# Two-Stage Hierarchical Beam Training for Near-Field Communications

Chenyu Wu, Changsheng You, Member, IEEE, Yuanwei Liu, Senior Member, IEEE, Li Chen, Senior Member, IEEE, and Shuo Shi, Member, IEEE

Abstract—Extremely large-scale array (XL-array) has emerged as a promising technology to improve the spectrum efficiency and spatial resolution of future wireless systems. However, the huge number of antennas renders the users more likely to locate in the near-field (instead of the far-field) region of the XL-array with spherical wavefront propagation. This inevitably incurs prohibitively high beam training overhead since it requires a two-dimensional (2D) beam search over both the angular and distance domains. To address this issue, we propose in this paper an efficient two-stage hierarchical beam training method for near-field communications. Specifically, in the first stage, we employ the central sub-array of the XL-array to search for a coarse user direction in the angular domain with conventional far-field hierarchical codebook. Then, in the second stage, given the coarse user direction, we progressively search for the finegrained user direction-and-distance in the polar domain with a dedicatedly designed codebook. Numerical results show that our proposed two-stage hierarchical beam training method can achieve over 99% training overhead reduction as compared to the 2D exhaustive search, yet achieving comparable rate performance.

Index Terms—Near-field communications, extremely largescale array (XL-array), beam training, hierarchical codebook.

## I. INTRODUCTION

Extremely large-scale array/surface (XL-array/surface) has emerged as a promising technology to achieve ultra-high spectrum efficiency and super spatial resolution in future sixthgeneration (6G) communication systems [1]–[7]. Specifically, with wireless communications migrating to higher frequency bands, such as millimeter-wave (mmWave) and even terahertz (THz) [8], XL-array/surface is expected to be widely deployed at the base station (BS) to compensate for the severe path-loss via highly directional beamforming.

With the sharp increase in the carrier frequency and number of antennas, the well-known *Rayleigh distance* will expand to

(Corresponding author: Changsheng You)

dozens to hundreds of meters, which means that the users are more likely to locate in the *near-field* region of the XLarray [1]. Different from conventional *far-field* communications with its electromagnetic (EM) field simply modeled as planar waves, the EM field of near-field communications should be accurately modeled by *spherical* waves. This unique channel characteristic can be delicately leveraged to enable flexible *beam focusing*, for which the beam energy can be focused on a specific spatial location rather than a specific direction as in conventional far-field communications [9]. Despite many advantages, near-field communications also face critical challenges in practical implementation, such as lowoverhead beam training. In this paper, we propose an efficient two-stage hierarchical beam training method to significantly reduce the training overhead of existing schemes.

#### A. Related Works

1) Near-field Wireless Communications: Near-field communications have attracted extensive research attention in recent years that exploits XL-arrays/surfaces for improving the performance of wireless systems. For example, a new concept of location division multiple access (LDMA) was proposed in [10] to enhance the multiple accessibility and system capacity in the near-field region. Particularly, the additional spatial resolution of near-field communications was exploited by the XL-array to serve multiple users simultaneously without introducing severe inter-user interference, even when the users are located in the same spatial directions. In addition, the near-field beam focusing property was utilized to improve the charging efficiency of wireless power transfer (WPT) systems [11]. In [12], an XL-surface aided communication system was studied, where the XL-surface is integrated into an access point to achieve transmit diversity and passive beamforming simultaneously. For near-field localization, the authors in [13] optimized the XL-array phase-shifts to improve the localization accuracy, while the Cramér-Rao bound (CRB) performance was studied in [14]. Moreover, the coexistence of far-field and near-field communications may also appear in future wireless communication systems [15]. In this context, a mixed near- and far-field communication scenario was investigated in [16], where the inter-user interference at the near-field user caused by the far-field beam was analytically characterized.

2) Conventional Far-field Beam Training: To fully reap the prominent beamforming gain brought by XL-array/surface, it is indispensable to perform efficient beam training at the BS to establish a high-quality initial link before implementing channel estimation and data transmission [17]. For far-field

C. Wu is with the School of Electronic and Information Engineering, Harbin Institute of Technology (HIT), Harbin 150001, China, and also with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology (SUSTech), Shenzhen 518055, China (e-mail: wuchenyu@hit.edu.cn).

C. You is with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology (SUSTech), Shenzhen 518055, China (email: youcs@sustech.edu.cn).

Y. Liu is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, UK (email: yuanwei.liu@qmul.ac.uk).

L. Chen is with the CAS Key Laboratory of Wireless-Optical Communications, University of Science and Technology of China, Hefei 230027, China (e-mail: chenli87@ustc.edu.cn).

S. Shi is with the School of Electronic and Information Engineering, Harbin Institute of Technology (HIT), Harbin 150001, China (e-mail: crcss@hit.edu.cn).

beam training, the BS conducts beam sweeping through multiple directional beams predefined in a codebook, while each user identifies its optimal beam (codeword) and feed its index back to the BS [18]. There have been substantial works on the far-field beam training (codebook) design with the aim to improve beam training accuracy and/or reduce its overhead, see e.g., [19]–[22]. Specifically, in [19], the authors proposed a hierarchical beam training method, where wide beams are generated by activating a sub-array of the full-array. The beam training accuracy was further improved in [20] by designing another hierarchical codebook that jointly utilizes the subarray and deactivation techniques. Moreover, a novel multibeam training method was proposed in [21] that steers multiple beams simultaneously at each time for reducing multi-user beam training overhead. In [22], the authors proposed a simultaneous multi-user hierarchical beam training scheme for mmWave massive multi-input multi-output (MIMO) systems, wherein an alternative minimization method was proposed to design the hierarchical codebook, based on which the beam training scheme at the BS was adaptively adjusted according to the beam training results of previous layers.

However, these far-field beam training methods may not be applicable to the near-field case, since the near-field user channel is related to not only the spatial angle-of-departure/arrival (AoD/AoA), but also the (effective) BS-user distance. In particular, when the far-field beams are applied to near-field beam training, there may appear the so-called *energy-spread* effect [23], in which the energy of a beam steering at a specific direction will be spread into multiple angles. As such, directly applying the far-field beam training methods [19]–[22] to the near-field scenario will result in severe performance degradation.

3) Near-field Beam Training: To tackle this issue, several research efforts have been made to design efficient nearfield beam training methods. Specifically, the authors in [23] proposed a new polar-domain codebook, wherein the angular domain is uniformly sampled while the distance domain is non-uniformly sampled. Based on this polar-domain codebook, the exhaustive-search based beam training method will incur prohibitively high training overhead, which is the product of the number of BS antennas and sampled distances. To reduce the training overhead, an efficient two-phase beam training method was recently proposed in [24], which decomposes the joint estimation of the spatial angle and distance into two separate phases. This scheme exploits a key observation that when far-field beams are used for beam training, the true user spatial direction approximately locates in the middle of a dominant angular region with high received signal powers. In [25], the authors proposed a deep-learning based beam training scheme which predicts the best near-field codeword by using neural networks with measurements of the beam gains of conventional far-field wide beams. For near-field wideband communications, the effect of beam split was utilized in [26] to enable fast beam training by focusing beams at different frequencies on different desired locations. Besides, the authors in [27] proposed a near-field hierarchical beam training method based on a new rectangular coordinate system. However, this scheme is highly sensitive to the sampling

steps and relies on the prior knowledge of the candidate region. In [28], the authors customized a near-field hierarchical beam training scheme for the digital precoding architecture by designing a near-field multi-resolution codebooks using the Gerchberg–Saxton (GS) based algorithm.

