

Channel Estimation for Extremely Large-Scale MIMO: Far-Field or Near-Field?

Mingyao Cui, *Student Member, IEEE*, and Linglong Dai^{id}, *Fellow, IEEE*

Abstract—Extremely large-scale multiple-input-multiple-output (XL-MIMO) is promising to meet the high rate requirements for future 6G. To realize efficient precoding, accurate channel state information is essential. Existing channel estimation algorithms with low pilot overhead heavily rely on the channel sparsity in the angular domain, which is achieved by the classical far-field planar-wavefront assumption. However, due to the non-negligible near-field spherical-wavefront property in XL-MIMO, this channel sparsity in the angular domain is not achievable. Therefore, existing far-field channel estimation schemes will suffer from severe performance loss. To address this problem, in this paper, we study the near-field channel estimation by exploiting the polar-domain sparsity. Specifically, unlike the classical angular-domain representation that only considers the angular information, we propose a polar-domain representation, which simultaneously accounts for both the angular and distance information. In this way, the near-field channel also exhibits sparsity in the polar domain, based on which, we propose on-grid and off-grid near-field XL-MIMO channel estimation schemes. Firstly, an on-grid polar-domain simultaneous orthogonal matching pursuit (P-SOMP) algorithm is proposed to efficiently estimate the near-field channel. Furthermore, an off-grid polar-domain simultaneous iterative gridless weighted (P-SIGW) algorithm is proposed to improve the estimation accuracy. Finally, simulations are provided to verify the effectiveness of our schemes.

Index Terms—Near-field, XL-MIMO, hybrid precoding, channel representation, channel estimation.

I. INTRODUCTION

MASSIVE multiple-input-multiple-output (MIMO) is one of the most critical technologies for current 5G communications [2]. Equipped with massive antenna arrays at the base station (BS), massive MIMO can improve the spectral efficiency by orders of magnitude through beamforming or multiplexing. For future 6G communications,

extremely large-scale MIMO (XL-MIMO), where the number of antennas can be much larger than that for massive MIMO [3], can effectively achieve 10-fold increases in spectral efficiency [4]. On the other hand, benefiting from the rich spectrum resource at millimeter-wave (mmWave) band or terahertz (THz) band, high-frequency communications can provide largely available bandwidth [5]. Meanwhile, the very small size of high-frequency antennas favorably enables the deployment of XL-MIMO with an extremely large number of antennas. Therefore, high-frequency XL-MIMO has been widely considered as a key enabling technology for future 6G [4].

Similar to the current 5G mmWave massive MIMO, hybrid precoding architecture is widely considered for high-frequency XL-MIMO [6], since the power consumption of the high-frequency radio-frequency (RF) chain is very high [7]. Efficient hybrid precoding requires accurate channel state information at the base station. Unfortunately, since the number of RF chains in hybrid precoding is much smaller than the number of antennas, the BS cannot observe the signals at each antenna simultaneously [8]. This will result in unacceptable pilot overhead, especially when the number of antennas is huge in XL-MIMO systems.

A. Prior Works

To alleviate the high pilot overhead in the channel estimation, in current 5G massive MIMO systems, by exploiting the channel sparsity in the angular domain, some compressive sensing (CS) based algorithms have been studied to accurately estimate the high-dimensional channels with low pilot overhead [8]–[13]. For example, by utilizing the angular-domain sparsity, the classical orthogonal matching pursuit (OMP) algorithm was used in [9] to recover the angular-domain channel, where the channel was transformed into its angular-domain representation through the standard spatial Fourier transform. As for wideband systems, a simultaneous OMP (SOMP) algorithm was proposed in [10], which jointly recovered the channels at different subcarriers based on the common support assumption, i.e., the sparse supports in the angular domain at different subcarriers were assumed to be the same. Moreover, by taking into account the colored noise induced by hybrid precoding, the pre-whitening procedure was introduced in the SOMP algorithm [8]. Furthermore, the message passing (MP) algorithms are utilized in [11], [12] to recover sparse angular-domain channel with prior information. Note that all solutions above assumed that the angle of departures/arrivals (AoDs/AoAs) lie in discrete points in the angular domain (i.e., “on-grid” AoDs/AoAs), while

Manuscript received August 29, 2021; revised December 1, 2021; accepted January 10, 2022. Date of publication January 26, 2022; date of current version April 18, 2022. This work was supported in part by the National Key Research and Development Program of China (Grant No.2020YFB1807201), in part by the National Natural Science Foundation of China (Grant No. 62031019), and in part by the European Commission through the H2020-MSCA-ITN META WIRELESS Research Project under Grant 956256. An earlier version of this paper was presented in part at the IEEE Global Communications Conference (IEEE GLOBECOM'21) [1] [DOI: 10.1109/GLOBECOM46510.2021.9685542]. The associate editor coordinating the review of this article and approving it for publication was S. Lambotharan. (Corresponding author: Linglong Dai.)

The authors are with the Beijing National Research Center for Information Science and Technology (BNRist) and the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: cmy20@mails.tsinghua.edu.cn; daill@tsinghua.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCOMM.2022.3146400>.

Digital Object Identifier 10.1109/TCOMM.2022.3146400

0090-6778 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

the actual AoDs are continuously distributed (i.e., “off-grid” AoDs/AoAs) in practical systems. To solve the resolution limitation of these on-grid algorithms, several off-grid solutions were studied in [14]–[17]. For instance, [17] proposed a simultaneous iterative gridless weighted (SIGW) algorithm to directly estimate the channel parameters, including the angles and the path gains, and thus the channel estimation accuracy could be improved.

It is worth noting that all solutions above heavily rely on the channel sparsity in the angular domain. However, this channel sparsity may not be achievable in XL-MIMO systems, and thus existing channel estimation schemes cannot be directly applied to XL-MIMO. Specifically, the change from massive MIMO to XL-MIMO not only means the increase in antenna number, but also leads to the fundamental change in the electromagnetic field structure. The radiation field of the electromagnetic wave can be divided into two regions, i.e., the far-field region and the near-field region [18]. The boundary to divide the near-field and the far-field is determined by the Rayleigh distance [18], which is proportional to the square of the number of antennas. In current 5G massive MIMO systems, as the number of antennas is not very large, the Rayleigh distance is usually several meters, which is negligible in practice. Therefore, the wavefront can be simply modeled under the far-field planar wavefront assumption, which only depends on the angle of departure/arrival (AoD/AoA) of the channel. Note that the channel sparsity in the angular domain is derived from this planar wavefront assumption. In future 6G XL-MIMO systems, due to the significant increase in the number of antennas, the Rayleigh distance can be up to several hundreds of meters, thus the near-field region in XL-MIMO systems becomes no longer negligible. When the receiver is located in the near-field region, the wavefront should be accurately modeled under the spherical wavefront assumption [19], which is decided by not only the AoD/AoA but also the distance between the BS and the scatter. For this near-field channel, a severe energy spread effect will be introduced for the classical angular-domain channel representation, i.e., the energy of a single near-field path component will be spread into multiple angles. In this case, the angular-domain channel may not be sparse, and thus existing far-field channel estimation schemes based on the channel sparsity in the angular domain will suffer from severe performance degradation. Consequently, to support the ultra-high data rate for future 6G, the efficient near-field channel estimation algorithm is essential for practical high-frequency XL-MIMO systems.

Unfortunately, up to now, there are no related works on near-field channel estimation for high-frequency XL-MIMO. Under the near-field spherical wavefront conditions, there are some existing works [20]–[23] investigating a similar problem in wireless sensing systems, i.e., the near-field localization problem. For instance, by exploiting the structure of the signal covariance matrix, several high-order statistic based methods, (e.g., the subspace-based algorithms [20], [21] or high-order MUSIC algorithms [22]) have been proposed to estimate the AoD and distance between the source and the receiver in the near-field. Furthermore, the OMP algorithm was improved in [23] to locate near-field scatters. However, all near-field

localization methods above assume that the dimension of received signals is equal to or larger than that of the channel, which is not valid in the practical high-frequency XL-MIMO systems. To the best of our knowledge, the near-field channel estimation in high-frequency XL-MIMO systems has not been studied in the literature.

B. Our Contributions

To fill in this gap, in this paper, the important problem of near-field channel estimation for XL-MIMO is studied, which is realized by replacing the classical angular-domain representation with a polar-domain representation.¹ Specifically, our contributions are summarized as follows.

- Firstly, by comparing the difference between the far-field and near-field channels, we reveal the energy spread effect when the practical near-field channel is transformed into the angular domain by using the classical spatial Fourier matrix. This energy spread effect indicates that the energy of a single near-field path component will be spread into multiple angles, and thus the near-field channel in the angular domain is no longer sparse.
- To deal with the energy spread effect, we propose a polar-domain representation of the near-field channel. Unlike the angular-domain representation of the far-field channel that only considers the angular information, the polar-domain representation accounts for both the angular and distance information simultaneously. To design the polar-domain transform matrix, we utilize the Fresnel approximation to approximate the near-field channel. Then, based on the Fresnel function, we analyze how to sample the angle and distance in the polar domain to reduce the column coherence of the transform matrix. Analytical results show that, the angle should be sampled uniformly, while the distance should be sampled non-uniformly. We further point out that both the far-field and near-field channels exhibit sparsity in the polar domain, and thus the classical angular-domain representation is a special case of the proposed polar-domain representation.
- By exploiting the channel sparsity in the polar domain, an on-grid polar-domain simultaneous orthogonal matching pursuit (P-SOMP) algorithm is proposed to estimate the near-field channel efficiently. To further improve the estimation accuracy, we propose an off-grid polar-domain simultaneous iterative gridless weighted (P-SIGW) algorithm, where the near-field channel parameters are directly estimated. Unlike existing off-grid channel estimation algorithms [14]–[17] that only estimate path gains and angles, the proposed P-SIGW algorithm simultaneously recovers the path gains, angles, and distances. Finally, numerical results are provided to verify the effectiveness of the proposed algorithms.

¹Simulation codes are provided to reproduce the results in this paper: <http://oa.ee.tsinghua.edu.cn/dailinglong/publications/publications.html>.

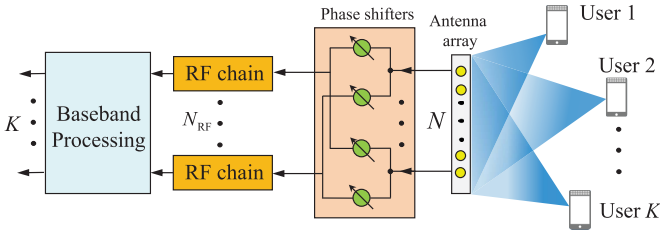


Fig. 1. XL-MIMO system with hybrid precoding.

C. Organization and Notation

Organization: The remainder of this paper is organized as follows. In section II, the system model is introduced, and the energy spread effect is revealed. In section III, the polar-domain representation is proposed, and the method to design the polar-domain transform matrix is provided. In section IV, the on-grid P-SOMP algorithm and the off-grid P-SIGW algorithm are proposed. Simulations are carried out in Section V, and finally conclusions are drawn in Section VI.

Notation: Lower-case and upper-case boldface letters represent vectors and matrices, respectively; $X_{p,q}$ denotes the (p,q) -th entry of the matrix \mathbf{X} ; $\mathbf{X}_{p,:}$ and $\mathbf{X}_{:,p}$ denote the p -th row and the p -th column of the matrix \mathbf{X} ; $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and conjugate transpose, respectively; $|\cdot|$ denotes the absolute operator; $\text{Tr}(\cdot)$ denotes the trace operator; $\mathcal{CN}(\mu, \Sigma)$ and $\mathcal{U}(a, b)$ denote the Gaussian distribution with mean μ and covariance Σ , and the uniform distribution between a and b , respectively.

II. SYSTEM MODEL

As shown in Fig. 1, we consider an uplink time division duplexing (TDD) based XL-MIMO OFDM communication system in this paper. The hybrid precoding architecture is employed at the BS. The BS is equipped with N_{RF} RF chains and an N -antenna uniform linear array, where $N_{\text{RF}} \ll N$. The antenna spacing is $d = \frac{\lambda_c}{2}$, where λ_c is the carrier wavelength. K single-antenna users are served with M subcarriers simultaneously, where $K \leq N_{\text{RF}}$. For uplink channel estimation, we assume the K users transmit mutual orthogonal pilot sequences to the BS [24], e.g., orthogonal time or frequency resources are utilized for different users to transmit pilot sequences. Therefore, channel estimation for each user is independent. Without loss of generality, we consider an arbitrary user.

