

Wideband Millimeter-Wave Massive MIMO Channel Training via Compressed Sensing

Tzu-Hsuan Chou, Nicolò Michelusi, David J. Love, and James V. Krogmeier

Abstract—In this work, a compressed sensing-aided wideband MIMO-OFDM channel training framework is proposed to reduce the training overhead in slowly-varying channels with frequency- and spatial-wideband (dual-wideband) effects. To combat the beam squint effect, a set of frequency-dependent array response matrices are constructed, enabling the recovery of the sparse beamspace channel from multiple observations across OFDM subcarriers, via multiple measurement vectors (MMV). A channel training algorithm (MMV-LS-CS) is proposed to estimate slowly-varying multipath channel parameters: MMV least squares (MMV-LS) is first used to estimate the channel on the previous beam index support, followed by MMV compressed sensing (MMV-CS) on the residual to estimate the time-varying multipath components. Finally, a channel refining algorithm is proposed to estimate the gains and time delays of the dominant channel paths jointly on pilot subcarriers. Numerical results show that MMV-LS-CS achieves more accurate and robust channel estimation than the state-of-the-art approach on slowly-varying dual-wideband MIMO-OFDM: given a moderate SNR of 20 dB, our algorithm attains NMSE = 0.15, as opposed to the state-of-the-art which attains NMSE = 0.43 in the same configuration. Besides, MMV-LS-CS necessitates SNR = 14 dB to achieve the spectral efficiency of 6 bit/s/Hz/stream, while the state-of-the-art scheme needs SNR = 17 dB to attain the same spectral efficiency.

I. INTRODUCTION

Millimeter-wave (mmWave) communication has been investigated as a promising solution to meet the capacity demands of future wireless systems, thanks to ample available bandwidth [1]. However, mmWave channels experience difficult channel conditions, requiring narrow beam communication with massive multiple-input multiple-output (MIMO) [2]. To reap the beamforming gain, accurate MIMO channel state information (CSI) is required, usually attained by beam scanning over a set of candidate beams to find the strongest one. Yet, this approach incurs an unacceptably large overhead due to the large beam space.

To reduce the training overhead, beam alignment has been largely investigated in recent years, ranging from *feedback-based schemes* [3], [4], *data-assisted schemes* [5], [6], to *multipath estimation* [7]–[9]. *Feedback-based schemes* adapt the beam-training according to the feedback information in an online fashion [3], or leverage the mobility of users as in [4]; *data-assisted schemes* use side information from an array of available sensors. Our work belongs to the class of *multipath estimation* schemes, which exploit the spatial

sparsity of the MIMO channel via compressed sensing (CS) to acquire the associated channel parameters (e.g., angle of arrivals (AOAs), angle of departures (AODs), etc.). The work [9] proposed an adaptive algorithm for mmWave massive MIMO channel estimation using a hierarchical multiresolution codebook. For wideband systems, MIMO orthogonal frequency division multiplexing (MIMO-OFDM) has been considered as the dominant structure to combat frequency selectivity, i.e., frequency-wideband effect. In addition, the time-delays across the aperture on large-scale antenna arrays are non-negligible, especially in massive MIMO, leading to the spatial-wideband effect [10], [11]. The AOAs/AODs observed by the receiver for a given propagation path are different across subcarriers, called a beam squint effect. The work [11] proposed a tensor-based channel training using the Vandermonde constraint and spatial smoothing for dual-wideband (frequency- and spatial-wideband effects) MIMO-OFDM channels. However, it does not leverage the temporal correlation of the channel, requiring periodic training with severe overhead.

The problem of exploiting the temporal correlation of the channel has been investigated for narrowband MIMO [12], [13] and frequency-wideband multiuser MIMO-OFDM [14], considering a common (or slightly-varying) support over time [15]. In [14], the problem of estimating the channel with a common slowly-varying support from multiple measurements among subcarriers is formulated as a multiple measurement vectors (MMV) problem, but without consideration of dual-wideband effects. However, this technique cannot be directly used on time-varying dual-wideband MIMO-OFDM channels: the frequency-dependent AOAs/AODs could break the common support across subcarriers, which may harm the estimation performance. This problem has not been investigated, and will be studied in this paper.

To address the problem, we construct frequency-dependent array response matrices based on the same AOA/AOD grid to preserve the common channel support across the OFDM subcarriers. To exploit the temporal correlation, we not only consider the previous support for the recovery process but also leverage the support to design the pilot signal and combining matrix to enhance the estimation performance. We propose a new channel training algorithm (MMV-LS-CS) for time-varying channels with dual-wideband effect and a channel refining algorithm to estimate the gains and time delays of the dominant paths. Numerical results show that MMV-LS-CS improves the channel estimation accuracy and spectral efficiency on the slowly-varying dual-wideband MIMO-OFDM.

II. SYSTEM MODEL

First, we describe a time-varying dual-wideband MIMO-OFDM channel model. Then, we introduce the signal model

This work was supported in part by the National Science Foundation under grants CNS-1642982, CCF-1816013, EEC-1941529 and CNS-2129015.

T.-H. Chou, D. J. Love and J. V. Krogmeier are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA; emails: {chou59, djlove, jvk}@purdue.edu.

N. Michelusi is with the School of Electrical, Computer and Energy Engineering, Arizona State University, AZ, USA; email: nicolo.michelusi@asu.edu.

of beam training, followed by the channel training protocol.

