Channel Estimation for mmWave Massive MIMO Based Access and Backhaul in Ultra-Dense Network

Zhen Gao, Linglong Dai, and Zhaocheng Wang

Tsinghua National Laboratory for Information Science and Technology (TNList),
Department of Electronic Engineering, Tsinghua University, Beijing 100084, P. R. China
Emails: gaozhen010375@foxmail.com

Abstract—Millimeter-wave (mmWave) massive MIMO used for access and backhaul in ultra-dense network (UDN) has been considered as the promising 5G technique. We consider such an heterogeneous network (HetNet) that ultra-dense small base stations (BSs) exploit mmWave massive MIMO for access and backhaul, while macrocell BS provides the control service with low frequency band. However, the channel estimation for mmWave massive MIMO can be challenging, since the pilot overhead to acquire the channels associated with a large number of antennas in mmWave massive MIMO can be prohibitively high. This paper proposes a structured compressive sensing (SCS)-based channel estimation scheme, where the angular sparsity of mmWave channels is exploited to reduce the required pilot overhead. Specifically, since the path loss for non-line-of-sight paths is much larger than that for line-of-sight paths, the mmWave massive channels in the angular domain appear the obvious sparsity. By exploiting such sparsity, the required pilot overhead only depends on the small number of dominated multipath. Moreover, the sparsity within the system bandwidth is almost unchanged, which can be exploited for the further improved performance. Simulation results demonstrate that the proposed scheme outperforms its counterpart, and it can approach the performance bound.

Index Terms—Millimeter-wave (mmWave), mmWave massive MIMO, compressive sensing (CS), hybrid precoding, channel estimation, access, backhaul, ultra-dense network (UDN), heterogeneous network (HetNet).

I. INTRODUCTION

It has been the consensus that future 5G networks should achieve the 1000-fold increase in system capacity [1]–[3]. To realize such an aggressive 5G version, millimeter-wave (mmWave) massive MIMO used for the access and backhaul in ultra-dense network (UDN) has been considered as a promising technique to enable gigabit-per-second user experience, seamless coverage, and green communication [2]. In this paper, we consider the heterogeneous network (HetNet) with the separation of control plane and data plane, as shown in Fig. 1. For such an HetNet, the macrocell base station (BS) provides the control signaling service for the large coverage area using conventional low frequency band, while the ultra-dense small BSs are specialized in data resources with limited coverage area for high-rate transmission, where the emerging mmWave massive MIMO technique is exploited for the access and backhaul [2], [4]. Moreover, the centralized radio access network (C-RAN) architecture can be integrated into the HetNet for the improved physical layer transmission, handover, scheduling, and etc., where the ultra-dense small cells are regarded as the
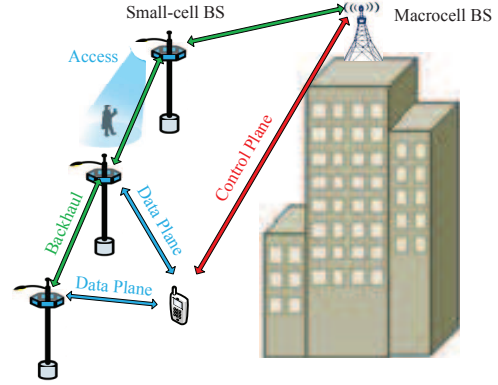

Fig. 1.   MmWave Massive MIMO based access and backhaul in UDN. transmission.

remote ratio head (RRH) and macrocell BS is considered as the baseband unit (BBU).

Due to the ultra-dense deployment, each user may receive the signal from multiple RRHs. To exploit the advantages of C-RAN architecture, the accurate channel state information (CSI) associated with multiple RRHs known at BBU is essential for the joint beamforming, scheduling, and cooperation among the ultra-dense small-cell BSs. However, how to require the reliable channel estimation with low overhead can be challenging, since the pilot overhead to estimate the channels associated with a large number of antennas in mmWave massive MIMO can be prohibitively high [5]. [6], [7] have proposed a multilevel codebook based joint channel estimation and beamforming for mmWave access and backhaul. However, this scheme only considers the mmWave multi-antenna systems with analog beamforming, which is limited to the point-to-point based single-stream transmission. To solve this problem, the mmWave massive MIMO has been emerging [2], [8]–[10], where the hybrid analog-digital beamforming scheme is proposed to support the multi-stream transmission with low hardware cost and energy consumption. In [9], an adaptive channel estimation has been proposed for mmWave massive MIMO. However, this scheme is limited to the single-cell scenario. Besides, [10] has proposed the reference signal design for the channel estimation in mmWave massive MIMO. However, the scheme fails to exploit the sparsity of mmWave channels, which may lead to a certain performance loss.

Recent study and experiments have shown that the mmWave massive MIMO channels appear the obviously sparsity in the

angular domain [2], [9], since the path loss for non-line-of-sight (NLOS) paths is much larger than that for line-of-sight (LOS) paths in mmWave [11], [12]. Moreover, since the spatial propagation characteristics of the mmWave channels within the system bandwidth are nearly unchanged, such sparsity is shared by subchannels of different subcarriers when the widely used orthogonal frequency-division multiplexing (OFDM) is considered. This phenomenon is referred to as the *spatially common sparsity within the system bandwidth* [13]. In this paper, we first propose the non-orthogonal pilot design at the small-cell BSs used for channel estimation. Furthermore, we propose a structured compressive sensing (SCS)-based sparse channel estimation scheme at the receiver to estimate the channels for the mmWave massive MIMO in UDN. Simulation results verify that the proposed scheme is superior to its counterpart, and it is capable of approaching the performance bound with low overhead.

Throughout our discussions, the boldface lower and uppercase symbols denote column vectors and matrices, respectively. The Moore-Penrose inversion, transpose, and conjugate transpose operators are given by $(\cdot)^\dagger$, $(\cdot)^{\mathrm{T}}$ and $(\cdot)^*$, respectively, while $\lceil \cdot \rceil$ is the integer ceiling operator. $(\cdot)^{-1}$ is the inverse operator. The ℓ_0-norm and ℓ_2-norm are given by $\|\cdot\|_0$ and $\|\cdot\|_2$, respectively, and $|\Gamma|$ is the cardinality of the set Γ. The support set of the vector $\mathbf{a}$ is denoted by $\mathrm{supp}\{\mathbf{a}\}$. The rank of $\mathbf{A}$ is denoted by $\mathrm{rank}\{\mathbf{A}\}$, while $\mathrm{E}\{\cdot\}$ is the expectation operator. $(\mathbf{a})_\Gamma$ denotes the entries of $\mathbf{a}$ whose indices are defined by Γ, while $(\mathbf{A})_\Gamma$ denotes a sub-matrix of $\mathbf{A}$ with column indices defined by Γ. $\otimes$ is the Kronecker product and $\mathrm{vect}(\cdot)$ is the vectorization operation according to the columns of the matrix. $[\mathbf{a}]_i$ denotes the ith entry of the vector $\mathbf{a}$, and $[\mathbf{A}]_{i,j}$ denotes the ith-row and jth-column element of the matrix $\mathbf{A}$.

