MmWave Channel Estimation via Atomic Norm Minimization for Multi-User Hybrid Precoding

Junquan Deng*, Olav Tirkkonen* and Christoph Studer†

*Department of Communications and Networking, Aalto University, Finland

†School of Electrical and Computer Engineering, Cornell University, NY, USA

Abstract—To perform multi-user multiple-input and multiple-output transmission in millimeter-wave (mmWave) cellular systems, the high-dimensional channels need to be estimated for designing the multi-user precoder. Conventional grid-based Compressed Sensing (CS) methods for mmWave channel estimation suffer from the basis mismatch problem, which prevents accurate channel reconstruction and degrades the precoding performance. This paper formulates mmWave channel estimation as an Atomic Norm Minimization (ANM) problem. In contrast to grid-based CS methods which use discrete dictionaries, ANM uses a continuous dictionary for representing the mmWave channel. We consider a continuous dictionary based on sub-sampling in the antenna domain via a small number of radio frequency chains. We show that mmWave channel estimation using ANM can be formulated as a Semidefinite Programming (SDP) problem, and the channel can be accurately estimated via off-the-shelf SDP solvers in polynomial time. Simulation results indicate that ANM can achieve much better estimation accuracy compared to grid-based CS, and significantly improves the spectral efficiency provided by multi-user precoding.

I. INTRODUCTION

In future millimeter-wave (mmWave) cellular networks, large antenna arrays are expected to be applied at the Base Station (BS) to serve multiple User Equipments (UE) in dense urban scenarios. To perform Multi-User Multiple-input and Multiple-Output (MU-MIMO) hybrid precoding [1], [2], accurate Channel State Information (CSI) is necessary. As mmWave channels typically compromise a few strong propagation paths, Compressed Sensing (CS) methods have been considered for mmWave channel estimation [1]–[5]. In order to apply CS to channel estimation, a discretization procedure is generally adopted to reduce the continuous angular or delay spaces to a finite set of grid points. Assuming that the Direction of Departure (DoD) and Direction of Arrival (DoA) lie on the grid, the virtual channel representation is sparse with few non-zero entries. The channel estimation problem then can be addressed with a specific measurement matrix and a recovery algorithm, e.g., the Orthogonal Matching Pursuit (OMP) [5]. However, as the actual signals are continuous and will not fall on the discrete points, basis mismatch will degrade recovery performance [6]. Although finer grids can reduce the reconstruction error, they require more computation resources.

In the literature, most of the multi-user hybrid precoding schemes assume perfect CSI at the transmitters (see e.g., [7], [8]). However, grid-based CS for channel estimation will lead to the basis mismatch problem and hence channel estimation errors, which will degrade the performance of MU-MIMO pre-coding such as Zero-Forcing (ZF) [2]. Recently, a continuous basis pursuit technique with auxiliary interpolation points [9] was used for mmWave channel estimation in [10], which shows that adding interpolation points to the original grids can improve channel estimation accuracy considerably. This paper considers a continuous dictionary for multi-user mmWave channel estimation in order to completely eliminate the basis mismatch error. We proposed a continuous dictionary based on antenna-domain sub-sampling via a much smaller number of Radio Frequency (RF) chains in the hybrid architecture. The sparse mmWave channel estimation is formulated as an Atomic Norm Minimization (ANM) problem [11], which can be solved via Semidefinite Programming (SDP) in polynomial time. ANM has been considered for massive MIMO channel estimation [12], [13], wherein a fully-digital architecture is assumed which has much higher complexity compared to the hybrid architectures. To show the advantage of ANM based on antenna-domain sub-sampling in the hybrid architecture, we investigate a low-complexity MU-MIMO hybrid precoding scheme using the estimated channel information. Finally, we evaluate the performance of ANM and the precoding scheme via simulations in realistic scenarios. Simulation results regarding the channel estimation error and user spectral efficiency are provided, which verify the efficacy of our solutions.

II. SYSTEM MODEL

A. Architectures of BS and UE

We consider a mmWave cellular system where the BS has N antennas and Q RF chains to serve $K = Q$ UEs in downlink (DL). In contrast to [4], [7], a BS architecture equipped with both Phase Shifters (PS) and switches is proposed, as shown in Fig. 1. The BS antennas are grouped into Q sub-arrays. Antennas in each sub-array are associated with one particular RF chain. With the proposed architecture, the BS can adopt two working modes: 1) the PS mode with a partially-connected PS network (as A2 in [7]); 2) the SW mode with a partially-connected switch network (as A6 in [7]). When working in SW mode, the BS contains a fully-digital architecture with Q antennas. The switch network enables accessing the received signal from an individual antenna, which is useful for channel estimation, especially when the Signal-to-Noise Ratio (SNR) is high. The cost of adding dedicated switches to the phase-shifter network is moderate as the implementation complexity of switches is typically lower than that of phase shifters [7]. In contrast, the PS network is used to perform analog beamforming and combining for both beam training and data transmission.
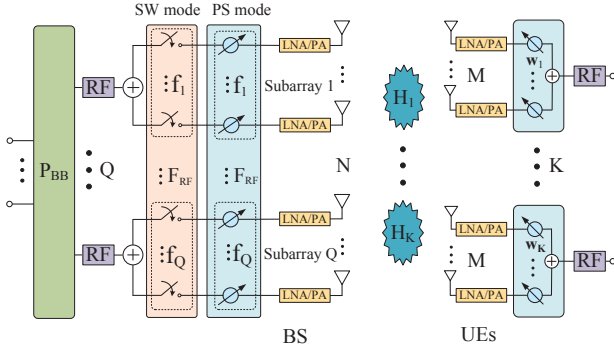
Fig. 1. System architectures for BS and UE, the RF part of BS has two modes of operation: 1) phase-shifter-based mode (PS mode) for directional data transmission and 2) switch-based mode (SW mode) for channel estimation.