A. Channel Model

We consider a MIMO-OFDM system with bandwidth B , carrier frequency f_c , K_o subcarriers, a receiver employing a uniform linear array (ULA) with N_r antennas, and a transmitter employing a ULA with N_t antennas. We consider a wideband geometric massive MIMO channel model with L scattering paths: let φ_ℓ and ν_ℓ be the physical AOA and physical AOD of the ℓ -th path, and β_ℓ be its fading coefficient. Given the antenna spacing d , the corresponding spatial AOA θ_ℓ and spatial AOD ϕ_ℓ are $\theta_\ell = \frac{d \sin \varphi_\ell}{\lambda_c}$ and $\phi_\ell = \frac{d \sin \nu_\ell}{\lambda_c}$, where λ_c is the carrier wavelength. Due to the increasing scale of antenna arrays, the propagation delay of waves traveling across the array aperture is non-negligible, leading to the spatial-wideband effect [10], [11]. The time delay of the ℓ -th path between the n_t -th transmit antenna and n_r -th receive antenna is $\tau_{\ell, n_r, n_t} = \tau_\ell + \tau_{n_r, n_t}^d(\theta_\ell, \phi_\ell)$, where τ_ℓ is the reference path delay of the ℓ -th scattering path on the first transmit and receive antenna pair ($n_r = n_t = 1$), and $\tau_{n_r, n_t}^d(\theta_\ell, \phi_\ell) = (n_r - 1) \frac{\theta_\ell}{f_c} - (n_t - 1) \frac{\phi_\ell}{f_c}$ represents the additional delay across the antenna aperture [16]. Given a baseband signal $x_{n_t}(t)$ transmitted at the n_t -th antenna, the baseband received signal at the n_r -th antenna is derived as

$$r_{n_r}(t) = \sum_{n_t=1}^{N_t} \sum_{\ell=1}^L \beta_\ell x_{n_t}(t - \tau_{\ell, n_r, n_t}) e^{-j2\pi f_c \tau_{\ell, n_r, n_t}}.$$

By taking the continuous Fourier transform, the baseband signal in the frequency domain is

$$R_{n_r}(f) = \sum_{n_t=1}^{N_t} \sum_{\ell=1}^L \alpha_\ell e^{-j2\pi(f_c+f)\tau_{n_r, n_t}^d(\theta_\ell, \phi_\ell)} e^{-j2\pi f \tau_\ell} X_{n_t}(f),$$

where we have defined the equivalent path gain $\alpha_\ell = \beta_\ell e^{-j2\pi f_c \tau_\ell}$ and $X_{n_t}(f)$ is the continuous Fourier transform of $x_{n_t}(t)$. By stacking up the MIMO signal on antennas, we have an $N_r \times 1$ receive signal vector

$$\mathbf{R}(f) = \mathbf{H}(f) \mathbf{X}(f) + \mathbf{n}(f),$$

where we have defined the transmit signal vector $\mathbf{X}(f) = [X_1(f), \dots, X_{N_t}(f)]^\top$ and the additive noise vector $\mathbf{n}(f) \in \mathbb{C}^{N_r \times 1}$. The frequency response of the baseband MIMO channel $\mathbf{H}(f) \in \mathbb{C}^{N_r \times N_t}$ is

$$\mathbf{H}(f) = \sqrt{N_r N_t} \sum_{\ell=1}^L \alpha_\ell \mathbf{a}_R(\theta_\ell; f) \mathbf{a}_T(\phi_\ell; f)^H e^{-j2\pi f \tau_\ell}, \quad (1)$$

where the receive and transmit spatial-frequency steering vectors are defined as

$$\begin{aligned} \mathbf{a}_R(\theta; f) &= \frac{1}{\sqrt{N_r}} [1, e^{-j2\pi(1+\frac{f}{f_c})\theta}, \dots, e^{-j2\pi(N_r-1)(1+\frac{f}{f_c})\theta}]^\top, \\ \mathbf{a}_T(\phi; f) &= \frac{1}{\sqrt{N_t}} [1, e^{-j2\pi(1+\frac{f}{f_c})\phi}, \dots, e^{-j2\pi(N_t-1)(1+\frac{f}{f_c})\phi}]^\top. \end{aligned}$$

On a MIMO-OFDM system with K_o subcarriers, we denote the baseband frequency of the k -th subcarrier as $f_k = (k - \frac{K_o+1}{2}) \frac{B}{K_o}$, $k = 1, \dots, K_o$. Thus, the MIMO channel on the k -th subcarrier is $\mathbf{H}_k = \mathbf{H}(f_k)$.

Considering the array geometry, the MIMO channel can be formulated as an extended virtual representation [1]. Assume the spatial AOAs (respectively, AODs) take values from a uniform grid $\mathcal{G}_R$ ($\mathcal{G}_T$) of size $G_r \geq L$ ($G_t \geq L$) on $[-1/2, 1/2)$, i.e.,

$$\mathcal{G}_R = \left\{ \theta_{g_r} = -\frac{1}{2} + \frac{g_r - 1}{G_r}, g_r = 1, \dots, G_r \right\}, \quad (2)$$

$$\mathcal{G}_T = \left\{ \phi_{g_t} = -\frac{1}{2} + \frac{g_t - 1}{G_t}, g_t = 1, \dots, G_t \right\}. \quad (3)$$

We construct the receive array response matrix on the k -th subcarrier $\mathbf{A}_{R,k} \in \mathbb{C}^{N_r \times G_r}$ by stacking the spatial-frequency steering vectors with AOAs taking value on the uniform grid $\mathcal{G}_R$,

$$\mathbf{A}_{R,k}(:, p) = \mathbf{a}_R(\theta_p; f_k), \quad \theta_p \in \mathcal{G}_R. \quad (4)$$

In the same manner, the transmit array response matrix on the k -th subcarrier $\mathbf{A}_{T,k} \in \mathbb{C}^{N_t \times G_t}$ is constructed as

$$\mathbf{A}_{T,k}(:, q) = \mathbf{a}_T(\phi_q; f_k), \quad \phi_q \in \mathcal{G}_T. \quad (5)$$

With this notation, the MIMO channel on the k -th subcarrier can be expressed in an extended virtual representation as [1]

$$\mathbf{H}_k = \mathbf{A}_{R,k} \mathbf{S}_k \mathbf{A}_{T,k}^H, \quad (6)$$

where $\mathbf{S}_k \in \mathbb{C}^{G_r \times G_t}$ is the beamspace channel matrix whose non-zero elements are located in the positions corresponding to the spatial AOAs/AODs of the propagation paths. Due to the mismatch between the spatial AOAs/AODs and quantized values, a grid-mismatch error may exist but can be diminished if the grid sizes (G_r, G_t) are chosen sufficiently large. With the compact antenna deployments and the limited scattering of the mmWave channel environment, the MIMO channels are spatially correlated and focus on certain spatial directions, so that the beamspace channel matrix $\mathbf{S}_k$ tends to be sparse. Therefore, we assume that $\mathbf{S}_k$ has at most $L < G_r G_t$ non-zero elements, and the remaining elements are negligible.