II. SYSTEM MODEL

Conventionally, mmWave is considered to be not suitable for access due to its high path loss, Moreover, the signal-to-noise-ratio (SNR) before beamforming is conventionally considered to be low due to the high path loss, which leads to the challenging channel estimation in mmWave communications. However, for the UDN scenarios, we will clarify this misunderstanding by comparing the path loss in UDN working at 30 GHz and that in conventional cellular networks working at 3 GHz. Specifically, considering the multipath fading, signal dispersion, and other loss factors, the path loss component of Friis equation in decibel (dB) can be provided as [6],

$$\eta = 32.5 + 20\log_{10}(f_c) + 10\alpha\log_{10}(d) + (\alpha_o + \alpha_r)d, \quad (1)$$

where f_c (MHz) is the carrier frequency, α (dB/km) is the path loss exponent, d (km) is the link distance, α_o (dB/km) is the atmospheric attenuation coefficient, and α_r (dB/km) is the rain attenuation coefficient. For the conventional cellular systems with $f_c = 3$ GHz and $d = 1$ km for example, we have $\eta = 192.62$ dB, where $\alpha = 2.2$ dB/km is considered in urban scenarios [6], and the atmospheric attenuation and rain attenuation are ignored. By contrast, for UDN with $f_c = 30$ GHz, we have $\eta = 188.27$ dB with $d = 100$ m for backhaul link and $\eta = 161.78$ dB with $d = 30$ m for access link, where $\alpha = 2.2$ dB/km, $\alpha_o = 0.1$ dB/km, and $\alpha_r = 5$ dB/km when
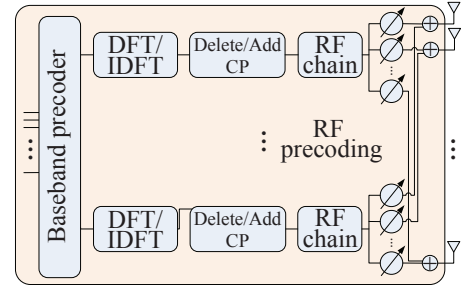

Fig. 2. Transceiver structure of the mmWave massive MIMO with analog phase shifter network.

the heavy rain with 25 mm/h is considered [1]. Due to the short link distance, the path loss in mmWave is even smaller than that in conventional cellular networks, which indicates the appropriate SNR for channel estimation in the mmWave access and backhaul even before beamforming.

On the other hand, mmWave massive MIMO has been emerging as the promising technique for access and backhaul [2], [8]–[10]. For the mmWave massive MIMO, as shown in Fig. 2, the number of baseband chains is larger than one but far smaller than that of the employed antennas, and the hybrid analog and digital beamforming can be exploited for the improved spatial multiplexing with low hardware cost and energy consumption. However, the large number of antennas lead to the challenging issue of channel estimation. Particularly, the delay-domain mmWave massive MIMO channel can be modeled as [6]

$$\mathbf{H}^d(\tau) = \sum_{l=0}^{L-1} \mathbf{H}_l^d \delta(\tau - \tau_l), \quad (2)$$

where L is the number of multipath, τ_l is the delay of the lth path, N_T, N_R are the numbers of antennas at the transmitter and receiver, respectively, and $\mathbf{H}_l^d \in \mathbb{C}^{N_R \times N_T}$ is given by

$$\mathbf{H}_l^d = \alpha_l \mathbf{a}_R(\theta_l)\mathbf{a}_T^*(\varphi_l), \quad (3)$$

with that α_l is the complex gain of the lth path, $\theta_l \in [0, 2\pi]$ and $\varphi_l \in [0, 2\pi]$ are azimuth angles of arrival or departure (AoA/AoD) if we consider the uniform linear array (ULA). In addition, $\mathbf{a}_R(\theta_l) = \left[1, e^{j2\pi d \sin(\theta_l)/\lambda}, \cdots, e^{j2\pi(N_T-1)d\sin(\theta_l)/\lambda}\right]^{\mathrm{T}}$ and $\mathbf{a}_T(\varphi_l) = \left[1, e^{j2\pi d \sin(\varphi_l)/\lambda}, \cdots, e^{j2\pi(N_R-1)d\sin(\varphi_l)/\lambda}\right]^{\mathrm{T}}$ are steering vectors at the receiver and transmitter, respectively, where λ and d are wavelength and antenna spacing, respectively.

Since the path loss for NLOS paths is much larger than that for LOS paths in mmWave, the mmWave channels appear the obviously sparsity in the angular domain. Here we consider frequency-domain subchannel $\mathbf{H}_n^f$ ($1 \leq n \leq N$) at the nth subcarrier, where N is the size of the OFDM symbol, and the relationship between the frequency-domain channel $\mathbf{H}_n^f$ and the delay-domain channel $\mathbf{H}^d(\tau)$ has been illustrated in our recent paper [13]. Moreover, we can obtain the sparse channel matrix in the angular domain $\mathbf{H}_n^a$ as [9]

$$\mathbf{H}_n^f = \mathbf{A}_R \mathbf{H}_n^a \mathbf{A}_T^*, \quad (4)$$

where $\mathbf{A}_T \in \mathbb{C}^{N_T \times N_T}$ and $\mathbf{A}_R \in \mathbb{C}^{N_R \times N_R}$ are the discrete Fourier transformation (DFT) matrices by quantizing the virtual angular domain with the resolutions of $2\pi/N_T$ at the transmitter

and $2\pi/N_R$ at the receiver, respectively. By vectorizing $\mathbf{H}_n^f$, we can further obtain

$$\mathbf{h}_n^f = \text{vect}\left(\mathbf{H}_n^f\right) = \left(\left(\mathbf{A}_T^*\right)^{\text{T}} \otimes \mathbf{A}_R\right) \text{vect}\left(\mathbf{H}_n^a\right) = \mathbf{A}\mathbf{h}_n^a, \quad (5)$$

where $\mathbf{A} = \left(\left(\mathbf{A}_T^*\right)^{\text{T}} \otimes \mathbf{A}_R\right)$ and $\mathbf{h}_n^a = \text{vect}\left(\mathbf{H}_n^a\right)$. Due to the sparsity of $\mathbf{h}_n^a$, we can obtain that

$$|\Theta_n| = |\text{supp}\{\mathbf{h}_n^a\}| = S_a \ll N_T N_R, \quad (6)$$

where Θ_n is the support set, and S_a is the sparsity level in the angular domain. Note that if we consider the quantized AoA/AoD with the same resolutions as $\mathbf{A}_T$ and $\mathbf{A}_R$, we have $S_a = L$ [9]. Since the spatial propagation characteristics of the channels within the system bandwidth are almost unchanged, $\{\mathbf{h}_n^a\}_{n=1}^N$ have the common sparsity, namely,

$$\text{supp}\{\mathbf{h}_1^a\} = \text{supp}\{\mathbf{h}_2^a\} = \cdots = \text{supp}\{\mathbf{h}_N^a\} = \Theta, \quad (7)$$

which is referred to as the spatially common sparsity within the system bandwidth.