Specifically, we denote $x_{m,p}$ as the transmit pilot at the m -th subcarrier in time slot p . Then, the received pilot $\mathbf{y}_{m,p} \in \mathbb{C}^{N_{\text{RF}} \times 1}$ is

$$\mathbf{y}_{m,p} = \mathbf{A}_p \mathbf{h}_m x_{m,p} + \mathbf{A}_p \mathbf{n}_{m,p}, \quad (1)$$

where $\mathbf{A}_p \in \mathbb{C}^{N_{\text{RF}} \times N}$ denotes the analog combining matrix satisfying the constant modulus constraint $|\mathbf{A}_p(i, j)| = \frac{1}{\sqrt{N}}$, and $\mathbf{n}_{m,p} \in \mathbb{C}^{N_{\text{RF}} \times 1}$ denotes the Gaussian complex noise following the distribution $\mathcal{CN}(0, \sigma^2 \mathbf{I}_{N_{\text{RF}}})$. Define P as the pilot length and assume $x_{m,p} = 1$ for $p = 1, 2, \dots, P$. Then the overall received pilot sequence $\mathbf{y}_m = [\mathbf{y}_{m,1}^T, \dots, \mathbf{y}_{m,P}^T]^T$ at

the m -th subcarrier can be denoted as

$$\mathbf{y}_m = \mathbf{A} \mathbf{h}_m + \mathbf{n}_m, \quad (2)$$

where $\mathbf{n}_m = [\mathbf{n}_{m,1}^T, \dots, \mathbf{n}_{m,P}^T]^T$ denotes the noise. $\mathbf{A} = [\mathbf{A}_1^T, \dots, \mathbf{A}_P^T]^T \in \mathbb{C}^{PN_{\text{RF}} \times N}$ denotes the overall observation matrix, where the elements in \mathbf{A} are independent and can be randomly generated from the set $\frac{1}{\sqrt{N}}\{-1, 1\}$ with equal probability. Since the BS antenna number N is very large, the dimension PN_{RF} of the received signal \mathbf{y}_m is usually much lower than N , which makes it challenging to estimate \mathbf{h}_m from \mathbf{y}_m .

Fortunately, the channel sparsity in the angular domain at high-frequency enables the compressive sensing (CS) based channel estimation methods, where the pilot length P can be significantly reduced. Specifically, users are assumed to be in the far-field region of the BS, where the channel is modeled under planar-wave assumption. At the m -th subcarrier, the classical far-field channel is expressed as [25]

$$\mathbf{h}_m^{\text{far-field}} = \sqrt{\frac{N}{L}} \sum_{l=1}^L g_l e^{-jk_m r_l} \mathbf{a}(\theta_l), \quad (3)$$

where $k_m = \frac{2\pi f_m}{c}$ denotes the wavenumber, L is the number of paths. Moreover, g_l , r_l , and θ_l are the complex path gain, the distance, and the angle of the l -th path, respectively. The steering vector $\mathbf{a}(\theta_l)$ on the angle $\theta_l \in [-1, 1]$ is derived from the planar-wave assumption, which is expressed as

$$\mathbf{a}(\theta_l) = \frac{1}{\sqrt{N}} [1, e^{j\pi\theta_l}, \dots, e^{j(N-1)\pi\theta_l}]^T. \quad (4)$$

Note that the phase of each element in the steering vector $\mathbf{a}(\cdot)$ is linear to the antenna index n , thus $\mathbf{a}(\cdot)$ is a discrete Fourier vector. Correspondingly, the channel $\mathbf{h}_m^{\text{far-field}}$ can be transformed into its angular-domain representation \mathbf{h}_m^A by the spatial Fourier transform matrix [11], i.e.,

$$\mathbf{h}_m^{\text{far-field}} = \mathbf{F} \mathbf{h}_m^A, \quad \text{远场的信道稀疏结构 用户是单天线 (5)}$$

where $\mathbf{F} \in \mathbb{C}^{N \times N}$ denotes the Fourier transform matrix. \mathbf{F} contains N orthogonal steering vectors uniformly sampled from the whole angular space as $\mathbf{F} = [\mathbf{a}(\theta_0), \dots, \mathbf{a}(\theta_{N-1})]$, where $\theta_n = \frac{2n-N+1}{N}$, $n = 0, 1, \dots, N-1$. As shown in Fig. 2, since the number of scatters is limited, the far-field channel \mathbf{h}_m^A in the angular domain is usually sparse. Utilizing this channel sparsity, some compressive sensing (CS) based channel estimation algorithms have been proposed to efficiently recover \mathbf{h}_m^A with low pilot overhead P [8]–[13].

However, this channel sparsity in the angular domain may no longer be achievable in XL-MIMO systems. The change from massive MIMO to XL-MIMO not only means the increase in antenna number, but also leads to the fundamental transformation of electromagnetic field structure. As shown in Fig. 3, the radiation field of electromagnetic wave can be divided into two regions, i.e., the far-field and near-field regions. The widely adopted boundary between these two fields is the Rayleigh distance $Z = \frac{2D^2}{\lambda_c}$ [18], where D and λ_c denote the array aperture and wavelength, respectively. As the antenna spacing is $d = \frac{\lambda_c}{2}$ and the number of

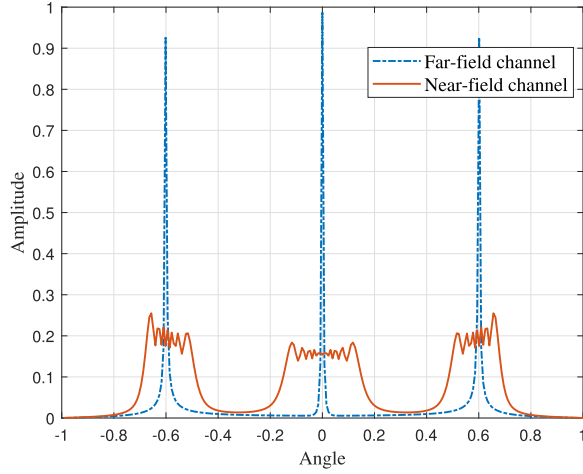


Fig. 2. Comparison of the angular-domain representations between the far-field and near-field channels. The BS antenna number is 256 and the carrier is 100 GHz. For far-field, the user is 100 meters from the BS, and for near-field, the user is 5 meters away from the BS while $L = 3$ paths are chosen.

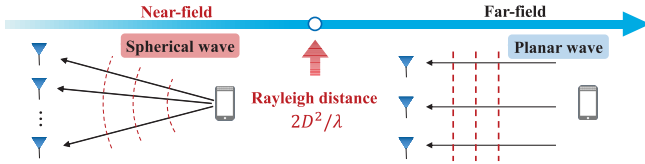


Fig. 3. The near-field region and the far-field region separated by the Rayleigh distance.

antennas is N , the array aperture of a uniform linear array is $D = Nd = N\frac{\lambda_c}{2}$. Therefore, the Rayleigh distance is $Z = \frac{1}{2}N^2\lambda_c$, which is proportional to N^2 . The physical meaning of Rayleigh distance Z is that, when the distance between the source and receiver is larger than Z , the radiation field is far-field, and the wavefronts can be approximated as planar waves. Otherwise, if the distance between the radiating source and the receiver is less than Z , the radiation field is near-field, and the wavefronts are spherical waves. In current 5G massive MIMO systems, as the array aperture is not very large, the Rayleigh distance is usually several meters, which is negligible in practice. However, in future 6G XL-MIMO systems, due to the significant increase in the number of antennas, the Rayleigh distance can be up to several hundreds of meters, thus the near-field region in XL-MIMO becomes not negligible. For instance, if the array aperture is 0.4 m and the carrier is 100 GHz, then the Rayleigh distance is around 107 meters, which covers a large part of a cell.

The spherical wave channel model in near-field region can be presented as [26]

$$\mathbf{h}_m = \sqrt{\frac{N}{L}} \sum_{l=1}^L g_l e^{-jk_m r_l} \mathbf{b}(\theta_l, r_l). \quad (6)$$

The main difference between the far-field channel (3) and the near-field channel (6) is the steering vector $\mathbf{b}(\cdot)$. The far-field steering vector \mathbf{a} is derived from the planar-wave assumption, while the near-field steering vector \mathbf{b} is derived from the

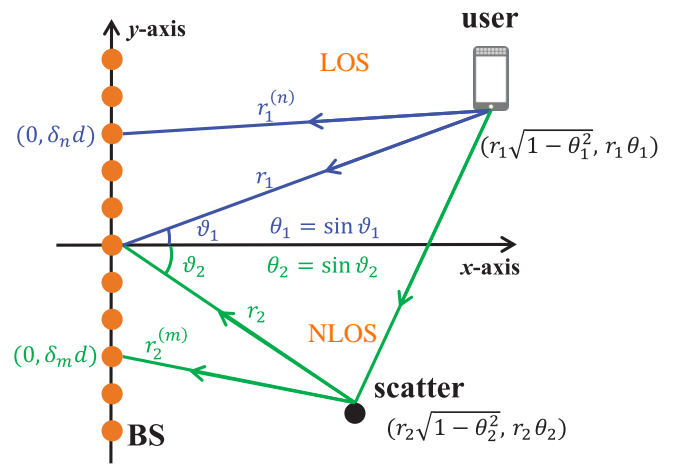


Fig. 4. The near-field channel model with two paths.

accurate spherical wave, i.e.,

$$\mathbf{b}(\theta_l, r_l) = \frac{1}{\sqrt{N}} [e^{-jk_c(r_l^{(0)} - r_l)}, \dots, e^{-jk_c(r_l^{(N-1)} - r_l)}]^T, \quad (7)$$

where $k_c = \frac{2\pi f_c}{c} = \frac{2\pi}{\lambda_c}$ denotes the wavenumber at the central carrier f_c . r_l denotes the distance between the BS and the scatter or user, and $r_l^{(n)}$ denotes the distance between the n -th BS antenna and the scatter or user. The schematic diagram of near-field channel model is plotted in Fig. 4. For expression simplicity, the channel is composed of a LOS path and an NLOS path. For the LOS path, $r_1^{(n)}$ denotes the distance between the n -th antenna and the user, while for the NLOS path, $r_2^{(m)}$ denotes the distance between the m -th antenna and the corresponding scatter. Suppose the coordinate of the n -th antenna is $(0, \delta_n d)$, where $\delta_n = \frac{2n-N+1}{2}$, $n = 0, 1, \dots, N-1$, then it can be derived from the geometry that $r_l^{(n)} = \sqrt{(r_l \sqrt{1-\theta_l^2} - 0)^2 + (r_l \theta_l - \delta_n d)^2} = \sqrt{r_l^2 + \delta_n^2 d^2 - 2r_l \theta_l \delta_n d}$, where $\theta_l \in [-1, 1]$ denotes the spatial angle. This spherical wave model indicates that the phase of each element in the steering vector $\mathbf{b}(\cdot)$ is nonlinear to the antenna index n , so $\mathbf{b}(\cdot)$ is not a Fourier vector. In this case, $\mathbf{b}(\cdot)$ cannot be described by a single far-field Fourier vector. As shown in Fig. 2, several far-field Fourier vectors should be jointly utilized to describe a near-field steering vectors $\mathbf{b}(\cdot)$. Consequently, the energy of one near-field path component is no longer concentrated in one angle, but spread towards multiple angles, which is called the energy spread effect in this paper. This energy spread effect implies that in the near-field, the channel \mathbf{h}_m^A in the angular domain may not be sparse. Thus existing far-field channel estimation schemes, based on angular-domain sparsity, will suffer from severe performance degradation in XL-MIMO systems.

III. NEAR-FIELD POLAR-DOMAIN REPRESENTATION

To realize efficient near-field channel estimation with reduced pilot overhead, in this section, we will propose a polar-domain representation of the XL-MIMO near-field channel to address the energy spread effect.

A. Polar-Domain Representation for the Near-Field Channel

Although it is observed from Fig. 2 that the near-field channel is not sparse in the angular-domain, as shown in (6), the number of paths is still limited, i.e., $L \ll N$. This indicates that the number of channel parameters to be estimated is still limited, and thus the near-field channel is also compressible.

To find a sparse representation for the near-field channel, we can refer to the derivation of angular-domain sparse representation (5) for the far-field channel. Specifically, as shown in (3), the far-field channel $\mathbf{h}_m^{\text{far-field}}$ can be regarded as the weighted sum of limited far-field steering vectors $\mathbf{a}(\theta)$, where $\mathbf{a}(\theta)$ only depends on the channel angle. Meanwhile, the angular-domain transform matrix \mathbf{F} in (5) is exactly composed of many far-field steering vectors $\mathbf{a}(\theta)$, where the angles θ are sampled from the entire angular domain to fully exploit the angular information of a far-field path component. Therefore, by utilizing this transform matrix \mathbf{F} , the angular-domain representation \mathbf{h}_m^A is sparse in the far-field.

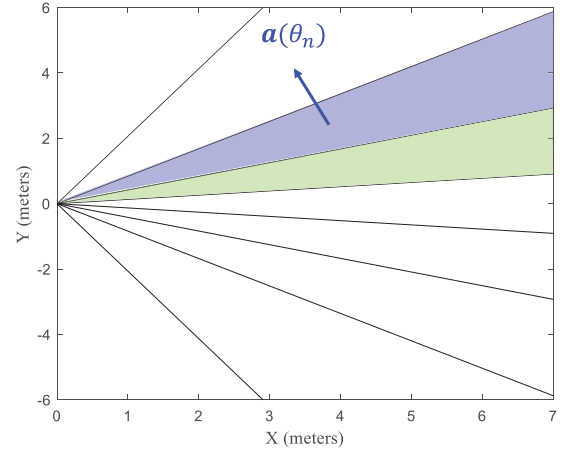
As for the sparse representation of the near-field channel \mathbf{h}_m , as shown in (6), \mathbf{h}_m can be regarded as the weighted sum of limited near-field steering vectors $\mathbf{b}(\theta, r)$, where $\mathbf{b}(\theta, r)$ depends not only the channel angle, but also the channel distance. Similar to the design of the existing matrix \mathbf{F} , we propose to design a new transform matrix \mathbf{W} , which is composed of many near-field steering vectors $\mathbf{b}(\theta, r)$, where the distances r and angles θ are sampled from the entire angular-distance domain. In this way, the angular and distance information of a near-field path component is fully exploited by \mathbf{W} . Since angles and distances represent the coordinates in the polar coordinate system, the angular-distance domain is called the polar domain. Accordingly, the matrix \mathbf{W} is called as the polar-domain transform matrix. This idea is intuitively shown in Fig. 5. Compared to the angular-domain transform matrix \mathbf{F} , which only samples different angles as shown in Fig. 5 (a), the polar-domain transform matrix \mathbf{W} simultaneously samples different angles and distances as shown in Fig. 5 (b).