For the support set information, we define the **beam-pair support** as $\Omega = \{(g_r, g_t) : \mathbf{S}_k(g_r, g_t) \neq 0\}$, which is the set of indices corresponding to the dominant elements in $\mathbf{S}_k$. Note that Ω is independent of the subcarrier k because we construct the array response matrices $\mathbf{A}_{R,k}$ ($\mathbf{A}_{T,k}$) based on the same uniform grid $\mathcal{G}_R$ ($\mathcal{G}_T$) containing the quantized spatial AOAs (AODs). For massive MIMO systems, the beam-pair support of the channel is mainly determined by the geometry (positions and antenna topology) of the transmitter and receiver, and also the scattering clusters in the environment, and may change over time as a result of the dynamics in the propagation environment [15]. Here, we use a block-fading model where the channel remains fixed over the duration of a fading block, but may change across subsequent blocks. We use the superscript (i) to indicate the terms in the i -th fading block, e.g., $\Omega^{(i)}$. Due to temporal correlation, Ω varies slowly over time, so that $\Omega^{(i-1)}$ and $\Omega^{(i)}$ share a common support. We assume that L_{cm} is the minimum number of channel elements (common paths) that remain fixed between the previous and current fading blocks, hence are shared between $\Omega^{(i-1)}$ and $\Omega^{(i)}$, so that $|\Omega^{(i)} \cap \Omega^{(i-1)}| \geq L_{cm}$. Given a fixed number of channel paths $L \geq L_{cm}$, there are at most $L - L_{cm}$ paths changing from one

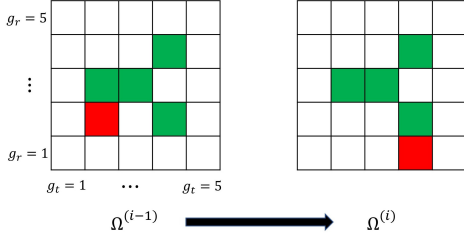


Fig. 1: The evolution of beamspace channel with $(L, L_{cm}) = (5, 4)$ is illustrated. Between two consecutive fading blocks, the green and red elements represent the common and changing elements, respectively.

fading block to the next one. Fig. 1 illustrates an example of the evolution of the beam index support, where the colored and white elements denote the dominant (non-zero) and negligible (zero) beamspace channel elements, respectively. Because of the slow channel variation, the beam-pair support sets $\Omega^{(i-1)}$ and $\Omega^{(i)}$ share $L_{cm} = 4$ common elements (shown in green), so that only one path may change from one fading block to the next (shown in red). This structure enables the use of the LS-CS approach to reduce the channel training overhead, which is discussed next.

B. Beam Training Scheme

The channels are assumed constant in each fading block with T_c consecutive subframes (channel uses), containing T_p pilot subframes for the channel training, and the remaining $T_c - T_p$ subframes for the data transmission, as in Fig. 2. For the pilot transmission, we exploit $K_p < K_o$ subcarriers with a comb-type arrangement, i.e., $\mathcal{P} = \{k = 1 + (v-1)\Delta_k : v = 1, \dots, K_p, \Delta_k = \lceil K_o/K_p \rceil\}$. We consider a hybrid precoder/combiner for MIMO-OFDM systems with N_t^{RF} and N_r^{RF} RF chains at the transmitter and receiver, respectively. We apply the frequency-flat beamforming training scheme [9], [11]. At the u -th subframe, the transmitter transmits the precoded signal $\mathbf{x}_u = \mathbf{F}_A \mathbf{F}_{D,u} \mathbf{s}_u$, where $\mathbf{F}_A \in \mathbb{C}^{N_t \times N_t^{RF}}$, $\mathbf{F}_{D,u} \in \mathbb{C}^{N_t^{RF} \times N_d}$, and $\mathbf{s}_u \in \mathbb{C}^{N_d \times 1}$. The receiver combines the measurement signal at the q -th stream by the combining vector $\mathbf{w}_q = \mathbf{W}_A \mathbf{w}_{D,q}$, where $\mathbf{W}_A \in \mathbb{C}^{N_r \times N_r^{RF}}$ and $\mathbf{w}_{D,q} \in \mathbb{C}^{N_r^{RF} \times 1}$, yielding the measurement signal on the k -th subcarrier denoted as

$$y_{q,u,k} = \mathbf{w}_q^H \mathbf{H}_k \mathbf{x}_u + \mathbf{w}_q^H \mathbf{n}_{q,u,k}, \quad (7)$$

where $\mathbf{n}_{q,u,k} \in \mathbb{C}^{N_r \times 1}$ is the additive noise vector. Assuming the transmitter employs distinct precoded pilots at T_p successive subframes and the receiver combines the signal through $Q_p \leq N_r^{RF}$ streams simultaneously, the $Q_p \times T_p$ combined signal on the k -th subcarrier is denoted as

$$\mathbf{Y}_k = \mathbf{W}^H \mathbf{H}_k \mathbf{X} + \tilde{\mathbf{N}}_k,$$

where $\mathbf{W} = [\mathbf{w}_1 \dots \mathbf{w}_{Q_p}] \in \mathbb{C}^{N_r \times Q_p}$ is the measurement combining matrix; $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_{T_p}] \in \mathbb{C}^{N_t \times T_p}$ is the pilot training matrix; $\tilde{\mathbf{N}}_k$ is the $Q_p \times T_p$ combined noise matrix.

C. Channel Training Protocol

We introduce the flow of the support tracking-based channel training. Initially, a conventional MIMO-OFDM channel estimation is used since no prior channel knowledge is available. Afterwards, a support-tracking based channel estimation

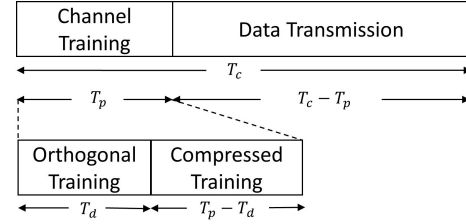


Fig. 2: The flow of MMV-LS-CS in one fading block.

(MMV-LS-CS) is used to estimate the channel aided by the previous support estimated in previous fading blocks. Then, the current support derived from the estimated channel is fed back to the transmitter for the next channel training. Note that the performance of the support tracking-based approach is related to the accuracy of the previous channel estimation. With an inaccurate previous support, the performance of the support tracking-based approach might deteriorate since the training is wasted on incorrect beam directions. To address this issue, we check the residual signal of the support-tracking based channel training algorithm. We use a predefined threshold that the remaining signal magnitude is compared to, and we declare that support tracking has failed if the magnitude exceeds the threshold.

