ABSTRACT In this paper, we investigate two novel beamspace channel estimation schemes for millimeter-wave MIMO-OFDM systems relying on a uniform linear antenna array. Based on the sparse nature of the beamspace channel, compressive sensing-based methods can be adopted. Most of the existing wideband beamspace channel estimation schemes take the common support assumption that the support of all subcarriers from a single path component keeps unchanged. However, due to the beam squint effect caused by time delays among different antennas, this assumption is no longer satisfied. Based on the above, we analyze the frequency-dependent sparse structure of the beamspace channel due to beam squint and propose the block simultaneous orthogonal matching pursuit algorithm where support is jointly estimated by using beamspace windows. Moreover, considering the power focusing effect of beam squint, we propose the two-stage beamspace channel estimation scheme, where the first stage is for the angle of arrival estimation, and the second stage is for updating path gain to reconstruct the beamspace channel. Simulation results demonstrate that proposed schemes can estimate the channel under the beam squint effect with high accuracy but low pilot overheads.

INDEX TERMS MIMO, millimeter-wave, compressive sensing, beamspace channel estimation, beam squint.

I. INTRODUCTION

With the rapid popularization of mobile Internet applications and the widespread use of smart terminals, the demand for mobile data rates has increased exponentially. The massive multiple-input multiple-output (MIMO) technique is designed to improve spectrum efficiency in wireless communication due to its degree of freedom in the spatial domain [1]–[3]. Meanwhile, the millimeter-wave (mmWave) at 30-300 GHz can provide rich spectrum resources to support wideband transmission and achieve high data rates [4]–[7]. By combining mmWave wideband communication and the massive MIMO technique, both spectrum efficiency and data rates can be improved, which has attracted humongous attention in academia and industry communities [6].

For wireless communication systems, acquiring the instantaneous channel state information (CSI) has a decisive impact on system performance. However, channel estimation becomes a challenging problem when massive MIMO and mmWave technologies are introduced. On the one hand, the dimension of the channel matrix becomes significant with a large number of antennas at the base station (BS), which leads to colossal pilot overheads and high complexity if classic algorithms are adopted, e.g., least squares (LS) and minimum mean square error (MMSE) [8]. On the other hand, the severe path loss of the mmWave channel and the mobility of users bring difficulties to channel estimation.

To reduce the complexity of channel estimation, some authors propose various channel estimation schemes on the aspect of exploiting the sparsity in the angle domain and delay domain of mmWave channel [9]–[15], which are developed by compressive sensing (CS)-based methods. Particularly, [9] provides an orthogonal matching pursuit (OMP) solution for hybrid massive MIMO systems by using the angle-domain channel sparsity. The authors in [10] propose a joint chan-
nel estimation and tracking scheme based on compressive sensing. In [11], a two-stage method is proposed for time-varying mmWave channels by exploiting the sparsity in the angle domain.

However, the methods in [9]–[11] are designed for narrowband systems. Although these methods can be extended to wideband systems via sub-carrier by sub-carrier processing, the complexity will increase due to the large number of subcarriers, especially in mmWave wideband systems. Recently, efficient wideband channel estimation schemes have been proposed for massive MIMO systems [12]–[15]. Specifically, an OMP based scheme is proposed in [12] by leveraging the sparsity of the mmWave channels in both time and frequency domains. [13] utilizes the basis expansion model to reduce the number of the coefficients to be estimated and then proposes a quasi-block simultaneous orthogonal matching pursuit algorithm to recover the channel. [16] investigates wideband beamspace channel estimation with a lens antenna array. An adaptive selecting network is designed at the first stage, and then a support detection-based channel estimation scheme is proposed by exploiting the structural characteristics of the mmWave beamspace channel. Besides, the authors in [14] propose a matrix-completion-based estimation approach based on a random spatial sampling structure. By exploiting the channel sparsity in both angle and delay domains, [15] adopts the maximum likelihood approach to obtain the super-resolution estimations of the AoAs.

However, the above-mentioned sparse estimation schemes have two vital drawbacks. On the one hand, they approximate the continuous parameter space to a finite set of grid points, introducing quantization error and degrading channel estimation performance. To deal with the grid mismatch problem, researchers propose off-grid schemes in [17], [18]. Specifically, [17] combines the smart perturbation mechanism and the simultaneous orthogonal matching pursuit (SOMP) algorithm to jointly solve off-grid parameters. In [18], the beamspace multiple signal classification (MUSIC) algorithm is implemented to estimate the angle of arrival (AoA) after antenna grouping. Although these methods have better performance than traditional ones, their complexity is generally high.

On the other hand, previous schemes [9]–[15] neglect the difference between the traditional MIMO channel model and the massive MIMO channel model. For large-scale antenna mmWave systems, time delay among antennas of the same physical path cannot be ignored due to a large number of antennas and a high sampling rate. The conventional massive MIMO channel model ignores delay differences among antennas due to a small number of antennas. Consequently, the channel estimation and precoding algorithms need to be revised. This phenomenon causes beam squint effect [19], [20] in the frequency domain, which is more obvious in wideband mmWave massive MIMO systems. The wideband solutions [12]–[15] suffer from obvious performance loss due to the common sparse support assumption is no longer satisfied. To deal with this problem, some solutions are proposed [21]–[26]. Specifically, [21] and [22] adopt the off-grid sparse Bayesian learning scheme to get continuous angle-delay parameters, and [23] divides the entire bandwidth into multiple subbands, then obtains the joint estimation of the path directions by MUSIC algorithm. But the complexity of these two schemes is relatively high. Besides, authors in [24] propose a CS-based scheme that exploits the shift-invariant block-sparsity of the beam squint channel model. In [25], with the help of the adaptive-updating dictionary, physical parameters of the channel can be extracted by a super-resolution CS algorithm. [26] proposes a simultaneous weighted OMP algorithm to deal with beam squint. However, none of these methods use the special sparse structure caused by the beam squint effect to reduce channel estimation complexity.

In this paper, we propose two schemes of beamspace channel estimation with beam squint effect for wideband mmWave massive MIMO systems: a CS-based algorithm and a two-stage framework. The main contributions of this paper are summarized as follows:

1) Based on the effect of beam squint, we firstly analyze a unique sparse structure of each path component of the beamspace channel, which is frequency-dependent. Besides, we define the beam distance as the support difference among different sub-carriers of each component, which can be mathematically calculated.

2) With the frequency-dependent sparse structure, we propose a block simultaneous orthogonal matching pursuit (BSOMP) scheme. For each path component, we jointly estimate sparse support to improve accuracy using the frequency-dependent sparse structure and then remove its influence to estimate the remaining path components.

3) We also propose a two-stage scheme, which decomposes the beamspace channel estimation into two subproblems, i.e., AoA estimation and path gain estimation. Due to the beam squint effect, we discover the property of power focusing, and then propose the off-grid AoA estimation algorithm by finding the index of antenna and sub-carrier with no power leakage. In the second stage, the channel gains are obtained by LS algorithm for reconstructing the beamspace channel.

The following of this paper is organized as follows. Section II introduces the channel model for the mmWave MIMO-OFDM system on a uniform linear array and introduces the effect of beam squint. Besides, the problem of beamspace channel estimation is formulated for single-antenna users. Section III proposes the BSOMP scheme to solve the problem of beamspace channel estimation with beam squint. In section IV, the two-stage scheme is proposed with the performance analysis. In Section V, simulation results are provided to demonstrate the advantages of our proposed scheme. Finally, the conclusion is given in Section VI.