#### B. Our Contributions

In this paper, we propose an efficient near-field hierarchical beam training design to further reduce the training overhead of existing near-field beam training methods. The main contributions of this paper are summarized as follows.

- First, we present one key observation that by properly activating a central sub-array of the XL-array for beam training, the near-field energy-spread effect can be effectively mitigated. This thus inspires us to design a simple yet efficient *two-stage hierarchical* beam training method. Specifically, in the first stage, we estimate the *coarse user direction* by employing the sub-array technique to implement the conventional far-field hierarchical beam training. Then, in the second stage, we further progressively resolve *the user direction and its effective distance* by devising a dedicated polar-domain hierarchical codebook as well as a new beam training method.
- Second, we show that the training overhead of the proposed two-stage hierarchical beam training design scales with the number of XL-array antennas, N, in the order of O(log<sub>2</sub>(N)). Moreover, we discuss the extensions of the proposed beam training design to other systems with planar arrays and metasurfaces.
- Finally, numerical results are provided to verify the effectiveness of our proposed two-stage hierarchical beam training design. In particular, we show that our proposed hierarchical beam training design can achieve over 99% and 95% beam training overhead reduction as compared with the exhaustive-search based [23] and two-phase [24] near-field beam training schemes, respectively, yet achieving comparable rate performance.

The rest of the paper is organized as follows. In Section II, we present the near-field communication system model. Several existing benchmark schemes for near-field beam training are introduced in Section III. In Section IV, we present our key observations and the proposed two-stage hierarchical beam training method. In Section V, we discuss several key parameters of our design and the extensions to other systems. Numerical results are shown in Section VI, followed by the conclusions in Section VII.

*Notations:* For a vector **a**,  $(\mathbf{a})^T$  and  $(\mathbf{a})^H$  denote its transpose and conjugate transpose, respectively. For a set S, S[n] returns its *n*-th element.  $\lceil \cdot \rceil$  and  $\lfloor \cdot \rfloor$  denote the ceil and floor operator, respectively. **0**<sub>n</sub> represents a vector with *n* zero elements.  $C\mathcal{N}(\mu, \sigma^2)$  denotes the circularly symmetric complex Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ .  $\mathbb{I}(\cdot)$  is the indicator function.  $\otimes$  stands for the Kronecker product.

#### **II. SYSTEM MODEL**

As shown in Fig. 1, we consider a narrow-band downlink communication system, where a BS equipped with an N-



Fig. 1: A narrow-band XL-array communication system.

antenna uniform linear array (ULA) communicates with a single-antenna user.<sup>1</sup> The antennas are placed along the *y*-axis with the coordinate of the *n*-th antenna given by  $(0, \delta_n d)$ , where  $\delta_n = \frac{2n-N-1}{2}$ , and  $d = \frac{\lambda_c}{2}$  is the antenna spacing with  $\lambda_c$  denoting the carrier wavelength. Then, based on the spherical wavefront propagation model, the near-field line-of-sight (LoS) channel from the BS to the user can be modeled as [23]

$$\mathbf{h} = \sqrt{N}h\mathbf{b}(\theta, r),\tag{1}$$

where  $h = \frac{\sqrt{\rho_0}}{r} e^{-j\frac{2\pi}{\lambda_c}r}$  is the complex channel gain with  $\rho_0$  denoting the reference channel power gain at a distance of 1 m,  $\theta$  is the sine of the spatial AoD  $\phi$ , and r denotes the distance between the geometric center of the BS and the user. Besides,  $\mathbf{b}(\theta, r)$  denotes the near-field steering vector, given by<sup>2</sup>

$$\mathbf{b}(\theta, r) = \frac{1}{\sqrt{N}} \left[ e^{-j\frac{2\pi}{\lambda_c}(r^{(1)} - r)}, ..., e^{-j\frac{2\pi}{\lambda_c}(r^{(N)} - r)} \right]^T, \quad (2)$$

where  $r^{(n)} = \sqrt{r^2(1-\theta^2) + (r\theta - \delta_n d)^2}$  represents the distance between the *n*-th antenna at the BS and the user. Note that we assume in this paper that the user is located in the Fresnel region, i.e.,  $r > r_{\min} = \max\{\frac{1}{2}\sqrt{\frac{D^3}{\lambda_c}}, 1.2D\}$ , where the amplitude variations across the antennas are negligible [29], [30]. It is worth pointing out that when *r* is sufficiently large,  $\mathbf{b}(\theta, r)$  is equivalent to the conventional far-field steering vector

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} \left[ 1, e^{-j\pi\theta}, ..., e^{-j\pi(N-1)\theta} \right]^T.$$
(3)

This is because  $r^{(n)} \approx r - \delta_n d\theta$  when r is sufficiently large. Thus, the near-field steering vector is a general model that characterizes the channel accurately.

Based on the BS-user channel in (1), the received signal at the user is given by

$$y(\mathbf{v}) = \mathbf{h}^H \mathbf{v} x + z_0, \tag{4}$$

where x denotes the transmitted symbol of the BS, v represents the analog phase-shifter (PS) based transmit beamforming vector at the BS, and  $z_0 \sim C\mathcal{N}(0, \sigma^2)$  is the additive white Gaussian noise (AWGN) at the user with power  $\sigma^2$ .

From (4), it can be easily obtained that the optimal BS beamforming vector is  $\mathbf{v}^* = \mathbf{b}(\theta, r)$ . In practice, the beam training method is usually performed to establish a high-quality initial link before channel estimation and data transmission, by finding the optimal beam codeword that matches  $\mathbf{v}^*$ .

#### **III. BENCHMARK SCHEMES**

In this section, we introduce existing beam training schemes for near-field communications and point out their limitations.

#### A. Exhaustive-Search Based Near-field Beam Training

Different from the far-field beam training, the near-field beam training requires the beam search in both the angular and distance domains. To this end, a polar-domain codebook was proposed in [23], which is given by

$$\mathbf{W}^{\text{pol}} = \left\{ \mathbf{w}_{n,s}^{\text{pol}} = \mathbf{b}(\theta_n, r_n^{(s)}) | n \in \mathcal{N}, s \in \mathcal{S} \right\},$$
(5)

where  $\mathcal{N} = \{1, ..., N\}$ ,  $\mathcal{S} = \{1, ..., S\}$ ,  $\theta_n = \frac{2n-N-1}{N}$ represents the pointing angle of each codeword in  $\mathbf{W}^{\text{pol}}$ , which is uniformly sampled in the spatial angular domain [-1, 1], S and  $r_n^{(s)}$  denote the number of sampled distances and the *s*-th sampled distance for each angle  $\theta_n$ , respectively. Specifically,  $r_n^{(s)}$  is given by

$$r_n^{(s)} = \frac{1}{s} S_\Delta (1 - (\theta_n)^2), s = 0, 1, ..., S - 1,$$
(6)

where  $S_{\Delta}$  is a constant [23]. Note that  $\forall n$ , the distance intervals,  $r_n^{(s+1)} - r_n^{(s)}$ , s = 0, ..., S-1, is larger as the distance increases since the near-field effect gradually diminishes with distance.