III. PROPOSED LS-CS CHANNEL ESTIMATION

In this section, we propose a support-tracking based channel training on dual-wideband MIMO-OFDM. Section III-A proposes a LS-CS channel estimation. Section III-B proposes a channel refining algorithm to estimate the gains and time delays of the dominant channel paths.

A. MMV-LS-CS Channel Estimation

Our goal is to design the combining matrix $\mathbf{W}$ and training signal matrix $\mathbf{X}$ to reduce the overhead of the channel training with the help of the previous beam-pair support Ω^{pr} . We estimate the AOAs/AODs of dominant paths, and then refine the gains and time delays of the estimated paths. Here, we introduce the MMV-LS-CS algorithm for the dual-wideband MIMO-OFDM channel estimation. For a given training sequence length T_p , we separate the channel training into two stages, **orthogonal training** and **compressed training**, as in Fig. 2. The orthogonal training, of duration T_d , exploits the previous support learned in the previous fading block to estimate the channel on few beam directions via a least squares approach. Next, for the remaining duration $T_p - T_d$ of channel training, the compressed training applies a CS-based approach to estimate the residual channel, expected to be sparse since most dominant channel elements have been estimated in the orthogonal training phase. The pilot signal structure can be expressed as $\mathbf{X} = [\mathbf{X}_d \mathbf{X}_s]$, where $\mathbf{X}_d \in \mathbb{C}^{N_t \times T_d}$ and $\mathbf{X}_s \in \mathbb{C}^{N_t \times (T_p - T_d)}$. To exploit the support information at the transmitter and receiver separately, given the previous beam-pair support Ω^{pr} , we define the **previous receive support** as $\Psi_R^{pr} = \{g_r : (g_r, g_t) \in \Omega^{pr}\}$, representing the index set of the dominant receive steering vectors. The **previous transmit support** is defined in the same manner as $\Psi_T^{pr} = \{g_t : (g_r, g_t) \in \Omega^{pr}\}$.

First, the **orthogonal training** (MMV-LS) seeks to estimate the beamspace channel on the beam directions corresponding to the combination of previous support $(\Psi_R^{pr}, \Psi_T^{pr})$. To construct the training signal and combiner, we pick the array response matrix on the middle frequency of the bandwidth, and then concentrate the beam training on the directions corresponding to the previous support. The receive combiner is designed as

$$\mathbf{W}_d = \left(\mathbf{A}_{R, \frac{K_o+1}{2}} \right) [\mathbf{I}_{N_r}]_{\Psi_R^{pr}} \in \mathbb{C}^{N_r \times |\Psi_R^{pr}|}, \quad (8)$$

where $[\mathbf{I}_{N_r}]_{\Psi_R^{pr}}$ is the submatrix consisting of the columns of $\mathbf{I}_{N_r}$ indicated by the indices in Ψ_R^{pr} ; the training matrix is

$$\mathbf{X}_d = \left(\mathbf{A}_{T, \frac{K_o+1}{2}} \right) [\mathbf{I}_{N_t}]_{\Psi_T^{pr}} \tilde{\mathbf{X}}_d \in \mathbb{C}^{N_t \times T_d}, \quad (9)$$

where $\tilde{\mathbf{X}}_d \in \mathbb{C}^{T_d \times T_d}$ is a known orthonormal matrix having the length $T_d = |\Psi_T^{pr}|$, satisfying $\tilde{\mathbf{X}}_d^H \tilde{\mathbf{X}}_d = P_t \mathbf{I}_{T_d}$ with the average transmit power P_t . The received signal on the k -th subcarrier is derived as

$$\begin{aligned} \mathbf{Y}_k^{LS} &= \mathbf{W}_d^H \mathbf{H}_k \mathbf{X}_d + \tilde{\mathbf{N}}_k^{LS} \\ &= (\mathbf{A}_{R,k}^H \mathbf{W}_d)^H \mathbf{S}_k^{LS} (\mathbf{A}_{T,k}^H \mathbf{X}_d) + \tilde{\mathbf{N}}_k^{LS}. \end{aligned}$$

Using the relationship $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$, we have the vectorization of $\mathbf{Y}_k^{LS}$ as

$$\mathbf{y}_k^{LS} = \mathbf{\Theta}_k \mathbf{s}_k^{LS} + \tilde{\mathbf{n}}_k^{LS}, \quad (10)$$

where $\tilde{\mathbf{n}}_k^{LS} = \text{vec}(\tilde{\mathbf{N}}_k^{LS})$, $\mathbf{s}_k^{LS} = \text{vec}(\mathbf{S}_k^{LS})$, the dictionary matrix is $\mathbf{\Theta}_k = (\mathbf{A}_{T,k}^H \mathbf{X}_d)^T \otimes (\mathbf{A}_{R,k}^H \mathbf{W}_d)^H$, and $\otimes$ is the Kronecker product. The MMV-LS greedily collects the columns of $\mathbf{\Theta}_k$ leading to the minimum residual error after orthogonalization, expressed as

$$\Gamma := \Gamma \cup \arg_j \min_{\substack{\Lambda = \Gamma \cup j \\ j \in \mathcal{Q} \setminus \Gamma}} \sum_{k \in \mathcal{P}} \|\mathbf{y}_k^{LS} - [\mathbf{\Theta}_k]_{\Lambda} [\mathbf{\Theta}_k]_{\Lambda}^+ \mathbf{y}_k^{LS}\|_F^2,$$

where $\mathcal{Q} = \{(g_t - 1)G_r + g_r : g_t \in \Psi_T^{pr}, g_r \in \Psi_R^{pr}\}$ is the set of indices in $\mathbf{\Theta}_k$ corresponding to the previous support. $(\cdot)^+$ stand for the pseudo-inverse of a matrix. Even with the frequency-dependent dictionary matrices $\mathbf{\Theta}_k$, the beamspace channel vectors share a common support, enabling the MMV problem to estimate the dominant paths. The greedy collection is terminated when $|\Gamma| = L_{cm}$. With the collected indices of estimated paths, we derive the gains and time delays by the channel refining algorithm (**Algorithm 2** in Section III-B), and then reconstruct the channel $\hat{\mathbf{H}}_k^{LS}$.

