

Beam Training and Tracking for Extremely Large-Scale MIMO Communications

Kangjian Chen, *Student Member, IEEE*, Chenhao Qi, *Senior Member, IEEE*,
Cheng-Xiang Wang, *Fellow, IEEE*, and Geoffrey Ye Li, *Fellow, IEEE*

Abstract—In this paper, beam training and beam tracking are investigated for extremely large-scale multiple-input-multiple-output communication systems with partially-connected hybrid combining structures. Firstly, we propose a two-stage hybrid-field beam training scheme for both the near field and the far field. In the first stage, each subarray independently uses multiple far-field channel steering vectors to approximate near-field ones for analog combining. To find the codeword best fitting for the channel, digital combiners in the second stage are designed to combine the outputs of the analog combiners from the first stage. Then, based on the principle of stationary phase and the time-frequency duality, the expressions of subarray signals after analog combining are analytically derived and a beam refinement based on phase shifts of subarrays (BRPSS) scheme with closed-form solutions is proposed for high-resolution channel parameter estimation. Moreover, a low-complexity near-field beam tracking scheme is developed, where the kinematic model is adopted to characterize the channel variations and the extended Kalman filter is exploited for beam tracking. Simulation results verify the effectiveness of the proposed schemes.

Index Terms—Beam tracking, beam training, extremely large-scale MIMO, hybrid combining, near field.

I. INTRODUCTION

Massive multiple-input-multiple-output (MIMO) is a key technology for the fifth-generation wireless communications [1], [2]. Equipped with large antenna arrays, the base station (BS) can achieve high spectral efficiency dramatically by exploiting the spatial degree of freedom [3], [4]. To further enhance the spectral efficiency for the future sixth-generation wireless communications, extremely large-scale MIMO (XL-MIMO) that uses far more antennas than the existing massive MIMO is developed [5]–[7].

This work was supported in part by the National Natural Science Foundation of China under Grants 62071116, U22B2007 and 61960206006, in part by the Fundamental Research Funds for the Central Universities under Grant 2242022k60006, in part by the Key Technologies R&D Program of Jiangsu (Prospective and Key Technologies for Industry) under Grants BE2022067 and BE2022067-1, in part by the EU H2020 RISE TESTBED2 project under Grant 872172, and in part by the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant KYCX23_0262. This article has been presented in part at the 2022 IEEE/CIC International Conference on Communications in China (ICCC), Foshan, China, Aug. 2022 [1]. (*Corresponding author: Chenhao Qi*)

Kangjian Chen and Chenhao Qi are with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: qch@seu.edu.cn).

Cheng-Xiang Wang is with the National Mobile Communications Research Laboratory, School of Information Science and Engineering, Southeast University, Nanjing 210096, China, and also with the Purple Mountain Laboratories, Nanjing 211111, China (e-mail: chxwang@seu.edu.cn).

Geoffrey Ye Li is with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, U.K. (e-mail: geoffrey.li@imperial.ac.uk).

Due to high power consumption, the fully-digital structure that allocates each antenna with a dedicated radio frequency (RF) chain is impractical for large antenna arrays [8]. Consequently, the hybrid precoding structure, where a small number of RF chains are connected to a large number of antennas, is developed for XL-MIMO systems [9]. According to the ways how the RF chains are connected to the antennas, the hybrid precoding structure can generally be divided into two categories, i.e., fully-connected hybrid precoding structure and partially-connected hybrid precoding structure. Although the fully-connected hybrid precoding structure can achieve better spectral efficiency than the partially-connected hybrid precoding structure, the latter is more practical than the former, owing to its low hardware complexity as well as its flexibility to be extended to different sizes of antennas in blocks [10].

One important difference between XL-MIMO and the existing massive MIMO is the channel features. According to the distance between the user and the BS, the radiation field can be divided into the near field and the far field. The boundary to separate the near field and the far field is referred to as the Rayleigh distance [11]–[13]. On one hand, the Rayleigh distance increases linearly with the wavelength. On the other hand, when fixing the wavelength, the Rayleigh distance increases quadratically with the number of antennas. As a result, the far field assumption in existing massive MIMO may not hold for the XL-MIMO, especially when a user is close to the BS. In this context, channel state information (CSI) acquisition methods for XL-MIMO should consider both the near-field and far-field effects.

In general, CSI acquisition includes channel estimation and beam training [14]. Channel estimation usually focuses on efficient estimation of the high-dimensional channel matrix by exploiting advanced signal processing techniques, such as compressed sensing. However, the beam training can avoid the estimation of the high-dimensional channel matrix and obtain considerable beamforming gain, especially in low signal-to-noise ratio (SNR). In [9], to estimate the near-field channels in XL-MIMO, a polar-domain simultaneous orthogonal matching pursuit (P-SOMP) algorithm is proposed, where random beamforming instead of the directional beamforming is used. For the two-phase beam training (TPBT) method in [15], the first phase determines the candidate channel angles while the second phase finds the channel distance based on shortlisted candidate angles from the first phase. For the distance-based hierarchical beam training (DHBT) method in [16], the codebook is designed by equally sampling

the distance. Although these methods work well for the fully-connected hybrid precoding structure, they are not tailored for the practical partially-connected hybrid precoding structure.

One common challenge in beam training is the limited resolution of the predefined codebook. To achieve high-resolution channel parameter estimation, the low-complexity beam refinement is widely adopted [17]–[19]. In [17], based on the monopulse signals, beam refinement with a closed-form solution is developed. In [18], an efficient angle-of-arrival estimator is designed by approximating the power of the array response as a Gaussian function. In [19], an auxiliary beam pair is designed to provide high-resolution estimates for the angles of the channel. These beam refinement methods are for far-field channels and more investigations are desired for near-field channels.

With far more antennas than massive MIMO, XL-MIMO systems suffer from the heavy burden of training overheads. An efficient way to reduce the training overhead is the beam tracking, which exploits the correlations of the channels at different time instants to narrow down the sets of training candidates [20]. To meet the demands for high reliability and massive connectivity in future wireless communications, beam tracking is usually characterized by high beamforming gain, low training overhead and low computational complexity. There are a variety of beam tracking schemes for the far-field channels [21]–[24]. In [21], well-designed pairs of auxiliary beams can capture the angle variations. The beam zooming-based beam tracking scheme in [22] exploits the delay-phase precoding structure to flexibly control the angular coverage of frequency-dependent beams. In the adaptive tracking framework in [23], beam direction is updated according to measurements of the current data beam. The grid-based hybrid tracking scheme in [24] searches the surroundings of the former beam and selects the best beam according to the changing trend of the previously used beams. For the near field, user tracking has also been investigated in a variety of works, such as [25]–[27]. However, the approaches of these works are incompatible with the beam training framework. More investigations are needed for efficient beam tracking in the near field.

In this paper, for XL-MIMO systems with partially-connected hybrid combining structure, we investigate beam training, beam refinement and beam tracking, which are three progressive techniques to accomplish the beam alignment. Our contributions are summarized as follows.

- We propose a two-stage hybrid-field beam training (THBT) scheme, which works for both the near and far fields. In the first stage, we use far-field channel steering vectors of subarrays to approximate the near-field ones for analog combining so that beam training for both the near and far fields can be performed simultaneously. In the second stage, digital combiners are designed to combine the outputs of the analog combiners from the first stage. Then from the predefined hybrid-field codebook, we select the codeword corresponding to the dedicated digital combiner that achieves the largest combining power as the result of the THBT.
- We propose a beam refinement based on phase shifts of

subarrays (BRPSS) scheme. Based on the principle of stationary phase (PSP) and the time-frequency duality, the expressions of subarray signals after analog combining are analytically derived, where the phases of these signals change quadratically with the subarray indices. By exploiting the phase shifts of subarrays, closed-form estimates of channel parameters can be obtained.

- We develop a low-complexity near-field beam tracking (NFBT) scheme. As the near-field channels are related to both the angle and distance of the radiation source, the kinematic model is adopted to characterize the channel variations. In addition, the BRPSS scheme is used to estimate the real-time channel parameters. Then the kinematic model and real-time estimates are exploited by the extended Kalman filter (EKF) to track and predict the near-field channel parameters.

The rest of this paper is organized as follows. Section II introduces the model of the XL-MIMO systems. In Section III, we propose the two-stage hybrid-field beam training scheme. The beam refinement based on phase shifts of subarrays scheme is provided in Section IV. The near-field beam tracking scheme is discussed in Section V. The simulation results are presented in Section VI, and the paper is concluded in Section VII.

The notations are defined as follows. Symbols for matrices (upper case) and vectors (lower case) are in boldface. The set is represented by bold Greek letters. $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and conjugate transpose (Hermitian), respectively. $[\mathbf{A}]_{:,m}$ denotes the m th column of a matrix \mathbf{A} . j denotes the square root of -1 . In addition, $|\cdot|$ and $\|\cdot\|_2$ denote the absolute value of a scalar and ℓ_2 -norm of a vector, respectively. \mathbb{C} denotes the set of complex numbers. The complex Gaussian distribution is denoted by \mathcal{CN} . $\lfloor \cdot \rfloor$ and $\text{blkdiag}\{\cdot\}$ denote the floor operation and the block diagonalization operation, respectively. Moreover, \mathcal{O} and $\text{mod}(\cdot)$ denote the order of complexity and the operation of modulo, respectively.

II. SYSTEM MODEL

We consider uplink beam training for an XL-MIMO system with a user and a BS. As shown in Fig. 1, the BS employs a large-scale uniform linear array (ULA) of N antennas with half wavelength interval and a partially-connected hybrid combining structure with N_{RF} RF chains, where the hybrid combining includes analog and digital combining. In practice, the BS with a hybrid structure has far more antennas than RF chains, i.e., $N \gg N_{\text{RF}}$. The ULA is formed by N_{RF} non-overlapping subarrays, where each subarray has $M = N/N_{\text{RF}}$ antennas and is solely connected to an RF chain after analog combining. Then all the N_{RF} RF chains are connected to a digital baseband processing unit for digital combining. In this work, we focus on the processing at the BS and consider a single-antenna user for simplicity. However, the proposed schemes can be extended to scenarios with multi-antenna users.

During uplink beam training, the p th transmit pilot signal by the user is denoted as x_p for $p = 1, 2, \dots, P$, where P is the pilot length. The channel vector between the user and the

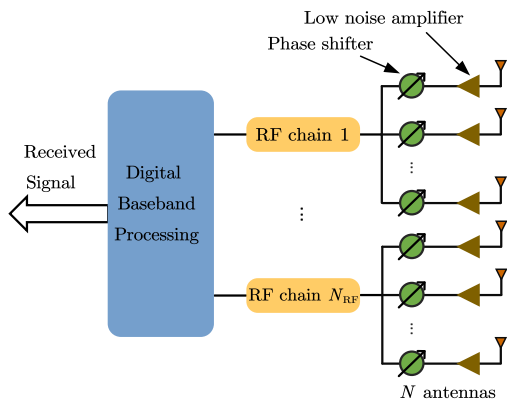


Fig. 1. Illustration of a partially-connected hybrid combining structure.

BS is denoted as $\mathbf{h} \in \mathbb{C}^N$. Then the signal after the analog combining at the BS side can be expressed as [10]

$$y_p = \mathbf{v}_p \mathbf{W}_p \mathbf{h} x_p + \mathbf{v}_p \mathbf{W}_p \boldsymbol{\eta}, \quad (1)$$

where $\mathbf{W}_p \in \mathbb{C}^{N_{\text{RF}} \times N}$ is the analog combiner, $\mathbf{v}_p \in \mathbb{C}^{1 \times N_{\text{RF}}}$ is the digital combiner, and $\boldsymbol{\eta} \in \mathbb{C}^N$ is an additive white Gaussian noise (AWGN) vector obeying $\boldsymbol{\eta} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$. If the digital combiner is a matrix, we may treat each row of it as a different \mathbf{v}_p , so that we can perform parallel baseband processing.

Generally, two kinds of channels, namely the near-field channel and the far-field channel, are considered in the existing literature according to the distance between the user and the BS [11], [12]. The commonly used boundary distance to distinguish the near field and the far field is the Rayleigh distance

$$Z = 2D^2/\lambda, \quad (2)$$

where D denotes the array aperture and λ denotes the wavelength. In other words, when the distance between the user and the BS is larger than Z , the wireless channel is regarded as the far-field channel; otherwise, it is the near-field channel. Since a half-wavelength-interval ULA is adopted in this work, the array aperture at the BS is $D = N\lambda/2$, and therefore $Z = N^2\lambda/2$.