Based on the polar-domain codebook in (5), we denote

$$(n^*, s^*) = \arg\max_{(n,s)} (|\mathbf{b}^H(\theta, r)\mathbf{w}_{n,s}^{\mathrm{pol}}|^2)$$
(7)

as the optimal codeword index for the user. In practice, the exhaustive-search based method can be applied to find the best beam that attains the maximum received signal-to-noise ratio (SNR) at the user. However, this method incurs a total overhead of NS, which is unaffordable when N is large. For example, when N = 512, S = 6, the total beam training overhead is 3072.

#### B. Two-phase Near-field Beam Training

To reduce the huge beam training overhead of the exhaustive-search based scheme, a fast near-field beam training method was proposed in [24] that separately estimates the spatial user direction and distance in two sequential

<sup>&</sup>lt;sup>1</sup>The proposed beam training method can be readily applied to the multiuser scenario by employing time division multiple access (TDMA). Therefore, without loss of generality, we consider a typical user in this paper.

<sup>&</sup>lt;sup>2</sup>In this paper, the near-field steering vector can represent the response of a ULA/sub-array with arbitrary numbers of antennas.

phases. Specifically, a key observation shows that when the far-field beams are used for beam training, the true spatial direction of the near-field user approximately lies in the middle of the dominant angular region with sufficiently high beam powers (as illustrated in the blue curve of Fig. 2). Given this observation, in the first phase, we employ the conventional angular-domain far-field codebook given by

$$\mathbf{W}^{\text{ang}} = \left\{ \mathbf{w}_n^{\text{ang}} = \mathbf{a}(\theta_n) \middle| n = 1, \dots, N, \theta_n = \frac{2n - N - 1}{N} \right\}$$
(8)

to perform beam sweeping over the whole angular domain. Then, the K directions that lie in the middle of the dominant angular region are selected as candidate spatial angles. In the second phase, the effective distance of the user is estimated by searching over the distance domain according to (6) for each candidate spatial angle. Compared with the exhaustive-search based beam training method, the beam training overhead of this method is reduced to N + KS. However, this overhead is still proportion to N, which is prohibitively high when N is large. For instance, when N = 512, S = 6, K = 1, the total overhead of the two-phase scheme in [24] is 518.

## C. Conventional Far-field Hierarchical Beam training

Consider the conventional far-field hierarchical beam training method and its issues when applied in the near-field beam training. The basic idea of hierarchical beam training is to first search for a coarse user direction by employing wide beams and then progressively identify fine-grained user spatial directions by using narrower beams. The corresponding farfield hierarchical codebook can be implemented by either the deactivation or the joint sub-array and deactivation techniques, with the details given in [19] and [20], respectively. When N = 512, the total overhead of far-field hierarchical beam training is  $2 \log_2(N) = 18$ .

However, the conventional far-field hierarchical beam training is no longer applicable to the near-field case due to the unique near-field *energy-spread* effect [23], which will be specified in the next section.

## IV. PROPOSED TWO-STAGE HIERARCHICAL BEAM TRAINING

In this section, we propose a new two-stage hierarchical beam training method to significantly reduce the overhead of the existing near-field beam training methods, yet achieving superior rate performance.

## A. Key Idea

First, we define the beam gain as follows.

**Definition 1.** Define  $g(\mathbf{b}(\theta, r), \mathbf{w})$  as the normalized beam gain of the beamforming vector  $\mathbf{w}$  along the channel steering vector  $\mathbf{b}(\theta, r)$ , which is given by

$$g(\mathbf{b}(\theta, r), \mathbf{w}) = |\mathbf{b}^H(\theta, r)\mathbf{w}|.$$
(9)

Next, consider a near-field user with a channel steering vector  $\mathbf{b}(\theta, r)$ . We present two key observations when the far-field beams based on the full- or sub-array are applied for beam sweeping.



Fig. 2: Normalized beam gain of the full-array and a central sub-array, respectively. The Xl-array has N = 512 antennas and the central sub-array includes  $N_c = 128$  antennas. The carrier frequency f is 100 GHz and the BS-user distance is r = 5 m.

**Observation 1** (Near-field energy-spread effect). Consider the full-array beam sweeping with far-field beams given in (8). The beam gain defined in (8) is large and fluctuates over a wide spatial angular region (see the blue curve of Fig. 2).

This is the so-called *energy-spread* effect for near-field communications [23], which can be explained as follows: The XL-array can be equally divided into multiple sub-arrays, such that the user locates in the far-field region of each sub-array, while it lies in the near-field region of the full-array. This thus leads to different AoDs (or called pointing directions) from each individual sub-array to the user. As such, when the full-array far-field beams are applied for beam sweeping in the neighborhood of the true direction  $\theta$ , the beam angle will align with the pointing directions of certain sub-arrays, hence yielding a wide angular region with strong beam gains. Therefore, it is expected that if we use the full XL-array for hierarchical beam training, the users cannot distinguish the optimal beam direction by finding the maximum received signal power among all beams.

**Observation 2** (Sub-array beam sweeping without nearfield effect). Consider the sub-array beam sweeping with far-field beams where only a central sub-array with  $N_c = N/4$  antennas are activated, i.e.,  $\mathbf{w}_i = \left[\mathbf{0}_{\frac{3N}{8}}^T, \mathbf{a}^T \left(-1 + \frac{2i-1}{N_c}\right), \mathbf{0}_{\frac{3N}{8}}^T\right]^T$ ,  $i = 1, ..., N_c$ . The beam gain of the central sub-array has a sharp peak pointing to the true direction of the user, and the normalized beam gain of the central sub-array approximates that of the full-array (see the red curve in Fig. 2).

**Observation 2** indicates that if we select a relatively small number of antennas for beam training, the users can be regarded as residing in the far-field region of the sub-array.

Motivated by the above, we propose a new *two-stage hierarchical beam training* method, as illustrated in Fig. 3. Specifically, in the first stage, we estimate the *coarse user direction* by employing the sub-array for implementing the



Fig. 3: Illustration of the proposed two-stage hierarchical beam training method (codebook).

conventional far-field hierarchical beam training based on an upper-level codebook in the angular domain. Then, in the second stage, we further resolve the user direction as well as its effective distance by designing a dedicated lower-level codebook in both the angular and distance domains as well as devising a new beam training method. Referring to the farfield binary-tree based hierarchical beam training [20], the total number of layers of our hierarchical codebook can be set as  $L_t = \log_2(N)$ , while the upper-level and lower-level codebooks include L and  $L_t - L$  layers, respectively. The details will be presented in subsequent subsections.

## B. Stage 1: Coarse User Direction Estimation

1) Upper-level 1D Hierarchical Codebook Design: For Stage 1, we first devise its upper-level one-dimensional (1D) hierarchical codebook in the angular domain and then present the beam training method for coarse user direction estimation.

