

RIS-aided Near-Field MIMO Communications: Codebook and Beam Training Design

Suyu Lv, *Graduate Student Member, IEEE*, Yuanwei Liu, *Fellow, IEEE*,
Xiaodong Xu, *Senior Member, IEEE*, Arumugam Nallanathan, *Fellow, IEEE*,
and A. Lee Swindlehurst, *Fellow, IEEE*

Abstract—Downlink reconfigurable intelligent surface (RIS)-assisted multi-input-multi-output (MIMO) systems are considered with far-field, near-field, and hybrid-far-near-field channels. According to the angular or distance information contained in the received signals, 1) a distance-based codebook is designed for near-field MIMO channels, based on which a hierarchical beam training scheme is proposed to reduce the training overhead; 2) a combined angular-distance codebook is designed for hybrid-far-near-field MIMO channels, based on which a two-stage beam training scheme is proposed to achieve alignment in the angular and distance domains separately. For maximizing the achievable rate while reducing the complexity, an alternating optimization algorithm is proposed to carry out the joint optimization iteratively. Specifically, the RIS coefficient matrix is optimized through the beam training process, the optimal combining matrix is obtained from the closed-form solution for the mean square error (MSE) minimization problem, and the active beamforming matrix is optimized by exploiting the relationship between the achievable rate and MSE. Numerical results reveal that: 1) the proposed beam training schemes achieve near-optimal performance with a significantly decreased training overhead; 2) compared to the angular-only far-field channel model, taking the additional distance information into consideration will effectively improve the achievable rate when carrying out beam design for near-field communications.

Index Terms—Beam training, codebook, multi-input-multi-output (MIMO), near-field communications (NFC), reconfigurable intelligent surface (RIS).

I. INTRODUCTION

Millimeter wave (mmWave) communications (30 - 300 GHz) and Terahertz (THz) communications (0.1 - 10 THz) have been recognized as promising candidate technologies for future 5G/6G networks to realize the communication goals of ultra-wide bandwidth and ultra-high transmission rate [1]–[3]. Although possessing extremely high bandwidth, mmWave and THz communication suffer from high propagation losses due

Suyu Lv is with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China (e-mail: lvsuyu@bupt.edu.cn).

Xiaodong Xu is with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China, and also with the Department of Broadband Communication, Peng Cheng Laboratory, Shenzhen 518066, Guangdong, China (e-mail: xuxiaodong@bupt.edu.cn).

Yuanwei Liu and Arumugam Nallanathan are with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: yuanwei.liu@qmul.ac.uk; a.nallanathan@qmul.ac.uk).

A. Lee Swindlehurst is with the Center for Pervasive Communications and Computing, Henry Samueli School of Engineering, University of California at Irvine, Irvine, CA 92697 USA (e-mail: swindle@uci.edu).

to penetration, reflection and diffraction, resulting in a reduced link budget compared to that in the sub-6 GHz scenario [4].

To tackle the problems caused by high-frequency communications, reconfigurable intelligent surface (RIS) technology has received extensive research attention due to its distinctive capability of restructuring the wireless propagation environment and establishing additional reliable reflection-based links between the base station (BS) and the user [5], [6]. RISs consist of numerous passive controllable reflecting elements that can adjust the phase shift of incident signals with extremely low power consumption, and thus have been regarded as one of the main enabling solutions to achieve green communication and coverage extension for next-generation wireless systems [7].

The low-cost and low-complexity characteristics of RIS reflecting elements have also promoted the emergence of extremely large-scale RIS to realize higher array gain. However, with an increase in the number of RIS reflecting elements, the boundary separating the near-field communication (NFC) region and far-field communication (FFC) region becomes more distant from the RIS, making the near-field region non-negligible [8]–[10]. In contrast to the *plane wave* model in the far-field region, the electromagnetic structure in the near-field region is fundamentally different and the *spherical wave* model should be used [8]. The spherical-wave propagation model for NFC contains both angular and distance information, which allows the resulting beam patterns to focus on a specific point [9]. Thus, NFC can utilize the new dimension of distance to realize more precise signal enhancement, interference management and user localization for RIS-assisted wireless networks [11]–[14].

However, reliable phase shift design for RIS-assisted communication systems relies heavily on accurate channel state information (CSI), which is challenging to obtain due to the passive property of RISs. In addition, the massive number of RIS elements in a multi-input-multi-output (MIMO) system will also lead to extremely large pilot overhead and high channel estimation complexity. Compared to cascaded channel estimation based on least squares or minimum mean squared error, codebook-based beam training has been widely adopted for mmWave or THz communication systems [15]–[17]. The beam training method can obtain CSI by estimating the physical directions of channel paths rather than the entire channel, thus realizing lower complexity. Moreover, beam training can directly achieve reliable beamforming by selecting a phase-shift vector from among a set of pre-defined “codewords”

1 during the training procedure, and thus is considered to be
 2 an efficient approach to acquire CSI in RIS-assisted wireless
 3 communication systems [18]–[20]. Existing codebook design
 4 and beam training schemes cannot be directly applied to RIS-
 5 assisted NFC systems or hybrid FFC & NFC systems.

6 Codebooks designed for FFC are mainly based on angular
 7 information, which is reasonable because the distance from
 8 any point on the transmitter to any point on the receiver
 9 can be considered equal, and the phase of electromagnetic
 10 waves received by different antennas is a linear function of
 11 the antenna index in FFC. However, in NFC, the phase of
 12 the electromagnetic waves received by different antennas is
 13 no longer a linear function of the antenna index, but is related
 14 to the distance between specific two points at the receiving
 15 and transmitting ends. In addition, channel responses can be
 16 decomposed into the Kronecker product of two vectors related
 17 to angle of arrival (AoA) and angle of departure (AoD) in FFC,
 18 which is no longer applicable in NFC due to the coupling
 19 between angle and distance, bringing more difficulties to the
 20 design of near-field codebooks.

21 Furthermore, in RIS-assisted communication systems, BS-
 22 RIS link and RIS-user link may be in either far-field or near-
 23 field situations, so there may be a mixture of far and near
 24 fields. In such cases, the degree of difficulty of codebook
 25 design will increase because it is necessary to comprehensively
 26 consider the channel conditions of the two links, and as
 27 a result, the size of the codebook will also become large.
 28 Designing an efficient and accurate beam training method is
 29 also a challenge.

30 A. Prior Works

31 1) *Channel modeling for NFC*: Signals in the far field of
 32 an array lead to planar wavefronts that produce linear phase
 33 variations and negligible amplitude differences across the
 34 array. However, the situation is different for near-field signals.
 35 In order to achieve better performance through reasonable
 36 resource allocation, accurate channel models are essential for
 37 NFC. Hybrid-field channel modeling schemes were studied in
 38 [21] and [22], where the far-field and near-field components
 39 are separately estimated. A polar-domain representation for
 40 near-field channel models was proposed in [23], accounting
 41 for the information in both the angular and distance domains.
 42 Considering the different multi-path characteristics observed
 43 by antennas at different distances in large-scale MIMO NFC
 44 systems, the authors in [24] proposed a non-stationary distance
 45 modeling framework.

46 2) *Beam training schemes*: The beam training process is
 47 implemented by searching for the optimal beam from multiple
 48 predefined directional beams (codewords). For beam training,
 49 testing all codewords in the codebook during the training
 50 process is the most straightforward approach. However, the
 51 codebook and the resulting training overhead can be extremely
 52 large for large-scale massive MIMO systems, especially when
 53 considering both angular and distance information in NFC.
 54 The authors in [15] proposed a new hierarchical codebook to
 55 achieve uniform beam alignment performance with low over-
 56 head for mmWave communication systems. For mmWave and

sub-THz systems with multiple users, [4] and [16] proposed
 hierarchical beamforming training strategies to carry out si-
 multaneous training for multiple users. Aiming at overcoming
 the resolution limit of directional angle estimation, [25] pro-
 posed a codebook-based beam search approach for mmWave
 massive MIMO transmissions. However, these codebooks are
 designed for the far-field case. Considering the spherical
 wavefront propagation model for NFC, [17], [26] proposed
 a two-phase beam training method to sequentially perform
 the beam sweeping in the angular and distance domains.
 The authors in [27] proposed a novel spatial-chirp codebook
 and subsequently proposed a slope-domain-based hierarchical
 update policy, where the upper-layer beam searching range
 was uniformly divided into several sub-ranges.

3) *Beam training design for RISs*: To implement beam
 training for RIS-based systems, one must first construct an
 appropriate RIS codebook based on information about the
 cascaded channel. Then, based on the information obtained
 during the training phase, the RIS selects the optimal code-
 word from the predesigned codebook to perform passive
 beamforming. For RISs with massive numbers of elements,
 highly efficient beam training schemes with good performance
 are required to reduce the training overhead. The authors in
 [19] first proposed a hierarchical codebook-generating method
 using pattern synthesis, then provided a hierarchical beam
 training method using two multi-mainlobe codewords in each
 layer. To train multiple users at the same time, a multi-lobe
 beam training mechanism was proposed in [28] based on the
 BS-RIS joint codebook, thus achieving reduced overhead. A
 multi-beam training method was proposed for a RIS-assisted
 multiuser system in [29], by dividing the RIS into multiple
 sub-surfaces and designing multi-beam directions over time.
 Although [20], [30], [31] proposed beam training schemes for
 RIS-assisted NFC systems, only limited scenarios with single-
 antenna users were considered.

56 B. Motivation and Contributions

57 As the size of the RIS increases, the near-field region
 becomes non-negligible. In NFC, CSI is not only related
 to the angle but also distance, thus the acquisition of CSI
 becomes very important for proper beam design. The methods
 for obtaining CSI can be roughly classified into two categories:
 channel estimation and beam training [31]. In this article, we
 consider the use of beam training to simultaneously estimate
 the channel parameters and perform passive beamforming
 [32], which is simpler than performing channel estimation first
 and then separately determining the optimal beamforming.

58 Although there are numerous beam training schemes for
 FFC, codebook and beam training design for NFC are still in
 their infancy, especially for RIS-assisted systems. Moreover,
 the effect of different channel modeling methods for FFC
 and NFC has not been well investigated. In order to fill
 this gap, in this article we focus on four different channel
 models for downlink MIMO RIS-aided systems. Furthermore,
 we design codebooks and beam training schemes for near-field
 and mixed-far-near-field channels, which are further used for
 RIS reflecting coefficient optimization. The main contributions
 of our work are listed as follows:

- 1) We consider four different channel models for a downlink RIS-assisted MIMO system, namely, i) FF: far-field model for both BS-RIS and RIS-user links; ii) NF: near-field model for BS-RIS link, far-field model for RIS-user link; iii) FN: far-field model for BS-RIS link, near-field model for RIS-user link; iv) NN: near-field model for both BS-RIS and RIS-user links. Here, the assumption of far-field propagation leads to a *planar-wave* channel model, while near-field propagation leads to a *spherical-wave* channel model.
- 2) According to the angular or distance information contained in the received signals, we design i) a distance-based codebook for the NN channel model, based on which we propose a hierarchical beam training scheme to reduce the training overhead; ii) a combined angular-distance codebook for NF and FN channel models, based on which we propose a two-stage beam training scheme to separately achieve alignment in the angular and distance domains.
- 3) To maximize the achievable rate, we propose an alternating optimization (AO) algorithm to carry out the joint optimization of the RIS coefficients and the transmit and receive beamformers in an iterative manner. Specifically, the RIS coefficient matrix is obtained by the proposed beam training schemes, the optimal combining matrix at the user is obtained from the closed-form solution for the mean square error (MSE) minimization problem, and the active beamforming matrix at the BS is optimized by exploiting the relationship between the achievable rate and the MSE.
- 4) We present numerical results revealing that: i) the proposed hierarchical and two-stage beam training approaches yield achievable rate performance similar to the exhaustive search (ES) method while significantly reducing the training overhead; ii) compared to angular-only far-field channel models, the achievable rate is improved by exploiting the additional distance information when performing the beam design for NFC.