At the UE side, each UE is assumed to have a single RF chain with M antennas and M phase shifters. Such UEs are able to perform single-data-stream analog beamforming. This UE architecture has low hardware complexity and energy consumption, making it suitable for mobile UEs [14].

The BS has two kinds of RF precoder/combiner depending on the working modes. In SW mode, the sub-array RF precoder/combiner $\mathbf{f}_q$ is an antenna selection vector with one non-zero element. In PS mode, $\mathbf{f}_q$ is a sub-array steering vector. For UEs, the RF precoder/combiner $\mathbf{w}_k$ of UE k is a beam steering vector. We assume that BS and UE have Q_{ps}-bits quantized PSs [15]. The number of beams that can be steered by BS or UE depends on the PS resolution Q_{ps} rather than the number of antennas, and is equal to $2^{Q_{\mathrm{ps}}}$. The coefficient set available for each PS is denoted by $\mathbb{P} = \{\omega^0, \omega^1, \ldots, \omega^{2^{Q_{\mathrm{ps}}}-1}\}$ where $\omega = \mathrm{e}^{j2\pi/2^{Q_{\mathrm{ps}}}}$ is the minimum angular domain separation. The quantized PSs can adjust the phases of the signals and have constant modulus. Specifically, the RF beamforming codewords are

$$\mathbf{f}_q = \mathbf{1}_{\Omega_q} \odot \boldsymbol{f}_\omega(c, N), \quad \mathbf{w}_k = \boldsymbol{f}_\omega(c, M), \quad (1)$$

where Ω_q is the antenna index set for qth sub-array, $\mathbf{1}_{\Omega_q}$ is a binary valued vector with $|\Omega_q|$ ones indexed by Ω_q, $\boldsymbol{f}_\omega(c, i) = [1, \omega^c, \omega^{2c}, \ldots, \omega^{(i-1)c}]^T$ with $c \in \{0, 1, \ldots, 2^{Q_{\mathrm{ps}}}-1\}$.

B. Channel Model

MmWave channels are highly directional and contain a few dominant paths in the angular domain [14]. Due to the extremely short wave-length, diffraction effects in mmWave frequencies are small; only Line-of-Sight (LoS) and reflected paths are significant. As a consequence, the number of paths is smaller than the number of BS antennas. Suppressing the vertical directivity of the channel, for the kth UE, the DL channel is modeled as [4],

$$\mathbf{H}_k = \sum_{l=1}^{L} \alpha_l \mathbf{a}_{\mathrm{UE}}(\theta_l) \mathbf{a}_{\mathrm{BS}}(\phi_l)^{\mathrm{H}} \in \mathbb{C}^{M \times N}, \quad (2)$$

where L represents the number of paths, and α_l denotes the complex gain of the lth path. In addition, $\mathbf{a}_{\mathrm{BS}}(\phi_l)$ and $\mathbf{a}_{\mathrm{UE}}(\theta_l)$ represent the BS and UE array response vectors for the lth path, where ϕ_l is the DoD at BS, and θ_l is DoA

at UE. Considering plane-wave model and Uniform Linear Array (ULA) for both BS and UE, the array steering vectors can be written as

$$\mathbf{a}_{\mathrm{BS}}(\phi) = [1, \mathrm{e}^{j\frac{2\pi}{\lambda} d \sin(\phi)}, \ldots, \mathrm{e}^{j(N-1)\frac{2\pi}{\lambda} d \sin(\phi)}]^{\mathrm{T}},$$
$$\mathbf{a}_{\mathrm{UE}}(\theta) = [1, \mathrm{e}^{j\frac{2\pi}{\lambda} d \sin(\theta)}, \ldots, \mathrm{e}^{j(M-1)\frac{2\pi}{\lambda} d \sin(\theta)}]^{\mathrm{T}}, \quad (3)$$

where λ is carrier wavelength and the antenna spacing here is assumed to be $d = \lambda/2$. We assume that the system is working in the Time Division Duplex (TDD) mode and the DL and uplink (UL) channels are reciprocal. Denote the BS baseband combiner or precoder as $\mathbf{P}_{\mathrm{BB}} = [\mathbf{p}_1, \ldots, \mathbf{p}_K] \in \mathbb{C}^{Q \times K}$, the BS RF combiner or precoder as $\mathbf{F}_{\mathrm{RF}} = [\mathbf{f}_1, \ldots, \mathbf{f}_Q] \in \mathbb{C}^{N \times Q}$, and the RF combiner or precoder for all K UEs as $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K] \in \mathbb{C}^{M \times K}$. The DL and the UL received signals including interference and noise for UE k are

$$z_{\mathrm{u}} = \mathbf{w}_k^{\mathrm{H}} \mathbf{H}_k \mathbf{F}_{\mathrm{RF}} \left(\mathbf{p}_k x_k + \sum_{j \neq k} \mathbf{p}_j x_j \right) + \mathbf{w}_k^{\mathrm{H}} \mathbf{n}_{\mathrm{u}},$$
$$z_{\mathrm{b}} = \mathbf{p}_k^{\mathrm{H}} \mathbf{F}_{\mathrm{RF}}^{\mathrm{H}} \left(\mathbf{H}_k^{\mathrm{H}} \mathbf{w}_k y_k + \sum_{j \neq k} \mathbf{H}_j^{\mathrm{H}} \mathbf{w}_j y_j + \mathbf{n}_{\mathrm{b}} \right), \quad (4)$$