Specifically, similar to the conventional far-field hierarchical codebook, the proposed upper-level codebook includes L layers of codewords based on the following two criteria. Firstly, for each layer in the upper-level codebook, each codeword should cover a specific angular region as well as the entire distance domain, i.e.,  $[r_{\min}, \infty]$ . Moreover, all the codewords in each layer should equally divide the entire spatial angular domain [-1, 1]. Secondly, the beam coverage of an arbitrary codeword within a layer should be covered by the union of those of two codewords in the next layer.

To satisfy the above two criteria, the upper-level codebook is devised as follows. For the  $\ell$ -th layer,  $N_{\ell} = 2^{\ell}$  antennas in the center of the BS XL-array are activated, which can generate a far-field beam with beam width of  $2/N_{\ell}$  [19]. Note that with a smaller  $N_{\ell}$ , the user is more likely to reside in the far-field region of the BS. As such, for the  $\ell$ -th layer, it has  $N_{\ell}$  codewords, each pointing to a spatial direction of  $-1 + \frac{2i-1}{N_{\ell}}, \ell = 1, ..., N_{\ell}$ . Let  $\mathbf{W}^{(\ell)} = \{\mathbf{w}_{1}^{(\ell)}, ..., \mathbf{w}_{N_{\ell}}^{(\ell)}\}$  denote the set of the  $\ell$ -th layer codewords in the first stage. Mathematically, the *i*-th codeword in  $\mathbf{W}^{(\ell)}$  is given by

$$\mathbf{w}_{i}^{(\ell)} = \left[\mathbf{0}_{\frac{N-N_{\ell}}{2}}^{T}, \mathbf{b}^{T}\left(-1+\frac{2i-1}{N_{\ell}},\infty\right), \mathbf{0}_{\frac{N-N_{\ell}}{2}}^{T}\right]^{T}, \quad (10)$$
$$i = 1, \dots, N_{\ell}.$$

It is worth noting that  $\mathbf{b}(\theta, \infty)$  is equal to  $\mathbf{a}(\theta)$ .

2) Hierarchical Beam Training of Stage 1: Based on the designed upper-level codebook, the angular-domain hierarchical beam training in Stage 1 can be implemented by using the binary-tree based beam search. Suppose that  $\mathbf{w}_{i^*}^{(L)}$  is the optimal codeword in the *L*-th layer of Stage 1 with  $i^*$  denoting the best codeword index. The coarse direction is then estimated as  $\theta^* = -1 + \frac{2i^* - 1}{2L}$ , while the true direction lies in the angular region  $[-1 + \frac{2i^* - 2}{2L}, -1 + \frac{2i^*}{2L}]$ .

For Stage 1, it takes two training beams for each layer of beam training and thus the total overhead is 2L, which is consistent with the conventional far-field case [19].

**Remark 1.** (Deactivation approach) Note that in the conventional far-field hierarchical codebook design [19], the activated sub-array can be selected arbitrarily from the entire array. In contrast, for the proposed first-stage codebook design, the activated sub-array are selected as those located in the middle of the full-array for improving the beam training performance. This is because the use of non-central sub-array may cause a larger estimation error as it may not point to the desired spatial directions during beam training. For the same reason, the joint sub-array and deactivation technique proposed in [20] is not applicable to our design.

### C. Stage 2: Fine-grained Direction-and-distance Estimation

Given the coarse user direction estimated in Stage 1, the optimal near-field beam vector for the user can be found



Fig. 4: Illustration of the sampled distances of the lower-level codebook.

by exhaustively searching over all candidate direction-anddistance pairs. However, the beam training overhead of this method is the product of the number of remaining candidate directions, i.e.,  $2^{L_t-L}$ , and the number of sampled distances S, and hence requiring  $2^{L_t-L}S$  training overhead in the second stage. To further reduce the overhead, we devise below a new lower-level *two-dimensional (2D)* hierarchical codebook as well as its dedicated beam training design.

1) Lower-Level 2D Hierarchical Codebook Design: As illustrated in Fig. 3, the lower-level codebook in Stage 2 consists of  $L_t - L$  layers of beam codewords, which is designed to satisfy the following criteria to facilitate the binary-tree search in the hierarchical beam training (see Section IV-C2). First, each codeword of the lower-level codebook should cover a specific angular region as well as a distance region in the polar domain. Moreover, all the codewords in each layer should jointly cover the entire angular domain [-1, 1] and the distance domain  $[r_{\min}, \infty]$ . Second, the beam coverage of an arbitrary codeword in one layer should be covered by the union of those of two codewords in the next layer.

To satisfy the above criteria, we design the lower-level codebook as follows. For the *u*-th layer,  $u = L + 1, ..., L_t$ ,  $N_u = 2^u$  antennas in the center of the BS XL-array are activated. Denote the set of the *u*-th layer codewords as  $\overline{\mathbf{W}}^{(u)} = \{\overline{\mathbf{w}}_{1,1}^{(u)}, ..., \overline{\mathbf{w}}_{1,S_u}^{(u)}, ..., \overline{\mathbf{w}}_{i,1}^{(u)}, ..., \overline{\mathbf{w}}_{i,S_u}^{(u)}, ...\}$ , wherein each codeword  $\overline{\mathbf{w}}_{i,j}^{(u)}$ ,  $i = 1, ..., N_u$ ,  $j = 1, ..., S_u$ , corresponds to a near-field steering vector pointing to a specific direction-and-distance pair, i.e.,  $\mathbf{b}(\theta_i^{(u)}, r_{i,j}^{(u)})$ , and  $S_u = 2^{u-L}$  is the number of sampled distances for  $\theta_i^{(u)}$  in the *u*-th layer. The design of codeword  $\overline{\mathbf{w}}_{i,j}^{(u)}$  is specified as follows:

Firstly, for the *u*-th layer, each codeword points to a spatial direction such that the sampled directions equally divide the angular domain [-1, 1], i.e.,  $\theta_i^{(u)} = -1 + \frac{2i-1}{N_u}$ ,  $i = 1, ..., N_u$ . Mathematically, we have

## Algorithm 1 Proposed two-stage hierarchical beam-training

- 1: Input:  $N, f, L, S_{\Delta}, S$ .
- 2: Stage 1: Coarse user direction estimation
- 3: **for**  $\ell = 1, ..., L$  **do**
- 4: Calculate the upper-level hierarchical codebook according to (10).
- 5: Test two codewords and find the best codeword index by measuring the received signal power of (4).
- 6: end for
- 7: Stage 2: Fine-grained direction-and-distance estimation
- 8: for  $u = L + 1, ..., L_t 1$  do
- 9: Calculate the lower-level hierarchical codebook according to (11).
- 10: Test four codewords and find the best codeword index by measuring the received signal power of (4).
- 11: end for

Ń

- 12: Calculate the sampled distances of the last layer by (15).
- 13: Test the signal power of the last-layer codewords.
- Output: The best codeword w\*, and the estimated user direction and distance.

