sum_{l=0}^{L_k-1} \mathbf{H}_{l,k}^d e^{-j2\pi f_s \tau_{l,k} p/P} \\ &= \sum_{l=0}^{L_k-1} \alpha_{l,k} e^{-\frac{j2\pi f_s \tau_{l,k} p}{P}} \mathbf{a}_{\text{BS}} \left( \frac{d \sin(\theta_{l,k})}{\lambda} \right) \mathbf{a}_{\text{UE}}^* \left( \frac{d \sin(\phi_{l,k})}{\lambda} \right), \end{aligned} \quad (5)$$

denotes the frequency-domain channel matrix associated with the  $p$ th pilot subcarrier for the  $k$ th UE,  $\mathbf{F}_{\text{RF},k}^{(t)} \in \mathbb{C}^{N_a^{\text{UE}} \times N_{\text{RF}}^{\text{UE}}}$ ,  $\mathbf{F}_{\text{BB},p,k}^{(t)} \in \mathbb{C}^{N_{\text{RF}}^{\text{UE}} \times N_{\text{RF}}^{\text{UE}}}$ ,  $\mathbf{s}_{p,k}^{(t)} \in \mathbb{C}^{N_{\text{RF}}^{\text{UE}} \times 1}$  are the RF precoding matrix, digital precoding matrix, and transmitted training sequence for the  $k$ th UE, respectively,  $\mathbf{f}_{p,k}^{(t)} = \mathbf{F}_{\text{RF},k}^{(t)} \mathbf{F}_{\text{BB},p,k}^{(t)} \mathbf{s}_{p,k}^{(t)} \in \mathbb{C}^{N_a^{\text{UE}} \times 1}$  is considered as the pilot signal transmitted by the  $k$ th user, and  $\mathbf{v}_p^{(t)}$  is the additive white Gaussian noise (AWGN) at the BS. Note that RF precoding/combining is the same for all subcarriers, since the RF phase shifter network can provide constant phase shift response over a wide frequency range [9].

Since the path loss for NLOS paths is much larger than that for LOS paths in mmWave systems, the mmWave channels appear the obvious sparsity in the angular domain, which

indicates small  $L_k$  and large  $K_{\text{factor}}$  in mmWave systems, e.g.,  $L_k = 4$  and  $K_{\text{factor}} = 20$  dB [2]. Hence, we can transform the frequency-domain channel matrix  $\mathbf{H}_{p,k}^f$  in (5) into the sparse angle-domain channel matrix  $\mathbf{H}_{p,k}^a$  as [7]

$$\mathbf{H}_{p,k}^a = \mathbf{A}_{\text{BS}}^* \mathbf{H}_{p,k}^f \mathbf{A}_{\text{UE}}, \quad (6)$$

where  $\mathbf{A}_{\text{BS}} \in \mathbb{C}^{N_a^{\text{BS}} \times N_a^{\text{BS}}}$  and  $\mathbf{A}_{\text{UE}} \in \mathbb{C}^{N_a^{\text{UE}} \times N_a^{\text{UE}}}$  are the DFT matrices by quantizing the virtual angular domain with the resolutions of  $1/N_a^{\text{BS}}$  at the BS and  $1/N_a^{\text{UE}}$  at the user, respectively. By vectorizing  $\mathbf{H}_{p,k}^f$ , we can further obtain

$$\mathbf{h}_{p,k}^f = \text{vect}(\mathbf{H}_{p,k}^f) = \mathbf{A} \mathbf{h}_{p,k}^a, \quad (7)$$

where  $\mathbf{A} = (\mathbf{A}_{\text{UE}}^*)^T \otimes \mathbf{A}_{\text{BS}}$  and  $\mathbf{h}_{p,k}^a = \text{vect}(\mathbf{H}_{p,k}^a)$ . Due to the sparsity of  $\mathbf{H}_{p,k}^a$ , only a minority of elements of  $\mathbf{h}_{p,k}^a$  dominate the majority of the channel energy, and thus we have

$$|\Theta_{p,k}| = |\text{supp}\{\mathbf{h}_{p,k}^a\}| = S_k \ll N_a^{\text{BS}} N_a^{\text{UE}}, \quad (8)$$

where  $\Theta_{p,k}$  is the support set, and  $S_k$  is the sparsity level in the angular domain.