where x_k and y_k are the DL and UL signals, satisfying $\mathbb{E}(x_k x_k^*) = \frac{\rho_{\mathrm{BS}}}{K}$ and $\mathbb{E}(y_k y_k^*) = \frac{\rho_{\mathrm{UE}}}{M}$ with transmit powers as ρ_{BS} and ρ_{UE}. In addition, $\mathbf{n}_{\mathrm{u}} \in \mathbb{C}^{M \times 1}$ is the $\mathcal{CN}(0, \sigma_{\mathrm{u}}^2)$ UE noise vector with noise power σ_{u}^2 per antenna, and $\mathbf{n}_{\mathrm{b}} \in \mathbb{C}^{N \times 1}$ is the $\mathcal{CN}(0, \sigma_{\mathrm{b}}^2)$ BS noise vector with noise power σ_{b}^2 per antenna. For DL transmission, $\mathbf{F} = \mathbf{F}_{\mathrm{RF}} \mathbf{P}_{\mathrm{BB}}$ satisfies power constraint $\mathrm{tr}(\mathbf{F}\mathbf{F}^{\mathrm{H}}) \leq 1$. For UL transmission, $\mathbf{w}_k$ satisfies $\|\mathbf{w}_k\|_2^2 \leq 1$.

III. PROBLEM FORMULATION FOR CHANNEL ESTIMATION

For channel training in UL, we assume that the kth UE transmits a pilot sequence $\mathbf{s}_k \in \mathbb{C}^{1 \times T_{\mathrm{s}}}$ using UE beam $\mathbf{w}_k$. Assuming $\mathbf{P}_{\mathrm{BB}} = \mathbf{I}_Q$, the received training signal at BS is

$$\mathbf{Y} = \mathbf{F}_{\mathrm{RF}}^{\mathrm{H}} \left(\sum_{k=1}^{K} \mathbf{h}_k \mathbf{s}_k + \mathbf{N}_{\mathrm{b}} \right). \quad (5)$$

Here $\mathbf{h}_k = \mathbf{H}_k^{\mathrm{H}} \mathbf{w}_k$ is the effective channel for UE k. The user pilots are assumed to satisfy $\mathbf{s}_k \mathbf{s}_{k'}^{\mathrm{H}} = T_{\mathrm{s}} \rho_{\mathrm{UE}} \delta_{k,k'}$, with $\delta_{i,j}$ the Kronecker delta function. Correlating $\mathbf{Y}$ with $\mathbf{s}_k$, we get

$$\mathbf{y}_k = \mathbf{Y} \mathbf{s}_k^{\mathrm{H}} = T_{\mathrm{s}} \rho_{\mathrm{UE}} \mathbf{F}_{\mathrm{RF}}^{\mathrm{H}} \mathbf{h}_k + \mathbf{F}_{\mathrm{RF}}^{\mathrm{H}} \mathbf{N}_{\mathrm{b}} \mathbf{s}_k^{\mathrm{H}}. \quad (6)$$

Assuming that the BS takes T snapshots of measurements with different combining matrices, stacking the $W = T \times Q$ measurement samples gives

$$\mathbf{z} = \begin{bmatrix} \mathbf{y}_{1,k} \\ \mathbf{y}_{2,k} \\ \vdots \\ \mathbf{y}_{T,k} \end{bmatrix} = T_{\mathrm{s}} \rho_{\mathrm{UE}} \underbrace{\begin{bmatrix} \mathbf{F}_{1,\mathrm{RF}}^{\mathrm{H}} \\ \mathbf{F}_{2,\mathrm{RF}}^{\mathrm{H}} \\ \vdots \\ \mathbf{F}_{T,\mathrm{RF}}^{\mathrm{H}} \end{bmatrix}}_{\boldsymbol{\Phi}} \mathbf{h}_k + \underbrace{\begin{bmatrix} \mathbf{F}_{1,\mathrm{RF}}^{\mathrm{H}} \mathbf{N}_{1,\mathrm{b}} \\ \mathbf{F}_{2,\mathrm{RF}}^{\mathrm{H}} \mathbf{N}_{2,\mathrm{b}} \\ \vdots \\ \mathbf{F}_{T,\mathrm{RF}}^{\mathrm{H}} \mathbf{N}_{T,\mathrm{b}} \end{bmatrix}}_{\mathbf{n}} \mathbf{s}_k^*, \quad (7)$$

where $\boldsymbol{\Phi}$ is the sensing matrix and $\mathbf{n}$ is the noise after combining. The objective of BS channel estimation is to estimate the effective channel. From (2), we have

$$\mathbf{h}_k = \mathbf{H}_k^{\mathrm{H}} \mathbf{w}_k = \sum_{l=1}^{L} \beta_l \mathbf{a}_{\mathrm{BS}}(\phi_l), \quad (8)$$

where $\beta_l = (\alpha_l (\mathbf{w}_k)^{\mathrm{H}} \mathbf{a}_{\mathrm{UE}}(\theta_l))^*$. The effective channel $\mathbf{h}_k$ contains fewer significant paths compared to the full MIMO

channel $\mathbf{H}_k$. To estimate $\mathbf{h}_k$, one can estimate DoAs $\{\phi_l\}$ and the complex gains $\{\beta_l\}$ for the paths and then reconstruct the channel. In grid-based CS methods, a discrete dictionary $\mathbf{\Psi}_{\mathrm{BS}} = [\mathbf{a}_{\mathrm{BS}}(\phi_1), \ldots, \mathbf{a}_{\mathrm{BS}}(\phi_{G_{\mathrm{b}}})]$ with G_{b} bases is used to represent the channel $\mathbf{h}_k$. We consider a grid in which $\{\sin(\phi_1), \ldots, \sin(\phi_{G_{\mathrm{b}}})\}$ are evenly distributed in $(-1, 1]$ [5]. The channel vector is represented as

