Dual-Wideband Time-Varying Sub-Terahertz Massive MIMO Systems: A Compressed Training Framework

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

Abstract

Designers of beyond-5G systems are planning to use new frequencies in the millimeter wave (mmWave) and sub-terahertz (sub-THz) bands to meet ever-increasing demand for wireless broadband access. Sub-THz communication, however, will come with many challenges of mmWave communication and new challenges associated with the wider bandwidths, larger numbers of antennas and harsher propagation characteristics. Notably the frequency- and spatial-wideband (dual-wideband) effects are significant at sub-THz. To address these challenges, this paper presents a compressed training framework to estimate the sub-THz time-varying MIMO-OFDM channels. A set of frequency-dependent array response matrices are constructed, enabling the channel recovery from multiple observations across subcarriers via multiple measurement vectors (MMV). Capitalizing on the temporal correlation, MMV least squares (MMV-LS) is designed to estimate the channel on the previous beam index support, followed by MMV compressed sensing (MMV-CS) on the residual signal to estimate the time-varying channel components. Furthermore, a channel refinement algorithm is proposed to estimate the path coefficients and time delays of the dominant paths. To reduce the computational complexity, a sequential search method using hierarchical codebooks is proposed for greedy beam selection. Numerical results show that MMV-LS-CS achieves a more accurate and robust channel estimation than state-of-the-art algorithms on time-varying dual-wideband MIMO-OFDM.

Index Terms

Wideband communication, sub-THz, MIMO, time-varying channel estimation, compressed sensing.

This work was supported in part by the National Science Foundation under grants CNS-1642982, CCF-1816013, EEC-1941529 and CNS-2129015. A preliminary version of this paper has been accepted at IEEE Globecom 2021 [1].

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.
I. INTRODUCTION

Future wireless applications will require networks that can provide high rates, improved power consumption for devices, and low latencies in a wide range of deployment scenarios [2]. These requirements motivate policy and technical research to make spectrum at higher frequencies, outside the popularly used sub-6 GHz spectrum, commercially available for wireless broadband thanks to large bandwidth availability. In 5G, this policy and technical work culminated in the standardization and (at least initial) deployment of communications in bands within the millimeter wave (mmWave) spectrum.

In the sub-terahertz (sub-THz) spectrum, roughly defined as $100 - 300$ GHz, there is approximately $21.2$ GHz available for wireless broadband [3]. Making these bands “usable” could alleviate backhaul and access concerns far into the future [3]–[5]. Unfortunately, sub-THz communication poses signal processing and wideband communication challenges unique to the spectrum. Most notable are the need for dramatic increases in antenna aperture and bandwidth and signal processing to deal with the unique propagation features.

**Signal Propagation and Large Antenna Arrays:** Friis’ equation dictates that the spreading loss increases with the operating frequency, which thus becomes a more severe issue in sub-THz than in mmWave or sub-6 GHz systems [6]. Molecular absorption loss, e.g., due to propagation in water vapor, also becomes more pronounced as the frequency increases [6], [7]. In non-line-of-sight (NLOS) scenarios, the sub-THz reflection coefficient [8] should be taken into account when modeling the channel conditions. Due to the high total pathloss when these effects combine together, sub-THz systems require the use of highly directional beamforming to compensate for the severe signal attenuation, which can be achieved using massive multiple-input multiple-output (MIMO) systems [9], [10]. Thanks to the short wavelength of sub-THz signals, massive MIMO is possible even with a small form factor.

**Orders-of-Magnitude Larger Bandwidths and Wideband Effects:** The frequency-wideband and spatial-wideband effects of the MIMO channel will be exacerbated in sub-THz deployments. The frequency-wideband effect is caused by the delay spread of the multipath channel and the accompanying frequency-selective channel responses. In addition, the spatial-wideband effect arises from differences in time-delays across the antenna aperture. This primarily arises because the angles of arrival of the propagation paths observed by the receiver vary with the operating frequencies, known as the beam squint effect [11]–[15]. Most existing works focusing on sub-6
GHz or mmWave systems [16]–[24] consider only the frequency-wideband effect and ignore the spatial-wideband effect since it is negligible when the available bandwidth is not significantly wide, which is common in current and past wireless deployments. In this paper, we are concerned with both effects, meaning that we address the dual-wideband or spatial-frequency wideband effect. A few recent works [11]–[14] investigate these combined effects by considering communication and channel estimation in mmWave systems with bandwidths up to 1GHz. However, sub-THz communication can utilize one order of magnitude more bandwidth than mmWave communication [3] and is thus more severely affected by the dual-wideband effect, which must be taken into account with an accurate model of the communication system’s channel.

**Few dominant paths and channel sparsity:** To reap the beamforming gain and compensate for the severe sub-THz channel conditions, massive MIMO requires accurate channel state information (CSI) acquired via channel estimation. Yet, traditional MIMO channel estimation techniques are impractical due to the prohibitive overhead that comes with (non-adaptive) omnidirectional training over a large number of antennas. The burden of MIMO channel estimation can be reduced by exploiting the fact that sub-THz bands exhibit a high degree of channel sparsity, with few dominant propagation clusters. For instance, [25] reported an average of 6 clusters and 4 multipath components (MPCs) per cluster at 140GHz, compared to 8 clusters and 5 MPCs/cluster at 28GHz. One approach is to utilize a predetermined set of beams (e.g., a codebook), usually designed to allocate power in specific directions, and to tailor the beam training algorithm to this lower-dimensional beam set instead of on the true higher-dimensional MIMO channel. For example, the most-discussed approach [26], [27] is to scan over some set of candidate beams and estimate the strongest (e.g., through received power measurement), but even this approach incurs a prohibitively large overhead due to the typically large size of the beamforming codebook.

To reap the promising performance gains available from the large sub-THz bandwidth, massive MIMO systems must compensate for the sub-THz channel impairments while dealing with dual-wideband effects caused by the wideband communications. MIMO orthogonal frequency division multiplexing (MIMO-OFDM) has been envisioned as an important tool in combating the frequency-wideband effect and the inter-symbol interference of the multipath channel. For MIMO-OFDM channel estimation, the work [21] addresses the problem of recovering the channel from multiple measurements among subcarriers sharing a common support, known as multiple measurement vectors (MMV). Yet, the work [21] neglects the spatial-wideband (or beam squint)
effect, which breaks the common support assumption across the subcarriers and degrades the estimation accuracy. This issue is addressed in the work [15] by designing a set of frequency-dependent dictionary matrices to preserve the common support across the subcarriers, mitigating the channel estimation losses incurred from the spatial-wideband effect.

Furthermore, massive MIMO CSI acquisition is challenging in a time-varying scenario since the channel needs to be periodically trained, causing a severe overhead. The MIMO channel is generally determined by the geometry (positions and antenna geometry) of the transmitter and receiver and the unique effects of the propagation environment on the electromagnetic signal, which are time-varying. One approach is to model the dynamic behavior of the channel as a birth-death process of MPCs [28], [29], which is a way of taking into account the temporal correlation of the surviving MPCs. The channel sparsity, combined with the slow temporal variations, result in a slowly-varying beamspace support, which could be exploited for MIMO channel estimation [30]. Channel estimation algorithms exploiting the temporal correlation of the channel, via the use of a common (or slightly varying) channel support over time, have been studied for the narrowband MIMO in [30] and for the frequency-wideband multiuser MIMO-OFDM in [21]. Nevertheless, these techniques cannot be directly applied to time-varying dual-wideband MIMO-OFDM channels due to the beam squint effect and frequency-dependent path gains, which may harm the estimation performance. To address these issues in our work, we propose a new channel training algorithm (MMV-LS-CS) for time-varying MIMO channels having the dual-wideband effect and a channel refinement algorithm to improve the estimation performance by deriving the path coefficients and time delays of the dominant paths across the pilot subcarriers.

A. Prior Works

Over the last decade, much research has focused on MIMO channel estimation in mmWave and (sub-)THz bands. Various approaches based on the idea of beam alignment have been investigated in recent years, including feedback-based schemes [31]–[38], data-assisted schemes [39]–[42], to multipath estimation [11]–[24], [30]. Feedback-based schemes adapt the beam training according to the feedback information sent from the receiver in an online fashion [31] or learn and leverage the UEs’ mobility as in [32], [33]. Data-assisted schemes perform the beam training by leveraging side information from other available sources, e.g., GPS positional information [42], lower-frequency communication [41], radar [39], and LIDAR [40]. Multipath
estimation schemes exploit the channel sparsity of the MIMO channel via compressed sensing (CS) to acquire the associated channel parameters, e.g., angles of arrival (AOAs), angles of departure (AODs), time delays, and path gains. The channel training proposed in this work is a form of multipath estimation.

The narrowband MIMO channel estimation problem has been investigated in [16], [17]. The work [16] proposes an adaptive algorithm for mmWave massive MIMO channel estimation using a hierarchical multi-resolution codebook. With an adaptive structure, the work [17] proposes channel estimation using a compressive beacon codebook with different pseudorandom phase settings of the antenna arrays. In contrast to [16], [17], our work focuses on wideband MIMO channel estimation, taking advantage of the abundant available spectrum in sub-THz bands.

Initial work on wideband MIMO channel estimation considering the frequency-wideband effect is available in [18]–[24]. Assuming channel sparsity, the work [18] formulates MIMO-OFDM channel estimation as a sparse recovery problem, solved via orthogonal matching pursuit (OMP). The work [19] applies a tensor decomposition to the training signal with multiple dimensions corresponding to the beams and subcarriers and proposes a CANDECOMP/PARAFAC decomposition-based algorithm. The work [20] proposes a CS-aided channel estimation using the Tucker tensor as a compressible representation and reconstruction by tensor-OMP (T-OMP).

Differently from these works, our work develops a channel training algorithm on dual-wideband MIMO-OFDM, also considered in [11]–[15]. The challenges of dual-wideband MIMO in mmWave are outlined in [11] and a channel estimation strategy exploiting the asymptotic characteristics of the channel is proposed. The work [12] presents a CS-aided channel estimation on dual-wideband MIMO-OFDM exploiting uplink/downlink channel reciprocity. The work in [13] considers the block sparsity of the beam squint effect to design a CS-based channel estimation. A tensor-based channel training using the Vandermonde constraint and spatial smoothing method is proposed in [14]. The dual-wideband effect in THz communications with uniform planar arrays is studied in [15], and an algorithm (GSOMP) to recover the channel by simultaneous OMP exploiting the common support across the subcarriers preserved by the frequency-dependent dictionary matrices is proposed (evaluated numerically in Section IV). Prior work, however, has not utilized the fact that in a time-varying scenario the channel tends to manifest temporal correlation between consecutive fading blocks, which may be leveraged to improve channel estimation. In our work, we devise a channel training algorithm in dual-wideband MIMO-OFDM that uses the temporal correlation and common support across the frequencies to improve the estimation performance.
B. Contributions

We develop a support tracking-based channel training framework that utilizes the previous estimated channel support in a time-varying dual-wideband MIMO-OFDM system in sub-THz bands. To address the spatial-wideband effect, a set of frequency-dependent array response matrices are constructed to preserve the common channel support, leading to the MMV problem formulation of recovering the sparse beamspace channel from multiple observations across the pilot subcarriers. Due to the slow channel variations, the channel supports of consecutive fading blocks tend to share many common elements, enabling the LS-CS residual approach [43], [44], substituting the CS on the measurements with the CS on the LS residual signal.