III. SCS-BASED CHANNEL ESTIMATION SCHEME

In this section, we propose the SCS-based channel estimation scheme for access in UDN, which can be also used to estimate the channels for backhaul. The procedure of the proposed channel estimation and the associated processing is first summarized. *Step 1*: Under the control of the macrocell BS, several continuous time slots are specialized for channel estimation, where ultra-dense small-cell BSs (RRH) transmit non-orthogonal pilots. *Step 2*: The user uses the proposed SCS-based channel estimator to acquire the channels associated with multiple small cells nearby. *Step 3*: The estimated channels are first quantized and then fed back to the macrocell (BBU) for the centralized processing including the beamforming, scheduling, cooperation, and etc.

Due to the hybrid transceiver structure as shown in Fig. 2, we consider each user employs N_a^{US} antennas and $N_{\text{BB}}^{\text{US}}$ baseband chains with $N_a^{\text{US}} \gg N_{\text{BB}}^{\text{US}}$, while each small-cell has N_a^{BS} antennas and $N_{\text{BB}}^{\text{BS}}$ baseband chains with $N_a^{\text{BS}} \gg N_{\text{BB}}^{\text{BS}}$. Each user can receive the signal from M small-cell BSs nearby. In downlink channel estimation for one user, the received pilot signal at the ξ_pth ($1 \le p \le P$) subcarrier in the tth time slot can be expressed as

$$\begin{aligned}\mathbf{r}_p^{(t)} &= (\mathbf{Z}_p^{(t)})^* \sum_{m=1}^M \tilde{\mathbf{H}}_{p,m}^f \mathbf{f}_{p,m}^{(t)} + \mathbf{v}_p^{(t)} \\ &= (\mathbf{Z}_p^{(t)})^* \sum_{m=1}^M \mathbf{A}_R \tilde{\mathbf{H}}_{p,m}^a \mathbf{A}_T^* \mathbf{f}_{p,m}^{(t)} + \mathbf{v}_p^{(t)},\end{aligned} \quad (8)$$

where $\mathbf{r}_p^{(t)} \in \mathbb{C}^{N_{\text{BB}}^{\text{US}} \times 1}$ is the received signal dedicated to the pth pilot subcarrier in the tth time slot, $\mathbf{Z}_p^{(t)} = \mathbf{Z}_{\text{RF},p}^{(t)} \mathbf{Z}_{\text{BB},p}^{(t)} \in \mathbb{C}^{N_a^{\text{US}} \times N_{\text{BB}}^{\text{US}}}$ is the combining matrix at the receiver with $\mathbf{Z}_{\text{BB},p}^{(t)} \in \mathbb{C}^{N_{\text{BB}}^{\text{US}} \times N_{\text{BB}}^{\text{US}}}$ and $\mathbf{Z}_{\text{RF},p}^{(t)} \in \mathbb{C}^{N_a^{\text{US}} \times N_{\text{BB}}^{\text{US}}}$ the baseband and RF combining, respectively, $\tilde{\mathbf{H}}_{p,m}^f = \mathbf{H}_{\xi_p,m}^f$ and $\tilde{\mathbf{H}}_{p,m}^a = \mathbf{H}_{\xi_p,m}^a$ are frequency-domain and angular-domain channel matrices associated with the pth pilot subcarrier from the mth small-cell BS, respectively, $\Omega_\xi = \{\xi_1, \xi_2, \cdots, \xi_P\}$ is the index set of the pilot subcarriers, ξ_p for $1 \le p \le P$ denotes the subcarrier index dedicated to the pth pilot subcarrier, $\mathbf{f}_{p,m}^{(t)} = \mathbf{F}_{\text{RF},p}^{(t)} \mathbf{F}_{\text{BB},p}^{(t)} \mathbf{s}_p^{(t)} \in \mathbb{C}^{N_a^{\text{BS}} \times 1}$ is the pilot signal transmitted by the mth small-cell BS, with $\mathbf{F}_{\text{RF},p}^{(t)} \in \mathbb{C}^{N_a^{\text{BS}} \times N_{\text{BB}}^{\text{BS}}}$,

$\mathbf{F}_{\text{BB},p}^{(t)} \in \mathbb{C}^{N_{\text{BB}}^{\text{BS}} \times N_{\text{BB}}^{\text{BS}}}$, $\mathbf{s}_p^{(t)} \in \mathbb{C}^{N_{\text{BB}}^{\text{BS}} \times 1}$ the RF precoding, baseband precoding, and training sequence, respectively, and $\mathbf{v}_p^{(t)}$ is the additive white Gaussian noise (AWGN) at the user. Moreover, (8) can be simplified as

$$\begin{aligned}\mathbf{r}_p^{(t)} &= (\mathbf{Z}_p^{(t)})^* \mathbf{A}_R \bar{\mathbf{H}}_p^a \bar{\mathbf{A}}_T^* \bar{\mathbf{f}}_p^{(t)} + \mathbf{v}_p^{(t)} \\ &= \left(\left(\bar{\mathbf{A}}_T^* \bar{\mathbf{f}}_p^{(t)}\right)^{\text{T}} \otimes (\mathbf{Z}_p^{(t)})^* \mathbf{A}_R\right) \text{vect}\left(\bar{\mathbf{H}}_p^a\right) = \mathbf{\Phi}^{(t)} \bar{\mathbf{h}}_p^a,\end{aligned} \quad (9)$$