Similar to the angular-domain representation $\mathbf{h}_m = \mathbf{F}\mathbf{h}_m^A$, the proposed polar-domain representation \mathbf{h}_m^P of the near-field channel is

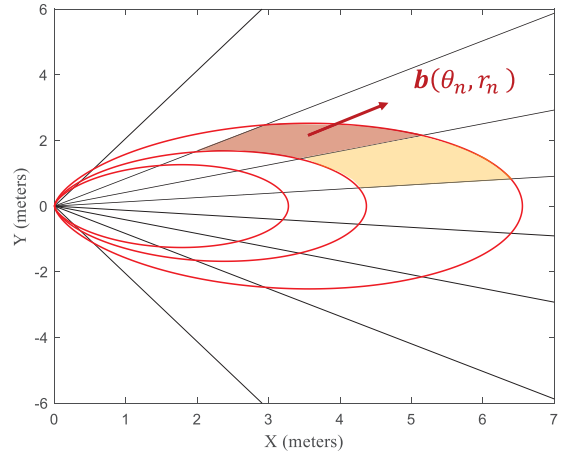
$$\mathbf{h}_m = \mathbf{W}\mathbf{h}_m^P, \quad (8)$$

where $\mathbf{W} \in \mathbb{C}^{N \times Q}$, and Q denotes the number of sampled near-field steering vectors in the polar domain. Analog to the angular-domain sparsity where the angular information in far-field is considered by the matrix \mathbf{F} , the proposed polar-domain transform matrix \mathbf{W} can account for both the angle and distance information of a near-field path component. Therefore, the near-field channel becomes sparse in the polar domain by utilizing the matrix \mathbf{W} . In this way, the energy spread effect for the near-field channel in the angular domain is avoided. Since the matrix \mathbf{W} simultaneously samples different angles and distances, the number of sampled vectors Q is not assumed to be equal to N , and generally, Q is larger than N . Therefore, \mathbf{W} is a wide matrix, but not a square matrix.

For the proposed polar-domain representation $\mathbf{h}_m = \mathbf{W}\mathbf{h}_m^P$, a fundamental question is how to sample the angles and



(a) The angular-domain representation



(b) The polar-domain representation

Fig. 5. Comparison between (a) the angular-domain representation and (b) the polar-domain representation.

distances to design the transform matrix \mathbf{W} ? Following the CS framework [27], to achieve the satisfying channel recovery accuracy, the sampling on angle and distance should make the column coherence $\mu = \max_{p \neq q} |\mathbf{b}(\theta_p, r_p)^H \mathbf{b}(\theta_q, r_q)|$ of the polar-domain transform matrix \mathbf{W} as small as possible, where $\mathbf{b}(\theta_p, r_p)$ and $\mathbf{b}(\theta_q, r_q)$ are two columns of \mathbf{W} . However, since the phase $-k_c(r^{(n)} - r)$ of n -th element in $\mathbf{b}(\theta, r)$ is non-linear to the antenna index n as shown in (7), it is intractable to get a close form of $|\mathbf{b}(\theta_p, r_p)^H \mathbf{b}(\theta_q, r_q)|$, which makes the design of \mathbf{W} difficult.

In the following discussions, we will approximately derive the column coherence between two near-field steering vectors, i.e., $f(\theta_p, \theta_q, r_p, r_q) = |\mathbf{b}(\theta_p, r_p)^H \mathbf{b}(\theta_q, r_q)|$. Based on the Fresnel approximation [26], the distance $r^{(n)}$ between the n -th antenna and the user or scatter can be approximated as $r^{(n)} = \sqrt{r^2 - 2r\delta_n d \theta + \delta_n^2 d^2} \stackrel{(a)}{\approx} r - \delta_n d \theta + \frac{\delta_n^2 d^2 (1 - \theta^2)}{2r}$, where (a) is derived by $\sqrt{1+x} \approx 1 + \frac{1}{2}x - \frac{1}{8}x^2$. It has been studied in [18] that the Fresnel approximation (a) is accurate when the distance between the BS and the user or scatter is larger than $0.5\sqrt{\frac{D^3}{\lambda_c}}$, which is much lower than the Rayleigh distance $\frac{2D^2}{\lambda_c}$. For instance, if the array aperture D is 0.4 m and the wavelength λ_c is 3 mm, then the Rayleigh

distance is 106.7 m, but $0.5\sqrt{\frac{D^3}{\lambda_c}}$ is only 2.3 m, which is nearly negligible. Therefore, the column coherence $f(\theta_p, \theta_q, r_p, r_q)$ can be approximated as

$$\begin{aligned} f(\theta_p, \theta_q, r_p, r_q) &= \left| \frac{1}{N} \sum_{\delta_n} e^{jk_c(r_p^{(n)} - r_q^{(n)})} \right| \\ &\approx \left| \frac{1}{N} \sum_{\delta_n} e^{jk_c \delta_n d(\theta_q - \theta_p) + jk_c \delta_n^2 d^2 \left(\frac{1-\theta_p^2}{2r_p} - \frac{1-\theta_q^2}{2r_q} \right)} \right| \\ &= \left| \frac{1}{N} \sum_{n=-(N-1)/2}^{(N-1)/2} e^{jn\pi(\theta_q - \theta_p) + jk_c n^2 d^2 \left(\frac{1-\theta_p^2}{2r_p} - \frac{1-\theta_q^2}{2r_q} \right)} \right|. \end{aligned} \quad (9)$$

It is still difficult to directly obtain the distance and angle sampling methods from (9). However, it can be observed that the phase of each item in the summation of (9) can be decoupled into two parts. The first part is the linear phase $n\pi(\theta_q - \theta_p)$ related to the angle, while the second part is the quadratic phase $jk_c n^2 d^2 \left(\frac{1-\theta_p^2}{2r_p} - \frac{1-\theta_q^2}{2r_q} \right)$ related to the angle and distance. Based on this observation, in the following two subsections, we will first derive the angular sampling method from the first linear phase part, and then derive the distance sampling method from the second quadratic phase part.

B. Angular Sampling Method

Firstly, to design the angular sampling method, we should focus on the first linear phase part $n\pi(\theta_q - \theta_p)$ only related to the angle. For arbitrary two locations (θ_p, r_p) and (θ_q, r_q) in the polar coordinates, if $\frac{1-\theta_p^2}{r_p}$ and $\frac{1-\theta_q^2}{r_q}$ are equal to a constant $\frac{1}{\phi}$, i.e., $\frac{1-\theta_p^2}{r_p} = \frac{1-\theta_q^2}{r_q} = \frac{1}{\phi}$, then it is obvious that the quadratic phase part $jk_c n^2 d^2 \left(\frac{1-\theta_p^2}{2r_p} - \frac{1-\theta_q^2}{2r_q} \right)$ becomes 0, and thus it can be removed. In this case, the column coherence $f(\theta_p, \theta_q, r_p, r_q)$ is dependent on the first linear phase part $n\pi(\theta_q - \theta_p)$, which only relies on the angles θ_p and θ_q , while the effect of distance r_p and r_q is removed. Moreover, note that $(\theta_p, r_p), (\theta_q, r_q)$ are two arbitrary locations satisfying $\frac{1-\theta_p^2}{r_p} = \frac{1-\theta_q^2}{r_q} = \frac{1}{\phi}$, which means they are sampled from the curve $\frac{1-\theta^2}{r} = \frac{1}{\phi}$. For expression simplicity, we name the curve $\frac{1-\theta^2}{r} = \frac{1}{\phi}$ as distance ring ϕ . As shown by the red curves in Fig. 5 (b), different constants ϕ correspond to different distance rings ϕ . Thus, if the locations are sampled on the distance ring ϕ , then the coherence $f(\theta_p, \theta_q, r_p, r_q)$ is only dependent on the angles θ_p and θ_q . Then, we can derive the following angular sampling method.

Specifically, since $\frac{1-\theta_p^2}{r_p} = \frac{1-\theta_q^2}{r_q}$, from (9), the coherence $f(\theta_p, \theta_q, r_p, r_q)$ can be expressed as

$$\begin{aligned} f(\theta_p, \theta_q, r_p, r_q) &= \left| \frac{1}{N} \sum_{n=-(N-1)/2}^{(N-1)/2} e^{jn\pi(\theta_q - \theta_p)} \right| \\ &= \left| \frac{\sin(\frac{1}{2}N\pi(\theta_q - \theta_p))}{N \sin(\frac{1}{2}\pi(\theta_q - \theta_p))} \right|, \end{aligned} \quad (10)$$

which is only related to the angles θ_p and θ_q . It can be found that (10) is exactly equivalent to the coherence between two far-field steering vectors [24]. The zero points of the function (10) satisfy $\theta_q - \theta_p = \frac{2m}{N}$, $m = 1, 2, \dots, N-1$. Therefore, the angular sampling method on distance ring ϕ is the same as the existing angular sampling method of the angular-domain transform matrix \mathbf{F} . In other words, angles should be uniformly sampled on distance ring ϕ as

$$\theta_n = \frac{2n-N+1}{N}, \quad n = 0, 1, \dots, N-1. \quad (11)$$

C. Distance Sampling Method

Similar to the derivation of the angular sampling method, since the distance information is contained in the second quadratic phase part $jk_c n^2 d^2 \left(\frac{1-\theta_p^2}{2r_p} - \frac{1-\theta_q^2}{2r_q} \right)$, we now focus on this part to derive the distance sampling method. For arbitrary two vectors sampled on the same angle θ , i.e., $\theta_p = \theta_q = \theta$, the first linear phase part $n\pi(\theta_q - \theta_p)$ becomes 0, thus being removed. In this case, the column coherence $f(\theta_p, \theta_q, r_p, r_q) = f(\theta, \theta, r_p, r_q)$ is dependent on the second quadratic phase part $jk_c n^2 d^2 \left(\frac{1-\theta_p^2}{2r_p} - \frac{1-\theta_q^2}{2r_q} \right) = k_c n^2 d^2 \left(\frac{1-\theta^2}{2r_p} - \frac{1-\theta^2}{2r_q} \right)$, which relies on the distance-related items $\frac{1-\theta^2}{r_p}$ and $\frac{1-\theta^2}{r_q}$. The physical significance behind the discussion above is that, if two locations are sampled on the same angle θ , the column coherence $f(\theta_p, \theta_q, r_p, r_q)$ is only decided by the distance-related items $\frac{1-\theta^2}{r_p}$ and $\frac{1-\theta^2}{r_q}$. Based on this observation, we can derive the distance sampling method on the angle θ .

Unfortunately, unlike the derivation of the angular sampling method, due to the quadratic phase property, it is difficult to get a close form like (10). To cope with problem, in the following **Lemma 1**, the Fresnel functions are introduced to approximate the column coherence $f(\theta, \theta, r_p, r_q)$.

Lemma 1: If two near-field steering vectors are sampled from the same angle θ but different distances r_p and r_q , then the column coherence $f(\theta, \theta, r_p, r_q)$ can be approximated as

$$f(\theta, \theta, r_p, r_q) \approx |G(\beta)| = \left| \frac{C(\beta) + jS(\beta)}{\beta} \right|, \quad (12)$$

where $\beta = \sqrt{\frac{N^2 d^2 (1-\theta^2)}{2\lambda_c} \left| \frac{1}{r_p} - \frac{1}{r_q} \right|}$, $C(\beta) = \int_0^\beta \cos(\frac{\pi}{2} t^2) dt$ and $S(\beta) = \int_0^\beta \sin(\frac{\pi}{2} t^2) dt$ are Fresnel functions [26].

Proof: See Appendix A. ■

Lemma 1 indicates that the column coherence heavily relies on the function $|G(\cdot)|$ and the parameter β . The function $|G(\cdot)|$ is composed of two Fresnel functions. Since the function $|G(\cdot)|$ does not contain any parameters, it is sufficient to obtain its numerical results through one numerical integration, and the result is shown in Fig. 6. With the increase of β , $|G(\beta)|$ shows a significant downward trend with slight fluctuation. Therefore, in order to make the column coherence as small as possible, i.e., to let $f(\theta, \theta, r_p, r_q) \approx |G(\beta)|$ lower than a desired threshold Δ , we should first calculate β_Δ satisfying $|G(\beta_\Delta)| = \Delta$. Then, due to the downward trend of the function $|G(\beta)|$, it can be approximately derived that $\beta \geq \beta_\Delta$.