$$\bar{\mathbf{v}}_{i,j}^{(u)} = \begin{bmatrix} \mathbf{0}_{\frac{N-N_u}{2}}^T, \mathbf{b}^T \left( \theta_i^{(u)}, r_{i,j}^{(u)} \right), \mathbf{0}_{\frac{N-N_u}{2}}^T \end{bmatrix}^T, \quad (11)$$
$$i = 1, ..., N_u, j = 1, ..., S_u.$$

Next, consider the sampled distances for the codewords with the same direction  $\theta_i^{(u)}$ . To cover the distance domain sampled in (6), we define a *distance index set* as

$$S \triangleq \{0, \dots, S-1\}. \tag{12}$$

Then, for each spatial direction  $\theta_i^{(u)}$  of the *u*-th layer codewords,  $S_u$  sampled distances are chosen such that the distance index set are progressively divided in the binary-tree manner, as illustrated in Fig. 4. Specifically, let I(u, j) denote a function that gives the index of the *j*-th sampled distance of the *u*-th layer codewords, which is given by

$$I(u,j) = \left\lceil \frac{2j-1}{2S_u} S \right\rceil.$$
 (13)

Consequently, the steering sampled distance of codeword  $\bar{\mathbf{w}}_{i,j}^{(u)}$  is

$$r_{i,j}^{(u)} = \frac{S_{\Delta}}{\mathcal{S}[I(u,j)]} (1 - (\theta_i^{(u)})^2).$$
(14)

2) *Hierarchical Beam Training of Stage 2:* Based on the devised lower-level codebook, the second-stage beam training design is detailed as the following two cases:

1) **Layers**  $u = L + 1, ..., L_t - 1$ . For the lower layers except the last layer, we implement the binary-tree based beam search over both the angular and the distance domain to gradually resolve the fine-grained direction-and-distance of the user. Specifically, for each layer of beam training, four codewords in (11) are tested, corresponding to the combinations of two steering directions and two steering distances. According to the received signal power at the user, we can decide the best beam in the *u*-th layer as well as its steering direction and distance. Then, we enter the next layer and repeat this procedure.

2) Last layer, i.e.,  $u = L_t$ . For the last layer, we aim to determine the best near-field beam for the user to maximize the beam training accuracy. Let  $\bar{\mathbf{w}}_{i^*,j^*}^{(L_t-1)}$  denote the best codeword obtained in layer  $(L_t - 1)$  with  $(i^*, j^*)$  denoting the corresponding beam index. Then, for the beam training in Layer  $L_t$ , the two beam directions are selected as  $\theta_1 = -1 + \frac{2(2i^*-1)-1}{N}$  and  $\theta_2 = -1 + \frac{2(2i^*)-1}{N}$ . For each direction, we select all the beam codewords in S whose steering distance indices are within  $[I(L_t - 1, j^* - 1), I(L_t - 1, j^* + 1)]$ . Mathematically, the steering distances of the selected codewords for direction  $\theta_i$  with i = 1, 2 are

$$r \in \left\{ \frac{S_{\Delta}}{\mathcal{S}[k]} (1 - (\theta_i)^2) \right\},\tag{15}$$

where  $I(L_t - 1, j^* - 1) < k < I(L_t - 1, j^* + 1)$ .

Accordingly, for Stage 2, it requires testing 4 codewords for each layer  $u = L + 1, ..., L_t - 1$  and the remaining  $\lceil \frac{S}{2^{L_t-L}} \rceil$  codewords for the last layer  $u = L_t$ . The total beam training overhead of Stage 2 is  $4(L_t - L - 1) + \lceil \frac{S}{2^{L_t-L}} \rceil$ . The detailed procedure of the proposed two-stage hierarchical beam training design is summarized in **Algorithm 1**.

An example is provided as below to demonstrate the detailed procedure of the proposed two-stage hierarchical beam training method.

**Example 1:** Consider the case with N = 512. The user is located at  $[\theta, r] = [0, 40 \, m]$ . The parameters are set as  $L_t = 9$ , L = 7, and  $S_{\Delta} = 70$ . As introduced in Section IV-B2, the coarse user direction is gradually estimated as -0.0078 after 7 layers of beam training in Stage 1 (the best codeword of Stage 1 is  $\mathbf{w}_{64}^{(7)}$ ). Then, we enter the second stage. Four codewords are tested in the 8-th layer, i.e.,  $\bar{\mathbf{w}}_{i,j}^{(8)}, i = 127, 128, j = 1, 2, \text{ with their sampled distances}$ given by  $r_{i,1}^{(8)} = 68(1 - (\theta_i^{(8)})^2), r_{i,2}^{(8)} = 17(1 - (\theta_i^{(8)})^2).$ Suppose that  $\bar{\mathbf{w}}_{128,1}^{(8)}$  is the best codeword for the 8-th layer. Then, in the last layer, six beam codewords are tested, i.e.,  $\bar{\mathbf{w}}_{i,j}^{(9)}, i = 256, 257, j = 1, 2, 3$ , corresponding to the sampled distances  $r \in [\infty, 68, 34]$  (see (15)). Finally, if  $\bar{\mathbf{w}}_{256,3}^{(9)}$  yields the maximum received signal power at the user, the best codeword of the proposed two-stage hierarchical beam training method is identified as  $\bar{\mathbf{w}}_{256,3}^{(9)}$ , which steers to the location  $[\theta, r] = [-0.0019, 35\,m].$ 

#### V. DISCUSSIONS AND EXTENSIONS

In this section, we discuss several key parameters of our proposed beam training design as well as its extensions to other systems with planar array and metasurface.

### A. Beam Training Overhead

Based on the above introduction, the total training overhead of the proposed two-stage hierarchical method is given by

$$T = 2L + 4(L_t - L - 1) + \left\lceil \frac{S}{2^{L_t - L}} \right\rceil$$

$$\stackrel{(a)}{\approx} 2L + 4(L_t - L) \qquad (16)$$

$$= 2\log_2(N_L) + 4\log_2(\frac{N}{N_L}) \sim \mathcal{O}(\log_2(N)),$$

where (a) is obtained by assuming that four codewords are tested in the last layer, similar as the preceding layers of Stage 2. Generally, it takes 2 training beams for each layer of Stage 1 for coarse user direction estimation, followed by 4 training beams for each layer of Stage 2 to fine-tune both the beam direction and distance. Note that the approximation in (16) is not affected by S since N is practically much larger than S. According to (16), the training overhead of our design scales in the order of  $\mathcal{O}(\log_2(N))$ , which is significantly lower than the benchmark near-field beam training schemes introduced in Section III.

**Example 2:** Consider the same system setup as that in **Example 1** with random user locations. The total beam training overhead of our proposed design is  $2 \times 7 + 4 + 6 = 24$ .