where $\bar{\mathbf{H}}_p^a = \left[\tilde{\mathbf{H}}_{p,1}^a, \tilde{\mathbf{H}}_{p,2}^a, \cdots, \tilde{\mathbf{H}}_{p,M}^a\right] \in \mathbb{C}^{N_a^{\text{US}} \times MN_a^{\text{BS}}}$, $\bar{\mathbf{A}}_T^* = \text{diag}\{\mathbf{A}_T^*, \mathbf{A}_T^*, \cdots, \mathbf{A}_T^*\} \in \mathbb{C}^{MN_a^{\text{BS}} \times MN_a^{\text{BS}}}$, $\bar{\mathbf{f}}_p^{(t)} = \left[\left(\mathbf{f}_{p,1}^{(t)}\right)^{\text{T}}, \left(\mathbf{f}_{p,2}^{(t)}\right)^{\text{T}}, \cdots, \left(\mathbf{f}_{p,M}^{(t)}\right)^{\text{T}}\right]^{\text{T}} \in \mathbb{C}^{MN_a^{\text{BS}} \times 1}$, $\bar{\mathbf{h}}_p^a = \text{vect}\left(\bar{\mathbf{H}}_p^a\right) \in \mathbb{C}^{MN_a^{\text{BS}} N_a^{\text{US}} \times 1}$, and $\mathbf{\Phi}^{(t)} = \left(\tilde{\mathbf{A}}_T^* \bar{\mathbf{f}}_p^{(t)}\right)^{\text{T}} \otimes (\mathbf{Z}_p^{(t)})^* \mathbf{A}_R \in \mathbb{C}^{N_{\text{BB}}^{\text{US}} \times MN_a^{\text{BS}} N_a^{\text{US}}}$. Due to the quasi-static property of the channel within the coherence time, the received signals in G successive time slots can be jointly exploited to acquire the downlink channel estimation at the user, which can be expressed as

$$\mathbf{r}_p^{[G]} = \mathbf{\Phi}_p^{[G]} \bar{\mathbf{h}}_p^a + \mathbf{v}_p^{[G]}, \quad (10)$$

where $\mathbf{r}_p^{[G]} = \left[\left(\mathbf{r}_p^{(1)}\right)^{\text{T}}, \left(\mathbf{r}_p^{(2)}\right)^{\text{T}}, \cdots, \left(\mathbf{r}_p^{(G)}\right)^{\text{T}}\right]^{\text{T}} \in \mathbb{C}^{GN_{\text{BB}}^{\text{US}} \times 1}$ is the aggregate received signal, $\mathbf{\Phi}^{[G]} = \left[\left(\mathbf{\Phi}^{(1)}\right)^{\text{T}}, \left(\mathbf{\Phi}^{(2)}\right)^{\text{T}}, \cdots, \left(\mathbf{\Phi}^{(G)}\right)^{\text{T}}\right]^{\text{T}} \in \mathbb{C}^{GN_{\text{BB}}^{\text{US}} \times MN_a^{\text{BS}} N_a^{\text{US}}}$ is the aggregate measurement matrix, and $\mathbf{v}_p^{[G]} = \left[\left(\mathbf{v}_p^{(1)}\right)^{\text{T}}, \left(\mathbf{v}_p^{(2)}\right)^{\text{T}}, \cdots, \left(\mathbf{v}_p^{(G)}\right)^{\text{T}}\right]^{\text{T}}$ is aggregate AWGN. The system's SNR is defined as $\text{SNR} = \text{E}\left\{\left\|\mathbf{\Phi}_p^{[G]} \bar{\mathbf{h}}_p^a\right\|_2^2\right\} / \text{E}\left\{\left\|\mathbf{v}_p^{[G]}\right\|_2^2\right\}$, according to (10).

To accurately estimate the channel from (10), the value of G used in conventional algorithms, such as the minimum mean square error (MMSE) algorithm, is heavily dependent on the dimension of $\bar{\mathbf{h}}_p^a$, i.e., $MN_a^{\text{US}} N_a^{\text{BS}}$. Usually, $GN_{\text{BB}}^{\text{US}} \ge MN_a^{\text{US}} N_a^{\text{BS}}$ is required, which leads G to be much larger than the coherence time and results in the poor channel estimation performance [5]. Moreover, to minimize the mean square error (MSE) of the channel estimate, $\mathbf{\Phi}_p^{[G]}$ should be a unitary matrix scaled by a transmit power factor [5]. Usually, $\mathbf{\Phi}_p^{[G]}$ is a diagonal matrix with equal-power diagonal elements. Such a pilot design is illustrated in Fig. 3 (a), which is called the time-domain orthogonal pilot. It should be pointed out that in MIMO-OFDM systems, to estimate the channel associated with one transmit antenna, P pilot subcarriers should be used, and usually $P = N_g$ is considered since $N_c = N/N_g$ adjacent subcarriers are correlated [5]. Hence the total pilot overhead to estimate the complete MIMO channel can be as large as $P_{\text{total}} = PG = N_g MN_a^{\text{US}} N_a^{\text{BS}}/N_{\text{BB}}^{\text{BS}}$. In this paper, we will propose an efficient non-orthogonal pilot scheme.

A. Non-Orthogonal Pilot for Downlink Channel Estimation

The proposed non-orthogonal pilot scheme is illustrated in Fig. 3 (b). Similar to the time-domain orthogonal pilot scheme, P subcarriers are dedicated to pilots in each OFDM symbol. However, the proposed scheme allows the non-orthogonal pilot
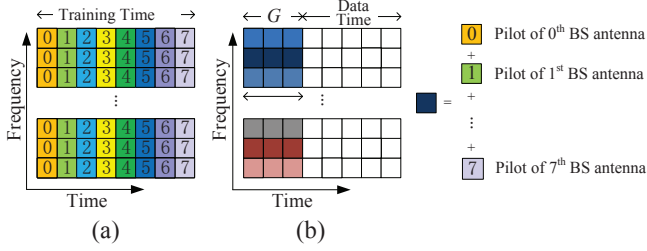
Fig. 3. (a) Orthogonal pilot [5], (b) Non-orthogonal pilot ($N_a^{\mathrm{BS}} = 8$).

signals associated with different BS antennas to occupy the completely identical frequency-domain subcarriers.

The orthogonal pilot based conventional designs usually require $GN_{\mathrm{BB}}^{\mathrm{US}} \geq MN_a^{\mathrm{US}}N_a^{\mathrm{BS}}$. By contrast, the proposed non-orthogonal pilot for SCS-based channel estimator is capable of providing the efficient compression and reliable recovery of sparse signals. Therefore, $GN_{\mathrm{BB}}^{\mathrm{US}}$ is mainly determined by $S_a \ll MN_a^{\mathrm{US}}N_a^{\mathrm{BS}}/N_{\mathrm{BB}}^{\mathrm{US}}$. The non-orthogonal pilot of the first stage is designed in advance, which will be discussed in Section IV-A. For the placement of pilot subcarriers, the widely used equi-spaced pilot is considered. It is worth pointing out that the pth pilot subcarrier is shared by the pilot signals of the N_a^{BS} transmit antennas of M small-cell BSs as illustrated in Fig. 3 (b).