$$\mathbf{h}_k = \mathbf{\Psi}_{\mathrm{BS}} \mathbf{h}_{\mathrm{v}}, \tag{9}$$

where $\mathbf{h}_{\mathrm{v}} \in \mathbb{C}^{G_{\mathrm{b}} \times 1}$ is the sparse virtual channel in dictionary $\mathbf{\Psi}_{\mathrm{BS}}$. Denoting $\mathbf{A} = \mathbf{\Phi} \mathbf{\Psi}_{\mathrm{BS}}$, to estimate the sparse virtual channel $\mathbf{h}_{\mathrm{v}}$, one can formulate the following optimization problem:

$$\underset{\mathbf{h}_{\mathrm{v}}}{\text{minimize}} \|\mathbf{h}_{\mathrm{v}}\|_0 \quad \text{s.t.} \quad \|\mathbf{z}_k - \mathbf{A}\mathbf{h}_{\mathrm{v}}\|_2^2 \leq \eta. \tag{10}$$

where η is an optimization parameter and typically set as the noise power after combining. In contrast to [5], we consider a switch-based method for constructing the sensing matrix, in which only Q individual antennas are sampled during one measurement snapshot. We call this method Antenna Domain Sub-Sampling (ADSS). The ADSS sensing matrix has the following structure:

$$\mathbf{\Phi} = T_{\mathrm{s}}\rho_{\mathrm{UE}}[\mathbf{e}_{i_{1,1}}, \mathbf{e}_{i_{1,2}}, \ldots, \mathbf{e}_{i_{1,Q}}, \ldots, \mathbf{e}_{i_{T,Q}}]^{\mathrm{H}}, \tag{11}$$

where the unit vector $\mathbf{e}_{i_{t,q}}$ has one entry equal to 1 at position $i_{t,q}$, and $i_{t,q}$ is the antenna index for the qth RF chain in the tth measurement snapshot. Furthermore, if we define the antenna index set $\Omega = \{i_{1,1}, i_{1,2}, \ldots, i_{1,Q}, \ldots, i_{T,Q}\}$ for all sampled antennas, we have

$$\mathbf{z} = \mathbf{\Phi}\mathbf{h}_k + \mathbf{n} = T_{\mathrm{s}}\rho_{\mathrm{UE}}(\mathbf{h}_k)_\Omega + \mathbf{n}. \tag{12}$$

A variety of algorithms have been proposed for obtaining an approximate solution to (10) at polynomial complexity. OMP is a preferred method due to its simplicity and fast implementation [5]. To solve (10) based on the measurement data $\mathbf{z}$ in (12), we use OMP as in Algorithm 1. The computational complexity of OMP is proportional to the number G_{b} of grid points. Step 3 in Algorithm 1 needs $\mathcal{O}(WG_{\mathrm{b}})$ computations and solving the Least Square (LS) problem in step 5 requires $\mathcal{O}(W|\mathcal{I}_t|^2)$.

The joint design of the sensing matrix $\mathbf{\Phi}$ and dictionary matrix $\mathbf{\Psi}$ plays a key role in achieving accurate channel estimation. To reconstruct $\mathbf{h}_{\mathrm{v}}$, $\mathbf{A}$ should have low mutual coherence $\mu(\mathbf{A})$, which is the largest normalized inner product for two different columns of $\mathbf{A}$. Specifically, when the path directions lie on the defined grid, it can be shown that OMP can recover $\mathbf{h}_{\mathrm{v}}$ in the noiseless case if [7]

$$\mu(\mathbf{A}) = \max_{i \neq j} \frac{|\mathbf{a}_i^{\mathrm{H}} \mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \cdot \|\mathbf{a}_j\|_2} < \frac{1}{2L-1}, \tag{13}$$

where $\mathbf{a}_i$ and $\mathbf{a}_j$ are two different columns of $\mathbf{A}$. The lower bound of $\mu(\mathbf{A})$ is given by the Welch bound [7], which decrease as the number of measurement W increases.

Algorithm 1 OMP for effective channel estimation

Input: The sensing matrix $\mathbf{\Phi}$, the dictionary matrix $\mathbf{\Psi}_{\mathrm{BS}}$ and $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}_{\mathrm{BS}} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{G_{\mathrm{b}}}]$, the measurement vector $\mathbf{z}$, and a threshold δ.

1: Iteration counter $t \leftarrow 0$, basis vector set $\mathcal{I}_t \leftarrow \emptyset$, virtual channel $\mathbf{h}_{\mathrm{v}} \leftarrow \mathbf{0} \in \mathbb{C}^{G_{\mathrm{b}} \times 1}$, residual error $\mathbf{r}_t \leftarrow \mathbf{z}$
2: **while** $\|\mathbf{r}\| > \delta$ and $t < G_{\mathrm{b}}$ **do**
3:    $g^\star \leftarrow \arg\max_g \|\mathbf{a}_g^{\mathrm{H}} \mathbf{r}_t\|_2$ ▷ Find new basis vector
4:    $t \leftarrow t+1$, $\mathcal{I}_t \leftarrow \mathcal{I}_{t-1} \bigcup \{g^\star\}$ ▷ Update vector set
5:    $\boldsymbol{x}^\star \leftarrow \arg\min_{\boldsymbol{x}} \|\mathbf{z} - (\mathbf{A})_{\mathcal{I}_t} \boldsymbol{x}\|_2$ ▷ LS method
6:    $\mathbf{r}_t \leftarrow \mathbf{z} - (\mathbf{A})_{\mathcal{I}_t} \boldsymbol{x}^\star$ ▷ Update residual error
7:    $(\mathbf{h}_{\mathrm{v}})_{\mathcal{I}_t} \leftarrow \boldsymbol{x}^\star$ ▷ Update the virtual channel
8: **end while**

Output: Estimated channel $\hat{\mathbf{h}} = \mathbf{\Psi}_{\mathrm{BS}}\mathbf{h}_{\mathrm{v}}$

IV. ATOMIC NORM MINIMIZATION FOR EQUIVALENT CHANNEL ESTIMATION

When using the dictionary $\mathbf{\Psi}_{\mathrm{BS}}$ for channel estimation via OMP, the estimated path angles are considered to be on a discrete grid $\{\phi_1, \ldots, \phi_{G_{\mathrm{b}}}\}$, which introduces the basis mismatch problem [6]. To avoid this, we consider a continuous dictionary for estimating the effective channel.