Notations: Scalar is denoted in italics, while the lower-case and upper-case boldface denote a vector and a matrix, respec-
between user $k$ and $n$-th antenna can be expressed as
\[
\begin{align*}
    h_{k,n}^{T}(t) &= \sum_{l=1}^{L_k} \tilde{g}_{k,l} e^{-j2\pi f_c \tau_{k,l,n}} \delta(t - \tau_{k,l,n}) \\
    &= \sum_{l=1}^{L_k} \tilde{g}_{k,l} e^{-j2\pi(n-1)\varphi_{k,l}} e^{-j2\pi f \tau_{k,l,n}}.
\end{align*}
\]  

where $\tilde{g}_{k,l} \triangleq g_{k,l} e^{-j2\pi f_{\tau_{k,l}}} \delta(t - \tau_{k,l,n})$ is the equivalent complex gain of $l$-th path for user $k$ and $L_k$ denotes the number of paths from user $k$ to the BS. The frequency-domain response can be given as
\[
\begin{align*}
    h_{k,n}^{F}(f) &= \sum_{l=1}^{L_k} \tilde{g}_{k,l} e^{-j2\pi(n-1)\varphi_{k,l}} e^{-j2\pi f \tau_{k,l,n}} \\
    &= \sum_{l=1}^{L_k} \tilde{g}_{k,l} e^{-j2\pi(n-1)\varphi_{k,l}} (1 + \frac{f}{f_c}) e^{-j2\pi f \tau_{k,l,n}}.
\end{align*}
\]

The channel response in frequency-domain of $N$ antennas is represented by $\mathbf{h}_{k,m}$, i.e.,
\[
\mathbf{h}_{k,m} = \sqrt{\frac{N}{L_k}} \sum_{l=1}^{L_k} \tilde{g}_{k,l} e^{-j2\pi \tau_{k,l} f m} \mathbf{\Xi}_{m}(\varphi_{k,l}),
\]

where $\mathbf{\Xi}_{m}(\varphi_{k,l}) = [1, e^{-j2\pi(1+\frac{m}{N})}\varphi_{k,l}, \ldots, e^{-j2\pi(N-1)(1+\frac{m}{N})}\varphi_{k,l}] \in \mathbb{C}^{N \times 1}$ represents the spatial-domain steering vector which is relevant to the frequency of sub-carriers, and $f_m$ denotes the frequency of $m$-th sub-carrier.

The widely-used MIMO channel model [28], [29] ignores the time delay among antennas, and the spatial-domain steering vector can be expressed as
\[
\mathbf{a}(\varphi_{k,l}) = [1, e^{-j2\pi \varphi_{k,l}}, \ldots, e^{-j2\pi(N-1)\varphi_{k,l}}],
\]

which is frequency-independent. But with the number of antennas and transmission bandwidth going large, the relationship $\max(\tau_{k,l,n} - \tau_{k,l}) \ll T_s$ does not hold any more where $T_s$ refers to sampling interval. For example, in the mmWave massive MIMO systems, the BS equipped with 128-antenna ULA of half-wavelength spacing and operate at 28GHz with 4GHz bandwidth can induce the maximum delay of $9.07T_s$ that is not negligible.

According to [30], we can get beamspace channel expression based on the discrete Fourier transform (DFT) matrix $\mathbf{U}$, which consists of $N$ orthogonal array response vectors
\[
\mathbf{U} = [\mathbf{a}(\varphi_1), \ldots, \mathbf{a}(\varphi_N)],
\]

where $\varphi_n = \frac{1}{N}(n - N + 1)$ for $n = 1, 2, \ldots, N$ refers to the pre-defined direction. Consider the effect of frequency-selective fading, the beamspace channel of the $m$-th sub-carrier can be expressed as
\[
\mathbf{h}_{k,m} = \mathbf{U}^{H} \mathbf{h}_{k,m} = \sqrt{\frac{N}{L_k}} \sum_{l=1}^{L_k} \tilde{g}_{k,l} e^{-j2\pi \tau_{k,l} f m} \tilde{c}_{k,l,m}.
\]
where \( \tilde{c}_{k,l,m} = U^H \Xi_m (\varphi_{k,l}) \) represents the \( l \)-th beamspace channel at sub-carrier \( m \) for user \( k \).

The channel model in (7) can be rewritten as a matrix form

\[
\tilde{h}_{k,m} = \tilde{C}_{k,m} \tilde{\beta}_{k,m},
\]

where \( \tilde{C}_{k,m} = [\tilde{c}_{k,1,m}, \tilde{c}_{k,2,m}, \ldots, \tilde{c}_{k,L,m}] \in \mathbb{C}^{N \times L} \), \( \tilde{\beta}_{k,m} = [\tilde{\beta}_{k,1,m}, \tilde{\beta}_{k,2,m}, \ldots, \tilde{\beta}_{k,L,m}]^T \in \mathbb{C}^{L \times 1} \) is the vector of path gains with \( \tilde{\beta}_{k,l,m} = \sqrt{N_{\text{RF}}} \bar{g}_{k,l} e^{-j2\pi \tau_{k,l} f_m} \).

### B. PROBLEM FORMULATION

In TDD systems, each user transmits their pilot in an orthogonal resource block. Therefore, the channel estimation of each user is independent of each other, and we omit subscript \( k \) in the following sections for notational simplicity. It should be noted that we focus on uplink channel estimation, and the downlink channel can be obtained by leveraging the channel reciprocity.

We denote \( x_{m,p} \) as the transmitted pilot at the \( m \)-th sub-carrier in the time slot \( p \). Then, after receiver combining, cyclic prefix removal, and \( M \)-point DFT, the received signal at BS can be obtained as

\[
y_{m,p} = W_p \hat{h}_{m} x_{m,p} + W_p n_{m,p}, m = 1, 2, \ldots, M, \tag{9}
\]

where \( W_p \in \mathbb{C}^{NRF \times N} \) is the receiver hybrid combining matrix, and \( n_{m,p} \sim \mathcal{CN}(0, \sigma^2 I_N) \) of size \( N \times 1 \) is the noise vector with \( \sigma^2 \) representing the noise power.

Let \( P \) denotes the length of pilots and assume \( x_{m,p} = 1 \) for \( p = 1, \ldots, P \). The overall received pilot signals at \( m \)-th sub-carrier \( \bar{y} = [y_{m,1}^T, y_{m,2}^T, \ldots, y_{m,P}^T]^T \in \mathbb{C}^{NRF \times 1} \) can be obtained as

\[
\bar{y}_m = \bar{W} \hat{h}_m + \bar{n}_m \tag{10}
\]

where \( \bar{W} = [W_1^T, W_2^T, \ldots, W_P^T]^T \in \mathbb{C}^{NRF \times PN} \) is the overall combining matrix, and \( \bar{n}_m = [n_{m,1}^T, n_{m,2}^T W_2^T, \ldots, n_{m,P}^T W_P^T]^T \in \mathbb{C}^{NRF \times 1} \) is the effective noise. Stacking received signals of \( M \) sub-carriers, we have

\[
\bar{Y} = [\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_M] = \bar{W} \mathbf{H} + \mathbf{N}, \tag{11}
\]

where \( \mathbf{H} = [\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_M] \in \mathbb{C}^{N \times M} \), and \( \mathbf{N} = [\bar{n}_1, \bar{n}_2, \ldots, \bar{n}_M] \in \mathbb{C}^{NRF \times M} \).

Now, our target is the reconstruction of \( \hat{h}_m \) based on \( \bar{y}_m \) and \( \bar{W} \) for \( m = 1, 2, \ldots, M \).