B. SCS-Based Channel Estimation Scheme

Given the measurements (10), the CSI can be acquired by solving the following optimization

$$
\begin{aligned}
& \min_{\bar{\mathbf{h}}_p^a, 1 \leq p \leq P} \left(\sum_{p=1}^P \left\| \bar{\mathbf{h}}_p^a \right\|_0^2 \right)^{1/2} \\
& \text{s.t. } \mathbf{r}_p^{[G]} = \mathbf{\Phi}_p^{[G]} \bar{\mathbf{h}}_p^a, \ \forall p \ \text{ and } \left\{ \bar{\mathbf{h}}_p^a \right\}_{p=1}^P \\
& \text{share the common sparse support set.}
\end{aligned}
\tag{11}
$$

To solve the optimization problem (11), we adopt the structured sparsity adaptive matching pursuit (SSAMP) algorithm to reconstruct the sparse angular domain channels of multiple pilot subcarriers [16]. The SSAMP algorithm, listed in Algorithm 1, is used to solve the optimization (11) to simultaneously acquire multiple sparse channel vectors at different pilot subcarriers. This algorithm is developed from the SAMP algorithm [14]. Compared to the classical SAMP algorithm [14] which recovers one high-dimensional sparse signal from single low-dimensional received signal, the SSAMP algorithm can simultaneously recover multiple high-dimensional sparse signals with the common support set by jointly processing multiple low-dimensional received signals.

By using the SSAMP algorithm at the user, we can acquire $\widehat{\bar{\mathbf{h}}}_p^a$ for $1 \leq p \leq P$. Consequently, the actual pth pilot subchannel associated with the mth small-cell BS $\left\{ \tilde{\mathbf{H}}_{p,m}^f \right\}_{m=1,p=1}^{M,P}$ can be acquired.

IV. Performance Analysis

The performance analysis includes the non-orthogonal pilot design and the theoretical limit of the required time slot overhead for the SCS-based channel estimation scheme.

A. Non-Orthogonal Pilot Design for Multi-Cell mmWave Massive MIMO Systems

The measurement matrices $\mathbf{\Phi}_p^{[G]} \ \forall p$ in (10) are very important for guaranteeing the reliable channel estimation at

---

Algorithm 1 SSAMP Algorithm

Input: Noisy received signals $\mathbf{r}_p^{[G]}$ and sensing matrices $\mathbf{\Phi}_p^{[G]}$ in (11), $1 \leq p \leq P$; termination threshold p_{th}.

Output: Estimated channel vectors in the virtual angular domain at multiple pilot subcarriers $\widehat{\bar{\mathbf{h}}}_p^a, \ \forall p$.

1: $\mathcal{T} = 1$; $i = 1$; $j = 1$. % $\mathcal{T}$, i, j are the sparsity level of the current stage, iteration index, and stage index, respectively.

2: $\mathbf{c}_p = \mathbf{t}_p = \mathbf{c}_p^{\mathrm{last}} = \mathbf{0} \in \mathbb{C}^{M \times 1}, \ \forall p.$ % $\mathbf{c}_p$ and $\mathbf{t}_p$ are intermediate variables, and $\mathbf{c}_p^{\mathrm{last}}$ is the channel estimation of the last stage.

3: $\Omega^0 = \Gamma = \widetilde{\Gamma} = \Omega = \widetilde{\Omega} = \emptyset$; $l_{\min} = \tilde{l} = 0$. % Ω^i is the estimated support set in the ith iteration, Γ, $\widetilde{\Gamma}$, Ω, and $\widetilde{\Omega}$ are sets, $l_{\min}$ and $\tilde{l}$ denote the support indexes.

4: $\mathbf{b}_p^0 = \mathbf{r}_p^{[G]} \in \mathbb{C}^{G \times 1}, \ \forall p.$ % $\mathbf{b}_p^i$ is the residual of the ith iteration.

5: $\sum_{p=1}^P \left\| \mathbf{b}_p^{\mathrm{last}} \right\|_2^2 = +\infty.$ % $\mathbf{b}_p^{\mathrm{last}}$ is the residual of the last stage.

6: **repeat**

7: $\quad \mathbf{a}_p = \left(\mathbf{\Phi}_p^{[G]} \right)^* \mathbf{b}_p^{i-1}, \ \forall p.$ % Signal proxy is saved in $\mathbf{a}_p$.

8: $\quad \Gamma = \arg \max_{\widetilde{\Gamma}} \left\{ \sum_{p=1}^P \left\| (\mathbf{a}_p)_{\widetilde{\Gamma}} \right\|_2^2, \left| \widetilde{\Gamma} \right| = \mathcal{T} \right\}.$ % Identify support.

9: $\quad (\mathbf{t}_p)_{\Omega^{i-1} \cup \Gamma} = \left(\left(\mathbf{\Phi}_p^{[G]} \right)_{\Omega^{i-1} \cup \Gamma} \right)^\dagger \mathbf{r}_p^{[G]}, \ \forall p.$ % LS estimation.

10: $\quad \Omega = \arg \max_{\widetilde{\Omega}} \left\{ \sum_{p=1}^P \left\| (\mathbf{t}_p)_{\widetilde{\Omega}} \right\|_2^2, \left| \widetilde{\Omega} \right| = \mathcal{T} \right\}.$ % Prune support.

11: $\quad (\mathbf{c}_p)_\Omega = \left(\left(\mathbf{\Phi}_p^{[G]} \right)_\Omega \right)^\dagger \mathbf{r}_p^{[G]}, \ \forall p.$ % LS estimation.

12: $\quad \mathbf{b}_p = \mathbf{r}_p^{[G]} - \mathbf{\Phi}_p^{[G]} \mathbf{c}_p, \ \forall p.$ % Compute the residual.

13: $\quad l_{\min} = \arg \min_{\tilde{l}} \left\{ \sum_{p=1}^P \left\| [\mathbf{c}_p]_{\tilde{l}} \right\|_2^2, \tilde{l} \in \Omega \right\}.$

14: $\quad$ **if** $\sum_{p=1}^P \left\| [\mathbf{c}_p]_{l_{\min}} \right\|_2^2 / P < p_{\mathrm{th}}$ **then**

15: $\quad \quad$ Quit iteration.

16: $\quad$ **else if** $\sum_{p=1}^P \left\| \mathbf{b}_p^{\mathrm{last}} \right\|_2^2 < \sum_{p=1}^P \left\| \mathbf{b}_p \right\|_2^2$ **then**

17: $\quad \quad$ Quit iteration.