The organization for the rest of this article is as follows. In Section II, we describe the proposed RIS-assisted MIMO system and the four considered channel models. In Section III, we design different codebooks for the different channel models based on the angular or distance information in the received signals. In Section IV, we propose an AO algorithm to carry out the joint optimization in an iterative manner for maximizing the achievable rate. In particular, two beam training algorithms are proposed in Section IV-A based on the predesigned codebooks, the combining matrix optimization is given in Section IV-B, and the active beamforming matrix optimization is described in Section IV-C. Simulation results and conclusions are given in Section V and Section VI, respectively.

II. SYSTEM MODEL

We consider a RIS-assisted downlink communication scenario, where a single multi-antenna BS serves a multi-antenna user. The direct path between the BS and the user is assumed

to be blocked due to the existence of propagation obstacles. The BS is equipped with an N_B -element uniform linear array (ULA) of antennas, and the user is equipped with an N_U -element ULA. The RIS consists of a uniform planar array (UPA) with $M = M_x M_y$ passive reflecting elements, where M_x and M_y denote the number of elements along the horizontal axis and vertical axis, respectively.

Four different channel models are considered for the RIS-assisted communication system, as shown in Fig. 1. Specifically, these four channel models are respectively abbreviated as (a) FF: far-field model for both BS-RIS and RIS-user links; (b) NF: near-field model for BS-RIS link, far-field model for RIS-user link; (c) FN: far-field model for BS-RIS link, near-field model for RIS-user link; (d) NN: near-field model for both BS-RIS and RIS-user links. We assume that the system operates at a frequency of f_c , so the wavelength of the signal is $\lambda_c = \frac{c}{f_c}$. Denote the array spacing as d , which we will assume here to be $d = \lambda_c/2$. Denote the array apertures of the BS, RIS and user as D_B , D_R and D_U , respectively. For the BS and the user with ULAs, the array apertures are calculated as $D_B = (N_B - 1)d$ and $D_U = (N_U - 1)d$, respectively; for the RIS with a UPA, the array aperture is calculated as $D_R = \sqrt{[(M_x - 1)d]^2 + [(M_y - 1)d]^2}$. The comparison between the far-field region and near-field region in RIS-assisted systems is summarized in **TABLE I**, including the Rayleigh boundary when $d = \lambda_c/2$.

A. Channel Model

In order to explore the differences among the four channel scenarios in terms of signal model, beam design, transmission performance, etc., we first provide mathematical descriptions of the models below. More specifically, we will first present the far-field model for the BS-RIS and RIS-user channels, then give the near-field model for the BS-RIS and RIS-user channels, and finally obtain the cascaded channel models for the four channel scenarios mentioned above.

1) *Far-field channel model*: When the far-field channel model is considered, the channel between the BS and RIS is characterized by the Saleh-Valenzuela channel model [33], given by

$$\mathbf{G}_{B,R}^{\text{far}} = \sqrt{\frac{MN_B}{L_B}} \sum_{l=0}^{L_B-1} \beta_l \mathbf{a}_R(\alpha_l^A, \varphi_l^A) \mathbf{a}_B^H(\alpha_l^D), \quad (1)$$

where L_B is the number of paths between the BS and RIS, β_0 represents the complex gain of the line-of-sight (LoS) component, β_l ($1 \leq l \leq L_B - 1$) denotes the complex gain of the l -th non-line-of-sight (NLoS) path [34], α_l^A and φ_l^A denote the azimuth and elevation AoA associated with the RIS, α_l^D denotes the AoD associated with the BS, and $\mathbf{a}_R(\alpha_l^A, \varphi_l^A)$ and $\mathbf{a}_B(\alpha_l^D)$ are the array response vectors associated with the RIS and the BS, respectively.

For a UPA with M_x and M_y elements ($M_x M_y = M$) on the horizontal and the vertical axes respectively, the RIS array

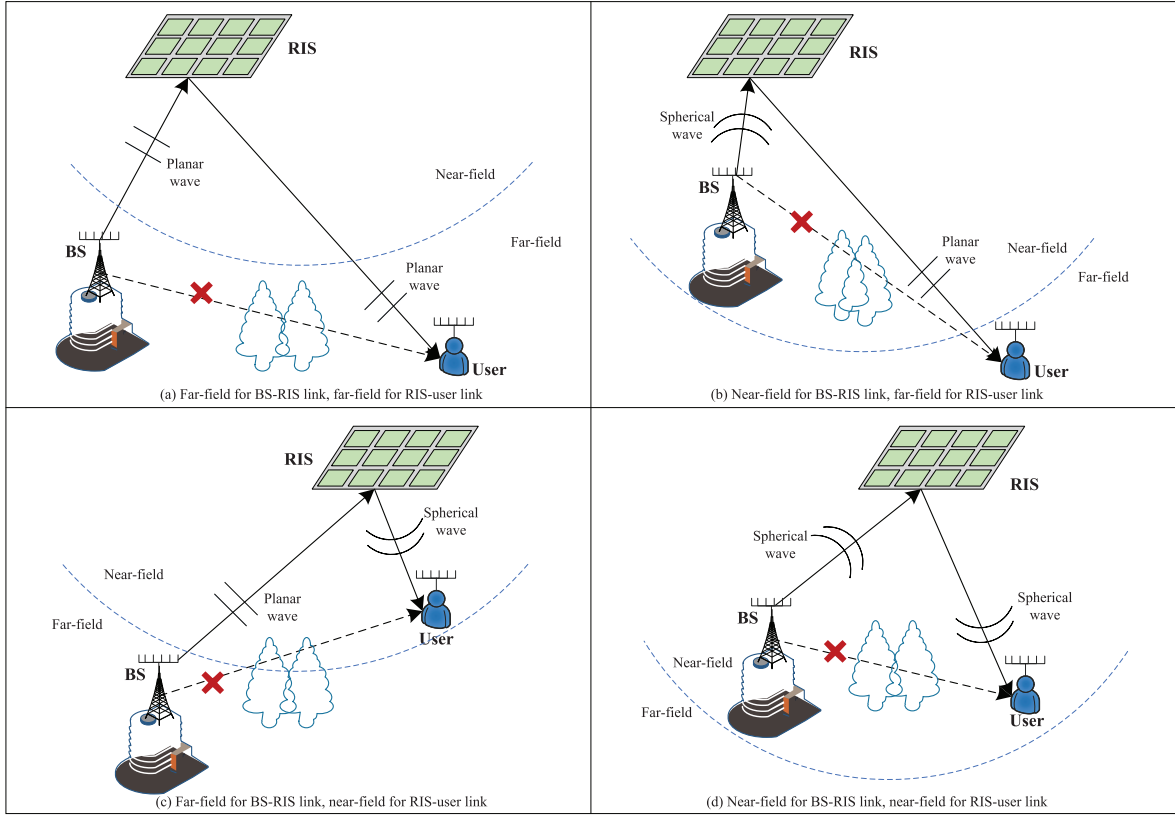


Fig. 1. Different channel models for RIS-assisted communication system.

 TABLE I
 COMPARISON BETWEEN THE FAR-FIELD AND NEAR-FIELD REGIONS IN RIS-ASSISTED SYSTEMS

	BS-RIS channel	RIS-user channel
Far-field	$r_{B,R} > \frac{2(D_B+D_R)^2}{\lambda_c}$	$r_{R,U} > \frac{2(D_R+D_U)^2}{\lambda_c}$
Near-field	$r_{B,R} \leq \frac{2(D_B+D_R)^2}{\lambda_c}$	$r_{R,U} \leq \frac{2(D_R+D_U)^2}{\lambda_c}$
Boundary ($d = \lambda_c/2$)	$\left[\frac{(N_B-1) + \sqrt{(M_x-1)^2 + (M_y-1)^2}}{2} \right]^2 \lambda_c$	$\left[\frac{(N_U-1) + \sqrt{(M_x-1)^2 + (M_y-1)^2}}{2} \right]^2 \lambda_c$

response vector is given by

$$\mathbf{a}_R(\alpha, \varphi) = \frac{1}{\sqrt{M}} \left[1, \dots, e^{-j \frac{2\pi d}{\lambda_c} (M_x-1) \sin \alpha \sin \varphi} \right]^T \otimes \left[1, \dots, e^{-j \frac{2\pi d}{\lambda_c} (M_y-1) \cos \varphi} \right]^T, \quad (2)$$

where $0 \leq x \leq (M_x - 1)$ and $0 \leq y \leq (M_y - 1)$, and \otimes is the Kronecker product. For the ULA with N_B elements at the BS, the array response vector is expressed as [35]

$$\mathbf{a}_B(\alpha) = \frac{1}{\sqrt{N_B}} \left[1, \dots, e^{-j(N_B-1) \frac{2\pi d}{\lambda_c} \sin \alpha} \right]^T, \quad (3)$$

where $0 \leq n_b \leq (N_B - 1)$. Since the response vector of a ULA is constant in the elevation domain, we omit the parameter φ in the response vector of the BS in (3).

Similarly, the channel between the RIS and user in the far-field region is given by

$$\mathbf{G}_{R,U}^{\text{far}} = \sqrt{\frac{MN_U}{L_U}} \sum_{l=0}^{L_U-1} \beta_l \mathbf{a}_U(\alpha_l^A) \mathbf{a}_R^H(\alpha_l^D, \varphi_l^D), \quad (4)$$

where L_U is the number of paths between the RIS and user, α_l^A denotes the azimuth AoA associated with the user, α_l^D and φ_l^D denote the azimuth and elevation AoD associated with the RIS respectively, and the array response vectors $\mathbf{a}_R(\alpha_l^D, \varphi_l^D)$ and $\mathbf{a}_U(\alpha_l^A)$ can be generated similarly to (2) and (3).

Remark 1. Assume that the number of RIS elements is far larger than the number of BS and user antennas. When the far-field channel model is considered (no matter whether the BS-RIS link or the RIS-user link follows the far-field model), the rank r of the cascaded BS-RIS-user channel satisfies $r \leq \min\{L_B, L_U\}$, and we say that the channel has r degrees of freedom (DoFs) [36]. Thus, the dimension of the transmitted signal vector should satisfy $q \leq r \leq \min\{L_B, L_U\}$. When environmental scattering is absent and no NLoS paths are present, the rank of the cascaded channel is 1.

2) Near-field channel model: For communications in the near-field domain, we consider a three-dimensional topology.

In FFC, the signal propagation paths from transmitters to receivers can be regarded as parallel to each other if they share the same AoA. However, in NFC, the angle distribution is also related to the distance between specific points of transmitters and receivers. Assume that all the BS antennas are located on the x -axis, where the coordinate of the midpoint of the BS array is denoted as $(0, 0, 0)$. Thus, the coordinate of the n_b -th antenna at the BS can be denoted as

$$\mathbf{q}_B(n_b) = (\tilde{n}_b d, 0, 0), \quad (5)$$

where $\tilde{n}_b = n_b - \frac{N_B - 1}{2}$.

Similarly, assume that all of the user antennas are parallel to the x -axis, where the coordinate of the midpoint of the user array is (x_U, y_U, z_U) . Thus, the location of n_u -th antenna at the user can be denoted as

$$\mathbf{q}_U(n_u) = (x_U + \tilde{n}_u d, y_U, z_U), \quad (6)$$

where $\tilde{n}_u = n_u - \frac{N_U - 1}{2}$.