#### B. Proper Value of L

In fact, there exists a trade-off between the beam training overhead and its accuracy with different L. In particular, when  $L = L_t$ , the proposed hierarchical codebook reduces to the conventional far-field hierarchical codebook [19]. Generally, for minimizing the beam training overhead, it is desirable to choose L as large as possible. However, for guaranteeing the beam training performance in the near field, we should properly choose a small number of activated antennas in Stage 1 (a small L) to avoid the *energy-spread* effect, which is affected by the number of activated antennas and user distance. To balance the above trade-off, we first show in Fig. 5(a) that if the user is located in the Fresnel region  $(r \ge r_{\min} = \frac{1}{2}\sqrt{\frac{D^3}{\lambda_c}} = \frac{\lambda_c}{4}\sqrt{\frac{N^3}{2}})$ , there is little energy-spread during the first-stage beam training when  $L \leq L_t - 2$ . Then, we plot in Fig. 5(b) the beam gain of all the codewords in (11) for  $L = L_t - 2$  and  $L = L_t - 3$ , respectively. It can be observed that the beam gain of the codewords with the first sampled distance (r = 9 m) approximates that of the codewords with the second sampled distances (r = 68 m) for N = 512, L = 7. This indicates that if Stage 2 starts at an earlier layer than  $L = L_t - 2$ , there is little enhancement in the spatial resolution for our proposed two-stage hierarchical beam training method, yet with increased training overhead as given in (16). Therefore, L can be practically chosen as  $L_t - 2$  for an arbitrary number of antennas in the XL-array without loss of generality. On the other hand, if the user locates in a region with the BS-user distance larger than  $R_c$  (not equal to  $r_{\min}$  but known as a prior), the maximum number of activated antennas in Stage 1,  $N_L$ , should be chosen such that the user locates within the Fresnel boundary of  $N_L$ , i.e.,  $N_L = \lfloor (\frac{4R_c}{\lambda_c})^2 \rfloor$ .



(a) Normalized beam gain of the central sub-array with  $N_L = N/4$  antennas  $(L = L_t - 2)$ .



(b) Normalized beam gain for the cases of  $L = L_t - 2$  and  $L = L_t - 3$  with N = 512 and the BS-user distance r = 5 m.

#### Fig. 5: Numerical results for the proper choice of L.

Correspondingly, the number of layers in Stage 1 should be  $L = \lceil \log_2(N_L) \rceil$ .

#### C. Extensions to UPA and IRS:

1) UPA: The proposed two-stage hierarchical beam training method for ULA can be readily extended to the systems with a uniform planar array (UPA) at the BS.

Without loss of generality, we assume that the UPA is composed of  $M = M_y \times M_z$  elements placed in the y-z plane, and the user is located in the x-y plane, as shown in Fig. 6. Denote  $(0, y_0, z_0)$  and  $(x_u, y_u, 0)$  as the coordinate of the (1,1)-th element at the left-bottom of the UPA and the coordinate of the user, respectively. Then, the coordinate of the  $(m_y, m_z)$ -th element of the UPA can be represented as  $(0, y_{m_y}, z_{m_z})$  with  $y_{m_y} = y_0 + (m_y - 1)d$ ,  $z_{m_z} = z_0 + (m_z - 1)d$ ,  $m_y = 1, ..., M_y$ ,  $m_z = 1, ..., M_z$ . Besides, the distance from the  $(m_y, m_z)$ -th element to the user is  $D_r(m_y, m_z) = \sqrt{(x_u)^2 + (y_{m_y} - y_u)^2 + (z_{m_z})^2}$ . Based on



Fig. 6: Extension to UPA and IRS.

the spherical wavefront propagation model, the near-field LoS channel between the UPA and the user can be represented by

$$\mathbf{h}_r^H = \alpha \mathbf{c}_r^H,\tag{17}$$

where  $\alpha = \frac{\sqrt{\rho_0}}{D_r(1,1)}$  is the complex channel gain between the IRS and user, and  $\mathbf{c}_r$  is the near-field steering vector, given by

$$\mathbf{c}_{r} = \left[ e^{-j\frac{2\pi}{\lambda_{c}}D_{r}(1,1)}, ..., e^{-j\frac{2\pi}{\lambda_{c}}D_{r}(1,M_{z})}, ..., e^{-j\frac{2\pi}{\lambda_{c}}D_{r}(M_{y},M_{z})}, ..., e^{-j\frac{2\pi}{\lambda_{c}}D_{r}(M_{y},M_{z})} \right]^{T}.$$
(18)

Based on the UPA-user channel in (17), the received signal at the user is given by

$$y(\mathbf{v}) = \mathbf{h}_r^H \mathbf{v} x + z_0. \tag{19}$$

To facilitate beam training for the user, we can decouple the joint three-dimensional (3D) XL-array beam training into two sequential phases, corresponding to the horizontal and vertical beam training for two mutually perpendicular ULAs, see e.g., ULA 1 and ULA 2 in Fig. 6. As such, our proposed method can be adopted directly to the UPA case.

2) *IRS/RIS:* Intelligent reflecting surface (IRS) [2], also known as reconfigurable intelligent surface (RIS) [3], [4], has emerged as a promising technology for enhancing the spectral efficiency of future wireless systems, by leveraging a massive number of low-cost passive reflecting elements to reconfigure the wireless propagation environment. In practice, the large-size IRS is usually deployed near the users to enhance their communication performance with associated BSs. This renders the users more likely to reside in the near field of the IRS, and hence making the beam training for the cascaded IRS channel more challenging.

First, the IRS-user channel can be modeled similarly as (17) and hence is omitted here. Assume that the BS is equipped with a ULA placed along the *x*-axis and the IRS is located in the far-field region of the BS. Then, the LoS channel between

the BS and IRS can be modeled as [2]

$$\mathbf{G} = \beta \mathbf{c}_t \mathbf{a}^H(\theta_t), \qquad (20)$$

where  $\beta = \frac{\sqrt{\rho_0}}{D_t(1,1)}$  is the complex channel gain between the BS and IRS,  $\theta_t$  is the spatial AoD from the BS antennas to the IRS,  $\mathbf{c}_t$  is the near-field steering vector, given by

$$\mathbf{c}_{t} = \left[e^{-j\frac{2\pi}{\lambda_{c}}D_{t}(1,1)}, ..., e^{-j\frac{2\pi}{\lambda_{c}}D_{t}(M_{y},M_{z})}\right]^{T}, \qquad (21)$$

where  $D_t(m_y, m_z) = \sqrt{(x_b)^2 + (y_{m_y})^2 + (z_{m_z} - z_b)^2}, m_y = 1, ..., M_y, m_z = 1, ..., M_z$  denotes the distance from the  $(m_y, m_z)$ -th IRS element to the BS,  $(x_b, 0, z_b)$  is the coordinate of the BS. Let  $\Phi \triangleq \operatorname{diag}(e^{j\phi_1}, e^{j\phi_2}, ..., e^{j\phi_M})$  denote the diagonal IRS reflecting matrix, where for simplicity we assume that the reflection amplitude of each element is set to one, and  $\phi_m, m = 1, ..., M$  denotes the phase-shift of the *m*-th element. Based on the channel models in (17) and (20), the received signal at the user is given by

$$y(\mathbf{v}, \mathbf{\Phi}) = \mathbf{h}_r^H \mathbf{\Phi} \mathbf{G} \mathbf{v} x + z_0.$$
(22)