18: $\quad$ **else if** $\sum_{p=1}^P \left\| \mathbf{b}_p^{i-1} \right\|_2^2 \leq \sum_{p=1}^P \left\| \mathbf{b}_p \right\|_2^2$ **then**

19: $\quad \quad j = j+1$; $\mathcal{T} = j$; $\mathbf{c}_p^{\mathrm{last}} = \mathbf{c}_p$, $\mathbf{b}_p^{\mathrm{last}} = \mathbf{b}_p, \ \forall p.$

20: $\quad$ **else**

21: $\quad \quad \Omega^i = \Omega$; $\mathbf{b}_p^i = \mathbf{b}_p, \ \forall p$; $i = i+1.$

22: $\quad$ **end if**

23: **until** $\sum_{p=1}^P \left\| [\mathbf{c}_p]_{l_{\min}} \right\|_2^2 / P < p_{\mathrm{th}}$

24: $\widehat{\bar{\mathbf{h}}}_p^a = \mathbf{c}_p^{\mathrm{last}}, \ \forall p.$ % Obtain the final channel estimation.

---

the user. Usually, $GN_{\mathrm{BB}}^{\mathrm{US}} \ll MN_a^{\mathrm{US}}N_a^{\mathrm{BS}}$. Since $\mathbf{\Phi}^{[G]} = \left[\left(\mathbf{\Phi}^{(1)} \right)^{\mathrm{T}}, \left(\mathbf{\Phi}^{(2)} \right)^{\mathrm{T}}, \cdots, \left(\mathbf{\Phi}^{(G)} \right)^{\mathrm{T}} \right]^{\mathrm{T}}$, $\mathbf{\Phi}^{(t)} = \left(\tilde{\mathbf{A}}_T^* \tilde{\mathbf{f}}_p^{(t)} \right)^{\mathrm{T}} \otimes \left(\mathbf{Z}_p^{(t)} \right)^* \mathbf{A}_R$, $\bar{\mathbf{A}}_T^* = \mathrm{diag}\left\{ \mathbf{A}_T^*, \mathbf{A}_T^*, \cdots, \mathbf{A}_T^* \right\}$, and $\mathbf{A}_T$, $\mathbf{A}_R$ are determined by the geometrical structure of the antenna arrays, both $\left\{ \mathbf{f}_{p,m}^{(t)} \right\}_{p=1,m=1,t=1}^{P,M,G}$ transmitted by the M small-cell BSs and $\left\{ \mathbf{Z}_p^{(t)} \right\}_{p=1,t=1}^{P,G}$ at the user should be elaborated to guarantee the desired robust channel estimation.

In CS theory, restricted isometry property (RIP) is used to evaluate the quality of the measurement matrix, in terms of the reliable compression and reconstruction of sparse signals. It is proven in [15] that the measurement matrix with its elements following the independent and identically distributed (i.i.d.) complex Gaussian distributions satisfies the RIP and enjoys a satisfying performance in compressing and recovering sparse signals, which provides the viable pilot design guideline.

On the other hand, the optimization problem (11) is essentially different from the single-measurement-vector (SMV)

and multiple-measurement-vector (MMV) problems in CS[1]. Typically, the MMV has the better recovery performance than the SMV, due to the potential diversity from multiple sparse signals [15]. Intuitively, the recovery performance of multiple sparse signals with different measurement matrices, as defined in the generalized MMV (GMMV), should be better than that with the common measurement matrix as given in the MMV. This is because the further potential diversity can benefit from different measurement matrices for the GMMV. To prove this, we investigate the uniqueness of the solution to the GMMV problem. First, we introduce the concept of 'spark' and the ℓ_0-minimization based GMMV problem.

Definition 1. *[15] The smallest number of columns of $\mathbf{\Phi}$ which are linearly dependent is the spark of the given matrix $\mathbf{\Phi}$, denoted by* $\mathrm{spark}(\mathbf{\Phi})$.

Problem 1. $\min\limits_{\mathbf{x}_p, \forall p} \sum\limits_{p=1}^{P} \|\mathbf{x}_p\|_0^2$, s.t. $\mathbf{y}_p = \mathbf{\Phi}_p \mathbf{x}_p$, $\mathrm{supp}\{\mathbf{x}_p\} = \Xi$, $\forall p$.

For the above ℓ_0-minimization based GMMV problem, $\mathbf{x}_p$, $\mathbf{y}_p$, $\mathbf{\Phi}_p$ are the high-dimensional sparse signal, low-dimensional measurement signal, and measurement matrix, respectively. Furthermore, the following result can be obtained.

Theorem 1. *[16] For $\mathbf{\Phi}_p$, $1 \leq p \leq P$, whose elements obey an i.i.d. continuous distribution, there exist full rank matrices $\mathbf{\Psi}_p$ for $2 \leq p \leq P$ satisfying $(\mathbf{\Phi}_p)_\Xi = \mathbf{\Psi}_p (\mathbf{\Phi}_1)_\Xi$ if we select $(\mathbf{\Phi}_1)_\Xi$ as the bridge, where Ξ is the common support set. Consequently, $\mathbf{x}_p$ for $1 \leq p \leq P$ will be the unique solution to Problem 1 if*

$$2S < \mathrm{spark}(\mathbf{\Phi}_1) - 1 + \mathrm{rank}\{\widetilde{\mathbf{Y}}\}, \qquad (12)$$

where $\widetilde{\mathbf{Y}} = \left[\mathbf{y}_1 \ \mathbf{\Psi}_2^{-1} \mathbf{y}_2 \cdots \mathbf{\Psi}_P^{-1} \mathbf{y}_P\right]$.

From Theorem 1, it is clear that the achievable diversity gain introduced by diversifying measurement matrices and sparse vectors is determined by $\mathrm{rank}\{\widetilde{\mathbf{Y}}\}$. The larger $\mathrm{rank}\{\widetilde{\mathbf{Y}}\}$ is, the more reliable recovery of sparse signals can be achieved. Hence, compared to the SMV and MMV, more reliable recovery performance can be achieved by the proposed GMMV. For the special case that multiple sparse signals are identical, the MMV reduces to the SMV since $\mathrm{rank}(\mathbf{Y}) = 1$, and there is no diversity gain by introducing multiple identical sparse signals. However, the GMMV in this case can still achieve diversity gain which comes from diversifying measurement matrices.

According to the discussions above, a measurement matrix whose elements follow an i.i.d. Gaussian distribution satisfies the RIP. Furthermore, diversifying measurement matrices can further improve the recovery performance of sparse signals. This enlightens us to appropriately design pilot signals.