According to (6)-(8), (4) can be further expressed as

$$\begin{aligned} \mathbf{r}_p^{(t)} &= (\mathbf{Z}_p^{(t)})^* \sum_{k=1}^K \mathbf{A}_{\text{BS}} \mathbf{H}_{p,k}^a \mathbf{A}_{\text{UE}}^* \mathbf{f}_{p,k}^{(t)} + \mathbf{v}_p^{(t)} \\ &= (\mathbf{Z}_p^{(t)})^* \mathbf{A}_{\text{BS}} \tilde{\mathbf{H}}_p^a \tilde{\mathbf{A}}_{\text{UE}}^* \tilde{\mathbf{f}}_p^{(t)} + \mathbf{v}_p^{(t)} \\ &= \left( (\tilde{\mathbf{A}}_{\text{UE}}^* \tilde{\mathbf{f}}_p^{(t)})^T \otimes (\mathbf{Z}_p^{(t)})^* \mathbf{A}_{\text{BS}} \right) \text{vect}(\tilde{\mathbf{H}}_p^a) + \mathbf{v}_p^{(t)} \\ &= \Psi_p^{(t)} \tilde{\mathbf{h}}_p^a + \mathbf{v}_p^{(t)}, \end{aligned} \quad (9)$$

where

$$\begin{aligned} \tilde{\mathbf{H}}_p^a &= [\mathbf{H}_{p,1}^a, \mathbf{H}_{p,2}^a, \dots, \mathbf{H}_{p,K}^a] \in \mathbb{C}^{N_a^{\text{BS}} \times K N_a^{\text{UE}}}, \\ \tilde{\mathbf{A}}_{\text{UE}}^* &= \text{diag}\{\mathbf{A}_{\text{UE}}^*, \mathbf{A}_{\text{UE}}^*, \dots, \mathbf{A}_{\text{UE}}^*\} \in \mathbb{C}^{K N_a^{\text{UE}} \times K N_a^{\text{UE}}}, \\ \tilde{\mathbf{f}}_p^{(t)} &= [(\mathbf{f}_{p,1}^{(t)})^T, (\mathbf{f}_{p,2}^{(t)})^T, \dots, (\mathbf{f}_{p,K}^{(t)})^T]^T \in \mathbb{C}^{K N_a^{\text{UE}} \times 1}, \\ \tilde{\mathbf{h}}_p^a &= \text{vect}(\tilde{\mathbf{H}}_p^a) \in \mathbb{C}^{K N_a^{\text{BS}} N_a^{\text{UE}} \times 1}, \\ \Psi_p^{(t)} &= (\tilde{\mathbf{A}}_{\text{UE}}^* \tilde{\mathbf{f}}_p^{(t)})^T \otimes (\mathbf{Z}_p^{(t)})^* \mathbf{A}_{\text{BS}} \in \mathbb{C}^{N_{\text{RF}}^{\text{BS}} \times K N_a^{\text{BS}} N_a^{\text{UE}}}. \end{aligned} \quad (10)$$

Furthermore, by jointly using the received pilot signals in  $G$  successive OFDM symbols, we can obtain

$$\tilde{\mathbf{r}}_p = \tilde{\Psi}_p \tilde{\mathbf{h}}_p^a + \tilde{\mathbf{v}}_p, \quad (11)$$

where  $\tilde{\mathbf{r}}_p = [(\mathbf{r}_p^{(1)})^T, (\mathbf{r}_p^{(2)})^T, \dots, (\mathbf{r}_p^{(G)})^T]^T \in \mathbb{C}^{G N_{\text{RF}}^{\text{BS}} \times 1}$  is the aggregate received signal,  $\tilde{\Psi}_p = [(\Psi_p^{(1)})^T, (\Psi_p^{(2)})^T, \dots, (\Psi_p^{(G)})^T]^T \in \mathbb{C}^{G N_{\text{RF}}^{\text{BS}} \times K N_a^{\text{BS}} N_a^{\text{UE}}}$  is the aggregate measurement matrix, and  $\tilde{\mathbf{v}}_p = [(\mathbf{v}_p^{(1)})^T, (\mathbf{v}_p^{(2)})^T, \dots, (\mathbf{v}_p^{(G)})^T]^T$  is aggregate AWGN.