Secondly, in the **compressed training** phase (MMV-CS), we apply the CS-based approach on the residual received signal obtained by subtracting the effect of the MMV-LS estimated channel $\hat{\mathbf{H}}_k^{LS}$. We design the combining matrix as

$$\mathbf{W}_s = \mathbf{\Phi}_W \in \mathbb{C}^{N_r \times Q_p}, \quad (11)$$

where $Q_p \leq N_r^{RF}$, and the pilot training matrix as

$$\mathbf{X}_s = \mathbf{\Phi}_X \in \mathbb{C}^{N_t \times (T_p - T_d)}, \quad (12)$$

where $\mathbf{\Phi}_W$ and $\mathbf{\Phi}_X$ are the measurement matrices satisfying the successful sparse recovery condition, e.g., RIP condition [17]. The combined signal matrix on the k -th subcarrier is

Algorithm 1 MMV-LS-CS Channel Estimation.

Input: measurement $(\mathbf{Y}_k^{LS}, \mathbf{Y}_k^{CS})$, combining and pilot training matrices $(\mathbf{W}_d, \mathbf{X}_d, \mathbf{W}_s, \mathbf{X}_s)$, the set of pilot subcarriers $\mathcal{P}$, previous support information $(\Psi_R^{pr}, \Psi_T^{pr})$

Output: $\hat{\mathbf{H}}_k^{rec}, \hat{\Psi}_R, \hat{\Psi}_T$

- 1: %%% MMV-LS aided by previous support %%%
 - 2: Calculate $\mathbf{\Theta}_k = (\mathbf{A}_{T,k}^H \mathbf{X}_d)^T \otimes (\mathbf{A}_{R,k}^H \mathbf{W}_d)^H$, $k \in \mathcal{P}$;
 - 3: $\mathbf{y}_k^{LS} = \text{vec}(\mathbf{Y}_k^{LS})$, $k \in \mathcal{P}$; $\Gamma = \emptyset$;
 - 4: $\mathcal{Q} = \{(g_t - 1)G_r + g_r : g_t \in \Psi_T^{pr}, g_r \in \Psi_R^{pr}\}$;
 - 5: **while** $|\Gamma| < L_{cm}$ **do**
 - 6: $\Gamma := \Gamma \cup \arg_j \min_{\substack{\Lambda = \Gamma \cup j \\ j \in \mathcal{Q} \setminus \Gamma}} \sum_{k \in \mathcal{P}} \|\mathbf{y}_k^{LS} - [\mathbf{\Theta}_k]_{\Lambda} [\mathbf{\Theta}_k]_{\Lambda}^+ \mathbf{y}_k^{LS}\|_F^2$;
 - 7: **end while**
 - 8: Reconstruct $\hat{\mathbf{H}}_k^{LS}$ by **Algorithm 2** with $(\mathbf{Y}_k^{LS}, \mathbf{\Theta}_k, \Gamma)$;
 - 9: %%% MMV-CS %%%
 - 10: Calculate $\mathbf{\Xi}_k = (\mathbf{A}_{T,k}^H \mathbf{X}_s)^T \otimes (\mathbf{A}_{R,k}^H \mathbf{W}_s)^H$, $k \in \mathcal{P}$;
 - 11: $\tilde{\mathbf{Y}}_k^{CS} = \mathbf{Y}_k^{CS} - \mathbf{W}_s^H \hat{\mathbf{H}}_k^{LS} \mathbf{X}_s$, $k \in \mathcal{P}$;
 - 12: $\mathbf{r}_k = \tilde{\mathbf{y}}_k^{CS} = \text{vec}(\tilde{\mathbf{Y}}_k^{CS})$, $k \in \mathcal{P}$;
 - 13: $\Upsilon = \emptyset$; $\mathcal{J} = \{1, \dots, G_r G_t\}$;
 - 14: **while** $|\Upsilon| < L - L_{cm}$ **do**
 - 15: $\Upsilon := \Upsilon \cup \arg \max_{j \in \mathcal{J} \setminus \Upsilon} \sum_{k \in \mathcal{P}} \|\mathbf{\Xi}_k(:, j)^H \mathbf{r}_k\|^2$;
 - 16: $\hat{\mathbf{g}}_k = \arg \min_{\mathbf{g}} \|\tilde{\mathbf{y}}_k^{CS} - [\mathbf{\Xi}_k]_{\Upsilon} \mathbf{g}\|_F^2$, $k \in \mathcal{P}$;
 - 17: $\mathbf{r}_k = \tilde{\mathbf{y}}_k^{CS} - [\mathbf{\Xi}_k]_{\Upsilon} \hat{\mathbf{g}}_k$, $k \in \mathcal{P}$;
 - 18: **end while**
 - 19: Reconstruct $\hat{\mathbf{H}}_k^{CS}$ by **Algorithm 2** with $(\tilde{\mathbf{Y}}_k^{CS}, \mathbf{\Xi}_k, \Upsilon)$;
 - 20: $\hat{\mathbf{H}}_k^{rec} = \hat{\mathbf{H}}_k^{LS} + \hat{\mathbf{H}}_k^{CS}$, $k = 1, \dots, K_o$;
 - 21: Derive $(\hat{\Psi}_R^{LS}, \hat{\Psi}_T^{LS})$ from Γ , and $(\hat{\Psi}_R^{CS}, \hat{\Psi}_T^{CS})$ from Υ ;
 - 22: $\hat{\Psi}_R = \hat{\Psi}_R^{LS} \cup \hat{\Psi}_R^{CS}$ and $\hat{\Psi}_T = \hat{\Psi}_T^{LS} \cup \hat{\Psi}_T^{CS}$;
-