Since the BS and IRS are usually deployed at fixed locations with a LoS path, the BS-IRS link can be considered as quasistatic and perfectly known at the BS by using existing channel estimation methods, see e.g., [31]–[34]. To facilitate beam training for the user, we first design the BS transmit beamforming to align with the BS-IRS channel, i.e.,  $\mathbf{v} = \mathbf{a}(\theta_t)$ . As such, the BS can be treated as having an equivalent single antenna, and the received signal in (22) can be written as

$$y(\mathbf{\Phi}) = \alpha \beta \mathbf{c}_r^H \mathbf{\Phi} \mathbf{c}_t x + z_0.$$
(23)

Next, to remove the effect of the near-field BS-IRS steering vector, i.e.,  $\mathbf{c}_t$ , we rewrite the IRS phase-shift matrix as  $\mathbf{\Phi} = \bar{\mathbf{\Phi}} \operatorname{diag}(\mathbf{c}_t^H)$ , where  $\bar{\mathbf{\Phi}} \triangleq \operatorname{diag}(e^{j\bar{\phi}_1}, e^{j\bar{\phi}_2}, ..., e^{j\bar{\phi}_M})$  is defined as a pseudo phase-shift matrix of the IRS used for modifying the directional beams during beam training, and  $\operatorname{diag}(\mathbf{c}_t^H)$  is a constant phase-shift matrix reserved for counteracting  $\mathbf{c}_t$ . By doing so, (23) can be further written as

$$y(\bar{\boldsymbol{\phi}}) = \alpha \beta \mathbf{c}_r^H \bar{\boldsymbol{\phi}} x + z_0, \qquad (24)$$

where  $\bar{\phi} = [e^{j\bar{\phi}_1}, e^{j\bar{\phi}_2}, ..., e^{j\bar{\phi}_M}]^T$ . We can easily observe that (24) has a similar form as (19) by treating  $\bar{\phi}$  as the transmit beamforming vector v. Hence, the beam training for the cascaded IRS channel reduces to the near-field beam training for the IRS [21], which is similar to the case of a UPA. Based on the above discussions, our proposed beam training algorithm can also be applied to IRS-aided near-field communication systems.

## VI. NUMERICAL RESULTS

We provide numerical results in this section to validate the effectiveness of our proposed two-stage hierarchical beam training method. Consider an XL-array communication system with N = 512 and f = 100 Ghz. Hence, we have  $\lambda_c = 0.003$  m,  $S_{\Delta} = 68.27$ , S = 6,  $r_{\min} = 6.14$ m [23]. The reference SNR is  $\gamma = \frac{PN\beta_0}{r^2\sigma^2}$  with  $\beta_0 = -72$ dB, P = 30 dBm. The noise power is set as  $\sigma^2 = -80$ dBm. To characterize the beam identification accurcy, we



Fig. 7: Success rate versus BS-user distance.



Fig. 8: Achievable rate versus BS-user distance.

define  $P_{\text{suc}} = \mathbb{E}(\mathbb{I}((\hat{n}^*, \hat{s}^*) = (n^*, s^*)))$  as the success rate under the polar-domain codebook, where  $(\hat{n}^*, \hat{s}^*)$  denotes the obtained codeword index by the specific beam training scheme and  $(n^*, s^*)$  is the optimal codeword index in (7). The achievable rate in bits/second/Hz (bps/Hz) is given by  $R = \log_2(1+\gamma|\mathbf{h}^H \mathbf{v}|^2)$ . All the numerical results are averaged over 5000 random channel realizations.

For performance comparison, the following benchmark schemes are considered:

- **Perfect-CSI based beamforming**: In this scheme, the transmit beamforming vector  $\mathbf{v}$  is designed to perfectly align with the user's channel, i.e.,  $\mathbf{v}^* = \mathbf{b}(\theta, r)$ . This scheme serves as a performance upper-bound when evaluating the achievable rate of our proposed scheme.
- Exhaustive-search based near-field beam training: as introduced in Section III-A.
- Two-phase near-field beam training with K = 1 [24]. The detailed procedures are presented in Section III-B.
- **Conventional far-field hierarchical beam training** [19]: In this scheme, the conventional far-field hierarchical codebook is adopted for beam training.
- 1) Overhead Comparison: We first compare the beam

Beam training method	Total overhead	In our setup
Exhaustive-search based near-field beam training	NS	3072
Two-phase near-field beam training	N+S	518
Far-field exhaustive-search based beam training	N	512
Far-field hierarchical beam training	$2\log_2(N)$	18
Proposed two-stage hierarchical beam training	$2\log_2(N_L) + 4\log_2(\frac{N}{N_L})$	24

TABLE I: Comparison of Beam Training Overhead of Different Schemes



Fig. 9: Success rate versus SNR.



Fig. 10: Achievable rate versus SNR.

training overhead of different beam training schemes in Table I. It is observed that the proposed two-stage hierarchical beam training scheme requires a much smaller number of training symbols than other near-field beam training schemes. Specifically, our proposed scheme can achieve over 99% and 95% overhead reduction as compared to the exhaustive search-based and two-phase beam training schemes, which, however, only suffers little rate performance loss as will be shown in Figs. 8 and 10.

2) Impact of BS-user Distance: Fig. 7 shows the effect of the BS-user distance on the success rate by different beam training schemes. It is observed that our proposed scheme with ex-

tremely small training overhead achieves comparable success rate with the two-phase scheme. Moreover, the proposed twostage hierarchical beam training scheme maintains a relatively high success rate (i.e., higher than 80%) for all distances. This implies that our design is not sensitive to the BS-user distance, thus being applicable to both far-field and near-field communication scenarios. In contrast, the conventional farfield hierarchical beam training scheme suffers a significantly low success rate when the BS-user distance is short (e.g., r < 50 m). This shows that the far-field hierarchical beam training scheme is not applicable to the near field since it neglects the spherical-wavefront channel characteristic.

Besides, Fig. 8 illustrates the rate performance of different beam training schemes versus the BS-user distance. Several key observations are made as follows. Firstly, the proposed scheme significantly outperforms the far-field hierarchical beam training scheme, especially when the user is close to the XL-array. This attributes to our second-stage design, which effectively takes into account the near-field channel characteristics. Secondly, our proposed scheme attains close rate performance with the two-phase based near-field beam training scheme especially when the BS-user distance is small, yet with substantial reduced training overhead (i.e., 24 versus 518). Finally, the rate performance of our proposed scheme approximates that of the exhaustive-search based scheme when the BS-user distance is small, and finally degenerates to that of the far-field hierarchical beam training scheme when the BS-user distance is large. This also verifies that our proposed two-stage hierarchical beam training method is universal for both near-field and far-field communications.

3) Impact of the Effective SNR: Fig. 9 shows the success rate of different beam training schemes versus the effective SNR. It is observed that the success rate of our proposed two-stage hierarchical beam training scheme monotonically increases with the SNR, and finally converges to that of the two-phase near-field beam training scheme. This is because the beam identification error caused by AWGN starts to be negligible in the high-SNR regime. In contrast, the success rate of conventional far-field hierarchical beam training scheme is always lower than 40%, even under a high SNR (over 30 dB). This is because the distance information of the user.