In a nutshell, the contributions of this paper are as follows:

- We propose a support tracking-based wideband channel training, MMV-LS-CS, on time-varying channel with the dual-wideband effect using the estimated previous channel support.
- We propose a channel refinement algorithm to improve the estimation performance by deriving the path gains and time delays of the estimated paths across the pilot subcarriers.
- We propose a sequential search method using a hierarchical codebook for greedy beam selection in the proposed channel training to reduce the computational complexity.

The rest of the paper is organized as follows. Section II introduces the channel and system model. Section III proposes the MMV-LS-CS channel training algorithm, followed by its complexity analysis. Section IV shows the numerical results, and Section V concludes the paper.

Notation: Bold lowercase letters \( \mathbf{x} \) and bold uppercase letters \( \mathbf{X} \) denote vectors and matrices, respectively; \( \mathbf{X}^\top \), \( \mathbf{X}^H \), \( \mathbf{X}^+ \), \( \text{vec}(\mathbf{X}) \), \( \det(\mathbf{X}) \) represent the transpose, conjugate transpose, Moore-Penrose pseudo-inverse, vectorization, and determinant of \( \mathbf{X} \), respectively; \( [\mathbf{X}]_\Gamma \) (respectively, \( [\mathbf{x}]_\Gamma \)) is the submatrix with columns of \( \mathbf{X} \) (the subvector with elements of \( \mathbf{x} \)) associated with the indices set \( \Gamma \); \( (\mathbf{X})_n \) is the \( n \)-th element of \( \mathbf{X} \); \( |\Psi| \) is the cardinality of the set \( \Psi \); \( \otimes \) denotes the Kronecker product; \( \mathbf{I}_M \) is an \( M \times M \) identity matrix; \( \mathcal{F}\{\cdot\} \) is the continuous Fourier transform.

II. System Model

Clearly defining the system model in sub-THz is critical. Section II-A describes a single-user time-varying wideband MIMO-OFDM channel model in sub-THz bands. Section II-B discusses its frequency selectivity. Section II-C introduces the signal model of the hybrid transceiver.

A. Channel Model

We consider a MIMO-OFDM system with bandwidth \( B \), carrier frequency \( f_c \), and \( K_o \) subcarriers. The receiver employs a uniform planar array (UPA) with \( N_r = N_{or} \times N_{hr} \) antennas,
and the transmitter employs a UPA with \( N_t = N_vt \times N_ht \) antennas. Both arrays are configured as depicted in Fig. 1 with antenna spacing \( d \). The receive and transmit antennas are indexed as \( \mathbf{n}_r = (n_{vr}, n_{hr}) \) and \( \mathbf{n}_t = (n_{vt}, n_{ht}) \), respectively. We adopt a wideband geometric massive MIMO channel model with \( L \) scattering paths between the transmitter and receiver \([45]\). The \( \ell \)-th propagation path (\( \ell = 1, \ldots, L \)) is characterized by the time-varying channel response \( \beta_\ell(u) \) due to the frequency-selective attenuation from the environment \([46]\), the physical AOA \( \theta^\ell_r = (\theta^\ell_{pr}, \theta^\ell_{ar}) \) and physical AOD \( \theta^\ell_t = (\theta^\ell_{pt}, \theta^\ell_{at}) \), specifying the polar and azimuth angles (indexed by “p” and “a”, respectively). The baseband signal received at the \( n_r \)-th antenna is

\[
\mathbf{r}_{n_r}(u) = \sum_{\ell=1}^{L} \sum_{n_t} \bar{x}_{\ell,n_t}(u - \tau_{\ell,n_r,n_t}) e^{-j2\pi f_c \tau_{\ell,n_r,n_t}} + v_{n_r}(u),
\]

where \( \tau_{\ell,n_r,n_t} \) is the time delay of the \( \ell \)-th channel path between the \( n_r \)-th transmit antenna and \( n_r \)-th receive antenna, \( v_{n_r}(u) \) is the additive noise at the \( n_r \)-th receive antenna, and \( \bar{x}_{\ell,n_r}(u) = x_{n_t} \ast \beta_\ell(u) \) is the convolution of the baseband signal \( x_{n_t}(u) \) transmitted at the \( n_t \)-th antenna with the channel response \( \beta_\ell(u) \). Due to the increasing scale of the two antenna arrays, the propagation delay of waves traveling across the array aperture is non-negligible, thus we denote the time delay as \([48]\)

\[
\tau_{\ell,n_r,n_t} = \tau_\ell + \tau_{n_r}(\theta^\ell_r) - \tau_{n_t}(\theta^\ell_t),
\]

where \( \tau_\ell \) is the reference path delay of the \( \ell \)-th scattering path on the first transmit and receive antenna pair \( \mathbf{n}_t = \mathbf{n}_r = (1, 1) \). The propagation delay of the \( n_r \)-th receive (\( n_t \)-th transmit) antenna across the UPA aperture with respect to the \( (1, 1) \)-th receive (transmit) antenna is represented as \( \tau_{n_\zeta}(\theta^\ell_\zeta) = \frac{d}{c} \left( (n_{h_\zeta} - 1) \cos \theta^\ell_{a_\zeta} \sin \theta^\ell_{p_\zeta} + (n_{v_\zeta} - 1) \sin \theta^\ell_{a_\zeta} \sin \theta^\ell_{p_\zeta} \right) \), for \( \zeta \in \{r, t\} \) \([47]\). By applying the continuous Fourier transform on (1), we obtain the baseband signal in the
frequency domain at the baseband frequency $f$, as

$$
(R(f))_n = \mathcal{F}\{r_n(u)\} = \sum_{n_t} \sum_{\ell=1}^L \alpha_\ell(f)e^{-j2\pi(f-f_c)(\tau_n(\phi^{\ell}_t) - \tau_n(\phi^{\ell}_r))}e^{-j2\pi f\tau}(X(f))_{n_t} + (V(f))_{n_r},
$$

where $(X(f))_{n_t} = \mathcal{F}\{x_{n_t}(u)\}$, $(V(f))_{n_r} = \mathcal{F}\{v_{n_r}(u)\}$, and we have defined the equivalent baseband path coefficient

$$
\alpha_\ell(f) \triangleq B_\ell(f)e^{-j2\pi f\tau},
$$

where $B_\ell(f) = \mathcal{F}\{\beta_\ell(u)\}$ is the frequency-domain path coefficient of the $\ell$-th path. By stacking up the MIMO signal on UPA antennas, we obtain the frequency-dependent input-output relationship for the MIMO channel as

$$
r(f) = H(f)x(f) + v(f),
$$

where $r(f) = \text{vec}(R(f)) \in \mathbb{C}^{N_r \times 1}$ is the received signal vector, $x(f) = \text{vec}(X(f)) \in \mathbb{C}^{N_t \times 1}$ is the transmit signal vector, and $v(f) = \text{vec}(V(f)) \in \mathbb{C}^{N_r \times 1}$ is the additive noise vector. The frequency response of the baseband MIMO channel $H(f) \in \mathbb{C}^{N_r \times N_t}$ is

$$
H(f) = \sqrt{N_rN_t}\sum_{\ell=1}^L \alpha_\ell(f)b_{N_r}(\psi^{\ell}_{hr},\psi^{\ell}_{vr};f)b^H_{N_t}(\psi^{\ell}_{ht},\psi^{\ell}_{vt};f)e^{-j2\pi f\tau},
$$

where $\psi^{\ell}_{hr}$ (respectively, $\psi^{\ell}_{ht}$) is the horizontal spatial AOA (AOD) of the $\ell$-th path, $\psi^{\ell}_{vr}$ ($\psi^{\ell}_{vt}$) is the vertical spatial AOA (AOD) of the $\ell$-th path, defined as $\psi^{\ell}_{hr} = \frac{d}{\lambda_c}\cos\theta^{\ell}_{ac}\sin\theta^{\ell}_{mc}$ and $\psi^{\ell}_{vr} = \frac{d}{\lambda_c}\sin\theta^{\ell}_{ac}\sin\theta^{\ell}_{mc}$, for $\zeta \in \{r,t\}$; $b_{N_r}(\psi^{\ell}_{hr},\psi^{\ell}_{vr};f)$, $b_{N_t}(\psi^{\ell}_{ht},\psi^{\ell}_{vt};f)$ denote the receive and transmit spatial-frequency UPA vectors, respectively, given by

$$
b_{N_r}(\psi^{\ell}_{hr},\psi^{\ell}_{vr};f) = a_{N_{hr}}(\psi^{\ell}_{hr};f) \otimes a_{N_{vr}}(\psi^{\ell}_{vr};f), \text{ for } \zeta \in \{r,t\}
$$

and the array response vector along the horizontal/vertical dimension is defined as

$$
a_N(\psi;f) = \frac{1}{\sqrt{N}}[1,e^{-j2\pi(1+1/f_c)\psi},\ldots,e^{-j2\pi(N-1)(1+1/f_c)\psi}]^T
$$

for $N$ antennas in that dimension.

### B. Frequency-Selectivity of the MIMO Channel in Sub-THz Bands

Eq. (6) models the frequency-domain MIMO channel $H(f)$, which is frequency-selective due to the frequency-wideband effect, spatial-wideband effect, and path gain. The frequency-
wideband effect derives from the frequency-selective channel response caused by the time delays \(\tau_\ell\) of the multipath fading channel and has been widely investigated in the existing works \([18]–[24]\) for wideband MIMO systems.

The spatial-wideband effect arises from the distinct time delays across the antenna array for a given channel path (as in (2)). Hence, for a given channel path associated with the array response vector \(a_N(\psi; f)\) having a spatial angle \(\psi\) (as in (8)), the effective spatial angle of the steering vector is \((1 + \frac{f}{f_c})\psi\), inducing a frequency-dependent phase shift \(\frac{f}{f_c}\psi\) that is called the beam squint effect. A system that encounters both the frequency- and spatial-wideband effects is called a dual-wideband system. Sub-THz system is more susceptible to the dual-wideband effect owing to the several orders of magnitude increase in the bandwidths in higher frequency spectrum.

For the path gain of the channel, signals propagating in sub-THz bands suffer from a spreading loss \(L_{\text{spread}}\), absorption loss \(L_{\text{abs}}\) and reflection coefficient \(\Gamma_\ell\) (with \(\Gamma_\ell = 1\) for a LOS path). The equivalent path gain of the \(\ell\)-th channel path (as in (4)) in the sub-THz bands \([6]\) is defined as

\[
|\alpha_\ell(f)|^2 = |B_\ell(f)|^2 = |\Gamma_\ell(f)|^2 L_{\text{spread}}(f, D_\ell) L_{\text{abs}}(f, D_\ell), \tag{9}
\]

where \(f\) is the baseband frequency, and \(D_\ell\) is the distance covered by the \(\ell\)-th path. The spreading loss (viewed as the path loss in conventional wireless communication) models the attenuation incurred during wave propagation \([6]\) and follows Friis’ transmission formula

\[
L_{\text{spread}}(f, D_\ell) = \left(\frac{c}{4\pi (f_c + f) D_\ell}\right)^2, \tag{10}
\]

where \(c\) is the speed of light. The absorption loss is the attenuation of the signal suffering from the molecular absorption in the sub-THz bands, which arises due to water vapor molecules \([6], [7]\), defined as

\[
L_{\text{abs}}(f, D_\ell) = e^{-\kappa_a(f_c + f) D_\ell}, \tag{11}
\]

where \(\kappa_a(\cdot)\) is the frequency-dependent molecular absorption coefficient and depends on the propagation medium at a molecular level \([7]\). The reflection coefficient of a LOS path \((\ell = 1)\) is assumed \(\Gamma_1(f) = 1\). For the NLOS paths \((\ell = 2, \ldots, L)\), we consider the single-bounce reflected rays model with the reflection coefficient defined according to \([6], [8]\) as

\[
\Gamma_\ell(f) = \frac{Z(f_c + f) \cos \varphi_{i,\ell} - Z_0 \cos \varphi_{r,\ell}}{Z(f_c + f) \cos \varphi_{i,\ell} + Z_0 \cos \varphi_{r,\ell}} - \frac{1}{2} \left(\frac{4\pi (f_c + f) \sigma_{\ell} \cos \varphi_{i,\ell}}{c}\right)^2, \tag{12}
\]
where $Z(\cdot)$ is the wave impedance of the reflecting material and a function of the operating frequency, $Z_o = 377 \, \Omega$ is the wave impedance in free space, $\varphi_{r,\ell} = \arcsin\left(\frac{Z}{Z_o} \sin \varphi_{i,\ell}\right)$ is the angle of refraction, $\varphi_{i,\ell}$ is the angle of incidence (or reflection), and $\sigma_r$ is the standard deviation of the reflecting surface characterizing the material roughness.