### III. THE BSOMP SCHEME FOR BEAMSPACE CHANNEL ESTIMATION WITH BEAM SIGHT

In this section, we first analyze the beamspace channel’s unique sparse structure with beam sight effect. Then, the BSOMP algorithm is proposed for beamspace channel estimation with beam sight.

#### A. SPARSE STRUCTURE OF BEAMSPACE CHANNEL WITH BEAM SIGHT

In this sub-section, we analyze the sparsity of \( \hat{h}_m \) and then prove that it exhibits a unique sparse structure, which is frequency-dependent.

We rewrite \( \tilde{c}_{l,m} \) as

\[
\tilde{c}_{l,m} = U^H \Xi_m (\varphi_l) = [\text{sinc}((1 + f_m f_c) \varphi_l - \varphi_1), \ldots, \text{sinc}((1 + f_m f_c) \varphi_l - \varphi_N)]^T,
\]

\[
m = 1, \ldots, M, \tag{12}
\]

where \( \text{sinc}(x) = \sin \pi x / \pi x \) represents the Dirichlet sinc function \([31]\). Most of the power of \( \tilde{c}_{l,m} \) is focused on a small number of elements according to the power-focusing property of \( \text{sinc}(x) \). Consequently, \( \hat{h}_m \) is a sparse vector due to the limited scattering in mmWave wireless systems, i.e., \( L \) is small \([5], [32]\).

According to (12), the focused power is related to the frequency, which means the support of \( \hat{h}_m \) is different among sub-carriers, i.e.,

\[
\text{supp}\left(\hat{h}_{m_1}\right) \neq \text{supp}\left(\hat{h}_{m_2}\right) \text{ for } m_1 \neq m_2, \tag{13}
\]

where \( \text{supp}(x) \) represents index set of non-zero elements. In the narrowband or small scale of antenna array systems, the support of \( \hat{h}_m \) is frequency-independent and usually takes the common support assumption that is no longer satisfied based on (13).

For example, considering a system with AoA \( \theta_{k,l} = \pi/8 \), the number of UL antennas \( N = 128 \), carrier frequency \( f_c = 28 \) GHz and bandwidth \( B = 4 \) GHz, the normalized beamspace channel power distributions with antenna index over different sub-carriers is shown in Fig.2. It can be observed that the index of the antenna with the most power changes with sub-carriers. Specifically, the first sub-carrier has a strong beam at 116-th antenna, and the middle sub-carrier at 119-th antenna, while the last sub-carrier at 123-th antenna, which is different. Fig.3 depicts the relationship between the index of the antenna with the strongest power and the sub-carriers. It shows that the index of the strongest antennas varies with sub-carriers, which means the support of the beamspace channel is related to frequency.

Both figures show the impact of beam sight, which needs to be considered when we design schemes for beamspace channel estimation.

Based on the above analysis, we know that the assumption of common support is no longer satisfied among sub-carriers because of the beam sight effect. However, the beamspace channel still shows a unique frequency-dependent sparse structure which is the basis of the proposed BSOMP scheme. 

**Lemma 1:** With the \( l \)-th path component of the beamspace channel, the beam distance from \( m_1 \)-th sub-carrier to \( m_2 \)-th sub-carrier can be calculated as \( d_{l,m_1,m_2} = (m_1 - m_2) \eta N f_c / f_c \) where \( \eta \) represents the sub-carrier spacing.

**Proof 1:** Firstly, we define the beam distance \( d_{l,m_1,m_2} \) as the index difference of the antenna that has the most power of \( l \)-th beamspace channel path between \( m_1 \)-th sub-carrier and \( m_2 \)-th sub-carrier, which can be expressed as

\[
d_{l,m_1,m_2} = n_{l,m_1}^* - n_{l,m_2}^*, \tag{14}
\]
where \( n_{l,m_1}^* \) and \( n_{l,m_2}^* \) represents the index of antenna with largest power for \( m_1 \)-th, \( m_2 \)-th sub-carrier respectively.

According to the sparsity of \( \tilde{\mathbf{h}}_m \), we can find the \( n_{l,m}^* \) by

\[
    n_{l,m}^* = \arg \min_n \left| \left( 1 + \frac{f_m}{f_c} \right) \varphi_l - \bar{\varphi}_n \right|. \tag{15}
\]

From (15), we can learn that the index \( n_{l,m}^* \) is relevant to the frequency of sub-carrier. It moves with \( f_m \), which indicates the common support assumption is no longer satisfied.

If the largest power antenna moves \( d_{l,m_1,m_2} \) with the sub-carriers, the value of pre-defined direction changes as:

\[
    \bar{\varphi}_{n_{l,m_1}} - \bar{\varphi}_{n_{l,m_2}}^* = \frac{1}{N} \left( n_{l,m_1}^* - \frac{N + 1}{2} \right) - \frac{1}{N} \left( n_{l,m_2}^* - \frac{N + 1}{2} \right) = \frac{d_{l,m_1,m_2}}{N}. \tag{16}
\]

So we have

\[
    \left( \left( 1 + \frac{f_{m_1}}{f_c} \right) \varphi_l - \left( 1 + \frac{f_{m_2}}{f_c} \right) \varphi_l \right) = \frac{d_{l,m_1,m_2}}{N}. \tag{17}
\]

According to (17), we have the conclusion \( d_{l,m_1,m_2} = (m_1 - m_2) \eta N \varphi_l / f_c \).

Consider the effect of beam squint and Lemma 1, the \( d_{l,m_1,m_2} \) doesn’t have to be an integer, and the fractional part \( d_{l,m_1,m_2} - \lfloor d_{l,m_1,m_2} \rfloor \) represents not all the beamspace channel power focus on the antenna of the index \( n_{l,m}^* \) where \( \lfloor x \rfloor \) represents that the largest integer smaller than \( x \). Due to the limited resolution of pre-defined directions, some power will distribute on adjacent antennas, which is called the power leakage effect from another perspective. On the other hand, we can calculate the support of all sub-carriers if one specific sub-carrier’s support is known.

**B. PROPOSED BSOMP SCHEME**

Since \( \tilde{\mathbf{h}}_m \) is a sparse vector, we intuitively adopt CS-based algorithms to solve the problem with a significantly reduced length of pilot signals, i.e., \( P \ll (N/N_{RF}) \) \cite{33, 34}. For the wideband systems, the CS-based schemes are proposed, such as SOMP \cite{35} and OMP-based \cite{12} algorithm. However, due to the beam squint effect, the common support among sub-carriers assumption of SOMP does not hold anymore, as shown in Fig.2 and Fig.3.

In this sub-section, we propose the BSOMP scheme to solve the beamspace channel estimation problem. Due to the frequency-dependent sparse structure of the beamspace channel, the BSOMP scheme jointly estimates the support of different sub-carriers for each path component to improve the accuracy. Although both SOMP and BSOMP estimate support jointly, the BSOMP scheme calculates each sub-carrier’s support after the jointly estimation result, and the SOMP scheme ignores the support difference among sub-carriers, which leads to performance degradation. After support estimation, the LS algorithm is adopted to get beamspace channel response.

According to (12), the index of the strongest element \( n_{l,m}^* \) can be found as

\[
    n_{l,m}^* = \arg \max_n \left| (1 + \frac{f_m}{f_c}) \varphi_l - \bar{\varphi}_n \right|. \tag{18}
\]

Based on Lemma 1, the index of the strongest element of the remaining sub-carriers can be calculated. For example, if we get \( n_{l,1}^* \), the others can be obtained as

\[
    n_{l,m}^* = n_{l,1}^* + \left( \frac{(m - 1) \eta \varphi_l N}{f_c} \right), m = 2, 3, \ldots, M. \tag{19}
\]

where we assume \( \varphi_l \approx \bar{\varphi}_{n_{l,m}^*} \).