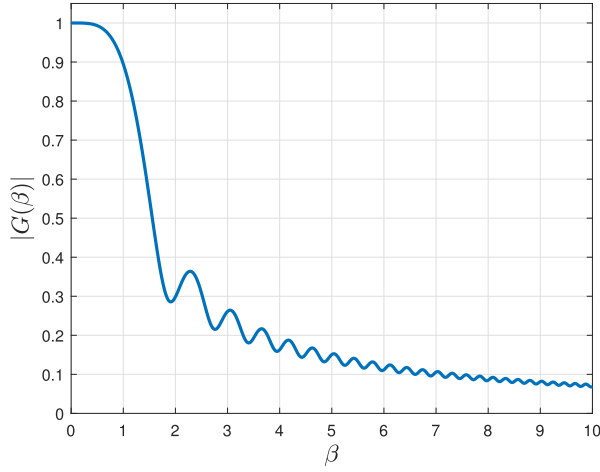


Fig. 6. The numerical results of $|G(\beta)|$ against β .

For example, if we desire that the column coherence is lower than $\Delta = 0.5$, then it can be solved from $|G(\beta_{0.5})| = 0.5$ that $\beta_{0.5} \approx 1.6$. Thus, we approximately have $\beta \geq 1.6$.

Based on the conditions $\beta \geq \beta_{\Delta}$ and $\beta = \sqrt{\frac{N^2 d^2 (1-\theta^2)}{2\lambda_c}} \left| \frac{1}{r_p} - \frac{1}{r_q} \right|$, the sampled distances r_p and r_q should satisfy

$$\left| \frac{1}{r_p} - \frac{1}{r_q} \right| \geq \frac{2\lambda_c \beta_{\Delta}^2}{N^2 d^2 (1-\theta^2)} = \frac{1}{Z_{\Delta} (1-\theta^2)}, \quad (13)$$

where $Z_{\Delta} = \frac{N^2 d^2}{2\lambda_c \beta_{\Delta}^2}$. By considering the array aperture $D = Nd$, we can rewrite Z_{Δ} as $Z_{\Delta} = \frac{D^2}{2\beta_{\Delta}^2 \lambda_c}$. We define Z_{Δ} as the threshold distance in this paper.

To make the column coherence lower than a given threshold, it is clear from (13) that the difference between the inverses of two distances should be larger than a constant. For example, if we set $r_q = \frac{1}{s} Z_{\Delta} (1-\theta^2)$, $s = 1, 2, 3, \dots$, then from (13) we have

$$r_p \geq \frac{1}{s-1} Z_{\Delta} (1-\theta^2) \quad \text{or} \quad r_p \leq \frac{1}{s+1} Z_{\Delta} (1-\theta^2). \quad (14)$$

It can be inferred from (14) that, on an arbitrary angle θ , if one steering vector is sampled from the distance $\frac{1}{s} Z_{\Delta} (1-\theta^2)$, then to keep the coherence lower than Δ , no other distances can be sampled in the range $\left[\frac{1}{s+1} Z_{\Delta} (1-\theta^2), \frac{1}{s-1} Z_{\Delta} (1-\theta^2) \right]$. In other words, $\frac{1}{s+1} Z_{\Delta} (1-\theta^2)$ is exactly the maximum feasible distance that is lower than $\frac{1}{s} Z_{\Delta} (1-\theta^2)$, while $\frac{1}{s-1} Z_{\Delta} (1-\theta^2)$ is exactly the minimum feasible distance that is higher than $\frac{1}{s} Z_{\Delta} (1-\theta^2)$.

Based on this observation, on the angle θ , if the sampled distances are

$$r_s = \frac{1}{s} Z_{\Delta} (1-\theta^2), \quad s = 0, 1, 2, 3, \dots, \quad (15)$$

then the column coherence of two near-field steering vectors sampled at two adjacent distances are exactly Δ . That is to say, the constraint that the column coherence is lower than Δ on any angle can be guaranteed. Comparing (11) and (15), it can be found that the angle should be sampled uniformly, while the distance should be sampled non-uniformly.

D. Design the Polar-Domain Transform Matrix

In this subsection, based on the angular and distance sampling methods derived in (11) and (15), we conclude how to generate the polar-domain transform matrix \mathbf{W} in **Algorithm 1**. As shown in Fig. 5 (b), the entire polar domain is divided into multiple distance rings r_s , while each distance ring is segmented by multiple angles. Therefore, the transform matrix \mathbf{W} can be composed S sub-matrices, i.e.,

$$\mathbf{W} = [\mathbf{W}_0, \mathbf{W}_2, \dots, \mathbf{W}_{S-1}] \in \mathbb{C}^{N \times NS}, \quad (16)$$

where S denotes the number of distance rings, and the sub-matrix $\mathbf{W}_s \in \mathbb{C}^{N \times N}$ consists of N near-field steering vectors sampled from the same curve $\frac{1-\theta^2}{r} = \frac{1}{r_s}$ (or distance ring r_s) but different angles. Therefore, the number of column vectors in \mathbf{W} is $Q = NS$.

Algorithm 1: The Generating Procedure of the Proposed Polar-Domain Transform Matrix \mathbf{W}

Require:

The minimum allowable distance ρ_{\min} ; threshold β_{Δ} ; antenna number N ; antenna spacing d ; wavelength λ_c

Ensure:

polar-domain transform matrix \mathbf{W}

- 1: $Z_{\Delta} = \frac{N^2 d^2}{2\beta_{\Delta}^2 \lambda_c}$
 - 2: $s = 0$
 - 3: **repeat**
 - 4: **for** $n \in \{0, 1, \dots, N-1\}$ **do**
 - 5: $\theta_n = \frac{2n-N+1}{N}$ according to (11)
 - 6: $r_{s,n} = \frac{1}{s} Z_{\Delta} (1-\theta_n^2)$ according to (15)
 - 7: **end for**
 - 8: $\mathbf{W}_s = [\mathbf{b}(\theta_0, r_{s,0}), \mathbf{b}(\theta_1, r_{s,1}), \dots, \mathbf{b}(\theta_{N-1}, r_{s,N-1})]$
 - 9: $S = s, \quad s = s + 1$
 - 10: **until** $\frac{1}{s} Z_{\Delta} < \rho_{\min}$
 - 11: $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_S]$
 - 12: **return** \mathbf{W} .
-

Specifically, in step 1 of **Algorithm 1**, the threshold distance Z_{Δ} is calculated. Next, in steps 3-10, different sub-matrices \mathbf{W}_s are sequentially generated. In steps 5-6, N angles and distances are sampled according to (11) and (15). After that, in step 8, the sub-matrix \mathbf{W}_s is generated based on the sampled angles and distances. It can be derived from steps 5-6 that θ_n and $r_{s,n}$ are sampled on the curve $\frac{1-\theta_n^2}{r_{s,n}} = \frac{1-\theta_n^2}{\frac{1}{s} Z_{\Delta} (1-\theta_n^2)} = \frac{s}{Z_{\Delta}}$, which is exactly equivalent to the distance ring $r_s = 1/\frac{s}{Z_{\Delta}} = \frac{1}{s} Z_{\Delta}$. Thus the s -th sub-matrices \mathbf{W}_s is sampled from the distance ring r_s . Furthermore, in step 10, since the actual distances between the BS and the users or scatters cannot be zero, we define ρ_{\min} as the minimum allowable distance ring. Then, only when the distance ring $r_s = \frac{1}{s} Z_{\Delta}$ is larger than ρ_{\min} can it be sampled, just as shown in step 10. Thus, the number of distances rings S is determined by ρ_{\min} . Finally, in step 11, all sub-matrices are concatenated to construct the polar-domain transform matrix \mathbf{W} .

It is worth noting that since the polar-domain transform matrix \mathbf{W} also samples distances in the far-field region, e.g., $s = 0$ (or $r_s = +\infty$) in (15), when the minimum

allowable distance ρ_{\min} is large enough, only the distances in the far-field will be sampled. Then, the polar-domain representation becomes the angular-domain representation, and therefore the angular-domain representation is a special case of the proposed polar-domain representation.

Based on the proposed polar-domain representation of the near-field XL-MIMO channel, we will propose the on-grid and off-grid near-field channel estimation schemes in the next section.

IV. PROPOSED NEAR-FIELD CHANNEL ESTIMATION SCHEMES

In this section, by exploiting the proposed polar-domain sparse representation of the near-field channel, we first propose an on-grid near-field channel estimation algorithm called polar-domain simultaneous orthogonal matching pursuit (P-SOMP) to efficiently estimate the near-field XL-MIMO channel. Then, an off-grid near-field channel estimation algorithm called polar-domain simultaneous iterative gridless weighted (P-SIGW) is further proposed to improve the channel estimation accuracy.

A. On-Grid Near-Field Channel Estimation

As we discussed in section II, since the K users transmit mutual orthogonal pilot sequences, uplink channel estimation for each user can be carried out independently. For an arbitrary user, based on the polar-domain representation (8), the received pilot \mathbf{y}_m at frequency f_m in (2) can be represented as

$$\mathbf{y}_m = \mathbf{A}\mathbf{W}\mathbf{h}_m^P + \mathbf{n}_m = \mathbf{\Psi}\mathbf{h}_m^P + \mathbf{n}_m, \quad (17)$$

where $\mathbf{\Psi} = \mathbf{A}\mathbf{W}$. Under the CS framework [27], each element of \mathbf{A} can be randomly selected from $\frac{1}{\sqrt{N}}\{-1, +1\}$ with equal probability. Since the polar-domain channel \mathbf{h}_m^P is sparse as discussed in Section III, the polar-domain channel estimation can be formulated as a sparse signal recovery problem.

However, since the received noise $\mathbf{n}_m = [\mathbf{n}_{m,1}^T \mathbf{A}_1^T, \dots, \mathbf{n}_{m,P}^T \mathbf{A}_P^T]^T$ is colored noise, a pre-whitening procedure should be carried out at first [8]. To be specific, the covariance matrix of the noise is $\mathbf{C} = \mathbb{E}(\mathbf{n}_m \mathbf{n}_m^H) = \text{blkdiag}(\sigma^2 \mathbf{A}_1 \mathbf{A}_1^H, \sigma^2 \mathbf{A}_2 \mathbf{A}_2^H, \dots, \sigma^2 \mathbf{A}_P \mathbf{A}_P^H)$. Then, this covariance matrix can be decomposed by Cholesky factorization as $\mathbf{C} = \sigma^2 \mathbf{D}\mathbf{D}^H$, where $\mathbf{D} \in \mathbb{C}^{PN_{\text{RF}} \times PN_{\text{RF}}}$ is a lower triangular matrix. Thus the pre-whitening matrix is \mathbf{D}^{-1} , and then the whitened received signal $\bar{\mathbf{y}}_m$ at frequency f_m is

$$\bar{\mathbf{y}}_m = \mathbf{D}^{-1} \mathbf{y}_m = \bar{\mathbf{\Psi}} \mathbf{h}_m^P + \bar{\mathbf{n}}_m, \quad (18)$$

where $\bar{\mathbf{\Psi}} = \mathbf{D}^{-1} \mathbf{A}\mathbf{W}$ and $\bar{\mathbf{n}}_m = \mathbf{D}^{-1} \mathbf{n}_m$. In this case, the covariance matrix of $\bar{\mathbf{n}}_m$ is $\bar{\mathbf{C}} = \mathbf{D}^{-1} \mathbf{C} \mathbf{D}^{-H} = \sigma^2 \mathbf{I}_{PN_{\text{RF}}}$, thus the noise $\bar{\mathbf{n}}_m$ becomes white.

Generally, the steering vectors at different sub-carriers are the same [10], just as the frequency-independent steering vector $\mathbf{b}(\cdot)$ and $\mathbf{a}(\cdot)$. Therefore, the sparsity support of the polar-domain channels \mathbf{h}_m^P at different subcarriers f_m are also

the same, and they can be simultaneously estimated to increase the estimation accuracy. Therefore, we rearrange (18) as

$$\bar{\mathbf{Y}} = \mathbf{D}^{-1} \mathbf{Y} = \bar{\mathbf{\Psi}} \mathbf{H}^P + \bar{\mathbf{N}}, \quad (19)$$

where $\bar{\mathbf{Y}} = [\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \dots, \bar{\mathbf{y}}_M]$, $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_M]$, $\mathbf{H}^P = [\mathbf{h}_1^P, \dots, \mathbf{h}_M^P]$, and $\bar{\mathbf{N}} = [\bar{\mathbf{n}}_1, \dots, \bar{\mathbf{n}}_M]$. The target is to estimate the channel $\mathbf{H} = \mathbf{W}\mathbf{H}^P$ from $\bar{\mathbf{\Psi}}$ and $\bar{\mathbf{Y}}$. Leveraging the channel sparsity in the polar domain, the row of \mathbf{H}^P is sparse. Therefore, the channel estimation problem can be solved by the existing simultaneous orthogonal matching pursuit (SOMP) algorithm [8].

In this paper, we extend the classical angular-domain SOMP algorithm to a polar-domain SOMP (P-SOMP) algorithm to recover the near-field XL-MIMO channel. The proposed P-SOMP algorithm is illustrated in **Algorithm 2**.