By combining the spreading loss, absorption loss, and reflection coefficient together, the overall pathloss $\frac{1}{|\alpha_\ell(f)|^2}$ exhibits the frequency-dependent behavior depicted in Fig 2. In the figure, we also plot the approximation $\propto (1 + f/f_c)^2$ (up to a scaling factor), which shows a good fit. Therefore, we approximate the complex baseband path coefficient as $\alpha_{\ell}(f) \approx \alpha_{\ell}'(0) \left(1 + f/f_c\right)^2$, where $\alpha_{\ell}'$ is the reference path coefficient at the carrier frequency ($\alpha_{\ell}(0) = \alpha_{\ell}'$), with frequency-independent phase shift [6].

C. Hybrid Transceiver and Signal Model

To facilitate the trade-off between performance and cost of the hardware implementation, we consider a hybrid precoder and combiner design for MIMO-OFDM systems with $N_{RF_t}^{t}$ and $N_{RF_r}^{r}$ RF chains at the transmitter and receiver, respectively, illustrated in Fig. 3 [14], [15]. The fully connected network of phase shifters are implemented in the RF precoder and combiner. Note that our proposed training framework in Section III-A is general and applicable to the new hybrid structures [15], [50] considering the beam squint effect in the THz bands. On a MIMO-OFDM system with $K_o$ subcarriers, we denote the baseband frequency of the $k$-th subcarrier as $f_k = \left(k - \frac{K_o + 1}{2}\right)\frac{B}{K_o}$, $k = 1, \ldots, K_o$, and the MIMO channel on the $k$-th subcarrier as $H_k = H(f_k)$.

At the transmitter, the signal sent on the $k$-th subcarrier at the $\nu$-th subframe is denoted as

$$x_{k,\nu} = F_A F_{BB,k} s_{k,\nu} \in \mathbb{C}^{N_t \times 1},$$  \hspace{1cm} (13)
where $s_{k,\nu} \in \mathbb{C}^{N_s \times 1}$ is the baseband signal on the $k$-th subcarrier at the $\nu$-th subframe, $F_A \in \mathbb{C}^{N_t \times N_{RF}}$ is a frequency-flat analog precoder and $F_{BB,k} \in \mathbb{C}^{N_{RF} \times N_s}$ are digital baseband precoders, possibly distinct on each subcarrier. At the receiver, the received signal is combined by the frequency-flat analog combiner $W_A \in \mathbb{C}^{N_r \times N_{RF}}$ and then processed through the cyclic prefix removal and the discrete Fourier transform. The received baseband combiner $W_{BB,k} \in \mathbb{C}^{N_{RF} \times N_s}$ is implemented on the baseband signal for each subcarrier separately, yielding the received signal on the $k$-th subcarrier at the $\nu$-th subframe

$$y_{k,\nu} = W_{BB,k}^H W_A^H H_k F_A F_{BB,k} s_{k,\nu} + W_{BB,k}^H W_A^H v_{k,\nu},$$

where $v_{k,\nu} \in \mathbb{C}^{N_r \times 1}$ is the additive noise vector, with independent and identically distributed (i.i.d.) zero-mean complex Gaussian components with variance $\sigma_n^2$. Note that the hybrid precoder ($F_A/F_{BB,k}$) and combiner ($W_A/W_{BB,k}$) can be chosen differently on distinct subframes.

### D. Extended Virtual Representation of the MIMO Channel

With the uniform antenna arrays, the MIMO channel can be formulated as an extended virtual representation \[51\]. For the UPA at the receiver, we consider the physical AOA of interest as $(\theta_{pr}, \theta_{ar}) \in [-\pi/2, \pi/2] \times [-\pi, \pi)$. Assuming, without loss of generality, half wave-length antenna spacing ($\frac{d}{\lambda_c} = \frac{c}{f_c} = 0.5$), the beam direction region of interest is $(\psi_{hr}, \psi_{vr}) \in [-0.5, 0.5] \times [-0.5, 0.5]$. We assume that the horizontal (vertical) spatial AOAs take values from the uniform grid $G_{hr} \ (G_{vr})$ of size $G_{hr} \geq L \ (G_{vr} \geq L)$, given by

$$G_{\rho R} = \left\{ \psi_{i_\rho} = \frac{i_\rho - G_{\rho r} + 1}{2G_{\rho r}}, \ i_\rho = 1, \ldots, G_{\rho r} \right\}, \text{ for } \rho \in \{h, v\}. \ (15)$$

Then, we define the grid of the receive UPA steering vectors as $G_R = G_{hr} \times G_{vr}$, whose size $|G_R|$ is $G_r = G_{hr} G_{vr}$. We construct the receive array response matrix on the $k$-th subcarrier
by collecting the spatial-frequency UPA steering vectors with the AOAs taking value on the grid $\mathcal{G}_R$ as

$$
A_{R,k}(::i_r) = a_{N_{hr}}(\psi_{i_h}; f_k) \otimes a_{N_{vr}}(\psi_{i_v}; f_k),
$$

(16)

where $i_r = (i_h - 1)G_{vr} + i_v$ with $(\psi_{i_h}, \psi_{i_v}) \in \mathcal{G}_R$. Similarly, for the UPA at the transmitter, we assume the horizontal and vertical spatial AODs take values from the uniform grid $\mathcal{G}_{hT}$ of size $G_{ht} \geq L$ and $\mathcal{G}_{vT}$ of size $G_{vt} \geq L$, respectively. The codebooks $\mathcal{G}_{hT}$ and $\mathcal{G}_{vT}$ are constructed as in (15) by substituting $(G_{ht}, G_{vt})$ for $(G_{hr}, G_{vr})$, and the grid of the transmit UPA steering vectors is defined as $\mathcal{G}_T = \mathcal{G}_{hT} \times \mathcal{G}_{vT}$, whose size $|\mathcal{G}_T|$ is $G_t = G_{ht}G_{vt}$. By collecting the spatial-frequency UPA steering vectors with the AODs taking value on the grid $\mathcal{G}_T$, the transmit array response matrix on the $k$-th subcarrier $A_{T,k} \in \mathbb{C}^{N_t \times G_t}$ is constructed as

$$
A_{T,k}(::i_t) = a_{N_{ht}}(\psi_{i_h}; f_k) \otimes a_{N_{vt}}(\psi_{i_v}; f_k),
$$

(17)

where $i_t = (i_h - 1)G_{vt} + i_v$ with $(\psi_{i_h}, \psi_{i_v}) \in \mathcal{G}_T$.

In this work, we consider a frame-based system, where each frame consists of $T_c$ consecutive subframes (channel uses). We assume that the frame duration is smaller than the channel coherence time, so that the channel remains constant during a frame duration (block-fading assumption). At the $F$-th frame, the MIMO channel on the $k$-th subcarrier can be expressed as an extended virtual representation [51]

$$
H_{k}^{(F)} = A_{R,k}D_{k}^{(F)}A_{T,k}^H,
$$

(18)

where $D_{k}^{(F)} \in \mathbb{C}^{G_r \times G_t}$ is the beamspace channel matrix whose non-zero elements are located in positions corresponding to the spatial AOAs/AODs of the propagation paths. Due to the mismatch between the spatial AOAs/AODs and the corresponding quantized values, a grid-mismatch error may exist but can be diminished if the grid sizes $(G_r, G_t)$ or $(G_{hr}, G_{vr}, G_{ht}, G_{vt})$ are chosen sufficiently large. With the compact antenna deployments and the limited scattering of the channel environment in sub-THz bands, the MIMO channels are spatially correlated and focus on certain spatial directions. The work [25] shows that the channel of a sub-THz band (140 GHz) has an average of 6 clusters and 4 MPCs per cluster, which are fewer compared to the channel of a mmWave band (28 GHz) having an average of 8 clusters and 5 MPCs/cluster. Therefore, the beamspace channel matrix $D_{k}^{(F)}$ tends to be more sparse. We assume that $D_{k}^{(F)}$ has at most
Fig. 4. The evolution of beamspace channel with \((L, L_{cm}) = (4, 3)\) is illustrated. Between two consecutive frames, the green and red elements represent the common and changing elements, respectively.

Fig. 5. The proposed channel training protocol.

\[ L < G_r G_t \text{ non-zero elements, and the remaining elements are negligible.} \]

For the support set information, we define the channel support of the \(F\)-th frame as \(\Omega^{(F)} = \{(i_r, i_t) : D_k^{(F)}(i_r, i_t) \neq 0\}\), which is the set of indices corresponding to the dominant elements in \(D_k^{(F)}\). Note that the channel support is independent of the subcarrier \(k\) because we construct the array response matrix \(A_{R,k}\), \(A_{T,k}\) based on the same frequency-independent uniform grid \(G_R\), \(G_T\) containing the quantized spatial AOAs (AODs). For the massive MIMO system, the channel support is mainly determined by the deployment geometry, hardware, and the time-varying effects of the propagation environment on the electromagnetic wave [28], [29].

Considering the time-varying channel in a frame-based system, the channel remains fixed over the duration of a frame, but may change across subsequent frames. With the temporal correlation, the channel support varies slowly over time, so that \(\Omega^{(F-1)}\) and \(\Omega^{(F)}\) share many common elements [30]. We assume that \(L_{cm}\) is the minimum number of channel elements (common paths) shared between \(\Omega^{(F-1)}\) and \(\Omega^{(F)}\), denoted as \(|\Omega^{(F)} \cap \Omega^{(F-1)}| \geq L_{cm}\), i.e., the AOA-AOD pairs that remain fixed between the previous and current frames. Given a fixed number of channel paths \(L \geq L_{cm}\), there are at most \(L - L_{cm}\) paths changing from one frame to the next one. In Fig. 4, an example of the channel support evolution is illustrated, where the colored and white elements denote the dominant (non-zero) and negligible (zero) beamspace channel elements, respectively. Given a slowly-varying channel with \(L = 4\) paths, the consecutive channel supports \(\Omega^{(F-1)}\) and \(\Omega^{(F)}\) share \(L_{cm} = 3\) common elements (shown in green), so that only one path may change from one frame to the next (shown in red). This structure enables the use of the LS-CS approach [43], [44] to reduce the channel training overhead, discussed in Section III.
III. PROPOSED LS-CS CHANNEL TRAINING

In this section, we propose a support tracking-based channel training on dual-wideband MIMO-OFDM. Section III-A depicts the beam training model, and Section III-B introduces the channel training protocol. Section III-C proposes a new MMV-LS-CS channel estimation method, and Section III-D proposes a channel refinement algorithm to improve the estimation performance. Section III-E proposes a sequential search method using a hierarchical codebook to reduce the computational overhead. Lastly, Section III-F provides the complexity analysis.