Fig. 10 presents the performance of achievable rate versus the effective SNR under different beam training schemes. It is observed that our proposed scheme always outperforms the conventional far-field hierarchical beam training scheme by over 1 bps/Hz, yet without compromising much rate performance compared with the exhaustive-search based near-field



Fig. 11: Achievable rate versus number of antennas, N.

#### beam training scheme.

4) Impact of the Number of BS Antennas: Finally, in Fig. 11, we plot the achievable rates of different schemes under different number of BS antennas N. The users are assumed to be located near the Fresnel boundary. It is observed that our proposed scheme always outperforms the conventional far-field counterparts. This again validates that our design is robust under different system setups.

#### VII. CONCLUSION

In this paper, we proposed a novel two-stage hierarchical beam training method for near-field communications, where a coarse user direction is found in the first stage, followed by a second stage that further resolves fine-grained directionand-distance of the user. We proposed efficient hierarchical codebook design for each stage as well as its corresponding beam training method. It was shown that our proposed method significantly reduces the training overhead of the existing nearfield beam training methods, yet achieving comparable rate performance.

#### REFERENCES

- M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-field communications for 6G: Fundamentals, challenges, potentials, and future directions," *IEEE Commun. Mag.*, pp. 1–7, 2022, early access, doi=10.1109/MCOM.004.2200136.
- [2] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-Aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.
- [3] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.
- [4] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE Commun. Surv. Tut.*, vol. 23, no. 3, pp. 1546–1577, thirdquarter 2021.
- [5] H. Lu and Y. Zeng, "Communicating with extremely large-scale array/surface: Unified modeling and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4039–4053, June 2022.
- [6] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, March 2020.

- [7] J. Xu, Y. Liu, X. Mu, and O. A. Dobre, "STAR-RISs: Simultaneous transmitting and reflecting reconfigurable intelligent surfaces," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3134–3138, Sept. 2021.
- [8] Z. Chen, X. Ma, B. Zhang, Y. Zhang, Z. Niu, N. Kuang, W. Chen, L. Li, and S. Li, "A survey on terahertz communications," *China Commun.*, vol. 16, no. 2, pp. 1–35, Feb. 2019.
- [9] H. Zhang, N. Shlezinger, F. Guidi, D. Dardari, M. F. Imani, and Y. C. Eldar, "Beam focusing for near-field multiuser MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7476–7490, Sept. 2022.
- [10] Z. Wu, M. Cui, and L. Dai, "Multiple access for near-field communications: SDMA or LDMA?" arXiv preprint arXiv:2208.06349, 2022.
- [11] H. Zhang, N. Shlezinger, F. Guidi, D. Dardari, M. F. Imani, and Y. C. Eldar, "Near-field wireless power transfer for 6G internet of everything mobile networks: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 60, no. 3, pp. 12–18, March 2022.
- [12] B. Zheng and R. Zhang, "Simultaneous transmit diversity and passive beamforming with large-scale intelligent reflecting surface," *IEEE Trans. Wireless Commun.*, pp. 1–1, early access, 2022, doi=10.1109/TWC.2022.3199426.
- [13] X. Gan, C. Huang, Z. Yang, C. Zhong, and Z. Zhang, "Near-field localization for holographic RIS assisted mmwave systems," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 140–144, Jan. 2023.
- [14] A. Chen, L. Chen, Y. Chen, C. You, G. Wei, and F. R. Yu, "Cramér-Rao bounds of near-field positioning based on electromagnetic propagation model," arXiv preprint arXiv:2207.00799, 2022.
- [15] C. Han, Y. Chen, L. Yan, Z. Chen, and L. Dai, "Cross far- and near-field wireless communications in terahertz ultra-large antenna array systems," arXiv preprint arXiv:2301.03035, 2022.
- [16] Y. Zhang, C. You, L. Chen, and B. Zheng, "Mixed near- and farfield communications for extremely large-scale array: An interference perspective," arXiv preprint arXiv:2301.07277, 2023.
- [17] Y. Wang, Z. Wei, and Z. Feng, "Beam training and tracking in mmwave communication: A survey," arXiv preprint arXiv:2205.10169, 2022.
- [18] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
- [19] T. He and Z. Xiao, "Suboptimal beam search algorithm and codebook design for millimeter-wave communications," *Mobile Netw. and Appl.*, vol. 20, no. 1, pp. 86–97, 2015.
- [20] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.
- [21] C. You, B. Zheng, and R. Zhang, "Fast beam training for IRS-assisted multiuser communications," *IEEE Wireless Commun. Lett.*, vol. 9, no. 11, pp. 1845–1849, Nov. 2020.
- [22] C. Qi, K. Chen, O. A. Dobre, and G. Y. Li, "Hierarchical codebookbased multiuser beam training for millimeter wave massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8142–8152, Dec. 2020.
- [23] M. Cui and L. Dai, "Channel estimation for extremely large-scale MIMO: Far-field or near-field?" *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2663–2677, April 2022.
- [24] Y. Zhang, X. Wu, and C. You, "Fast near-field beam training for extremely large-scale array," *IEEE Wireless Commun. Lett.*, vol. 11, no. 12, pp. 2625–2629, Dec. 2022.
- [25] W. Liu, H. Ren, C. Pan, and J. Wang, "Deep learning based beam training for extremely large-scale massive MIMO in nearfield domain," *IEEE Commun. Lett.*, pp. 1–1, 2022, early access, doi=10.1109/LCOMM.2022.3210042.
- [26] M. Cui, L. Dai, Z. Wang, S. Zhou, and N. Ge, "Near-field rainbow: Wideband beam training for XL-MIMO," *IEEE Trans. Wireless Commun.*, pp. 1–1, early access, 2022, doi=10.1109/TWC.2022.3222198.
- [27] X. Wei, L. Dai, Y. Zhao, G. Yu, and X. Duan, "Codebook design and beam training for extremely large-scale RIS: Far-field or near-field?" *China Commun.*, vol. 19, no. 6, pp. 193–204, June 2022.
- [28] Y. Lu and L. Dai, "Hierarchical beam training for extremely large-scale MIMO: From far-field to near-field," *arXiv preprint arXiv:2223.14705*, 2022.
- [29] K. T. Selvan and R. Janaswamy, "Fraunhofer and fresnel distances: Unified derivation for aperture antennas," *IEEE Antennas and Propag. Mag.*, vol. 59, no. 4, pp. 12–15, Aug. 2017.
- [30] E. Björnson, O. T. Demir, and L. Sanguinetti, "A primer on near-field beamforming for arrays and reconfigurable intelligent surfaces," in 2021 55th Asilomar Conference on Signals, Systems, and Computers, 2021, pp. 105–112.

- [31] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Commun. Surv. Tut.*, vol. 24, no. 2, pp. 1035–1071, Secondquarter 2022.
- [32] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2604–2620, Nov. 2020.
- [33] C. Wu, C. You, Y. Liu, X. Gu, and Y. Cai, "Channel estimation for STAR-RIS-aided wireless communication," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 652–656, March 2022.
- [34] Y. Wei, M.-M. Zhao, M.-J. Zhao, and Y. Cai, "Channel estimation for IRS-aided multiuser communications with reduced error propagation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2725–2741, April 2022.