#### B. DCS-Based Channel Estimation

To accurately estimate channels from (11),  $G$  in conventional algorithms, is heavily dependent on the dimension of  $\tilde{\mathbf{h}}_p^a$ , i.e.,  $K N_a^{\text{UE}} N_a^{\text{BS}}$ . Usually,  $G N_{\text{RF}}^{\text{BS}} \geq K N_a^{\text{UE}} N_a^{\text{BS}}$  is required, which leads  $G$  to be much larger than the channel coherence time [7]. Fortunately, the sparsity of mmWave massive MIMO channels motivates us to leverage the CS theory to estimate channels with much reduced pilot overhead. Moreover, according to (5), it can be observed that  $\{\mathbf{H}_{p,k}^f\}_{p=1}^P$

**Algorithm 1** Proposed DGMP Algorithm

**Input:** Received signals  $\tilde{\mathbf{r}}_p$  and sensing matrices  $\tilde{\Psi}_p$  in (11)  $\forall p$ , AoA/AoD resolution factor  $J$ , and error threshold  $\varepsilon$ .

**Output:** The steering vector estimation of  $k$ th UE's LOS path  $\hat{\mathbf{a}}_{\text{BS}}^{k,\text{LOS}}$  and  $\hat{\mathbf{a}}_{\text{UE}}^{k,\text{LOS}}$ , and the estimation set of path gains  $\hat{\alpha} \in \mathbb{C}^{1 \times K}$ , where  $[\hat{\alpha}]_k$  denotes the gain estimate of  $k$ th UE's LOS path.

• **Step 1 (Initialization)** The residue  $\mathbf{b}_p = \tilde{\mathbf{r}}_p$ , the iteration index  $k = 1$ ,  $[\tilde{\Psi}_p]_{:,j} = [\tilde{\Psi}_p]_{:,j} / \left\| [\tilde{\Psi}_p]_{:,j} \right\|_2$  for  $1 \leq j \leq KN_a^{\text{UE}}N_a^{\text{BS}}$ ,  $\forall p$ , and the matrix  $\Xi_p$  and set  $\mathcal{K}$  are set to be empty.

• **Step 2 (Estimate steering vectors and gains of  $K$  UEs' LOS paths)** for  $k \leq K$  do