Assume that all elements of the RIS are parallel to the XY-plane. Denote the coordinate of the midpoint of the RIS as (x_R, y_R, z_R) . Thus, the coordinate of the (m_x, m_y) -th element of the RIS can be expressed as

$$\mathbf{q}_R(m_x, m_y) = (x_R + \tilde{m}_x d, y_R + \tilde{m}_y d, z_R), \quad (7)$$

where $\tilde{m}_x = m_x - \frac{M_x - 1}{2}$ and $\tilde{m}_y = m_y - \frac{M_y - 1}{2}$. Further, assume that the location and topology of the RIS are known.

Based on the above geometry, the LoS link for the near-field channel between BS and RIS can be modeled as [37]

$$\mathbf{G}_{B,R}^{\text{near}} = \left[\mathbf{g}_{B,R}^1, \dots, \mathbf{g}_{B,R}^{m_{xy}}, \dots, \mathbf{g}_{B,R}^M \right]^T, \quad (8)$$

where $\mathbf{g}_{B,R}^{m_{xy}} = \left[\kappa_{m_{xy},1} e^{-j \frac{2\pi}{\lambda_c} d_{m_{xy},1}^B}, \dots, \kappa_{m_{xy},N_B} e^{-j \frac{2\pi}{\lambda_c} d_{m_{xy},N_B}^B} \right]^T$ and $\kappa_{m_{xy},n_b} \propto \frac{1}{d_{m_{xy},n_b}^B}$ denotes the free-space large-scale path loss between the n_b -th antenna of the BS and the (m_x, m_y) -th element of the RIS. The distance between the n_b -th antenna of the BS and the (m_x, m_y) -th element of the RIS is given by

$$d_{m_{xy},n_b}^B = \|\mathbf{q}_B(n_b) - \mathbf{q}_R(m_x, m_y)\|_2, \quad (9)$$

which can be expanded as

$$\begin{aligned} d_{m_{xy},n_b}^B &= \sqrt{(x_R + \tilde{m}_x d - \tilde{n}_b d)^2 + (y_R + \tilde{m}_y d)^2 + z_R^2} \\ &= \sqrt{r_{m_{xy}}^B{}^2 + (\tilde{n}_b d)^2 - 2\tilde{n}_b d r_{m_{xy}}^B \sin \alpha_{m_{xy}}^B \sin \varphi_{m_{xy}}^B} \\ &\stackrel{(a)}{\approx} r_{m_{xy}}^B - \tilde{n}_b d \sin \alpha_{m_{xy}}^B \sin \varphi_{m_{xy}}^B \\ &\quad + \frac{(\tilde{n}_b d)^2 (1 - \sin^2 \alpha_{m_{xy}}^B \sin^2 \varphi_{m_{xy}}^B)}{2r_{m_{xy}}^B}, \end{aligned} \quad (10)$$

with $r_{m_{xy}}^B$ denoting the distance between $(0, 0, 0)$ and the (m_x, m_y) -th element of the RIS, $\alpha_{m_{xy}}^B, \varphi_{m_{xy}}^B$ denoting the corresponding azimuth and elevation angle, respectively, and (a) is obtained based on the first-order Taylor expansion $\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \mathcal{O}(x^3)$.

Similarly, the LoS near-field channel between the RIS and user can be modeled as

$$\mathbf{G}_{R,U}^{\text{near}} = \left[\mathbf{g}_{R,U}^1, \dots, \mathbf{g}_{R,U}^{n_u}, \dots, \mathbf{g}_{R,U}^{N_U} \right]^T, \quad (11)$$

where $\mathbf{g}_{R,U}^{n_u} = \left[\kappa_{n_u,1} e^{-j \frac{2\pi}{\lambda_c} d_{n_u,1}^U}, \dots, \kappa_{n_u,M} e^{-j \frac{2\pi}{\lambda_c} d_{n_u,M}^U} \right]^T$ and $\kappa_{n_u,m_{xy}} \propto \frac{1}{d_{n_u,m_{xy}}^U}$ denote the free-space large-scale path loss between the n_u -th antenna of the user and the (m_x, m_y) -th element of the RIS. The distance between the n_u -th antenna of the user and the (m_x, m_y) -th element of the RIS is given by

$$d_{n_u,m_{xy}}^U = \|\mathbf{q}_U(n_u) - \mathbf{q}_R(m_x, m_y)\|_2, \quad (12)$$

which can be approximated as

$$\begin{aligned} d_{n_u,m_{xy}}^U &\approx r_{m_{xy}}^U - \tilde{n}_u d \sin \alpha_{m_{xy}}^U \sin \varphi_{m_{xy}}^U \\ &\quad + \frac{(\tilde{n}_u d)^2 (1 - \sin^2 \alpha_{m_{xy}}^U \sin^2 \varphi_{m_{xy}}^U)}{2r_{m_{xy}}^U}, \end{aligned} \quad (13)$$

with $r_{m_{xy}}^U$ denoting the distance between the midpoint of the user array and the (m_x, m_y) -th element of the RIS, $\alpha_{m_{xy}}^U, \varphi_{m_{xy}}^U$ denoting the corresponding azimuth and elevation angle, respectively.

Remark 2. When both the BS-RIS and RIS-user links are in the near-field region, the rank of the cascaded LoS channel is $r = \min\{N_B, N_U, M\}$, which will be much larger than 1 in the case of large-scale arrays, even in the absence of NLoS paths. Thus, one of the main advantages of near-field communications is that they possess higher DoFs and can support q ($q > 1$) data streams without relying on abundant environmental scattering¹.

3) *Cascaded channel model:* Denote the phase-shift matrix of the RIS as $\Theta = \text{diag}(\phi_1, \dots, \phi_m, \dots, \phi_M)$, with $\phi_m = e^{-j\theta_m}, \theta_m \in [0, 2\pi)$ denoting the phase-shift coefficient of the m -th RIS element. Therefore, the cascaded channels of the four scenarios mentioned above can be expressed as

$$1) \mathbf{H}_{\text{FF}} = \mathbf{G}_{R,U}^{\text{far}} \Theta \mathbf{G}_{B,R}^{\text{far}}, \quad (14a)$$

$$2) \mathbf{H}_{\text{NF}} = \mathbf{G}_{R,U}^{\text{far}} \Theta \mathbf{G}_{B,R}^{\text{near}}, \quad (14b)$$

$$3) \mathbf{H}_{\text{FN}} = \mathbf{G}_{R,U}^{\text{near}} \Theta \mathbf{G}_{B,R}^{\text{far}}, \quad (14c)$$

$$4) \mathbf{H}_{\text{NN}} = \mathbf{G}_{R,U}^{\text{near}} \Theta \mathbf{G}_{B,R}^{\text{near}}. \quad (14d)$$

B. Signal Model

We assume that fully-digital beamforming is adopted at the BS, and thus the BS transmitted signal is given by

$$\mathbf{s} = \mathbf{W}\mathbf{x}, \quad (15)$$

where $\mathbf{x} \in \mathbb{C}^{q \times 1}$ is the symbol vector, satisfying $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}_q$, and $\mathbf{W} \in \mathbb{C}^{N_B \times q}$ is the beamforming matrix or precoder.

Assuming the receiver adopts a linear combiner, the received signal at the user is given by

$$\mathbf{y} = \mathbf{U}^H \mathbf{H} \mathbf{W} \mathbf{x} + \mathbf{U}^H \mathbf{n}, \quad (16)$$

where $\mathbf{U} \in \mathbb{C}^{N_U \times q}$ is the combining matrix, $\mathbf{H} \in \{\mathbf{H}_{\text{FF}}, \mathbf{H}_{\text{NF}}, \mathbf{H}_{\text{FN}}, \mathbf{H}_{\text{NN}}\}$ is the cascaded channel between

¹In this article, we mainly consider the transmission of LoS path to compare the performance of far- and near-field communications in a non-scattering environment, i.e., $L_B = L_U = 1$. This setting is to demonstrate that NFC can support multiple data streams and have higher DoFs in the same non-scattering environments. In this case, only a single RF chain is required in FFC due to the low rank of far-field channels [38], and $2q$ RF chains are required in NFC to achieve the same performance as fully-digital beamforming [39].

the BS and the user based on the specific channel model, and $\mathbf{n} \in \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_U})$ is additive white Gaussian noise (AWGN).

The maximum achievable data rate (bits/s/Hz) for the user is given by

$$R(\mathbf{W}, \Theta) = \log_2 |\mathbf{I}_{N_U} + \mathbf{H}\mathbf{W}\mathbf{W}^H \mathbf{H}^H \sigma^{-2}|, \quad (17)$$

where $|\mathbf{X}|$ represents the determinant of the matrix \mathbf{X} .

Remark 3. Equation (17) characterizes the maximum achievable rate when the optimal combiner at the user, denoted as \mathbf{U}_{opt} , is adopted. Given $\{\mathbf{W}, \Theta\}$, closed-form solutions for \mathbf{U}_{opt} can be obtained, as detailed in the sequel.

III. RIS CODEBOOK DESIGN

The precise design of the RIS coefficient matrix requires accurate channel information, which is difficult to obtain due to the passivity of RISs. Therefore, in this article, we consider exploiting beam training, which can perform channel estimation and RIS coefficient optimization simultaneously. To begin with, in this section, we will first design codebooks for different channel models. Based on the information contained in the received signals, angular-domain and distance-based codebooks are designed for FF and NN models respectively, while a combined angular-distance (AD) codebook is designed for NF and FN models. Each column in the designed codebook corresponds to one candidate RIS passive beamforming vector.

A. Angular-domain Codebook for FF

In this subsection, we consider an angular-domain codebook that is universal in the far-field domain. Generally, existing RIS codebooks for FF channel models aim at aligning the beam to the strongest propagation path to maximize the beamforming gain. Ignoring weak paths, the signal at the user can be written as

$$\mathbf{y}_{\text{FF}} = \beta_0' \mathbf{U}^H \mathbf{a}_U(\alpha_0^A) \mathbf{a}_R^H(\alpha_0^D, \varphi_0^D) \text{diag}(\phi) \mathbf{a}_R(\alpha_0^A, \varphi_0^A) \mathbf{a}_B^H(\alpha_0^D) \mathbf{W}\mathbf{x} + \mathbf{U}^H \mathbf{n}, \quad (18)$$

where β_0' denotes the complex gain of the cascaded LoS channel. By substituting the array response vectors $\mathbf{a}_R(\alpha_0^D, \varphi_0^D)$ and $\mathbf{a}_R(\alpha_0^A, \varphi_0^A)$ into (18), the user's signal can be reformulated as

$$\mathbf{y}_{\text{FF}} = \beta_0' \mathbf{U}^H \mathbf{a}_U(\alpha_0^A) \sum_{m=1}^M e^{-j(\theta_m + \nu_m^A - \nu_m^D)} \mathbf{a}_B^H(\alpha_0^D) \mathbf{W}\mathbf{x} + \mathbf{U}^H \mathbf{n}, \quad (19)$$

where $\nu_m^A = \frac{2\pi d}{\lambda_c} [(m_x - 1) \sin \alpha_0^A \sin \varphi_0^A + (m_y - 1) \cos \varphi_0^A]$ and $\nu_m^D = \frac{2\pi d}{\lambda_c} [(m_x - 1) \sin \alpha_0^D \sin \varphi_0^D + (m_y - 1) \cos \varphi_0^D]$. Thus, the optimal RIS phase-shift coefficients are $\theta_m = \nu_m^D - \nu_m^A$. All candidate angle pairs on the beam grid should be considered when accurate channel state information is unavailable, which is one of the primary principles for far-field codebook design.