As shown in Fig. 2, a near-field channel composed of one line-of-sight (LoS) path and several non-line-of-sight (NLoS) paths between the BS and the user is considered. N antennas of the BS are placed along the y-axis in the Cartesian coordinate system and the coordinate of the n th antenna is $(0, \delta_n \lambda)$, where $\delta_n \triangleq (2n - N - 1)/4$ for $n = 1, 2, \dots, N$. The coordinate of the user is denoted as $\mathbf{p}_1 = (r_1 \cos \theta_1, r_1 \sin \theta_1)$, where r_1 is the distance between the user and the origin, and $\theta_1 \in [-\pi/2, \pi/2]$ is the angle of the user relative to the positive x-axis. Similarly, the coordinate of the scatterer on the l th path for $l \geq 2$ is denoted as $\mathbf{p}_l = (r_l \cos \theta_l, r_l \sin \theta_l)$, where r_l is the distance between the scatterer on the l th path and the coordinate origin, and $\theta_l \in [-\pi/2, \pi/2]$ is the angle of the scatterer on the l th path relative to the positive x-axis. The distance between \mathbf{p}_l and the n th antenna can be expressed as

$$r_l^{(n)} = \sqrt{r_l^2 + \delta_n^2 \lambda^2 - 2r_l \Omega_l \delta_n \lambda}, \quad (3)$$

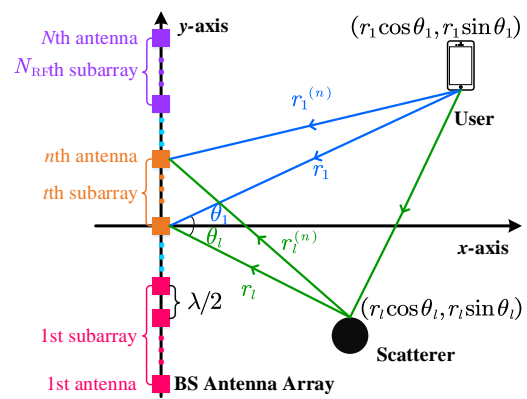


Fig. 2. Illustration of a near-field channel model.

for $l \geq 1$, where $\Omega_l \triangleq \sin \theta_l \in [-1, 1]$. Then the channel vector between the user and the BS can be expressed as

$$\mathbf{h} = \sum_{l=1}^L g_l \boldsymbol{\alpha}(N, \Omega_l, r_l), \quad (4)$$

where L and g_l denote the number of paths and the path gain of the l th path, respectively. The channel steering vector, $\alpha(\cdot)$, is defined as

$$\boldsymbol{\alpha}(N, \Omega_l, r_l) = \frac{1}{\sqrt{N}} \left[e^{-j \frac{2\pi}{\lambda} (r_l^{(1)} - r_l)}, \dots, e^{-j \frac{2\pi}{\lambda} (r_l^{(N)} - r_l)} \right]^T. \quad (5)$$

Note that the channel steering vector in (5) can be used to describe both the far-field channel and the near-field channel.

If $r_l > Z$, $r_l^{(n)}$ in (3) can be approximated as $r_l^{(n)} \approx r_l - \Omega_l \delta_n \lambda$ because $\delta_n \lambda / r_l \approx 0$ and $\sqrt{1 + \epsilon} \approx 1 + \frac{1}{2} \epsilon$ [9]. As a result, $\alpha(N, \Omega_l, r_l)$ in (5) can be approximated as

$$\boldsymbol{\alpha}(N, \Omega_l, r_l) \approx \frac{1}{\sqrt{N}} [e^{j2\pi\Omega_l\delta_1}, \dots, e^{j2\pi\Omega_l\delta_N}]^T. \quad (6)$$

Multiplying a constant phase factor $e^{-j2\pi\Omega_l\delta_1}$ to all the entries of $\alpha(N, \Omega_l, r_l)$ in (6), we have

$$e^{-j2\pi\Omega_l\delta_1}\boldsymbol{\alpha}(N, \Omega_l, r_l) \approx \frac{1}{\sqrt{N}} \left[1, e^{j\pi\Omega_l}, \dots, e^{j\pi(N-1)\Omega_l} \right]^T \\ \triangleq \boldsymbol{\beta}(N, \Omega_l). \quad (7)$$

In fact, $\beta(N, \Omega_l)$ is the far-field channel steering vector independent of r_l .

If $r_l \leq Z$, the approximation in (7) is not accurate enough. Alternatively, we approximate $r_l^{(n)}$ as $r_l^{(n)} \approx r_l - \Omega_l \delta_n \lambda + \frac{\delta_n^2 \lambda^2 (1 - \Omega_l^2)}{2r_l}$ according to $\sqrt{1 + \epsilon} \approx 1 + \frac{1}{2}\epsilon - \frac{1}{8}\epsilon^2$, which is verified to be accurate if $r_l^{(n)} \geq 0.5\sqrt{D^3/\lambda}$ [9], [12]. Note that the ratio of Z to $0.5\sqrt{D^3/\lambda}$ is $\sqrt{8N}$, which indicates that the latter is much smaller than the former, especially for XL-MIMO systems. For example, if $N = 512$, $0.5\sqrt{D^3/\lambda}$ is only $1/64$ of Z , which is nearly negligible. Therefore, in this work, we focus on the near field with $r_l^{(n)} \geq 0.5\sqrt{D^3/\lambda}$. Then $\alpha(N, \Omega_l, r_l)$ in (5) can be approximated as

$$\alpha(N, \Omega_l, r_l) \approx \frac{1}{\sqrt{N}} [e^{j2\pi(\Omega_l \delta_1 - \rho_l \delta_1^2)}, \dots, e^{j2\pi(\Omega_l \delta_N - \rho_l \delta_N^2)}]^T \quad (8)$$

where $\rho_l \triangleq \frac{\lambda(1-\Omega_l^2)}{2r_l}$. Define $b_l \triangleq \Omega_l + \rho_l(N + 1)/2$ and $k_l \triangleq -\rho_l/2$. Multiplying a constant fac-

for $e^{-j2\pi(\Omega_l(\delta_1-1/2)-\rho_l(N+1)^2/16)}$ to all the entries of $\alpha(N, \Omega_l, r_l)$ in (6), we have

$$\begin{aligned} & e^{-j2\pi(\Omega_l(\delta_1-1/2)-\rho_l(N+1)^2/16)} \alpha(N, \Omega_l, r_l) \\ & \approx \frac{1}{\sqrt{N}} \left[e^{j\pi(k_l+b_l)}, \dots, e^{j\pi(k_l n^2+b_l n)}, \dots, e^{j\pi(k_l N^2+b_l N)} \right]^T \\ & \triangleq \gamma(N, \Omega_l, r_l). \end{aligned} \quad (9)$$

Note that $\beta(N, \Omega_l)$ is a special case of $\gamma(N, \Omega_l, r_l)$ with $\rho_l = 0$. Therefore, $\gamma(N, \Omega_l, r_l)$ can be used to approximate both the far-field and the near-field channel steering vectors.

III. TWO-STAGE HYBRID-FIELD BEAM TRAINING

In this section, we will use the far-field channel steering vectors of a subarray to approximate its near-field channel steering vectors. Then we propose a THBT scheme for both the near field and far field.

A. Hybrid-Field Beam Sweeping

To estimate the multipath channel, codebook-based beam training is widely adopted [3]. Since the user is either in the far field or near field, a hybrid-field codebook considering both the near-field and far-field effects will be developed in the following. First, we establish a far-field codebook $\mathbf{C}_f \in \mathbb{C}^{N \times Q}$ based on the far-field channel steering vectors [3], where Q denotes the quantized number of the angle and the q th column of \mathbf{C}_f is denoted as

$$[\mathbf{C}_f]_{:,q} \triangleq \beta \left(N, \frac{2q-1-Q}{Q} \right) \quad (10)$$

for $q = 1, 2, \dots, Q$. Then we establish a near-field codebook $\mathbf{C}_n \in \mathbb{C}^{N \times (QS)}$. Since the near-field channel is relevant to both the distance and the angle, we quantize the distance and the angle by S samples and Q samples, respectively. The q th angle sample is $\Theta_q = (2q-1-Q)/Q$ for $q = 1, 2, \dots, Q$. According to the polar-domain sparsity in [9], the s th distance sample can be expressed as

$$d_{q,s} = \frac{N^2 \lambda (1 - \Theta_q^2)}{8\beta s} \quad (11)$$

for $s = 1, 2, \dots, S$, where β is a factor to adjust the coherence of the adjacent codewords. By setting $N^2 \lambda / (8\beta S) = 0.5 \sqrt{D^3 / \lambda}$, we have $\beta = \sqrt{\frac{N}{2S^2}}$. Then (11) can be rewritten as

$$d_{q,s} = \frac{N^{3/2} \lambda S (1 - \Theta_q^2)}{4\sqrt{2} s}. \quad (12)$$

Let $\mathbf{C}_n \triangleq \{\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_Q\}$, where the s th column of $[\mathbf{C}_q]$ is $[\mathbf{C}_q]_{:,s} = \alpha(N, \Theta_q, d_{q,s})$. Accordingly, the hybrid-field codebook is defined as

$$\mathbf{C}_h \triangleq \{\mathbf{C}_n, \mathbf{C}_f\} \in \mathbb{C}^{N \times (QS+Q)}. \quad (13)$$

Based on (1), we have

$$\mathbf{y}_p = [\mathbf{C}_h]_{:,p}^H \mathbf{h} x_p + [\mathbf{C}_h]_{:,p}^H \boldsymbol{\eta} \quad (14)$$

for $p = 1, 2, \dots, QS+Q$. The beam training aims at finding the codeword in \mathbf{C}_h best fitting for the multipath channel, which can be expressed as

$$\tilde{p} = \arg \max_{p=1,2,\dots,QS+Q} |[\mathbf{C}_h]_{:,p}^H \mathbf{h}|. \quad (15)$$

To solve (15), we need to test all the codewords in \mathbf{C}_h one by one, which is called the hybrid-field beam sweeping (HFBS). To perform the HFBS, $(QS+Q)$ times of beam training are needed. Comparing (7) with (5), larger overhead is needed by the near-field beam training than by the far-field beam training because the former needs S times beam training for different distances even for the same angle while the latter needs only one time of beam training for the same angle. Therefore, it would be interesting to consider how to use the far-field beam training for the near-field channel, which will be discussed subsequently.

B. Subarray Approximation

We define an analog combiner as

$$\mathbf{W} \triangleq \text{blkdiag}\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{N_{\text{RF}}}\} \quad (16)$$

and a digital combiner as $\mathbf{v} \in \mathbb{C}^{1 \times N_{\text{RF}}}$, where $\mathbf{w}_t \in \mathbb{C}^{1 \times M}$ for $t = 1, 2, \dots, N_{\text{RF}}$. Given \mathbf{h} , the optimal hybrid combiner to achieve the maximum received power can be designed via solving the problem

$$\begin{aligned} & \max_{\mathbf{W}, \mathbf{v}} |\mathbf{v} \mathbf{W} \mathbf{h}| \\ & \text{s.t. } \|\mathbf{v} \mathbf{W}\|_2 = 1, |[w_t]_m| = 1 \end{aligned} \quad (17)$$

for $m = 1, 2, \dots, M$ and $t = 1, 2, \dots, N_{\text{RF}}$. Due to the much smaller path gain of the NLoS paths than that of the LoS path especially in millimeter wave or terahertz band, we omit the NLoS paths. Then (17) can be rewritten as

$$\max_{\mathbf{W}, \mathbf{v}} |\mathbf{v} \mathbf{W} \boldsymbol{\alpha}(N, \Omega, r)| \quad (18a)$$

$$\text{s.t. } \|\mathbf{v} \mathbf{W}\|_2 = 1, |[w_t]_m| = 1 \quad (18b)$$

for $m = 1, 2, \dots, M$ and $t = 1, 2, \dots, N_{\text{RF}}$. In (18a), the subscript “ l ” is omitted for simplicity, which goes the same for the rest of this paper.

According to the Cauchy-Schwartz inequality, we have

$$|\mathbf{v} \mathbf{W} \boldsymbol{\alpha}(N, \Omega, r)| \leq \|\mathbf{v}\|_2 \|\mathbf{W} \boldsymbol{\alpha}(N, \Omega, r)\|_2. \quad (19)$$

The equality in (19) holds if $\mathbf{v}^H = \mu \mathbf{W} \boldsymbol{\alpha}(N, \Omega, r)$, where μ is a scaling factor. Since \mathbf{W} and \mathbf{v} are independent of each other, we can achieve the equality of (19). Consequently, we may first determine \mathbf{W} and then design \mathbf{v} accordingly.