1.  $\rho = \arg \max_{\tilde{\rho}} \left\{ \sum_{p=1}^P \left\| [(\tilde{\Psi}_p)^* \mathbf{b}_p]_{\tilde{\rho}} \right\|_2^2, [\tilde{\rho}/(N_a^{\text{UE}}N_a^{\text{BS}})] \notin \mathcal{K} \right\}$ ;
2.  $\tilde{k} = \lceil \rho/(N_a^{\text{UE}}N_a^{\text{BS}}) \rceil$ ,  $\mathcal{K} = \mathcal{K} \cup \tilde{k}$ ;
3.  $\tilde{\mathbf{r}}_p = [\tilde{\Psi}_p]_{:,(k-1)N_a^{\text{BS}}N_a^{\text{UE}}+1:\tilde{k}N_a^{\text{BS}}N_a^{\text{UE}}}$ ;  
**repeat**
4.  $\rho = \arg \max_{\tilde{\rho}} \left\{ \sum_{p=1}^P \left\| [(\tilde{\mathbf{r}}_p)^* \mathbf{b}_p]_{\tilde{\rho}} \right\|_2^2 \right\}$ ;
5.  $\beta_{\text{last}} = \beta$ ,  $\beta = \sum_{p=1}^P \left\| [(\tilde{\mathbf{r}}_p)^* \mathbf{b}_p]_{\rho} \right\|_2^2$ ;
6.  $n^{\text{UE}} = \lceil \rho/N_a^{\text{BS}} \rceil$ ,  $n^{\text{BS}} = \rho - (n^{\text{UE}} - 1)N_a^{\text{BS}}$ ;
7.  $\tilde{\mathbf{A}}_{\text{UE}} = [\mathbf{a}_{\text{UE}}(n^{\text{UE}} + \frac{j_{\text{UE}}}{2J})/N_a^{\text{UE}}]_{j_{\text{UE}} \in [-J, -J+1, \dots, J]}$ ;
8.  $\tilde{\mathbf{A}}_{\text{BS}} = [\mathbf{a}_{\text{BS}}(n^{\text{BS}} + \frac{j_{\text{BS}}}{2J})/N_a^{\text{BS}}]_{j_{\text{BS}} \in [-J, -J+1, \dots, J]}$ ;
9.  $\tilde{\Upsilon}_p^{(i)} = (\tilde{\mathbf{A}}_{\text{UE}}^* \mathbf{f}_{p,\tilde{k}}^{(i)})^T \otimes (\mathbf{Z}_p^{(i)})^* \tilde{\mathbf{A}}_{\text{BS}}$ ;
10.  $\tilde{\Upsilon}_p = [(\tilde{\Upsilon}_p^{(1)})^T, (\tilde{\Upsilon}_p^{(2)})^T, \dots, (\tilde{\Upsilon}_p^{(G)})^T]^T$ ;
11.  $[\tilde{\Upsilon}_p]_{:,j} = [\tilde{\Upsilon}_p]_{:,j} / \left\| [\tilde{\Upsilon}_p]_{:,j} \right\|_2$ ,  $1 \leq j \leq (2J-1)^2$ ,  $\forall p$ ;
12.  $\eta = \arg \max_{\tilde{\eta}} \left\{ \sum_{p=1}^P \left\| [(\tilde{\Upsilon}_p)^* \mathbf{b}_p]_{\tilde{\eta}} \right\|_2^2 \right\}$ ;
13.  $m^{\text{UE}} = \lceil \eta/(2J-1) \rceil$ ,  $m^{\text{BS}} = \eta - (m^{\text{UE}} - 1)(2J-1)$ ;
14.  $\tilde{\mathbf{A}}_{\text{UE}} = [\mathbf{a}_{\text{UE}}(n^{\text{UE}} + \frac{-J+m^{\text{UE}}-1}{2J})/N_a^{\text{UE}}]_{n^{\text{UE}} \in [0,1,\dots,N_a^{\text{UE}}-1]}$ ;
15.  $\tilde{\mathbf{A}}_{\text{BS}} = [\mathbf{a}_{\text{BS}}(n^{\text{BS}} + \frac{-J+m^{\text{BS}}-1}{2J})/N_a^{\text{BS}}]_{n^{\text{BS}} \in [0,1,\dots,N_a^{\text{BS}}-1]}$ ;
16.  $\Upsilon_p^{(i)} = (\tilde{\mathbf{A}}_{\text{UE}}^* \mathbf{f}_{p,\tilde{k}}^{(i)})^T \otimes (\mathbf{Z}_p^{(i)})^* \tilde{\mathbf{A}}_{\text{BS}}$ ;
17.  $\Upsilon_p = [(\Upsilon_p^{(1)})^T, (\Upsilon_p^{(2)})^T, \dots, (\Upsilon_p^{(G)})^T]^T$ ;
18.  $[\Upsilon_p]_{:,j} = [\Upsilon_p]_{:,j} / \left\| [\Upsilon_p]_{:,j} \right\|_2$ ,  $1 \leq j \leq N_a^{\text{UE}}N_a^{\text{BS}}$ ,  $\forall p$ ;
- until**  $|\beta_{\text{last}} - \beta| < \varepsilon$
19.  $\hat{\mathbf{a}}_{\text{BS}}^{k,\text{LOS}} = \mathbf{a}_{\text{BS}}((n^{\text{BS}} + \frac{-J+m^{\text{BS}}-1}{2J})/N_a^{\text{BS}})$ ;
20.  $\hat{\mathbf{a}}_{\text{UE}}^{k,\text{LOS}} = \mathbf{a}_{\text{UE}}((n^{\text{UE}} + \frac{-J+m^{\text{UE}}-1}{2J})/N_a^{\text{UE}})$ ;
21.  $\Xi_p = [\Xi_p, [\Upsilon_p]_{:,j}]$ ,  $\hat{\alpha}_{\mathcal{K}} = (\Xi_p)^{\dagger} \tilde{\mathbf{r}}_p$ ,  $\mathbf{b}_p = \tilde{\mathbf{r}}_p - \alpha_{\mathcal{K}} \Xi_p$ ;
- end for**

share the same AoA/AoD, and thus  $\{\mathbf{h}_{p,k}^a\}_{p=1}^P$  obtained after (6) and (7) have the structured sparsity, i.e.,

$$\text{supp}\{\mathbf{h}_{1,k}^a\} = \text{supp}\{\mathbf{h}_{2,k}^a\} = \dots = \text{supp}\{\mathbf{h}_{P,k}^a\} = \Theta_k. \quad (12)$$

Specifically, given (11) and the sparse constraints of (8) and (12), the channels can be estimated with standard DCS tool. However, due to the continuous AoA/AoD and the limited angle-domain resolution of  $\mathbf{A}_{\text{BS}}$  and  $\mathbf{A}_{\text{UE}}$ , the sparsity of  $\mathbf{h}_p^a$  may be impaired due to the power leakage problem [7], which will result in the poor channel estimation performance.