A. Beam Training Scheme

With each frame divided into $T_c$ equal-sized subframes, we assume that $T_p$ subframes are used for pilot-based channel training and the remaining $T_c - T_p$ subframes are used only for data transmission. For the pilot transmission, we exploit $K_p < K_o$ subcarriers with a comb-type arrangement, i.e., $\mathcal{P} = \{k = 1 + (i - 1)\Delta_k : i = 1, \ldots, K_p\}$, and the remaining subcarriers are used for data transmission. At the $\nu$-th subframe, the transmitter transmits the precoded signal $x_{k,\nu}$ on the $k$-th subcarrier, and the receiver combines the measurement signal at the $q$-th stream on the $k$-th subcarrier by the combining vector $w_{k,q}$. Assuming that the transmitter employs distinct precoded pilots of $T_p$ subframes and the receiver combines the signal through $Q_p$ streams, the $Q_p \times T_p$ combined signal on the $k$-th subcarrier, $k \in \mathcal{P}$, is denoted as

$$Y_k = W_k^H H_k X_k + \tilde{V}_k = (A_{R,k}^H W_k)^H D_k (A_{T,k}^H X_k) + \tilde{V}_k,$$

where $W_k = [w_{k,1}, \ldots, w_{k,Q_p}]$ and $X_k = [x_{k,1}, \ldots, x_{k,T_p}]$ are the measurement and pilot matrices, respectively. With the extended virtual representation of the MIMO channel in (18), we have the combined signal form in (19). For the channel training, we adopt the random beamforming method [14], [19] to design the pilot and measurement matrices, i.e., $X_k(m,n) = \frac{1}{\sqrt{N_t}} e^{j\zeta_{m,n}}$, $\zeta_{m,n} \sim \mathcal{U}(0, 2\pi)$ and $W_k(m,n) = \frac{1}{\sqrt{N_r}} e^{j\eta_{m,n}}$, $\eta_{m,n} \sim \mathcal{U}(0, 2\pi)$. With sufficiently large $N_r$ and $N_t$, the elements of $A_{R,k}^H W_k$ and $A_{T,k}^H X_k$ are approximately i.i.d. $\mathcal{CN}(0, 1/N_r)$ (respectively, $\mathcal{CN}(0, 1/N_t)$) according to the central limit theorem [19, Appendix B], leading to the measurement matrices satisfying the successful sparse recovery condition [52].

To design such measurement and pilot matrices based on the hybrid transceiver structure, we assume for simplicity that the number of training subframes $T_p$ and the number of streams $Q_p$ satisfy $T_p = N_{ST}N_t^{RF}$ and $Q_p = N_{SR}N_r^{RF}$, for some $N_{ST}, N_{SR} \in \mathbb{N}$ (generalizations are possible). We then design a sequence of measurement matrices $W_k = [W_{k,1}, \ldots, W_{k,N_{SR}}] \in$...
\( \mathbb{C}^{N_r \times Q_p} \) and of pilot matrices \( X_k = [X_{k,1}, \ldots, X_{k,N_{ST}}] \in \mathbb{C}^{N_t \times T_p} \) so as to match the beamforming values generated randomly as explained in the previous paragraph. Each submatrix \( W_{k,i} = W_{A,k,i} W_{BB,k,i} \in \mathbb{C}^{N_r \times N_{RF}}, i = 1, \ldots, N_{SR} \) can be directly obtained by setting the receive baseband combiner as \( W_{BB,k,i} = I_{N_{RF}} \) and the analog combiner as \( W_{A,k,i} = W_{k,i} \). Similarly, each submatrix \( X_{k,i} = F_{A,k,i} F_{BB,k,i} S_{k,i} \in \mathbb{C}^{N_t \times N_{T_{RF}}}, i = 1, \ldots, N_{ST} \) can be directly obtained by setting the baseband precoder and pilot signals as \( F_{BB,k,i} = S_{k,i} = I_{N_{T_{RF}}} \), and the analog precoder as \( F_{A,k,i} = X_{k,i} \).

**B. Channel Training Protocol**

We introduce the protocol of the support tracking-based channel training to reduce the training overhead by exploiting the temporal correlation, shown in Fig. 5. Initially, a conventional MIMO-OFDM channel estimation is applied since no prior channel knowledge is available. Next, a support tracking-based channel estimation (e.g., MMV-LS-CS) is used to estimate the channel aided by the previous channel support estimated in the preceding frame. Note that the support tracking-based approach hinges upon accurate estimation of the support in the previous frame. With an inaccurate previous channel support, the performance of the support tracking-based approach might deteriorate since more channel paths are required to be estimated in the compressed sensing stage. To ensure a satisfactory quality of the support tracking, we check the residual signal of the support-tracking based channel training algorithm. If its magnitude is larger than a predefined threshold, we declare failure of the support tracking, and the initial channel estimation is necessitated. Otherwise, we continue the support tracking-based channel training on the next frame.

**C. MMV-LS-CS Channel Estimation**

Our goal is to develop an approach to reduce the overhead of channel training using the estimated previous channel support \( \hat{\Omega}^{pr} \). Note that we denote the real previous channel support as \( \Omega^{pr} \). To estimate the channel, we first estimate the spatial AOAs/AODs of the dominant channel paths and then refine the time delays and path coefficients of the estimated paths. Here, we introduce the **MMV-LS-CS** algorithm for the dual-wideband MIMO-OFDM channel estimation problem. We separate the MMV-LS-CS channel training into two stages: MMV-LS and MMV-CS. MMV-LS exploits the slowly-varying channel by performing the LS estimate on the estimated previous channel support. Next, MMV-CS applies the CS-based approach on the LS residual, expected to be sparse since most dominant channel elements have been estimated.
in MMV-LS. Given the training signal (19) with the measurement and pilot matrices \((W_k, X_k)\) and the relationship \(\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)\), we have the vectorization of \(Y_k\) as

\[
y_k = \text{vec}(Y_k) = \Theta_k z_k + \tilde{v}_k,
\]

where \(\Theta_k = (A^H_{I,k} X_k)^T \otimes (A^H_{R,k} W_k)^H\) is the equivalent dictionary matrix on the \(k\)-th subcarrier, \(z_k = \text{vec}(D_k)\) and \(\tilde{v}_k = \text{vec}(\tilde{V}_k)\). The MMV-LS-CS algorithm is shown in Algorithm 1 and operates as follows.

First, MMV-LS seeks to estimate the beamspace channel on the paths corresponding to the estimated previous channel support \(\hat{\Omega}^{pr}\). We denote the set of column indices (of \(\Theta_k\)) corresponding to \(\hat{\Omega}^{pr}\) as \(Q = \{(i_t - 1)G_r + i_r : (i_r, i_t) \in \hat{\Omega}^{pr}\}\). Given the frequency-dependent dictionary matrices \(\Theta_k\), the beamspace channel vectors share a common support, enabling the MMV to estimate the dominant paths. The MMV-LS algorithm is initialized as \(\Gamma = \emptyset\) (line 3). Then, the set \(\Gamma\) is reconstructed recursively by collecting the indices from \(Q\) (line 5-6), which leads to the minimum residual error after orthogonalization, until \(|\Gamma| = L_{cm}\) (lines 4-7). The estimated beamspace channel in MMV-LS is then derived (line 8) as

\[
[\hat{z}_{k}^{LS}]_\Gamma = [\Theta_k]_\Gamma^T y_k \quad \text{and} \quad [\hat{z}_{k}^{LS}]_{\Gamma_c} = 0,
\]

where \(\Gamma_c = \{1, \ldots, G_r G_t\}\) \(\backslash\Gamma\).

Secondly, MMV-CS executes the CS-based approach on the LS residual signal obtained by subtracting the channel effect of the MMV-LS estimated beamspace channel \(\hat{z}_{k}^{LS}\) (line 10), given as

\[
y_k^{CS} = y_k - \Theta_k \hat{z}_{k}^{LS} = \Theta_k z_k^{CS} + \tilde{v}_k,
\]

where \(z_k^{CS} = z_k - \hat{z}_{k}^{LS}\) is expected to be sparser than the beamspace channel \(z_k\) since most dominant path components \((L_{cm} \text{ out of } L)\) are expected to be detected in \(\hat{z}_{k}^{LS}\), so that \(z_k^{CS}\) is expected to have only \((L - L_{cm})\) non-zero components. The sparse recovery problem on the pilot subcarriers is formulated as

\[
\arg\min_{z_k^{CS}} \sum_{k \in \mathcal{P}} \|z_k^{CS}\|_1, \quad \text{s.t.} \quad \|y_k^{CS} - \Theta_k z_k^{CS}\|_2 \leq \epsilon, \quad k \in \mathcal{P},
\]

where \(\epsilon > 0\) is a constant threshold. Among many available sparse recovery algorithms exploiting the common channel support across the pilot subcarriers, we adopt the simultaneous OMP.
algorithm (SOMP) \cite{53} (line 12-18). The algorithm is initialized as $\Upsilon = \emptyset$ and the residual $r_k = y_k^{CS}$, $k \in \mathcal{P}$ (line 11). Then, the set $\Upsilon$ is constructed by collecting the index of the column (of $\Theta_k$) with the largest correlation with the residual $r_k$ (line 13-14), until attaining the maximum iteration number or the mean squared error between the current and previous residuals are less than a constant threshold (line 12). In each iteration, the residual $r_k$ is updated by removing the channel effect of the collected indices in $\Upsilon$ (line 15-16). The estimated beamspace channel in MMV-CS is derived (line 19) as

$$[\hat{z}_k^{CS}]_\Upsilon = [\Theta_k]_+^\top y_k^{CS} \text{ and } [\hat{z}_k^{CS}]_{\Upsilon^c} = 0,$$

where $\Upsilon^c = \{1, \ldots, G_r G_t\} \setminus \Upsilon$.

As discussed in \cite{44, 54}, with the $\ell_1$-norm minimization subject to the noise constraint (e.g., MMV-CS in (22)), the estimate $\hat{z}_k^{CS}$ tends to be biased towards zero. Moreover, MMV-CS is terminated when the stop criterion is satisfied, which might lead to false detections of the channel paths, whose beamspace channel value is zero but is detected as non-zero due to noise or some detection error. To reduce the bias and improve the performance in practical settings, we use a similar trick inspired by \cite{44, 54}: we derive the channel support by combining the new detection $\hat{z}_k^{CS}$ in MMV-CS with the estimated $\hat{z}_k^{LS}$ from MMV-LS (line 21) as

$$\Xi = \left\{ i : i \in \{1, \ldots, G_r G_t\}, \frac{1}{|\mathcal{P}|} \sum_{k \in \mathcal{P}} |(\hat{z}_k^A)_i| > \zeta_A \right\},$$

where $\hat{z}_k^A = \hat{z}_k^{LS} + \hat{z}_k^{CS}$, and $\zeta_A > 0$ is a constant threshold excluding the negligible components. Then, the LS estimate on $\Xi$ is computed as $[\hat{z}_k^{\text{det}}]_\Xi = [\Theta_k]_+^\top y_k$, $[\hat{z}_k^{\text{det}}]_{\Xi^c} = 0$, $k \in \mathcal{P}$, where $\Xi^c = \{1, \ldots, G_r G_t\} \setminus \Xi$. Next, to remove likely wrongly detected elements from the estimated channel support, likely false detections are detected and removed (line 23) as

$$\tilde{\Xi} = \Xi \setminus \left\{ i : i \in \Xi, \frac{1}{|\mathcal{P}|} \sum_{k \in \mathcal{P}} |(\hat{z}_k^{\text{det}})_i| \leq \zeta_D \right\},$$

where $\zeta_D > 0$ is a constant deletion threshold. The set $\tilde{\Xi}$ is the estimated current channel support, which can be considered as the estimated previous channel support in the next frame.