Then the support of \( \tilde{\mathbf{h}}_m \) for \( l \)-th path component is

\[
    \Omega_{l,m} = \Theta_N \left\{ n_{l,m}^* - \omega, \ldots, n_{l,m}^* + \omega \right\}, \tag{20}
\]

where \( \Theta_N (x) = \text{mod}_N (x - 1) + 1 \) represents the mod function whose target is to guarantee all elements in \( \Omega_{l,m} \) belong to \( \{1, 2, \ldots, N\} \). It should be noted that the size of

![Figure 2](image-url) Normalized beamspace channel power distributions with antennas.

![Figure 3](image-url) The index of most channel powered antennas with sub-carriers.
the support is determined by the parameter $\omega$, and 90% beamspace channel power contained for $\omega = 4$ and $N = 256$ scenario under common support assumption [16].

Removing the influence of former $l$ paths, i.e., $\tilde{h}_m(\Omega) = 0$, the support of $(l + 1)$-th path component can be obtained by repeating (18) and (20). Finally, the support of $\tilde{h}_m$ is $\Omega = \Omega_1 \cup \ldots \cup \Omega_L$.

Consider $\tilde{H} = [\tilde{h}_1, \ldots, \tilde{h}_M]$, we design a beamspace window (BWIn) to capture the beamspace channel power of $l$-th component, which is the region enclosed by several antennas and all $M$ sub-carriers that expressed as

$$\Gamma_n = \Theta_N \{n, \ldots, n + d_n - 1\}. \quad (21)$$

The purpose of BWIn is to determine the antenna region that most beamspace channel power-focused of $l$-th path component. Due to the beam squint effect, we use the average power focused on each antenna to measure the power of BWIn captured, i.e.,

$$n_l^* = \arg \max_n \frac{\|\tilde{H}(\Gamma_n,:)\|^2_F}{|\Gamma_n|}. \quad (22)$$

The key of BWIn is to design $d_n$. Based on Lemma 1, the maximal beam distance happens when $m_1 = M$ and $m_2 = 1$, i.e.,

$$d_{l,max} = d_{l,M,1} = (M - 1) \eta N \varphi_l / f_c, \quad (23)$$

It indicates the optimal $d_n$ is

$$d_n = \left[ \frac{(M - 1) \eta \varphi_l N}{f_c} \right]. \quad (24)$$

The relationship between target in (22) and $d_n$ is shown in Fig.4 when $N = 128$, $M = 512$, $f_c = 28$ GHz and $B = 4$ GHz. It can be observed that target in (22) shows a good performance when $d_n$ satisfies (24). Besides, (24) and Fig.4 both show the optimal $d_n$ varying with different values of normalized AoA, which is caused by reason that different values of normalized AoA lead to different levels of power leakage because of beam squint.

After getting $n_l^*$, the sparse channel support of $l$-th component of other sub-carriers is obtained by (19). Then, the LS algorithm is adopted to get beamspace channel response. The pseudo-code of the BSOMP scheme is summarized in Algorithm 1.

Firstly, we initialize $R = [r_1, r_2, \ldots, r_M] = Y$, where $r_m$ is the residual at sub-carrier $m$.

For $l$-th path component, we firstly adopt the BWIns to capture beamspace channel power. Specifically, we calculate the the correlation matrix as $A_l = \tilde{W}^H R$ in step 1. Note that the combine matrix $\tilde{W}$ need to satisfied the condition of restricted isometry property (R.I.P), i.e., $\tilde{W}^H \tilde{W} \approx \mathbf{I}_N$. Then, we generate $N$ BWIns $\Gamma_n = \Theta_N \{n, \ldots, n + d_n - 1\}$ based on Lemma 1 in step 4, and obtain the index of strongest element of the $l$-th component at the first sub-carrier in step 6, i.e.,

$$n_l^* = \arg \max_n \frac{\|A_l(\Gamma_n,:)\|^2_F}{|\Gamma_n|}. \quad (25)$$

Then, we get the frequency-dependent sparse channel support $\Omega_{l,m}$ of the $l$-th path component for $m = 1, 2, \ldots, M$ in step 9. After the support estimation, the nonzero elements of $l$-th component at sub-carrier $m$ can be obtained by the LS algorithm, shown in step 10. Then, the procedure of residual update is proceeded in step 11, i.e.,

$$r_m = r_m - \tilde{W}(\cdot, \Omega_{l,m}) \hat{s}_{l,m}(\Omega_{l,m}). \quad (26)$$

The procedure above will be repeated until all $L$ path components have been estimated. Finally, we combine the support of all $L$ path components and independently estimate the beamspace channel, which is executed in steps 15 and 16.

C. ALGORITHM SUMMARY AND PERFORMANCE ANALYSIS

In this subsection, we summarize the proposed BSOMP scheme and analyze its Cramer-Rao lower bound (CRLB) and the computational complexity.

The main difference between the BSOMP scheme and conventional CS-based schemes is the support estimation. For instance, the OMP-based scheme estimates the support of different sub-carriers independently [12], which is vulnerable to noise and shows a poor performance at low signal-noise ratio (SNR) [36]. And the SOMP-based scheme assumes all sub-carriers share common support [35], which leads to serious performance loss due to the beam squint effect. By contrast, the BSOMP scheme jointly detects the support with beam squint by exploiting the frequency-dependent sparse structure of the wideband beamspace channel, which achieves a better performance. In addition, we assume the number of paths $L$ is known as prior information that channel measurements can obtain [5]. For the cases where $L$ is unknown, we can modify the loop condition of the algorithm [34]. Specifically, we define the channel estimation difference as $\frac{1}{MN} \|\tilde{H}^t - \hat{H}^{-1}\|^2_F$, where superscript $t$ represents
the number of loops. If the difference is smaller than a threshold \( \epsilon \), the procedure will be stopped.

Then, we discuss the the CRLB of the BSOMP scheme. We know that the noise in (10) after combining process-
ing is still subject to Gaussian distribution, i.e., \( \tilde{\mathbf{n}}_m \sim \mathcal{CN} \left( \mathbf{0}_{PN_{RF} \times 1}, \sigma^2 \mathbf{I}_{PN_{RF}} \right) \) [11]. Therefore, we give the conditional probability density function of \( \tilde{y}_m \) with given \( \hat{h}_m \), which is

\[
P_{\tilde{y}_m | \hat{h}_m} (\tilde{y}_m; \hat{h}_m) = \frac{1}{(2\pi \sigma^2)^{\frac{PN_{RF}}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \left\| \tilde{y}_m - \mathbf{W} \hat{h}_m \right\|^2 \right\} .
\]

(27)

According to [37], the Fisher information matrix of (27) can be calculated as

\[
\begin{align*}
[J]_{i,j} & \triangleq -\mathbb{E} \left\{ \frac{\partial^2 \ln p_{\tilde{y}_m | \hat{h}_m} (\tilde{y}_m; \hat{h}_m)}{\partial \hat{h}_m(i) \partial \hat{h}_m(j)} \right\} \\
& = \frac{1}{\sigma^2} \left[ (\mathbf{W}^H \mathbf{W}) \right]_{i,j}.
\end{align*}
\]

(28)

Based on the estimation theory [37], we have

\[
\mathbb{E} \left\{ \left\| \hat{h}_m - \tilde{h}_m \right\|^2 \right\} \geq \text{Tr} \left\{ \mathbf{J}^{-1} \right\} \\
= \sigma^2 \text{Tr} \left\{ \left( \mathbf{W}^H \mathbf{W} \right)^{-1} \right\} \\
\geq \sigma^2 \text{Tr} \left\{ \mathbf{I}_N \right\} \\
= N\sigma^2.
\]