To this end, we propose a distributed grid matching pursuit (DGMP) algorithm as listed in **Algorithm 1** including

outer loop and inner loop. In each iteration of outer loop (steps 2.1-2.3 and 2.19-2.21), according to correlation operation (step 2.1), the UE index  $\tilde{k}$  (step 2.2) and adaptive measurement matrix  $\tilde{\Upsilon}_p$  (step 2.3) associated with the most possible path are acquired and input to inner loop; according to the output of inner loop, the  $\tilde{k}$ th UE's transmit/receive steering vectors are acquired (steps 2.19-2.20), and  $|\mathcal{K}|$  UEs' LOS path gains and residue  $\mathbf{b}_p$  are updated (step 2.21). The iteration of outer loop stops when AoA/AoD and path gains of all  $K$  UEs' LOS paths are estimated. For inner loop (steps 2.4-2.18), the AoA/AoD estimation associated with the  $\tilde{k}$ th UE's LOS path is improved with the grid matching strategy. Specifically, according to the inputs  $\tilde{k}$  and  $\tilde{\Upsilon}_p$  from outer loop, AoA/AoD indices  $n^{\text{BS}}$  and  $n^{\text{UE}}$  of the most possible path are acquired (step 2.6), and the corresponding correlation value is recorded as  $\beta$  (step 2.5); we construct the local over-complete measurement matrix  $\tilde{\Upsilon}_p$  (steps 2.7-2.11), where the local resolution of AoA associated with the index  $n^{\text{BS}}$  and AoD associated with the index  $n^{\text{UE}}$  is increased by  $(2J-1)$  times; according to correlation operation (step 2.12), finer AoA/AoD indices  $m^{\text{BS}}$  and  $m^{\text{UE}}$  can be acquired (step 2.13); finally,  $\tilde{\Upsilon}_p$  is updated, where the grid of AoA/AoD candidates is adjusted according to  $m^{\text{BS}}$  and  $m^{\text{UE}}$  (step 2.14-2.18). The inner loop stops if  $|\beta_{\text{last}} - \beta| < \varepsilon$ .

With the joint process of  $\tilde{\Psi}_p$  and  $\mathbf{b}_p$  for  $1 \leq p \leq P$ , the DGMP algorithm exploits the structured sparsity for improved performance, which can be found in steps 2.1, 2.4, and 2.12. Moreover, the adaptive measurement matrix  $\tilde{\Upsilon}_p$  with grid matching pursuit strategy can achieve high resolution estimation of AoA/AoD. Additionally, the near-LOS mmWave channel property is exploited, where only  $K$  UEs' LOS paths are estimated. Compared to the adaptive CS algorithm [7] estimating single sparse narrow-band channel from single received signal, the proposed DGMP algorithm jointly estimates multiple sparse subchannels from multiple received signals. Moreover, the grid matching pursuit strategy (steps 2.4-2.18) with adaptive  $\tilde{\Upsilon}_p$  can solve the problem of power leakage caused by the continuous AoA/AoD, which is different from the classical DCS algorithms [10].

### C. Pilot Design According to DCS Theory

The measurement matrices  $\tilde{\Psi}_p$ ,  $\forall p$  in (11) are very important for guaranteeing the reliable channel estimation. Usually, we have  $GN_{\text{RF}}^{\text{BS}} \ll KN_a^{\text{UE}}N_a^{\text{BS}}$ . Since  $\tilde{\Psi}_p = [(\Psi_p^{(1)})^T, (\Psi_p^{(2)})^T, \dots, (\Psi_p^{(G)})^T]^T$ ,  $\Psi_p^{(i)} = (\tilde{\mathbf{A}}_{\text{UE}}^* \tilde{\mathbf{f}}_p^{(i)})^T \otimes (\mathbf{Z}_p^{(i)})^* \mathbf{A}_{\text{BS}}$ ,  $\tilde{\mathbf{A}}_{\text{UE}}^* = \text{diag}\{\mathbf{A}_{\text{UE}}^*, \mathbf{A}_{\text{UE}}^*, \dots, \mathbf{A}_{\text{UE}}^*\}$ , and  $\mathbf{A}_{\text{UE}}$ ,  $\mathbf{A}_{\text{BS}}$  are determined by the geometrical structure of the antenna arrays, both  $\{\mathbf{f}_{p,k}^{(i)}\}_{p=1,k=1,i=1}^{P,K,G}$  transmitted by the  $K$  users and  $\{\mathbf{Z}_p^{(i)}\}_{p=1,i=1}^{P,G}$  at the BS should be elaborated to guarantee the desired robust channel estimation.