With the estimated channel support $\tilde{\Xi}$, we derive the path coefficients and time delays of the estimated paths corresponding to $\tilde{\Xi}$ from the received signal jointly on the pilot subcarriers using the channel refinement algorithm (Algorithm 2), and the estimated channel $\hat{H}_k$ is reconstructed.
Algorithm 1 MMV-LS-CS Channel Estimation.

**Input:** measurement $Y_k$, dictionary matrix $\Theta_k$, estimated previous channel support $\hat{\Omega}^{pr}$

**Output:** estimated channel $\hat{H}_k$, estimated channel support $\hat{\Omega}$.

1. **MMV-LS aided by the estimated previous channel support**
2. $y_k = \text{vec}(Y_k), \ k \in \mathcal{P}$;
3. $\Gamma = \emptyset$; $Q = \{(i_t - 1)G_r + i_r : (i_r, i_t) \in \hat{\Omega}^{pr}\}$;
4. **while** $|\Gamma| < L_{\text{em}}$ **do**
5. $j^* = \arg\min_{j \in Q \setminus \Gamma} \sum_{k \in \mathcal{P}} \|y_k - [\Theta_k][r_{\Gamma \cup j}]^+ y_k\|_F^2$;
6. $\Gamma := \Gamma \cup j^*$;
7. **end while**
8. $[\hat{z}_k^{LS}]_r = [\hat{\Omega}_k]^+ y_k$ and $[\hat{z}_k^{LS}]_{\hat{J}} = 0, \ k \in \mathcal{P}$;

9. **MMV-CS on the LS residual**
10. $y_k^{CS} = y_k - [\Theta_k] \hat{z}_k^{LS}, \ k \in \mathcal{P}$;
11. $\Upsilon = \emptyset$; $r_{k,t} = y_k^{CS}, \ k \in \mathcal{P}$; $J = \{1, \ldots, G_r, G_t\}$; $\varepsilon_t = \infty$;
12. **while** $\varepsilon_t > \varepsilon$ **and** $t < t_{\text{max}}$ **do**
13. $j^* = \arg\max_{j \in J \setminus \Upsilon} \sum_{k \in \mathcal{P}} \|y_k - [\Theta_k] \hat{g}_k(:, j)^T r_{k,t}\|_F^2$;
14. $\Upsilon := \Upsilon \cup j^*$;
15. $\hat{g}_k = \arg\min_{g} \|y_k^{CS} - [\Theta_k]^T g\|_F^2, \ k \in \mathcal{P}$;
16. $r_{k,t+1} = y_k^{CS} - [\Theta_k]^T \hat{g}_k, \ k \in \mathcal{P}$;
17. $\varepsilon_{t+1} = \frac{1}{|P|} \sum_{k \in \mathcal{P}} \|r_{k,t+1} - r_{k,t}\|_F^2$; $t := t + 1$;
18. **end while**
19. $[\hat{z}_k^{CS}]_r = [\Theta_k]^+ y_k^{CS}$ and $[\hat{z}_k^{CS}]_{\hat{J}} = 0, \ k \in \mathcal{P}$;

20. **Combine the detections from MMV-LS and MMV-CS**
21. $\Xi = \{i : i \in \{1, \ldots, G_r, G_t\}, \frac{1}{|P|} \sum_{k \in \mathcal{P}} |(\hat{z}_k^A)_i| > \zeta_A\}$, where $\hat{z}_k^A = \hat{z}_k^{LS} + \hat{z}_k^{CS}$, $k \in \mathcal{P}$;
22. **Delete the false detections**
23. $\hat{\Xi} = \Xi \setminus \{i : \Xi, \frac{1}{|P|} \sum_{k \in \mathcal{P}} |(\hat{z}_k^{det})_i| \leq \zeta_D\}$, where $[\hat{z}_k^{det}]_r = [\Theta_k]^+ y_k, \ [\hat{z}_k^{det}]_{\hat{J}} = 0, \ k \in \mathcal{P}$;
24. **Derive the estimated channel support $\hat{\Omega}$ from $\hat{\Xi}$**;
25. **Reconstruct $\hat{H}_k$ by Algorithm 2** with $(Y_k, \Theta_k, \hat{\Xi})$;

Algorithm 2 Channel Refinement Algorithm.

**Input:** measurement $Y_k$, dictionary matrix $\Theta_k$, support set $\hat{\Xi}$

**Output:** $\hat{H}_k$, $k = 1, \ldots, K_o$

1. Derive $(i_{0,\ell}, i_{\phi,\ell})$ from $\hat{\Xi}$, $\ell = 1, \ldots, L'$;
2. Derive $\hat{q}_k$ by solving (26), $k = 1, \ldots, K_p$;
3. Derive $\hat{z}_\ell$ by (27), $\hat{Q}_\ell$ by (28), $\ell = 1, \ldots, L'$;
4. Reconstruct $\hat{H}_k$ by (29), $k = 1, \ldots, K_o$;

D. Channel Reconstruction And Path Coefficient/Time Delay Refinement

Given the estimated channel support $\hat{\Xi}$, the effective path coefficient vector can be directly estimated by the LS estimator as

$$\arg\min_{q_k} \|y_k - [\Theta_k]_{\hat{\Xi}} q_k\|_F^2, \ k \in \mathcal{P},$$  \hspace{1cm} (26)
where $y_k = \text{vec}(Y_k)$ and $\Theta_k = (A^H_{T,k}X_k) \otimes (A^H_{R,k}W_k)H$ as in (20). Assuming $|\Xi| = L'$, the vector $q_k = [\gamma_{1,k}, \ldots, \gamma_{L',k}]^\top$ is the effective path coefficient vector. The problem (26) is a least squares problem, solved by $\hat{q}_k = [\hat{\gamma}_{1,k}, \ldots, \hat{\gamma}_{L',k}]^\top = ([\Theta_k]_\Xi)^+ y_k$. Note that we have the indices of the estimated AOAs/AODs as $(i_{\theta,\ell}, i_{\phi,\ell})$, $\ell = 1, \ldots, L'$, corresponding to the indices in $\Xi$. Thus, the MIMO-OFDM channel is reconstructed as $\hat{H}_k = \sum_{\ell=1}^{L'} \hat{\gamma}_{\ell,k} A_{R,k}(; i_{\theta,\ell}) A^H_{T,k}(; i_{\phi,\ell})$.

Note however, that such LS estimate does not embed any structure in the path coefficients and time delays across the pilot subcarriers (frequency), as expressed in (6). To leverage such structure and further improve the estimation performance, we proceed as follows. With the generator set $\{z_{\ell} = e^{-j2\pi \frac{f_{\ell,k}}{f_{\ell,k}} \tau_{\ell}}\}$, the effective path coefficient is expressed as $\gamma_{\ell,k} = \sqrt{N_{f}N_{i}} \alpha_{\ell}(f_k) z_{\ell} k^{-\frac{K_{o}+1}{2}}$.

In Section II-B we have proposed and justified the use of $\alpha_{\ell}(f) \approx \frac{\alpha'_{\ell}}{1+f_k/f_c}$, where $\alpha'_{\ell}$ is the reference path coefficient at the carrier frequency. Therefore, the effective path coefficient can be approximated as $\gamma_{\ell,k} \approx \sqrt{N_{f}N_{i}} \frac{\alpha'_{\ell}}{1+f_k/f_c} z_{\ell} k^{-\frac{K_{o}+1}{2}}$. Given the estimated $\hat{q}_k = [\hat{\gamma}_{1,k}, \ldots, \hat{\gamma}_{L',k}]^\top$ by (26) and $\frac{\gamma_{\ell,k}+1}{\gamma_{\ell,k}} \approx \frac{1+f_k/f_c}{1+f_{k+1}/f_c} z_{\ell}$, the estimation of $z_{\ell}$ can be formulated as

$$\arg \min_{z_{\ell}} \sum_{i=1}^{K_{p}-1} \left( z_{\ell} - \left( \frac{\hat{\gamma}_{\ell,1+i\Delta_{k}} \cdot (1+f_{1+i\Delta_{k}}/f_c)}{\hat{\gamma}_{\ell,1+(i-1)\Delta_{k}} \cdot (1+f_{1+(i-1)\Delta_{k}}/f_c)} \right) \right)^2,$$

which is solved by

$$\hat{z}_{\ell} = \frac{1}{K_{p}-1} \sum_{i=1}^{K_{p}-1} \left( \frac{\hat{\gamma}_{\ell,1+i\Delta_{k}} \cdot (1+f_{1+i\Delta_{k}}/f_c)}{\hat{\gamma}_{\ell,1+(i-1)\Delta_{k}} \cdot (1+f_{1+(i-1)\Delta_{k}}/f_c)} \right)^{\frac{1}{2}} \Delta_{k}.$$

The refined time delay is derived by $\hat{\tau}_{\ell} = -\frac{K_{o}}{2\pi B} \angle \hat{z}_{\ell}$, where $\angle \hat{z}_{\ell}$ denotes the phase angle of $\hat{z}_{\ell}$.

Next, we estimate the reference path coefficient $\alpha'_{\ell}$ by formulating the optimization problem as

$$\arg \min_{\alpha'_{\ell}} \|p_{\ell} - c_{\ell} \alpha_{\ell}'\|^2_{F}, \; \ell = 1, \ldots, L,$$

where $p_{\ell}, c_{\ell} \in \mathbb{C}^{K_{p} \times 1}$ with $p_{\ell}(k) = \hat{\gamma}_{\ell,k}$ and $c_{\ell}(k) = \sqrt{N_{f}N_{i}} \frac{1+f_k/f_c}{1+f_{k+1}/f_c} z_{\ell} k^{-\frac{K_{o}+1}{2}}$, $k \in \mathcal{P}$. The problem (28) is a least squares problem, solved by $\hat{\alpha}_{\ell}' = (c_{\ell})^\top p_{\ell}$. Given the refined version of reference path coefficients $\hat{\alpha}_{\ell}'$ and generators $\hat{z}_{\ell}$ (time delays $\hat{\tau}_{\ell}$) accompanied with the indices of the estimated AOAs/AODs, we reconstruct the MIMO channel by

$$\hat{H}_k = \sqrt{N_{f}N_{i}} \sum_{\ell=1}^{L'} \frac{\hat{\alpha}_{\ell}'}{1+f_k/f_c} \hat{z}_{\ell} k^{-\frac{K_{o}+1}{2}} A_{R,k}(; i_{\theta,\ell}) A^H_{T,k}(; i_{\phi,\ell}).$$

The channel refinement algorithm is shown in Algorithm 2.
Fig. 6. An example of the hierarchical codebook with \((G_{\text{sub}}, M) = (3, 2)\).

E. Sequential Search Method With Hierarchical Codebook

Note that, with the extended virtual representation of the MIMO channel in Section II-D, the grid sizes \((G_{hr}, G_{vr}, G_{ht}, G_{vt})\) or \((G_r, G_t)\) should be chosen sufficiently large to minimize the grid-mismatch error between the spatial AOAs/AODs and their quantized values. However, the greedy selection (line 13, Algorithm 1) searches over four different grids \((G_{hT}, G_{vT}, G_{hR}, G_{vR})\) jointly, which can be reformulated into the problem as

\[
\arg \max \left( \sum_{k \in P} \left| H_r k \right|^2 \right),
\]

where the equivalent dictionary vector is defined as

\[
u_k(\psi_A, \psi_B, \psi_C, \psi_D) = \left( a_{N_{ht}}(\psi_A; f_k) \otimes a_{N_{vt}}(\psi_B; f_k) \right)^H X_k \otimes \ldots \left( a_{N_{hr}}(\psi_C; f_k) \otimes a_{N_{vt}}(\psi_D; f_k) \right)^H W_k \right)^H.
\]

The problem is computationally prohibitive to have large grid sizes because the greedy selection of MMV-LS-CS necessitates searching over all combinations of \(\{i_{hr}, i_{vr}, i_{ht}, i_{vt}\} \in [1, G_{hr}] \times [1, G_{vr}] \times [1, G_{ht}] \times [1, G_{vt}]\). To address the issue, a sequential search method using the hierarchical codebook is developed to reduce the computational overhead of the greedy selection.