To estimate the effective channel based on UL measurements, the following constrained optimization problem can be formulated:

$$\underset{\mathbf{h}}{\text{minimize}} \, \mathcal{M}(\mathbf{h}) \quad \text{s.t.} \quad \|\mathbf{z} - \mathbf{\Phi}\mathbf{h}\|_2^2 \leq \eta. \tag{14}$$

Here $\mathcal{M}(\mathbf{h})$ denotes a sparse metric to be minimized for a channel vector $\mathbf{h}$. The noise is assumed to be bounded by $\|\mathbf{n}\|_2^2 \leq \eta$, and $\mathbf{h}$ can be treated as the signal of interest which needs to be estimated based on the observed data $\mathbf{z}$ in the context of line spectral estimation [11]. Problem (14) differs from (10) as it is formulated without a discrete dictionary. Instead of (14), one can solve a regularized optimization problem as

$$\underset{\mathbf{h}}{\text{minimize}} \, \xi \mathcal{M}(\mathbf{h}) + \frac{1}{2} \|\mathbf{z} - \mathbf{\Phi}\mathbf{h}\|_2^2, \tag{15}$$

where $\xi > 0$ is a regularization parameter related to η. There are different choices for $\mathcal{M}(\mathbf{h})$ [6]. Here, we consider the atomic norm $\mathcal{M}(\mathbf{h}) = \|\mathbf{h}\|_{\mathcal{A}}$ proposed in [6]. The atomic norm has been used in grid-less compressive sensing for a range of applications [6], [11]–[13], including DoA estimation, line spectral estimation and massive MIMO channel estimation. The continuous set of atoms used to represent $\mathbf{h}$ is defined as

$$\mathcal{A} = \left\{ \mathbf{a}_{\mathrm{BS}}(\phi)\alpha : \phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right], \alpha \in \mathbb{C}, |\alpha| = 1 \right\}. \tag{16}$$

The atomic norm is the gauge function [16] of the convex hull conv$(\mathcal{A})$. Formally, the atomic norm can be written as

$$\begin{aligned} \|\mathbf{h}\|_{\mathcal{A}} &= \inf \{ g > 0 : \mathbf{h} \in g \cdot \mathrm{conv}(\mathcal{A}) \} \\ &= \inf \left\{ \sum_i b_i : \mathbf{h} = b_i \sum_i \mathbf{a}_i, b_i > 0, \mathbf{a}_i \in \mathcal{A} \right\}. \end{aligned} \tag{17}$$

The atomic norm is a continuous counterpart of the ℓ_1-norm used in on-grid CS methods. It can be computed efficiently via semidefinite programming [6]. Specifically, $\|\mathbf{h}\|_{\mathcal{A}}$ defined

in (17) is the optimal value of the following matrix trace minimization problem:

$$\underset{\boldsymbol{u},t}{\text{minimize}} \ \frac{1}{2}(t + u_1) \quad \text{s.t.} \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \mathbf{h} \\ \mathbf{h}^{\mathrm{H}} & t \end{bmatrix} \succeq \mathbf{0}, \qquad (18)$$

where $\mathcal{T}(\boldsymbol{u})$ is a Hermitian Toeplitz matrix with the first row as $\boldsymbol{u} = [u_1, \dots, u_N]^{\mathrm{T}}$. In the noisy case, using the atomic norm, we can rewrite the optimization problem (15) as

$$\underset{\mathbf{D} \succeq \mathbf{0}}{\text{minimize}} \ \frac{\xi}{2}(t + u_1) + \frac{1}{2} \|\mathbf{z} - \boldsymbol{\Phi}\mathbf{h}\|_2^2$$

$$\text{s.t. } \mathbf{D} = \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \mathbf{h} \\ \mathbf{h}^{\mathrm{H}} & t \end{bmatrix}. \qquad (19)$$

The above problem is a SDP which can be solved by off-the-shelf convex optimization tools in polynomial time [17]. The computational complexity is $\mathcal{O}\left((N+1)^6\right)$ in each iteration using the interior point method.

To guarantee that $\mathbf{h}$ can be exactly recovered via ANM, the measurement matrix $\boldsymbol{\Phi}$ and the number of atoms in $\mathbf{h}$ (i.e., the number of paths L in the user channel) should satisfy a condition which is similar to the mutual coherence condition in on-grid CS. To define this condition for ANM, a concept called spark for the continuous set of atoms is introduced. Let us define the transformed continuous dictionary based on the measurement matrix $\boldsymbol{\Phi}$ as

$$\mathcal{A}_{\boldsymbol{\Phi}} = \left\{ \boldsymbol{\Phi}\mathbf{a}_{\mathrm{BS}}(\phi) : \phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right] \right\}. \qquad (20)$$

Similar to the definition of matrix spark, $\text{spark}(\mathcal{A}_{\boldsymbol{\Phi}})$ is defined as the smallest number of atoms which are linearly dependent in $\mathcal{A}_{\boldsymbol{\Phi}}$. From [16], the problem (14) with $\eta = 0$ in the noiseless setting has a unique solution if the number of paths satisfies

$$L < \frac{\text{spark}(\mathcal{A}_{\boldsymbol{\Phi}})}{2}. \qquad (21)$$

Note that the spark in (13) satisfies $\text{spark}(\mathbf{A}) \geq 1 + (\mu(\mathbf{A}))^{-1}$, so the condition above is equivalent to (13) in the case of discrete atoms. Thus the $\text{spark}(\mathcal{A}_{\boldsymbol{\Phi}})$ should be as large as possible so that ANM can recover a channel with as many paths as possible. For a specific BS array, one may notice that $\text{spark}(\mathcal{A}_{\boldsymbol{\Phi}})$ depends only on $\boldsymbol{\Phi}$. For ADSS via switches, $\text{spark}(\mathcal{A}_{\boldsymbol{\Phi}})$ depends on the number of sampled antennas $|\Omega|$ and the sampled antenna indexes. It was shown in [16] that $2 \leq \text{spark}(\mathcal{A}_{\boldsymbol{\Phi}}) \leq |\Omega| + 1$, and a random Ω or a Ω with $|\Omega|$ consecutive integers can achieve a sufficiently large spark.