The design of \mathbf{W} can be formulated as

$$\begin{aligned} & \max_{\mathbf{W}} \|\mathbf{W} \boldsymbol{\alpha}(N, \Omega, r)\|_2 \\ & \text{s.t. } |[w_t]_m| = 1 \end{aligned} \quad (20)$$

for $m = 1, 2, \dots, M$ and $t = 1, 2, \dots, N_{\text{RF}}$. The entries of \mathbf{W} are mutually independent, implying that the maximization of $\|\mathbf{W} \boldsymbol{\alpha}(N, \Omega, r)\|_2$ is essentially the maximization of the absolute value of each entry of $\mathbf{W} \boldsymbol{\alpha}(N, \Omega, r)$. Therefore, the problem in (20) is divided into N_{RF} independent subproblems and the t th subproblem for $t = 1, 2, \dots, N_{\text{RF}}$ can be expressed as

$$\begin{aligned} & \max_{\mathbf{w}_t} |\mathbf{w}_t \mathbf{G}_t \boldsymbol{\alpha}(N, \Omega, r)| \\ & \text{s.t. } |[w_t]_m| = 1 \end{aligned} \quad (21)$$

for $m = 1, 2, \dots, M$, where we define $\mathbf{G}_t \triangleq [\mathbf{0}_{M \times (t-1)M}, \mathbf{I}_M, \mathbf{0}_{M \times (N_{\text{RF}}-t)M}]$. The optimal solution of

(21) is

$$\bar{\mathbf{w}}_t = \sqrt{N}(\mathbf{G}_t \boldsymbol{\alpha}(N, \Omega, r))^H \quad (22)$$

for $t = 1, 2, \dots, N_{\text{RF}}$.

Note that $\mathbf{G}_t \boldsymbol{\alpha}(N, \Omega, r)$ is exactly the channel steering vector between the user and the t th subarray. From (2), the entire array at the BS achieves larger Z than each subarray. The Rayleigh distance decreases quadratically with the reduction of the antenna number. In fact, the Rayleigh distance of a subarray is only $1/N_{\text{RF}}^2$ of the Rayleigh distance of the entire array at the BS, which implies that a user in the near field of the entire array may be in the far field of a subarray. For example, if $N = 256$, $\lambda = 0.003$ m, and $N_{\text{RF}} = 4$, the Rayleigh distance of the entire array is 98.3m while the Rayleigh distance of a subarray is only 6.1m. In other words, the near-field effect of subarrays is much weaker than that of the entire array. Based on the above discussion, we try to use the far-field channel steering vectors to approximate the near-field steering vectors for subarrays. To evaluate the deviation of the approximation, we have the following lemmas proved in Appendices A and B.

Lemma 1: The maximum beam gain loss of approximating near-field channel steering vector with far-field channel steering vector for a subarray can be approximated as

$$\Gamma_{\max} = \max\{1 - N_{\text{RF}}/\sqrt[4]{2N}, 0\}. \quad (23)$$

Lemma 2: Denote the sine result of the angle that points from the center of the t th subarray to the user as Ψ_t and denote the beam center of $\mathbf{G}_t \boldsymbol{\alpha}(N, \Omega, r)$ as B_t . Then, we have $B_t \approx \Psi_t$.

According to **Lemma 1**, the beam gain loss of approximating the near-field channel steering vector with the far-field channel steering vector is limited. For example, if $N = 256$ and $N_{\text{RF}} = 4$, the maximum beam gain loss is only 16%. From **Lemma 2**, considerable beamforming gain can be obtained if the t th subarray receives the signal from channel $\mathbf{G}_t \boldsymbol{\alpha}(N, \Omega, r)$ with $\beta(M, \Psi_t)$.

Based on **Lemma 1** and **Lemma 2**, for each subarray, we use far-field channel steering vectors,

$$\hat{\mathbf{w}}_t = \sqrt{M}\beta(M, \Psi_t)^H \quad (24)$$

for $t = 1, 2, \dots, N_{\text{RF}}$ to approximate the near-field channel steering vectors in (22), which is termed the subarray approximation. Therefore, we can use (24) for both the far-field and near-field channels. Note that the subarray approximation has also been studied in the existing works, such as [28]. The existing works aim to design the wideband beamforming based on the ideal CSI. However, our work performs the effective beam training to obtain the CSI.

By substituting (24) into (16), we express the designed analog combiner for (18) as

$$\hat{\mathbf{W}} = \text{blkdiag}\{\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, \dots, \hat{\mathbf{w}}_{N_{\text{RF}}}\}. \quad (25)$$

Since $\hat{\mathbf{W}}\hat{\mathbf{W}}^H = M\mathbf{I}_{N_{\text{RF}}}$, we have $\|\mathbf{v}\hat{\mathbf{W}}\|_2 = \sqrt{M}\|\mathbf{v}\|_2$. Then the design of \mathbf{v} according to (18) can be expressed as

$$\max_{\mathbf{v}} \mathbf{v}\hat{\mathbf{W}}\boldsymbol{\alpha}(N, \Omega, r) \quad (26a)$$

$$\text{s.t. } \|\mathbf{v}\|_2 = 1/\sqrt{M}. \quad (26b)$$

Note that (18a) can be rewritten as (26a) because we can

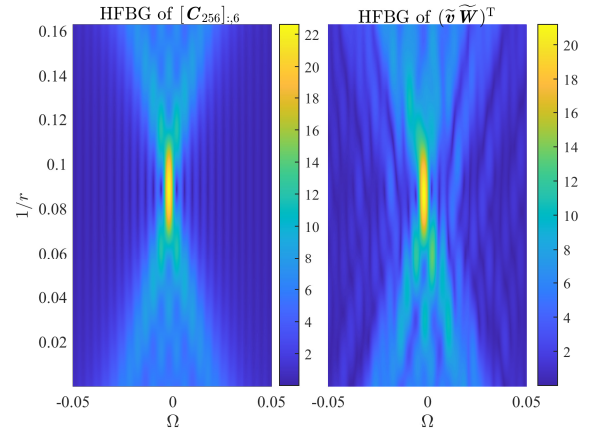


Fig. 3. Illustration of the subarray approximation.

always adjust the phase of \mathbf{v} so that $\mathbf{v}\hat{\mathbf{W}}\boldsymbol{\alpha}(N, \Omega, r)$ is real positive and the maximum of $|\mathbf{v}\hat{\mathbf{W}}\boldsymbol{\alpha}(N, \Omega, r)|$ is still achieved. The optimal \mathbf{v} for (26) is

$$\hat{\mathbf{v}} = \frac{(\hat{\mathbf{W}}\boldsymbol{\alpha}(N, \Omega, r))^H}{\sqrt{M}\|\hat{\mathbf{W}}\boldsymbol{\alpha}(N, \Omega, r)\|_2}. \quad (27)$$

Note that (25) and (27) are designed based on the continuous space. To facilitate the implementation of beam training, hybrid combiners for the quantized space need to be designed. For each subarray, the commonly-used DFT codebook for beam training is

$$\Phi = \{\sqrt{M}\beta(M, \Phi_1), \sqrt{M}\beta(M, \Phi_2), \dots, \sqrt{M}\beta(M, \Phi_M)\}, \quad (28)$$

where

$$\Phi_m = (2m - 1 - M)/M \quad (29)$$

for $m = 1, 2, \dots, M$. In fact, the DFT codebook equally samples the full space $[-1, 1]$ by M angles, where the m th angle is Φ_m . For the t th subarray, the index of the codeword in Φ best fitting for the channel is

$$\tilde{m}_t = \arg \min_{m=1,2,\dots,M} |\Phi_m - \Psi_t|. \quad (30)$$

Therefore, the designed analog combiner and digital combiner according to (25) and (27), respectively, are

$$\tilde{\mathbf{W}} = \text{blkdiag}\{\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2, \dots, \tilde{\mathbf{w}}_{N_{\text{RF}}}\}, \quad (31)$$

$$\tilde{\mathbf{v}} = \frac{(\tilde{\mathbf{W}}\boldsymbol{\alpha}(N, \Omega, r))^H}{\sqrt{M}\|\tilde{\mathbf{W}}\boldsymbol{\alpha}(N, \Omega, r)\|_2}, \quad (32)$$

where

$$\tilde{\mathbf{w}}_t = [\Phi]_{:, \tilde{m}_t}^H \quad (33)$$

for $t = 1, 2, \dots, N_{\text{RF}}$ based on (24).

Fig. 3 illustrates the performance of the subarray approximation, where $N = 512$, $N_{\text{RF}} = 4$, $Q = 512$ and $S = 11$. We take the subarray approximation of $[\mathbf{C}_{256}]_{:,6} = \boldsymbol{\alpha}(N, \Theta_{256}, d_{256,6})$ as an example. Based on (30), we can obtain $\tilde{m}_1 = 63$, $\tilde{m}_2 = 64$, $\tilde{m}_3 = 65$ and $\tilde{m}_4 = 66$. The analog combiner and the digital combiner can be designed

based on (31) and (32), respectively. Then the combined channel steering vector is $(\tilde{\mathbf{v}}\tilde{\mathbf{W}})^T$. We define the hybrid-field beam gain (HFBG) of a channel steering vector \mathbf{u} as

$$B(\mathbf{u}, \Omega, r) = |\alpha(N, \Omega, r)^H \mathbf{u}|. \quad (34)$$

In Fig. 3, we illustrate the HFBGs of $[\mathbf{C}_{256}]_{:,6}$ and $(\tilde{\mathbf{v}}\tilde{\mathbf{W}})^T$. From the figure, the HFBG of $(\tilde{\mathbf{v}}\tilde{\mathbf{W}})^T$ is similar to that of $[\mathbf{C}_{256}]_{:,6}$, which verifies the effectiveness of the subarray approximation.

C. Beam Training

Now we propose a two-stage hybrid-field beam training scheme.

Note that for the partially-connected structure, each subarray is exclusively connected to an RF chain, which indicates that each RF chain can support an independent beam training based on a subarray. In addition, the near-field effect is substantially weakened once the ULA is divided into N_{RF} subarrays. Therefore, for each angle, we only need one time of beam training no matter whether the user is in the near field or far field.

Based on the above discussions, in the first stage, each subarray independently performs beam training with codewords in Φ . When performing the m th beam training, the analog combiner according to (28) can be expressed as

$$\bar{\mathbf{W}}_m = \text{blkdiag}\{\sqrt{M}\beta(M, \Phi_m)^H, \dots, \sqrt{M}\beta(M, \Phi_m)^H\}. \quad (35)$$

for $m = 1, 2, \dots, M$. Based on (1), the output signal of the k th combiner is

$$\mathbf{z}_m = \bar{\mathbf{W}}_m \mathbf{h} x_m + \bar{\mathbf{W}}_m \boldsymbol{\eta} \quad (36)$$

for $m = 1, 2, \dots, M$.

Note that according to the analysis in Section III-B, the received signals of subarrays can be combined so that the power of combined signals can approximate the power of signals received by the entire array. In addition, for any codeword in \mathbf{C}_h , the analog combiners of subarrays come from Φ . Moreover, all subarrays have performed beam training with Φ in the first stage. Therefore, by reusing the received signals in the first stage and designing the digital combiners, the power of the combined signals can approximate the power of signals received by codewords in \mathbf{C}_h .