In each spatial dimension, we construct a hierarchical codebook to do the index selection on multiple levels. Considering an example of a uniform grid on \([-0.5, 0.5)\) of size \(G\) along a spatial dimension (same as in (15)) as \(\mathcal{G} = \{ \psi_i = \frac{i - G_{\text{sub}} + 1}{G}, \ i = 1, \ldots, G \},\) we divide the index selection of \(\mathcal{G}\) into the hierarchical search on \(M\) levels, where each level has its own sub-codebook of size \(G_{\text{sub}}\), satisfying \(G = G_{\text{sub}}^M\). We design the sub-codebook in the \(m\)-th level as

\[
\mathcal{G}^{(m)}(\hat{\psi}^{(m-1)}) = \left\{ \psi_{(m,i)} = \hat{\psi}^{(m-1)} + \frac{i - G_{\text{sub}} + 1}{2 G_{\text{sub}}^m}, \ i = 1, \ldots, G_{\text{sub}} \right\},
\]

where \(\hat{\psi}^{(m-1)}\) is the selected codeword from the previous level. The initial value of the selected codeword is assumed as \(\hat{\psi}^{(0)} = 0\). In the first level \((m = 1)\), the index search is implemented with
the resolution $G_{\text{sub}}^{-1}$. Then, in the subsequent levels ($m > 1$), the index search is implemented with a finer resolution $G_{\text{sub}}^{-m}$ centered around the previously selected point $\hat{\psi}^{(m-1)}$. Hence, the total number of codewords to be searched can be greatly reduced from $G = G_{\text{sub}}^M$ to $MG_{\text{sub}}$.

In Fig. 6, an example of the hierarchical codebook with $(G_{\text{sub}}, M) = (3, 2)$ is illustrated. A codebook of size $G = 9$ is separated into $M = 2$ levels, where each level has its sub-codebook with $G_{\text{sub}} = 3$. Due to the narrow beam property of massive MIMO, the sub-codebook size $G_{\text{sub}}$ should be selected to guarantee the beam coverage of $G^{(1)}$ on all possible spatial angles.

To choose $G_{\text{sub}}$, we analyze the beamwidth of the effective dictionary vector in (19). For ease of exposition, we consider the effective dictionary vector along one spatial dimension $W_k^H a_N(\psi; f_k)$, where $W_k \in \mathbb{C}^{N \times Q}$ is the measurement matrix and $a_N(\psi; f_k) \in \mathbb{C}^{N \times 1}$ is the array response vector (as in (8)). The beamforming pattern of the effective dictionary vector $W_k^H a_N(\psi_i; f_k)$, $\psi_i \in G^{(1)}$ is defined as

$$p(\psi, \psi_i; f_k) = (W_k^H a_N(\psi; f_k))^H W_k^H a_N(\psi_i; f_k).$$

Note that we adopt the random beamforming method for the channel training, i.e., $W_k(n, q) = \frac{1}{\sqrt{N}} e^{j\eta_{n,q}}$, $\eta_{n,q} \sim \mathcal{U}(0, 2\pi)$. Thus, the expected beamforming pattern of the effective dictionary vector is $\mathbb{E}[p(\psi, \psi_i; f_k)] = \frac{Q}{N} a_N(\psi; f_k)^H a_N(\psi_i; f_k)$, demonstrating that the expected beamwidth of the effective dictionary vector is the same as the beamwidth of $a_N(\psi_i; f_k)$, represented by the Half Power Beamwidth (HPBW) $= 0.5 \times$ First Null Beamwidth (FNBW) \cite{47}. We can derive $\text{FNBW} = \frac{1}{(1 + f_k / f_c) \frac{2}{N}}$ by analyzing the beamforming pattern $p(\psi, \psi_i; f_k)$, so the beamwidth of $a_N(\psi_i; f_k)$ is $\text{HPBW} = \frac{1}{(1 + f_k / f_c) \frac{1}{N}} \approx \frac{1}{N}$. To guarantee the beam coverage of the sub-codebook
on all possible spatial angles, the sub-codebook resolution $G_{sub}^{-1}$ is required to be at least finer than the beamwidth $N^{-1}$. Therefore, we choose the sub-codebook size as $G_{sub} \geq N$. In Fig. 7a, we illustrate the beamforming gain of the effective dictionary vectors $\mathbf{W}_k^H \mathbf{a}_N(\psi_i; f_k)$ with $\psi_i \in G^{(1)}$ (as in (31)), given $N = G_{sub} = 16$, $Q = 30$, $B = 8$ GHz, $f_c = 142$ GHz, and $k = 50$ among $K_o = 128$ subcarriers. The expected beamforming gain is shown in Fig. 7b.

Since solving the problem (30) may contribute to a high computing overhead due to the extremely large dictionary matrices ($G_{ht}G_{vt}G_{hr}G_{vr}$ index sets to be searched), we propose a sequential search method to efficiently solve (30), detailed as below:

\begin{align}
(\hat{\psi}_A^{(1)}, \hat{\psi}_B^{(1)}, \hat{\psi}_C^{(1)}, \hat{\psi}_D^{(1)}) = \arg \max_{(\psi_A, \psi_B, \psi_C, \psi_D)} \sum_{k \in P} |\mathbf{u}_k(\psi_A, \psi_B, \psi_C, \psi_D) H \mathbf{r}_k|^2; \quad (33)
\end{align}

\begin{align}
\hat{\psi}_A^{(m)} = \arg \max_{\psi \in \mathcal{G}_{HT}(\hat{\psi}_A^{(m-1)})} \sum_{k \in P} |\mathbf{u}_k(\psi, \hat{\psi}_B^{(1)}, \hat{\psi}_C^{(1)}, \hat{\psi}_D^{(1)}) H \mathbf{r}_k|^2, \quad m = 2, \ldots, M; \quad (34)
\end{align}

\begin{align}
\hat{\psi}_B^{(m)} = \arg \max_{\psi \in \mathcal{G}_{VT}(\hat{\psi}_B^{(m-1)})} \sum_{k \in P} |\mathbf{u}_k(\hat{\psi}_A^{(1)}, \psi, \hat{\psi}_C^{(1)}, \hat{\psi}_D^{(1)}) H \mathbf{r}_k|^2, \quad m = 2, \ldots, M; \quad (35)
\end{align}

\begin{align}
\hat{\psi}_C^{(m)} = \arg \max_{\psi \in \mathcal{G}_{HR}(\hat{\psi}_C^{(m-1)})} \sum_{k \in P} |\mathbf{u}_k(\hat{\psi}_A^{(1)}, \hat{\psi}_B^{(1)}, \psi, \hat{\psi}_D^{(1)}) H \mathbf{r}_k|^2, \quad m = 2, \ldots, M; \quad (36)
\end{align}

\begin{align}
\hat{\psi}_D^{(m)} = \arg \max_{\psi \in \mathcal{G}_{VR}(\hat{\psi}_D^{(m-1)})} \sum_{k \in P} |\mathbf{u}_k(\hat{\psi}_A^{(1)}, \hat{\psi}_B^{(1)}, \hat{\psi}_C^{(1)}, \psi) H \mathbf{r}_k|^2, \quad m = 2, \ldots, M; \quad (37)
\end{align}

First, the problem (33) is solved by searching over the index set in $\mathcal{G}_{HT}^{(1)} \times \mathcal{G}_{VT}^{(1)} \times \mathcal{G}_{HR}^{(1)} \times \mathcal{G}_{VR}^{(1)}$ to maximize the objective function, where $\mathcal{G}_{HT}^{(1)}, \mathcal{G}_{VT}^{(1)}, \mathcal{G}_{HR}^{(1)}, \mathcal{G}_{VR}^{(1)}$ are the first-level hierarchical codebooks of $\mathcal{G}_{HT}, \mathcal{G}_{VT}, \mathcal{G}_{HR}, \mathcal{G}_{VR}$, respectively. We assume the hierarchical codebooks are constructed with the same grid size $G_{sub}$ and number of levels $M$. Note that for the distinct spatial dimensions, $G_{sub}$ and $M$ could be selected as different values subject to the desired resolutions. With the rough estimate $(\hat{\psi}_A^{(1)}, \hat{\psi}_B^{(1)}, \hat{\psi}_C^{(1)}, \hat{\psi}_D^{(1)})$ by (33), we sequentially solve the one-dimensional search problems using the hierarchical search (as in (31)), starting with the search on the horizontal spatial AODs (34), the vertical spatial AODs (35), the horizontal spatial AOAs (36), and then the vertical spatial AOAs (37). With the sequential search method, we reduce the number of the index sets to be searched from $G_{ht}G_{vt}G_{hr}G_{vr} = G_{sub}^{4M}$ to $G_{sub}^4 + 4(M - 1)G_{sub}$.

F. Computational Complexity

For MMV-LS, the complexity is dominated by the subset selection of $\Gamma$ and its pseudo-inverse operation (line 5, Algorithm 1), searching over the combinations of $\{i_t, i_l\} \in \hat{\Omega}^{pr}$, with complexity $O\left(|\hat{\Omega}^{pr}|K_p L_{cm}^3\right)$ [55]. For MMV-CS, the complexity is dictated by the greedy selection (line 13, Algorithm 1), searching over all combinations of $\{g_r, g_t\} \in \mathcal{J}$, with
TABLE I

COMMON SIMULATION PARAMETERS [3], [4]

| Parameter                  | Symbol | Value                      |
|----------------------------|--------|----------------------------|
| Carrier frequency          | $f_c$  | 142 GHz                    |
| Bandwidth                  | $B$    | 8 GHz                      |
| Total subcarriers          | $K_o$  | 1600                       |
| Channel paths              | $(L, L_{con})$ | (4, 3)                     |
| Ref. path coefficient      | $\alpha'_L$ | $CN(0, \sigma^2)$        |
| Polar AOA/AOD              | $\theta_{pr}/\theta_{pt}$ | $U(-\pi/2, \pi/2)$ |
| Azimuth AOA/AOD            | $\theta_{ar}/\theta_{at}$ | $U(-\pi, \pi)$           |
| Tx antenna (UPA)           | $N_t$  | (4, 4)                     |
| Rx antenna (UPA)           | $N_r$  | 256 (16, 16)               |
| Tx/Rx sub-codebook         | $(G_{sub,t}, G_{sub,r})$ | (4, 16)                   |
| Pilot subcarriers          | $K_p$  | 10                         |
| Training parameters        | $(Q_p, T_p)$ | (30, 30)                  |
| Subframe/Frame duration    | $\delta_s/T_{frame}$ | 10µs/10ms                  |

complexity $O\left(G_rG_tK_pQ_pT_p\hat{L}\right)$, where $\hat{L}$ is the number of estimated paths in MMV-CS. Note that the sequential search method greatly alleviates the overhead of the greedy selection, leading to the complexity $O\left((G_{sub}^4 + 4(M - 1)G_{sub})K_pQ_pT_p\hat{L}\right)$, where $G_{sub}$ is the sub-codebook size and $M$ is the level number. For the channel refinement algorithm, the complexity is dominated by the least squares operations in (26). Assuming $L'$ estimated paths in Algorithm [2] the update of $q_k$ requires the pseudo-inverse operations on pilot subcarriers with complexity $O\left(K_p(L')^3\right)$.