V. MULTI-USER DOWNLINK PRECODING/COMBINING BASED ON THE ESTIMATED CHANNELS

In this section, we consider the designs of downlink BS precoder and UE combiners. To design the BS precoder $\{\mathbf{F}_{\mathrm{RF}}, \mathbf{P}_{\mathrm{BB}}\}$ and UE combiners $\mathbf{W}$ optimally, full CSI $\mathcal{H} = \{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_K\}$ needs to be known. However, acquiring $\mathcal{H}$ would require each UE to send huge amount of pilots which consumes lots of resources. A sub-optimal method, which works well when user channels are sparse, is that each UE first finds its best beam which aligns with the most significant path, then UEs transmit their pilots to BS for multi-user channel estimation as discussed in sections III and IV. As a result,
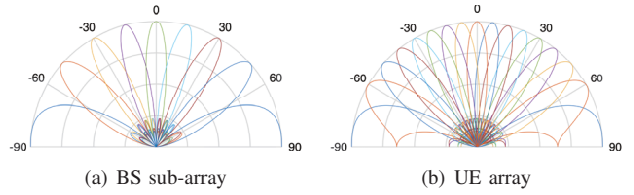

Fig. 2. (a) Designed Q DL sub-array beams with $N = 64, Q = 8, Q_{\mathrm{ps}} = 4$; (b) Designed 16 UE beams with $M = 8, Q_{\mathrm{ps}} = 4$.

the BS only needs to estimate the user effective channel $\mathbf{h}_k$ which has much fewer coefficients and significant paths to be estimated. The BS will then design the its MU-MIMO precoder based on those estimated effective channels. For UEs, the beam codebook is

$$\mathcal{U} = \left\{ \sqrt{1/M} \boldsymbol{f}_{\omega}(c, M) : c \in \{0, \dots, 2^{Q_{\mathrm{ps}}} - 1\} \right\}. \qquad (22)$$

For BS, the DL training beam codebook used by the qth sub-array is

$$\mathcal{F}_q = \left\{ \mathbf{1}_{\Omega_q} \odot \boldsymbol{f}_{\omega}(c, N) : c \in \{0, \dots, 2^{Q_{\mathrm{ps}}} - 1\} \right\}. \qquad (23)$$

Sub-array q will use one beam from $\mathcal{F}_q$ for UE beam training.

The objective of the UE beam training is to find a UE beam which maximizes the norm of the effective channel

$$\underset{\mathbf{w}_k \in \mathcal{U}}{\text{maximize}} \|\mathbf{w}_k^{\mathrm{H}} \mathbf{H}_k\|_2^2, \qquad (24)$$

The problem in (24) is equivalent to finding the best steering vector $\mathbf{w}_k \in \mathcal{U}$ that aligns with the paths that have large path gains. In principle, to find the best UE codeword, BS needs to transmit DL training beams which cover all the paths. However, this would increase the training overhead. We propose using all sub-arrays to transmit Q DL pilots with Q sub-array beams simultaneously which are evenly distributed in $[-\pi/2, \pi/2)$. Fig. 2 (a) shows the designed DL beams used for the BS sub-arrays and Fig. 2 (b) shows the UE beam codebook we used. The MU-MIMO precoding scheme based on the UE beam training and the effective channel estimation is summarised and shown in Algorithm 2. Noticing that DoAs and DoDs of the propagation paths change much slower than the channel coefficients [14], the best UE beam remains the same during a period which is much longer than channel coherence time, and the UE beam tracking can be performed less frequently.

VI. NUMERICAL RESULTS

In this section, the performance of the proposed channel estimation method and the precoding scheme are evaluated via numerical simulations. We assume $M = 8$ for UEs, and $N = 64$, $Q = 8$ for BS. The BS antennas are grouped into Q sub-arrays with equal size. We consider an outdoor small cell where BS is lower than buildings, and signals will be reflected or blocked by the walls. A ray-tracing channel model [18] is considered to generate the channels. The reflection coefficients are computed based on the Fresnel equation and up to 3rd order reflections are taken into account. The relative permittivities of building walls are uniformly distributed between 3 and 7. The BS is at the center while UEs