Algorithm 2: The Proposed Polar-Domain SOMP Algorithm

Require:

Received pilot \mathbf{Y} ; combining matrix \mathbf{A} ; the minimum distance ρ_{\min} ; number of paths \hat{L}

Ensure:

The estimated near-field channel $\hat{\mathbf{H}}$

- 1: Construct the polar-domain transform matrix \mathbf{W} by **Algorithm 1**
 - 2: Covariance matrix $\mathbf{C} = \text{blkdiag}(\mathbf{A}_1 \mathbf{A}_1^H, \dots, \mathbf{A}_P \mathbf{A}_P^H)$
 - 3: Calculate the pre-whitening matrix \mathbf{D} by solving $\mathbf{C} = \mathbf{D}\mathbf{D}^H$
 - 4: Pre-whitening: $\bar{\mathbf{Y}} = \mathbf{D}^{-1} \mathbf{Y}$, $\bar{\mathbf{\Psi}} = \mathbf{D}^{-1} \mathbf{A}\mathbf{W}$
 - 5: Initialization: $\mathbf{R} = \bar{\mathbf{Y}}$, $\mathcal{Y} = \{\emptyset\}$
 - 6: **for** $l \in \{1, 2, \dots, \hat{L}\}$ **do**
 - 7: Calculate the correlation matrix: $\mathbf{\Gamma} = \bar{\mathbf{\Psi}}^H \mathbf{R}$
 - 8: Detect new support: $p^* = \arg \max_p \sum_{m=1}^M |\Gamma(p, m)|^2$
 - 9: Update support set: $\mathcal{Y} = \mathcal{Y} \cup p^*$
 - 10: Orthogonal projection: $\hat{\mathbf{H}}_{\mathcal{Y},:}^P = \bar{\mathbf{\Psi}}_{\mathcal{Y},:}^\dagger \bar{\mathbf{Y}}$
 - 11: Update residual: $\mathbf{R} = \mathbf{R} - \bar{\mathbf{\Psi}}_{\mathcal{Y},:} \hat{\mathbf{H}}_{\mathcal{Y},:}^P$
 - 12: **end for**
 - 13: $\hat{\mathbf{H}} = \mathbf{W}_{:, \mathcal{Y}} \hat{\mathbf{H}}_{\mathcal{Y},:}^P$
 - 14: **return** $\hat{\mathbf{H}}$.
-

Specifically, in step 1, we first construct the polar-domain transform matrix \mathbf{W} according to **Algorithm 1**. Next in steps 2-4, the pre-whitening procedure is carried out to whiten the received signal. Then, in steps 5-12, we utilize the SOMP algorithm to successively estimate the physical channel angle and distance in the polar domain. In step 5, we initialize the residual matrix $\mathbf{R} = \bar{\mathbf{Y}}$ and the sparse support set $\mathcal{Y} = \{\emptyset\}$. Then, for the l -th path component, we first calculate the correlation matrix $\mathbf{\Gamma} = \bar{\mathbf{\Psi}}^H \mathbf{R}$ in step 7. Next in steps 8, based on the assumption that the support sets at different subcarriers are the same, the power of the correlation matrix on the p -th row is $\sum_{m=1}^M |\Gamma(p, m)|^2$. Thus, the index p^* of the physical location of the l -th path component can be determined as $p^* = \arg \max_p \sum_{m=1}^M |\Gamma(p, m)|^2$. Then, we add p^* to the sparse support set \mathcal{Y} . After that, in step 10, the path gain $\hat{\mathbf{H}}_{\mathcal{Y},:}^P$ on the support set \mathcal{Y} is calculated through orthogonal least square, and in step 11, we update the residual matrix \mathbf{R} . The steps above are carried out for \hat{L} times until all path

components are detected. Finally, the near-field channel $\hat{\mathbf{H}}$ is recovered as $\hat{\mathbf{H}} = \mathbf{W}_{:,T} \hat{\mathbf{H}}_{T,:}^P$.

The main difference between the proposed P-SOMP algorithm and the existing SOMP algorithm is that, the proposed P-SOMP algorithm is carried out in the polar domain, so it can efficiently estimate the near-field XL-MIMO channel. Moreover, since the polar-domain transform matrix \mathbf{W} also samples distances in the far-field region, e.g., $s = 0$ in (15), the P-SOMP algorithm also works well in the far-field, which will be verified by simulations in Section V.

However, the proposed P-SOMP algorithm assumes that the angles and distances exactly lie in the sampled points in the polar domain, i.e., on-grid angles and distances. In contrast, the actual angles and distances are continuously distributed, i.e., off-grid angles and distances. Then, the estimation accuracy of the proposed on-grid P-SOMP algorithm is limited, which will be improved in the next Subsection IV-B.

B. Off-Grid Near-Field Channel Estimation

To cope with the estimation error introduced by the on-grid sample points, inspired by the classical off-grid simultaneous gridless weighted (SIGW) algorithm in the angular-domain, we propose an off-grid near-field channel estimation algorithm called polar-domain simultaneous gridless weighted (P-SIGW) to improve the channel estimation performance. Unlike the existing SIGW algorithm that only refines the estimated path gains and angles, the proposed P-SIGW algorithm simultaneously refines the path gains, angles, and distances by following the maximum likelihood principle. Specifically, the proposed P-SIGW algorithm is provided in **Algorithm 3**.

Algorithm 3: The Proposed Polar-Domain SIGW Algorithm

Require:

Received pilot sequences \mathbf{Y} ; combining matrix \mathbf{A} ; the minimum distance ρ_{\min} ; number of detected paths \hat{L} , number of iterations N_{iter}

Ensure:

The estimated near-field channel $\hat{\mathbf{H}}$

Initialization stage

- 1: Obtain the initial value of the distances $\hat{\mathbf{r}}^0 = [\hat{r}_1^0, \hat{r}_2^0, \dots, \hat{r}_{\hat{L}}^0]$ and the angles $\hat{\boldsymbol{\theta}}^0 = [\hat{\theta}_1^0, \hat{\theta}_2^0, \dots, \hat{\theta}_{\hat{L}}^0]$ by **Algorithm 2**.

Refinement stage

- 2: **for** $n \in \{1, 2, \dots, N_{\text{iter}}\}$ **do**
 - 3: Choose Armijo backtracking line search step length l_1
 - 4: Update the angles by $\hat{\boldsymbol{\theta}}^n = \hat{\boldsymbol{\theta}}^{n-1} - l_1 \nabla_{\hat{\boldsymbol{\theta}}} \mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}^{n-1})|_{\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}^{n-1}}$ by (23)
 - 5: Choose Armijo backtracking line search step length l_2
 - 6: Update the distances by $\frac{1}{\hat{\mathbf{r}}}^n = \frac{1}{\hat{\mathbf{r}}^{n-1}} - l_2 \nabla_{\frac{1}{\hat{\mathbf{r}}}} \mathcal{L}(\hat{\boldsymbol{\theta}}^n, \hat{\mathbf{r}})|_{\hat{\mathbf{r}}=\hat{\mathbf{r}}^{n-1}}$ by (24)
 - 7: Update the path gains $\hat{\mathbf{G}}^n$ by (21)
 - 8: **end for**
 - 9: $\hat{\mathbf{H}} = [\mathbf{b}(\hat{\theta}_1^n, \hat{r}_1^n), \mathbf{b}(\hat{\theta}_2^n, \hat{r}_2^n), \dots, \mathbf{b}(\hat{\theta}_{\hat{L}}^n, \hat{r}_{\hat{L}}^n)] \hat{\mathbf{G}}^n$
 - 10: **return** $\hat{\mathbf{H}}$.
-

The proposed P-SIGW algorithm is composed of an initialization stage and a refinement stage. Firstly, in step 1,

we regard the P-SOMP algorithm as an initialization stage of the P-SIGW algorithm. After carrying out **Algorithm 2**, we obtain the initial value of the estimated distances $\hat{\mathbf{r}} = [\hat{r}_1, \hat{r}_2, \dots, \hat{r}_{\hat{L}}]$, angles $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_{\hat{L}}]$, and complex path gains $\hat{\mathbf{G}} = \hat{\mathbf{H}}_{T,:}^P \in \mathbb{C}^{\hat{L} \times M}$.

Then, in the refinement stage, we concatenate the detected near-field paths to construct the matrix $\tilde{\mathbf{W}}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) = [\mathbf{b}(\hat{\theta}_1, \hat{r}_1), \mathbf{b}(\hat{\theta}_2, \hat{r}_2), \dots, \mathbf{b}(\hat{\theta}_{\hat{L}}, \hat{r}_{\hat{L}})]$, thus the recovered channel can be written as $\hat{\mathbf{H}} = \tilde{\mathbf{W}}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \hat{\mathbf{G}}$. In steps 2-7, the distances $\hat{\mathbf{r}}$, angles $\hat{\boldsymbol{\theta}}$, and path gains $\hat{\mathbf{G}}$ are alternatively optimized to maximize the likelihood, which is formulated as

$$\min_{\hat{\mathbf{G}}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}} \|\bar{\mathbf{Y}} - \tilde{\Psi}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \hat{\mathbf{G}}\|_F^2, \quad (20)$$

where $\tilde{\Psi}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) = \mathbf{D}^{-1} \mathbf{A} \tilde{\mathbf{W}}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$. Since the optimization problem (20) is non-convex, we utilize the alternating minimization method to solve this problem. For fixed $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$, the optimal solution for $\hat{\mathbf{G}}$ is given by

$$\hat{\mathbf{G}}^{\text{opt}} = \tilde{\Psi}^\dagger(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \bar{\mathbf{Y}}. \quad (21)$$

Substitute (21) into (20), the maximum-likelihood problem is reformulated as

$$\begin{aligned} & \min_{\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}} \|\bar{\mathbf{Y}} - \tilde{\Psi}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \tilde{\Psi}^\dagger(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \bar{\mathbf{Y}}\|_F^2 \\ & \Leftrightarrow \min_{\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}} \text{Tr} \left\{ \bar{\mathbf{Y}}^H \left(\mathbf{I} - \mathbf{P}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \right)^H \left(\mathbf{I} - \mathbf{P}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \right) \bar{\mathbf{Y}} \right\} \\ & \stackrel{(a)}{\Leftrightarrow} \min_{\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}} \mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) = -\text{Tr} \left\{ \bar{\mathbf{Y}}^H \mathbf{P}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \bar{\mathbf{Y}} \right\}, \end{aligned} \quad (22)$$

where $\mathbf{P}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) = \tilde{\Psi}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \tilde{\Psi}^\dagger(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$, and (a) is derived by $\mathbf{P}^H(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) \mathbf{P}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) = \mathbf{P}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$. The new objective function $\mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$ can be optimized using an iterative gradient descent approach. In the n -th iteration, the angles are updated as

$$\hat{\boldsymbol{\theta}}^n = \hat{\boldsymbol{\theta}}^{n-1} - l_1 \nabla_{\hat{\boldsymbol{\theta}}} \mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}^{n-1})|_{\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}^{n-1}}, \quad (23)$$

where l_1 denotes the step length for the angles. Moreover, as for the distance, we found that if the gradient with respect to $\hat{\mathbf{r}}$ is directly utilized to update $\hat{\mathbf{r}}$, the estimation performance fluctuates with distance. This fact can be explained by the distance-sampling method (15) that r is non-uniformly sampled, or in other words, $\frac{1}{r}$ is uniformly sampled. Motivated by this observation, we define $\frac{1}{\hat{\mathbf{r}}} = [\frac{1}{\hat{r}_1}, \frac{1}{\hat{r}_2}, \dots, \frac{1}{\hat{r}_{\hat{L}}}]$, and utilize the gradient with respect to $\frac{1}{\hat{\mathbf{r}}}$ to indirectly update $\hat{\mathbf{r}}$, i.e.,

$$\frac{1}{\hat{\mathbf{r}}}^n = \frac{1}{\hat{\mathbf{r}}^{n-1}} - l_2 \nabla_{\frac{1}{\hat{\mathbf{r}}}} \mathcal{L}(\hat{\boldsymbol{\theta}}^n, \hat{\mathbf{r}})|_{\hat{\mathbf{r}}=\hat{\mathbf{r}}^{n-1}}, \quad (24)$$

where l_2 denotes the step length for the inverse of distances. To guarantee the objective function is non-increasing, the step lengths l_1 and l_2 are chosen by Armijo backtracking line search. The gradient $\nabla \mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$ is given in Appendix B. Based on (21), (23), and (24), the parameters are updated in steps 3-7 in **Algorithm 3**. Finally, all refined paths are concatenated to reconstruct the near-field channel $\hat{\mathbf{H}}$. In section V, simulation results will be provided to verify the effectiveness of the proposed on-grid P-SOMP algorithm and off-grid P-SIGW algorithm.