Assume that \mathbf{U} and \mathbf{W} have already been designed to align with the strongest path [20], and focus on coefficient optimization at RIS for codebook design. \mathbf{U} and \mathbf{W} can be designed with the channel information obtained through beam

training, and we will show how \mathbf{U} and \mathbf{W} can be obtained later. Thus, the RIS codebook for the FF channel model can be designed as [29] [31]

$$\mathcal{F}_{\text{FF}} = \left[\mathbf{b}(\beta_1, \delta_1), \dots, \mathbf{b}(\beta_{m_x}, \delta_{m_y}), \dots, \mathbf{b}(\beta_{M_x}, \delta_{M_y}) \right]^*, \quad (20)$$

where $\beta_{m_x} = \frac{2m_x - M_x - 1}{M_x}$ with $m_x = 1, \dots, M_x$, $\delta_{m_y} = \frac{2m_y - M_y - 1}{M_y}$ with $m_y = 1, \dots, M_y$, and $\mathbf{b}(\beta, \delta)$ is the following far-field steering vector

$$\mathbf{b}(\beta, \delta) = \begin{bmatrix} 1, \dots, e^{-j\frac{2\pi d}{\lambda_c}(M_x-1)\beta\delta} \\ 1, \dots, e^{-j\frac{2\pi d}{\lambda_c}(M_y-1)\delta} \end{bmatrix}^T \otimes \quad (21)$$

Each column of \mathcal{F}_{FF} is a candidate beam codeword for the RIS in the far-field domain. This angular-domain codebook design makes full use of the angular information of the far-field paths.

B. Distance-based Codebook Design for NN

Angular-only far-field codebooks are not applicable to near-field channels, even if only one or the other of the BS-RIS and RIS-user links is in the near-field domain because both distance and angular information are embedded in the received signal. Therefore, novel codebooks are required to match the channels in the near-field region. When both the BS-RIS link and RIS-user link are in the near-field domain, the combined signal at the user is given by (22) at the bottom of the next page, where $\nu_{n_u, m, n_b} = \frac{2\pi}{\lambda_c} (d_{n_u, m}^U + d_{m, n_b}^B)$, determined by the sum of the distances from the BS to RIS and from the RIS to the user.

Define the following array steering vector related to the distance

$$\mathbf{f}(x, y, z) = \exp \left[-j\frac{2\pi}{\lambda_c} (r_{1,1}, \dots, r_{m_x, m_y}, \dots, r_{M_x, M_y}) \right]^T, \quad (23)$$

where (x, y, z) denotes the coordinate of the BS or the user, and $r_{m_x, m_y} = \|(x, y, z), \mathbf{q}_R(m_x, m_y)\|_2$. Generally, the height of the BS or the user is constant. Therefore, for the design of the near-field codebooks, we mainly consider the exploration of x and y , and the z component in $\mathbf{f}(x, y, z)$ will be omitted hereafter.

Definition 1. $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ and $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ are two sets of vectors with the same dimension. The \star operator is defined as follows for vectors of equal dimensions:

$$\begin{aligned} & [\mathbf{x}_1, \dots, \mathbf{x}_M] \star [\mathbf{y}_1, \dots, \mathbf{y}_N] \\ &= [\mathbf{x}_1 \odot \mathbf{y}_1, \dots, \mathbf{x}_1 \odot \mathbf{y}_N, \dots, \mathbf{x}_M \odot \mathbf{y}_N], \end{aligned} \quad (24)$$

where \odot denotes the Hadamard product.

We use a pair of coordinates to represent the rectangular sampling range on the XY-plane as (x_{\min}^k, y_{\min}^k) and (x_{\max}^k, y_{\max}^k) , where x_{\min}^k (y_{\min}^k) and x_{\max}^k (y_{\max}^k) are the minimum and the maximum x (y)-coordinate of the sampling points respectively, and $k \in \{\text{B}, \text{U}\}$ indicates the BS or user. Denote the number of points sampled on the x -axis and y -axis as S_x and S_y , respectively. Taking the corresponding distance

between the sampling points into account, the codebook at the RIS for the NN channel model is given by

$$\mathcal{F}_{\text{NN}} = \left[\mathbf{f}(x_1^{\text{U}}, y_1^{\text{U}}), \dots, \mathbf{f}(x_{S_x}^{\text{U}}, y_{S_x}^{\text{U}}), \dots, \mathbf{f}(x_{S_x}^{\text{U}}, y_{S_y}^{\text{U}}) \right]^* \star \left[\mathbf{f}(x_1^{\text{B}}, y_1^{\text{B}}), \dots, \mathbf{f}(x_{S_x}^{\text{B}}, y_{S_x}^{\text{B}}), \dots, \mathbf{f}(x_{S_x}^{\text{B}}, y_{S_y}^{\text{B}}) \right]^* \quad (25)$$

where $x_{s_x}^k$ and $y_{s_x}^k$ are given by

$$x_{s_x}^k = x_{\min}^k + \left(s_x - \frac{1}{2} \right) \frac{x_{\max}^k - x_{\min}^k}{S_x}, \quad s_x = 1, \dots, S_x, \quad (26a)$$

$$y_{s_y}^k = y_{\min}^k + \left(s_y - \frac{1}{2} \right) \frac{y_{\max}^k - y_{\min}^k}{S_y}, \quad s_y = 1, \dots, S_y. \quad (26b)$$

The size of the codebook will increase with the number of sampling points in each direction, and the accuracy of the RIS beamforming will improve correspondingly.

C. Combined AD Codebook Design for Hybrid FFC & NFC

In this section, we consider hybrid cascaded channel models, where one side of the RIS is in the far-field and the other is in the near-field. As such, the link on one side of RIS only needs to consider angular information, while the link on the other side needs to consider both angular and distance information. Taking the FN channel model as an example, i.e., with the BS-RIS link in the far-field and the RIS-user link in the near-field, the received signal at the user is given by

$$\mathbf{y}_{\text{FN}} = \mathbf{U}^H \mathbf{G}_{\text{R,U}}^{\text{near}} \text{diag}(\phi) \mathbf{G}_{\text{B,R}}^{\text{far}} \mathbf{W} \mathbf{x} + \mathbf{U}^H \mathbf{n}. \quad (27)$$

By substituting $\mathbf{G}_{\text{R,U}}^{\text{near}}$ in (11) and $\mathbf{G}_{\text{B,R}}^{\text{far}}$ in (1) into (27), \mathbf{y}_{FN} can be expanded into the following form

$$\mathbf{y}_{\text{FN}} = \frac{\beta_0 \lambda_c}{4\pi \sqrt{M}} \mathbf{U}^H \left[\sum_{m=1}^M \frac{1}{d_{1,m}^{\text{U}}} e^{-j(\theta_m + \nu_{1,m}^{\text{U}})}, \dots, \sum_{m=1}^M \frac{1}{d_{N_{\text{U}},m}^{\text{U}}} e^{-j(\theta_m + \nu_{N_{\text{U}},m}^{\text{U}})} \right]^T \mathbf{a}_{\text{B}}^H(\alpha_0^{\text{D}}) \mathbf{W} \mathbf{x} + \mathbf{U}^H \mathbf{n}, \quad (28)$$

where

$$\nu_{n_u,m}^{\text{A}} = \frac{2\pi}{\lambda_c} \left[(m_x - 1) d \sin \alpha_0^{\text{A}} \sin \varphi_0^{\text{A}} + (m_y - 1) d \cos \varphi_0^{\text{A}} + d_{n_u,m}^{\text{U}} \right]. \quad (29)$$

Similar to the assumptions above for FF channels, we assume that the transmit or receive beam related to the far-field link has

been designed to align with the steering vector of the strongest path. That is to say, \mathbf{W} has been aligned to $\mathbf{a}_{\text{B}}^H(\alpha_0^{\text{D}})$ when designing the RIS codebook.

Considering the array response in both the distance and angular domains, we define the array steering vector as

$$\mathbf{c}(x, y, \beta, \delta) = \left[e^{-j \frac{2\pi}{\lambda_c} d_{1,1}}, \dots, e^{-j \frac{2\pi}{\lambda_c} d_{M_x, M_y}} \right]^T \odot \left\{ \left[1, \dots, e^{-j \frac{2\pi d}{\lambda_c} (M_x - 1) \beta \delta} \right]^T \otimes \left[1, \dots, e^{-j \frac{2\pi d}{\lambda_c} (M_y - 1) \delta} \right]^T \right\}, \quad (30)$$

where $d_{m_x, m_y} = \|(x, y, z), \mathbf{q}_{\text{R}}(m_x, m_y)\|_2$ and z is assumed to be fixed and known.

A codebook design for hybrid cascaded channel models should cover not only the angular region of all possible signals, but their distances as well. Therefore, based on the received signal model given in (27), the codebook for hybrid cascaded channel models should be designed to include the following two components.

1) *Angular-domain sweeping, similar to FFC*: To align the phase on the FFC side, the full angular range will be covered in the predesigned codebook. Therefore, the first component of the RIS codebook for hybrid cascaded channel models is given by

$$\mathcal{F}_{\text{Hb}}^1 = \left[\mathbf{b}(\beta_1, \delta_1), \dots, \mathbf{b}(\beta_{m_x}, \delta_{m_y}), \dots, \mathbf{b}(\beta_{M_x}, \delta_{M_y}) \right]^*, \quad (31)$$

where $\beta_{m_x} = \frac{2m_x - M_x - 1}{M_x}$ with $m_x = 1, \dots, M_x$, $\delta_{m_y} = \frac{2m_y - M_y - 1}{M_y}$ with $m_y = 1, \dots, M_y$.

2) *Distance-domain sampling, similar to NFC*: For each given angle pair β and δ , all possible sampling points in the feasible range should be included to align the phase on the NFC side. Therefore, the second component of the RIS codebook for hybrid cascaded channel models is given by

$$\mathcal{F}_{\text{Hb}}^{2,k} = \left[\mathbf{f}(x_1^k, y_1^k), \dots, \mathbf{f}(x_{s_x}^k, y_{s_x}^k), \dots, \mathbf{f}(x_{S_x}^k, y_{S_y}^k) \right]^*, \quad (32)$$

where $x_{s_x}^k$ and $y_{s_x}^k$, $k \in \{\text{B}, \text{U}\}$, are defined as in (26). In particular, $k = \text{B}$ for the NF case, and $k = \text{U}$ for the FN case.