IV. NUMERICAL RESULTS

We evaluate the performance of MMV-LS-CS channel estimation in dual-wideband MIMO-OFDM. The numerical parameters are listed in Table I. We consider the multipath channel with delays $\tau_L \sim U(0, 100)$ns, where the delay spread is $D_s = 100$ns [4]. This yields a coherence bandwidth $B_c = \frac{1}{2D_s} = 5$ MHz [48], leading to $K_o = B/B_c = 1600$ subcarriers. The receive and transmit array response matrices ($A_{R,k}$ and $A_{T,k}$) are constructed as in (16) and (17), with the uniform grid $G_R = G_{hR} \times G_{vR}$ of grid size $G_r = G_{hr}G_{vr}$ and the uniform grid $G_T = G_{hT} \times G_{vT}$ of grid size $G_t = G_{ht}G_{vt}$, respectively. To reduce the computational complexity, we apply a sequential search method with the hierarchical codebook (as in Section [III-E]) for the greedy selection. We consider a hierarchical codebook with $(G_{sub,t}, M)$ for the uniform grids $G_{hT}, G_{vT}$ on the transmitter, i.e., $G_{ht} = G_{vt} = G_{sub,t}^M$. Similarly, a hierarchical codebook with $(G_{sub,r}, M)$ is considered for the uniform grids $G_{hR}, G_{vR}$ on the receiver, i.e., $G_{hr} = G_{vr} = G_{sub,r}^M$. The receiving signal-to-noise ratio (SNR) is defined as

$$SNR = \frac{\sum_k \|Y_k - \tilde{V}_k\|^2_F}{\sum_k \|\tilde{V}_k\|^2_F},$$

where $Y_k$ is the received signal matrix, and $\tilde{V}_k$ is the combined noise matrix, as in (19).
We compare the proposed MMV-LS-CS with the estimation schemes listed as follows:

**MMV-LS-CS**: MMV-LS-CS is applied with the "estimated" previous channel support.

**MMV-LS-CS w/ prev. support**: MMV-LS-CS is applied with the previous channel support.

**Random LS**: The LS estimator is applied with the random orthonormal training matrices.

**Genie-aided LS**: The LS estimator is applied with the current channel support.

**Genie-aided LS w/ refinement**: The channel refinement algorithm is applied with the current channel support.

**GSOMP [15]**: The GSOMP estimator recovers the channel (with frequency-dependent dictionary matrices) using the common support across the pilot subcarriers.

Note that the number of training measurements with \((Q_p, T_p) = (30, 30)\) (partial training of 21.9% \(N_r N_t\) beams) is considered. Note that the work [15] evaluates the GSOMP estimator under the partial training of 80% \(N_r N_t\) beams, which is much larger than the setting in our experiments.

First, to evaluate the channel estimation accuracy, we define the normalized mean squared error (NMSE) of the channel as

\[
\text{NMSE} = \frac{\sum_k \| \mathbf{H}_k - \hat{\mathbf{H}}_k \|_F^2}{\sum_k \| \mathbf{H}_k \|_F^2}.
\]  

In Fig. 8, we evaluate the NMSE of the channel versus the SNR. Genie-aided LS w/ refinement attains \(\text{NMSE} = 4 \times 10^{-3}\) above \(\text{SNR} = -5\) dB, which provides a lower bound for NMSE, while Genie-aided LS attains \(\text{NMSE} = 10^{-2}\) in the same configuration. It shows that given the correct AOAs/AODs of the channel paths, the channel refinement algorithm improves the estimation accuracy by deriving the frequency-independent reference path coefficient and time delays. We evaluate the **MMV-LS-CS w/ prev. support** scheme with different level numbers (of the hierarchical codebooks), \(M = 1, \ldots, 4\). For \(M = 1\), the MMV-LS-CS w/ prev. support attains \(\text{NMSE} = 0.64\) with \(\text{SNR} = 10\) dB. In the same configuration, the NMSE performance improves with a larger \(M\), attaining 0.08 for \(M = 2\), 0.051 for \(M = 3\), and 0.049 for \(M = 4\). We observe that there is only minor improvement as we increase the level number from \(M = 3\) to \(M = 4\), so we choose \(M = 3\) for the following experiments. Given an \(\text{SNR} = 10\) dB, MMV-LS-CS attains \(\text{NMSE} = 0.1\), as opposed to 0.05 for MMV-LS-CS w/ prev. support, 0.36 for GSOMP, and 0.88 for Random LS. MMV-LS-CS is inferior to MMV-LS-CS w/ prev. support due to the potential errors in the estimated previous channel support. With an inaccurate previous estimated channel support, more channel elements are expected to be detected in MMV-CS (i.e., fewer correct
components are estimated in MMV-LS), which deteriorates the estimation performance. The previous estimated channel support depends on the previous channel estimation, which tends to be more inaccurate at the low SNR region. Nonetheless, given an SNR $= -5$ dB (i.e., a low SNR region), our proposed MMV-LS-CS attains $NMSE = 0.22$, which outperforms the state-of-the-art approach (GSOMP) with $NMSE = 0.5$ in the same configuration.

Fig. 8. The NMSE versus the SNR.

Next, we evaluate the spectral efficiency (SE) for different receive SNRs. The achievable rate used is given by

\[
R = \sum_{k=1}^{K_o} \frac{B}{K_o} \log_2 \det \left( I_{N_s} + \frac{P_t}{N_s} R_{\hat{v}_k}^{-1} \hat{W}_k^H \hat{H}_k \hat{F}_k^H \hat{H}_k^H \hat{W}_k \right), \tag{40}
\]

where $P_t$ is the average transmit power for each transmission. The number of data streams is assumed as $N_s = 4$. The combiner $\hat{W}_k$ (or the precoder $\hat{F}_k$) is derived by the directions of the eigenvectors of $\hat{H}_k \hat{H}_k^H$ (or $\hat{H}_k^H \hat{H}_k$). The post-processing noise covariance matrix is $R_{\hat{v}_k} = E[\hat{v}_k \hat{v}_k^H]$, where $\hat{v}_k = \hat{W}_k^H v$ with the additive complex Gaussian noise $v$. The fraction of time for the pilot transmission is $\iota = \frac{T_{\text{train}}}{T_{\text{frame}}}$, where the duration of $T_{\text{train}} = T_p \delta_s$ on the $K_p$ pilot subcarriers is the resulting training overhead. The data transmission is allowed on all bandwidth at the transmission time, and also on the bandwidth other that the pilot subcarriers at the training time. We denote the achievable rate on the training time as $R_{\text{train}}$, which is constructed as in (40) without the pilot subcarriers. Thus, the SE is defined as $\frac{\iota R_{\text{train}} + (1-\iota)R}{BN_s}$ (bit/s/Hz/stream), which includes the loss due to the training overhead. In Fig. 9, we evaluate the SE versus the SNR. Full CSI attains the largest SE because its $\hat{W}_k$ and $\hat{F}_k$ are derived using the perfect CSI. Genie-aided LS w/ refinement has the performance only inferior to that of full CSI since the correct AOAs/AODs of the channel paths are given. The MMV-LS-CS w/ prev. support scheme attains a better SE with a larger level number $M$. And there is only minor difference

Fig. 9. Spectral efficiency versus the SNR.
of the SE between the schemes with $M = 3$ and $M = 4$, consistent with the NMSE evaluation. For SNR = 10dB, the SE of MMV-LS-CS is 4.45bit/s/Hz/stream, which outperforms the state-of-the-art (GSOMP) by 0.92bit/s/Hz/stream, and Random LS by 2.43bit/s/Hz/stream. As opposed to MMV-LS-CS w/ prev. support with $M = 3$, the SE loss of MMV-LS-CS is around 0.1bit/s/Hz/stream, originating from the potential errors in the estimated previous channel support. Compared with Full CSI, the SE loss of MMV-LS-CS is only 0.32bit/s/Hz/stream; in contrast, a larger SE gap (1.24bit/s/Hz/stream) exists between GSOMP and Full CSI because our proposed MMV-LS-CS provides a more accurate channel estimation as shown in Fig. 8.

In Fig. 10 we evaluate the NMSE versus the number of training subcarriers $K_p$, with SNR = −15dB. In the case with $K_p = 1$, we do the LS estimator to reconstruct the channel for the schemes w/ refinement since there are no multiple subcarriers to implement the channel refinement algorithm. With $K_p = 1$, we have NMSE = 0.94 for MMV-LS-CS, as opposed to NMSE = 0.09 for Genie-aide LS, NMSE = 0.6 for MMV-LS-CS w/ prev. support, NMSE = 1.16 for GSOMP, and NMSE = 1.92 for Random LS. As we increase to $K_p = 5$, we observe that the NMSE of MMV-LS-CS, MMV-LS-CS w/ prev. support and GSOMP significantly improve to NMSE = 0.49, NMSE = 0.24 and NMSE = 0.93, respectively. The improvement originates from the MMV approach using the common channel support shared across the pilot subcarriers. Besides, the channel refinement algorithm provides a more accurate estimation when the number of pilot subcarriers $K_p$ increases, especially in the low SNR region. We also observe that Genie-aided LS w/ refinement attains NMSE = 0.04 when $K_p = 5$, as opposed to Genie-aided LS has only NMSE = 0.08 in the same configuration. The improvement arises from fitting the frequency-independent reference path coefficients and time delays across subcarriers.
For $K_p > 5$, all approaches have fairly minor improvement with additional pilot subcarriers.

In Fig. 11, given an SNR $= 10$dB, we evaluate the NMSE versus the measurement ratio $M_p = N_p/(N_rN_t)$, where $N_p$ is the number of training measurements. For a channel training with $N_p$ measurements, we assume the training parameters as $Q_p = T_p = \sqrt{N_p}$. The NMSE of channel decreases if there are more channel training measurements. MMV-LS-CS attains NMSE $= 0.48$ with a measurement ratio $M_p = 3.5\%$ (or $Q_p = T_p = 12$), as opposed to the state-of-the-art (GSOMP), which achieves NMSE $= 0.98$ in the same configuration. As we increase the training measurements to $M_p = 21.9\%$ (or $Q_p = T_p = 30$), the NMSE performance improves (NMSE $= 0.1$ for MMV-LS-CS and NMSE $= 0.36$ for GSOMP), which shows the advantage of our proposed MMV-LS-CS channel training algorithm.

In Fig. 12, we evaluate the NMSE of the channel versus the bandwidth $B$, given an SNR $= 10$dB. The dual-wideband effect becomes more severe with larger bandwidth. As the bandwidth increases from 0.7 to 8 GHz, the NMSE of MMV-LS-CS increases by only 0.008, as opposed to 0.09 for GSOMP. Both the proposed MMV-LS-CS and the state-of-the-art (GSOMP) are not too sensitive to the change of bandwidths owing to the frequency-dependent dictionary matrices. We observe that MMV-LS-CS is slightly more robust to bandwidth changes since it estimates a part of the channel paths on the previous channel support, and only few channel components are estimated in the CS stage. In contrast, the GSOMP estimates all channel paths in the CS stage, which exacerbates the detriments of the dual-wideband effect.