$\mathbf{Y}_k^{CS} = \mathbf{W}_s^H \mathbf{H}_k \mathbf{X}_s + \tilde{\mathbf{N}}_k^{CS}$. Given the estimated $\hat{\mathbf{H}}_k^{LS}$, the residual signal matrix is

$$\begin{aligned} \tilde{\mathbf{Y}}_k^{CS} &= \mathbf{Y}_k^{CS} - \mathbf{W}_s^H \hat{\mathbf{H}}_k^{LS} \mathbf{X}_s \\ &= (\mathbf{A}_{R,k}^H \mathbf{W}_s)^H \tilde{\mathbf{S}}_k^{CS} (\mathbf{A}_{T,k}^H \mathbf{X}_s) + \tilde{\mathbf{N}}_k^{CS}, \end{aligned}$$

where $\tilde{\mathbf{S}}_k^{CS}$ is expected to be sparse because most elements of dominant channel paths have been estimated in $\hat{\mathbf{H}}_k^{LS}$. The vectorization of $\tilde{\mathbf{Y}}_k^{CS}$ is expressed as

$$\tilde{\mathbf{y}}_k^{CS} = \mathbf{\Xi}_k \tilde{\mathbf{s}}_k^{CS} + \tilde{\mathbf{n}}_k^{CS}, \quad (13)$$

where $\tilde{\mathbf{n}}_k^{CS} = \text{vec}(\tilde{\mathbf{N}}_k^{CS})$, $\tilde{\mathbf{s}}_k^{CS} = \text{vec}(\tilde{\mathbf{S}}_k^{CS})$, and the dictionary matrix is $\mathbf{\Xi}_k = (\mathbf{A}_{T,k}^H \mathbf{X}_s)^T \otimes (\mathbf{A}_{R,k}^H \mathbf{W}_s)^H$. The sparse recovery problem on the pilot subcarriers is formulated as

$$\arg \min_{\tilde{\mathbf{s}}_k^{CS}} \sum_{k \in \mathcal{P}} \|\tilde{\mathbf{s}}_k^{CS}\|_1, \text{ s.t. } \|\tilde{\mathbf{y}}_k^{CS} - \mathbf{\Xi}_k \tilde{\mathbf{s}}_k^{CS}\|_2 \leq \epsilon, \quad k \in \mathcal{P},$$

where $\epsilon > 0$ is a constant threshold. Therefore, the sparse signal $\tilde{\mathbf{s}}_k^{CS}$ can be estimated from $\tilde{\mathbf{y}}_k^{CS}$ using the sparse recovery. Among many available sparse recovery algorithms used to exploit the common channel support across subcarriers, we adopt the simultaneous OMP algorithm (SOMP) [18]. Similarly, with the collected indices of estimated paths, the gains and time delays of these paths are derived using the

Algorithm 2 Channel Refining Algorithm.

Input: measurement $\mathbf{Y}_k$, dictionary matrix $\mathbf{D}_k$, support set Γ

Output: $\hat{\mathbf{H}}_k, k = 1, \dots, K_o$

- 1: Derive $(i_{\theta,\ell}, i_{\phi,\ell})$ from $\Gamma, \ell = 1, \dots, L'$;
 - 2: Derive $\hat{\mathbf{q}}_k$ by solving (14), $k = 1, \dots, K_p$;
 - 3: Derive $\hat{z}_\ell$ by solving (15), $\ell = 1, \dots, L'$;
 - 4: Derive $\hat{\alpha}_\ell$ by solving (16), $\ell = 1, \dots, L'$;
 - 5: $\hat{\mathbf{H}}_k = \sqrt{N_r N_t} \sum_{\ell=1}^{L'} \hat{\alpha}_\ell (\hat{z}_\ell)^{k - \frac{K_o+1}{2}} \mathbf{A}_{R,k}(:, i_{\theta,\ell}) \mathbf{A}_{T,k}^H(:, i_{\phi,\ell})$;
-

channel refining algorithm (**Algorithm 2**), and the MMV-CS channel $\hat{\mathbf{H}}_k^{CS}$ is reconstructed. Finally, the MIMO-OFDM channel is $\hat{\mathbf{H}}_k^{rec} = \hat{\mathbf{H}}_k^{LS} + \hat{\mathbf{H}}_k^{CS}$. The proposed **MMV-LS-CS** algorithm is shown in **Algorithm 1**.

B. Path Gain And Time Delay Refinement

Here, we propose an approach to refine the gains and time delays of the estimated paths, to compensate the beam squint effect. We assume that the received signal matrix $\mathbf{Y}_k$ is measured by the combining matrix $\mathbf{W}$ and precoding matrix $\mathbf{X}$ on the set of pilot subcarriers $\mathcal{P} = \{k = 1 + (v-1)\Delta_k : v = 1, \dots, K_p, \Delta_k = \lceil K_o/K_p \rceil\}$. The optimization problem for the effective path gain vector is

$$\arg \min_{\mathbf{q}_k} \|\mathbf{y}_k - [\mathbf{D}_k]_\Gamma \mathbf{q}_k\|_F^2, \quad k \in \mathcal{P}, \quad (14)$$

where $\mathbf{y}_k = \text{vec}(\mathbf{Y}_k)$, and $\mathbf{D}_k = (\mathbf{A}_{T,k}^H \mathbf{X})^\top \otimes (\mathbf{A}_{R,k}^H \mathbf{W})^H$ is the dictionary matrix on the k -th subcarrier. The set Γ contains the indices of columns in $\mathbf{D}_k$ corresponding to the estimated paths. Assuming $|\Gamma| = L'$, the vector $\mathbf{q}_k = [\gamma_{1,k}, \dots, \gamma_{L',k}]^\top$ is the effective path gain vector, where $\gamma_{\ell,k} = \sqrt{N_r N_t} \alpha_\ell e^{-j2\pi f_k \tau_\ell}$. This is a least squares problem, which can be solved by $\hat{\mathbf{q}}_k = ([\mathbf{D}_k]_\Gamma)^\dagger \mathbf{y}_k$. With the generator $\{z_\ell = e^{-j2\pi \frac{B}{K_o} \tau_\ell}\}$, the effective path gain can be expressed as $\gamma_{\ell,k} = \sqrt{N_r N_t} \alpha_\ell (z_\ell)^{k - \frac{K_o+1}{2}}$. Therefore, given the estimated $\hat{\mathbf{q}}_k = [\hat{\gamma}_{1,k}, \dots, \hat{\gamma}_{L',k}]^\top$, the estimation of z_ℓ is formulated as