(29)

As for complexity analysis, we observe that the complexity of BSOMP is dominated by steps 2, 6, 10, 11 and 16 according to Algorithm 1. In step 2, the complexity of correlation matrix calculation is \( O(\text{NMN}_{RF} M) \). In step 6, the optimal \( n^*_m \) is found, which has the complexity in order of \( O(NM) \). In step 10, the size of \( \mathbf{W}(:, \Omega_{l,m}) = P\mathbf{N}_{RF} \times (2\omega + 1) \) and the size of \( r_m \) is \( P\mathbf{N}_{RF} \times 1 \). Hence, the multiplication gives a complexity in order of \( O(NP_{RF}P\omega^2) \). Same as step 10, the step 11 calculate the multiplication between \( \mathbf{W}(:, \Omega_{l,m}) \) and \( \tilde{s}_{l,m} \), which gives a complexity in order of \( O(P\mathbf{N}_{RF}P\omega^2) \). In step 16, we compute the multiplication between \( \mathbf{W}^\dagger(:, \Omega_m) \) and \( \tilde{y}_m \), where the size of \( \mathbf{W}^\dagger(:, \Omega_m) \) is \( P\mathbf{N}_{RF} \times |\Omega_m| \). So the complexity is \( O(P\mathbf{N}_{RF}P|\Omega_m|) \), where \( |\Omega_m| \leq L(2\omega + 1) \).

Considering that there are \( L \) path components, steps 2, 6, 10 and 11 are executed for \( L \) times, while step 16 is executed once. So the total complexity of the BSOMP scheme is given as

\[
O(NML) + O(PN_{RF}P\omega^2) + O(NP_{RF}P\omega^2).
\]

(30)

By contrast, the complexity of both the OMP-based and SOMP-based schemes can be presented as \( O(PN_{RF}P\omega^3) + O(NPN_{RF}P\omega) \) [11], [35], which is higher than that of the BSOMP scheme.

IV. THE PROPOSED TWO-STAGE BEAMSPACE CHANNEL ESTIMATION WITH BEAM SQUINT

For the BSOMP scheme, the size of support of all sub-carriers are same, i.e., \( \omega = \omega_1 = \cdots = \omega_M \). According to Dirichlet sinc form of elements in (12), the level of power focusing is different among sub-carriers. Hence, \( \omega \) is not suitable for all sub-carriers, which causes performance degradation. Considering this issue, in this section, we will introduce another property of beam squint, i.e., power focusing effect. Combining with the matrix form of beamspace channel expression, we propose a novel two-stage framework for beamspace channel estimation, including AoA estimation followed by path gain estimation.

A. THE POWER FOCUSING EFFECT OF BEAM SQUINT

In this subsection, we demonstrate the effect of power focusing, which can help us estimate the normalized AoA of \( l \)-th path component.
The elements of \( \mathbf{\hat{c}}_{l,m} \) present a form of Dirichlet sinc function so that the largest power element for \( l \)-th path component at \( m \)-th sub-carrier are located at

\[
n_{l,m}^\ast = \arg \min_n \left( 1 + \frac{f_m}{f_c} \right) \varphi_l - \varphi_n, \quad m = 1, \ldots, M.
\]

Due to the limited number of antennas, the pre-defined direction \( \varphi_n \) are discrete values. In most cases, the value of \( \left( 1 + \frac{f_m}{f_c} \right) \varphi_l - \varphi_n \) is not equal to zero, which causes the power leakage phenomenon. In traditional MIMO channel models, the level of power leakage of all sub-carriers is same. However, we can observe that the level of power leakage varies with frequency when the beam squint effect is taken into consideration.

Based on (31), we define the beam squint matrix \( \mathbf{Q}_l \), whose column is \([n_{l,m}^\ast, m]^T\) for \( m = 1, 2, \ldots, M \). The value of \( \left( 1 + \frac{f_m}{f_c} \right) \varphi_l - \varphi_n \) is different for different \( \mathbf{Q}_l(:, m) \). This variation rule is consistent with the Dirichlet sinc function. As shown in Fig. 5, \( \left( 1 + \frac{f_m}{f_c} \right) \varphi_l \) is located at the red region when \( n_{l,m}^\ast \) is found. The larger \( f_m \) is, the value of \( \left( 1 + \frac{f_m}{f_c} \right) \varphi_l \) goes more closer to \( \varphi_n \), if \( \left( 1 + \frac{f_m}{f_c} \right) \varphi_l < \varphi_n \).

The step size depends on sub-carrier spacing \( \eta \) and the resolution of pre-defined direction that relies on the number of antennas \( N \). When \( N \) and \( M \) are large, we will get the point where the entire channel power of the \( l \)-th path component is focused on one element with no power leakage, i.e., existing \( f_{\hat{m}} \) satisfies

\[
\left( 1 + \frac{f_{\hat{m}}}{f_c} \right) \varphi_l = \varphi_{n_{l,m}^\ast}.
\]

For the sub-carriers whose index satisfies \( m > \hat{m} \), the channel power of the \( l \)-th path component will gradually focus on the \((n_{l,m}^\ast + 1)\)-th element. If \( M \) is large enough, we will get the next point with the whole channel power of the \( l \)-th path component, i.e.,

\[
\left( 1 + \frac{f_{\hat{m} + 1}}{f_c} \right) \varphi_l = \varphi_{n_{l,m}^\ast + 1},
\]

where \( f_{\hat{m} + 1} \) refers to the frequency of sub-carrier that entire channel power is focused on.

For example, Fig.6 shows the normalized beamspace channel power distribution of \( l \)-th path component when the system parameters are \( \varphi_l = 0.4, N = 128, M = 512 \) and \( B = 4 \) GHz. Specifically, the 116-th antenna focuses more channel power with the index of sub-carriers goes larger. Besides, the entire beamspace channel power is concentrated at sub-carrier 22 for the first time. The same phenomenon is observed from antenna 117 to antenna 123, and the whole process presents the periodicity, which is defined as beam squint period (BSP). The BSP \( F_{bsp} \) refers to the number of sub-carriers that satisfies \( d_{l,m_1,m_2} = 1 \), which can be calculated as

\[
F_{bsp} = m_1 - m_2 = \frac{f_c}{\eta N \varphi_l}.
\]

Besides, the Fig.6 also shows the last entire beamspace channel power focuses at sub-carrier 512, antenna 123, and there are 7 complete BSPs when \( \varphi_l = 0.4 \).

**B. THE PROPOSED TWO-STAGE SCHEME**

In this subsection, we propose a two-stage scheme to estimate the beamspace channel. The first phase is the normalized AoA estimation based on the power focusing effect, and the channel gain is obtained by the LS algorithm in phase two.

For the \( l \)-th path component, the normalized AoA estimation is tightly related to the index of antennas that beamspace channel power focused. Meanwhile, the values of practical AoAs are continuous, resulting in beamspace channel power distribute more than one beam (antenna). Hence, the results of on-grid angle estimation approaches are inaccurate.

However, due to the power focusing effect of beam squint, the level of power leakage is not the same for different sub-carriers, but it shows a tendency that the beamspace channel power slowly focuses on one antenna and then diverges with sub-carriers. The key idea of normalized AoA estimation is to find the index of antenna and sub-carrier that the entire channel power of \( l \)-th path component is focused on. Based on the analysis of the power focusing effect, we know that at
least one complete BSP is needed to obtain the entire channel power and the corresponding normalized AoA, and with a larger number of sub-carriers or BSPs, the result becomes more accurate. The pseudo-code of the proposed two-stage beamspace channel estimation scheme is summarized in Algorithm 2.