Algorithm 2 Joint UE RF combining and BS MU-MIMO precoding

1: BS transmits Q DL pilots using Q sub-arrays
2: Each UE $k \in \{1, 2, \ldots, K\}$ receives DL pilots using codewords from UE codebook $\mathcal{U} = \{\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(P)}\}$ and calculate the received power with the pth UE beam and qth BS DL beam, each UE then has a received power matrix $\mathbf{P}_k = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_Q] \in \mathbb{R}^{P \times Q}$. Let $\mathbf{c} \leftarrow \mathbf{0}_{P \times 1}$,
3: **for** $q = 1, 2, \ldots, Q$ **do**
4:     Find the best UE beam $i^\star$ for the qth BS beam:
5:     $i^\star \leftarrow \arg\max_i (\mathbf{p}_q)_i$, $(\mathbf{c})_{i^\star} \leftarrow (\mathbf{c})_{i^\star} + (\mathbf{p}_q)_{i^\star}$.
6:     Find the final best UE beam:
7:     $p^\star \leftarrow \arg\max_i (\mathbf{c})_i$, $\mathbf{w}_k \leftarrow \mathbf{w}^{(p^\star)}$
8: **end for**
9: UE k transmits UL pilot to BS using $\mathbf{w}_k$, the BS estimates the effective user channel $\mathbf{h}_k$ using OMP or ANM methods as described in sections III and IV; the BS has an estimated multi-user channel $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \ldots, \hat{\mathbf{h}}_K]^{\mathrm{H}}$. Let $\mathbf{F}_{\mathrm{RF}} = [\mathbf{f}_1, \ldots, \mathbf{f}_Q] \leftarrow \mathbf{0}_{N \times Q}$,
10: **for** $q \in \{1, 2, \ldots, Q\}$ **do**
11:     Select UE $k = q$, find $\mathbf{f}_q = \arg\max_{\mathbf{f} \in \mathcal{F}_q} |\hat{\mathbf{h}}_k^{\mathrm{H}} \mathbf{f}|^2$
12: **end for**
13: Design $\mathbf{P}_{\mathrm{BB}}$ as $\mathbf{P}_{\mathrm{BB}} = \dfrac{(\hat{\mathbf{H}} \mathbf{F}_{\mathrm{RF}})^{-1}}{\|\mathbf{F}_{\mathrm{RF}} (\hat{\mathbf{H}} \mathbf{F}_{\mathrm{RF}})^{-1}\|_{\mathrm{F}}}$ using ZF.

TABLE I
SIMULATION PARAMETERS

Parameter	Symbol	value
Carrier frequency	f_c	28 GHz
System bandwidth	b_w	256 MHz
OFDM subcarrier number	N_c	256
UE total Tx power	$N_c \times \rho_{\mathrm{UE}}$	23 dBm
BS total Tx power	$N_c \times \rho_{\mathrm{BS}}$	40 dBm
UE noise power	$N_c \times \sigma_{\mathrm{u}}^2$	-89 dBm
BS noise power	$N_c \times \sigma_{\mathrm{b}}^2$	-86 dBm
UE antenna pattern	$g_1(\theta)$	omnidirectional
BS antenna pattern	$g_2(\phi)$	defined in [20]
Phase-shifter resolution	Q_{ps}	4 bits
Pilot sequences	$\mathbf{s}_k$	Zadoff-Chu, $T_s = 63$

are dropped outdoor with distances to BS smaller than 150 meters, as shown in Fig. 3. Blocked UEs which cannot find a LOS or a reflected path are not considered in the simulation. The path coefficient in (2) is

$$\alpha_l = \mathrm{e}^{j\psi} \sqrt{G_0 x^{-2} g_1(\theta_l) g_2(\phi_l) \prod_{i=0}^{R} |r_i|^2}, \qquad (25)$$

where $G_0 = 10^{-6.14}$ is the omnidirectional path gain [19] at reference distance one meter, x is the path propagation distance in meter, $\psi \sim U(0, 2\pi)$ is a random phase, $g_1(\theta_l)$ and $g_2(\phi_l)$ are UE and BS antenna element patterns, respectively, R is the reflection order and r_i is the ith reflection coefficient. For a LOS path, we have $R = 0$, $r_0 = 1$. More details of simulation parameters are given in Tabe I.

A. Effective User Channel Estimation via OMP and ANM

All UEs transmit orthogonal pilots to BS using their best beams, and BS receives them via ADSS. In these simulations,
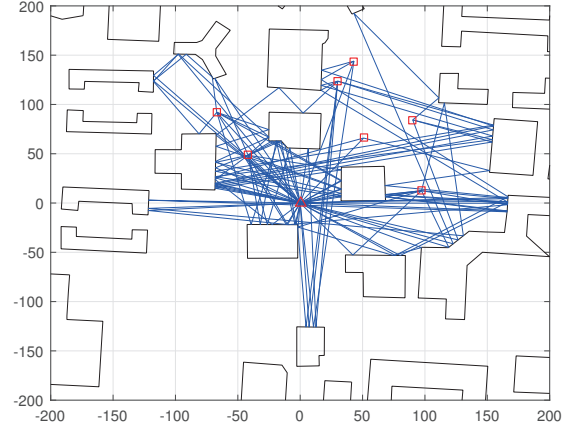

Fig. 3. Simulation scenario in a 400×400 m² area, the polygons represent the buildings with walls, red triangle is the BS and red squares are the UEs. Ray-tracing is used to generate the propagation paths which are shown in blue lines.

$T = 2$ measurement snapshots are used, so we have $T \times Q = 16$ antennas sampled pseudo-randomly via the switch network. The ANM problem (19) is solved via the SDPT3 [17] solver. The OMP algorithm applys a non-redundant dictionary with $G_b = N$ and redundant dictionaries with $G_b = 2N, 4N$. We estimate channel estimation performance, in terms of the Normalized Mean Square Error (NMSE) $\mathbb{E}\{\|\mathbf{h}_k - \hat{\mathbf{h}}_k\|_2^2 / \|\mathbf{h}_k\|_2^2\}$. In total 8000 UEs are simulated. Results are presented as a function of SNR $= \rho_{\mathrm{UE}} \|\mathbf{H}_k\|_{\mathrm{F}}^2 / (MN\sigma_{\mathrm{b}}^2)$. As illustrated in Fig. 4, increasing G_{b} directly results in better estimation accuracy for OMP. The ANM method outperforms OMP since it avoids basis mismatch by using the continuous dictionary. As G_{b} increases, the gap between OMP and ANM decreases at the cost of increasing computational complexity. When a non-redundant dictionary with $G_b = N$ is used for OMP, OMP suffers severely from basis mismatch; an error floor develops in the hign SNR regime. In comparison, ANM achieves much better estimation accuracy, especially in the high SNR regime, effectively removing the error floor. Inaccurate channel estimation degrades the interference mitigation performance in MU-MIMO precoding techniques. As a result, ANM would be a strong candidate for the hybrid precoding solution discussed in Section V. We also found that in the simulated scenario, the user effective channel $\mathbf{h}_k$ is sparse in angular domain. In most cases, there are one to three significant paths even when the original channel $\mathbf{H}_k$ has much more paths. This ensures that the condition (21) is satisfied and ANM can recover the channel with high probability. In the case that the condition (21) is not satisfied, ANM aims to recover those strongest paths and moderate channel estimation accuracy is attainable. Compared to OMP, solving the ANM problem using SDPT3 entails higher complexity. In practice, an efficient solver based on the Alternating Direction Method of Multipliers (ADMM) [21] technique could be adopted to solve the ANM problem.