Accordingly, the codebook at the RIS for hybrid cascaded channel models can be designed as

$$\mathcal{F}_{\text{NF}} = \mathcal{F}_{\text{Hb}}^1 \star \mathcal{F}_{\text{Hb}}^{2,\text{B}}, \quad (33a)$$

$$\mathcal{F}_{\text{FN}} = \mathcal{F}_{\text{Hb}}^1 \star \mathcal{F}_{\text{Hb}}^{2,\text{U}}, \quad (33b)$$

where the definition of \star is given in (24).

$$\mathbf{y}_{\text{NN}} = \left(\frac{\lambda_c}{4\pi} \right)^2 \mathbf{U}^H \begin{bmatrix} \sum_{m=1}^M \frac{1}{d_{1,m}^{\text{U}} d_{m,1}^{\text{B}}} e^{-j(\theta_m + \nu_{1,m,1})} & \dots & \sum_{m=1}^M \frac{1}{d_{1,m}^{\text{U}} d_{m,N_{\text{B}}}^{\text{B}}} e^{-j(\theta_m + \nu_{1,m,N_{\text{B}}})} \\ \vdots & & \vdots \\ \sum_{m=1}^M \frac{1}{d_{n_u,m}^{\text{U}} d_{m,1}^{\text{B}}} e^{-j(\theta_m + \nu_{n_u,m,1})} & \dots & \sum_{m=1}^M \frac{1}{d_{n_u,m}^{\text{U}} d_{m,N_{\text{B}}}^{\text{B}}} e^{-j(\theta_m + \nu_{n_u,m,N_{\text{B}}})} \\ \vdots & & \vdots \\ \sum_{m=1}^M \frac{1}{d_{N_{\text{U}},m}^{\text{U}} d_{m,1}^{\text{B}}} e^{-j(\theta_m + \nu_{N_{\text{U}},m,1})} & \dots & \sum_{m=1}^M \frac{1}{d_{N_{\text{U}},m}^{\text{U}} d_{m,N_{\text{B}}}^{\text{B}}} e^{-j(\theta_m + \nu_{N_{\text{U}},m,N_{\text{B}}})} \end{bmatrix} \mathbf{W} \mathbf{x} + \mathbf{U}^H \mathbf{n}, \quad (22)$$

IV. AO-BASED ALGORITHM FOR MAXIMIZING RATE

We aim to maximize the achievable rate by jointly optimizing the combining matrix \mathbf{U} at the user, the digital beamforming matrix \mathbf{W} at the BS, and the reflecting coefficient matrix Θ at the RIS. Suppose each passive beamforming vector at the RIS is selected from the predesigned codebooks. Therefore, the mathematical form of the optimization problem can be given as

$$(P) \max_{\mathbf{W}, \phi} R(\mathbf{W}, \text{diag}(\phi)), \quad (34a)$$

$$s.t. \|\mathbf{W}\|_F^2 \leq P_{\max}, \quad (34b)$$

$$\phi \in \mathcal{F}, \quad (34c)$$

where $\mathcal{F} \in \{\mathcal{F}_{FF}, \mathcal{F}_{NF}, \mathcal{F}_{FN}, \mathcal{F}_{NN}\}$ in (34c) is the pre-designed codebook based on the specific channel model.

To handle the highly-coupled non-convex problem in (34), we propose an alternating optimization algorithm in this section, where three sub-problems will be solved in an iterative manner. First, based on the predesigned codebooks, different beam training schemes will be proposed to perform the channel estimation and the RIS optimization for the specific channel models. Second, a closed-form solution for the receive combining matrix at the user is obtained. Finally, the active beamforming problem is solved by exploiting the relationship between the achievable rate and the mean square error. Details for solving these sub-problems are given in the following subsections.

A. Beam Training Design for RIS

1) *Angular-domain Beam Sweeping for FF* : The most direct method to perform beam training for this case is sweeping through all possible angles if the resulting training overhead is acceptable. The size of the FF codebook in (20) depends on the number of RIS reflecting elements. Based on the given angle pair $\langle \beta_{m_x}, \delta_{m_y} \rangle$, the corresponding codeword is $\mathbf{b}(\beta_{m_x}, \delta_{m_y})$, and the achievable rate can be calculated as

$$R(\mathbf{W}, \text{diag}(\mathbf{b}(\beta_{m_x}, \delta_{m_y}))), \quad (35)$$

with given beamforming matrix \mathbf{W} . The optimal codeword is selected as the one that yields the maximum rate after sweeping through all candidate codewords in (20).

2) *Hierarchical Beam Training Scheme for NN* : From (25) and (26), we see that the NN codebook size is determined by the number of sampled points on the x -axis and y -axis. The beam training overhead can be reduced by reducing the codebook size (i.e., increasing the step size between the codebook samples), but this will reduce the performance of the beam training. In order to solve this problem, we design a hierarchical beam training approach that consists of several different levels of sub-codebooks. These sub-codebook levels are determined using different sampling ranges and sampling step sizes.

As illustrated in Fig. 2, we divide the sampling range parallel to the XY-plane into four sub-ranges in each layer, and find the range corresponding to the optimal codeword as the sampling range of the next layer. More particularly, in the l -th layer, the rectangular sampling range defined by the

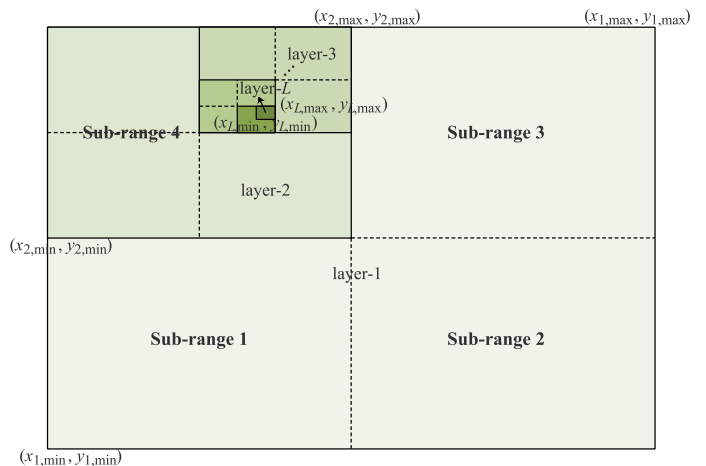


Fig. 2. Illustration for the proposed hierarchical beam training procedure.

coordinate pairs (x_{\min}^k, y_{\min}^k) and (x_{\max}^k, y_{\max}^k) , $k \in \{B, U\}$, is divided into the following four sub-ranges:

$$1) (x_{\min}^k, y_{\min}^k), (x_{\min}^k + x_{\Delta}, y_{\min}^k + y_{\Delta}), \quad (36a)$$

$$2) (x_{\min}^k + x_{\Delta}, y_{\min}^k), (x_{\max}^k, y_{\min}^k + y_{\Delta}), \quad (36b)$$

$$3) (x_{\min}^k + x_{\Delta}, y_{\min}^k + y_{\Delta}), (x_{\max}^k, y_{\max}^k), \quad (36c)$$

$$4) (x_{\min}^k, y_{\min}^k + y_{\Delta}), (x_{\min}^k + x_{\Delta}, y_{\max}^k), \quad (36d)$$

where $x_{\Delta} = \frac{1}{2} [x_{\max}^k - x_{\min}^k]$ and $y_{\Delta} = \frac{1}{2} [y_{\max}^k - y_{\min}^k]$. Therefore, the sub-codebook for sub-range i of B and sub-range j of U in the l -th layer, denoted by $\mathcal{F}_{NN}(l, i, j)$, $i, j = 1, 2, 3, 4$, can be generated based on (25). Each column in $\mathcal{F}_{NN}(l, i, j)$ corresponds to a codeword $\phi_{l, i, j, s}$, with s denoting the column index in sub-codebook $\mathcal{F}_{NN}(l, i, j)$. Based on the given beamforming matrix \mathbf{W} and codeword $\phi_{l, i, j, s}$, the achievable rate can be calculated as

$$R(\mathbf{W}, \text{diag}(\phi_{l, i, j, s})). \quad (37)$$

In fact, R can be calculated based on the corresponding received signal strength without accurate channel information.

By adopting the ES method, the optimal codeword can be obtained for the first layer, as well as the optimal sampling sub-range pair $\langle i, j \rangle$, which will be the sampling range in the next layer. This process is repeated until the maximum number of layers L is reached.

The hierarchical beam training procedure is summarized in **Algorithm 1**. We define the size of the codebook as the number of candidate codewords, i.e., the number of column vectors. There are 16 sub-codebooks in each layer, and the size of each sub-codebook is $(S_x S_y)^2$. Therefore, with a given maximum number of training layers L , the training overhead can be calculated as $16L(S_x S_y)^2$. For the same sampling grid resolution and hence beam training accuracy, the training overhead of ES will be $4^{L+1}(S_x S_y)^2$.

3) *Two-Stage Beam Training Scheme for FN and NF* : The size of codebooks \mathcal{F}_{NF} and \mathcal{F}_{FN} is determined by the product of the number of RIS elements and the number of sample points, which can be extremely large especially for large RISs. Thus, while ES is a straightforward way to find

Algorithm 1 Hierarchical Beam Training Scheme for NN

Input: sampling range $x(y)_{\min(\max)}^k$, $k \in \{B, U\}$, number of sampled points in each layer S_x and S_y , maximum layer L_{\max} .

Output: Θ_{opt} , R_{opt} .

- 1: **Initialization:** $l = 1$, $x(y)_{\min(\max)}^k(l) = x(y)_{\min(\max)}^k$, $k \in \{B, U\}$, $R_{\text{opt}} = 0$.
- 2: **repeat**
- 3: Divide the sampling ranges corresponding to both the BS and the user into four sub-ranges according to (36);
- 4: Generate sub-codebooks $\mathcal{F}_{\text{NN}}(l, i, j)$, $i, j = 1, 2, 3, 4$, based on (25), where i is the sampling sub-range index for B and j is the sampling sub-range index for U;
- 5: **Based on the received signal strength, calculate the corresponding $R(\mathbf{W}, \Theta)$ for each codeword;**
- 6: Find the maximum $R(\mathbf{W}, \Theta)$, the corresponding codeword Θ_{opt} , and the corresponding optimal sampling sub-range index pair $\langle i, j \rangle$;
- 7: Update $x(y)_{\min(\max)}^B(l+1) = x(y)_{\min(\max)}^B(l, i)$;
- 8: Update $x(y)_{\min(\max)}^U(l+1) = x(y)_{\min(\max)}^U(l, j)$;
- 9: $l \leftarrow l + 1$;
- 10: **until** $l = L_{\max}$.

the optimal codeword from the predefined codebooks, the training overhead and consumed time can be unacceptable. We therefore propose a two-stage beam training scheme for the FN and NF channel models. Specifically, all candidate angles are swept through in *stage one*, and the distance is determined in *stage two* based on the selected direction. The details are given as follows.

Stage 1 Angular-domain beam sweeping: In the first stage, we aim to find the optimal beamforming angle based on the far-field codebook given in (20), with the objective of maximizing the achievable rate in (17). All sampled angles will be swept through in this stage, and the one yielding the maximum achievable rate will be selected. Therefore, the selected indices in the angular-domain sweeping stage can be expressed mathematically as

$$\langle m_x, m_y \rangle_{\text{opt}} = \operatorname{argmax} \left\{ \mathbf{b}(\beta_{m_x}, \delta_{m_y}) \in \mathcal{F}_{\text{FF}} \mid R(\mathbf{W}, \Theta_{m_x, m_y}) \right\}, \quad (38)$$

where $\Theta_{m_x, m_y} = \operatorname{diag}(\mathbf{b}(\beta_{m_x}, \delta_{m_y}))$. Denote the optimal index pair as $\langle m'_x, m'_y \rangle$.

Stage 2 Distance-domain sampling: In NFC, the beam can be designed to point at a particular location, rather than just a specific angle. Thus, in the second stage, we aim to determine the optimal focusing point for the optimal direction. With a given angle pair $\langle m'_x, m'_y \rangle$, the selected sampling point in the distance domain can be determined by

$$\langle x_{s_x}^k, y_{s_y}^k \rangle_{\text{opt}} = \operatorname{argmax} \left\{ \left(x_{s_x}^k, y_{s_y}^k \right) \in \mathcal{D}_k \mid R(\mathbf{W}, \Theta_{x_{s_x}^k, y_{s_y}^k}) \right\}, \quad (39)$$

where \mathcal{D}_k is the sampling range for k , $k \in \{B, U\}$, and $\Theta_{x_{s_x}^k, y_{s_y}^k} = \operatorname{diag}(\mathbf{c}(x_{s_x}^k, y_{s_y}^k, \beta_{m'_x}, \delta_{m'_y}))$.

Similar to the analysis in Section IV-A2, we adopt a hierarchical beam training scheme in this stage to strike a balance

Algorithm 2 Two-Stage Beam Training Scheme for Hybrid Cascaded Channel Models

Input: M_x, M_y , sampling range $x(y)_{\min(\max)}^k$, $k \in \{B, U\}$, number of sampled points in each layer S_x and S_y , maximum layer L_{\max} .

Output: Θ_{opt} , R_{opt} .