Specifically, we firstly get the beam squint matrix \( Q_l \) of the \( l \)-th path component based on Algorithm 1. Then, the sliding window (SWin) is adopted to capture the largest power point based on the effect of power focusing.

The SWin is defined as \( \Theta_m = \{ m - \Delta_m, \ldots, m + \Delta_m \} \), where \( \Delta_m \) represents the size of SWin. For each complete BSP, we will get one point that the entire channel power of the \( l \)-th path component is focused on. Based on beam squint matrix \( Q_l \), we assume the range of \( n_{l,m}^i \) is from \( n_{l,e} \) to \( n_{l,s} \) (\( n_{l,e} > n_{l,s} \)), and at least have \([n_{l,e} - n_{l,s}] + 1\) complete BSPs. For \([n_{l,s} < n_{l,e}] < [n_{l,e}] \), we find the center point \( (n_{l,i}, n_{l,i}) \) by

\[
(n_{l,i}, n_{l,i}) = \arg \max_m \| A_l(n_{l,i}, \Theta_m) \|^2. \tag{35}
\]

After getting \((n_{l,i}, n_{l,i})\), the normalized AoA can be calculated based on the following equations:

\[
\begin{align*}
\hat{\theta}_{l,i} &= \frac{n_{l,i} + \frac{f_c}{N} - (n_{l,s} + \frac{1}{2})}{N} - \frac{f_c}{N} + \frac{f_c}{n_{l,s} + 1} - \frac{n_{l,e} - 1}{N}, \\
\hat{\phi}_{l,i} &= \frac{2n_{l,e} - (N + 1) \eta N \hat{\phi}_{l,i}/f_c}{2N - f_c + (n_{l,s} - 1) \eta}.
\end{align*} \tag{37}
\]

The final result for the \( l \)-th path component is obtained by average function, i.e.,

\[
\hat{\theta}_l = \frac{\sum_{n_{l,s} = n_{l,s} - 1} n_{l,e}}{n_{l,e} - n_{l,s} - 1}. \tag{38}
\]

After the AoA estimation, we can get \( \hat{c}_{l,m} \) by

\[
\hat{c}_{l,m} = U^H \Xi_m(\hat{\phi}_l), m = 1, \ldots, M. \tag{39}
\]

Then, the channel path gain can be obtained by the LS algorithm as

\[
\hat{\beta}_{l,m} = (W \hat{c}_{l,m})^H y_m. \tag{40}
\]

Then, the influence of the \( l \)-th path components is removed in step 11. The beamspace channel can be recovered based on (8) after all path components have been estimated.

\section*{C. ALGORITHM SUMMARY AND PERFORMANCE ANALYSIS}

In this subsection, we summarize the proposed two-stage scheme and analyze its CRLB and computational complexity.

Note that the proposed two-stage scheme is different from conventional CS-based schemes. The CS-based schemes estimate the support by exploiting the sparsity in the angle or the time domain. By contrast, the proposed two-stage scheme aims to find out the largest element with no power leakage by using the power focusing effect of beam squint, which results in the normalized AoA can be estimated accurately. The advantage of the proposed two-stage scheme can be verified in the simulation results section.

Next, we give the the CRLB of the two-stage scheme when the prior knowledge of AoA is known.

Firstly, we rewrite (10) as

\[
\hat{y}_m = \hat{W} \hat{C}_m \hat{\beta}_m + \hat{n}_m. \tag{41}
\]

where \( \hat{C}_m \) is the ideal receive steering vector, and \( \hat{W} \) is the combining matrix at receiver. The CRLB of the two-stage scheme can be given as follows.

\[ \text{Lemma 2:} \quad \text{By assuming the prior knowledge of AoA at BS, the CRLB of the two-stage scheme can be expressed as} \]

\[
\mathbb{E}\left(\left\| \hat{h}_m - h_m \right\|^2_2 \right) \geq \frac{\sigma^2 L^2 \lambda_{\min}(D)}{\text{Tr}\left\{ (G_m)^H G_m \right\}}, \tag{42}
\]

where \( D = \hat{C}_m^H \hat{C}_m, G_m = \hat{W} \hat{C}_m \), and \( \lambda_{\min}(D) \) denotes the minimum eigenvalue of \( D \).

\[ \text{Proof 2:} \quad \text{Please refer to Appendix A.} \]

For the complexity analysis, the complexity of the proposed two-stage scheme is dominated by steps 2, 5, 9, 10, 11 and 17.
Step 2 has the complexity in order of $O(NN_{RF}MP)$. In step 5, the complexity is dominated by (35), which has complexity in order of $O(M)$. The complexity of step 9, 10, 11 is $O(N^2)$, $O(NP^2N_{RF}^2)$ and $O(NPN_{RF})$, respectively. In the end, the multiplication between $C_m$ of size $N \times L$ and $\hat{\beta}_m$ of size $L \times 1$ is required. Therefore, step 17 has complexity in order of $O(NL)$.

Consider steps 2, 5, 9, 10 and 11 are executed $L$ times, and step 17 is executed once. So the overall complexity of the two-stage scheme is

$$O(MN^2L) + O(MNP^2N_{RF}^2L)$$  \(43\)

Although the complexity of the two-stage scheme is higher than that of the other three schemes, it has the best performance among those schemes, which is demonstrated in the simulation section.

V. SIMULATION RESULTS

In this section, the simulation results are presented to evaluate the performance of the proposed beamspace channel estimation schemes. For the wideband mmWave systems, the BS equips the $N = 128$ ULA array and $N_{RF} = 8$ RF chains to serve $K = 8$ single-antenna users with OFDM modulation, which has $M = 512$ sub-carriers. The carrier frequency is $f_c = 28$ GHz, and the bandwidth is $B = 4$ GHz. For the channel parameters, the number of scattering paths is $L = 3$, the AoA is uniformly distributed in $\left(-\frac{\pi}{3}, \frac{\pi}{3}\right)$, the path gain $\beta_l$ is based on Gaussian distribution $CN(0, 1)$, and the time delay is uniformly distributed in $[0, 20]$ ns. While the SNR is defined as $SNR = 1/\sigma^2$. Besides, we use $P = 10$ time slots for pilot transmission and set $\omega = 4$ for the CS-based schemes. All simulation results are calculated over 1000 Monte Carlo trials.

A. AOA ESTIMATION PERFORMANCE COMPARISON

Firstly, we evaluate the performance of normalized AoA estimation. The Euclidean distance between the actual steering vector and estimated steering vector is defined to evaluate the accuracy of normalized AoA estimation [11].

$$d_{AoA} = \frac{1}{ML} \mathbb{E}\left(\sum_{m=1}^{M} \sum_{l=1}^{L} ||\Xi_m(\hat{\phi}_l) - \Xi_m(\hat{\phi}_l)||_2\right).$$  \(44\)

Fig.7 shows the performance of normalized AoA estimation at different SNR levels. It can be observed that the performance of the proposed scheme is better than all considered schemes at all SNR regions, which shows the advantage of the AoA estimation scheme. Note that the proposed scheme in [11] and MUSIC algorithm [18] are designed for narrowband without the parameter for the number of sub-carriers. Fig.7 also shows that the performance of the proposed scheme becomes more accurate with $N$ or $M$ going larger. The reason is that the quantization error decreases with the increase of the number of antennas and sub-carriers, which is caused by the calculation of pre-defined directions and beam distance. Moreover, the proposed scheme has a preformance floor when SNR is high, caused by the quantization error due to the limited number of antennas and sub-carriers.