TABLE I
COMPUTATIONAL COMPLEXITY COMPARISON

Algorithm	Computational complexity $\mathcal{O}(\cdot)$
P-SOMP	$\mathcal{O}(\hat{L}PN_{\text{RF}}NSM)$
SWOMP	$\mathcal{O}(\hat{L}PN_{\text{RF}}NM)$
P-SIGW	$\mathcal{O}(\hat{L}PN_{\text{RF}}NSM) + \mathcal{O}(N_{\text{iter}}(P^2N_{\text{RF}}^2M + PN_{\text{RF}}M^2))$
SS-SIGW-OLS	$\mathcal{O}(\hat{L}PN_{\text{RF}}NM) + \mathcal{O}(N_{\text{iter}}(P^2N_{\text{RF}}^2M + PN_{\text{RF}}M^2))$

C. Complexity and Convergence Analysis

1) *Convergence*: The convergence of the proposed optimization algorithm P-SIGM is discussed here. Since the objective function $\|\mathbf{Y} - \hat{\Psi}(\hat{\theta}, \hat{\mathbf{r}})\hat{\mathbf{G}}\|_F^2$ is larger than 0, it has a lower-bound. Then in each iteration, $\hat{\mathbf{G}}^{\text{opt}}$ is the optimal solution to update $\hat{\mathbf{G}}$. Moreover, for the steps of updating $\hat{\theta}$ and $\hat{\mathbf{r}}$, their step lengths l_1 and l_2 are chosen by Armijo backtracking line search. Therefore, the steps for updating $\hat{\mathbf{G}}$, $\hat{\theta}$, and $\hat{\mathbf{r}}$ are monotonically non-increasing, and the alternating procedure will converge.²

2) *Complexity*: The computational complexities of the proposed P-SOMP and P-SIGW schemes are analyzed, which are summarized in Table I. For **Algorithm 2**, the overall complexity is mainly determined by operations of the SOMP procedure, i.e., the steps 6-12 in **Algorithm 2**. Thus we only consider these steps. Since $\hat{\Psi} \in \mathbb{C}^{PN_{\text{RF}} \times NS}$, $\mathbf{R} \in \mathbb{C}^{PN_{\text{RF}} \times M}$, and $\mathbf{Y} \in \mathbb{C}^{PN_{\text{RF}} \times M}$, the computation complexities from step 7-11 are $\mathcal{O}(PN_{\text{RF}}NSM)$, $\mathcal{O}(NSM)$, $\mathcal{O}(1)$, $\mathcal{O}(\hat{L}^2PN_{\text{RF}} + \hat{L}PN_{\text{RF}}M)$, and $\mathcal{O}(\hat{L}PN_{\text{RF}}M)$, respectively. Generally, the number of paths \hat{L} is small, thus the computation complexity from steps 7-11 is decided by $\mathcal{O}(PN_{\text{RF}}NSM)$. After \hat{L} times iterations, the overall complexity is $\mathcal{O}(\hat{L}PN_{\text{RF}}NSM)$. As shown in Table I, the computation complexity of the proposed P-SOMP is S times that of the far-field on-grid scheme SWOMP [8]. However, the number of sampled distance rings S is generally very small. For instance, in our simulations, S is chosen from 4 to 6. Thus, this computation complexity is acceptable.

For **Algorithm 3**, its complexity is determined by the initialization stage and the refinement stage. The complexity in the initialization stage is the same as that of **Algorithm 2**, i.e., $\mathcal{O}(\hat{L}PN_{\text{RF}}NSM)$. Moreover, the complexity in the refinement stage is mainly introduced by the updates of the variables $\hat{\theta}$, $\hat{\mathbf{r}}$, and $\hat{\mathbf{G}}$. As shown in (21), the complexity to update $\hat{\mathbf{G}}$ is $\mathcal{O}(\hat{L}^2PN_{\text{RF}} + \hat{L}PN_{\text{RF}}M)$. As shown in Appendix B, the complexities of (29), (30), (31), and (32) to calculate the gradients with respect to $\hat{\theta}$ or $\hat{\mathbf{r}}$ are $\mathcal{O}(P^2N_{\text{RF}}^2M + PN_{\text{RF}}M^2)$, $\mathcal{O}(\hat{L}^2PN_{\text{RF}} + \hat{L}(PN_{\text{RF}})^2)$, $\mathcal{O}(\hat{L}^3 + \hat{L}^2PN_{\text{RF}})$, and $\mathcal{O}(PN_{\text{RF}}N)$. Since \hat{L} is very small, the complexity for each iteration in the refinement stage is determined by $\mathcal{O}(P^2N_{\text{RF}}^2M + PN_{\text{RF}}M^2)$. After N_{iter} times iterations, the complexity for refinement becomes $\mathcal{O}(N_{\text{iter}}(P^2N_{\text{RF}}^2M + PN_{\text{RF}}M^2))$. In conclusion, the overall complexity is $\mathcal{O}(\hat{L}PN_{\text{RF}}NSM + N_{\text{iter}}(P^2N_{\text{RF}}^2M + PN_{\text{RF}}M^2))$. It can be observed from Table I that, the complexity for refinement of the proposed P-SIGW scheme is the

²Notice that the proposed optimization algorithm converges to a feasible solution, while only the convergence rate is fully analyzed can its local optimality be strictly proved [28], [29].

TABLE II
SIMULATION CONFIGURATIONS

The number of BS antennas N	256
The number of Users K	4
The number of RF chains N_{RF}	4
Carrier frequency f_c	100 GHz
Bandwidth B	100 MHz
The number of subcarriers M	256
The minimum allowable distance ρ_{\min}	3 meters
The number of channel paths L per user	6
The distribution of θ	$\mathcal{U}\left(-\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}\right)$
Signal-to-noise ratio SNR	$1/\sigma^2$
Parameter β_{Δ}	1.2
The number of iterations N_{iter}	10
The number of paths to be detected \hat{L}	12

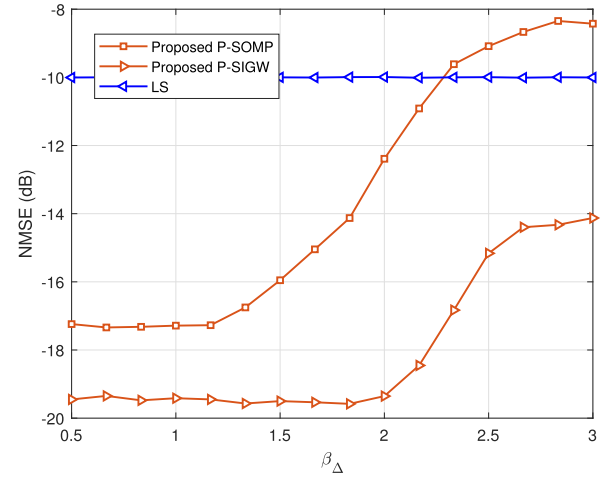


Fig. 7. NMSE performance against the parameter β_{Δ} .

same as that of the far-field off-grid SS-SIGW-OLS [17]. The complexity difference between these two algorithms is introduced by their initialization stages, which has been discussed before.

V. SIMULATION RESULTS

In this section, simulation results are provided to verify the performance of the proposed near-field channel estimation schemes. The performance is evaluated by the normalized mean square error (NMSE), which is defined as $\text{NMSE} = \mathbb{E}\left(\frac{\|\mathbf{H} - \hat{\mathbf{H}}\|_2^2}{\|\mathbf{H}\|_2^2}\right)$. A multi-user XL-MIMO OFDM system is considered, and the simulation configurations are shown in Table II.

Firstly, to determine the value of parameter β_{Δ} , the algorithm performance with respect to β_{Δ} is evaluated in Fig. 7. The distances between the BS and the users or scatters are chosen from $\mathcal{U}(5 \text{ m}, 10 \text{ m})$, the pilot length is 32, and the SNR is 10 dB. The proposed on-grid scheme will not suffer from a severe performance loss if $\beta_{\Delta} < 1.2$, while the proposed off-grid scheme will not suffer from a severe performance loss if $\beta_{\Delta} < 2$. Therefore, to guarantee the performance of the proposed two algorithms, we set $\beta_{\Delta} = 1.2$ in the following simulations. Under such settings, the number of sampled distances on each angle is $S = 6$.

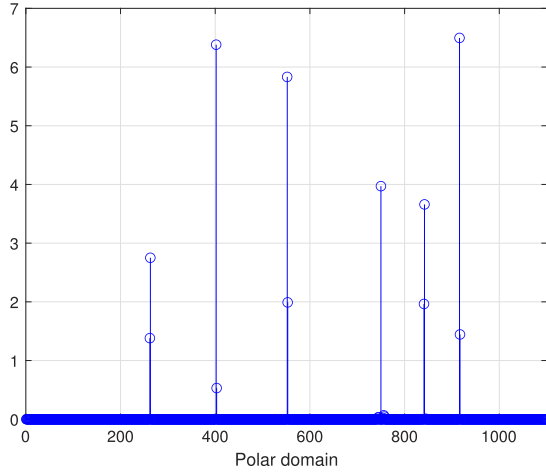


Fig. 8. The proposed polar domain channel representation.

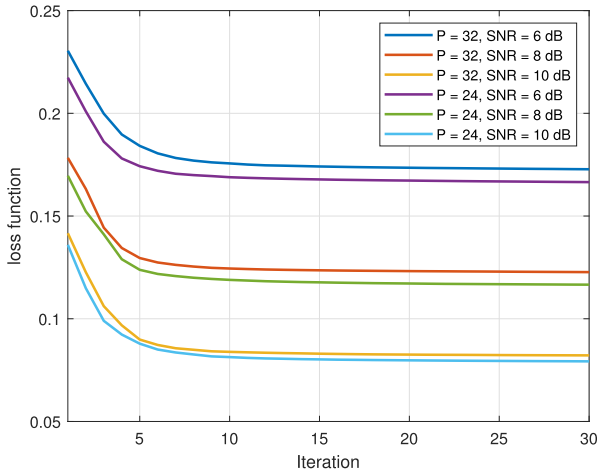


Fig. 9. Objective function with respect to the number of iterations.

In Fig. 8, the polar-domain channel in near-field is plotted to show the polar-domain sparsity. The distances between the BS and the users or scatters are randomly chosen from $\mathcal{U}(5 \text{ m}, 10 \text{ m})$. Since the angular-domain transform matrix $\mathbf{F} \in \mathbb{C}^{N \times N}$ is an orthogonal matrix, the angular-domain channel \mathbf{h}_m^A can be obtained as $\mathbf{h}_m^A = \mathbf{F}^H \mathbf{h}_m$. However, as for the polar-domain transform matrix $\mathbf{W} \in \mathbb{C}^{N \times NS}$, since the number of columns NS is larger than N , the polar-domain channel \mathbf{h}_m^P can not be directly obtained by $\mathbf{h}_m^P = \mathbf{W}^H \mathbf{h}_m$. To explicitly illustrate \mathbf{h}_m^P , we utilize compressed sensing methods to obtain $\hat{\mathbf{h}}_m^P$ from \mathbf{h}_m by solving $\mathbf{h}_m = \mathbf{W} \mathbf{h}_m^P$, which is plotted in Fig. 8. In our simulations, we found that the NMSE between $\hat{\mathbf{h}}_m = \mathbf{W} \hat{\mathbf{h}}_m^P$ and \mathbf{h}_m is lower than -20 dB , so we can assume $\hat{\mathbf{h}}_m^P \approx \mathbf{h}_m^P$. The number of paths L is 6, and there are exactly 6 peaks in the polar-domain channel, which shows obvious sparsity.

The convergence behavior of the proposed P-SIGW algorithm is shown in Fig. 9. The objective function $\|\bar{\mathbf{Y}} - \tilde{\Psi}(\hat{\theta}, \hat{\mathbf{r}})\|_F^2$ decreases monotonically over iteration. Under different parameters, the algorithm convergence is guaranteed. The simulation results are consistent with the convergence analysis in section IV-C. Taking into account the algorithm

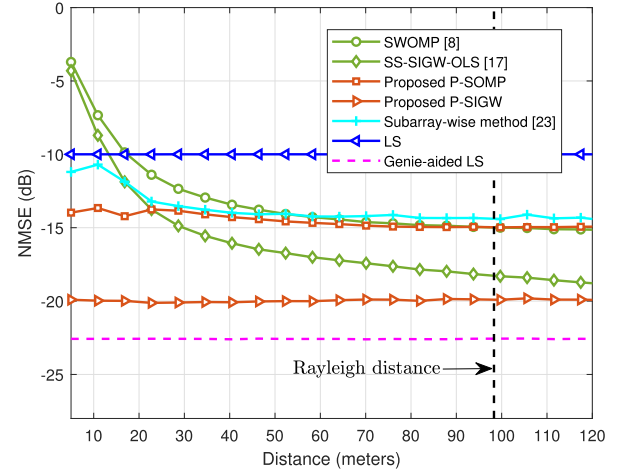


Fig. 10. NMSE performance comparison against the distance.

performance and complexity, we set the maximum iterations N_{iter} as 10 in the following simulations.