Specifically, as discussed above, we have $\mathbf{Z}_p^{(t)} = \mathbf{Z}_{\mathrm{RF},p}^{(t)} \mathbf{Z}_{\mathrm{BB},p}^{(t)}$, $\mathbf{f}_{p,m}^{(t)} = \mathbf{F}_{\mathrm{RF},p,m}^{(t)} \mathbf{F}_{\mathrm{BB},p,m}^{(t)} \mathbf{s}_{p,m}^{(t)} = \mathbf{F}_{\mathrm{RF},p,m}^{(t)} \widetilde{\mathbf{s}}_{p,m}^{(t)}$ by defining $\widetilde{\mathbf{s}}_{p,m}^{(t)} = \mathbf{F}_{\mathrm{BB},p,m}^{(t)} \mathbf{s}_{p,m}^{(t)}$ for $1 \leq m \leq M$ and $1 \leq t \leq G$,

[1]In CS, SMV recovers single high-dimensional sparse signal from its low-dimensional measurement signal, while MMV recovers multiple high-dimensional sparse signals with the common support set from multiple low-dimensional measurement signals with the identical measurement matrix.

and $1 \leq p \leq P$, each element of pilot signals is given by

$$\left[\mathbf{Z}_{\mathrm{RF},p}^{(t)}\right]_{i_1,j_1} = e^{j\phi_{i_1,j_1,t}^1}, 1 \leq i_1 \leq N_a^{\mathrm{US}}, 1 \leq j_1 \leq N_{\mathrm{BB}}^{\mathrm{US}}, \quad (13)$$

$$\left[\mathbf{F}_{\mathrm{RF},p}^{(t)}\right]_{i_2,j_2} = e^{j\phi_{i_2,j_2,t,m}^2}, 1 \leq i_2 \leq N_a^{\mathrm{BS}}, \leq j_2 \leq N_{\mathrm{BB}}^{\mathrm{BS}}, \quad (14)$$

$$\left[\mathbf{s}_p^{(t)}\right]_{i_3} = e^{j\phi_{i_3,p,t,m}^3}, 1 \leq i_3 \leq N_{\mathrm{BB}}^{\mathrm{BS}}, \quad (15)$$

$$\left[\mathbf{Z}_{\mathrm{BB},p}^{(t)}\right]_{i_4,j_4} = e^{j\phi_{i_4,j_4,p,t}^4}, 1 \leq i_4 \leq N_a^{\mathrm{BS}}, 1 \leq j \leq N_{\mathrm{BB}}^{\mathrm{BS}}, \quad (16)$$

where $\phi_{i_1,j_1,t}^1$, $\phi_{i_2,j_2,t,m}^2$, $\phi_{i_3,p,t,m}^3$, and $\phi_{i_4,j_4,p,t}^4$ follow the i.i.d. uniform distribution $\mathcal{U}[0, 2\pi]$. Note that $\mathbf{F}_{\mathrm{RF},1}^{(t)} = \mathbf{F}_{\mathrm{RF},p}^{(t)}$ and $\mathbf{Z}_{\mathrm{RF},1}^{(t)} = \mathbf{Z}_{\mathrm{RF},p}^{(t)}$, since different subcarriers share the same RF precoding/combining. It is readily seen that the designed pilot signals guarantee that the elements of $\mathbf{\Phi}_p^{[G]}$, obey the i.i.d. complex Gaussian distribution with zero mean. Moreover, $\mathbf{\Phi}_p^{[G]}$ with different p are diversified. Hence, the proposed pilot signal design is 'optimal', in terms of the reliable compression and recovery of sparse angular domain channels.

B. Required Time Slot Overhead for SCS-Based Channel Estimation

According to Theorem 1, for the optimization problem (11), $\widetilde{\mathbf{Y}} = \mathbf{\Phi}_1^{[G]} \mathbf{X}$ with $\widetilde{\mathbf{Y}} = \left[\mathbf{r}_1^{[G]} \ \mathbf{\Psi}_2^{-1} \mathbf{r}_2^{[G]} \cdots \mathbf{\Psi}_P^{-1} \mathbf{r}_P^{[G]}\right]$ and $\mathbf{X} = \left[\bar{\mathbf{h}}_1^a \ \bar{\mathbf{h}}_2^a \cdots \bar{\mathbf{h}}_P^a\right]$. Since $\left|\mathrm{supp}\{\bar{\mathbf{h}}_p^a\}\right| = S_a$, it is clear that

$$\mathrm{rank}\{\widetilde{\mathbf{Y}}\} \leq \mathrm{rank}\{\mathbf{X}\} \leq S_a. \qquad (17)$$

Moreover, as $\mathbf{\Phi}_1^{[G]} \in \mathbb{C}^{GN_{\mathrm{BB}}^{\mathrm{US}} \times M N_a^{\mathrm{BS}} N_a^{\mathrm{US}}}$,

$$\mathrm{spark}(\mathbf{\Phi}_1^{[G]}) \in \left\{2, 3, \cdots, GN_{\mathrm{BB}}^{\mathrm{US}} + 1\right\}. \qquad (18)$$

Substituting (17) and (18) into (12) yields $GN_{\mathrm{BB}}^{\mathrm{US}} \geq S_a + 1$. Therefore, the smallest required time slot overhead is $G = \left\lceil \frac{S_a+1}{N_{\mathrm{BB}}^{\mathrm{US}}} \right\rceil$. By increasing the number of measurement vectors P, the required time slot overhead for reliable channel estimation can be reduced, since more measurement matrices and sparse signals can increase $\mathrm{rank}\{\widetilde{\mathbf{Y}}\}$.

V. SIMULATION RESULTS

We consider the ULA-based mmWave massive MIMO system with $d = \lambda/2$. In the simulations, $f_c = 30\,\mathrm{GHz}$, bandwidth $B_s = 0.25\,\mathrm{GHz}$, $N_a^{\mathrm{US}} = 32$, $N_{\mathrm{BB}}^{\mathrm{US}} = 2$, $N_a^{\mathrm{BS}} = 512$, $N_{\mathrm{BB}}^{\mathrm{BS}} = 8$, $\tau_{\max} = 100\,\mathrm{ns}$, $N = P = 64 > \tau_{\max} B_s$, the user simultaneously estimates the channels associated with $M = 4$ small-cell BSs nearby, $L = 4$ for each link between the user and small-cell BS. For mmWave massive MIMO channels, we consider Rican fading consisting of one LOS path and $L - 1$ equal-power NLOS paths with $K_{\mathrm{factor}} = 10$ dB, where path gains follow the mutually independent complex Gaussian distribution with zero means, and K_{factor} denotes the ratio between the power of LOS path and the power of NLOS path. We set p_{th} to 0.06, 0.02, 0.01, 0.008, and 0.005, respectively, at the SNR of 10 dB, 15 dB, 20 dB, 25 dB and ≥ 30 dB. The oracle LS estimator was used as the benchmarks for the SCS-based channel estimation scheme. The adaptive orthogonal matching pursuit (OMP)-based channel estimation scheme [9] was also adopted for comparison.