Fig.8 illustrates the $d_{AoA}$ against pilot length $P$ of the proposed normalized AoA estimation scheme when SNR is set to 0 dB. It can be shown that systems of considered parameters have lower $d_{AoA}$ as the pilot of length $P$ going larger. Besides, the accuracy of the proposed normalized AoA estimation scheme becomes more precise with the increase of the number of antennas and sub-carrier, which is consistent with the previous analysis. Moreover, Fig.8 indicates that we can increase the number of antennas and sub-carriers to get more accurate results instead of increasing the pilot length.

In addition to the Euclidean distance as evaluation criteria, we also evaluate the performance of normalized AoA estimation from another perspective. The probability of successful normalized AoA estimation is defined as the ratio of the
The normalized mean square error (NMSE) is adopted as performance evaluation criteria, which is defined as

\[
\text{NMSE} = \mathbb{E} \left( \frac{1}{M} \sum_{m=1}^{M} \frac{\| \tilde{h}_m - \hat{h}_m \|^2}{\| \tilde{h}_m \|^2} \right)
\]  

(45)

Fig.10 shows the performance of all considered schemes in wideband wireless communication systems. From Fig.10, we can see that the performance of the proposed two-stage scheme is better than other schemes in all SNR regions, and the performance of the BSOMP scheme is better than other CS-based schemes and close to the oracle LS scheme. Specifically, the NMSE of the OMP-based scheme is not satisfying when SNR is low because it does not use the unique sparse structure of the wideband beamspace channel. For the SOMP-based scheme, the performance decline when SNR goes high, which is caused by the reason the common support assumption is no longer valid because of the beam squint effect. By contrast, the BSOMP scheme has better accuracy than both OMP and SOMP based schemes since it considers the sparse structure of the wideband beamspace channel. The BSOMP scheme can achieve the NMSE quite close to the oracle LS scheme, which means the support detection is accurate. Besides, the proposed two-stage scheme has the best performance since it exploits the sparse structure in the normalized AoA estimation phase and reconstructs the channel according to the channel model. The four CS-based schemes reach the NMSE floor when SNR is high since they only consider the wideband beamspace channel’s nonzero elements and ignore low power elements, which induce the error. And the proposed two-stage scheme also reaches the NMSE floor when SNR is high. This can be explained by the fact that there is quantization error at the normalized AoA estimation phase due to the limited number of antennas and sub-carriers, which cannot be eliminated.

Fig.11 depicts NMSE comparison against the bandwidth \( B \) when the SNR is 20 dB, and other simulation parameters remain the same as Fig.10. When \( B \) is small, the NMSE of SOMP is close to the BSOMP scheme. However, the performance of SOMP decreases as \( B \) increases and is even worse than the OMP scheme when \( B \) is larger than 3GHz. This can be explained by the fact that for small \( B \), the level of beam squint is less pronounced and the support at different sub-carriers is similar. As \( B \) increases, the support of the wideband beamspace channel at different sub-carriers will be more divergent, and the common support assumption will be more unreliable. However, the performance of BSOMP and two-stage scheme do not degrade with \( B \), which means both our schemes are robust to bandwidth even if the level of beam squint is not pronounced.

Fig.12 provides NMSE performance against pilot length \( P \) when SNR is 20dB, and other parameters remain the same as in Fig.10. It can be observed that all considered schemes have better NMSE as pilot length becomes longer.
For the same pilot length $P$, the two-stage scheme always has better NMSE than all considered schemes, and the BSOMP scheme is better than other CS-based schemes. For different lengths of the pilot, the NMSE gap between the proposed two-stage scheme and other schemes is quite large. This suggests that the two-stage can efficiently reduce the pilot overheads for the beamspace channel estimation.

Fig.13 shows the achievable sum-rate with different channel estimation schemes, where the hybrid precoding method in [38] is adopted. From Fig.13, we learn that the system achieves a higher sum rate by utilizing the proposed two-stage scheme, which is very close to one with the perfect channel when SNR is low (e.g., 0 dB) and moderate (e.g., 15 dB). The system using the BSOMP scheme has a higher sum rate than that uses other CS-based schemes. Besides, the sum rate of using a two-stage scheme at 0 dB is greater than that of other CS-based schemes at 15 dB, which indicates the two-stage scheme performs well, especially at low SNR. Moreover, the performance of all schemes varies with the channel estimation SNR. This can be explained by the fact that the sum rate is dominated by channel estimation error when the SNR for data transmission is high.

**VI. CONCLUSION**

In this paper, we investigate the beamspace channel estimation problem for the mmWave MIMO systems relying on the ULA array. We propose a BSOMP scheme and a novel two-stage beamspace channel estimation framework. For the BSOMP scheme, the support is detected by selecting the BWins generated to capture a single path beamspace channel power. For the proposed two-stage scheme, the normalized AoA is estimated by utilizing frequency-dependent sparse structure and power focusing effect of beamspace channel, aiming to find the no power leakage antenna among all sub-carriers. After getting normalized AoA, the path gain is updated in the second phase. The simulation results show that the performance of the BSOMP scheme is better than other CS-based schemes, and the two-stage scheme is better than other schemes in all considered SNR regions at the same pilot overheads. In future work, we will consider the beam squint effect when the reconfigurable intelligent surface (RIS) is introduced and propose a suitable algorithm for channel estimation and phase design of RIS.

**APPENDIX A PROOF OF LEMMA 2**

Based on (8) and (10), the mean square error (MSE) is

$$
\mathbb{E} \left( \| \hat{h}_m - \tilde{h}_m \|_2^2 \right) = \mathbb{E} \left( \| \tilde{C}_m (\beta - \hat{\beta}) \|_2^2 \right) \geq \lambda_{\min}(\Lambda) \mathbb{E} \left( \| (\beta - \hat{\beta}) \|_2^2 \right)
$$

where inequality \((a)\) obtained by Rayleigh-Ritz ratio [9].
Same as in performance analysis of the BSOMP scheme, the Fisher matrix is
\[
J = \frac{1}{\sigma^2} G_m^H G_m, \quad (47)
\]
where \( G_m = \mathbf{W} \tilde{C}_m \). Therefore, the CRLB of \( \hat{\beta} \) is obtained by
\[
E\left( \left\| \hat{\beta} - \beta \right\|^2 \right) \geq \text{Tr} \left( J^{-1} \right) \quad (48)
\]
and the \( \text{Tr} \left\{ \left( G_m^H G_m \right)^{-1} \right\} \) can be calculated as
\[
\text{Tr} \left\{ \left( G_m^H G_m \right)^{-1} \right\} = \sum_{l=1}^{L} \lambda_l \left( \left( G_m^H G_m \right)^{-1} \right) = \sum_{l=1}^{L} \lambda_l \left( \left( G_m^H G_m \right)^{-1} \right) \geq \frac{L^2}{\text{Tr} \left\{ (G_m^H G_m) \right\}} \quad (49)
\]
where the inequality \((a)\) derives from the arithmetic-harmonic means inequality \([39]\) and \( \lambda_i(\cdot) \) denotes the \( i \)-th eigenvalue. Substituting (49) and (48) into (46), we finally get the CRLB of the two-stage scheme.