- 1: **Initialization:** $m_x = 1$, $m_y = 1$, $R_{\text{opt}} = 0$.
- 2: **Stage 1:**
- 3: Generate sub-codebook $\mathcal{F}_{\text{Hb}}^1$ based on (31);
- 4: **repeat**
- 5: **repeat**
- 6: **if** $R(\mathbf{W}, \Theta_{m_x, m_y}) > R_{\text{opt}}$ **then**
- 7: $R_{\text{opt}} = R(\mathbf{W}, \Theta_{m_x, m_y})$;
- 8: $\langle m_x, m_y \rangle_{\text{opt}} = \langle m_x, m_y \rangle$;
- 9: **end if**
- 10: $m_x \leftarrow m_x + 1$;
- 11: **until** $m_x = M_x$;
- 12: $m_y \leftarrow m_y + 1$.
- 13: **until** $m_y = M_y$;
- 14: **Stage 2:**
- 15: **repeat**
- 16: Divide the sampling ranges corresponding to the BS (NF Case) or the user (FN Case) into four sub-ranges according to (36);
- 17: Generate sub-codebooks $\mathcal{F}_{\text{Hy}}^2(l, i)$, $i = 1, 2, 3, 4$, based on (32), where i is the sampling sub-range index for B (NF Case) or U (FN Case);
- 18: Generate $\mathcal{F}_{\text{NF/FN}} = \{ \mathcal{F}_{\text{Hb}}^1 \}_{\langle m_x, m_y \rangle_{\text{opt}}} \star \mathcal{F}_{\text{Hb}}^{2, k}$;
- 19: **Based on the received signal strength, calculate the corresponding $R(\mathbf{W}, \Theta_{x_{s_x}^k, y_{s_y}^k})$ for each codeword;**
- 20: Find the maximum $R(\mathbf{W}, \Theta_{x_{s_x}^k, y_{s_y}^k})$, the corresponding codeword Θ_{opt} , and the corresponding optimal sampling sub-range index i ;
- 21: Update $x(y)_{\min(\max)}^k(l+1) = x(y)_{\min(\max)}^k(l, i)$;
- 22: $l \leftarrow l + 1$;
- 23: **until** $l = L_{\max}$.

between the training overhead and accuracy. Thus, the two-stage beam training procedure is summarized in **Algorithm 2**. The overhead of the proposed two-stage beam training scheme is $M + 4LS_xS_y$, while that of the ES method is $M \times 4LS_xS_y$.

Remark 4. The proposed beam training algorithms can be easily extended to scenarios with multiple users by dividing the RIS into multiple sub-surfaces, where each sub-surface implements the proposed beam training schemes to realize beam alignment for each user.

B. Closed-form Solution for U

With the estimated channel information acquired through the beam training, we can perform the BS beamforming and the user combining design. Although the combining matrix \mathbf{U} is not present in the objective function (17), it will have an impact on the BS beamforming design and the achievable rate. Therefore, we derive the optimal combining matrix \mathbf{U} at the user below. The principle of combining matrix design is to recover the original transmitted signals, which can be

transformed into a mean square error (MSE) minimization problem. The MSE matrix of the user is given by

$$\begin{aligned} \mathbf{E} &= \mathbb{E} \left[(\mathbf{y} - \mathbf{x})(\mathbf{y} - \mathbf{x})^H \right] \\ &= (\mathbf{U}^H \mathbf{H} \mathbf{W} - \mathbf{I}_q) (\mathbf{U}^H \mathbf{H} \mathbf{W} - \mathbf{I}_q)^H + \sigma^2 \mathbf{U}^H \mathbf{U}, \end{aligned} \quad (40)$$

and thus the mathematical form of the combining matrix optimization problem can be written as

$$(P1) \min_{\mathbf{U}} \text{Tr}(\mathbf{E}). \quad (41)$$

With given \mathbf{W} and Θ , the optimal solution \mathbf{U} of problem (41) can be obtained by solving $\partial \text{Tr}(\mathbf{E}) / \partial \mathbf{U} = \mathbf{0}$, where $\frac{\partial \text{Tr}(\mathbf{E})}{\partial \mathbf{U}} = 2(\mathbf{H} \mathbf{W} \mathbf{W}^H \mathbf{H}^H + \sigma^2 \mathbf{I}_{N_U}) \mathbf{U} - 2 \mathbf{H} \mathbf{W}$. After some simple derivations, the optimal combining matrix \mathbf{U} is found to be

$$\mathbf{U}_{\text{opt}} = (\mathbf{H} \mathbf{W} \mathbf{W}^H \mathbf{H}^H + \sigma^2 \mathbf{I}_{N_U})^{-1} \mathbf{H} \mathbf{W}, \quad (42)$$

which is necessary for the optimal value of \mathbf{W} to achieve the maximum of (17).

C. Active Beamforming Matrix Optimization

To solve the transmit beamforming optimization problem, we introduce an auxiliary matrix $\mathbf{F} \succeq \mathbf{0}$, and reformulate the original optimization as follows by exploiting the relationship between the achievable rate and the MSE [40]

$$(P') \max_{\mathbf{F}, \mathbf{W}, \phi} f(\mathbf{F}, \mathbf{W}, \phi), \quad (43)$$

s.t. (34b), (34c),

where $f(\mathbf{F}, \mathbf{W}, \phi)$ is given by

$$f(\mathbf{F}, \mathbf{W}, \phi) = \log |\mathbf{F}| - \text{Tr}(\mathbf{F} \mathbf{E}). \quad (44)$$

The optimal \mathbf{F} of problem (43) can be obtained by solving $\partial f / \partial \mathbf{F} = \mathbf{0}$, which is given by $\mathbf{F} = \mathbf{E}^{-1}$. With given \mathbf{F} , \mathbf{U} and ϕ , the beamforming optimization problem can be expressed by substituting \mathbf{E} into (44) and removing the constant terms:

$$(P2) \min_{\mathbf{W}} \text{Tr}(\mathbf{W}^H \mathbf{H}^H \mathbf{U} \mathbf{F} \mathbf{U}^H \mathbf{H} \mathbf{W}) - 2 \text{Tr}(\Re(\mathbf{F} \mathbf{U}^H \mathbf{H} \mathbf{W}))$$

s.t. (34b),

(45)

which is a convex optimization problem and its optimal solution can be obtained using standard methods.

D. Overall Algorithm

The overall alternating optimization algorithm for maximizing the achievable rate is summarized in **Algorithm 3**, where the optimal solutions in each iteration are the inputs for the next iteration. First, with given $\mathbf{U}^{(t)}$ and $\Theta^{(t)}$ at the t -th iteration, the optimal $\Theta^{(t+1)}$ can be obtained by carrying out the beam training scheme based on **Algorithm 2** or **Algorithm 1**. The estimated channel information can also be obtained in this step. Then, with given $\Theta^{(t+1)}$ and $\mathbf{W}^{(t)}$, we calculate and update the combining matrix $\mathbf{U}^{(t+1)}$ according to (42). Although \mathbf{U} is not included in the expression of the objective function in (17), it will serve as input for the next step in solving \mathbf{W} . Last, with given $\mathbf{U}^{(t+1)}$, $\Theta^{(t+1)}$ and $\mathbf{W}^{(t)}$, we first obtain the reformulated objective function

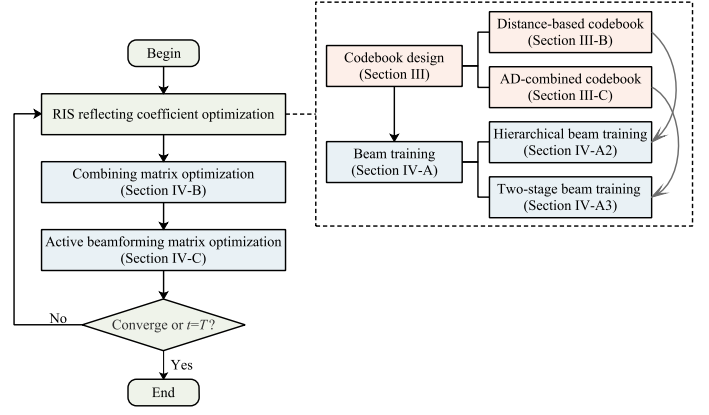


Fig. 3. Illustration for the proposed algorithm.

Algorithm 3 Alternating Optimization-based Algorithm for Maximizing Achievable Rate

Input: $R(\mathbf{W}, \Theta)$, maximum iteration times T , convergence threshold ζ .

Output: $\mathbf{U}_{\text{opt}}, \Theta_{\text{opt}}, \mathbf{W}_{\text{opt}}$.

- 1: **Initialization:** $t = 0$, $\mathbf{U}^{(t)}$, $\Theta^{(t)}$, $\mathbf{W}^{(t)}$, $R^{(t)}(\mathbf{W}, \Theta)$.
- 2: **while** $t \leq T$ and $\Gamma \geq \zeta$ **do**
- 3: For given $\mathbf{U}^{(t)}$ and $\mathbf{W}^{(t)}$, perform beam training based on **Algorithm 2** or **Algorithm 1**, and obtain $\Theta^{(t+1)}$;
- 4: For given $\Theta^{(t+1)}$ and $\mathbf{W}^{(t)}$, calculate and update $\mathbf{U}^{(t+1)}$ according to (42);
- 5: For given $\mathbf{U}^{(t+1)}$, $\Theta^{(t+1)}$ and $\mathbf{W}^{(t)}$, update $\mathbf{W}^{(t+1)}$ by solving problem (45);
- 6: Calculate $R(\mathbf{W}^{(t+1)}, \Theta^{(t+1)})$;
- 7: Calculate $\Gamma = \frac{|R(\mathbf{W}^{(t+1)}, \Theta^{(t+1)}) - R(\mathbf{W}^{(t)}, \Theta^{(t)})|}{|R(\mathbf{W}^{(t)}, \Theta^{(t)})|}$;
- 8: $t \leftarrow t + 1$;
- 9: **end while**

by substituting $\mathbf{E}^{(t)}$ in (40) into (44), and then obtain the optimized beamforming matrix $\mathbf{W}^{(t+1)}$ by solving problem (45). This process is repeated until a maximum number of iterations is reached or the convergence condition is satisfied. The proposed algorithm is summarized in Fig. 3.

The computational complexity of the beam training mainly comes from sorting, which requires $\mathcal{O}(M)$ operations for angular-domain beam sweeping, $\mathcal{O}(16L(S_x S_y)^2)$ operations for the hierarchical beam training scheme, and $\mathcal{O}(M + 4LS_x S_y)$ operations for the two-stage beam training algorithm. Solving problem (P1) for the combining matrix involves matrix inversion, leading to a computational complexity of $\mathcal{O}(N_U^3)$ [41]. According to [42], the worst-case complexity for solving problem (P2) is $\mathcal{O}(q^3 N_B^3)$ using efficient interior-point methods. Thus, the overall computational complexity of the proposed algorithm is $\mathcal{O}(I_3(M + q^3 N_B^3 + N_U^3))$ for FF with angular-domain beam sweeping, $\mathcal{O}(I_3(16L(S_x S_y + q^3 N_B^3 + N_U^3)))$ for NN with hierarchical beam training, and $\mathcal{O}(I_3(M + 4LS_x S_y + q^3 N_B^3 + N_U^3))$ for NF and FN with two-stage beam training, with I_3 representing the number of iterations required for the convergence of **Algorithm 3**.