According to [10], a measurement matrix whose elements follow an independent identically distributed (i.i.d.) Gaussian distribution can achieve the good performance for sparse signal recovery. Furthermore, diversifying measurement matrices  $\tilde{\Psi}_p$ ,  $\forall p$  can further improve the recovery performance of sparse signals according to DCS theory [10]. This enlightens us to appropriately design pilot signals for mmWave massive MIMO systems. Specifically, as discussed above,

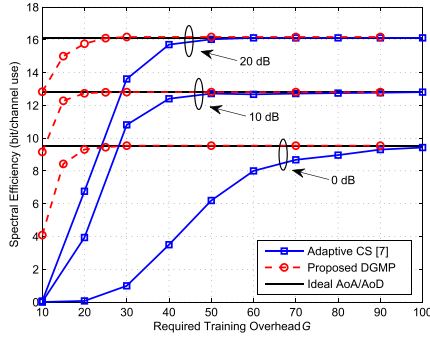


Fig. 2. Comparison of spectral efficiency performance of different channel estimation schemes against the training overhead  $G$  and SNR.

$\mathbf{Z}_p^{(t)} = \mathbf{Z}_{\text{RF}}^{(t)} \mathbf{Z}_{\text{BB},p}^{(t)}$ ,  $\mathbf{f}_{p,k}^{(t)} = \mathbf{F}_{\text{RF},k}^{(t)} \mathbf{F}_{\text{BB},p,k}^{(t)} \mathbf{s}_{p,k}^{(t)} = \mathbf{F}_{\text{RF},k}^{(t)} \tilde{\mathbf{s}}_{p,k}^{(t)}$  if we define  $\tilde{\mathbf{s}}_{p,k}^{(t)} = \mathbf{F}_{\text{BB},p,k}^{(t)} \mathbf{s}_{p,k}^{(t)}$  ( $1 \leq k \leq K$ ,  $1 \leq t \leq G$ ,  $1 \leq p \leq P$ ). Hence, we propose that each element of pilot signals is given by  $[\mathbf{Z}_{\text{RF}}^{(t)}]_{i_1,j_1} = e^{j\phi_{i_1,j_1}^1}$ ,  $1 \leq i_1 \leq N_a^{\text{BS}}$ ,  $1 \leq j_1 \leq N_{\text{RF}}^{\text{BS}}$ ,  $[\mathbf{F}_{\text{RF},k}^{(t)}]_{i_2,j_2} = e^{j\phi_{i_2,j_2}^2}$ ,  $1 \leq i_2 \leq N_a^{\text{UE}}$ ,  $1 \leq j_2 \leq N_{\text{RF}}^{\text{UE}}$ ,  $[\mathbf{Z}_{\text{BB},p}^{(t)}]_{i_4,j_4} = e^{j\phi_{i_4,j_4}^4}$ ,  $1 \leq i_4 \leq N_{\text{RF}}^{\text{BS}}$ ,  $1 \leq j_4 \leq N_{\text{RF}}^{\text{BS}}$ ,  $[\mathbf{s}_{p,k}^{(t)}]_{i_3} = e^{j\phi_{i_3,p,t,k}^3}$ ,  $1 \leq i_3 \leq N_{\text{RF}}^{\text{UE}}$ , where  $\phi_{i_1,j_1}^1$ ,  $\phi_{i_2,j_2}^2$ ,  $\phi_{i_3,p,t,k}^3$ , and  $\phi_{i_4,j_4,p,t}^4$  follow the i.i.d. uniform distribution  $\mathcal{U}[0, 2\pi)$ . Note that elements of RF precoding/combining matrices should meet the constant modulus property, and different subcarriers share the same RF precoding/combining. It is readily seen that the designed pilot signals guarantee that the elements of  $\tilde{\mathbf{\Psi}}_p$  obey the i.i.d. complex Gaussian distribution with zero mean. Moreover,  $\tilde{\mathbf{\Psi}}_p$  with different  $p$  are diversified. Hence, the proposed pilot signal design is optimal in terms of the joint recovery of multi-user's sparse angle-domain channels in the uplink.