The NMSE performance with respect to distance is evaluated in Fig. 10. We compare the proposed on-grid polar-domain P-SOMP algorithm and the off-grid polar-domain P-SIGW algorithm with the existing methods, including the on-grid angular-domain SW-OMP algorithm [8], the off-grid angular-domain SS-SIGW-OLS algorithm [17], the subarray-wise near-field localization method proposed in [23], and the LS method. Moreover, the Genie-aided LS method, where the true distances and angles of the receivers and scatters are assumed to be available, is also compared as the NMSE performance bound. The SNR is 10 dB, and the pilot length is $P = 32$, i.e., the compressive ratio is $\frac{P}{N} = \frac{1}{2}$. The distances between the BS and the users or scatters are increasing from 3 meters to 120 meters. Since the number of antennas is $N = 256$, and the frequency is $f_c = 100 \text{ GHz}$, then the Rayleigh distance is around 100 meters. To highlight the impact of the near-field, we ignore the large-scale path loss in the channel. It can be observed from Fig. 10 that as the distance decreases, all of the far-field algorithms will suffer severe performance degradation, especially when the distance is lower than the Rayleigh distance. As for the subarray-wise near-field localization method, this method is equivalent to uniformly sampling multiple distances from the polar-domain, its NMSE performance is not robust to distance. By contrast, the proposed P-SOMP and P-SIGW algorithms outperform existing methods, and are robust to the small distance. Moreover, since the P-SIGW algorithm refines the estimated channel parameters with a much higher resolution, its NMSE performance is much better than the P-SOMP algorithm.

Then, Fig. 11 compares NMSE performance against SNR, where the length of pilot is $P = 32$. In Fig. 11 (a), the distances between the BS and the users or scatters are randomly sampled from $\mathcal{U}(5 \text{ m}, 10 \text{ m})$. In Fig. 11 (b), the corresponding distances are randomly sampled from $\mathcal{U}(100 \text{ m}, 120 \text{ m})$. It can be observed from Fig. 11 (a) that when the distance is small, the proposed near-field channel estimation schemes

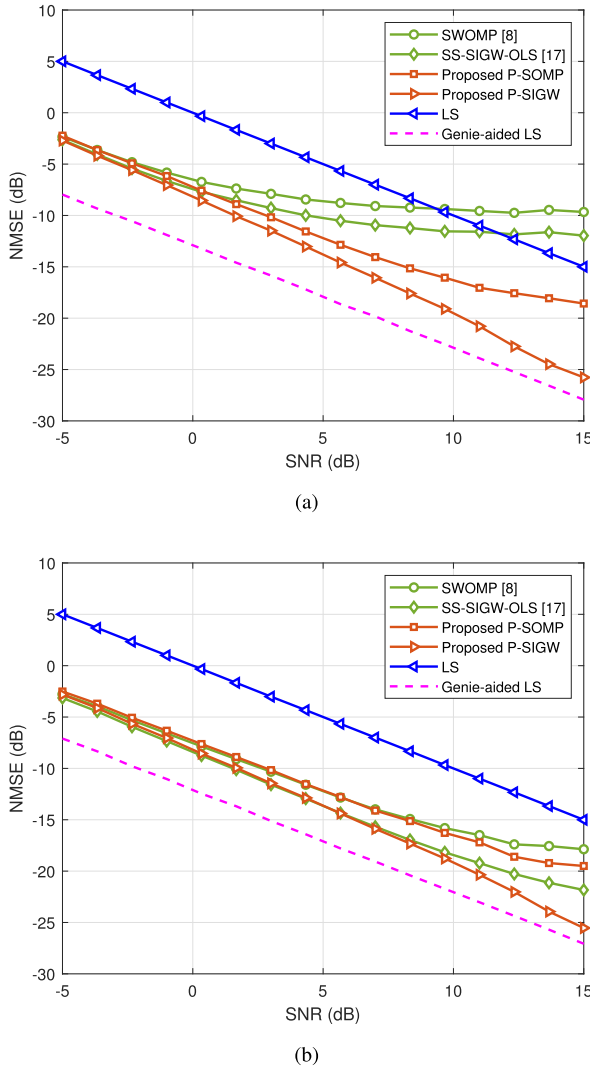


Fig. 11. The NMSE performance comparison against the SNR, where the distances between BS and the users or scatters are chosen from $\mathcal{U}(5\text{m}, 10\text{m})$ in (a), and the distances are chosen from $\mathcal{U}(100\text{m}, 120\text{m})$ in (b).

significantly outperform existing far-field channel estimation algorithms at all considered SNR. For far-field scenario in Fig. 11 (b), the proposed near-field channel estimation schemes can achieve the similar NMSE performance compared with the existing angular-domain based algorithms. The reason is that, the designed polar-domain transform matrix \mathbf{W} also samples distances in the far-field, e.g. $s = 0$ in (15), so the polar-domain transform matrix can also extract the far-field information. Moreover, since the Fresnel approximation is more accurate than the far-field approximation, when the SNR is high (>10 dB), the proposed methods outperform the far-field methods, even for far-field channels, and approaches the performance bound achieved by Genie-aided LS scheme. In conclusion, the proposed near-field channel estimation algorithms can accurately recover the channel in both the near-field and far-field, while the existing far-field channel estimation algorithms can only accurately recover the far-field channel.

Fig. 12 provides the NMSE performance against the pilot length P , where the SNR is 10 dB. In Fig. 11 (a), the distances

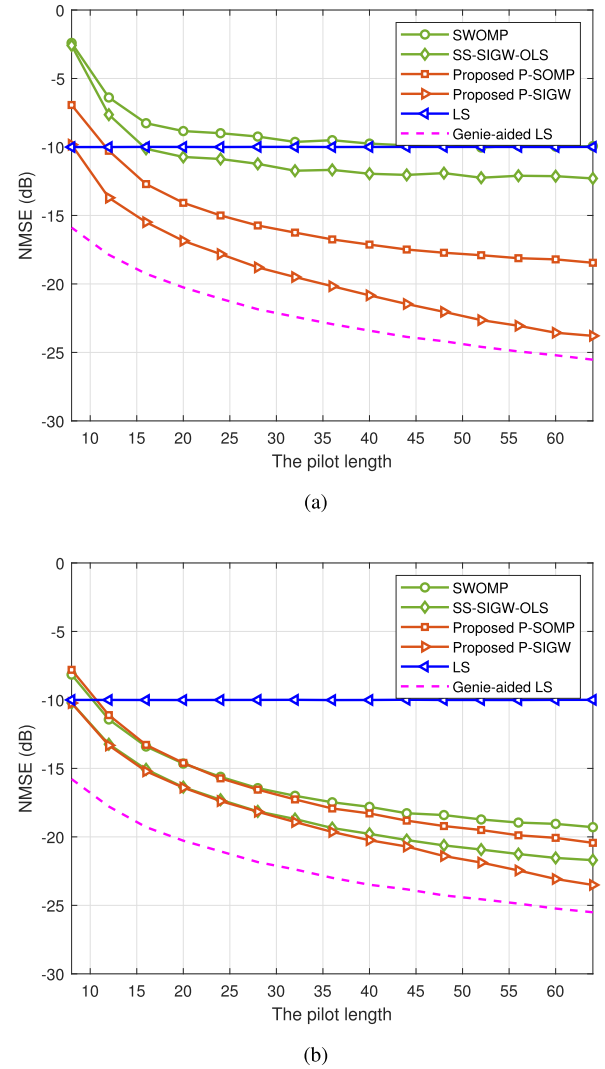


Fig. 12. The NMSE performance against the length of pilot, where the distances between the BS and the users or the scatters are chosen from $\mathcal{U}(5\text{m}, 10\text{m})$ in (a), and the distances are chosen from $\mathcal{U}(100\text{m}, 120\text{m})$ in (b).

are randomly sampled from $\mathcal{U}(5\text{m}, 10\text{m})$, while in Fig. 11 (b), the distances are randomly sampled from $\mathcal{U}(100\text{m}, 120\text{m})$. The length of pilot sequence P is increasing from 8 to 64, so that the compressive ratio $\frac{PN_{\text{RE}}}{N}$ is increasing from $\frac{1}{8}$ to 1. It can be observed from Fig. 12 that the NMSE performance achieved by all considered schemes improves as the pilot length P becomes longer. As shown in Fig. 12 (a), when the distance is small, the proposed P-SOMP and P-SIGW algorithms significantly outperform other angular-domain based algorithms. For example, when the pilot length is $P = 24$ or $P = 32$, the performance gap between the proposed P-SIGW algorithm and the existing algorithms is quite large. This indicates that the proposed algorithms can reduce the overhead for near-field XL-MIMO channel estimation. Moreover, for far-field scenario in Fig. 12 (b), when the pilot length P is shorter than 32, the NMSE achieved by the proposed near-field channel estimation schemes and existing far-field channel estimation schemes are similar, for both the on-grid and off-grid conditions. When the pilot length P is longer than 32,

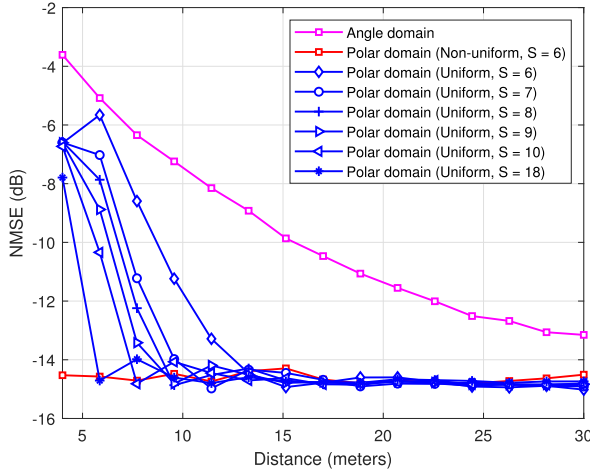


Fig. 13. NMSE performance comparison between the proposed non-uniform distance-sampling method and the uniform distance-sampling method.

the NMSE performance of the proposed algorithms slightly outperform the far-field algorithms. This is because the Fresnel approximation adopted in the polar-domain transform is more accurate than the far-field approximation. In conclusion, the proposed near-field channel estimation schemes can accurately recover the channel both in the near-field and far-field with low pilot overhead.

Finally, to verify the conclusion proved in (15) that the distances should be sampled non-uniformly, we evaluate the NMSE performance achieved by the P-SOMP method with uniform distance-sampling method in Fig. 13. For the uniform distance-sampling method, on each angle, S distances are uniformly sampled from the predefined range $[\rho_{\min}, \rho_{\max}]$, i.e., $r_s = \rho_{\min} + \frac{s}{S}(\rho_{\max} - \rho_{\min})$, $s = 0, 1, \dots, S-1$, where we set ρ_{\min} as 3 meters and ρ_{\max} as the Rayleigh distance. As shown in Fig. 13, although the estimation performance can be improved to some degree by uniformly sampling multiple distances in the polar domain, the NMSE performance is severely degraded when the distance is small, even if the number of sampled distances S on each angle is 18. However, with only $S = 6$ distance rings, the NMSE performance achieved by the proposed non-uniform distance-sampling method is robust for all considered distances.

VI. CONCLUSION

In this paper, we have investigated the near-field channel estimation problem in XL-MIMO systems with hybrid precoding for the first time, where the near-field channel property was taken into account. The energy spread effect for the near-field channel in the angular domain was revealed at first, where one near-field path component would spread towards multiple angles, and thus the angular-domain sparsity was not achievable in the near-field region. We further showed that the energy spread effect would severely degrade the performance of existing compressed sensing based channel estimation algorithms. To address this issue, we have proposed a polar-domain representation of the near-field XL-MIMO channel and designed the angular and distance sampling rules for the polar-domain transform matrix. Since the polar-domain

transform matrix was able to simultaneously extract information of the angle and distance, both the far-field channel and near-field channel will be sparse in the polar domain. Based on this polar-domain sparsity, an on-grid P-SOMP algorithm and an off-grid P-SIGW algorithm were proposed to estimate the near-field XL-MIMO channels. Simulation results show that in the near-field region, our proposed near-field channel estimation schemes can achieve much better NMSE performance than existing far-field channel estimation schemes. In addition, the proposed near-field channel estimation schemes also performed well in the far-field region. For future works, one may consider extending the proposed polar-domain representation to the near-field channel estimation in reconfigurable intelligent surface (RIS) aided communications [30].

APPENDIX A PROOF OF LEMMA 1

Substituting $k_c = \frac{2\pi}{\lambda_c}$ and $\theta_p = \theta_q$ into (9), we get

$$f(\theta, \theta, r_p, r_q) = \left| \frac{1}{N} \sum_{n=-(N-1)/2}^{(N-1)/2} e^{j\pi n^2 \frac{d^2(1-\theta^2)}{\lambda_c} \left(\frac{1}{r_p} - \frac{1}{r_q} \right)} \right| = |F(x)|, \quad (25)$$

where $x = \frac{d^2(1-\theta^2)}{\lambda_c} \left(\frac{1}{r_p} - \frac{1}{r_q} \right)$, and the function $F(x)$ is

$$F(x) = \frac{1}{N} \sum_{n=-(N-1)/2}^{(N-1)/2} e^{j\pi n^2 x} \approx \frac{1}{N} \int_{-N/2}^{N/2} e^{j\pi n^2 x} dn. \quad (26)$$

If $r_p \leq r_q$, then $x = \frac{d^2(1-\theta^2)}{\lambda_c} \left(\frac{1}{r_p} - \frac{1}{r_q} \right)$ is larger than 0, so we have

$$F(x) \stackrel{(a)}{\approx} \frac{2}{\sqrt{2xN}} \int_0^{\sqrt{2xN/2}} e^{j\frac{\pi}{2}t^2} dt = \frac{\int_0^{\sqrt{2xN/2}} \cos(\frac{\pi}{2}t^2) dt + j \int_0^{\sqrt{2xN/2}} \sin(\frac{\pi}{2}t^2) dt}{\sqrt{2xN/2}}, \quad (27)$$

where (a) is derived by letting $n^2 x = \frac{1}{2}t^2$ in (26). Note that $\int_0^{\sqrt{2xN/2}} \cos(\frac{\pi}{2}t^2) dt$ and $\int_0^{\sqrt{2xN/2}} \sin(\frac{\pi}{2}t^2) dt$ are Fresnel functions [26]. We denote $\beta = \frac{\sqrt{2xN}}{2}$, $C(\beta) = \int_0^\beta \cos(\frac{\pi}{2}t^2) dt$, and $S(\beta) = \int_0^\beta \sin(\frac{\pi}{2}t^2) dt$, then the function $F(x)$ can be rewritten as

$$F(x) = \frac{C(\beta) + jS(\beta)}{\beta} = G(\beta), \quad (28)$$

where $\beta = \frac{\sqrt{2xN}}{2} = \sqrt{\frac{N^2 d^2 (1-\theta^2)}{2\lambda_c} \left(\frac{1}{r_p} - \frac{1}{r_q} \right)}$.