Based on the above discussions, in the second stage, by utilizing $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M\}$, we design the digital combiner, \mathbf{v}_p , for $p = 1, 2, \dots, QS + Q$, to test all the codewords in the hybrid-field codebook \mathbf{C}_h . To be detailed, for each codeword $[\mathbf{C}_h]_{:,p}$, we can design a dedicated digital combiner \mathbf{v}_p to combine the output of the analog combiner from the first stage. From (13), we have

$$[\mathbf{C}_h]_{:,p} = \begin{cases} \alpha(N, \Theta_{\bar{q}}, d_{\bar{q}, \bar{s}}), & p \leq QS, \\ \beta(N, \frac{2(p-QS)-1-Q}{Q}), & p > QS, \end{cases} \quad (37)$$

where

$$\bar{q} = \lceil p/S \rceil \text{ and } \bar{s} = p - (\bar{q} - 1)S. \quad (38)$$

Replacing $\alpha(N, \Omega, r)$ in (18) by $[\mathbf{C}_h]_{:,p}$, we can obtain $\tilde{\mathbf{w}}_t^{(p)}$ via (33). Similar to (31) and (32), we design the analog

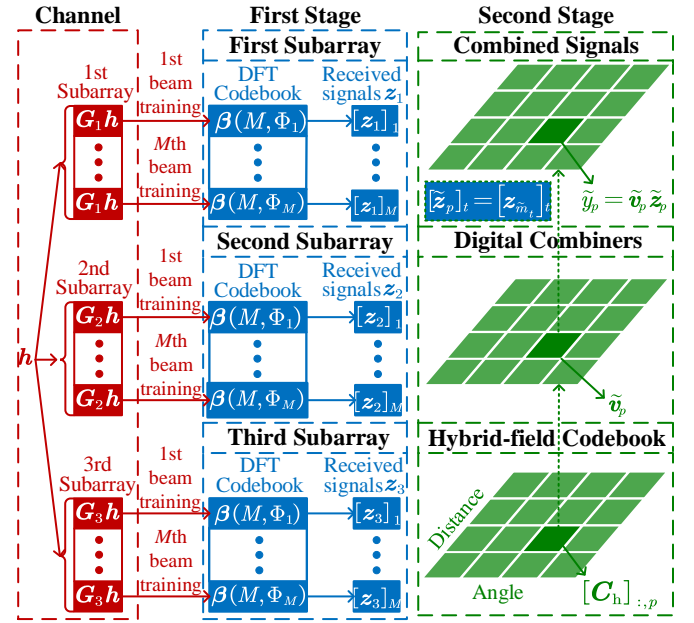


Fig. 4. Illustration of the procedure for the THBT scheme.

combiner and the digital combiner, respectively, as

$$\mathbf{F}_p = \text{blkdiag}\{\tilde{\mathbf{w}}_1^{(p)}, \tilde{\mathbf{w}}_2^{(p)}, \dots, \tilde{\mathbf{w}}_{N_{\text{RF}}}^{(p)}\}, \quad (39)$$

$$\tilde{\mathbf{v}}_p = \frac{(\mathbf{F}_p [\mathbf{C}_h]_{:,p})^H}{\sqrt{M} \|\mathbf{F}_p [\mathbf{C}_h]_{:,p}\|_2}. \quad (40)$$

If we use \mathbf{F}_p and $\tilde{\mathbf{v}}_p$ for combining, we have

$$\tilde{\mathbf{z}}_p = \mathbf{F}_p \mathbf{h} x_p + \mathbf{F}_p \boldsymbol{\eta}, \quad (41)$$

and

$$\tilde{y}_p = \tilde{\mathbf{v}}_p^H \tilde{\mathbf{z}}_p, \quad p = 1, 2, \dots, QS + Q. \quad (42)$$

In fact, $\tilde{\mathbf{z}}_p$ can not be obtained directly because we do not actually perform beam training with \mathbf{F}_p . However, each entry of $\tilde{\mathbf{z}}_p$ can be obtained from the beam training in (36) because both \mathbf{F}_p and $\bar{\mathbf{W}}_m$ are composed of channel steering vectors from the same set Φ in (28). We can obtain $\tilde{\mathbf{z}}_p$ by setting

$$[\tilde{\mathbf{z}}_p]_t = [\mathbf{z}_{\tilde{m}_t}]_t, \quad (43)$$

where \tilde{m}_t can be obtained from (30) during the design of \mathbf{F}_p in (39). From $\{\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_{QS+Q}\}$, we select the one with the largest power, which can be expressed as

$$\tilde{p} = \arg \max_{p=1,2,\dots,QS+Q} |\tilde{y}_p|^2. \quad (44)$$

From the hybrid-field codebook, \mathbf{C}_h , we select the codeword $[\mathbf{C}_h]_{:, \tilde{p}}$ corresponding to the dedicated digital combiner $\tilde{\mathbf{v}}_{\tilde{p}}$ that can achieve the largest combining power $|\tilde{y}_{\tilde{p}}|^2$ as the result of the THBT.

Furthermore, if the user is in the near field, we can also roughly locate the user based on the beam training results. According to (38), the indices of the quantized angle and the quantized distance can be calculated as

$$\tilde{q} = \lceil \tilde{p}/S \rceil \text{ and } \tilde{s} = \tilde{p} - (\tilde{q} - 1)S. \quad (45)$$

Then the angle and the distance of the user can be roughly

Algorithm 1 Two-Stage Hybrid-Field Beam Training (THBT) Scheme

```

1: Input:  $N$ ,  $N_{\text{RF}}$ ,  $M$ ,  $Q$ ,  $S$ ,  $\lambda$ .
2: First Stage:
3: Obtain  $\mathbf{z}_m$ ,  $m = 1, 2, \dots, M$  via (36).
4: Second Stage:
5: for  $p = 1, 2, \dots, QS + Q$  do
6:   Obtain  $[\mathbf{C}_h]_{:,p}$  via (37).
7:   Obtain  $\mathbf{F}_p$  and  $\tilde{\mathbf{v}}_p$  via (39) and (40), respectively.
8:   Obtain  $\tilde{\mathbf{z}}_p$  via (43).
9:   Obtain  $\tilde{\mathbf{y}}_p$  via (42).
10: end for
11: Obtain  $\tilde{\mathbf{p}}$  via (44).
12: Obtain  $\tilde{\Omega}$  and  $\tilde{r}$  via (46).
13: Output:  $[\mathbf{C}_h]_{:, \tilde{\mathbf{p}}}$ ,  $\tilde{\Omega}$  and  $\tilde{r}$ .

```

estimated as

$$\tilde{\Omega} = (2\tilde{q} - 1 - Q)/Q \text{ and } \tilde{r} = \frac{N^{3/2}\lambda S(1 - \Theta_{\tilde{q}}^2)}{4\sqrt{2}\tilde{s}}. \quad (46)$$

Finally, The detailed steps of the proposed THBT scheme are summarized in **Algorithm 1**.

Fig. 4 illustrates the procedure for the THBT scheme, where $N_{\text{RF}} = 3$. First, each subarray receives the uplink pilots independently based on the DFT codebook. Then, we design the digital combiners to combine the outputs of subarrays to test codewords in the predefined codebook. Finally, from the predefined hybrid-field codebook, we select the codeword that achieves the largest combining power as the result of the THBT.

Now we evaluate the overhead of the THBT scheme. The THBT scheme performs digital combining in the second stage, which does not need any pilots. Before digital combining, in the first stage, each subarray independently performs beam training based on the DFT codebook and needs M times of beam training. Therefore, the total overhead of the THBT scheme is M .

IV. BEAM REFINEMENT BASED ON PHASE SHIFTS OF SUBARRAYS

In this section, we focus on refining the quantized beam training results of the THBT. Based on the PSP and the time-frequency duality, the expressions of the subarray signals after analog combining are analytically derived and a BRPSS scheme with closed-form solutions is proposed for high-resolution channel parameter estimation.

After the THBT, the initial estimation of k and b in (9) can be expressed as

$$\tilde{k} = -\frac{\lambda(1 - \tilde{\Omega}^2)}{4\tilde{r}} \text{ and } \tilde{b} = \tilde{\Omega} - \tilde{k}(N + 1). \quad (47)$$

Note that the resolutions of \tilde{k} and \tilde{b} are limited by the quantization of the hybrid codebook. To improve the estimation accuracy of k and b , the user continues to transmit one uplink pilot and the BS receives that with the analog combiner

$$\overline{\mathbf{W}} = \text{blkdiag}\{\overline{\mathbf{w}}_1^H, \overline{\mathbf{w}}_2^H, \dots, \overline{\mathbf{w}}_{N_{\text{RF}}}^H\}, \quad (48)$$

where

$$[\overline{\mathbf{w}}_t]_m = e^{j\pi(\tilde{k}((t-1)M+m)^2 + \tilde{b}((t-1)M+m))} \quad (49)$$

for $m = 1, 2, \dots, M$ and $t = 1, 2, \dots, N_{\text{RF}}$. Note that $\overline{\mathbf{W}}$ is designed based on \tilde{k} and \tilde{b} . Thus, the near-field effects can be alleviated and directional beams are formed to provide high beamforming gains. Then the received signal of the t th subarray after analog combining can be expressed as

$$\begin{aligned} \hat{\mathbf{z}}_t^{(a)} &\approx \sum_{n=(t-1)M+1}^{tM} e^{-j\pi(\tilde{k}n^2 + \tilde{b}n)} [\mathbf{h}]_n \\ &\stackrel{(b)}{\approx} g \sum_{n=(t-1)M+1}^{tM} e^{-j\pi(\tilde{k}n^2 + \tilde{b}n)} [\boldsymbol{\alpha}(N, \Omega, r)]_n \\ &\stackrel{(c)}{\approx} \tilde{g} \sum_{n=(t-1)M+1}^{tM} e^{-j\pi(\tilde{k}n^2 + \tilde{b}n)} [\boldsymbol{\gamma}(N, \Omega, r)]_n \\ &= \tilde{g} \sum_{n=(t-1)M+1}^{tM} e^{j\pi(\Delta k n^2 + \Delta b n)} \\ &= \tilde{g} \sum_{m=1}^M e^{j\pi(\Delta k(m+(t-1)M)^2 + \Delta b(m+(t-1)M))} \\ &= \tilde{g} e^{j\pi(\Delta k(t-1)^2 M^2 + \Delta b(t-1)M)} \\ &\quad \cdot \sum_{m=1}^M e^{j\pi(\Delta k m^2 + \Delta b m + 2\Delta k m M(t-1))}, \end{aligned} \quad (50)$$

for $t = 1, 2, \dots, N_{\text{RF}}$, where $\Delta k \triangleq k - \tilde{k}$, $\Delta b \triangleq b - \tilde{b}$ and $\tilde{g} \triangleq g e^{j2\pi(\Omega(\delta_1 - 1/2) - \rho(N+1)^2/16)}$. In (50), we omit the noise term and set $x_p = 1$ in (a), omit the effects of the NLoS paths in (b), and approximate $\boldsymbol{\alpha}(N, \Omega, r)$ with $\boldsymbol{\gamma}(N, \Omega, r)$ in (c) to simplify the analysis. The summation in (50) is tricky due to the quadratic phase term $\Delta k m^2$. To obtain a deeper insight of $\hat{\mathbf{z}}_t$, we first focus on $e^{j\pi\Delta k m^2}$. Define $\mathbf{s} \triangleq [e^{j\pi\Delta k}, e^{j\pi4\Delta k}, \dots, e^{j\pi M^2\Delta k}]^T$. According to the PSP [29]–[31], the discrete-time Fourier transform of \mathbf{s} is expressed as

$$\begin{aligned} \mathcal{F}(\omega) &= \sum_{m=1}^M [\mathbf{s}]_m e^{-j\pi m \omega} \\ &\approx \begin{cases} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})} \sqrt{\frac{1}{-\Delta k}}, & 2\Delta k M \leq \omega \leq 2\Delta k, \\ 0, & \text{others,} \end{cases} \end{aligned} \quad (51)$$

where we assume $\Delta k < 0$ without loss of generality. Then, according to the time-frequency duality, we have

$$\begin{aligned} [\mathbf{s}]_m &= e^{j\pi\Delta k m^2} \\ &\approx \frac{1}{2} \int_{-1}^1 \mathcal{F}(\omega) e^{j\pi m \omega} d\omega \\ &= \frac{1}{2} \int_{2\Delta k M}^{2\Delta k} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})} \sqrt{\frac{1}{-\Delta k}} e^{j\pi m \omega} d\omega. \end{aligned} \quad (52)$$

Substituting (52) into (50), we have (53), which is shown at the top of the next page. In (53), we have

$$\begin{aligned} \bar{g} &\triangleq \frac{1}{2\sqrt{-\Delta k}} \tilde{g} e^{j\pi((M+1)\Delta b/2 - 1/4)}, \\ \Delta \tilde{k} &\triangleq \Delta k M^2, \end{aligned}$$

$$\begin{aligned}
\hat{z}_t &\approx \tilde{g} e^{j\pi(\Delta k(t-1)^2 M^2 + \Delta b(t-1)M)} \cdot \sum_{m=1}^M \left(\frac{1}{2} \int_{2\Delta kM}^{2\Delta k} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})} \sqrt{\frac{1}{-\Delta k}} e^{j\pi m\omega} d\omega \right) e^{j\pi(\Delta b m + 2\Delta k m M(t-1))} \\
&= \tilde{g} e^{j\pi(\Delta k(t-1)^2 M^2 + \Delta b(t-1)M)} \frac{1}{2} \int_{2\Delta kM}^{2\Delta k} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})} \sqrt{\frac{1}{-\Delta k}} \left(\sum_{m=1}^M e^{j\pi(\Delta b m + 2\Delta k m M(t-1))} e^{j\pi m\omega} \right) d\omega \\
&= \tilde{g} e^{j\pi(\Delta k(t-1)^2 M^2 + \Delta b(t-1)M)} \frac{1}{2} \sqrt{\frac{1}{-\Delta k}} \int_{2\Delta kM}^{2\Delta k} e^{-j\pi(\frac{\omega^2}{4\Delta k} + \frac{1}{4})} \left(e^{j(M+1)(\pi\phi_t + \pi\omega)/2} \frac{\sin(M(\pi\phi_t + \pi\omega)/2)}{\sin((\pi\phi_t + \pi\omega)/2)} \right) d\omega \\
&= \tilde{g} e^{j\pi(\Delta \tilde{k}(t-1)^2 + \Delta \tilde{b}(t-1))} \int_{2\Delta kM}^{2\Delta k} \underbrace{e^{j\pi(\frac{(M+1)\omega}{2} - \frac{\omega^2}{4\Delta k})}}_{\mathcal{P}(\omega)} \underbrace{\frac{\sin(M(\pi\phi_t + \pi\omega)/2)}{\sin((\pi\phi_t + \pi\omega)/2)}}_{\mathcal{A}(\phi_t, \omega)} d\omega \\
&= \tilde{g} \underbrace{e^{j\pi(\Delta \tilde{k}(t-1)^2 + \Delta \tilde{b}(t-1))}}_{\mathcal{C}(t)} \underbrace{\int_{2\Delta kM}^{2\Delta k} \mathcal{P}(\omega) \mathcal{A}(\phi_t, \omega) d\omega}_{\mathcal{B}(\phi_t)} \\
&= \tilde{g} \mathcal{C}(t) \mathcal{B}(\phi_t).
\end{aligned} \tag{53}$$

$$\begin{aligned}
\Delta \tilde{b} &\triangleq (\Delta b + \Delta k(M+1))M, \\
\phi_t &\triangleq \Delta b + 2\Delta k M(t-1).
\end{aligned} \tag{54}$$