V. CONCLUSION

We proposed a CS training framework for time-varying channels suffering from a dual-wideband effect in sub-THz MIMO-OFDM. We constructed the frequency-dependent array
response matrices based on the same grid to preserve the common channel support among OFDM subcarriers and recovered the sparse beamspace channel from multiple observations sharing common support. For slowly-varying channels, we developed a new MMV-LS-CS channel estimation, applying MMV-LS to do the LS estimate on the estimated previous channel support, and MMV-CS on the residual signal to detect the time-varying channel components. We proposed a channel refinement algorithm to reconstruct the channel by estimating the time delays and path coefficient jointly on the pilot subcarriers. To reduce the computational complexity, we proposed a sequential search method using the hierarchical codebook for the greedy selection. Numerical results showed that MMV-LS-CS provides a more efficient and accurate channel estimation and improved SE than the state-of-the-art [15] in time-varying dual-wideband MIMO-OFDM.

REFERENCES

[1] T.-H. Chou, N. Michelusi, D. J. Love, and J. V. Krogmeier, “Wideband Millimeter-Wave Massive MIMO Channel Training via Compressed Sensing,” to appear at IEEE Globecom 2021.
[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, “What will 5G be?” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
[3] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhattee, and G. C. Trichopoulos, “Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond,” IEEE Access, vol. 7, pp. 78 729–78 757, Apr. 2019.
[4] Y. Xing, T. S. Rappaport, and A. Ghosh, “Millimeter Wave and sub-THz Indoor Radio Propagation Channel Measurements, Models, and Comparisons in an Office Environment,” IEEE Commun. Lett., pp. 1–1, Oct. 2021.
[5] S. Ju, Y. Xing, O. Kanhere, and T. S. Rappaport, “Millimeter Wave and Sub-Terahertz Spatial Statistical Channel Model for an Indoor Office Building,” IEEE J. Sel. Areas Commun., vol. 39, no. 6, pp. 1561–1575, Jun. 2021.
[6] C. Lin and G. Y. Li, “Adaptive Beamforming With Resource Allocation for Distance-Aware Multi-User Indoor Terahertz Communications,” IEEE Trans. Commun., vol. 63, no. 8, pp. 2985–2995, Aug. 2015.
[7] J. M. Jornet and I. F. Akyildiz, “Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band,” IEEE Trans. Wireless Commun., vol. 10, no. 10, pp. 3211–3221, Oct. 2011.
[8] R. Piesiewicz, C. Jansen, D. Mittleman, T. Kleine-Ostmann, M. Koch, and T. Kurner, “Scattering Analysis for the Modeling of THz Communication Systems,” IEEE Trans. Antennas Propag., vol. 55, no. 11, pp. 3002–3009, Nov. 2007.
[9] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” IEEE Commun. Mag., vol. 52, no. 2, pp. 186–195, Feb. 2014.
[10] E. Björnson, L. Van der Perre, S. Buzzi, and E. G. Larsson, “Massive MIMO in Sub-6 GHz and mmWave: Physical, Practical, and Use-Case Differences,” IEEE Wireless Commun., vol. 26, no. 2, pp. 100–108, Apr. 2019.
[11] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, “Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems,” IEEE Trans. Signal Process., vol. 66, no. 13, pp. 3393–3406, Jul. 2018.
[12] B. Wang, M. Jian, F. Gao, G. Y. Li, and H. Lin, “Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems,” IEEE Trans. Signal Process., vol. 67, no. 23, pp. 5893–5908, Dec. 2019.
[13] M. Wang, F. Gao, N. Shlezinger, M. F. Flanagan, and Y. C. Eldar, “A block sparsity based estimator for mmWave massive MIMO channels with beam squint,” IEEE Trans. Signal Process., vol. 68, pp. 49–64, Nov. 2019.
[14] 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., vol. 68, no. 7, pp. 4218–4232, Jul. 2020.

[15] K. Dovelos, M. Matthaiou, H. Q. Ngo, and B. Bellalta, “Channel Estimation and Hybrid Combining for Wideband Terahertz Massive MIMO Systems,” IEEE J. Sel. Areas Commun., vol. 39, no. 6, pp. 1604–1620, Jun. 2021.

[16] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, “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.

[17] Z. Marzi, D. Ramasamy, and U. Madhow, “Compressive Channel Estimation and Tracking for Large Arrays in mm-Wave Picocells,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 514–527, Apr. 2016.

[18] J. P. González-Coma, J. Rodriguez-Fernandez, N. González-Prelcic, L. Castedo, and R. W. Heath, “Channel Estimation and Hybrid Precoding for Frequency Selective Multiuser mmWave MIMO Systems,” IEEE J. Sel. Topics Signal Process., vol. 12, no. 2, pp. 353–367, May 2018.

[19] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, “Low-Rank Tensor Decomposition-Aided Channel Estimation for Millimeter Wave MIMO-OFDM Systems,” IEEE J. Sel. Areas Commun., vol. 35, no. 7, pp. 1524–1538, Jul. 2017.

[20] D. C. Araújo, A. L. De Almeida, J. P. Da Costa, and R. T. de Sousa, “Tensor-Based Channel Estimation for Massive MIMO-OFDM Systems,” IEEE Access, vol. 7, pp. 42 133–42 147, Mar. 2019.

[21] Z. Gao, L. Dai, Z. Wang, and S. Chen, “Spatially Common Sparsity Based Adaptive Channel Estimation and Feedback for FDD Massive MIMO,” IEEE Transactions on Signal Processing, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[22] J. Choi and R. W. Heath, “Interpolation based transmit beamforming for MIMO-OFDM with limited feedback,” IEEE Trans. Signal Process., vol. 53, no. 11, pp. 4125–4135, Nov. 2005.

[23] T. Pande, D. J. Love, and J. V. Krogmeier, “A Weighted Least Squares Approach to Precoding With Pilots for MIMO-OFDM,” IEEE Trans. Signal Process., vol. 54, no. 10, pp. 4067–4073, Oct. 2006.

[24] ——, “Reduced Feedback MIMO-OFDM Precoding and Antenna Selection,” IEEE Trans. Signal Process., vol. 55, no. 5, pp. 2284–2293, May 2007.

[25] S. L. H. Nguyen, J. Järveläinen, A. Karttunen, K. Haneda, and J. Putkonen, “Comparing radio propagation channels between 28 and 140 GHz bands in a shopping mall,” in Proc. Eur. Conf. Antennas Propag. (EuCAP), Apr. 2018, pp. 1–5.

[26] T. Nitsche, C. Cordeiro, A. B. Flores, E. W. Knightly, E. Perahia, and J. C. Widmer, “IEEE 802.11 ad: directional 60 GHz communication for multi-Gigabit-per-second Wi-Fi,” IEEE Commun. Mag., vol. 52, no. 12, pp. 132–141, Dec. 2014.

[27] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, “A Tutorial on Beam Management for 3GPP NR at mmWave Frequencies,” IEEE Commun. Surveys Tuts., vol. 21, no. 1, pp. 173–196, Sep. 2018.

[28] S. Wu, C.-X. Wang, H. Haas, e.-H. M. Aggoune, M. M. Alwakeel, and B. Ai, “A non-stationary wideband channel model for massive MIMO communication systems,” IEEE Trans. Wireless Commun., vol. 14, no. 3, pp. 1434–1446, Mar. 2015.

[29] Y. Chen and C. Han, “Time-Varying Channel Modeling for Low-Terahertz Urban Vehicle-to-Infrastructure Communications,” in Proc. IEEE Glob. Commun. Conf., 2019, pp. 1–6.

[30] 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.

[31] 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.

[32] M. Hussain, M. Scalabrin, M. Rossi, and N. Michelusi, “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.

[33] M. Hussain and N. Michelusi, “Learning and Adaptation for Millimeter-Wave Beam Tracking and Training: a Dual Timescale Variational Framework,” IEEE J. Sel. Areas Commun., pp. 1–1, Nov. 2021.
[34] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, “Millimeter Wave Beamforming for Wireless Backhaul and Access in Small Cell Networks,” IEEE Trans. Commun., vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[35] M. B. Booth, V. Suresh, N. Michelusi, and D. J. Love, “Multi-Armed Bandit Beam Alignment and Tracking for Mobile Millimeter Wave Communications,” IEEE Commun. Lett., vol. 23, no. 7, pp. 1244–1248, Jul. 2019.

[36] D. J. Love and R. W. Heath, “Equal gain transmission in multiple-input multiple-output wireless systems,” IEEE Trans. Commun., vol. 51, no. 7, pp. 1102–1110, Jul. 2003.

[37] D. J. Love, R. W. Heath, and T. Strohmer, “Grassmannian beamforming for multiple-input multiple-output wireless systems,” IEEE Trans. Inf. Theory, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[38] D. J. Love, R. W. Heath, V. K. Lau, D. Gesbert, B. D. Rao, and M. Andrews, “An overview of limited feedback in wireless communication systems,” IEEE J. Sel. Areas Commun., vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[39] N. González-Prelcic, R. Méndez-Rial, and R. W. Heath, “Radar aided beam alignment in mmWave V2I communications supporting antenna diversity,” in Proc. Inf. Theory Appl. Workshop, Feb. 2016, pp. 1–7.

[40] A. Klautau, N. González-Prelcic, and R. W. Heath, “LIDAR Data for Deep Learning-Based mmWave Beam-Selection,” IEEE Wireless Commun. Lett., vol. 8, no. 3, pp. 909–912, Jun. 2019.

[41] M. Hashemi, C. E. Koksal, and N. B. Shroff, “Out-of-Band Millimeter Wave Beamforming and Communications to Achieve Low Latency and High Energy Efficiency in 5G Systems,” IEEE Trans. Commun., vol. 66, no. 2, pp. 875–888, Feb. 2018.

[42] T.-H. Chou, N. Michelusi, D. J. Love, and J. V. Krogmeier, “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.

[43] N. Vaswani and J. Zhan, “Recursive Recovery of Sparse Signal Sequences From Compressive Measurements: A Review,” IEEE Trans. Signal Process., vol. 64, no. 13, pp. 3523–3549, Jul. 2016.

[44] N. Vaswani, “LS-CS-Residual (LS-CS): Compressive Sensing on Least Squares Residual,” IEEE Trans. Signal Process., vol. 58, no. 8, pp. 4108–4120, Aug. 2010.

[45] A. Alkhateeb and R. W. Heath, “Frequency Selective Hybrid Precoding for Limited Feedback Millimeter Wave Systems,” IEEE Trans. Commun., vol. 64, no. 5, pp. 1801–1818, May 2016.

[46] A. Molisch, “Ultrawideband propagation channels—theory, measurement, and modeling,” IEEE Trans. Veh. Technol., vol. 54, no. 5, pp. 1528–1545, Sep. 2005.

[47] C. A. Balanis, Antenna theory: analysis and design. John wiley & sons, 2016.

[48] D. Tse and P. Viswanath, Fundamentals of wireless communication. Cambridge university press, 2005.

[49] C. Han, A. O. Bicen, and I. F. Akyildiz, “Multi-Ray Channel Modeling and Wideband Characterization for Wireless Communications in the Terahertz Band,” IEEE Trans. Wireless Commun., vol. 14, no. 5, pp. 2002–2142, May 2014.

[50] J. Tan and L. Dai, “Delay-Phase Precoding for THz Massive MIMO with Beam Split,” in Proc. IEEE Glob. Commun. Conf., 2019, pp. 1–6.

[51] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, “An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 436–453, Apr. 2016.

[52] E. J. Candes, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” Commun. Pure Appl. Math., vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[53] 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.

[54] E. Candes and T. Tao, “The Dantzig selector: Statistical estimation when p is much larger than n,” Annals of statistics, vol. 35, no. 6, pp. 2313–2351, Dec. 2007.

[55] G. H. Golub and C. F. Van Loan, Matrix computations, 4th ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2012.