B. Spectral Efficiency Performance of Multi-User Precoding

To evaluate the multi-user precoding performance, 8000 UEs are randomly divided into 1000 groups and 8 UEs in each
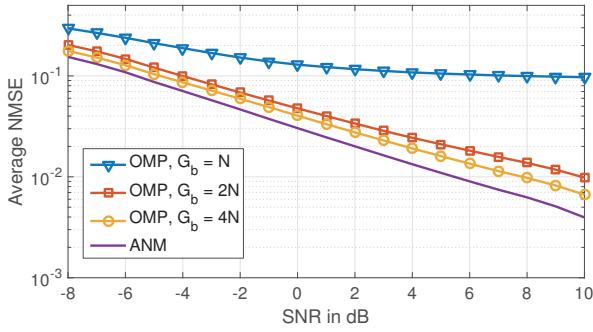
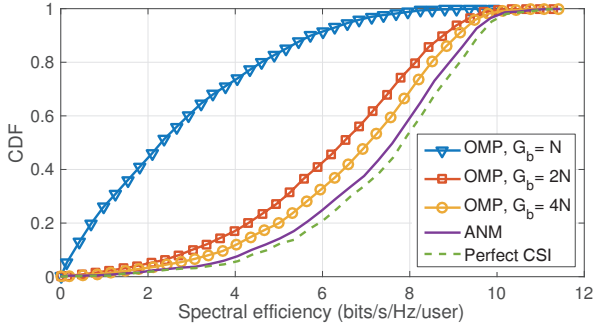Fig. 4. The average channel normalized estimation error for different channel estimation schemes.

Fig. 5. Performance of multi-user hybrid precoding with sub-arrays, with different channel estimates.

group are served simultaneously. The Cumulative Distribution Function (CDF) for user spectral efficiency is collected, where SE is estimated as $\log_2(1 + \gamma)$ from the Shannon formula with γ the signal-to-interference-plus-noise ratio. As shown in Fig. 5, hybrid precoding with ANM-based channel estimation can provide much better performance compared to an OMP-based scheme since it provides accurate channel estimates for the effective user channel. The baseband ZF precoding is sensitive to channel estimation error and the error will damage the user orthogonality provided by ZF. On average, OMP with $G_b = N$, $2N$ and $4N$ loses 63%, 18% and 11% of the spectral efficiency with perfect CSI. By comparison, ANM loses only 3.4%, which verify the efficacy of ANM.

VII. CONCLUSION

This paper has proposed a low-complexity hybrid architecture in which an inexpensive switch network is added to the subarrays to facilitate channel estimation. We have formulated the mmWave channel estimation as an ANM problem which can be solved via SDP with polynomial complexity. Compared to the grid-based CS methods, the proposed ANM approach uses a continuous dictionary with sub-sampling in antenna domain via the switch network. Simulation results show that ANM can achieve much better channel estimation accuracy compared to grid-based CS methods, and can help MU-MIMO hybrid precoding to achieve significantly better user spectral efficiency performance.

REFERENCES

[1] A. Alkhateeb et al., "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.

[2] A. Alkhateeb et al., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[3] Y. Han and J. Lee, "Two-stage compressed sensing for millimeter wave channel estimation," in *Proc. IEEE Int. Symp. Infor. Theory (ISIT)*, July 2016, pp. 860–864.

[4] A. Alkhateeb et al., "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, Apr. 2015, pp. 2909–2913.

[5] J. Lee et al., "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, June 2016.

[6] G. Tang et al., "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[7] R. Méndez-Rial et al., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.

[8] X. Yu et al., "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[9] C. Ekanadham et al., "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process*, vol. 59, no. 10, pp. 4735–4744, Oct. 2011.

[10] S. Sun and T. S. Rappaport, "Millimeter wave MIMO channel estimation based on adaptive compressed sensing," in *Proc. IEEE Int. Conf. Commun. Workshops (ICCW)*, May 2017, pp. 47–53.

[11] B. N. Bhaskar et al., "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.

[12] P. Zhang et al., "Atomic norm denoising-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, June 2015, pp. 4564-4569.

[13] Y. Wang et al., "Efficient channel estimation for massive MIMO systems via truncated two-dimensional atomic norm minimization," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1-6.

[14] T. S. Rappaport et al., *Millimeter wave wireless communications*. Pearson Education, 2014.

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

[16] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5145-5157, Oct.1, 1 2016.

[17] K.-C. Toh et al., "SDPT3 - A MATLAB software package for semidefinite programming, version 1.3," *Optimization methods and software*, vol. 11, no. 1-4, pp. 545–581, 1999.

[18] S. Hur et al., "28 GHz channel modeling using 3D ray-tracing in urban environments," in *Proc. European Conf. Antennas Propag. (EuCAP)*, May 2015, pp. 1–5.

[19] 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, June 2014.

[20] 3GPP, "Study on channel model for frequency spectrum above 6 GHz," TR 38.900 V14.3.1, July 2017.

[21] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1257-1269, Mar. 2016.