Remark 1: In this part, we focus on providing high beamforming gains for all subarrays by analyzing the relations between the space quantization in (11) and the received signals of subarrays in (53). From (53), the power of \hat{z}_t is determined by the value of $|\mathcal{B}(\phi_t)|$. Moreover, $\mathcal{B}(\phi_t)$ is essentially the summation of $\mathcal{A}(\phi_t, \omega)$ with weighted factors $\mathcal{P}(\omega)$. To provide high beam gains for all subarrays, the maximum peak shift of $\mathcal{A}(\phi_t, \omega)$ should be less than the half-power width so that each term of $\mathcal{A}(\phi_t, \omega)$ can provide high beam gain for summation and the maximum phase shift of $\mathcal{P}(\omega)$ should be less than $\pi/2$ so that $\mathcal{A}(\phi_t, \omega)$ for different ω will not cancel each other. Then we have

$$|\phi_t + \omega| \leq \frac{1}{M}, \tag{55a}$$

$$\left| \frac{(M+1)\omega}{2} - \frac{\omega^2}{4\Delta k} \right| \leq 1/2, \tag{55b}$$

for $\omega \in [2\Delta kM, 2\Delta k]$. Substituting ϕ_t in (54) into (55a), we have

$$|\Delta k| \leq \frac{1}{M(N-1)} - \frac{1}{Q(N-1)}. \tag{56}$$

From (55b), we have

$$|\Delta k| \leq \frac{1}{4M^2}. \tag{57}$$

Note that

$$\frac{1}{M(N-1)} - \frac{1}{Q(N-1)} < \frac{1}{M(N-1)} \approx \frac{1}{N_{\text{RF}} M^2} \leq \frac{1}{4M^2} \tag{58}$$

for $N_{\text{RF}} \geq 4$, which is easy to be satisfied for XL-MIMO systems. Therefore, we omit the constraint in (57) and focus on (56) to simplify the expression. Note that

$$|\Delta k| \leq \frac{1}{4} \left| \frac{\lambda(1 - \Theta_q^2)}{2d_{q,s}} - \frac{\lambda(1 - \Theta_q^2)}{2d_{q,s+1}} \right|$$

$$= \frac{\sqrt{2}}{2N^{3/2}S}. \tag{59}$$

Combining (56) and (59), we have

$$Q \geq M, \text{ and } S \geq \frac{\sqrt{2}(N-1)}{2N^{3/2} \left(\frac{1}{M} - \frac{1}{Q} \right)}, \tag{60}$$

which provides guidance for the settings of S and Q .

In (53), the expression of \hat{z}_t is analytically derived, where the phase of \hat{z}_t shifts along with t . Specifically, the phase of \hat{z}_t is related to \tilde{g} , $\mathcal{C}(t)$, and $\mathcal{B}(\phi_t)$, where \tilde{g} is a constant irrelevant to t . In addition, the phase of $\mathcal{B}(\phi_t)$ is also irrelevant to t because $\mathcal{P}(\omega)$ is irrelevant to t and $\mathcal{A}(\phi_t, \omega)$ consistently takes on positive values under the constraint in (55a). Therefore, the unwrapped phase of \hat{z}_t is a quadratic function of t with $\Delta k M^2$ and $(\Delta b + \Delta k(M+1))M$ as its coefficients, which indicates that the parameters of the channels are estimated if we can obtain the coefficients of the quadratic function.

Define $\Upsilon_t \triangleq \angle \hat{z}_t$, where $\angle(\cdot)$ denotes the phase of a complex value. The first-order difference of Υ_t can be expressed as

$$\Delta \Upsilon_t \triangleq \Upsilon_{t+1} - \Upsilon_t = \pi(\Delta \tilde{k}(2t-1) + \Delta \tilde{b}), \tag{61}$$

for $t = 1, 2, \dots, N_{\text{RF}} - 1$. Then the second-order difference of Υ_t can be expressed as

$$\Delta^2 \Upsilon_t \triangleq \Delta \Upsilon_{t+1} - \Delta \Upsilon_t = 2\pi \Delta \tilde{k}, \tag{62}$$

for $t = 1, 2, \dots, N_{\text{RF}} - 2$. Note that

$$|\Delta^2 \Upsilon_t| = 2\pi |\Delta \tilde{k}| \stackrel{(a)}{<} \frac{2\pi M^2}{M(N-1)} = \frac{2\pi}{N_{\text{RF}}} \cdot \frac{N}{N-1} \stackrel{(b)}{<} \pi \tag{63}$$

where (a) can be obtained via (56) and (b) holds for $N_{\text{RF}} > 2$. To avoid phase wrap, we can shift $\Delta^2 \Upsilon_t$ to $[-\pi, \pi]$ by

$$\bar{\Upsilon}_t \triangleq \text{mod}(\Delta^2 \Upsilon_t + \pi, 2\pi) - \pi. \tag{64}$$

Then the estimation of Δk can be expressed as

$$\Delta \hat{k} = \frac{1}{N_{\text{RF}} - 2} \sum_{t=1}^{N_{\text{RF}}-2} \frac{\bar{\Upsilon}_t}{2\pi M^2}. \tag{65}$$

Algorithm 2 Beam Refinement based on Phase Shifts of Subarrays (BRPSS) Scheme

- 1: **Input:** N , N_{RF} , M , λ , \hat{z}_t , $\hat{\Omega}$ and \hat{r} .
- 2: Obtain \hat{k} and \hat{b} via (47).
- 3: Obtain $\mathcal{Y}_t = \angle \hat{z}_t$, for $t = 1, 2, \dots, N_{\text{RF}}$.
- 4: Obtain $\Delta \mathcal{Y}_t$ and $\Delta^2 \mathcal{Y}_t$ via (61) and (62), respectively.
- 5: Obtain $\hat{\Delta k}$ and $\hat{\Delta b}$ via (65) and (69), respectively.
- 6: Obtain \hat{k} and \hat{b} via (70).
- 7: Obtain $\hat{\Omega}$ and \hat{r} via (71).
- 8: **Output:** \hat{k} , \hat{b} , $\hat{\Omega}$ and \hat{r} .

Now we turn to the estimation of Δb . Define

$$\tilde{\mathcal{Y}}_t \triangleq \Delta \mathcal{Y}_t - (2t - 1)M^2 \Delta \hat{k} \pi - M(M + 1) \Delta \hat{k} \pi. \quad (66)$$

Then according to (61), we have

$$\tilde{\mathcal{Y}}_t = M \Delta b \pi + 2u\pi, \quad u \in \mathbb{N}. \quad (67)$$

Note that

$$|M \Delta b \pi| \leq \frac{M\pi}{Q} + \frac{(N+1)\pi}{(N-1)} - \frac{M(N+1)\pi}{Q(N-1)} \stackrel{(a)}{\approx} \pi. \quad (68)$$

where (a) holds because $(N+1)/(N-1) \approx 1$ for XL-MIMO systems. Therefore, we can also shift the results in (67) to $[-\pi, \pi]$ to avoid phase wrap. Define $\hat{\mathcal{Y}}_t \triangleq \text{mod}(\tilde{\mathcal{Y}}_t + \pi, 2\pi) - \pi$. Then the estimation of Δb can be expressed as

$$\Delta \hat{b} = \frac{1}{N_{\text{RF}} - 1} \sum_{t=1}^{N_{\text{RF}}-1} \frac{\hat{\mathcal{Y}}_t}{M\pi}. \quad (69)$$

Then the refined estimation of k and b can be expressed as

$$\hat{k} = \hat{\Delta k} + \tilde{k} \text{ and } \hat{b} = \Delta \hat{b} + \tilde{b}. \quad (70)$$

Accordingly, the refined estimation of r and Ω can be expressed as

$$\hat{\Omega} = \hat{b} + \hat{k}(N+1) \text{ and } \hat{r} = -\frac{\lambda(1 - \hat{\Omega}^2)}{2\hat{k}}. \quad (71)$$

Finally, we summarize the detailed procedures of the BRPSS scheme in **Algorithm 2**.

Remark 2: The overheads, computational complexity, and hardware requirements of the proposed BRPSS scheme are remarked here. Since the BRPSS is developed based on the phase shifts of the subarrays, only one pilot is needed for beam refinement benefiting from the subarray architecture of the partially-connected hybrid combining structure. The proposed BRPSS has closed-form expressions, which no longer needs running any algorithms and therefore has very low computational complexity. The computation of the BRPSS mainly comes from the computing of (61), (62), (65) and (69), which are all only related to N_{RF} . Therefore, the computational complexity of the BRPSS is $\mathcal{O}(N_{\text{RF}})$. Note that the unwrapped phase of \hat{z}_t is the quadratic function of the indices of subarrays. Therefore, at least three subarrays are needed for the BRPSS scheme.

V. NEAR-FIELD BEAM TRACKING

In this section, we propose a low-complexity NFBT scheme, where a kinematic model is adopted to characterize

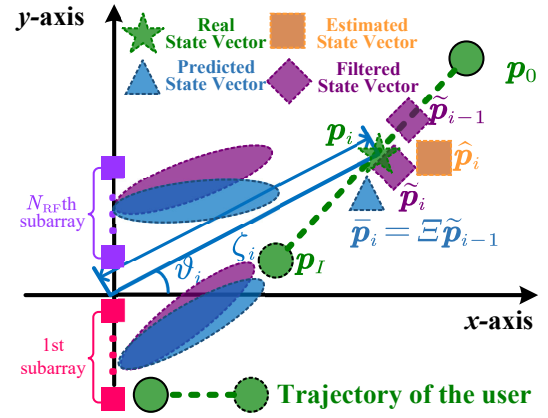


Fig. 5. Illustration of the near-field beam tracking procedure.

the channel variations at different time instants and the BRPSS scheme is used to estimate the real-time channel parameters. Then the kinematic model and real-time estimates are exploited by the EKF to track and predict the near-field channel parameters.

Different from the far-field channel that only depends on the angle of the user, the near-field channel is related to both the angle and the distance of the user. By performing the BRPSS scheme with the channel measurements, we can estimate the angles and distances of the user. Then, based on the estimated parameters, the real-time position of the user can be calculated. Given this, we adopt the kinematic model to characterize the channel variations at different time instants. In Fig. 5, we illustrate the near-field beam tracking procedure, where the center of the BS antennas is set as the origin, the normal direction of the BS antenna array is set as the x-axis and the direction along the BS antennas array is set as the y-axis. We assume the user communicates with the BS in a block-wise way, where the duration of one block is ΔT . The user moves along the trajectory for I blocks. Denote the kinematic parameters of the user at the i th block as $\mathbf{p}[i] = [a_x[i], a_y[i], v_x[i], v_y[i]]^T$, for $i = 0, 1, 2, \dots, I$, where $a_x[i]$, $a_y[i]$, $v_x[i]$, and $v_y[i]$ are the x-axis coordinate, the y-axis coordinate, the velocity component in the x-axis direction, and the velocity component in the y-axis direction, respectively. The angle and distance of the user at the i th block are expressed as ϑ_i and ζ_i , respectively. We then propose the NFBT scheme, which includes initialization, beam prediction, beam refinement, filtering, and stop conditions.

1) *Initialization:* We denote the initial kinematic parameters of the user as $\mathbf{p}[0] = [a_x[0], a_y[0], 0, 0]^T$, which can be obtained via the THBT and BRPSS schemes.