$$\arg \min_{z_\ell} \sum_{v=1}^{K_p-1} \left(z_\ell - \left(\frac{\hat{\gamma}_{\ell, 1+v\Delta_k}}{\hat{\gamma}_{\ell, 1+(v-1)\Delta_k}} \right)^{1/\Delta_k} \right)^2, \quad (15)$$

which is solved by $\hat{z}_\ell = \frac{1}{K_p-1} \sum_{v=1}^{K_p-1} \left(\frac{\hat{\gamma}_{\ell, 1+v\Delta_k}}{\hat{\gamma}_{\ell, 1+(v-1)\Delta_k}} \right)^{1/\Delta_k}$. The refined time delay $\hat{\tau}_\ell$ is derived by $\hat{\tau}_\ell = -\frac{K_o}{2\pi B} \angle \hat{z}_\ell$, where $\angle \hat{z}_\ell$ is the phase angle of $\hat{z}_\ell$. Given $\hat{z}_\ell$, we then estimate the gains α_ℓ by the following optimization problem

$$\arg \min_{\alpha_\ell} \sum_{k \in \mathcal{P}} \left(\hat{\gamma}_{\ell,k} - \sqrt{N_r N_t} \alpha_\ell (\hat{z}_\ell)^{k - \frac{K_o+1}{2}} \right)^2, \quad (16)$$

which can be solved by the least square approach, and the refined path gains $\hat{\alpha}_\ell$ are derived. Note that we could have the indices of estimated AOAs/AODs as $(i_{\theta,\ell}, i_{\phi,\ell})$, $\ell = 1, \dots, L'$, corresponding to the indices in Γ . Given the refined version of gains and time delays accompanied with the estimated AOAs/AODs, we reconstruct the MIMO channel by

$$\hat{\mathbf{H}}_k = \sqrt{N_r N_t} \sum_{\ell=1}^{L'} \hat{\alpha}_\ell (\hat{z}_\ell)^{k - \frac{K_o+1}{2}} \mathbf{A}_{R,k}(:, i_{\theta,\ell}) \mathbf{A}_{T,k}^H(:, i_{\phi,\ell}).$$

IV. NUMERICAL RESULTS

A. Experiment Setting

We consider a single-user MIMO scenario with $f_c = 60$ GHz and $B = 1$ GHz, employing half wave-length antenna spacing ULAs with array sizes $N_r = N_t = 64$. The number of receive and transmit RF chains are $N_r^{RF} = N_t^{RF} = 8$. The transmitter uses precoded pilots with $T_p = 14$ subframes and the receiver combines the signal through $Q_p = 8$ streams simultaneously. The number of subcarriers is $K_o = 128$, among which $K_p = 10$ subcarriers with a comb-type arrangement are selected for the pilot transmission. We consider the wideband geometric channel model with $L = 6$ channel paths, whose physical AOAs and AODs follow i.i.d. $\mathcal{U}(0, 2\pi)$; the delay spreads follow i.i.d. $\mathcal{U}(0, 100)$ ns; the gains α_ℓ are i.i.d. $\mathcal{CN}(0, 1)$. The matrices $\mathbf{A}_{R,k}/\mathbf{A}_{T,k}$ are constructed as in (4) and (5) with the uniform grids $\mathcal{G}_R/\mathcal{G}_T$ of sizes $G_r = G_t = 256$. To model the slow channel variation, we assume the number of common paths is $L_{cm} = 5$, which means that one channel path (among $L = 6$ paths in the previous fading block) disappears and another randomly selected path appears. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = \frac{\sum_k \|\mathbf{Y}_k - \tilde{\mathbf{N}}_k\|_F^2}{\sum_k \|\tilde{\mathbf{N}}_k\|_F^2},$$

where $\mathbf{Y}_k$ is the received signal matrix, and $\tilde{\mathbf{N}}_k$ is the combined noise matrix.

We compare MMV-LS-CS with the existing works **SCPD** [11] and **DWE-SCPD** [11]. SCPD neglects the spatial-wideband effect. DWE-SCPD is designed for the dual-wideband MIMO-OFDM, but the temporal correlation is not exploited; in contrast, the proposed MMV-LS-CS exploits the temporal correlation to provide a more accurate channel estimation for the dual-wideband MIMO-OFDM. In addition, we consider two baseline schemes as follows: **Genie-aided LS** assumes the current channel support is known, with the training on the subspace of the support, and the channel refining algorithm is applied. **MMV-CS** estimates the channel without previous support by the MMV-CS in **Algorithm 1**.

B. Estimation Accuracy

To evaluate the channel estimation accuracy, we define the normalized mean square error (NMSE) of the channel as

$$\text{NMSE} = \frac{\sum_{k=1}^{K_p} \|\mathbf{H}_k - \hat{\mathbf{H}}_k\|_F^2}{\sum_{k=1}^{K_p} \|\mathbf{H}_k\|_F^2}.$$

In Fig. 3, we plot the NMSE of the channel versus the SNR. Genie-aided LS attains NMSE = 0.04 when SNR ≥ 10 dB, which provides a lower bound for NMSE. For an SNR = 20dB, MMV-LS-CS attains NMSE = 0.15, as opposed to 0.43 for DWE-SCPD, 0.45 for SCPD, and 0.89 for MMV-CS. The NMSE of MMV-LS-CS outperforms the state-of-the-art approach (DWE-SCPD) in the moderate and low SNR regions. DWE-SCPD is susceptible to noise since its beam training procedure is not performed jointly on pilot subcarriers; in comparison, MMV-LS-CS does the channel estimation based on the measurements on all pilot subcarriers.

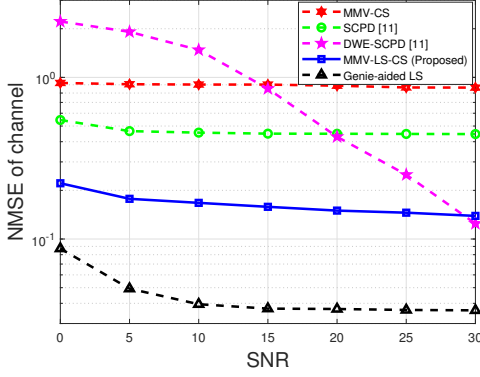


Fig. 3: NMSE of channel versus the SNR.