Fig. 4 compares the MSE performance of the adaptive OMP scheme and the SSAMP algorithm, where $S_a = LM = 16$ was
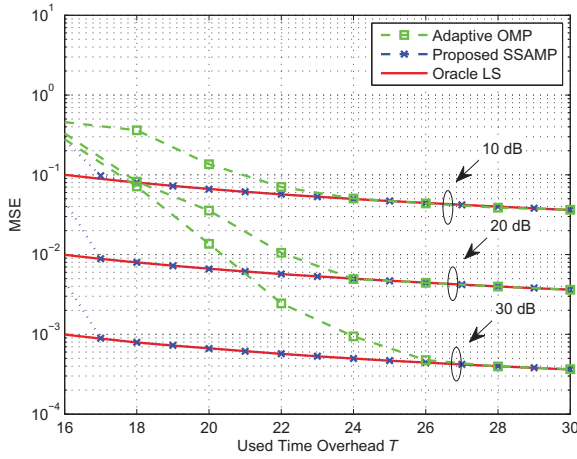
Fig. 4. MSE performance of different channel estimation schemes as functions of the time overhead G and SNR.

Fig. 5. Downlink BER performance, where the used channel state information at the BS is acquired by different channel estimation schemes.

considered. The oracle LS estimator with the known support set of the sparse channel vectors was adopted as the performance bound. From Fig. 4, it can be seen that the adaptive OMP scheme performs poorly. By contrast, the SSAMP algorithm is capable of approaching the oracle LS performance bound when $G > 2S_a/N_{BB}^{US}$. This is because the proposed SCS-based scheme fully exploits the spatially common sparsity of mmWave channels within the system bandwidth.

Fig. 5 compares the downlink bit error rate (BER) performance by using the hybrid analog-digital beamforming proposed in our previous work [2], where 16 QAM is used, and the precoding and combining are based on the estimated channels[2]. In the simulations, we consider two best LOS paths are used to serve the user due to $N_{BB}^{US} = 2$, which indicates two out of four small-cell BSs with the optimal channel quality jointly serve the user. It can be observed that the proposed channel estimation scheme outperforms its counterpart, and its BER performance is capable of approaching that of the performance bound.

VI. CONCLUSIONS

In this paper, we have proposed the SCS-based channel estimation scheme for the mmWave massive MIMO based access and backhaul in UDN. We first demonstrate that the SNR before beamforming in mmWave is appreciate for channel estimation due to the short link distance in UDN, although the path loss in mmWave is high. Moreover, by exploiting the sparsity of mmWave channels in the angular domain due to the high path loss for NLOS paths in mmWave, we propose the non-orthogonal pilot at the transmitter and the SCS-based channel estimator at the receiver. The proposed scheme can simultaneously estimate the channels associated with multiple small-cell BSs, and the required pilot overhead is only dependent on the small number of the dominated multipath. Simulation results have confirmed that our scheme can reliably acquire the mmWave massive MIMO channels with much reduced pilot overhead.

VII. ACKNOWLEDGMENTS

This work was supported in part by the International Science & Technology Cooperation Program of China (Grant

REFERENCES

[1] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 136-143, Dec. 2014.

[2] Z. Gao, L. Dai, Z. Wang, M. Imran, and M. Shakir, "MmWave massive MIMO based wireless backhaul for 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13-21, Oct. 2015..

[3] X. Gao, L. Dai, S. Han, C.-L. I, and R. Heath, "Energy efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," to appear in *IEEE J. Sel. Areas Commun.*

[4] R. Taori and A. Sridharan, "Point-to-multipoint in-band mmWave backhaul for 5G networks," *IEEE Wireless Commun.*, vol. 53, no. 1, pp. 195-201, Jan. 2015.

[5] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.

[6] S. Hur, T Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, A. Ghosh, "Multilevel millimeter wave beamforming for wireless backhaul," in *Proc. IEEE GLOBECOM Workshops 2011*, Dec. 5-9 2011, pp. 253-257.

[7] S. Hur, T Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, 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.

[8] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid precoding analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186-194, Jan. 2015.

[9] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. Heath, "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.

[10] 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.

[11] J. Zhang, L. Dai, X. Zhang, E. Bjornson, and Z. Wang, "Achievable rate of rician large-scale MIMO channels with transceiver hardware impairments," to appear in *IEEE Trans. Veh. Technol.*, 2015.

[12] W. Shen, L. Dai, B. Shim, S. Mumtaz, and Z. Wang, "Joint CSIT acquisition based on low-rank matrix completion for FDD massive MIMO systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2178-2181, Dec. 2015.

[13] Z. Gao, L. Dai, C. Yuen, and Z. Wang, "Asymptotic orthogonality analysis of time-domain sparse massive MIMO channels," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1826-1829, Oct. 2015.

[14] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proc. 42nd Asilomar Conf. Signal Syst. Comput.* (Pacific Grove, CA), Oct. 26-29, 2008, pp. 581–587.

[15] 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. 2009.

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

[2]We consider the estimated channels are perfectly fed back to the macrocell BS (BBU) with the control plane using low frequency band. The feedback overhead can be small, since only the estimated support set and the associated path gains ar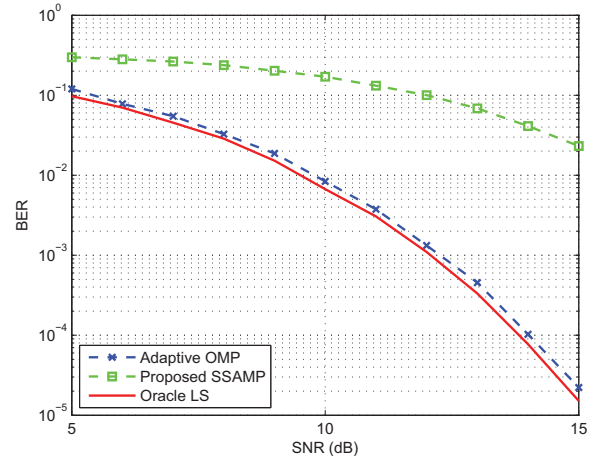e enough due to the sparsity of mmWave channels.