2) *Beam Prediction:* Denote the filtered state vector after performing the EKF at the $(i-1)$ th block for $i \geq 1$ as $\tilde{\mathbf{p}}[i-1]$, where $\tilde{\mathbf{p}}[0] = \mathbf{p}[0]$. Then the predicted state vector at the i th block can be expressed as

$$\bar{\mathbf{p}}[i] = \underbrace{\begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{E}} \tilde{\mathbf{p}}[i-1] \quad (72)$$

Algorithm 3 Near-Field Beam Tracking (NFBT) Scheme

```

1: Input:  $N, N_{\text{RF}}, M, \lambda, \mathbf{p}[0]$  and  $\Delta T$ .
2: Initialization:  $i \leftarrow 1$ .
3: while stop conditions in Sec. V-5 are not satisfied do
4:   Obtain  $\bar{\mathbf{p}}[i]$  via (72).
5:   Obtain  $\hat{\mathbf{p}}[i]$  via (78).
6:   Obtain  $\tilde{\mathbf{p}}[i]$  via the EKF.
7:   Obtain  $\tilde{\mathbf{h}}_i$  via (79).
8:    $i \leftarrow i + 1$ .
9: end while
10: Output:  $\tilde{\mathbf{h}}_i$ .

```

where the predicted distance and angle can be respectively expressed as

$$\bar{\zeta}_i = \sqrt{\bar{a}_x[i]^2 + \bar{a}_y[i]^2} \text{ and } \bar{\vartheta}_i = \arctan(\bar{a}_y[i]/\bar{a}_x[i]). \quad (73)$$

3) *Beam Refinement:* In the stage of beam refinement, the user transmits one uplink pilot and the BS receives the signal with the analog combiner

$$\vec{\mathbf{W}}_i = \text{blkdiag}\{\vec{\mathbf{w}}_1^H, \vec{\mathbf{w}}_2^H, \dots, \vec{\mathbf{w}}_{N_{\text{RF}}}^H\}, \quad (74)$$

where

$$[\vec{\mathbf{w}}_t]_m = e^{j\pi(\bar{k}_i((t-1)M+m)^2 + \bar{b}_i((t-1)M+m))} \quad (75)$$

for $m = 1, \dots, M$ and $t = 1, 2, \dots, N_{\text{RF}}$. In (75),

$$\bar{k}_i = -\frac{\lambda(1 - \sin^2 \bar{\vartheta}_i)}{4\bar{\zeta}_i} \text{ and } \bar{b}_i = -\sin \bar{\vartheta}_i - \bar{k}_i(N+1) \quad (76)$$

which are calculated based on the predicted distance and angle in (73). Then the received signal of the t th subarray after analog combining can be expressed as

$$\bar{z}_t = \vec{\mathbf{w}}_t^H \mathbf{G}_t \bar{\mathbf{h}}_i + \vec{\mathbf{w}}_t^H \mathbf{G}_t \bar{\boldsymbol{\eta}}_i, \quad (77)$$

for $t = 1, 2, \dots, N_{\text{RF}}$, where $\bar{\mathbf{h}}_i$ denotes the channel between the user and the BS at the i th block and $\bar{\boldsymbol{\eta}}_i$ denotes the AWGN. Substituting $N, N_{\text{RF}}, M, \lambda, \bar{z}_t, -\sin \bar{\vartheta}_i$ and $\bar{\zeta}_i$ into **Algorithm 2**, we can obtain the estimated angle $\hat{\vartheta}_i$ and the estimated distance $\hat{\zeta}_i$ of the user at the i th block. Then the estimated state vector can be expressed as

$$\hat{\mathbf{p}}[i] = [\hat{\zeta}_i \cos \hat{\vartheta}_i, \hat{\zeta}_i \sin \hat{\vartheta}_i, 0, 0]^T. \quad (78)$$

4) *Filtering:* The estimated state vector, $\hat{\mathbf{p}}[i]$, and the predicted state vector, $\bar{\mathbf{p}}[i]$, can be exploited by the widely-adopted EKF for effective tracking. Since the EKF has been exhaustively introduced in a variety of works [32]–[34], we omit the details and denote the filtered state vector at the i th block after performing the EKF as $\tilde{\mathbf{p}}[i]$. Then the estimated channel steering vector of the LoS path at the i th block can be expressed as

$$\begin{aligned} [\tilde{\mathbf{h}}_i]_n &= \frac{1}{\sqrt{N}} e^{j\pi(\tilde{k}_i n^2 + \tilde{b}_i n)} \\ \tilde{k}_i &= -\frac{\lambda(1 - \sin^2 \tilde{\vartheta}_i)}{4\tilde{\zeta}_i}, \quad \tilde{b}_i = -\sin \tilde{\vartheta}_i - \tilde{k}_i(N+1), \\ \tilde{\zeta}_i &= \sqrt{\tilde{a}_x[i]^2 + \tilde{a}_y[i]^2}, \text{ and } \tilde{\vartheta}_i = \arctan(\tilde{a}_y[i]/\tilde{a}_x[i]), \end{aligned} \quad (79)$$

for $n = 1, 2, \dots, N$, which can be used to design the hybrid combiners for communication.

5) *Stop Condition:* We repeat Section V-2, Section V-3, and Section V-4 until the predefined maximum number of blocks I is achieved or the communication process is completed.

In Fig. 5, we illustrate the process of beam training using the i th block as an example. Suppose we have obtained the filtered state vector in the $(i-1)$ th block, $\tilde{\mathbf{p}}[i-1]$. Based on the kinematic model in (72), we can obtain the predicted state vector in the i th block, $\bar{\mathbf{p}}[i]$. Using $\bar{\mathbf{p}}[i]$ as a reference, the BS designs the analog combiner $\vec{\mathbf{W}}_i$ and each subarray independently receives the uplink pilot with $\vec{\mathbf{W}}_i$. Based on the measurements of subarrays, the BS performs the BRPSS scheme to obtain the estimated state vector, $\hat{\mathbf{p}}[i]$. Then, by exploiting the EKF, the BS fuses $\bar{\mathbf{p}}[i]$ and $\hat{\mathbf{p}}[i]$ to obtain the more accurate filtered state vector $\tilde{\mathbf{p}}[i]$. During the process of beam tracking, for each block, the BS only receives one uplink pilot to obtain the estimated state vector. Therefore, the overhead of the proposed NFBT scheme is one.

Finally, we summarize the detailed procedures of the proposed NFBT scheme in **Algorithm 3**.

Remark 3: The proposed schemes are adaptable to the multi-user scenarios with three extensions. Firstly, following the works in [9], we can allocate mutually orthogonal pilots to N_{UE} users and separately perform the proposed schemes for each user. Secondly, following the works in [35], we allocate mutually orthogonal pilots to the N_{RF} RF chains and perform the downlink beam training/tracking. Then, the users can concurrently perform the proposed schemes. Thirdly, following the works in [10], we perform uplink beam training for the N_{UE} users simultaneously, where \hat{L} channel paths are detected sequentially based on the beam training results. Then, the users transmit orthogonal pilots to determine the association of the \hat{L} detected paths.

VI. SIMULATION RESULTS

Now we evaluate the performance of the proposed schemes. We consider an XL-MIMO system equipped with $N = 512$ antennas [7]. The antenna array is composed of $N_{\text{RF}} = 4$ subarrays with each subarray having $M = 128$ antennas [8]. The wavelength is set to be $\lambda = 0.003$ m corresponding to the carrier frequency of 100 GHz [9]. The channel between the user and the BS is set up with $L = 3$ channel paths with one LoS path and two NLoS paths, where the channel gain of the LoS path obeys $g_1 \sim \mathcal{CN}(0, 1)$ and the NLoS paths obey $g_2 \sim \mathcal{CN}(0, 0.01)$ and $g_3 \sim \mathcal{CN}(0, 0.01)$ [3]. Channel angle Ω_l of the l th path obeys the uniform distribution between $[-\sqrt{3}/2, \sqrt{3}/2]$ [9]. The HFBS, far-field beam sweeping (FFBS) [36], TPBT [15], DHBT [16] and P-SOMP [9] are adopted as the benchmarks. The parameters of different schemes are set in the **Parameter Settings** of Table I, where K, V and R denote the number of candidate angles, the number of layers in the hierarchical codebook and the number of space samples in each layer.

TABLE I
OVERHEADS FOR DIFFERENT BEAM TRAINING SCHEMES.

Schemes	Overhead	Parameter Settings	Calculated Training Overhead
HFBS	$Q(S + 1)$	$Q = 512, S = 11$	6144
FFBT [36]	Q	$Q = 512$	512
TPBT [15]	$Q + K(S + 1)$	$Q = 512, K = 3, S = 11$	548
DHBT [16]	VR	$V = 2, R = 256$	512
P-SOMP [9]	P	$P = 128$	128
Proposed THBT	M	$M = 128$	128
Proposed THBT with BRPSS	$M + 1$	$M = 128$	129

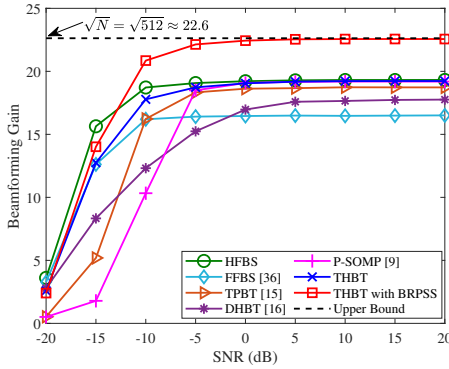


Fig. 6. Comparisons of beamforming gains for different schemes.

A. Evaluation of the THBT and BRPSS Schemes

In Fig. 6, we compare the proposed schemes with the HFBS, FFBS, TPBT, DHBT and P-SOMP in terms of beamforming gains for different SNRs. Suppose \mathbf{f} is the estimated channel steering vector. Then, the beamforming gain after beam alignment is defined as

$$\xi = \max_l \frac{|g_l|}{g_m} |\alpha(N, \Omega_l, r_l)^H \mathbf{f}|, \quad (80)$$

where $g_m = \max_l |g_l|$. The distances between the BS and the user or scatterers obey the uniform distribution between [6, 150] m, where the lower bound is set according to the reactive distance [9] while the upper bound is set according to the effective Rayleigh distance [28]. The hybrid-field codebook is designed with $Q = 512$ and $S = 11$, which is also used to generate the dictionary of the P-SOMP for fair comparison. From Fig. 6, the HFBS can achieve better performance than the other schemes when the SNR is less than -10 dB, which lies in the fact that the HFBS exhaustively tests all the codewords in \mathcal{C}_h and needs far more times of beam training than the other schemes. In addition, the proposed THBT with the BRPSS scheme performs the best among all the schemes when the SNR is larger than -10 dB because the performance of all the other schemes is limited by the quantized error of the codebook or dictionary while the proposed THBT with the BRPSS scheme can achieve high-resolution estimation thanks to the beam refinement of the BRPSS. Moreover, the performance of the proposed THBT scheme can approach the performance of the HFBS at various SNR conditions. The performance of P-SOMP is worse than that of the other schemes at low SNRs, such as -15 dB, because the random beamforming of P-SOMP cannot achieve

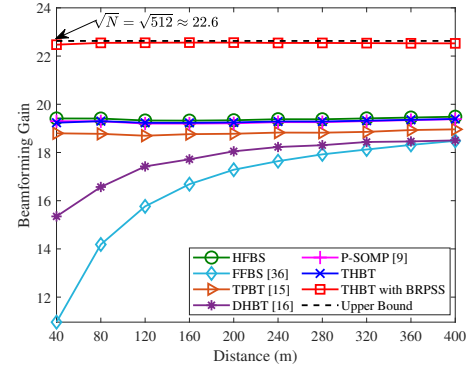


Fig. 7. Comparisons of the beamforming gains for different schemes with varying distances.

enough beamforming gain and will significantly degrade the performance. The performance of the FFBS is worse than that of the other schemes at high SNRs because the beamforming gain of the far-field channel steering vector will decrease in the near field. Furthermore, the TPBT performs worse than the proposed THBT; the justification is that the first phase of TPBT may select the incorrect middle- K angles due to the effects of the noise and the NLoS paths. The DHBT performs worse than the proposed THBT due to the neglect of the polar-domain sparsity.

In Fig. 7, we compare the proposed schemes with the HFBS, FFBS, TPBT, DHBT and P-SOMP in terms of beamforming gains for different distances. The distances between the BS and the user or scatterers obey the uniform distribution between [6, r] m, where r ranges from 40 to 400. Note that 400 m is approximately the Rayleigh distance of the considered XL-MIMO system. The SNR is fixed to be 10 dB. From Fig. 7, the proposed THBT with the BRPSS scheme achieves the highest beam gain at different distances thanks to the high-resolution BRPSS scheme. As the distance decreases, the FFBS will suffer from severe loss of beamforming gain because it only considers the far-field channels. In addition, the performance of DHBT deteriorates with the decrease of the distance as it neglects the polar-domain sparsity, which makes it sensitive to the near-field effects. By contrast, both the THBT scheme and the THBT with BRPSS scheme are robust to the distance, which indicates that the proposed beam training and beam refinement schemes are adaptive to both the near field and the far field.