Furthermore, if $r_p > r_q$, then we have $-x = \frac{d^2(1-\theta^2)}{\lambda_c} \left(\frac{1}{r_q} - \frac{1}{r_p} \right) > 0$. Similar to the derivation of (28), we can obtain that if $-x > 0$, then $F(x) = G^*(\beta)$, where $\beta = \sqrt{\frac{N^2 d^2 (1-\theta^2)}{2\lambda_c} \left(\frac{1}{r_q} - \frac{1}{r_p} \right)}$.

To sum up, the column coherence can be approximated as $f(\theta, \theta, r_p, r_q) \approx |F(x)| \approx |G(\beta)|$, where $\beta = \sqrt{\frac{N^2 d^2 (1-\theta^2)}{2\lambda_c} \left| \frac{1}{r_q} - \frac{1}{r_p} \right|}$. Therefore, the proof of **Lemma 1** is completed.

APPENDIX B DERIVATION OF THE GRADIENT

In this appendix, the explicit derivation of the gradient of $\mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$ with respect to $\hat{\mathbf{r}} = [\hat{r}_1, \hat{r}_2, \dots, \hat{r}_L]$ and $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_L]$ is provided. For the l -th angle $\hat{\theta}_l$, the gradient of $\mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$ is given by

$$\frac{\partial \mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})}{\partial \hat{\theta}_l} = -\text{Tr} \left\{ \tilde{\mathbf{Y}}^H \frac{\partial \mathbf{P}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})}{\partial \hat{\theta}_l} \tilde{\mathbf{Y}} \right\}. \quad (29)$$

For expression simplicity, in the following derivation, we ignore the term $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$. Since $\mathbf{P} = \tilde{\Psi} \tilde{\Psi}^\dagger$ and $\tilde{\Psi}^\dagger = (\tilde{\Psi}^H \tilde{\Psi})^{-1} \tilde{\Psi}^H$, the gradient of \mathbf{P} is given by

$$\begin{aligned} \frac{\partial \mathbf{P}}{\partial \hat{\theta}_l} &= \frac{\partial \tilde{\Psi}}{\partial \hat{\theta}_l} (\tilde{\Psi}^H \tilde{\Psi})^{-1} \tilde{\Psi}^H + \tilde{\Psi} \frac{\partial (\tilde{\Psi}^H \tilde{\Psi})^{-1}}{\partial \hat{\theta}_l} \tilde{\Psi}^H \\ &\quad + \tilde{\Psi} (\tilde{\Psi}^H \tilde{\Psi})^{-1} \frac{\partial \tilde{\Psi}^H}{\partial \hat{\theta}_l}. \end{aligned} \quad (30)$$

According to the inverse matrix differentiation law that $d\mathbf{A}^{-1} = -\mathbf{A}^{-1}(d\mathbf{A})\mathbf{A}^{-1}$, the gradient of $(\tilde{\Psi}^H \tilde{\Psi})^{-1}$ is given by

$$\begin{aligned} \frac{\partial (\tilde{\Psi}^H \tilde{\Psi})^{-1}}{\partial \hat{\theta}_l} &= -(\tilde{\Psi}^H \tilde{\Psi})^{-1} \left(\frac{\partial \tilde{\Psi}^H}{\partial \hat{\theta}_l} \tilde{\Psi} + \tilde{\Psi}^H \frac{\partial \tilde{\Psi}}{\partial \hat{\theta}_l} \right) \\ &\quad \times (\tilde{\Psi}^H \tilde{\Psi})^{-1}. \end{aligned} \quad (31)$$

Since $\tilde{\Psi}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}}) = \mathbf{D}^{-1} \mathbf{A} \tilde{\mathbf{W}}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$, the gradient of $\tilde{\Psi}$ is given by

$$\frac{\partial \tilde{\Psi}}{\partial \hat{\theta}_l} = \mathbf{D}^{-1} \mathbf{A} \frac{\partial \tilde{\mathbf{W}}}{\partial \hat{\theta}_l}, \quad (32)$$

where the gradient of $\tilde{\mathbf{W}}$ is

$$\frac{\partial \tilde{\mathbf{W}}}{\partial \hat{\theta}_l} = \left[\mathbf{0}, \dots, \mathbf{0}, \frac{\partial \mathbf{b}(\hat{\theta}_l, \hat{r}_l)}{\partial \hat{\theta}_l}, \mathbf{0}, \dots, \mathbf{0} \right]. \quad (33)$$

Combining (29)-(33), we can derive the gradient of $\mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$ against the $\hat{\theta}_l$. The same procedure can be carried out to calculate the gradients of the remaining angle parameters and distance parameters. The only difference for the gradients of $\frac{1}{r}$ is that (33) is changed to

$$\frac{\partial \tilde{\mathbf{W}}}{\partial \frac{1}{r_i}} = \left[\mathbf{0}, \dots, \mathbf{0}, \frac{\partial \mathbf{b}(\hat{\theta}_l, \hat{r}_l)}{\partial \frac{1}{r_i}}, \mathbf{0}, \dots, \mathbf{0} \right]. \quad (34)$$

Finally, stacking all of the distance and angle terms in a column vector, we can obtain the gradient $\nabla_{\frac{1}{\mathbf{r}}} \mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$ and $\nabla_{\hat{\boldsymbol{\theta}}} \mathcal{L}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{r}})$.

REFERENCES

- [1] M. Cui and L. Dai, "Near-field channel estimation for extremely large-scale MIMO with hybrid precoding," in *Proc. IEEE Global Commun. Conf. (IEEE GLOBECOM)*, Madrid, Spain, Dec. 2021, pp. 1–6, doi: 10.1109/GLOBECOM46510.2021.9685542.
- [2] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
- [3] S. Hu, F. Rusek, and O. Edfors, "Beyond massive MIMO: The potential of data transmission with large intelligent surfaces," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2746–2758, May 2018.
- [4] T. S. Rappaport *et al.*, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78729–78757, 2019.
- [5] H. Elayan, O. Amin, B. Shihada, R. M. Shubair, and M.-S. Alouini, "Terahertz band: The last piece of RF spectrum puzzle for communication systems," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1–32, 2019.
- [6] L. Dai, B. Wang, M. Peng, and S. Chen, "Hybrid precoding-based millimeter-wave massive MIMO-NOMA with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 131–141, Jan. 2019.
- [7] B. Ning, Z. Tian, Z. Chen, C. Han, J. Yuan, and S. Li, "Prospective beamforming technologies for ultra-massive MIMO in terahertz communications: A tutorial," 2021, *arXiv:2107.03032*.
- [8] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, Jr., "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.
- [9] J. Lee, G.-T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.
- [10] 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 Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
- [11] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mmWave MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 245–259, Jan. 2019.
- [12] X. Wei, C. Hu, and L. Dai, "Deep learning for beamspace channel estimation in millimeter-wave massive MIMO systems," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 182–193, Jan. 2021.
- [13] X. Gao, L. Dai, S. Zhou, A. M. Sayeed, and L. Hanzo, "Wideband beamspace channel estimation for millimeter-wave MIMO systems relying on lens antenna arrays," *IEEE Trans. Signal Process.*, vol. 67, no. 18, pp. 4809–4824, Oct. 2019.
- [14] J. Rodríguez-Fernández, N. González-Prelcic, and R. W. Heath, Jr., "A compressive sensing-maximum likelihood approach for off-grid wideband channel estimation at mmWave," in *Proc. IEEE 7th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Dec. 2017, pp. 1–5.
- [15] 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.
- [16] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, Sep. 2018.
- [17] N. Gonzalez-Prelcic, H. Xie, J. Palacios, and T. Shimizu, "Wideband channel tracking and hybrid precoding for mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2161–2174, Apr. 2021.
- [18] K. T. Selvan and R. Janaswamy, "Fraunhofer and Fresnel distances: Unified derivation for aperture antennas," *IEEE Antennas Propag. Mag.*, vol. 59, no. 4, pp. 12–15, Aug. 2017.
- [19] Z. Zhou, X. Gao, J. Fang, and Z. Chen, "Spherical wave channel and analysis for large linear array in LoS conditions," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2015, pp. 1–6.
- [20] W. Zuo, J. Xin, N. Zheng, and A. Sano, "Subspace-based localization of far-field and near-field signals without eigendecomposition," *IEEE Trans. Signal Process.*, vol. 66, no. 17, pp. 4461–4476, Sep. 2018.
- [21] H. Liu, H. Meng, L. Gan, D. Li, Y. Zhou, and T.-K. Truong, "Subspace and sparse reconstruction based near-field sources localization in uniform linear array," *Digit. Signal Process.*, vol. 106, p. 102824, 2020.
- [22] B. Friedlander, "Localization of signals in the near-field of an antenna array," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 3885–3893, Aug. 2019.
- [23] Y. Han, S. Jin, C.-K. Wen, and X. Ma, "Channel estimation for extremely large-scale massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 5, pp. 633–637, May 2020.
- [24] D. Tse and P. Viswanath, *Fundamentals Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

- [25] H. Wang, J. Fang, P. Wang, G. Yue, and H. Li, "Efficient beam-forming training and channel estimation for millimeter wave OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 2805–2819, May 2021.
- [26] J. Sherman, "Properties of focused apertures in the **Fresnel region**," *IRE Trans. Antennas Propag.*, vol. 10, no. 4, pp. 399–408, Jul. 1962.
- [27] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
- [28] J. C. Bezdek and R. J. Hathaway, "Some notes on alternating optimization," in *Proc. Conf. Adv. Soft Comput.*, Feb. 2002, pp. 288–300.
- [29] D. B. Fogel and C. J. Robinson, *Computational Intelligence: The Experts Speak*. Piscataway, NJ, USA: IEEE Press, 2003. [Online]. Available: <https://ieeexplore.ieee.org/book/5271147>
- [30] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144–4157, Mar. 2021.



Mingyao Cui (Student Member, IEEE) received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2020, where he is currently pursuing the M.S. degree in electronic engineering. His research interests include massive MIMO and millimeter-wave communications.



Linglong Dai (Fellow, IEEE) received the B.S. degree from Zhejiang University, Hangzhou, China, in 2003, the M.S. degree (Hons.) from the China Academy of Telecommunications Technology, Beijing, China, in 2006, and the Ph.D. degree (Hons.) from Tsinghua University, Beijing, in 2011.

From 2011 to 2013, he was a Post-Doctoral Research Fellow with the Department of Electronic Engineering, Tsinghua University, where he was an Assistant Professor from 2013 to 2016 and has been an Associate Professor since 2016. He has coauthored the book *MmWave Massive MIMO: A Paradigm for 5G* (Academic Press, 2016). He has authored or coauthored over 70 IEEE journal articles and over 40 IEEE conference papers. He also holds 19 granted patents. His current research interests include reconfigurable intelligent surface (RIS), massive MIMO, millimeter-wave and terahertz communications, and machine learning for wireless communications. Particularly, he is dedicated to reproducible research and has made a large amount of simulation codes publicly available.

Dr. Dai has received five IEEE Best Paper Awards at the IEEE ICC 2013, the IEEE ICC 2014, the IEEE ICC 2017, the IEEE VTC 2017-Fall, and the IEEE ICC 2018. He has also received the Tsinghua University Outstanding Ph.D. Graduate Award in 2011, the Beijing Excellent Doctoral Dissertation Award in 2012, the China National Excellent Doctoral Dissertation Nomination Award in 2013, the URSI Young Scientist Award in 2014, the IEEE TRANSACTIONS ON BROADCASTING Best Paper Award in 2015, the *Electronics Letters* Best Paper Award in 2016, the National Natural Science Foundation of China for Outstanding Young Scholars in 2017, the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2017, the IEEE ComSoc Asia-Pacific Outstanding Paper Award in 2018, the *China Communications* Best Paper Award in 2019, the IEEE ACCESS Best Multimedia Award in 2020, and the IEEE Communications Society Leonard G. Abraham Prize in 2020. He was listed as a Highly Cited Researcher by Clarivate Analytics in 2020 and 2021. He is an Area Editor of IEEE COMMUNICATIONS LETTERS, and an Editor of IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.