TABLE II
SIMULATION CONFIGURATIONS

Parameter	Value
Carrier frequency	30GHz
Noise power	-105dBm
Power budget at the BS	30 dBm
Transmitted signal dimension q of NN	2
Transmitted signal dimension q of FF/NF/FN	1
BS's coordinate	(0, 0, 0)m
User's coordinate	(24, 0, 0)m
RIS's coordinate	(12, 0, 8)m
Number of BS antennas	8
Number of user antennas	8
Number of RIS elements along x/y -axis	60, 2
Antenna spacing	$\lambda_c/2$

V. NUMERICAL RESULTS

The simulation configurations listed in TABLE II are used unless stated otherwise. The following parameter settings were used when executing the beam training schemes for both the hierarchical beam training algorithm for the NN channel model and the second beam training stage for hybrid cascaded channel models. For hierarchical beam training, the sampled range in each direction is set as $[-1000\lambda_c, 1000\lambda_c]$. For example, denote the coordinate of the BS antenna midpoint as (x_{BS}, y_{BS}, z_{BS}) , where z_{BS} is fixed. The sampled ranges of the BS in the x - and y -directions are $[x_{BS} - 1000\lambda_c, x_{BS} + 1000\lambda_c]$ and $[y_{BS} - 1000\lambda_c, y_{BS} + 1000\lambda_c]$, respectively. The same applies to the sampled ranges of the user. The numbers of sampled points for hierarchical beam training in the x - and y -directions are both set as $S_x = S_y = 2$. The simulation results are obtained by averaging over 500 channel realizations.

To compare the impact of different channel modeling methods and validate the effectiveness of proposed approach, we conduct numerical simulations on the following schemes:

- NN, Perfect CSI: upper bound, where RIS beamforming is performed with perfect CSI [43];
- NN, Hierarchical: NN channel model, proposed distance-based codebook and hierarchical beam training scheme;
- NN, Uniform ES: NN channel model, the whole sampled range is uniformly divided into several sub-ranges, and the optimal codeword is obtained by exhaustive search²;
- NN, Mismatch: NN channel model, angular-domain codebook and beam sweeping;
- NF/FN, Two-stage: hybrid-far-near-field channel model, proposed combined AD codebook and two-stage beam training scheme;
- FF, Angular-domain: FF channel model, angular-domain codebook and beam sweeping.

To determine the number of training layers of the hierarchical beam training scheme, we show the relationship between the achievable rate and the number of training layers in Fig. 4. It can be seen that when $L_{\max} \geq 12$, the achievable rate no longer increases with L_{\max} . Although the achievable rate

²If the sampling step is too small, the cost of the exhaustive search is unacceptable for the communication system. Therefore, here we divide the entire sampling range into 16 equal intervals in the x and y directions, respectively.

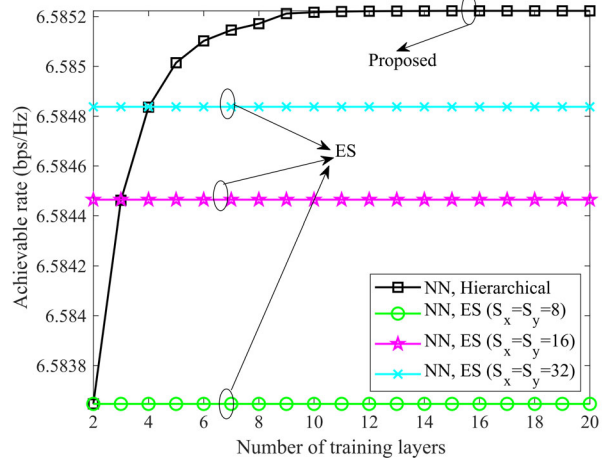


Fig. 4. Achievable rate vs. training layers of hierarchical beam training scheme.

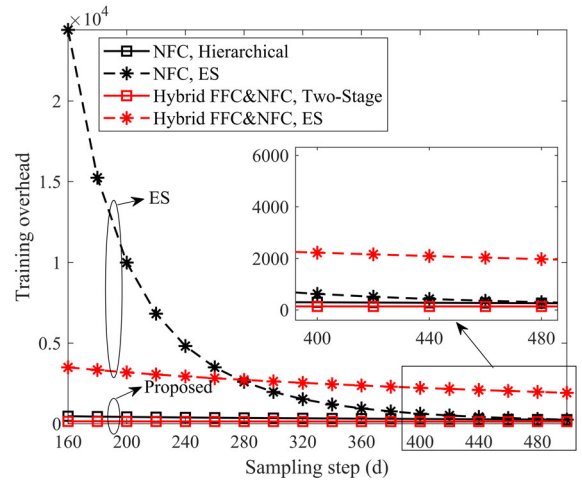


Fig. 5. Beam training overhead vs. sampling step size in multiples of d .

only improves slightly from $L_{\max} = 2$ to $L_{\max} = 12$, we set the total number of training layers hereafter as $L_{\max} = 12$ for better performance. Also, we depict the achievable rate obtained by the ES approach when the number of sampled points is set as 8 and 16 in each direction. The size of the NFC codebook is $(S_x S_y)^2$, so it can be extremely large for large S_x and S_y . Thus, we only considered two scenarios with small S_x and S_y to compare the performance between the proposed hierarchical beam training and ES approaches. Note that we divide the whole sampling range into two sub-ranges in each direction, so there are 4 sampled points in each direction when $S_x = S_y = 2$ for hierarchical beam training. Therefore, the proposed hierarchical algorithm achieves the same accuracy as ES with $S_x = S_y = 8$ when the number of training layers is 2, and it is also the same for $S_x = S_y = 16$ when the number of training layers is 3 and for $S_x = S_y = 32$ when the number of training layers is 4, which verifies the effectiveness of our proposed approach which has significantly reduced training costs.

The required beam training overhead of various training algorithms for different sampling step sizes is compared in Fig. 5. For NN channel modeling, the proposed hierarchical

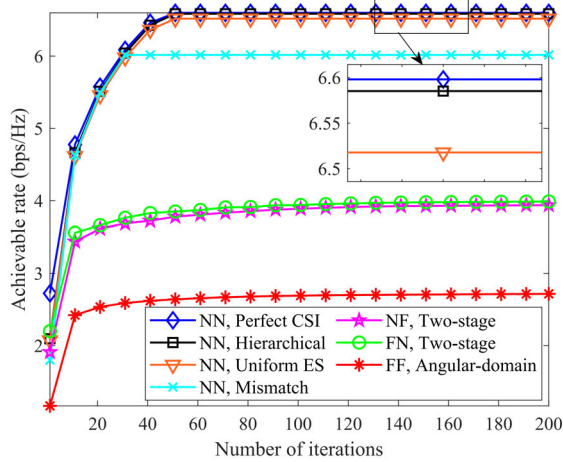


Fig. 6. Convergence performance of the proposed algorithm.

near-field beam training approach greatly reduces the training overhead compared to ES beam training, especially when the sampling step size is small. When the step size is $160d$, the overhead of the hierarchical beam training approach is only about 2% of that required for ES. When the step size is $500d$, these two schemes have the same training overhead, and the required number of training layers for hierarchical beam training is 1. For hybrid FFC & NFC channel models, hierarchical beam training always requires less overhead than ES, regardless of the sampling step size.

The convergence performance of the proposed AO-based algorithm is shown in Fig. 6. Based on the simulation parameters provided earlier, we can calculate that both the BS-RIS and RIS-user links are within the near-field region. Since the NN channel model with hierarchical beam training is more accurate since it considers both angular and distance information of the cascaded links, it will provide a higher achievable rate. Furthermore, it can be seen that the obtained rate by the proposed hierarchical beam training is very close to that of the RIS beamforming with perfect CSI, while the computational complexity significantly reduced. Although the uniform ES scheme is also very close to the upper bound, its overhead is much higher than the proposed scheme. Even though the NN channel model is adopted, the performance obtained with mismatched beam training is poor, because distance information is not considered. The FF channel model has the worst performance due to the accumulative channel estimation error of links on both sides of the RIS.

For the NN channel model, the normalized received signal strength is depicted in Fig. 7. We can see that the maximum received power appears at coordinate (24, 0), which is exactly where the user is located. Since the user's antennas are distributed parallel to the x -axis, the received signal strength is also dispersed along the x -axis. Moreover, the RIS reflecting elements are more distributed along the x -axis direction, which will also lead to such a distribution of reflected signals. Unlike FFC, NFC can achieve beam focusing, concentrating the energy of the beam at a specific location.

In Fig. 8, we show the relationship between the achievable rate and the power budget at the BS. As expected, regardless of

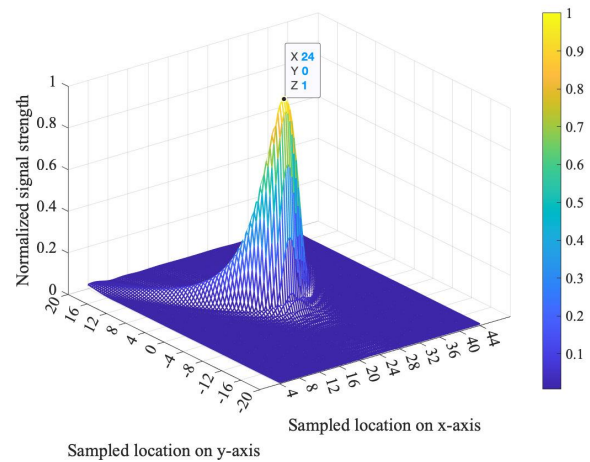


Fig. 7. Normalized received signal strength.

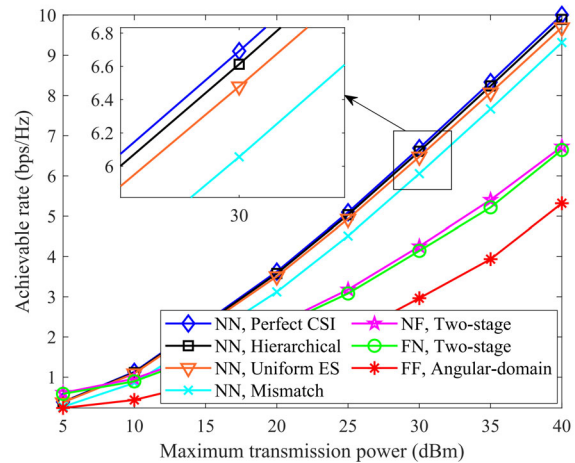


Fig. 8. Achievable rate vs. maximum transmission power at the BS.

the channel model or beam training approach, the achievable rate increases with the BS power budget. In addition, for low BS power, NN with hierarchical beam training cannot fully leverage its advantages for accurate channel estimation, thus having a lower achievable rate than NF and FN. As the power increases, the achievable rate of NN increases the fastest, and its performance advantages become increasingly pronounced compared to the other models. The rates achieved with the NF and FN channel models are nearly the same for all power levels, since the BS-RIS and RIS-user distances remain the same, the number of antennas equipped by BS and user is the same, and the same beam training methods are used for both. Furthermore, the performance of the proposed hierarchical beam training is always approximately equal to that under perfect CSI and superior to other comparison schemes, which verifies the effectiveness of the proposed scheme.