#### IV. SIMULATION RESULTS

In simulations, carrier frequency  $f_c = 30$  GHz,  $f_s = 0.25$  GHz, the maximum delay spread  $\tau_{\text{max}} = 100$  ns,  $L_{\text{CP}} = \tau_{\text{max}} f_s = 25$ ,  $P = 32$ ,  $N_a^{\text{UE}} = 32$ ,  $N_{\text{RF}}^{\text{UE}} = 1$ ,  $N_a^{\text{BS}} = 128$ ,  $N_{\text{RF}}^{\text{BS}} = 4$ ,  $d = \lambda/2$ ,  $K_{\text{factor}} = 20$  dB,  $J = 10$ ,  $\varepsilon = 10^{-3}$ ,  $K = 4$ ,  $L_k = 4$  for  $1 \leq k \leq K$ . The case with the ideal AoD/AoD known at the BS is used as the performance benchmark for comparison. The adaptive CS-based channel estimation scheme [7] is also adopted for comparison.

Fig. 2 investigates the downlink spectral efficiency by using the hybrid analog-digital precoding scheme in [3], where the channels were estimated by the adaptive CS scheme [7] and the proposed DGMP algorithm. The case with ideal AoA/AoD was adopted as the performance bound. From Fig. 2, it can be observed that the adaptive CS scheme performs poorly, since it does not exploit the structured sparsity of mmWave massive MIMO channels. In contrast, the proposed DGMP algorithm can approach the performance bound with ideal AoA/AoD when  $G \geq 20$ . This is because the proposed DCS-based channel estimation scheme can leverage the angle-domain structured sparsity of mmWave FSF channels within the system bandwidth.

Fig. 3 compares the downlink bit error rate (BER) performance, where 16-QAM is used, and  $G$  for adaptive CS

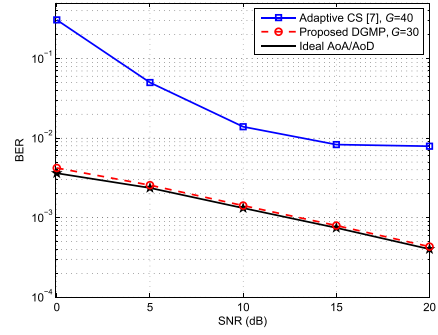


Fig. 3. BER performance comparison of different schemes.

algorithm and DGMP algorithm are 40, and 30, respectively. It can be observed that the proposed channel estimation scheme outperforms its counterpart with reduced training overhead, and its BER performance is very close to the performance bound with ideal AoA/AoD.

#### V. CONCLUSIONS

In this letter, we have proposed a DCS-based uplink channel estimation scheme for the multi-user mmWave massive MIMO with hybrid precoding, and it can effectively combat mmWave FSF channels. Specifically, we have designed an efficient pilot scheme and proposed a reliable DGMP algorithm under the framework of DCS theory, whereby the angle-domain structured sparsity of mmWave FSF channels is exploited for the reduced training overhead. Moreover, by using the grid matching pursuit strategy with adaptive measurement matrix, the proposed algorithm can effectively solve the power leakage problem. Our scheme can accurately estimate the FSF channels in mmWave massive MIMO with much lower pilot overhead than the existing scheme.

#### REFERENCES

- [1] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, Jr., "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, Apr. 2016.
- [2] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shaker, "mmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, Oct. 2015.
- [3] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
- [4] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis, and measurements," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3814–3827, Jul. 2013.
- [5] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Trans. Commun.*, vol. 64, no. 4, pp. 1557–1571, Apr. 2016.
- [6] S. Han, C.-L. I, Z. Xu, and S. Wang, "Reference signals design for hybrid analog and digital beamforming," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1191–1193, Jul. 2014.
- [7] 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.
- [8] H. Ghahouch, M. Bengtsson, T. Kim, and M. Skoglund, "Subspace estimation and decomposition for hybrid analog-digital millimeter-wave MIMO systems," in *Proc. IEEE 16th SPWAC*, Jun./Jul. 2015, pp. 395–399.
- [9] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, to be published.
- [10] Y. C. Eldar and G. Kutyniok, Eds., *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, May 2012.