In Fig. 8, we evaluate the positioning performance of the proposed schemes, where the SNR is set to be 20 dB. The

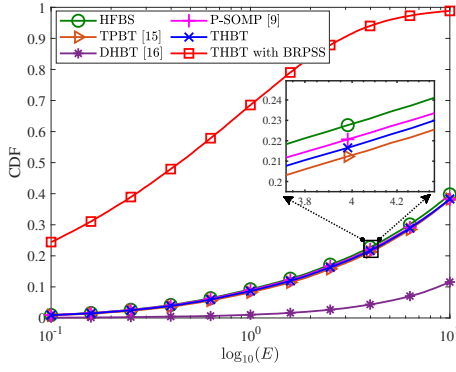


Fig. 8. Evaluation of the positioning performance for different schemes.

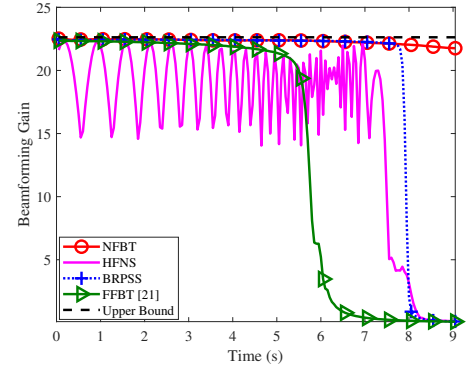


Fig. 10. Comparisons of the beamforming gains during the tracking process.

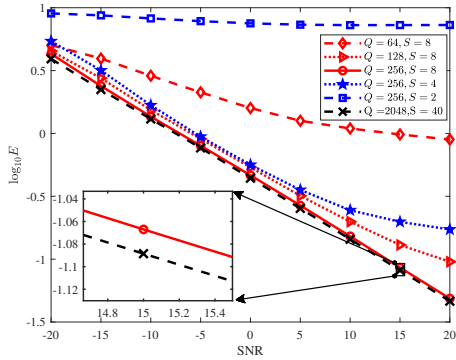


Fig. 9. Evaluation of the BRPSS for different settings of Q and S .

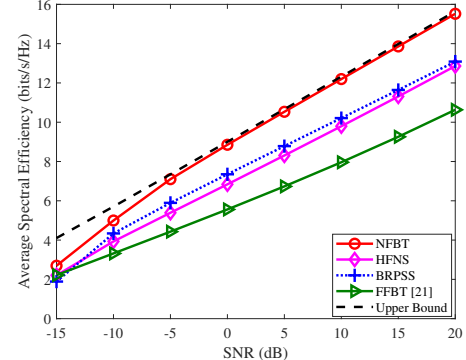


Fig. 11. Comparisons of the spectral efficiency during the tracking process.

distances between the BS and the user or scatterers obey the uniform distribution between $[6, 150]$ m, which is the same as in Fig. 6. Since the FFBS cannot obtain the position of the user, we only compare the proposed schemes with HFBS, TPBT, DHBT and P-SOMP. We denote the distance between the real position and the estimated position as E and use 10^5 times of Monte Carlo simulation to calculate the cumulative distribution function (CDF). From Fig. 8, the proposed THBT with the BRPSS scheme has the smallest positioning error among all the schemes due to the high-resolution estimation of the BRPSS scheme. In addition, the HFBS, TPBT, THBT and P-SOMP share similar performance because they are based on the same quantized angles and distances. The DHBT performs worse than the other schemes due to the neglect of the polar-domain sparsity.

In Table. I, we compare the training overheads of different schemes. The training overheads of the HFBS, FFBS, TPBT, DHBT, P-SOMP, the proposed THBT, and the proposed THBT with the BRPSS are $Q(S+1)$, Q , $Q+K(S+1)$, VR , P , M and $M+1$, respectively. Under the simulation settings, these seven schemes require 6144, 512, 548, 512, 128, 128, and 129 times of beam training, respectively. Specifically, the performance of the proposed THBT scheme can approach that of the HFBS with 97.92% reduction in training overhead, and the proposed THBT with the BRPSS scheme outperforms the HFBS with 97.90% reduction in training overhead for high SNRs.

In Fig. 9, we evaluate the performance of the BRPSS for different settings of Q and S . To avoid the effects of the NLoS paths and the prior beam training failure, we set $L = 1$ and suppose that the beam training has found the codeword best fitting for the channel in the hybrid-field codebook. The distances between the BS and the user obey the uniform distribution between $[10, 30]$ m. From Fig. 9, if we fix $Q = 256$ or $S = 8$, the positioning error decreases with the increase of S or Q . This is because beam gain losses of subarrays will occur if the conditions of Q and S in (60) are not satisfied, which will deteriorate the performance of beam refinement. However, if the conditions of Q and S in (60) are satisfied, i.e. $Q = 256$ and $S = 8$, continuing to increase Q and S can only slightly improve the performance of beam refinement, which verifies that (60) is appropriate for the settings of Q and S .

B. Evaluation of the NFBT Scheme

Now we evaluate the performance of the proposed NFBT scheme. We set $\mathbf{p}_0 = [50, 50\sqrt{3}, 0, 0]$, $\Delta T = 0.05$ s, $I = 180$, and $[v_x[i], v_y[i]]^T = [-5, -5\sqrt{3}]^T$ for $i = 1, 2, \dots, I$.

In Fig. 10, we evaluate the tracking performance of different schemes in terms of the beamforming gains during the tracking process. We extend the neighboring search in [24] to the hybrid-field neighboring search (HFNS), which tests the additional four codewords adjacent to the best codeword

TABLE II
OVERHEADS FOR DIFFERENT BEAM TRACKING SCHEMES.

Schemes	Overhead
NFBT	1
HFNS	5
BRPSS	1
FFBT [21]	3

of the previous block. The BRPSS and the far-field beam tracking (FFBT) [21] are also adopted as benchmarks, where the BRPSS is performed based on the estimation of the previous block. The SNR is set to be 0 dB. From the figure, the performance of the HFNS drops dramatically at the 7.5th second and that of the BRPSS drops dramatically at the 8th second because the deviation between the positions of the adjacent blocks is too large and the deteriorated beamforming gains cannot support effective beam tracking. In addition, the performance of the FFBT drops dramatically at the 6th second because the near-field channel differs gradually from the far-field channel with the decrease of the distance. However, the proposed NFBT scheme can maintain high beamforming gains during the whole process of beam tracking. It is also worth noting that the performance of the proposed NFBT scheme decreases slightly with the increase of time. This is because the near-field effect strengthens gradually with the decrease of distance and the same tracking error will lead to larger beamforming gain loss for smaller distances.

In Fig. 11, we compare the proposed NFBT with the HFNS, BRPSS and FFBT in terms of the average spectral efficiency during the tracking process for different SNRs. From the figure, the proposed NFBT performs the best among the four schemes and can approach the upper bound when SNR is larger than -5 dB, which demonstrates the effectiveness of the proposed schemes.

In Table II, we also compare the overheads of different beam tracking schemes. Since the HFNS tests the best codeword of the previous block and additional four codewords adjacent to the best codeword, it needs five times of beam training. In addition, the FFBT tests the best codeword of the previous block and the additional two codewords adjacent to the best codeword, and it needs three times of beam training. According to Section IV, the overhead of the BRPSS scheme is one. Similarly, the overhead of the NFBT scheme, which is based on the BRPSS scheme, is also one. In summary, the proposed NFBT scheme outperforms existing ones with lower overhead.

VII. CONCLUSION

In this paper, a THBT scheme has been proposed. In the first stage, each subarray independently uses multiple far-field channel steering vectors to approximate near-field channel steering vectors for analog combining. In the second stage, digital combiners are designed to combine the outputs of the analog combiners from the first stage to find the codeword in the predefined hybrid-field codebook best fitting for the channel. Then, based on the PSP and the time-frequency duality, the expressions of subarray signals after

analog combining have been analytically derived, and the BRPSS scheme with closed-form solutions has been proposed for high-resolution channel parameter estimation. Moreover, a low-complexity NFBT scheme has been proposed, where the kinematic model has been adopted to characterize the channel variations at different time instants and the EKF has been exploited for beam tracking. For future works, we will try to perform effective beam refinement for XL-MIMO systems with the power of received signals.

APPENDIX A

First of all, we calculate the maximum received powers with the near-field and the far-field steering vectors, respectively. Suppose the t th subarray receives signals with far-field steering vector $\sqrt{M}\beta(M, \Omega)$. Then, the absolute value of the received signal after analog combining can be calculated as

$$\begin{aligned}
 |G(\mathbf{G}_t \mathbf{h}, \Omega)| &= \sqrt{M}\beta(M, \Omega)^H \mathbf{G}_t \mathbf{h} \\
 &\approx \left| \sum_{n=(t-1)M+1}^{tM} e^{j\pi(kn^2 + bn)} e^{-j\pi n\Omega} \right| \\
 &= \left| \sum_{n=1}^M e^{j\pi(kn^2 + \bar{b}n)} e^{-j\pi n\Omega} \right| \\
 &\approx \left| \int_1^M e^{j\pi(k\chi^2 + (\bar{b}-\Omega)\chi)} d\chi \right| \\
 &= \left| \int_{-\infty}^{\infty} U(\chi) e^{j\pi(k\chi^2 + (\bar{b}-\Omega)\chi)} d\chi \right|, \quad (81)
 \end{aligned}$$

where

$$\bar{b} \triangleq b + 2k(t-1)M, \quad U(\chi) = \begin{cases} 1, & 1 \leq \chi \leq M, \\ 0, & \text{others.} \end{cases} \quad (82)$$

According to the PSP [29]–[31], (81) can be approximated as

$$\begin{aligned}
 |G(\mathbf{G}_t \mathbf{h}, \Omega)| &\approx \left| \sqrt{\frac{-2\pi}{\Phi''(\chi_0, \Omega)}} U(\chi_0) \right| \\
 &= \begin{cases} \sqrt{\frac{1}{-k}}, & \bar{b} + 2kM \leq \Omega \leq \bar{b} + 2k, \\ 0, & \text{others,} \end{cases} \quad (83)
 \end{aligned}$$

where $\Phi(\chi, \Omega) \triangleq \pi(k\chi^2 + (\bar{b} - \Omega)\chi)$ and $\chi_0 = \frac{\Omega - \bar{b}}{2k}$. The relation in (83) indicates that the maximum received powers with near-field channel steering vectors can be approximated as $\sqrt{\frac{1}{-k}}$. On the other hand, when the analog combiner is aligned with the channel steering vector, the modulus of the maximum received signal after analog combining is M . Therefore the beam gain loss of replacing the near-field channel steering vector with the far-field channel steering vector can be approximated as

$$\begin{aligned}
 \Gamma &= 1 - \frac{\max_{\Omega} |G(\mathbf{G}_t \mathbf{h}, \Omega)|}{M} \\
 &= 1 - \sqrt{\frac{4r}{\lambda(1 - \Omega^2)M^2}} \\
 &\stackrel{(a)}{\leq} 1 - \sqrt{\frac{4r}{\lambda M^2}}
 \end{aligned}$$

$$\begin{aligned} &\stackrel{(b)}{\leq} 1 - \sqrt{\frac{2\sqrt{D^3/\lambda}}{\lambda M^2}} \\ &= 1 - \frac{N_{\text{RF}}}{\sqrt[4]{2N}}, \end{aligned} \quad (84)$$

where the equation in (a) holds when $\Omega = 0$ and the equation in (b) holds when $r = 0.5\sqrt{D^3/\lambda}$. Since the beam gain loss cannot be less than 0, we normalize the maximum beam gain loss as

$$\Gamma_{\max} = \max\{1 - N_{\text{RF}}/\sqrt[4]{2N}, 0\}, \quad (85)$$

which completes the proof.