C. Spectral Efficiency

Here, we first define the achievable transmission rate as

$$R = \sum_{k=1}^{K_o} \frac{B}{K_o} \log_2 \det \left(\mathbf{I}_{N_s} + \frac{P_t}{N_s} \mathbf{R}_{\hat{\mathbf{n}}_k}^{-1} \hat{\mathbf{W}}_k^H \mathbf{H}_k \hat{\mathbf{F}}_k \hat{\mathbf{F}}_k^H \mathbf{H}_k^H \hat{\mathbf{W}}_k \right),$$

where P_t is the average transmit power for each transmission. The number of data streams is assumed as $N_s = 6$. The combiner $\hat{\mathbf{W}}_k$ (or precoder $\hat{\mathbf{F}}_k$) is derived by the directions of eigenvectors of $\hat{\mathbf{H}}_k \hat{\mathbf{H}}_k^H$ (or $\hat{\mathbf{H}}_k^H \hat{\mathbf{H}}_k$). The post-processing noise covariance matrix is $\mathbf{R}_{\hat{\mathbf{n}}_k} = \mathbb{E}[\hat{\mathbf{n}}_k \hat{\mathbf{n}}_k^H]$, where $\hat{\mathbf{n}}_k = \hat{\mathbf{W}}_k^H \mathbf{n}$ with the additive complex Gaussian noise $\mathbf{n}$. We assume the length of training sequence $T_p = 14$, so the fraction of time for data transmission is identical for all schemes. Thus, we define the spectral efficiency as $\frac{R}{BN_s}$ (bit/s/Hz/stream). In Fig. 4, we evaluate the spectral efficiency versus the SNR. **Full CSI** attains the largest spectral efficiency because its $\hat{\mathbf{W}}_k/\hat{\mathbf{F}}_k$ are derived from the perfect CSI. Genie-aided LS has almost the same performance as Full CSI, showing that the channel refining algorithm accurately estimates the gains and time delays if the correct AOA/AODs of channel paths are given. For an SNR = 15 dB, the spectral efficiency of MMV-LS-CS is 6.24 bit/s/Hz/stream, which outperforms the state-of-the-art (DWE-SCPD) by 0.73 bit/s/Hz/stream, SCPD by 0.74 bit/s/Hz/stream, MMV-CS by 3.99 bit/s/Hz/stream. Although DWE-SCPD has better spectral efficiency in the high SNR regime (SNR > 20dB), MMV-LS-CS is more robust to noise due to the low NMSE in this configuration as in Fig. 3.

V. CONCLUSION

In this work, we proposed a CS-aided training for time-varying channels in dual-wideband MIMO-OFDM. We constructed the frequency-dependent array response matrices based on the same grid to maintain the common channel support across OFDM subcarriers, and then recovered the sparse signal from multiple observations sharing a common support. We developed a novel channel estimation procedure that leverages slow variations in the beam support, termed MMV-LS-CS, which applies MMV-LS on the measurements based on the previous support, and MMV-CS on the residual error to estimate time-varying components. We proposed a channel refining algorithm to reconstruct the channel by estimating the gains and time delays from the measurements on pilot subcarriers. Numerical results showed that MMV-LS-CS

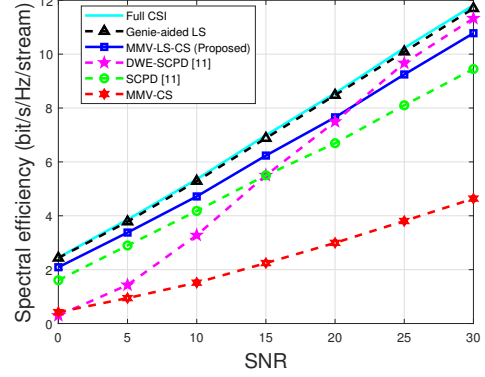


Fig. 4: Spectral efficiency (bit/s/Hz/stream) versus the SNR.

provides more robust and accurate channel estimation and improved spectral efficiency than the state-of-the-art approach in the time-varying dual-wideband MIMO-OFDM.

REFERENCES

- [1] R. W. Heath *et al.*, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.
- [2] E. G. Larsson *et al.*, "Massive MIMO for Next Generation Wireless Systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
- [3] M. Hussain and N. Michelusi, "Energy-Efficient Interactive Beam Alignment for Millimeter-Wave Networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 838–851, Feb. 2018.
- [4] M. Hussain *et al.*, "Mobility and Blockage-aware Communications in Millimeter-Wave Vehicular Networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13 072–13 086, Nov. 2020.
- [5] T.-H. Chou *et al.*, "Millimeter Wave Beam Recommendation via Tensor Completion," in *Proc. IEEE Int. Conf. Commun.*, 2020, pp. 1–6.
- [6] —, "Fast Position-Aided MIMO Beam Training via Noisy Tensor Completion," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 3, pp. 774–788, Apr. 2021.
- [7] S. Hur *et al.*, "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] A. J. Duly, T. Kim, D. J. Love, and J. V. Krogmeier, "Closed-Loop Beam Alignment for Massive MIMO Channel Estimation," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1439–1442, Aug. 2014.
- [9] A. Alkhateeb *et al.*, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [10] B. Wang *et al.*, "Spatial-Wideband Effect in Massive MIMO with Application in mmWave Systems," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 134–141, Dec. 2018.
- [11] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Tensor-Based Channel Estimation for Millimeter Wave MIMO-OFDM with Dual-Wideband Effects," *IEEE Trans. Commun.*, Jul. 2020.
- [12] Y. Han, J. Lee, and D. J. Love, "Compressed Sensing-Aided Downlink Channel Training for FDD Massive MIMO Systems," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2852–2862, Jul. 2017.
- [13] S. G. Larew and D. J. Love, "Adaptive Beam Tracking With the Unscented Kalman Filter for Millimeter Wave Communication," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1658–1662, Nov. 2019.
- [14] Z. Gao *et al.*, "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.
- [15] S. Wu *et al.*, "A Non-Stationary Wideband Channel Model for Massive MIMO Communication Systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1434–1446, Mar. 2015.
- [16] C. A. Balanis, *Antenna Theory: Analysis and Design*. Hoboken, NJ, USA: Wiley, 2016.
- [17] E. J. Candes, J. K. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.
- [18] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal processing*, vol. 86, no. 3, pp. 572–588, Mar. 2006.