REFERENCES
[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” IEEE Communications Magazine, vol. 52, no. 2, pp. 186–195, 2014.
[2] T. L. Marzetta, “Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas,” IEEE Transactions on Wireless Communications, vol. 9, no. 11, pp. 3590–3600, 2010.
[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An Overview of Massive MIMO: Benefits and Challenges,” IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 5, pp. 742–758, 2014. [Online]. Available: https://dx.doi.org/10.1109/jstsp.2014.2317671
[4] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” IEEE Communications Magazine, vol. 49, no. 6, pp. 101–107, 2011.
[5] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!” IEEE Access, vol. 1, pp. 335–349, 2013.
[6] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, “Millimeter-wave massive MIMO: the next wireless revolution?” IEEE Communications Magazine, vol. 52, no. 9, pp. 56–62, 2014.
[7] S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges,” Proceedings of the IEEE, vol. 102, no. 3, pp. 366–385, 2014.
[8] J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, “On channel estimation in ofdm systems,” in 1995 IEEE 45th Vehicular Technology Conference. Countdown to the Wireless Twenty-First Century, vol. 2, 1995, pp. 815–819 vol.2.
[9] J. Lee, G. Gil, and Y. H. Lee, “Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems in Millimeter Wave Communications,” IEEE Transactions on Communications, vol. 64, no. 6, pp. 2370–2386, 2016.
[10] Z. Marzi, D. Ramasamy, and U. Madhow, “Compressive Channel Estimation and Tracking for Large Arrays in mm-Wave Picocells,” IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 3, pp. 514–527, 2016.
[11] Q. Qin, L. Gui, P. Cheng, and B. Gong, “Time-varying channel estimation for millimeter wave multiuser MIMO systems,” IEEE Transactions on Vehicular Technology, vol. 67, no. 10, pp. 9435–9448, 2018.
[12] V. Venugopal, A. Alkhateeb, N. G. Prelic, and T. W. Heath, “Channel estimation for hybrid architecture-based wideband millimeter wave systems,” IEEE Journal on Selected Areas in Communications, vol. 35, no. 9, pp. 1996–2009, 2017.
[13] Q. Qin, L. Gui, B. Gong, and S. Luo, “Sparse Channel Estimation for Massive MIMO-OFDM Systems Over Time-Varying Channels,” IEEE Access, vol. 6, pp. 33 740–33 751, 2018.
[14] E. Vlachos, G. C. Alexandropoulos, and J. Thompson, “Wideband mimo channel estimation for hybrid beamforming millimeter wave systems via random spatial sampling,” IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 5, pp. 1136–1150, 2019.
[15] A. Liao, Z. Gao, H. Wang, S. Chen, M.-S. Alouini, and H. Yin, “Closed-loop sparse channel estimation for wideband millimeter-wave full-dimensional mimo systems,” IEEE Transactions on Communications, vol. 67, no. 12, pp. 8329–8345, 2019.
[16] X. Gao, L. Dai, S. Han, I. Chih-Lin, and X. Wang, “Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array,” IEEE Transactions on Wireless Communications, vol. 16, no. 9, pp. 6010–6021, 2017.
[17] C. K. Anjinappa, A. C. Gurbuz, Y. Yaprıcı, and I. Guvenc, “Off-grid aware channel and covariance estimation in mmwave networks,” IEEE Transactions on Communications, vol. 68, no. 6, pp. 3908–3921, 2020.
[18] F. Dong, W. Wang, Z. Huang, and P. Huang, “High-resolution angle-of-arrival and channel estimation for mmwave massive mimo systems with lens antenna array,” IEEE Transactions on Vehicular Technology, vol. 69, no. 11, pp. 12 963–12 973, 2020.
[19] J. H. Brady and A. M. Sayeed, “Wideband communication with high-dimensional arrays: New results and transceiver architectures,” in 2015 IEEE International Conference on Communication Workshop (ICCWC), 2015, pp. 1042–1047.
[20] 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 Journal of Selected Topics in Signal Processing, vol. 10, no. 3, pp. 436–453, 2016.
[21] M. Jian, F. Gao, Z. Tian, S. Jin, and S. Ma, “Angle-domain aided UL/DL channel estimation for wideband mmWave massive MIMO systems with beam squint,” IEEE Transactions on Wireless Communications, vol. 18, no. 7, pp. 3515–3527, 2019.
[22] M. Jian, F. Gao, S. Jin, H. Lin, and L. Xing, “Wideband channel estimation for mmwave massive mimo system with off-grid sparse bayesian learning,” in 2018 IEEE Global Communications Conference (GLOBECOM), 2018, pp. 1–6.
[23] H. Yu, P. Guan, Y. Wang, and Y. Zhao, “Spatial spectrum-based channel estimation for wideband mimo system with beam squint,” IEEE Access, vol. 9, pp. 16 164–16 172, 2021.
[24] 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 Transactions on Signal Processing, vol. 68, pp. 49–64, 2019.
[25] B. Wang, M. Jian, F. Gao, G. Y. Li, and H. Lin, “Beam squint and channel estimation for wideband mmWave massive MIMO systems,” IEEE Transactions on Signal Processing, vol. 67, no. 23, pp. 5903–5908, 2019.
[26] J. Rodriguez-Fernandez and N. Gonzalez-Prelcic, “Channel Estimation for Frequency-Selective mmWave MIMO Systems with Beam-Squint,” in 2018 IEEE Global Communications Conference (GLOBECOM), 2018, pp. 1–6.
[27] B. Wang, F. Gao, G. Y. Li, S. Jin, and H. Lin, “Wideband Channel Estimation for mmWave Massive MIMO Systems with Beam Squint Effect,” in 2018 IEEE Global Communications Conference (GLOBECOM), 2018, pp. 1–6.
[28] 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 Journal on Selected Areas in Communications, vol. 35, no. 7, pp. 1524–1538, 2017.
[29] X. Ma, F. Yang, S. Liu, J. Song, and Z. Han, “Design and optimization on training sequence for mmWave communications: A new approach for sparse channel estimation in massive MIMO,” IEEE Journal on Selected Areas in Communications, vol. 35, no. 7, pp. 1486–1497, 2017.
[30] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, “Spatial-and frequency-wideband effects in millimeter-wave massive MIMO systems,” IEEE Transactions on Signal Processing, vol. 66, no. 13, pp. 3395–3406, 2018.
[31] A. Sayeed and J. Brady, “Beamspace MIMO for high-dimensional multiuser communication at millimeter-wave frequencies,” in 2013 IEEE global communications conference (GLOBECOM). IEEE, 2013, pp. 3679–3684.
[32] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter wave channel modeling and cellular capacity evaluation,” IEEE journal on selected areas in communications, vol. 32, no. 6, pp. 1164–1179, 2014.
[33] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, “Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels,” Proceedings of the IEEE, vol. 98, no. 6, pp. 1058–1076, 2010.
[34] M. F. Duarte and Y. C. Eldar, “Structured compressed sensing: From theory to applications,” IEEE Transactions on signal processing, vol. 59, no. 9, pp. 4053–4085, 2011.
[35] Z. Gao, C. Hu, L. Dai, and Z. Wang, “Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels,” IEEE Communications Letters, vol. 20, no. 6, pp. 1259–1262, 2016.
[36] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” IEEE Transactions on information theory, vol. 53, no. 12, pp. 4655–4666, 2007.
[37] S. M. Kay, Fundamentals of statistical signal processing: estimation theory. Prentice-Hall, Inc., 1993.
[38] L. Liang, W. Xu, and X. Dong, “Low-Complexity Hybrid Precoding in Massive Multiuser MIMO Systems,” IEEE Wireless Communications Letters, vol. 3, no. 6, pp. 653–656, 2014.
[39] M. Varanasi, C. Mullis, and A. Kapur, “On the Limitation of Linear MMSE Detection,” IEEE Transactions on Information Theory, vol. 52, no. 9, pp. 4282–4286, 2006.