APPENDIX B

First of all, we calculate the beam center of $\mathbf{G}_t\boldsymbol{\alpha}(N, \Omega, r)$. Based on (83), we can approximate the beam gain of $\mathbf{G}_t\boldsymbol{\alpha}(N, \Omega, r)$ as

$$|G(\mathbf{G}_t\boldsymbol{\alpha}(N, \Omega, r), \Omega)| = \begin{cases} \sqrt{\frac{1}{-k}}, & \bar{b} + 2kM \leq \Omega \leq \bar{b} + 2k, \\ 0, & \text{others,} \end{cases} \quad (86)$$

which indicates that the beam coverage of $\mathbf{G}_t\boldsymbol{\alpha}(N, \Omega, r)$ is $[\bar{b} + 2kM, \bar{b} + 2k]$. Therefore, the beam center of $\mathbf{G}_t\boldsymbol{\alpha}(N, \Omega, r)$ can be calculated as

$$\begin{aligned} B_t &= \bar{b} + k(M + 1) \\ &= \Omega + \frac{\lambda(1 - \Omega^2)(N - (2t - 1)M)}{4r}. \end{aligned} \quad (87)$$

Then we calculate Ψ_t , which denotes the sine result of the angle that points from the center of the t th subarray to the user. As shown in Fig. 2, the center of the t th subarray is $(0, \Delta_t\lambda)$, where $\Delta_t = [(2t - 1)M - N]/4$. Then Ψ_t can be computed as

$$\begin{aligned} \Psi_t &= \frac{r\Omega - \Delta_t\lambda}{\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}} \\ &= \frac{r\Omega - \Omega^2\Delta_t\lambda}{\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}} + \frac{(\Omega^2 - 1)\Delta_t\lambda}{\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}} \\ &\stackrel{(a)}{\approx} \frac{r\Omega - \Omega^2\Delta_t\lambda}{r - \Omega\Delta_t\lambda} + \frac{(\Omega^2 - 1)\Delta_t\lambda}{r} \\ &= \Omega + \frac{\lambda(1 - \Omega^2)(N - (2t - 1)M)}{4r}, \end{aligned} \quad (88)$$

where the former of (a) holds because we approximate $\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}$ with its first-order Taylor series and the latter of (a) holds because we approximate $\sqrt{r^2 + \Delta_t^2\lambda^2 - 2r\Omega\Delta_t\lambda}$ with r . We use different approximations in (a) of (88) because the numerator of the former is much larger than that of the latter, which implies that the denominator of the former needs a more accurate approximation than that of the latter. Comparing (87) and (88), we have

$$B_t \approx \Psi_t, \quad (89)$$

which completes the proof.

REFERENCES

[1] K. Chen, C. Qi, and C.-X. Wang, "Two-stage hybrid-field beam training for ultra-massive MIMO systems," in *IEEE/CIC Int. Conf. Commun. China (ICCC)*, Foshan, China, Aug. 2022, pp. 1074–1079.

[2] Z. Chen, B. Ning, C. Han, Z. Tian, and S. Li, "Intelligent reflecting surface assisted Terahertz communications toward 6G," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 110–117, Dec. 2021.

[3] 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, Sep. 2020.

[4] B. Ning, Z. Tian, W. Mei, Z. Chen, C. Han, S. Li, J. Yuan, and R. Zhang, "Beamforming technologies for ultra-massive MIMO in terahertz communications," *IEEE Open J. Commun. Society*, vol. 4, pp. 614–658, Feb. 2023.

[5] C.-X. Wang, X. You, X. Gao *et al.*, "On the road to 6G: Visions, requirements, key technologies and testbeds," *IEEE Commun. Surv. Tut.*, vol. 25, no. 2, pp. 905–974, 2nd Quart., 2023.

[6] X. You, C.-X. Wang, J. Huang *et al.*, "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, no. 1, pp. 1–74, Nov. 2020.

[7] Z. Wu and L. Dai, "Multiple access for near-field communications: SDMA or LDMA?" *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1918–1935, June 2023.

[8] X. Song, T. Kühne, and G. Caire, "Fully-/partially-connected hybrid beamforming architectures for mmWave MU-MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1754–1769, Mar. 2020.

[9] 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, Jan. 2022.

[10] S. Noh, J. Song, and Y. Sung, "Fast beam search and refinement for millimeter-wave massive MIMO based on two-level phased arrays," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6737–6751, Oct. 2020.

[11] 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.

[12] K. T. Selvan and R. Janaswamy, "Fraunhofer and fresnel distances : Unified derivation for aperture antennas," *IEEE Antennas Propag. Mag.*, vol. 59, no. 4, pp. 12–15, Aug. 2017.

[13] 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., early access*, 2023.

[14] C. Qi, P. Dong, W. Ma, H. Zhang, Z. Zhang, and G. Y. Li, "Acquisition of channel state information for mmWave massive MIMO: Traditional and machine learning-based approaches," *Sci. China Inf. Sci.*, vol. 64, no. 8, p. 181301, Aug. 2021.

[15] 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.

[16] 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.

[17] H.-L. Song and Y.-C. Ko, "Robust and low complexity beam tracking with monopulse signal for UAV communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3505–3513, Mar. 2021.

[18] J. A. Zhang, K. Wu, X. Huang, and Y. J. Guo, "Beam alignment for analog arrays based on Gaussian approximation," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8152–8157, June 2023.

[19] D. Zhu, J. Choi, and R. W. Heath, "Auxiliary beam pair enabled AoD and AoA estimation in closed-loop large-scale millimeter-wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4770–4785, July 2017.

[20] C. You, Y. Zhang, C. Wu, Y. Zeng, B. Zheng, L. Chen, L. Dai, and A. L. Swindlehurst, "Near-field beam management for extremely large-scale array communications," *arXiv preprint arXiv:2306.16206*, 2023.

[21] D. Zhu, J. Choi, Q. Cheng, W. Xiao, and R. W. Heath, "High-resolution angle tracking for mobile wideband millimeter-wave systems with antenna array calibration," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7173–7189, Nov. 2018.

[22] J. Tan and L. Dai, "Wideband beam tracking in THz massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1693–1710, June 2021.

[23] C. Liu, M. Li, L. Zhao, P. Whiting, S. V. Hanly, I. B. Collings, and M. Zhao, "Robust adaptive beam tracking for mobile millimetre wave communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1918–1934, Mar. 2021.

[24] B. Ning, Z. Chen, Z. Tian, C. Han, and S. Li, "A unified 3D beam training and tracking procedure for terahertz communication," *IEEE Trans. Wireless Commun.*, vol. 21, no. 4, pp. 2445–2461, Apr. 2022.

- [25] W. Zuo, J. Xin, H. Ohmori, N. Zheng, and A. Sano, "Subspace-based algorithms for localization and tracking of multiple near-field sources," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 156–171, Mar. 2019.
- [26] A. Guerra, F. Guidi, D. Dardari, and P. M. Djurić, "Near-field tracking with large antenna arrays: Fundamental limits and practical algorithms," *IEEE Trans. Signal Process.*, vol. 69, pp. 5723–5738, Aug. 2021.
- [27] S. Palmucci, A. Guerra, A. Abrardoand, and D. Dardari, "Two-timescale joint precoding design and RIS optimization for user tracking in near-field MIMO systems," *arXiv preprint arXiv:2212.07333*, 2023.
- [28] M. Cui, L. Dai, R. Schober, and L. Hanzo, "Near-field wideband beamforming for extremely large antenna arrays," *arXiv preprint arXiv:2109.10054*, 2021.
- [29] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*. New York, NY, USA: McGraw-Hill, 1978.
- [30] F. W. J. Olver, *Asymptotics and Special Functions*. Boca Raton, FL, USA: CRC Press, 1997.
- [31] N. M. Temme, *Asymptotic Methods for Integrals*. Singapore: World Scientific, vol. 11, 2014.
- [32] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Cliffs, NJ, USA: Prentice-Hall, 1998.
- [33] F. Liu, Y. Cui, C. Masouros *et al.*, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Mar. 2022.
- [34] W. Yuan, F. Liu, C. Masouros, J. Yuan, D. W. K. Ng, and N. González-Prelcic, "Bayesian predictive beamforming for vehicular networks: A low-overhead joint radar-communication approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1442–1456, Nov. 2021.
- [35] Z. Xiao, P. Xia, and X.-G. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 141–153, Jan. 2017.
- [36] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

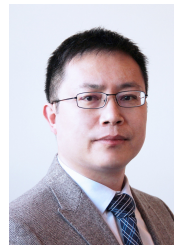


Kangjian Chen (Student Member, IEEE) received the B.S. degree from Nanjing University of Science and Technology, Nanjing, China, in 2017, and the M.S. degree in signal processing from Southeast University, Nanjing, in 2020, where he is currently pursuing the Ph.D. degree. His current research interests include integrated sensing and communication (ISAC), massive MIMO, and reconfigurable intelligent surface (RIS). He received the Best Paper Awards from IEEE Global Communications Conference (GLOBECOM) in 2019 and IEEE/CIC International Conference on Communications in China (ICCC) in 2022.



Chenhao Qi (Senior Member, IEEE) received the B.S. degree (Hons.) in information engineering from the Chien-Shiung Wu Honored College, Southeast University, China, in 2004, and the Ph.D. degree in signal and information processing from Southeast University in 2010.

From 2008 to 2010, he visited the Department of Electrical Engineering, Columbia University, New York, USA. Since 2010, he has been a Faculty Member with the School of Information Science and Engineering, Southeast University, where he is currently a Professor and the Head of Jiangsu Multimedia Communication and Sensing Technology Research Center. He received Best Paper Awards from IEEE GLOBECOM in 2019, IEEE/CIC ICC in 2022, and the 11th International Conference on Wireless Communications and Signal Processing (WCSP) in 2019. He has served as an Associate Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE COMMUNICATIONS LETTERS, IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY, and CHINA COMMUNICATIONS.



Cheng-Xiang Wang (Fellow, IEEE) received the B.Sc. and M.Eng. degrees in communication and information systems from Shandong University, China, in 1997 and 2000, respectively, and the Ph.D. degree in wireless communications from Aalborg University, Denmark, in 2004.

He was a Research Assistant with the Hamburg University of Technology, Hamburg, Germany, from 2000 to 2001, a Visiting Researcher with Siemens AG Mobile Phones, Munich, Germany, in 2004, and a Research Fellow with the University of Agder, Grimstad, Norway, from 2001 to 2005. He has been with Heriot-Watt University, Edinburgh, U.K., since 2005, where he was promoted to a Professor in 2011. In 2018, he joined Southeast University, Nanjing, China, as a Professor. He is also a part-time Professor with Purple Mountain Laboratories, Nanjing. He has authored 4 books, 3 book chapters, and 520 papers in refereed journals and conference proceedings, including 27 highly cited papers. He has also delivered 24 invited keynote speeches/talks and 16 tutorials in international conferences. His current research interests include wireless channel measurements and modeling, 6G wireless communication networks, and electromagnetic information theory.

Dr. Wang is a Member of the Academia Europaea (The Academy of Europe), a Member of the European Academy of Sciences and Arts (EASA), a Fellow of the Royal Society of Edinburgh (FRSE), IEEE, IET and China Institute of Communications (CIC), an IEEE Communications Society Distinguished Lecturer in 2019 and 2020, a Highly-Cited Researcher recognized by Clarivate Analytics in 2017–2020. He is currently an Executive Editorial Committee Member of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He has served as an Editor for over ten international journals, including the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, from 2007 to 2009, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, from 2011 to 2017, and the IEEE TRANSACTIONS ON COMMUNICATIONS, from 2015 to 2017. He was a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, Special Issue on Vehicular Communications and Networks (Lead Guest Editor), Special Issue on Spectrum and Energy Efficient Design of Wireless Communication Networks, and Special Issue on Airborne Communication Networks. He was also a Guest Editor for the IEEE TRANSACTIONS ON BIG DATA, Special Issue on Wireless Big Data, and is a Guest Editor for the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, Special Issue on Intelligent Resource Management for 5G and Beyond. He has served as a TPC Member, a TPC Chair, and a General Chair for more than 30 international conferences. He received 15 Best Paper Awards from IEEE GLOBECOM 2010, IEEE ICCT 2011, ITST 2012, IEEE VTC 2013 Spring, IWCMC 2015, IWCMC 2016, IEEE/CIC ICC 2016, WPMC 2016, WOCC 2019, IWCMC 2020, WCSP 2020, CSPS2021, WCSP 2021, and IEEE/CIC ICC 2022.



Geoffrey Ye Li (Fellow, IEEE) is currently a Chair Professor at Imperial College London, UK. Before joining Imperial in 2020, he was a Professor at Georgia Institute of Technology, USA, for 20 years and a Principal Technical Staff Member with AT&T Labs – Research (previous Bell Labs) in New Jersey, USA, for five years. He made fundamental contributions to orthogonal frequency division multiplexing (OFDM) for wireless communications, established a framework on resource cooperation in wireless networks, and introduced deep learning

to communications. In these areas, he has published over 600 journal and conference papers in addition to over 40 granted patents. His publications have been cited over 63,000 times with an H-index of 116. He has been listed as a Highly Cited Researcher by Clarivate/Web of Science almost every year.

Dr. Geoffrey Ye Li was elected to IEEE Fellow and IET Fellow for his contributions to signal processing for wireless communications. He won 2024 IEEE Eric E. Sumner Award and several awards from IEEE Signal Processing, Vehicular Technology, and Communications Societies, including 2019 IEEE ComSoc Edwin Howard Armstrong Achievement Award.