Fig. 9 shows the achievable rate for different numbers of RIS reflecting elements M . The achievable rate for all channel model and beam training approaches increases with M , due to the higher passive array gains that can be obtained from larger RIS. The NN channel model achieves the fastest increase in rate as M increases since the near-field effect is

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

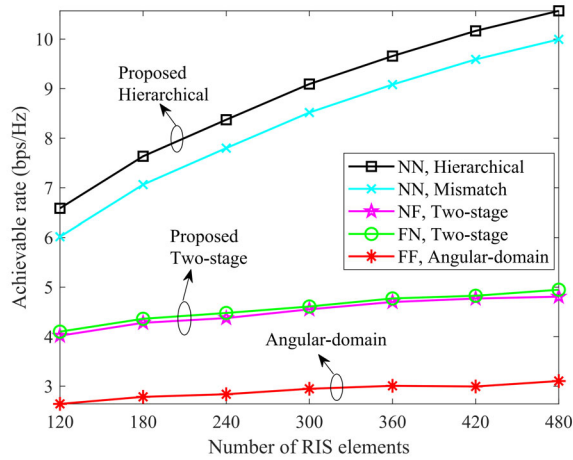
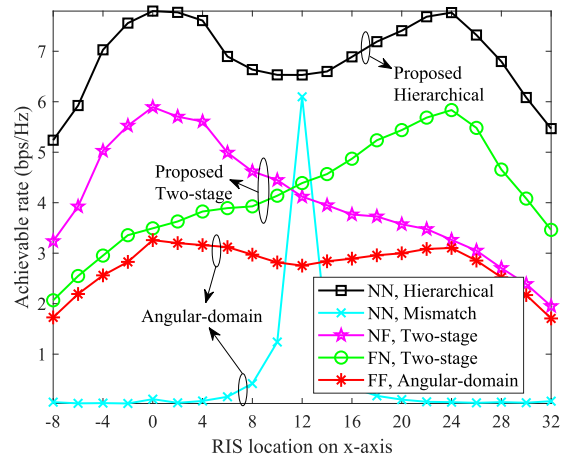
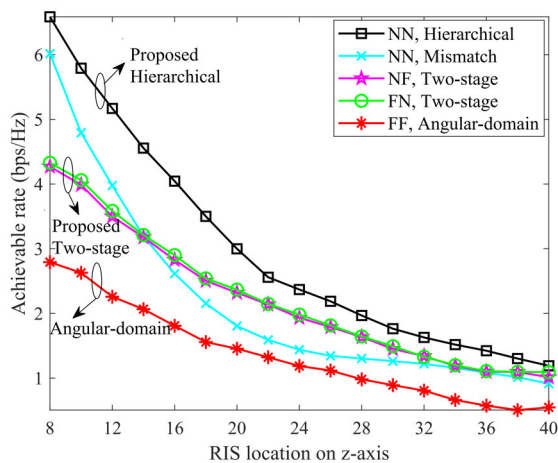


Fig. 9. Achievable rate vs. number of RIS elements.

Fig. 11. Achievable rate vs. RIS location on x -axis.Fig. 10. Achievable rate vs. RIS location on z -axis.

greater for larger RISs, the NN model describes the channel more accurately, and the hierarchical beam training approach realizes a more effective beam focusing. Furthermore, the achievable rate obtained using the NN channel model with angular-domain beam training will be superior to that of the FF model due to the more accurate channel information, even if the utilized beam training approach is not specifically designed for the NN model.

Fig. 10 shows the achievable rate as the RIS is located at different points in the z -direction. Note that the z coordinates of both the BS and the user are 0. For smaller z values, the closer the RIS is to the BS and user, the more obvious the near-field effect. Therefore, the performance gain of adopting NN channel models and hierarchical beam training is more significant with smaller z , and the results are consistent with that in Fig. 6 when $z = 8$. Moreover, as the distance between the RIS and BS and the user gradually increases, the performance gap due to different channel models and beam training methods gradually decreases. It can be predicted that when z is large enough, these curves will eventually coincide, which also verifies that the near-field channel can be approximated as the far-field channel.

The achievable rate obtained for different RIS locations along the x -direction is depicted in Fig. 11. It is not difficult to see that the NN channel model with hierarchical beam training always has the highest rate, regardless of the RIS location. In addition, for the NN channel model, a peak appears on both the BS and user side, which is consistent with the conclusions for FFC. For the NF channel model, the maximum achievable rate is obtained with the RIS near the BS, due to the more obvious near-field effect of the BS-RIS link. As the RIS moves away from the BS, the achievable rate decreases, although this degradation in rate slows near the user's location. For the FN channel model, in contrast to NF, the peak appears around the user. When adopting the NN channel model and angular beam training, a peak occurs only when the RIS is deployed at an equal distance between the BS and the user, which means that angular beam training for the NN model is only useful in restricted scenarios.

VI. CONCLUSIONS

In this article, we considered four different channel models for a RIS-assisted downlink MIMO system. According to the angular or distance information embedded in the received signals under these specific channel models, we designed different codebooks to match the beam steering vectors for the RIS. Based on the predesigned codebooks, we proposed two beam training approaches, which were further used for RIS coefficient optimization. More specifically, for the NN channel model, we designed a distance-based codebook and proposed a hierarchical beam training algorithm to realize beam alignment while reducing the training overhead; for the NF and FN channel models, we designed a combined angular-distance codebook and proposed a two-stage beam training approach to separately realize beam alignment in the angular- and distance domains. To maximize the achievable rate, we proposed an AO-based algorithm to carry out the multi-resource optimization in an iterative manner. Numerical results show that the proposed beam training approaches can obtain achievable rate performance similar to the ES method, while significantly reducing the training overhead. Our results demonstrate that the use of distance in addition to angular

information can effectively improve the system performance with near-field channels.

REFERENCES

- [1] I. F. Akyildiz, C. Han, and S. Nie, "Combating the distance problem in the millimeter wave and terahertz frequency bands," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 102–108, 2018.
- [2] C. Castro, R. Elschner, T. Merkle, C. Schubert, and R. Freund, "Experimental demonstrations of high-capacity THz-wireless transmission systems for beyond 5G," *IEEE Commun. Mag.*, vol. 58, no. 11, pp. 41–47, 2020.
- [3] K. M. S. Huq, J. Rodriguez, and I. E. Otung, "3D network modeling for THz-enabled ultra-fast dense networks: A 6G perspective," *IEEE Commun. Stand. Mag.*, vol. 5, no. 2, pp. 84–90, 2021.
- [4] C. Lin, G. Y. Li, and L. Wang, "Subarray-based coordinated beamforming training for mmWave and sub-THz communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2115–2126, 2017.
- [5] C. Pradhan, A. Li, L. Song, B. Vucetic, and Y. Li, "Hybrid precoding design for reconfigurable intelligent surface aided mmWave communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 1041–1045, 2020.
- [6] C. Huang, Z. Yang, G. C. Alexandropoulos, K. Xiong, L. Wei, C. Yuen, Z. Zhang, and M. Debbah, "Multi-hop RIS-empowered terahertz communications: A DRL-based hybrid beamforming design," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1663–1677, 2021.
- [7] 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, 2020.
- [8] 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, 2022.
- [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, 2022.
- [10] Y. Liu, Z. Wang, J. Xu, C. Ouyang, X. Mu, and R. Schober, "Near-field communications: A tutorial review," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1999–2049, 2023.
- [11] Y. Jiang, F. Gao, M. Jian, S. Zhang, and W. Zhang, "Reconfigurable intelligent surface for near field communications: Beamforming and sensing," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3447–3459, 2023.
- [12] P. Mei, Y. Cai, K. Zhao, Z. Ying, G. F. Pedersen, X. Q. Lin, and S. Zhang, "On the study of reconfigurable intelligent surfaces in the near-field region," *IEEE Trans. Antennas Propag.*, vol. 70, no. 10, pp. 8718–8728, 2022.
- [13] X. Zhang and H. Zhang, "Hybrid reconfigurable intelligent surfaces-assisted near-field localization," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 135–139, 2023.
- [14] Q. Yang, R. He, and B. Ai, "SNR analysis for RIS-aided communication system in near- and far-fields," in *2022 14th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2022, pp. 1001–1006.
- [15] J. Zhang, Y. Huang, Q. Shi, J. Wang, and L. Yang, "Codebook design for beam alignment in millimeter wave communication systems," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4980–4995, 2017.
- [16] C. Qi, K. Chen, O. A. Dobre, and G. Y. Li, "Hierarchical codebook-based multiuser beam training for millimeter wave massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8142–8152, 2020.
- [17] 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, 2022.
- [18] P. Wang, J. Fang, W. Zhang, Z. Chen, H. Li, and W. Zhang, "Beam training and alignment for RIS-assisted millimeter-wave systems: State of the art and beyond," *IEEE Wireless Commun.*, vol. 29, no. 6, pp. 64–71, 2022.
- [19] J. Wang, W. Tang, S. Jin, C.-K. Wen, X. Li, and X. Hou, "Hierarchical codebook-based beam training for RIS-assisted mmWave communication systems," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3650–3662, 2023.
- [20] W. Liu, C. Pan, H. Ren, F. Shu, S. Jin, and J. Wang, "Low-overhead beam training scheme for extremely large-scale RIS in near field," *IEEE Trans. Commun.*, vol. 71, no. 8, pp. 4924–4940, 2023.
- [21] X. Wei and L. Dai, "Channel estimation for extremely large-scale massive MIMO: Far-field, near-field, or hybrid-field?" *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 177–181, 2022.
- [22] Z. Hu, C. Chen, Y. Jin, L. Zhou, and Q. Wei, "Hybrid-field channel estimation for extremely large-scale massive MIMO system," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 303–307, 2023.
- [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, 2022.
- [24] Z. Yuan, J. Zhang, Y. Ji, G. F. Pedersen, and W. Fan, "Spatial non-stationary near-field channel modeling and validation for massive MIMO systems," *IEEE Trans. Antennas Propag.*, vol. 71, no. 1, pp. 921–933, 2023.
- [25] S.-G. Yoon and S. J. Lee, "Improved hierarchical codebook-based channel estimation for mmwave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 10, pp. 2095–2099, 2022.
- [26] C. Wu, C. You, Y. Liu, L. Chen, and S. Shi, "Two-stage hierarchical beam training for near-field communications," *arXiv preprint arXiv:2302.12511*.
- [27] X. Shi, J. Wang, Z. Sun, and J. Song, "Spatial-chirp codebook-based hierarchical beam training for extremely large-scale massive MIMO," *IEEE Trans. Wireless Commun.*, 2023.
- [28] Y. Zhang, B. Di, H. Zhang, M. Dong, L. Yang, and L. Song, "Dual codebook design for intelligent omni-surface aided communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9232–9245, 2022.
- [29] 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, 2020.
- [30] D. Shen, L. Dai, X. Su, and S. Suo, "Multi-beam design for near-field extremely large-scale RIS-aided wireless communications," *IEEE Trans. Green Commun. Netw., early access*, 2023.
- [31] 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, 2022.
- [32] D. De Donno, J. Palacios, and J. Widmer, "Millimeter-wave beam training acceleration through low-complexity hybrid transceivers," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3646–3660, 2017.
- [33] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, March 2014.
- [34] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, June 2014.
- [35] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, April 2016.
- [36] C. Ouyang, Y. Liu, X. Zhang, and L. Hanzo, "Near-field communications: A degree-of-freedom perspective," *arXiv preprint arXiv:2308.00362*.
- [37] M. Cui, L. Dai, R. Schober, and L. Hanzo, "Near-field wideband beamforming for extremely large antenna arrays," *arXiv preprint arXiv:2109.10054*, 2021.
- [38] Y. Liu, J. Xu, Z. Wang, X. Mu, and L. Hanzo, "Near-field communications: What will be different?" *arXiv preprint arXiv:2303.04003*, 2023.
- [39] F. Sahrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, 2016.
- [40] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, 2011.
- [41] B. Zheng, C. You, and R. Zhang, "Double-IRS assisted multi-user MIMO: Cooperative passive beamforming design," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4513–4526, July 2021.
- [42] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebert, "Applications of second-order cone programming," *Linear Algebra and its Applications*, vol. 284, no. 1, pp. 193–228, 1998, International Linear Algebra Society (ILAS) Symposium on Fast Algorithms for Control, Signals and Image Processing. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0024379598100320>
- [43] J. He, H. Wymeersch, and M. Juntti, "Channel estimation for RIS-aided mmWave MIMO systems via atomic norm minimization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5786–5797, 2021.